

## A STUDY: MAMMOGRAM IMAGE SEGMENTATION AND CLASSIFICATION BASED ON ABC ALGORITHM AND ARTIFICIAL NEURAL NETWORKS

**Mamindla Ajay Kumar**

Department of CSE, GCET, Hyderabad.  
ajaymamindla@gmail.com

**Dr.Y.Ramadevi**

Department of CSE, CBIT, Hyderabad yrdcse.cbit@gmail.com

**Abstract:** Adopting Nature inspired optimization algorithms for image processing is on the double growth in the last decade. Artificial Bee Colony (ABC) approach is highly potential nature inspired optimization method mimics the bee's foraging behaviour. Moreover, the popularity of classification and artificial intelligence in different fields leads to the employment of ABC algorithm in upsurge. Notably, Early detection of breast cancer through digital mammogram images is essential as it is the one the most common cause of humankind cancer deaths. the aim of this comprehensive survey was to methodically analyse the effectiveness of using ABC algorithm in medical image enhancement, segmentation, and classification. This study firstly gives introduction of ABC algorithm and its basic mathematical and biological principles and operations respectively. Furthermore, this academic study summarizes the ABC applications on image segmentation techniques like Otsu and image classification approaches like Artificial Neural Networks (ANN). Finally, this far-reaching study come up with the challenges while exercising ABC algorithm in medical image processing, especially in mammogram images.

**Keywords:** Mammogram images, Artificial Bee Colony, Artificial Neural Networks.

### 1. Introduction

Breast cancer is the commonest cancer among Indian women. one woman out of two newly breast cancer diagnosed dies in India, mortality rate is 50percent that is 87,090 breast cancer patients out of 1,62,468 newly diagnosed were reportedly died in 2018 [1]. Despite of gender, breast cancer is the second leading cancer next to lung cancer [2]. Early detection of malicious tumour in breast cancer decreased the prevalence, morbidity of disease and reduces the human effort to cure the any kind of cancer disease. Practicing medical image processing techniques is the first step in detection of breast cancer by identifying the breast tissue abnormality [3]. In the process of detection, Radiologists come across with large amount of digital mammogram images, diagnosis accuracy is direct propositional to the radiologists' fatigue and workload. Therefore, it is fascinating to have a highly adequate automated mechanism to improve the efficiency, non-erroneous and accuracy in detection of abnormal tissue in breast and ensure the patient safety. Consequently, preventing needless treatment and decreasing the mortality rate of breast cancer patient by identifying benign tumour accurately. Nowadays, Medical image shows a crucial role in disease diagnosis, gauging treatment and pinpointing of abnormalities in different human body parts such as the breast, lungs, brain, stomach, and the eye. Scrutinizing the human body to diagnose, track or treat a disorder using specific methods

or techniques is referred as medical imaging [11]. Each medical imaging technology bids specific information about the province of the body to be investigated or treated, injury, illness, or the adequacy of the medical treatment in which part of the human body was investigated for abnormality detection and consideration [12].

Survivability is the primary thing for any living kind; therefore, every living should endeavour some mechanisms to survive. Fauna and flora presented many ways to secure or protect when their food sources insufficient [4]. These tactics could be used to enhance the performance of optimization algorithms. In last few decades, various NIOA have been proposed by imitating the biological or natural systems behaviour, but No Free Lunch (NFL) theorem states that no single NIOA based optimization algorithm is sufficient to solve all problems [5]. These algorithms are called as Meta Heuristic Nature inspired algorithms.

Features and merits of NIOA are as follows [6].

- i) As NIOA are global optimizers they have the potential for discovering the proper optimality
- ii) Nonlinear, multi modal problems could be dealt by gradient free NIOA or NIMA methods discontinuously.
- iii) These methods are suitable to deal with extensive dimension of problems without relative knowledge as they see all the problem as block box problems
- iv) Many of these algorithms are escape from Local minima due to the stochastic component. Unlike traditional methods, they use stochastic components like random walks and random numbers, as a consequence it obtains the different solutions despite of similar initial points.
- v) Being global optimizers, find the optimal solution with the balanced exploration (global search) and exploitation (local search) components. Exploration or global search implies the new region visit in a global search. Exploitation or local search implies the visiting previously visited points new neighbourhood regions. Furthermore, various strategies could be applied to increase the balance between exploitation and exploration.

As a result of above-mentioned feature, real world NP-Hard problems could be solvable with NIOAs efficiently and effectively. Time complexity of NP-Hard problems is  $O(2^N)$ . it grows exponentially when problem size 'n' increases [7].

Artificial neural network (ANN) is a connection of nodes inspired by human biological neurons and their interconnections. The ANN accuracy is depending on the parameters like number of input layers, output layers, hidden layers, number of neurons, weighted value, and activation function [8]. ANN is the most fortunate technique to find the link between input and outputs. ANN are the most popular method for load estimation. Usually, feed forward multilayer ANN have single input, single output and one or more than one hidden Layers [9]. Output Function calculation:

$$y_i = f_i \left( \sum_{j=1}^n w_{ij} x_j + \theta_i \right) \quad (1)$$

Where, weight of input J to neuron I is denoted with  $w_{ij}$ , Input neuron number is denoted with  $x_j$ ,  $\theta_i$  is the bias neuron I and  $f_i$  is the activation function.

Some frequently used activation records are shown in figure1 and figure2.

All the nodes in network are logically interconnected with associated weights. Every node has responsible to give some result with an activation or transfer function [10]. figure 3 gives a basic ANN structure.

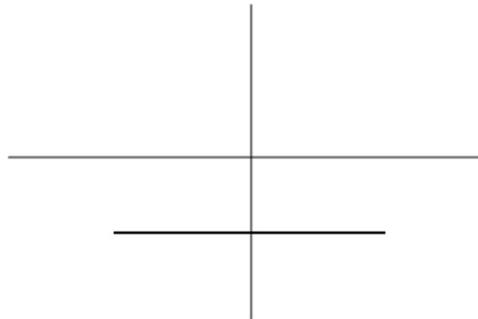


Figure 1. Linear activation function

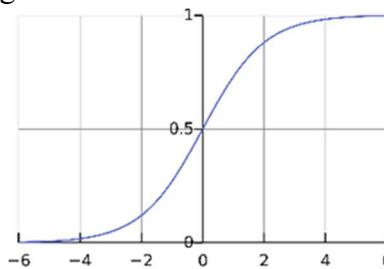


Figure 2. Sigmoid activation function

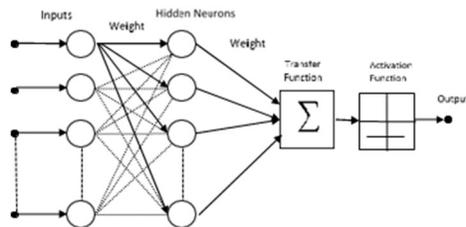


Figure 3: Neural Network Structure.

The primary objective of this paper is to review the ABC algorithm effectiveness in the medical image processing problems especially in mammogram-based breast cancer detection. Medical image processing is key factor for human health. Key variants of ABC algorithm and ANN applications are reported. Thus, an organized review is presented. Moreover, researchers showing interest in these areas could be benefited from and enquire when selecting an optimization technique. In this paper, a sufficient review is performed on the applications of the ABC algorithm with ANN in the field of mammogram image processing and image analysis. To achieve the aim, research publications published in the last few decades have been extensively reviewed

The remaining parts of this study are presented as follows: In section 2 The literature and discussion on recent NIOA publication is presented. In section 3 applications of ANN for optimized problems have been reported. Employment of NIOAs methods especially ABC is presented in section 4. In section 5, Effect of ABC, ANN in Mammogram image processing is reported using metrics. Lastly, Future research areas and conclusion are discussed in section 6.

## 2. Survey on NIOA

Instead of inappropriate traditional algorithms to solve the complex optimization problems alternate optimization methods called Nature-inspired optimization algorithms are proposed. These algorithms are developed by imitating or mimic the behaviour of individuals cooperate each other to find the solution to a certain problem using Artificial Intelligence (AI). This motivational nature behaviour-based algorithms increases the interest in several researchers to use nature-based optimization algorithm to solve the complex problems. In last couple of decades, lot many NIOAs are presented based on social, biological, and physical principles in addition to nature laws [13]. Therefore, this study classifies the NIOAs techniques into mainly four nonhomogeneous classes (i). Bio inspired methods (ii). Chemistry and Physics based methods (iii). Evolutionary computation (ECT) techniques (iv) other Nature based algorithms.

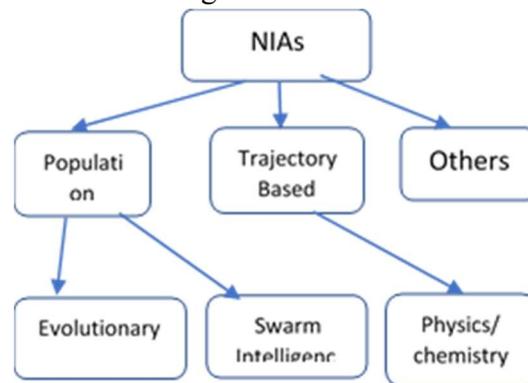


Figure4: NIOAs Classification [13]

Out of all developed NIAOs algorithms most of them are belongs to class Bio-Inspired Algorithms (BIA) optimization. In BIAs, novel exploration and exploitation operators are produced in abstraction form [15]. These BIAs are again classified into two broad sub-classes called Swarm based Intelligence (SI) algorithms and Non-Swarm Based Intelligence (NSI) algorithms. The primary principle behind this method of SI algorithms is mimic the collective behaviour of many agents to find the solution to complex problems. These algorithms are inspired by the collective behaviour of bees, ants and other animal or insects' swarms with their own environment and are also called as population-based algorithms. In this sub-class, algorithms performance improved by information sharing among agents, organised in own or self, concept of co-evolution and many, due to this many developed NIOAs are comes under this subclass. Some very familiar optimization algorithms developed in this sub-class are as follows, based on fishes and birds swarming behaviour Particle Swarm Optimization (PSO) algorithm is developed. Similarly, based on honeybees foraging behaviour Artificial Bee Colony (ABC) algorithm is developed. Additionally, based on fireflies flashing behaviour Firefly Algorithm (FA) is developed. Cooperative behaviour of ants leads to development of the Ant Colony Optimization (ACO) algorithm whereas based on brooding parasitism of few cuckoo species Cuckoo Search (CS) algorithm is conceived [13]. NSI algorithms are also belongs to bio-inspired class of methodologies however in their structure they use lone population but not to be considered as swarm. Unlike SI, NSI algorithms acknowledge the various communication mechanisms or biological metaphors inspired from natural phenomenality. Invasive

Weeds Optimization (IWO) algorithm, Flower Pollination (FPA) and Biogeography – Based Optimization (BBO) are some optimization mechanisms belongs to this class of NSI algorithms. Wherein, IWO imitate the weeds spreading strategy, FPA uses the concept of insects pollinating process and BBO uses the abstraction concept of migration, extinction, and speciation with an island migration concept to solve the optimization problems [16]. Also, in the year of 2021 nearly two dozen of NIOAs are presented, Table 1 list those algorithms.

ECT are based on Darwinian Theory of Evolution [DTE] principle in terms of computational [14]. these algorithms adopt different approaches like selection mechanisms, population, stop criteria, global mutation and local mutation operators and cost function. Based on their characteristics, ECTs are breached into different approaches like Differential Evolution [DE] based on search space weighted sum, Evolutionary Programming [EP] used Finite State Automata [FA] transition operations, Evolution Strategies [ES] based on mutation operators and recombination, Genetic Algorithms [GA] produces better future generations genetic material by using recombination of genes and Finally Genetic Programming [GP] are computer instruction set of genes [13]. Table 1 summarizes the EA with its features [41].

Table2: summary of Evolutionary algorithms

Algorithm	Representation	Search operators	Selection	Order of operators
<b>Genetic Algorithms</b>	Bit Numbers	Crossover Mutation	Stochastic	Selection, mutation Crossover,
<b>Evolutionary Strategy</b>	Real numbers	Mutation, Crossover (lately)	Deterministic	Selection, mutation Crossover,
<b>Genetic Programming</b>	Programs	Crossover	Stochastic	Selection, Crossover
<b>Evolutionary Programming</b>	Finite State Machines	Mutation	Stochastic	Mutation, Selection
<b>Differential Evolution</b>	Real numbers	Crossover, Mutation	Stochastic	Selection, mutation Crossover,

S.No	Algorithm Name	Author’s Name	Application Area
1	African Vulture’s Optimization Algorithm	Abdollahzadeh et. al. [17]	Mathematical and Engineering optimization problems, Feature selection and classification [17]
2	Artificial Gorilla Troops Optimizer	Abdollahzadeh et. al. [18]	Mathematical and Engineering optimization problems [18]
3	Aquila Optimizer	Abualigah et. al. [19]	Image classification and Engineering optimization problems [19]
4	Atomic Orbital Search	Azizi [20]	Mathematical and Engineering optimization problems and Feature selection [20]
5	Arithmetic Optimization Algorithm	Abualigah et. al. [21]	Mathematical and Engineering optimization problems and Feature selection [21]

6	Archimedes Optimization Algorithm	Hashim et. Al [22]	Mathematical and Engineering optimization problems, Feature selection and classification [22]
7	Aptenodytes Forsteri Optimization Algorithm	Yang et. al. [23]	Mathematical and Engineering optimization problems [23]
8	Bonobo Optimizer	Das et. al.[24]	Mathematical and Engineering optimization problems [24]
9	Battle Royale Optimization Algorithm	Farshi [25]	Artificial Neural Network and optimization problems [25]
10	Chameleon Swarm Algorithm	Braik [26]	Image Classification and optimization problems [26]
11	Child Drawing Development Optimization	Abdulhameed and Rashid [27]	Mathematical and Engineering optimization problems [27]
12	Cat and Mouse Based Optimizer	Dehghani et. al. [28]	Mathematical and Engineering optimization problems [28]
13	Dingo Optimizer	Bairwa et. al. [29]	Mathematical and Engineering optimization problems [29]
14	Flow Direction Algorithm	Karami et. Al [30]	Mathematical and Engineering optimization problems [30]
15	Hunger Games Search Optimizer	Nguyen et al. [31]	Artificial Neural Networks, Feature selection and image segmentation and [31]
16	Horse Herd Optimization Algorithm	MiarNaeimi et al. [32]	Feature selection and scheduling of nano grids [32]
17	Preaching-inspired Optimization Algorithm	Wei et.al.[33]	MLT based image segmentation and optimization problems [33]
18	Past Present Future: a new Human-based Algorithm	Naik and Satapathy [34]	Mathematical and Engineering optimization problems [34]
19	QANA: Quantum-based Avian Navigation Optimizer	Zamani et. Al [35]	Mathematical and Engineering optimization problems [35]
20	Remora Optimization Algorithm	Jia et. al. [36]	Mathematical and Engineering optimization problems [36]
21	Red Colobuses Monkey	Al-Kubaisy et al [37]	Design Optimization Problems, Engineering and Mathematical optimization problems [37]
22	Rat Swarm Optimizer	Nguyen et al. [38]	Image Classification and Deep Neural Networks [38]
23	Reptile Search Algorithm	Abualigah et. Al [39]	Neuro-Fuzzy Inference System and optimization Problems, Engineering and Mathematical optimization problems [39]
24	Tuna Swarm Optimization	Xie et. al. [40]	Mathematical and Engineering optimization problems [40]

Table2: List of Nature inspired optimization algorithms recently

Algorithms developed based on music, gravity, electrical-charges, and many more chemical and physical phenomenon comes under third category called physics and chemistry. In this Sub-Class of NIOAs, based on musicians Harmony Search (HS) algorithm is developed to produce new music compositions. Also, based on lone population attraction-repulsion forces Gravitational Search Algorithms (GSA) is proposed. Finally, by following electromagnetism principles in electrical relationships Electromagnetism-like optimization (EMO) technique is presented.

Other than the phenomena mentioned above, several others NIOAs are developed. This category of NIOAs includes social, emotional, and so on. These algorithms are overperformed BIAs and ECTs in some applications. Differential Search (DS), Fractal Search (FS), League Championship (LC) and Social Emotional Search (SES) have been developed by using non-biological principles. Figure 5

shows NIOAs general step by step procedure.

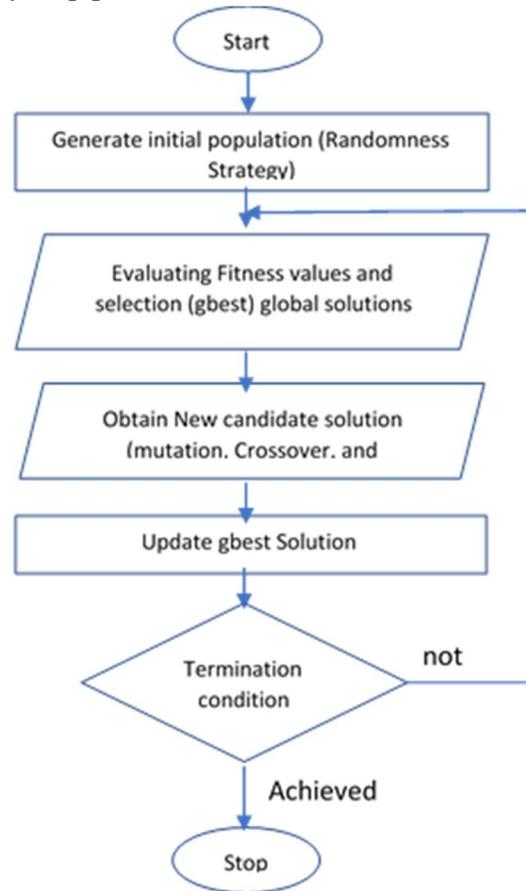


Figure5: NIOAs General Steps

Firstly, individuals' initial population is generated using randomness principle then gbest (global best) solution is selected based on evolution of individual fitness values. Later it obtain the new candidate solution based on operators like selection, crossover and mutation. Next, gbest solution is updated using offspring solution fitness values. Finally, process is exited if termination condition is achieved otherwise process repeats from generation of gbest solution.

### 3. Artificial Bee colony (ABC)

ABC is a novel bio-inspired heuristic optimization algorithm imitating the foraging behaviour of honeybee in a hive [42]. Since, it has less control parameter, better robustness, and easy model advantageous grabs the highest attention from researchers. Moreover, Among the NIOAs ABC is the most widely applied and preformed research to find the solution to the real-world engineering and optimization problems. Firstly, Karaboga presented Swarm intelligence-based ABC algorithm in 2005 [43]. The key feature in any SI class methodologies is Self-organization. Level of intelligence in the Swarm varies from swarm to swarm, depends on the self-organized interaction behaviour among agents. Self-organization in the swarm is achieved with the four characteristics: Positive feedback, Negative feedback, Fluctuation and Multiple interactions [44]. In Addition to the mentioned characteristics Swarm should satisfy the five principles given below to maintain the

required intelligence. First, proximity principle, swam could be able to perform time and space computations. Second, quality principle, should be able to give response to environment quality factors. Third, Diverse response principle, should not commit unreasonably narrow channels activities. Fourth, Stability principle, should not change its behavioural mode for environment fluctuations. Finally, adaptability principle, should be able to change mode of behaviour when needed [44]. In the ABC algorithm, bees have swarm-based intelligence in food sources quest and produce several qualities like sharing information through waggle dance, remembering the distance and direction of food source, and making decision to accept food source or not base on fitness value. This algorithm can be described with food source, dances, foraging behaviour, employed bees and unemployed bees [45]. Firstly, bee's search for the flower (food source) and gather information about difficulty in the nectar amount extraction, location of food source from the hive in terms of direction and distance and quantity of nectar. The quality of food source relies on richness, extraction difficulty level and proximity. Figure6 shows the different types of bees in the hive.

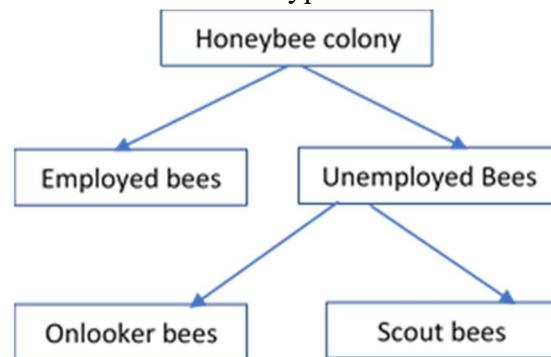


Figure6: ABC members

Unemployed and employed bees are also treated as working bees. Employed bee is responsible for searching of food sources (exploitation) and auditing the amount of nectar in flower. In the dancing area of hive, they share information with the unemployed bees about flower direction from the hive, distance to the hive, and its profitability. Further, unemployed bees are divided into scout bees and onlooker bees. Upon exhaustion of present food source, the scout bee is instructed to search for the next food source. The onlooker will establish the food source using share information by staying the nest. The ratio of employee, scout and onlooker bees in hive is 50:10:40.

The starting step in foraging process is searching flower source. Extraction of nectar amount may take half an hour to over two hours [46]. Information sharing among bees is through the different dance movements. the primary clue of new food source location is forager body's floral smell. Significance of communication between the bees is depends on the floral odor and dance language. Other bees touch the antenna of the employee bee during the dance to know the nectar taste in the source [46]. There are three different types of dance movements: waggle, round, and tremble. Direction and distance of food source is indicated by the waggle dance. Speed of dance proportional to the hive distance to the food source. When food source is near to the hive then it performs the round dance with out indicate the food source direction. Bee performs the tremble dance If the nectar amount and profitability are unknown and takes more time to nectar unloading. Figure7 shows the ABC algorithm.

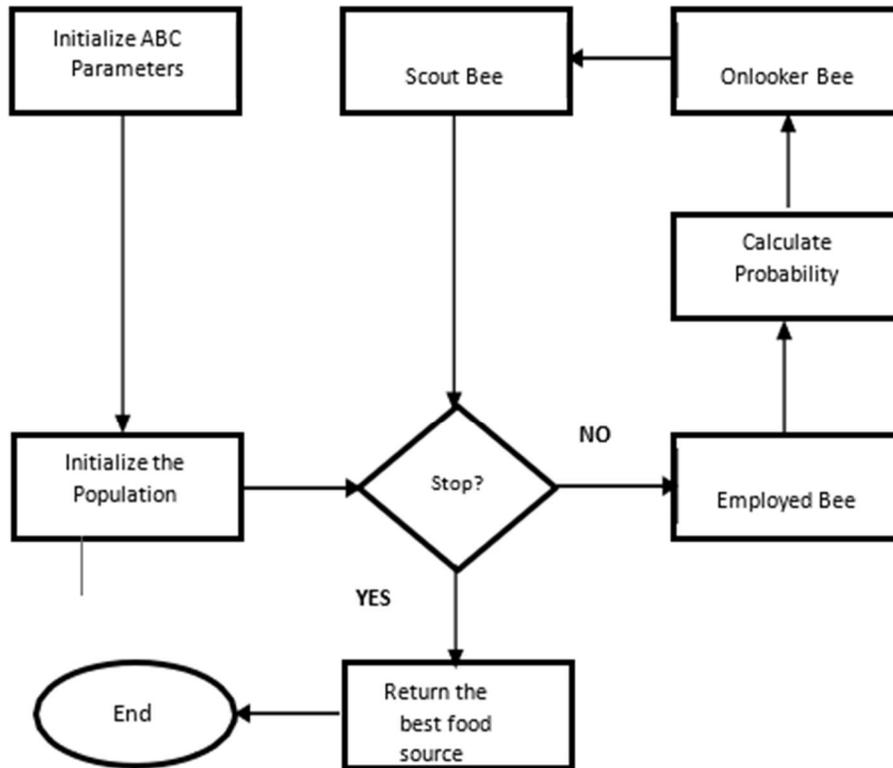


Figure 7: ABC algorithm used by [Ayadallah et al. \[47\]](#).

**ABC Algorithm**

ABC is a meta heuristic optimization algorithm based on population and honeybees foraging behaviour. ABC has less control parameters like solution number which indicates the number of food sources, maximum cycle number which signifies the number of maximum generations and limit represents the abandonment of food source. Selection and population are two processes linked with ABC. The process of explores the various area within the search space is called population and ensuring the previous experiences exploitation is called selection process. There are four phases in ABC 1. Initialization Phase 2. Employed bees’ phase 3. Onlooker bees’ phase 4. Scout bee phase. Fitness of food source is depending on the nectar amount and position of source. Each food source is assigned with a one employed bee thus number of employed bees and solutions (food sources) are equal. Firstly, in the initialization phase scout bees initiate the food sources (solutions) population and control parameters. In the second phase, employed bee phase, new solutions are found with in the nearby areas, evaluate the fitness values using roulette wheel (greedy selection) method if found the solution. In the hive dancing area employed bees and onlooker bees shares the information about the food sources. In the final scout bee phase, a scout bee abandons the food source (solution) exhausted maximum number of trials for improvement (called limit). Later the scout bee randomly start search for new food sources. Figure 8 shows the ABC algorithm pseudo-code [48].

**Pseudo-Code**

1. Defining the problem
2. ABC parameters initialization
3. Construct initially bee colony
4. Fitness value evolution for each bee
5. Repeat
6. N=0
7. Repeat
8. (K= one I neighbourhood solution)
9. ( $\Phi$ = random number in [-1,1])
10. Generate new solution by using equation two
11. Perform the greedy selection
12. Using Equation 3 evaluate probability values
13. For positions  $v_i$  and  $x_i$  assign new onlooker bees
14. Do for all onlooker bees
15. Perform greedy selection
16. If onlooker fitness probability is < employed fitness probability
17. Employed probability replaces the onlooker value
18. END
19. END
20. If BEST onlooker fitness < Best Fitness
21. Replace onlooker fitness with best fitness
22. END
23. N=N+1
24. Until N= Employed Bee Number
25. Using Equation 1 identify the abandoned solution
26. Replace employed solution with scout solution if scout solution is efficient than employed.
27. Repeat Until Max iteration

Figure8: ABC Pseudo Code [48]

Equations used in ABC are

$$x_j = x_j + rand(0,1)(x_j - x_j), \forall j = 1, 2, \dots, D \quad (1)$$

where  $x_{j_{NIN}}$  and  $x_{j_{NAS}}$  are bounds of  $x_i$  in  $j$ th direction

$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \quad (2)$$

where  $\Phi_{ij}(x_{ij}-x_{kj})$  is size of step and  $k \in \{1, 2, \dots, SN\}, j \in \{1, 2, \dots, D\}$ , are randomly chosen two indices and  $\Phi_{ij} \in [-1, 1]$ .

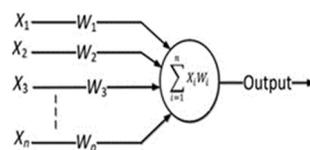
S.No	Year	Authors	Description
1	2005	Karaboga	ABC First Invention for optimization [43]
2	2006	Basturk B, Karaboga D	First conference Paper was published [49]

3	2007	Karaboga , Basturk	First journal article ,Performance evaluation of ABC [50]
4	2008	Karaboga , Basturk	Compared ABC algorithm with DE, PSO [51]
5	2009	Mala et al	Applied for Test suit optimization & compared with ACO [52]
6	2010	Akay et al	Modified versions of ABC, Real parameter optimization [53]
7	2011	Donfli et al	Modified ABC for numerical optimization problems [54]
8	2012	Mohammaed et al	Performance assessment: ABC & evolutionary algorithms [55]
9	2013	T.Liao et al	Applied ABC for continuous optimization [56]
10	2014	A.Draa et al	Used ABC in Image contrast enhancement [57]
11	2015	C.Ozturk et al	Improved Binary ABC in Soft computing [58]
12	2016	A. Bose, K. Mali	ABC applied in grey Image Segmentation [59]
13	2017	K. Balasubramani,et al	Brain image processing using hybridized ABC [60]
14	2018	B. Khomri et al	Presented elite guided multi-Objective ABC [61]
15	2019	D.Karaboga et al	Combinational ABC for Travelling Salesman Problem [62]
16	2020	Z. Hassan, et al	Integrated ABC with PCO algorithm for BLRP [63]

**Table3: some important studies on ABC in ascending of year.**

**4. Artificial Neural Networks**

In early decade of 1940s, A neurophysiologist, Warren McCulloch, and a young mathematician, Walter Pitts purported how neurons in human brain communicate and modelled this communication as an electric circuit network [73]. This model shown in figure9 is simple and is applied to classify the input into two classes.



**Figure89: McCullochs & Pitts Model [73].**

ANN become more popular due to its ability to find the solution and broad applicability in various distinct fields. In spite of ample ANN application in different disciplines, one of the most important aspects to concern is its time complexity and accuracy (Performance). The major limitation of Perceptron type of ANN model can be suitable for linear classification only, due to this limitation young researchers lost interest in ANN. But, after couple of decades, ANN is remodel or improved to solve the nonlinear problems, this led to regain interest in ANN among researchers [72]. Further, ANN is classified into two broad categories, one Feedforward Neural Networks [FNN] where ANN is upgraded to be able to solve the linear problems and complex problems with the added hidden layer and two Recurrent Neural Networks [RNN] where ANN is updated to give the output of present layer neurons to previous layer neurons. Figure11 show the basic FNN.

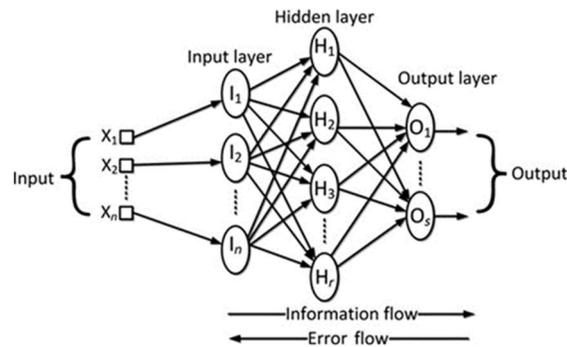


Figure11: Feed Forward Neural Network [72].

FNN consists a processing input layer, one or more than one hidden layers, for the result one output layer, connection between  $n$ th,  $(n-1)$ th and  $(n+1)$ th layers and some weight associated with every connection. In addition, there exist no cycle, that is neurons in same layer should not have connections. In RNN, the current layer neuron output is given as input next layer neuron output. RNN is widely applied in dynamic information processing like processing control, prediction of time series, and so on. Its popular types are feedback perceptron and Hopfield network. Energy minimization function and Generalized delta rule algorithms are mainly used for training RNN model [74]. Figure10 shows the structure of RNN.

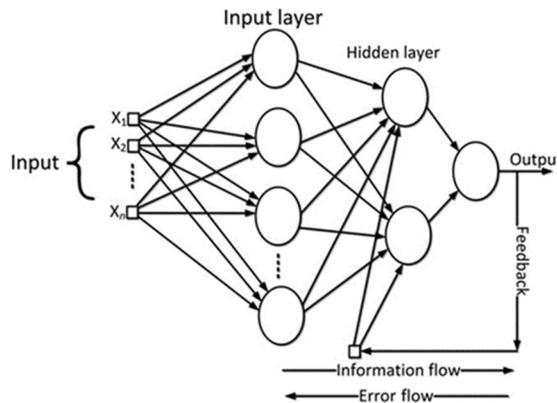


Figure10: Recurrent Neural Network [72].

FNN and RNN model layers can be optimized by adopting the NIOAs [71]. M. Carvalho, T.B. Ludermir presented a technique for classification problem and integration of more than one PSO algorithms execution and this study also discussed weighted and optimized architecture simultaneously [75]. The network is treated as it consists only one layer where in outer PSO algorithm optimizes FNN architecture and inner PSO performs weight optimization. Mendes, R et al, discussed about PSO paradigm for optimization of real valued functions based on group interaction social dynamics and discussed how PSO is used in the NN training [64]. Brajevic, I., Tuba, M discussed how a Firefly Algorithm [FA] is used in FNN training for classification, in the experiments, they used three most popular classification problems for performance evolution and result shows that FA outperforms the GA and some other NIOA in few aspects [67]. Milan Tuba, et all, presented how to train the FNN with the help of Cuckoo Search and modified version of Bat Algorithm, their model minimize the function errors to search for the neural network optimal synaptic weights by using three mentioned algorithms an [68]. Nabipour et al, newly presented an

optimization algorithms by integrating Salp Swarm algorithm (SSA), Biogeography-based optimization (BBO), Grasshopper Optimization Algorithm (GOA), and Particle Swarm Optimization (PSO) hybridized using ANN [69]. Cuthbert Shang Wui et al, implemented Machine Learning [ML] technique to make dynamic proxy models and described the well control rates optimization on couple of waterflooding cases with the help of ANN, PSO and grey wolf optimization (GWO) metaheuristic algorithms [70]. In [71], a comprehensive study on hybrid and multi objective optimizations is presented and compared many NIOAs with pros and cons in ANN training. Nimbark et al, discussed novel ABC technique is used to ANN architecture optimizations on transfer function and synaptic weight properties [76]. Karaboga et al, discussed how to overcome the drawbacks like computational complexity and local minima of traditional algorithms for model training, employed ABC for NN training with optimal weight set [65]. Same authors also applied ABC for ANN training to classify most popular data sets used in Machine learning applications and result says that ANN training can be done efficiently with ABC in pattern classification [66].

**5. Medical Image Processing using ABC**

**5.1 Image Enhancement using ABC**

Image visual appearance can be improved with the lot of diverse techniques called image enhancement. Image information perception, visual appearance and interpretability are improved by enhancement techniques. In lot of application like noise filtering and region of interest, image enhancement is pre-processing. It retains required image contents like edges and other image information. Image enhancement can be achieved with various techniques like edge enhancement, improving the region of interest, featuring and filtering enhancement and contrast enhancement [78]. Piyush J et al, proposed modified ABC method for quality enhancement of image, their major two contributions are ABC associated with direct constraints (DC) for exploring right direction and contrast estimation method based on retinal visual system [77]. The result of the paper also compared with the genetic algorithm and basic ABC are shown in Table4.

Test Image	Genetic Algorithm	Basic ABC	Their ABC
Image No:1	14.16	23.64	24.40
Image No:2	13.21	12.18	24.11
Image No:3	13.75	17.42	24.14
Image No:4	20.12	20.16	31.24
Image No:5	17.92	24.84	26.40

Table 4: Image PSNR value comparison [77].

Many researchers use the variants of ABC algorithm for smoothing and sharpening images and those medical image enhancement techniques are used explicit and assist doctors accurately to diagnose the disease, a crucial foundation for better interpretation and treatment [45]. Saadi et al, proposed a technique to deblurring images form blurred radiological images. ARMA (Aggressive moving Average) model is applied for images degraded nonlinearly. In ARMA method NN training is optimized by applying the ABC algorithm [79].

### 5.2 Image Segmentation using ABC

Image samples	I-ABC	ABC	GA	AFS
Sample A	0.8124	0.8243	0.4941	0.5024
Sample B	0.8274	0.8238	0.3249	0.3056
Sample C	0.8627	0.4861	0.2543	0.2147
Sample D	0.7641	0.7259	0.2413	0.3148

In image segmentation, raw image is partitioning into a group of homogeneous and non-overlapping partitions consists of similar kind of objects. This process is performed by considering the thresholding values or the adjacent regions discontinuity. Thresholding based methods further classified into bi-level where image is divided into two different classes and multi-level partitioning methods where image is categorized into more classes. Images segmentation is an essential step in all areas of image processing. Ankita Bose et al, presented FABC is a type of unsupervised classification image segmentation technique by combining the fuzzy c means (FCM) and ABC algorithms (FABC). In their study they compared FABC, EM, GA and PSO results over different synthetic, medical and texture kind of grayscale images and it is proven its efficiency in terms of qualitative and quantitative measures [80]. Dakshitha et al, proposed a combination of SEFF (Sinusoidal Evaluation of Fitness Function) and bi-level ABC with Tsallis entropy [81]. in their method global and local minima is searched in every iteration and initially maximize the Tsallis entropy to select the threshold value of histogram, followed by optimization using ABC. The result of proposed method is compared with BF, GA and PSO with respect to PSNR values. Bhandari et al, developed a MABC (modified ABC) technique used with Tsallis, Otsu and Kapur’s objective function. the MABC initialize the population using learning routine based on opposition and chaotic system [82]. In short time , MABC-Otsu’s, MABC-Kapur’s and MABC-Tsallis results outperforms the rest of the algorithms like ABC,GA and PSO in terms of optimal convergence rate, computational efficiency, few number of control parameters and accuracy. Hanbay et al, reported Synthetic aperture radar (SAR) multi class segmentation algorithm, in which feature extration is achieved with cooccurrence matrix and Neutrosophic set (NS) and Improve ABC (I-ABC) is applied for optimization [83]. The results of I-ABC in comparison with SI methods are given in Table5.

Table5 : I-ABC and SI comparison results [83].

Medical image segmentation is mainly used to diagnoses the decease , tumors detection, tumor volume, disease measurement anatomical region classification in human body parts. Decision making process in disease diagnosis is most successful with medical image segmentation. In recent years , using ABC in medical image segmentation results efficient techniques to diagnosis. ABC is applied for segmentation in body parts such as: MRI images of Images[84], MR or CT images of Liver [85], blood vessel segmentation [86], lung cancer detection [87] are the few parts covered to reduce the human mistakes and improve the segmentation results. Mostafa et al, used image clustering with ABC for MR or CT images, this approach calculate the clusters centroid in the images using ABC optimization algorithm [85]. Fu et al, introduced a method for White Blood Cell (WBC) images processing. This algorithm includes different steps like cell segmentation, feature extraction,

WBC classification in color blood images cell [88]. Experimental results of this algorithm in comparison with PSO and DE are shown

Algorithm		Image A	Image B	Image C
PSO	Mean	-0.922	-0.978	-0.754
	Std	0.091	0.023	0.226
	Time	157.979	11.312	14.733
DE	Mean	-0.919	-0.960	-0.754
	Std	0.085	0.023	0.080
	Time	160.775	12.148	14.941
ABC	Mean	-0.0923	-0.972	-0.760
	Std	0.0641	0.018	0.108
	Time	156.549	11.388	14.63

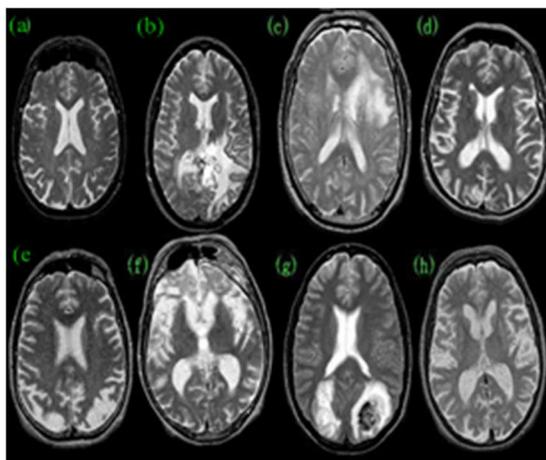
in the Table6.

Table 6: WBC detection results comparison [88]

### 5.3 Image Classification using ABC

In CAD (Computer Aided Diagnosis), classification for analysing medical images is crucial step. The process of medical image Classification makes the doctors better in disease diagnosis, to keep track of disease and human organ effected area. R Sharma et al, developed a ABC trained BPNN technique to detect Leukemia, this technique firstly perform PCA dimensionality reduction on Leukemia dataset then Optimal feature set is obtained by using ABC and finally BPNN is trained with ABC for classification [89]. The proposed ABC-BPNN results shows 98.72% average accuracy in classification. Y. Zhang et al developed a SCABC-FNN novel classifier to find the abnormality in brain MRI images [90]. The SCABC-FNN method shows almost 100% accuracy in image datasets of weighted MRI images. Proposed method performed on a total of 66 images with 6-fold cross validation, out of all 8 images are detected as abnormal. Figure12 shows a sample of 8 images.

Figure 12. Sample of brain MRIs: (a) Normal brain; (b) glioma;(c) meningioma; (d) Alzheimer’s disease; (e) Alzheimer’s disease with visual agnosia; (f) Pick’s disease; (g) sarcoma; (h) Huntington’s disease [90].



After classification of 66 images using 6-fold stratified validation the SCABC is showed the following confusion matrix. Moreover, presented Approach outperforms other approaches like

DWT+SOM, DWT+SVM etc with almost accuracy of 100% shown in figure 12.

18 27.3%	0 0.0%	100% 0.0%
0 0.0%	48 72.7%	100% 0.0%
100% 0.0%	100% 0.0%	100% 0.0%

Figure 12 . Confusion matrix of the SCABC method [43].

M. Sornam et al, presented a hybrid method by combining back – propagation NN with logit base ABC [LB-ABC] to reduce the diagnostic errors and dental caries misclassification [91]. Figure13 shows the dental caries image segmentation process.

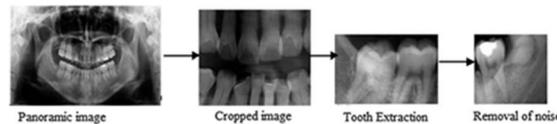


Figure13. X-ray image segmentation process [91]

Table7 shows the dental X-ray images GLCM feature extraction.

Algorithm	Colon		Leukemia	
	Accuracy (%)	Number of genes	Accuracy (%)	Number of genes
ABC-SVM	95.61	20	95.83	20
GA-SVM	95.61	83	93.1	51
PSO-SVM	93.55	78	95.83	53
MPSO-SVM	85.48	20	94.44	23
PSO-SVMA	87.01	2000	93.06	7129

Table7. GLCM feature extraction for dental X-ray images [91].

Hala Alshamlan et al, proposed a ABC-SVM optimized classification technique to deal with high dimensionality microarray data, Colon and Leukemia dataset are considered for ABC-SVM classification and it outperforms the other approaches and results are shown in table8 [92].

**Table 8: SVM and NIOA performance comparison.**

Network	Hidden nodes	Epochs	Approximate learning rate for the BPNN	Accuracy (%)	MSE					
BPNN classifier	40	500	0.1	89.76	2.134					
	40	1000	0.2	91.85	1.982					
	40	1500	0.3	96.50	0.126					
Proposed LB-ABC with BPNN	Hidden nodes	Epochs	Learning rate generated by the LB-ABC	Accuracy (%)	MSE					
						20	2000	0.04	98.7	0.108
						15	4000	0.05	99.16	0.008
						10	5000	0.06	99.1667	0.00339

**6. Discussion and Future Scope**

This study focussed on the medical image processing especially using NIOAs, particularly using ABC algorithm. feature selection, segmentation and classifications steps of image processing is mainly focussed for this study. ABC has been employed in several application in many fields. Training NN is one of the most interesting fields that is finding weight set in optimistic way. In machine learning society, ABC is widely used to feedforward NN for classification due to its few control parameters and strong capabilities for global search. ABC convergence rate and ability of finding global optimum level is medium but it is powerful against population initialization.

Unlike ABC algorithm ANN techniques easily fell into local minimum, therefore applying optimizing algorithm often solving these problems of ANN. Feature selection is the plays a crucial role in classification. Despite, considerable research done in the field of classification still young researchers come across with challenges like accuracy, stability, scalability, and computational cost. In recent times, there is scope for researchers to propose efficient and optimized techniques for choosing classification features. A standardized classification technique needs consideration for development, or a powerful classifier is still required to implement on various data sets to achieve needed results.

This study also suggests to use of hybridized methods to improve the classification performance and improve the accuracy of classifies using NIOA like ABC. In Neuro-Optimization technique, Optimization techniques like ABC trains the ANN by changing networks weights and based. It is clear from the literature study; ABC is highly powerful and more flexible optimization method that can be applied in many fields from hyperspectral images to medical images. Compare to other SI algorithms it has less control parameters, easy to apply. Moreover, it is hybridizing with other SI algorithms easily and deals with stochastic objective function. However, slow convergence rate due its stochastic nature and tripped into local optima while solving complex problems are main drawbacks of ABC algorithms. Besides, it inherits the drawbacks of SI like slowing speed of population size, changing objective function and problem based.

It has been noticed that many modifications and improvements such as scout bee addition,

improvement global and local search ability, plug in for fitness functions are added to the ABC basic structure. Also, finding solution for Slow convergence problem and slowdown examination in iterations process are the open-ended problems.

Medical images are different from other images like nature, satellite images etc. Today, ANN techniques come across with multi-class problems due to medical imaging methods. It is observed that there are very less methods proposed for Mammogram classification using NIOAs based ANN. Besides, NIOAs based ANN is used in classification of medical images like MRI images, dental caries, lung cancer images and so on. This study also suggests that there is still requirement of efficient technique to mammogram image processing such as optimized image enhancement, optimized image segmentation, optimized image clustering and optimized image classification separately.

## **7. Conclusion**

The comparative study always strengthens the knowledge in the required field. This study reviews the basic concept, variants, application of ABC optimization to solve the real time complex problems. Apart from that, how ABC is used to train ANN classifier for medical image processing. Besides, it pointed out that mammogram imaging based on metaheuristics optimized algorithm is hot area for researchers in recent times. A comparative study of the NIOAs, ABC improvements and ABC based ANN classifiers are proposed and illustrated the investigators implementations on those. An evaluation of ABC and ANN methods have been reported with conclusion.

ANN is most popular machine learning method applied in the biomedical imaging. Earlier to ANN, several parameters and architecture should be mentioned first. So, optimizing the ANN architecture and its several parameters become more important. This paper summarizes optimization techniques and their use in ANN parameter optimization. It also depicts the hybridized ABC and NN classifiers experimental results on the same dataset. From the Literature, it is easily observed that the expectations of ABC growth in medical imaging field have exceeded. Overall, it was noticed that hybridizing ABC and ANN improves the performance in medical imaging. Finally, it is suggested to have a ABC optimized ANN classifier for Breast cancer detection As there is less research on metaheuristic-based mammogram image segmentation and classification.

## **References**

1. <http://cancerindia.org.in/breast-cancer/>
2. C. Li, Breast Cancer Epidemiology. New York, NY, USA: Springer, 2010.
3. S. Nass, I. Henderson, and J. Lashof, "Mammography and beyond:Developing technologies for the early detection of breast cancer," Nat.Cancer Policy Board, Inst. Med., Commission Life Stud, Nat. Res. Council, Washington, DC, USA, 2001.
4. Khairuzzaman AKM, Chaudhury S (2020) Modified moth-fame optimization algorithm-based multilevel minimum cross entropy thresholding for image segmentation. Int J Swarm Intell Res IJSIR 11(4):123–139
5. Dhal KG, Ray S, Das A, Das S (2019a) A survey on nature-inspired optimization algorithms and their application in image enhancement domain. Arch Comput Methods Eng 26(5):1607–1638
6. Yang XS (2018) Mathematical analysis of nature-inspired algorithms. In: Yang XS (ed) 3586

- Nature-inspired algorithms and applied optimization. Springer, Cham, pp 1–25
7. Garey M, Johnson D (1979) Computers and intractability: a guide to the theory of NP completeness. W.H. Freeman, New York
  8. H. Chiroma et al., "Progress on Artificial Neural Networks for Big Data Analytics: A Survey," in IEEE Access, vol. 7, pp. 70535-70551, 2019, doi: 10.1109/ACCESS.2018.2880694.
  9. Guo, Z., et al., multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy, 2012. 37(1): p. 241-249.
  10. Çevik, Hasan & Harmancı, Hüseyin & Cunkas, Mehmet. (2016). Short-term Load Forecasting based on ABC and ANN for Smart Grids. International Journal of Intelligent Systems and Applications in Engineering. 4. 38-38. 10.18201/ijisae.266014.
  11. Essam H. Houssein, Marwa M. Emam, Abdelmgeid A. Ali, Ponnuthurai Nagaratnam Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, Expert Systems with Applications, Volume 167, 2021, 114161, ISSN 0957-4174,
  12. Ashour, A. S., Dey, N., & Mohamed, W. S. (2016). Abdominal imaging in clinical applications: Computer aided diagnosis approaches. In Medical imaging in clinical applications (pp. 3–17). Springer
  13. K. G. Dhal, A. Das, S. Ray, J. Gálvez, and S. Das, Nature-Inspired Optimization Algorithms and Their Application in Multi-Thresholding Image Segmentation, vol. 27, no. 3. Springer Netherlands, 2020.
  14. Fister Jr I (2013) A comprehensive review of bat algorithms and their hybridization. asters thesis. University of Maribor, Slovenia
  15. Fister Jr I, Yang XS, Fister I, Brest J, Fister D (2013) A brief review of nature-inspired algorithms for optimization. arXiv preprint arXiv:1307.4186
  16. Yang XS (2012) Flower pollination algorithm for global optimization. In: International conference on unconventional computing and natural computation. Springer, Berlin, pp 240–249
  17. Abdollahzadeh B, Gharehchopogh FS, Mirjalili S (2021) African vultures optimization algorithm: a new nature-inspired metaheuristic algorithm for global optimization problems. Comput Ind Eng 158:107408
  18. Abdollahzadeh B, Soleimani Gharehchopogh F, Mirjalili S (2021) Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems. Int J Intell Syst 36(10):5887–5958
  19. Abualigah L, Yousri D, Abd Elaziz M, Ewees AA, Al-qaness MA, Gandomi AH (2021) Aquila optimizer: a novel meta-heuristic optimization algorithm. Comput Ind Eng 157:107250
  20. Azizi M (2021) Atomic orbital search: a novel metaheuristic algorithm. Appl Math Model 93:657–683
  21. Abualigah L, Diabat A, Mirjalili S, Abd Elaziz M, Gandomi AH (2021b) The arithmetic

- optimization algorithm. *Comput Meth- ods Appl Mech Eng* 376:113609
22. Hashim FA, Hussain K, Houssein EH, Mabrouk MS, Al-Atabany W (2021) Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. *Appl Intell* 51(3):1531–1551
  23. Yang Z, Deng L, Wang Y, Liu J (2021a) Aptenodytes forsteri optimization: algorithm and applications. *Knowl Based Syst* 232:107483
  24. Das AK, Pratihari DK (2021) Bonobo optimizer (BO): an intelligent heuristic with self-adjusting parameters over continuous spaces and its applications to engineering problems. *Appl Intell* 1–33. <https://doi.org/10.1007/s10489-021-02444-w>
  25. RahkarFarshi T (2021) Battle royale optimization algorithm. *Neural Comput Appl* 33:1139–1157
  26. Braik MS (2021) Chameleon swarm algorithm: a bio-inspired optimizer for solving engineering design problems. *Expert Syst Appl* 174:114685
  27. Abdulhameed S, Rashid TA (2021) Child drawing development optimization algorithm based on child's cognitive development. *Arab J Sci Eng*:1–15. <https://doi.org/10.1007/s13369-021-05928-6>
  28. Dehghani M, Hubálovský Š, Trojovský P (2021) Cat and mouse based optimizer: a new nature-inspired optimization algorithm. *Sensors* 21(15):5214
  29. Bairwa AK, Joshi S, Singh D (2021) Dingo optimizer: a nature-inspired metaheuristic approach for engineering problems. *Math Probl Eng* 2021:1–12
  30. Karami H, Anaraki MV, Farzin S, Mirjalili S (2021) Flow Direction Algorithm (FDA): a novel optimization approach for solving optimization problems. *Comput Ind Eng* 156:107224
  31. Nguyen H, Bui XN (2021) A novel Hunger Game search optimization-based artificial neural network for predicting ground vibration intensity induced by mine blasting. *Nat Resour Res* 30(5):3865–3880
  32. MirNaeimi F, Azizyan G, Rashki M (2021) Horse herd optimization algorithm: a nature-inspired algorithm for high-dimensional optimization problems. *Knowl Based Syst* 213:106711
  33. Wei D, Wang Z, Si L, Tan C (2021) Preaching-inspired swarm intelligence algorithm and its applications. *Knowl Based Syst* 211:106552
  34. Naik A, Satapathy SC (2021) Past present future: a new human-based algorithm for stochastic optimization. *Soft Comput* 25(20):12915–12976
  35. Zamani H, Nadimi-Shahraki MH, Gandomi AH (2021) QANA: quantum-based avian navigation optimizer algorithm. *Eng Appl Artif Intell* 104:104314
  36. Jia H, Peng X, Lang C (2021) Remora optimization algorithm. *Expert Syst Appl* 185:115665
  37. Al-Kubaisy WJ, Yousif M, Al-Khateeb B, Mahmood M, Le DN (2021) The red colobus monkey: a new nature-inspired metaheuristic optimization algorithm. *Int J Comput Intell Syst* 14(1):1108–1118
  38. Dhiman G, Garg M, Nagar A, Kumar V, Dehghani M (2021) A novel algorithm for global optimization: rat swarm optimizer. *J Ambient Intell Humaniz Comput* 12(8):8457–8482
  39. Abualigah L, Abd Elaziz M, Sumari P, Geem ZW, Gandomi AH (2021c) Reptile Search

- Algorithm (RSA): a nature-inspired meta-heuristic optimizer. *Expert Syst Appl* 191:116158
40. Xie L, Han T, Zhou H, Zhang ZR, Han B, Tang A (2021) Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization. *Comput Intell Neurosci*. <https://doi.org/10.1155/2021/9210050>
  41. Kulkarni AJ, Krishnasamy G, Abraham A (2017) Cohort intelligence: a socio-inspired optimization method. Springer, Heidelberg. [https://doi.org/10.1007/978-3-319-44254-9\\_1](https://doi.org/10.1007/978-3-319-44254-9_1)
  42. Liu, Ying, Lianbo Ma, and Guangming Yang. "A survey of artificial bee colony algorithm." 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2017.
  43. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report. Computer Engineering Department, Engineering Faculty, Erciyes University
  44. Karaboga, Dervis, et al. "A comprehensive survey: artificial bee colony (ABC) algorithm and applications." *Artificial Intelligence Review* 42.1 (2014): 21-57.
  45. Öztürk, Şaban, Rehan Ahmad, and Nadeem Akhtar. "Variants of Artificial Bee Colony algorithm and its applications in medical image processing." *Applied soft computing* 97 (2020): 106799.
  46. J.C. Bansal, H. Sharma, S.S. Jadon, Artificial bee colony algorithm: a survey, in: *International Journal of Advanced Intelligence Paradigms*, 5 (2013). DOI: 10.1504/IJAIP.2013.054681
  47. M.A. Awadallah, M.A. Al-Betar, A.L.a. Bolaji, E.M. Alsukhni, H. Al-Zoubi, Natural selection methods for artificial bee colony with new versions of onlooker bee, in: *Soft Computing*, 23 (2018) 6455-6494. <https://doi.org/10.1007/s00500-018-3299-2>.
  48. B. Alatas, Chaotic bee colony algorithms for global numerical optimization, in: *Expert Systems with Applications*, 37 (2010) 5682-5687. <https://doi.org/10.1016/j.eswa.2010.02.042>
  49. Basturk B, Karaboga D (2006) An artificial bee colony (abc) algorithm for numeric function optimization. In: *IEEE swarm intelligence symposium 2006*, Indianapolis, IN, USA
  50. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *J Glob Optim* 39:459–471
  51. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (abc) algorithm. *Appl Soft Comput* 8(1):687–697
  52. Mala DJ, Kamalpriya M, Shobana R, Mohan V (2009) A non-pheromone based intelligent swarm optimization technique in software test suite optimization. In: *IAMA: 2009 international conference on intelligent agent and multi-agent systems*, IEEE Madras Section; IEEE Computer Society, Madras Chapter; Computer Society of India Div II; Council of Science & Industrial Research; Govt India, Department of Information Technology, pp 188–192
  53. Akay B, Karaboga D (2010) A modified artificial bee colony algorithm for real-parameter optimization. *Inf Sci*. doi:10.1016/j.ins.2010.07.015
  54. Dongli Z, Xinping G, Yinggan T, Yong T (2011) Modified artificial bee colony algorithms for numerical optimization. In: *2011 3rd international workshop on intelligent systems and*

- applications (ISA), pp 1–4
55. Mohammed , El-Abd (2012) Performance assessment of foraging algorithms vs. evolutionary algorithms. *Inf Sci* 182(1):243–263
  56. T. Liao, D. Aydm, T. Stützle, Artificial bee colonies for continuous optimization: Experimental analysis and improvements, in: *Swarm Intelligence*, 7 (2013) 327-356. <https://doi.org/10.1007/s11721-013-0088-5>.
  57. A. Draa, A. Bouaziz, An artificial bee colony algorithm for image contrast enhancement, in: *Swarm and Evolutionary Computation*, 16 (2014) 69-84. <https://doi.org/10.1016/j.swevo.2014.01.003>
  58. J. C. Ozturk, E. Hancer, D. Karaboga, Dynamic clustering with improved binary artificial bee colony algorithm, in: *Applied Soft Computing*, 28 (2015) 69-80. <https://doi.org/10.1016/j.asoc.2014.11.040>.
  59. A. Bose, K. Mali, Fuzzy-based artificial bee colony optimization for gray image segmentation, in: *Signal, Image and Video Processing*, 10 (2016) 1089-1096. <https://doi.org/10.1007/s11760-016-0863-z>
  60. K. Balasubramani, P. Madhura, R.V. Kulkarni, P. Pavithra, Hybridized approach of artificial bee colony algorithm for detection of suspicious brain pattern using magnetic resonance images, in: *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017, pp. 451-455. DOI: 10.1109/ICPCSI.2017.8392336
  61. B. Khomri, A. Christodoulidis, L. Djerou, M.C. Babahenini, F. Cheriet, Retinal blood vessel segmentation using the elite-guided multi-objective artificial bee colony algorithm, in: *IET Image Processing*, 12 (2018) 2163-2171. DOI: 10.1049/iet-ipr.2018.5425
  62. D. Karaboga, B. Gorkemli, Solving Traveling Salesman Problem by Using Combinatorial Artificial Bee Colony Algorithms, in: *International Journal on Artificial Intelligence Tools*, 28(2019). <https://doi.org/10.1142/S0218213019500040>.
  63. Z. Hassan, An Integration Based Optimization Approach (ABC and PSO) for Parameter Estimation in BLRP Model for Disaggregating Daily Rainfall, in: *Pertanika Journal of Science & Technology*, 2020 385-402.
  64. Mendes, R., Cortez, P., Rocha, M., Neves, J.: Particle swarm for feedforward neural network training. In: *Proceedings of the International Joint Conference on Neural Networks 2*, 1895–1899 (2002)
  65. Karaboga, D., Akay, B., Ozturk, C.: Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks. *Lecture Notes in Computer Science: Modeling Decisions for Artificial Intelligence* 4617, 318–329 (2007)
  66. Karaboga, D., Ozturk, C.: Neural networks training by artificial bee colony algorithm on pattern classification. *Neural Netw. World* 19(3), 279–292 (2009)
  67. Brajevic, I., Tuba, M.: Training feed-forward neural networks using firefly algorithm. In: *Proceedings of the 12th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED '13)*, pp. 156–161 (2013)
  68. Milan Tuba, Adis Alihodzic, and Nebojsa Bacanin. Cuckoo Search and Bat Algorithm 3590

- Applied to Training Feed-Forward Neural Networks, pages 139–162. Springer International Publishing, 2015 (Accessed June 20th, 2019)
69. Nabipour, Narjes, et al. "Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks." *IEEE Access* 8 (2020): 15210-15222.
  70. Ng, Cuthbert Shang Wui, Ashkan Jahanbani Ghahfarokhi, and Menad Nait Amar. "Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization." *Journal of Petroleum Exploration and Production Technology* 11.7 (2021): 3103-3127.
  71. Hemeida, Ashraf Mohamed, et al. "Nature-inspired algorithms for feed-forward neural network classifiers: a survey of one decade of research." *Ain Shams Engineering Journal* 11.3 (2020): 659-675.
  72. Gupta, Tarun Kumar, and Khalid Raza. "Optimization of ANN architecture: a review on nature-inspired techniques." *Machine learning in bio-signal analysis and diagnostic imaging* (2019): 159-182.
  73. McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5.4 (1943): 115-133.
  74. D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back propagating errors, *Nature* 323 (6088) (1986) 533–536.
  75. M. Carvalho, T.B. Ludermir, in: Particle swarm optimization of neural network architectures and weights, 7th Int. Conf. Hybrid Intell. Syst. (HIS), 2007, 2007, pp. 336–339.
  76. H. Nimbark, R. Sukhadia, P.P. Kotak, Optimizing architectural properties of Artificial Neural Network using proposed Artificial Bee Colony algorithm, in: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014, pp. 1285–1289.
  77. P. Joshi, S. Prakash, An efficient technique for image contrast enhancement using artificial bee colony, in: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015), 2015, pp. 1-6. DOI:10.1109/ISBA.2015.7126363
  78. Akhtar, N., Choubey, N. S., & Ragavendran, U. (2019). Investigation of Non-natural Information from Remote Sensing Images: A Case Study Approach. In *Computational Intelligence and Sustainable Systems* (pp. 165-199). Springer, Cham
  79. S. Saadi, A. Guessoum, M. Bettayeb, ABC optimized neural network model for image deblurring with its FPGA implementation, in: *Microprocessors and Microsystems*, 37 (2013) 52-64. <https://doi.org/10.1016/j.micpro.2012.09.013>
  80. A. Bose, K. Mali, Fuzzy-based artificial bee colony optimization for gray image segmentation, in: *Signal, Image and Video Processing*, 10 (2016) 1089-1096. <https://doi.org/10.1007/s11760-016-0863-z>
  81. B.A. Dakshitha, V. Deekshitha, K. Manikantan, A novel Bi-level Artificial Bee Colony algorithm and its application to image segmentation, in: 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCI), 2015, pp. 1-7. DOI: 3591

- 10.1109/ICCIC.2015.7435656
82. A.K. Bhandari, A. Kumar, G.K. Singh, Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions, in: *Expert Systems with Applications*, 42 (2015) 1573-1601. <https://doi.org/10.1016/j.eswa.2014.09.049>
  83. K. Hanbay, M.F. Talu, Segmentation of SAR images using improved artificial bee colony algorithm and neutrosophic set, in: *Applied Soft Computing*, 21 (2014) 433-443. <https://doi.org/10.1016/j.asoc.2014.04.008>
  84. M. Shokouhifar, G. S. Abkenar, An Artificial Bee Colony Optimization for MRI Fuzzy Segmentation of Brain Tissue, in: *Proceedings of International Conference on Management and Artificial Intelligence IPEDR*, 2011 6-10. DOI: 10.13140/2.1.2866.1761
  85. A. Mostafa, A. Fouad, M.A. Elfattah, A.E. Hassanien, H. Hefny, S.Y. Zhu, G. Schaefer, CT Liver Segmentation Using Artificial Bee Colony Optimisation, in: *Procedia Computer Science*, 60 (2015) 1622-1630. doi: 10.1016/j.procs.2015.08.272
  86. A.E. Hassanien, E. Emary, H.M. Zawbaa, Retinal blood vessel localization approach based on bee colony swarm optimization, fuzzy c-means and pattern search, in: *Journal of Visual Communication and Image Representation*, 31 (2015) 186-196. <https://doi.org/10.1016/j.jvcir.2015.06.019>
  87. S. Perumal, T. Velmurugan, Lung Cancer Detection and Classification on CT scan images using enhanced artificial bee colony, in: *Proceeding of International Journal of Engineering & Technology*, (2018) 74-79. DOI: 10.14419/ijet.v7i2.26.12538
  88. Z. Fu, Y. Liu, H. Hu, D. Wu, H. Gao, An efficient method of white blood cells detection based on artificial bee colony algorithm, in: *2017 29th Chinese Control And Decision Conference (CCDC)*, 2017, pp. 3266-3271. DOI: 10.1109/CCDC.2017.7979070
  89. R. Sharma, R. Kumar, A Novel Approach for the Classification of Leukemia Using Artificial Bee Colony Optimization Technique and Back-Propagation Neural Networks, in: *Proceedings of 2nd International Conference on Communication, Computing and Networking*, 2019, pp. 685-694. [https://doi.org/10.1007/978-981-13-1217-5\\_68](https://doi.org/10.1007/978-981-13-1217-5_68)
  90. Y. Zhang, L. Wu, S. Wang, Magnetic Resonance Brain Image Classification by an Improved Artificial Bee Colony Algorithm, in: *Progress In Electromagnetics Research*, 116 (2011) 65-79. <https://doi.org/10.2528/PIER11031709>
  91. M. Sornam, M. Prabhakaran, Logit-Based Artificial Bee Colony Optimization (LB-ABC) Approach for Dental Caries Classification Using a Back Propagation Neural Network, in: *Integrated Intelligent Computing, Communication and Security*, 2019, pp. 79-91. [https://doi.org/10.1007/978-981-10-8797-4\\_9](https://doi.org/10.1007/978-981-10-8797-4_9)
  92. H. Alshamlan, G. Badr, Y. Alohal, Microarray Gene Selection and Cancer Classification Method Using Artificial Bee Colony and SVM Algorithms (ABC-SVM), in: *Proceedings of the International Conference on Data Engineering 2015 (DaEng-2015)*, 2019, pp. 575-584. [https://doi.org/10.1007/978-981-13-1799-6\\_59](https://doi.org/10.1007/978-981-13-1799-6_59)

## ORIGINAL ARTICLE

 OPEN ACCESS

Received: 22-03-2022

Accepted: 04-08-2022

Published: 26-08-2022

**Citation:** Reddy MVK, Srinivas PVS, Mohan MC (2022) A Novel Trust Adaptability Approach for Secure Data Transmissions using Enhanced Collaborative Nodes

Trustworthiness in Mobile Ad-hoc Networks. Indian Journal of Science and Technology 15(33): 1613-1623.

[https://doi.org/](https://doi.org/10.17485/IJST/v15i33.663)

[10.17485/IJST/v15i33.663](https://doi.org/10.17485/IJST/v15i33.663)

\* **Corresponding author.**

[krishnareddy\\_cse@cbit.ac.in](mailto:krishnareddy_cse@cbit.ac.in)

**Funding:** None

**Competing Interests:** None

**Copyright:** © 2022 Reddy et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.indjst.org/))

**ISSN**

Print: 0974-6846

Electronic: 0974-5645

# A Novel Trust Adaptability Approach for Secure Data Transmissions using Enhanced Collaborative Nodes Trustworthiness in Mobile Ad-hoc Networks

M Venkata Krishna Reddy<sup>1,2\*</sup>, P V S Srinivas<sup>3</sup>, M Chandra Mohan<sup>4</sup>

**1** Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana, 500075, India

**2** Assistant Professor, Department of CSE, Chaitanya Bharathi Institute of Technology(A), Hyderabad, Telangana, 500075, India

**3** Professor, Department of Computer Science and Engineering, Vignana Bharathi Institute of Technology(A), Hyderabad, Telangana, 500075, India

**4** Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana, 500075, India

## Abstract

**Objectives:** To find an efficient security routing model based on trust adaptability by considering various transmission parameters that influence the node's behavior. **Methods:** Enhanced Collaborative Trust Based Approach (ECTBA) was applied to isolate the malicious nodes from routing by computing their enhanced collaborative trust value based on the node's behavior using transmission parameters. Parameters that influence the node's behavior like the number of data packets and control packets forwarded, dropped, or misrouted by the node are quantified to compute direct trust value and neighbor reputation. Node's Enhanced Collaborative trust value was generated by the combination of direct and neighbor observations. **Findings:** The proposed strategy is compared with several cases like the Direct Trust Based Approach (DTBA), where routing involves trustworthy nodes categorized based on only direct trust, existing methods like Belief-dependent trust evolution method(BETM), Novel extended trust-dependent method (NETM) where routing is done with nodes that are categorized as trustworthy depending on direct and indirect observations and simple AODV routing performed with all the possible random nodes without any trust detection. The performance parameters of the proposed ECTBA exhibit a success rate of 10.2% in false positives detection (FPD), network throughput of 438.12 Kbps, and packet delivery ratio (PDR) of 92.3%. This method proves to be a better method when compared with the traditional trust-based security methods(BETM and NETM) in terms of efficiency. **Novelty:** This research suggested a novel and fine-tuned method for quantifying a node's trustworthiness and for secure routing that coupled the direct and indirect observations into enhanced collaborative trust

based on the node's behavior by considering the transmission parameters. The present work highlights the combination of data packets forwarding behavior and control packets forwarding behavior in computing enhanced collaborative trust to decide the involvement of intermediate trustworthy nodes' which is not tried before but this study does.

**Keywords:** Dynamic Topology; Direct Trust; Neighbor Trust; Secure Routing; Enhanced Collaborative Trust

---

## 1 Introduction

Security is always a important factor in MANET due to frequent connection interruptions, bandwidth, resource constraint and high mobility of the wireless nodes. The nodes may behave selfishly and maliciously due to the energy constraints in forwarding other nodes' packets as they have to use their energy<sup>(1-4)</sup>. Many conventional cryptography-based approaches are in existence to resolve security issues in MANET but these approaches are inefficient in isolating the untrustworthy nodes. Several trust-dependent mechanisms are presented to isolate the malicious nodes from routing. These methods are treated as an efficient measure to encounter the security threats caused by pernicious nodes<sup>(5)</sup>. But all these trust-dependent mechanisms are computing the value of the trust using either direct or indirect observations and do not consider the network parameters while evaluating the node's trustworthiness.

Anwar et al. presented BETM Belief-dependent trust evolution method for Mobile Adhoc Networks<sup>(6)</sup>. This mechanism differentiates the malicious nodes from good nodes. It also provides security against On-Off attacks, Denial of Service attacks, Bad-mouth attacks. This mechanism supports an estimation approach based on Bayesian formulae for calculating sensor nodes' direct and neighbor observations for secure data transmission without including the malicious nodes in routing. The drawback of this methodology is that only the packet forwarding nature of the node is considered in evaluating the trust. Syed and Shahzad<sup>(7)</sup> proposed a novel extended trust dependent method NETM for secure data transmission in the MANET. It uses a trust-dependent approach that takes into consideration both blind-direct trust and referential-neighbour trust. The drawback of this approach is that it computes trust value to classify a node's trustworthiness purely based on previous experience trust. Usha and Radha proposed a CLAODV method coupled with adaptive RSA which is used to avoid malicious nodes<sup>(8)</sup>. This method isolates the malicious nodes based on their past experiences but does not consider the present behavior which is a pitfall. A trust based security solution in Mobile ad hoc networks is given by Alrahhhal et al.<sup>(9)</sup>, where trust is generated based on direct and neighbor observations. The proposed method calculates trust based on only the packet transmission nature of the Node. In the Dynamic Bargain Theory methodlogy proposed by Sumathi<sup>(10)</sup>, trustworthiness of data is calculated based on nodes' trustworthiness, whereas the latter is calculated using only the direct behaviour of the node. Here the drawback is that neighbor observations and other parameters are not considered for evaluating the node's trustworthiness. In the work proposed by Sathesh and Prasad<sup>(11)</sup>, a security mechanism to minimize the chances of a weak node becoming a cluster head is proposed. The method purely depends on the observations collected directly from the node which is a drawback in realty. The direct observations are computed based on the source node's observations but not considering the packet forwarding nature of the node. A energy-efficient trust-centered multipath routing scheme is presented by Alappatt et al.<sup>(12)</sup>, primarily depends on direct and indirect trust and different path trust factors. In this method, trust is evaluated majorly based on the communication behavior of the nodes and does not consider any network parameters which is the strong limitation. Khan et al. presented a hybrid and a multifactor trust

model which depends on sensor nodes trust value, residual energy, no of hops, and routing methodology<sup>(13)</sup>. The multifactor methodology identifies trustworthy nodes to forward data in turn to reduce energy. However, trust is generated using communication behavior of the node. Transmission parameters are not considered for computing the trust value in this approach which is a major disadvantage. A Long Short-Term Memory model (LSTM) based on adaptive trust model proposed by Du et al.<sup>(14)</sup> in which the trust is evaluated in two steps: data collection and evaluation of trust. The limitation of this method is calculating the trust using a direct strategy. Alnumayet al. discussed a quantitative method for an IoT integrated with MANET<sup>(15)</sup> that clubs both direct and neighbor trust opinions to compute the resultant trust value. Here the different trust evidence and direct trust are combined using beta probabilistic distribution. The authors claimed that this method depends on the good and bad characteristics of the node. However, a more sophisticated approach is required to compute trust value for efficient isolation of the malicious nodes. With this context, a new algorithm for routing using reputation of the nodes is proposed by Guaya-Delgado et al.<sup>(16)</sup>. The reputation values are assigned to each node in the network and based upon these values, nodes are distinguished as good/bad. The method assumes the statutory behavior of the node which is a major drawback in MANET's.

Gopala Krishnan aims to produce a power management scheme for MANET that protects energy while routing is performed based on the cluster<sup>(17)</sup>. The cluster heads may not work properly and sometimes fail causing power issues. In this method, three key elements viz. the information gathering, computing trust levels, and trust-based configuration are used to form a trust management system which coordinates with each node to maintain reliable network connections. This scheme depends only on direct observations and becomes a drawback at the time of calculating the trust factor. A new cross-layer-based trust estimation method proposed by Dhage and Vemuru provides a defense against multiple attacks<sup>(18)</sup>. The trust is evaluated based on direct observation and the major drawback of the method is not considering the network parameters in the evaluation of the trust value.

Therefore, from the previous studies, it is seen that all the existing trust-based mechanisms proposed for secure data transmission in MANETs are either dependent on direct or indirect observations for trust evaluation. It is also observed that trust evaluation is performed based on node's communication behavior as well and they are not considering all the transmission parameters for node's trust quantification. Most of the researchers are not concentrating on the isolation of malicious nodes using the trust factor. It is observed and quite necessary that an efficient and secure trust-based mechanism is required to ensure a secured routing by isolating malicious nodes using the node's trustworthiness. Node's trust value should be evaluated based upon nodes' packet forwarding behavior in terms of transmission parameters. Both the control and data packets forwarding behavior should be taken into consideration.

In this paper, a new and efficient model based on trust that combines direct and neighbor trust values is presented. The major contribution of the article is 1) to ensure a secured routing by isolating malicious nodes using trustworthiness; 2) Quantifying the trust value based on the direct and indirect observations using transmission parameters. The direct trust observations are quantified by considering transmission parameters. Neighbor trust observations are computed by considering the weights allotted to the neighbor nodes depending on their distance. The proposed enhanced collaborative scheme (ECTBA) couples direct observations of the nodes and various neighbor recommendations collected for calculating the resultant trust. To provide good performance and trustable links for the secure transmission of data, the proposed methodology believes in trust factor. The presented method is compared with existing methods, i.e., BETM, NETM, Direct Trust based approach, and simple AODV routing without any trust calculation to evaluate the performance. From the end results, it is seen that the proposed method ECTBA outperforms all the above-mentioned methods in terms of performance parameters like packet delivery ratio, throughput, and false positive detection.

## 2 Materials and Methods

### 2.1 Proposed Model

In this proposed model, enhanced collaborative trust is computed based on direct and indirect observations using network parameters such as nodes data, control packets forwarding count, drop ratio, and misrouted number. Further enhanced collaborative trust of the node is mapped with the threshold value to identify malicious nodes and later they are isolated for efficient routing.

The existing methods BETM and NETM compute the trust of the node based on only the node's packet forwarding behavior. In these existing methods, neighbor node locations in the network are ignored while computing the neighbor observations. The proposed model ECTBA computes the trust value for isolation of the misbehaving nodes based on transmission parameters considering both direct and neighbor observations. While computing neighbor observations, neighbor nodes are given weights based on their location in the network.

## 2.2 Trust Model

The trust model explained in Fig.1 illustrates the generation of trust, propagation, node categorization, and routing decision. Trust is generated based on direct and indirect observations using transmission parameters. Trust generated for all the nodes is propagated across the network. A node is categorized into malicious and trustworthy based upon the threshold value. Threshold value is fixed considering the network parameters.

The routing decision is completely based on the node categorization, and the node categorization depends on trust calculation.

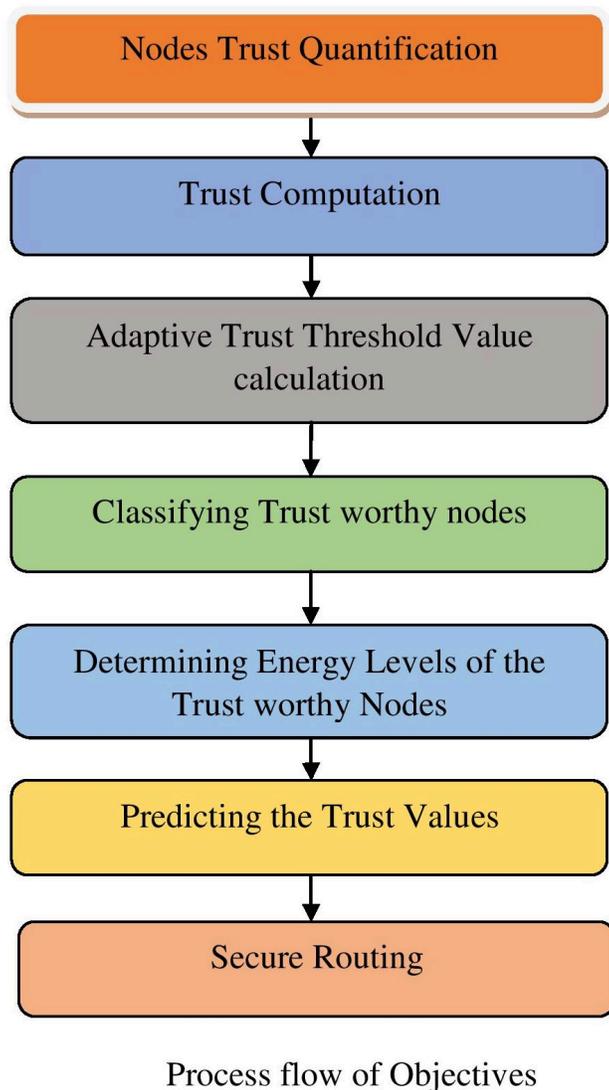


Fig 1. Framework for Trust Calculation and Routing Decision

## 2.3 Trust Calculation

### 2.3.1 Direct Trust Based Approach (DTBA)

A node records all its observations on the trustee node intended for trust calculation directly. Based on the behavior of the node, it records all its findings directly. The packet forward behavior of the nodes is taken into consideration for evolving the trust using DTBA.

A node should listen to its neighbor nodes to estimate the trust. Due to the distributed characteristics of mobile ad-hoc networks, a node can monitor the neighbor node to evaluate its behavior during their direct communications in a passive mode. The proposed scheme uses direct observations to calculate direct trust-related values ( $D_{Trust}$ ) of neighbor nodes by applying the below parameters.

- a. Number of data packets forwarded(considered as Good)
- b. Number of data packets dropped (considered as Bad)

Node A collects the information regarding the above parameters for calculating trust by observing traffic that goes through each neighbor of Node B. Then Node A uses the above information to generate the direct trust value in a time period ( $\tau_{Time}$ ). After running out of each time period( $\tau_{Time}$ ), the trust parameters information is gathered again and direct trust value ( $D_{Trust}$ ) is evaluated. Trust is calculated time to time due to the adhoc nature of MANETs and the nodes are grouped depending on their good and bad characterizations.

After gathering the information about the good and bad behaviors of the nodes, the direct trust can be calculated by Node A on Node B using Equation 1.

$$D_{Trust} = \alpha / (\alpha + \beta) \quad (1)$$

where  $\alpha$  represents Good behavior,  $\beta$  represents Bad behavior, and  $0 \leq D_{Trust} \leq 1, \alpha, \beta > 0$

### 2.3.2 DTBA Algorithm

The algorithm for DTBA considers the following steps:

- Step 1: Consider the node for which trust should be calculated.
- Step 2: Observe the behavior of the node using parameters like packet forwarding and packet dropping.
- Step 3: Compute the direct trust based on equation 1.
- Step 4: Direct trust value can be propagated to all the nodes.
- Step 5: Compare the trust value evaluated with a threshold value fixed based upon network performance.
- Step 6: Node categorization can be done based on Step 5.
- Step 7: Repeat the procedure for all the nodes to be involved in routing.
- Step 8: Routing will be performed only with those nodes categorized as trustworthy.

### 2.3.3 Proposed Enhanced Collaborative Trust-Based Approach (ECTBA)

The proposed enhanced collaborative trust can be computed based upon the direct and neighbor observations on that node.

For computing, the proposed scheme, the following parameters are taken into consideration.

#### 1. For Data Packets

- Total no of data packets received at the node ( $D_t$ )
- Number of data packets forwarded correctly by the node ( $D_f$ )
- Number of data packets dropped by the node ( $D_d$ )
- Number of data packets misrouted by the node ( $D_m$ )

#### 2. Control Packets Forwarded

- Total no of Route Request packets received at the Node ( $R_{tr}$ )
- Total no of Route Reply packets received at the Node ( $R_{tp}$ )
- Number of the Route Request packets forwarded by the Node ( $R_r$ )
- Number of the Route Request packets forwarded by the Node ( $R_p$ )

After gathering the information using the above parameters regarding the node, the data packets and control packets ratio can be estimated by Node A on Node B using Equation 2 and Equation 3.

Data packets ratio, DR

$$DR = w_1 \left( \frac{D_f}{D_t} \right) + w_2 \left( \frac{D_d}{D_t} \right) + w_3 \left( \frac{D_m}{D_t} \right)$$

where,  $w_1 + w_2 + w_3 = 1$

Control packets ratio, CR

$$CR = w_1 \left( \frac{R_r}{R_{tr}} \right) + w_2 \left( \frac{R_p}{R_{tp}} \right)$$

where,  $w_1 + w_2 = 1$

$D_{Trust}$  = Direct Trust and then,

$$D_{Trust} = w_1 \times DFR + w_2 \times CFR$$

where,  $w_1 + w_2 = 1$

### 2.3.4 Neighbor Node Reputation Trust Calculation

In the process of trust evolution on Node B, Node A also considers the recommendations of neighboring nodes on Node B in the 1-Hop distance<sup>(19)</sup>. The Neighbor Node Trust ( $ND_{Trust}$ ) is calculated through Equation 5.

$$ND_{Trust} = \sum_{i=1 \text{ to } n} (w_i \times D_{Trust})$$

Where  $w_i$  is weights allotted to the nodes based on their location in 1-Hop,  $0 \leq w_i \leq 1$ , and  $D_{Trust}$  represents the Direct Trust Observations of the Neighboring Nodes on Node B.

### 2.3.5 Enhanced Collaborative Trust

The final enhanced collaborative trust ( $C_T$ ) of a node is computed using Direct Trust  $D_{Trust}$  and Neighbor Node Trust Calculation  $ND_{Trust}$  (20) through Equation 6.

$$C_T = w_1 \times D_{Trust} + w_2 \times ND_{Trust}$$

where,  $w_1 + w_2 = 1$

### 2.3.6 ECTBA Algorithm

Algorithm for Enhanced Collaborative Trust Based Approach (ECTBA)

- Step 1: Consider the node for which trust should be calculated.
- Step 2: Observe the behavior of the node in terms of parameters: Data Packets and Control Packets
- Step 3: Compute the Data Packet ratio and Control Packet ratio using equations 2& 3 respectively.
- Step 4: Compute the Direct trust based on Equation 4.
- Step 5: Compute Neighbor Trust based on Equation 5
- Step 6: Compute Enhanced Collaborative Trust using Equation 6.
- Step 7: Final Enhanced Collaborative trust value computed can be propagated to all the nodes.
- Step 8: Compare final trust value evaluated with threshold value fixed based upon network performance.
- Step 9: Node categorization can be done based on Step 7.
- Step 10: Repeat the procedure for all the nodes to be involved in routing.
- Step 11: Routing will be performed only with those nodes categorized as trustworthy.

## 2.4 Trust Propagation

Once the trust is evaluated on target node by any of the nodes, the resources used for recomputation of trust by other nodes can be minimized if the evaluated trust gets propagated in the network.

The Enhanced Collaborative Trust ( $C_T$ ) determined for the target node is broadcasted across the network so that the other nodes can update.

## 2.5 Node Categorization

A node is said to be a bad one if it randomly drops packets intentionally but not for intrinsic network issues. The nodes are clustered into two groups Good or Bad, depending on their Direct Trust evaluated using Equation 3 and compared with Trust threshold (TH). These threshold limits are fixed depending on network configuration. Static trust threshold value is taken into consideration and average threshold trust value of 0.6 is used to isolate malicious nodes.

- Good: if  $C_T \geq TH$
- Bad: if  $C_T < TH$

## 2.6 Routing Decision

Source finds trusted nodes using the proposed scheme to establish the secure route to the destination. Each node consists of a list of dependable(trusted) neighbor nodes as well as their latest calculated trust values. Good nodes are used to form path between source and destination.

### 3 Simulation Carried

The performance of the proposed ECTBA is evaluated by comparing it with the existing BETM, NETM, Direct Trust-based approach, and simple AODV routing without any trust calculation. The proposed model of this article computes the enhanced collaborative trust value of a node by looking into its data transmission nature. Various transmission parameters like the number of data packets and control packets forwarded or dropped are considered for the resultant trust calculation. In this proposed solution, the resultant enhanced collaborative trust is evaluated using a computable approach by considering the direct and other trust observations on the node using network parameters. Simulation is carried out in a  $700 \times 500 m^2$  network area and IEEE 802.11 MAC for 500s with 100 nodes<sup>(6,21)</sup>. Table 1 summarizes the simulation parameters.

**Table 1.** Simulation Parameters

Simulation Parameter	Value
Simulator	NS2.34
No of Nodes	100
Network area	700 x 500
Packet Size	512 bytes
No. of malicious nodes	05
Traffic Type	CBR/UDP
Mobility	4–25 m/s
Pause Time	5s
Simulation Time	500s

## 4 Results and Discussions

### 4.1 Results

The performance of the proposed model when mapped with methods BETM and NETM are shown in Table 2 with computed trust values for the proposed and existing models.

**Table 2.** Trust Value Computations of proposed ECTBA, DTBA, BETM and NETM

	ECTBA			DTBA	BETM	NETM
Node No.	Direct Trust Calculation	Neighbor Trust calculation	Node enhanced collaborative Trust Value	Direct & Final Trust Calculation	Final Trust Value	Final Trust Value
Node1	0.89	0.39	0.81	0.78	0.79	0.75
Node2	0.67	0.37	0.71	0.64	0.73	0.34
Node3	0.44	0.56	0.65	0.45	0.54	0.65
Node4	0.27	0.49	0.23	0.31	0.65	0.61
Node5	0.39	0.56	0.69	0.36	0.52	0.61
Node6	0.19	0.37	0.73	0.25	0.65	0.32
Node7	0.09	0.15	0.29	0.69	0.35	0.67
Node8	0.02	0.71	0.32	0.12	0.65	0.69
Node9	0.59	0.56	0.66	0.55	0.52	0.53
Node10	0.79	0.15	0.54	0.65	0.53	0.57
Node11	0.65	0.52	0.75	0.63	0.63	0.73
Node12	0.57	0.33	0.73	0.59	0.69	0.66
Node13	0.45	0.32	0.65	0.67	0.61	0.72
Node14	0.58	0.35	0.77	0.55	0.54	0.24
Node15	0.49	0.32	0.65	0.52	0.61	0.32

The proposed ECTBA method classifies the malicious nodes and trustworthy nodes depending on the computed enhanced collaborative trust value. Here average trust threshold value taken into consideration based on network conditions. Node classification details are shown in Table 3.

Table 3. Classification of the Nodes

Node	ECTBA - Node enhanced collaborative Trust Value	Static Trust Threshold	Decision
Node1	0.81	0.6	Trusted Node
Node2	0.71	0.6	Trusted Node
Node 3	0.65	0.6	Trusted Node
Node 4	0.23	0.6	Malicious Node
Node5	0.69	0.6	Trusted Node
Node6	0.73	0.6	Trusted Node
Node7	0.29	0.6	Malicious Node
Node8	0.32	0.6	Malicious Node
Node9	0.66	0.6	Trusted Node
Node10	0.54	0.6	Malicious Node
Node11	0.75	0.6	Trusted Node
Node12	0.73	0.6	Trusted Node
Node13	0.65	0.6	Trusted Node
Node14	0.77	0.6	Trusted Node
Node15	0.65	0.6	Trusted Node

It is noticed that around 22% of the packets (non-intentional-malicious drops) are found lost because of the environmental glitches in the network. For the performance evolution of the proposed model, metrics like Packet Delivery Ratio (PDR), Detection of False Positives, and Throughput are considered.

### 4.2 False Positive Detection Ratio

FPD ratio is the count of good nodes falsely identified as malicious nodes to the total available number of nodes in the network. It is used to calculate the (False positive)<sup>(22)</sup>. False positive detection ratio of the nodes for DTBA and proposed ECTBA with existing BETM and NETM methods as shown in Figure 2. From the graph it is seen that 10.6% of nodes are falsely detected as false positives, malicious when 25% of packet collisions occur in ECTBA, 9.2% false positive detection in BETM, 8.6% false positive detection in NETM, and 4.8% nodes are wrongly detected as false positives in DTBA. However, the results also show that ECTBA is efficient in case of False positives detection.

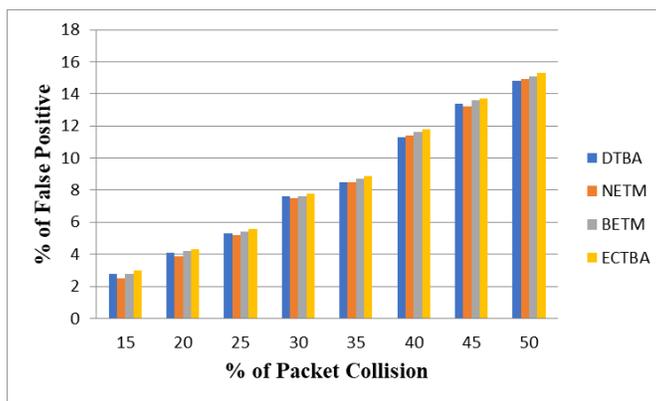


Fig 2. False Positive Detection Ratio

### 4.3 Packet Delivery Ratio

Figure 3 depicts how packet delivery is affected by the presence of malicious nodes. The graph depicts the mapping between the routing protocol without any trust, DTBA, NETM, BETM, and the proposed model ECTBA. It shows high Packet Delivery Ratio. In the presence of 5% of malicious nodes, packet delivery ratio for the routing protocol without any trust calculation is 72.2%, 82% for DTBA, 85% for NETM, 87.2% for BETM and 92.3% for ECTBA. It is observed that in the proposed scheme ECTBA is detecting malicious nodes using enhanced collaborative trust evaluation efficiently and avoiding them in routing thus increasing the packet delivery ratio.

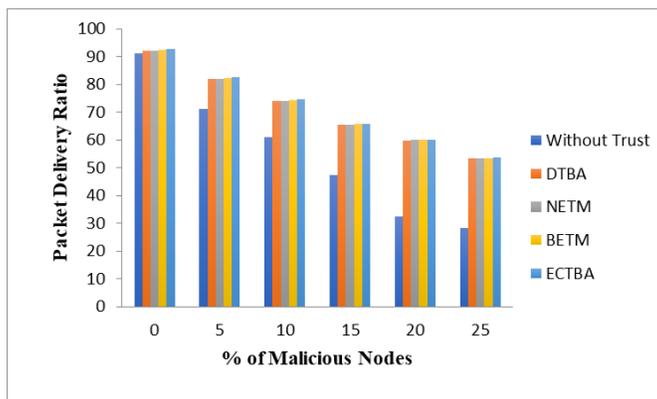


Fig 3. Packet delivery ratio mapped with number of malicious nodes

### 4.4 Throughput

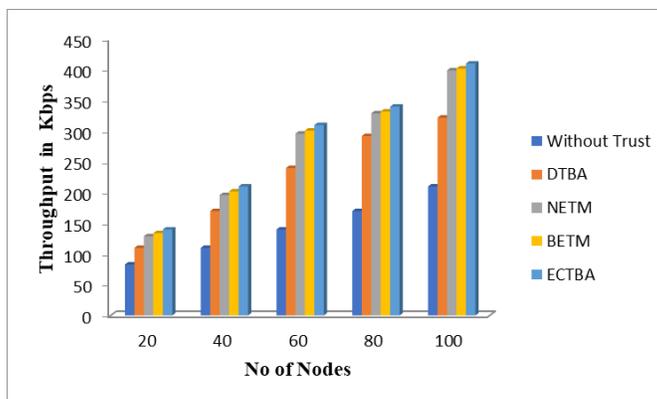


Fig 4. Throughput comparison

Throughput of the routing protocol without any trust computation, Direct Trust-based approach DTBA, NETM, BETM, and proposed method ECTBA with enhanced collaborative trust calculations are shown in Figure 4. Throughput is the number of units delivered in a stipulated time. The graph shows that the proposed method ECTBA has a throughput performance of 438.12 Kbps, BETM has 432.2 Kbps, and NETM has 428.24Kbps. whereas the scheme DTBA has 422.72 Kbps and the scheme without any trust calculation has 349.34 Kbps. It shows that the presented scheme is efficient throughput wise.4.5

#### Discussion

The proposed model ECTBA is compared with existing methods BETM and NETM. Existing methods compute the trust value using only the node’s packet forwarding behavior. The proposed ECTBA estimated the node’s trust value depending on network parameters and assigns weights for the neighbor nodes while computing neighbor observations. The performance metrics comparison between proposed and existing methods is tabulated and shown in Table 4 .

**Table 4.** Comparison of results – Efficiency of the proposed method ECTBA

S.No	Performance Parameter	Proposed Method – ECTBA (%)	DTBA (%)	NETM (%)	BETM (%)	Without any Trust Calculation(AODV) (%)
1	Packet Delivery Ratio	92.3	82	85	87.2	71
2	False Positive Detection Ratio	10.6	4.8	8.6	9.2	-
3	Throughout	438.12 Kbps	422.72 Kbps	428.24 Kbps	432.2 Kbps	349.34 Kbps

The proposed method ECTBA shows a high packet delivery ratio when compared with the existing methods BETM and NETM. It clearly shows more than a 5% of increment in packet delivery ratio. The proposed ECTBA detects false positives at the rate of 10.6 % when compared with existing methods. It shows increased throughput of 438.12 kbps and exhibits high performance.

### 5 Conclusion

This paper presents a quantitative trust model ECTBA by the integration of various data and control packets, and network parameters using direct and neighbour observations in calculating collaborative trust. The proposed trust method calculates resultant trust using the combination of data packets and control packets ratio along with indirect observations. Results obtained from simulation show that the model performs with good efficiency with respect to performance metrics like false positive detection, packet delivery ratio, and throughput as compared to the routing protocol without any trust calculation for DTBA or other existing NETM and BETM methods. Based on the comparison, the proposed model ECTBA has achieved a 10.2% success rate in False Positives detection, 438.12 Kbps Throughput, and 92.3% Packet Delivery Ratio. This shows that trustworthy nodes can be easily detected through this method efficiently. So the path between source and destination is connected with the identified trusted nodes.

In the future, this method can be used in making efficient routing decisions. The enhanced collaborative trust factor after computation can be compared with the adaptive threshold for secured and efficient routing decisions. The proposed trust scheme can be applied to provide security in mobile ad-hoc networks.

### 6 Acknowledgement

The authors thank all the members for having provided their continuous support to this work.

### References

- Zhihan LV, Song H. Trust Mechanism of Feedback Trust Weight in Multimedia Network. *ACM Transactions on Multimedia Computing, Communications, and Applications*. 2021;17(4):1–26. Available from: <https://doi.org/10.1145/3391296>.
- Kotteeswaran C, Patra I, Nagaraju R, Sungeetha D, Kommula BN, Algani YMA, et al. Autonomous detection of malevolent nodes using secure heterogeneous cluster protocol. *Computers and Electrical Engineering*. 2022;100:107902–107902. Available from: <https://doi.org/10.1016/j.compeleceng.2022.107902>.
- Mbarek B, Ge M, Pitner T. An adaptive anti-jamming system in HyperLedger-based wireless sensor networks. *Wireless Networks*. 2022;28(2):691–703. Available from: <https://doi.org/10.1007/s11276-022-02886-1>.
- Mohammadi V, Rahmani AM, Darwesh AM, Sahafi A. Trust-based recommendation systems in Internet of Things: a systematic literature review. *Human-centric Computing and Information Sciences*. 2019;9:1–61. Available from: <https://doi.org/10.1186/s13673-019-0183-8>.
- Yang H. A Study on Improving Secure Routing Performance Using Trust Model in MANET. *Mobile Information Systems*. 2020;2020:1–17. Available from: <https://doi.org/10.1155/2020/8819587>.
- Anwar RW, Zainal A, Outay F, Yasar A, Iqbal S. BTEM: Belief based trust evaluation mechanism for wireless sensor networks. *Future generation computer systems*. 2019;96:605–621. Available from: <https://doi.org/10.1016/j.future.2019.02.004>.
- Syed SA, Ali S. Enhanced dynamic source routing for verifying trust in mobile ad hoc network for secure routing. *International Journal of Electrical and Computer Engineering (IJECE)*. 2022;12(1):425–425. Available from: <https://doi.org/10.11591/ijece.v12i1.pp425-430>.
- Usha S, Radha S. Detection and avoidance of node misbehavior in MANET based on CLAODV. *Indian Journal of Science and Technology*. 2011;4(10):1340–1346. Available from: <https://doi.org/10.17485/ijst/2011/v4i10.12>.
- Alrahhah H, Jamous R, Ramadan R, Alayba AM, Yadav K. Utilising Acknowledge for the Trust in Wireless Sensor Networks. *Applied Sciences*. 2022;12(4):2045–2045. Available from: <https://doi.org/10.3390/app12042045>.
- Sumathi AC, Akila M, De Prado RP, Wozniak M, Divakarachari P. Dynamic Bargain Game Theory in the Internet of Things for Data Trustworthiness. *Sensors*. 2021;21(22):7611–7611. Available from: <https://doi.org/10.3390/s21227611>.
- Satheesh N, Prasadh K. Improvements in Cluster-Based Routing to Protect Malicious Node Attacks on TAODV Routing Protocol using MANET. *Applied Mathematics & Information Sciences*. 2019;13(6):899–911. Available from: <https://doi.org/10.18576/amis/130603>.

- 12) Alappatt V, M JPP. Trust-Based Energy Efficient Secure Multipath Routing in MANET Using LF-SSO and SH2E. *International Journal of Computer Networks and Applications*. 2021;8(4):400–400. Available from: <https://doi.org/10.22247/ijcna/2021/209706>.
- 13) Khan T, Singh K, Manjul M, Ahmad MN, Zain AM, Ahmadian A. A Temperature-Aware Trusted Routing Scheme for Sensor Networks: Security Approach. *Computers & Electrical Engineering*. 2022;98:107735–107735. Available from: <https://doi.org/10.1016/j.compeleceng.2022.107735>.
- 14) Du J, Han G, Lin C, Martinez-Garcia M. LTrust: An Adaptive Trust Model Based on LSTM for Underwater Acoustic Sensor Networks. *IEEE Transactions on Wireless Communications*. 2022;p. 1–1. Available from: <https://doi.org/10.1109/TWC.2022.3157621>.
- 15) Alnumay W, Ghosh U, Chatterjee P. A Trust-Based Predictive Model for Mobile Ad Hoc Network in Internet of Things. *Sensors*. 2019;19(6):1467–1467. Available from: <https://doi.org/10.3390/s19061467>.
- 16) Guaya-Delgado L, Pallarès-Segarra E, Mezher AM, Forné J. A novel dynamic reputation-based source routing protocol for mobile ad hoc networks. *EURASIP Journal on Wireless Communications and Networking*. 2019;2019(1):1–6. Available from: <https://doi.org/10.1186/s13638-019-1375-7>.
- 17) Krishnan CG, Nishan AH, Gomathi S, Swaminathan GA. Energy and Trust Management Framework for MANET using Clustering Algorithm. *Wireless Personal Communications*. 2022;122. Available from: <https://doi.org/10.1002/dac5138>.
- 18) Dhage MR, Vemuru S. Trust based Secure Routing using Cross layer for Heterogeneous Environment in WSN. *International Journal of Emerging Trends in Engineering Research*. 2020;8(7):3241–3246. Available from: <https://doi.org/10.30534/ijeter/2020/59872020>.
- 19) Mukhedkar MM, Kolekar U. Hybrid PSGWO Algorithm for Trust-Based Secure Routing in MANET. *Journal of Networking and Communication Systems (JNACS)*. 2019;2(3). Available from: <https://doi.org/10.46253/jnacs.v2i3.a1>.
- 20) Kumar N. Battery power and trust based routing strategy for MANET. *In 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*. 2014;p. 1559–1562. Available from: <https://doi.org/10.1109/icimia48430.2020.9074870>.
- 21) Sarbhukan VV, Raha L. Establishing Secure Routing Path Using Trust to Enhance Security in MANET. *Wireless Personal Communications*. 2020;110(1):245–255. Available from: <https://doi.org/10.1007/s11277-019-06724-0>.
- 22) Wahi C, Chakraverty S, Vandana Bhattacharjee. A trust-based secure AODV routing scheme for MANET. *International Journal of Ad Hoc and Ubiquitous Computing*. 2021;38(4):231–231. Available from: <https://doi.org/10.1504/ijahuc.2021.119853>.

## Enhancing the Routing Security through Node Trustworthiness using Secure Trust Based Approach in Mobile Ad Hoc Networks

<https://doi.org/10.3991/ijim.v16i14.30651>

M. Venkata Krishna Reddy<sup>1,2(✉)</sup>, P.V.S. Srinivas<sup>3</sup>, M. Chandra Mohan<sup>1</sup>

<sup>1</sup>Jawaharlal Nehru Technological University, Hyderabad, India

<sup>2</sup>Chaitanya Bharathi Institute of Technology(A), Hyderabad, India

<sup>3</sup>Vignan Bharathi Institute of Technology(A), Hyderabad, India

krishnareddy\_cse@cbit.ac.in

**Abstract**—Mobile Ad Hoc Networks, also known as MANET's are the part of many heterogeneous networks which utilizes the technologies like Internet of Things. Internet is filled with known as well as unknown sources which are still considered as a challenge. Secure routing is always a major concern in MANET's. Among all the existing and proposed cryptographic approaches to provide security to these networks seemed lengthy, complex and inefficient in eliminating malicious nodes. Many trust based approaches are proposed to replace these traditional cryptographic security methods for secure routing in MANET's. But all those trust based approaches concentrate on either direct observations or hybrid observations to determine the node's trustworthiness without taking into count network parameters. Considering the security challenges that arise due to the topology, infrastructure and bandwidth of MANET's, a novel secure trust based approach (STBA) is proposed in this article to strengthen the evolution of trust component for effective isolation of malicious nodes and secure routing. This work focuses on the computation of the node's trust factor based on network parameters and node's behavior to simulate the challenge of providing the secure transmission. The proposed method, STBA computes secure trust of a node depending on three tier observations. The performance of the proposed secure trust mechanism STBA is evaluated by comparing it with routing without any trust calculation, with existing Belief based Trust Evaluation Mechanism (BTEM) and Novel extended trust based mechanism (NETM) where routing is performed involving only with direct and indirect trust computation for node's distribution in both cases. Results show the proposed method is performing well.

**Keywords**—Mobile Ad Hoc networks, secure routing, node trustworthiness, direct trust, indirect trust, secure trust

### 1 Introduction

MANET are portable adhoc networks, which in general forms a dynamic routing virtual network. They are collection of remote self-organized nodes fueled by battery

power where other shapes of communication are unreasonable to convey [1]. They are made up of a group of portable nodes linked electronically in a self-configuring, identity system that does not rely on a wired network. Because of the distributed structure of the MANET, connection between nodes alters regularly and they are ready to roam in and around the network at their will. The node is considered as a gateway, which forwards information to all other devices in the network by dissipating its own resources. The MANET's key problem lies in equipping all these devices with the necessary information to provide services continually. Now a day they are used for monitoring the atmosphere, the house wellness, relief operations, wind defence, weaponry, drones, and other applications like accident prevention. Most of this applications demand certain security levels and raise a basic issue, especially Wireless Sensor Networks WSN is effortlessly defenseless to attacks when compared to wired systems due to its remote broadcasting characterization and constrained assets [2]. The features which make MANET's unique are Dynamic Topology, Autonomy Conduct and Resource Intervention [3].

MANET's are especially susceptible to security because of their lack of dispersed design of the encryption, wireless connectivity and centralized control. They need low latency to set up the connection, making them constantly independent. Isolation from centralized control management made MANET's more vulnerable to security.

### **1.1 Security issues in MANETs**

MANET routing performance is affected by capabilities of mobile nodes [4]. Each unit can act as a gateway in the network as well as a server, demonstrating their independent nature. They have to dissipate their own energy resources for other node's packet forwarding which may lead to behave them as selfish and can act as malicious nodes. Isolation of malicious nodes from the routing in MANET's is always a critical security concern. A secure environment necessitates a set of well-behaving and fair nodes. Because of noise factor in the network, due to lack of centralized control, transmission and supplies are restricted. All connectivity activities, such as filtering and data packets are self-organized in a MANET [5]. Ensuring secure routing in MANET's has become difficult due to these factors.

Node Mobility is also a major security issue in MANET. Packets are routed by establishing the path with available nodes in the network [6]. Nodes may enter and exit the network at any movement and any time [7]. Secure routing is always a challenge with the presence of the malicious nodes that behave selfishly to save their energy resources from being consumed for forwarding other node's data in the network. Many existing encryption algorithms like digital signature and authentication based schemes proved to be inefficient in terms of protecting against attacks from these malicious nodes [8]. Various security solutions based on the trust idea were developed in supplement to the old cryptographic methods. By applying the trust concept in ad hoc networks context, there was a significant trend toward enhancing security in MANET. Quantifying the nodes trustworthiness plays the key role in isolating the malicious nodes thus establish the secure routing and data transmission. Trust factor evaluated confirms each node's

fair participation in routing. Many of the existing and proposed trust based methods are relying on either direct or hybrid observations in deciding the trust factor for node categorization. These methods are not considering network parameters and nodes behavior to evaluate the node's trust factor. These trust based mechanisms are proven to be inefficient in isolating the malicious nodes from their involvement in the routing.

In this article, the node trust worthiness is quantified based on three tier observations, direct, neighbour and self appraisal of the node using the method, STBA, secure trust based approach to enrich the trust factor in isolating the malicious nodes. Results obtained show the satisfactory performance of the proposed method STBA. Routing after nodes trustworthiness evolution using proposed STBA is compared with routing without any trust computation, with existing method BTEM and Novel extended trust based mechanism NETM where routing is performed after evaluating nodes trustworthiness using only direct observation and hybrid observations to show the performance of the proposed method. The main aim of the proposed STBA method in this article is to provide secure routing for data transmission by simulating the important factor node trustworthiness.

This work is organized as follows: Introduction and security issues of the MANET are presented in Section 1; Related Research work on MANET is described under Section 2. Section 3 discusses the proposed STBA method and in Section 4 simulation results and discussions are presented. The concluding remarks of the work and future research recommendations are given in Section 5.

## 2 Related work

In [9], authors presented a belief based trust evolution mechanism BETM for MANET's. This method classifies the malicious nodes and trust-worthy nodes. It defends against several attacks like Denial of Service (DoS), On-Off and Bad-mouth attacks. In this method, authors employed Bayesian estimation approach for computing direct and neighbor trust values of the sensor nodes which estimates imprecise knowledge in decision making by considering the data correlation collected over a period of time for secure transmission of data thus isolating and keeping away the malicious nodes from routing. However, this method considers only packet forwarding behavior of the node in estimating the trust. In [10], authors proposed a estimation-based trust model, Novel extended trust based mechanism NETM which aims on estimating each node's trust level in the network. This mechanism uses blind and referential trust based on previous experiences of the node. This method is not considering network parameters and packet forwarding behavior of the node in estimating the trust. Authors in [11] presented a new evaluated and administration (TEAM) paradigm that provides a distinctive template for the construction, maintenance, and assessment of Trust Models in a variety of situations and the context of malicious nodes. Various trust models (TMs) were actively developed, but presently there exists no practical process of comparing how they might perform in practice in adversarial situations. Nodes in MANET actively communicate critical data such as pre-collision signals. As a result, this data must remain secure, trustworthy and legitimate. For recognizing unscrupulous nodes and identifying the communications containing dangerous data, trust formation between nodes is essential.

Author proposed a trust based approach to ensure MANET’s integrity which focus on direct observations. Similarly, authors in [12] state that wireless adhoc sensor networks (WSNs) are specialized networks with a huge number of sensor units (SNs) which are used to manage multiple natural and physical phenomena. The SNs can indeed be employed in a variety of technical, security, and agricultural purposes, such as mobility tracking and combat monitoring.

Sensors are installed on ad hoc basis and act independently in these systems; additionally, there is a growing demand for encrypted communication across SNS. The authors applied energy-efficient routing protocols (ERPSs) for isolation of selfish nodes based on trust factor where hybrid approach is taken into consideration as a result of the SNs’ structure and constraints [13]. Authors in [14], proposed cluster based trust methods where the network is divided into regions. Clustering is generated based on how comparable SNs are. Each clustering seems to have a set of cluster members (CMs), with one or more designated as cluster chiefs (CHs). CHs evolves the trustworthiness of CMs using direct observations. The data transmission effectiveness is legitimate in this method and also be analyzed through nodes trust factor. In [15] trust based mechanism is illustrated to resolve the congestion problem in the network. Nodes trust is evaluated using hybrid approach to avoid dropping of the packets due to congestion. In [16], authors presented a trust based control mechanism which depends on direct trust factor. In [17], authors proposed a methodology for secure routing taking into consideration, MANET characteristics by assessing the node’s trustworthiness. It is noted that all the trust dependent approaches proposed are taking into account either direct or hybrid observations. Hence a better evaluation of Node Trustworthiness based on Node’s behavior and Network parameters is required to maintain the secure transmission in wireless networks. To strengthen the computation of trust factor for establishing secure routing, a novel method based on three tier observations, STBA is presented in this article.

### 3 Proposed model

The proposed work STBA, Secure Trust Based Approach improvises the trustworthiness of nodes for secure transmission. It computes the secure trust value of the node depending on three level observations: direct, indirect and self appraisal.

#### 3.1 Model for secure trust computation

The model for secure trust computation is given in Figure 1. These processes are basically dependent. The resultant secure trust is combination of three tier observations on the node under consideration. It includes direct observations, neighbor observations and nodes historical trust, Self appraisal. Secure trust value is calculated according to equation 1.

$$\text{Resultant Secure Trust} = \frac{\text{Direct Trust} + \text{Neighbour Trust} + \text{Historical Trust}}{\text{Self appraisal of Node}} \quad (1)$$

The step-by-step calculations for the overall resultant trust i.e secure trust is given in Figure 1.

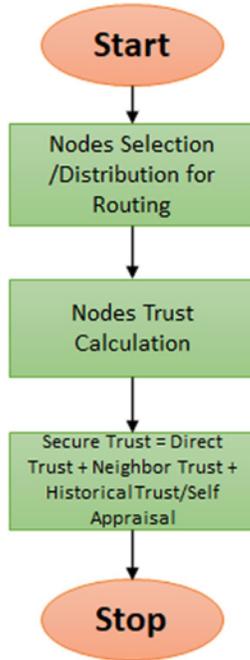


Fig. 1. Flowchart of the proposed STBA method

### 3.2 Direct trust calculation

Direct Trust is evaluated depending upon the direct observations on the node in the network whom trust is being calculated. Direct observations capture the nodes behavior. Several network parameters are taken into consideration here for quantifying the nodes behavior to evaluate direct trust factor. Parameters considered are

*Data Packets Forwarded:*

- Total no of data packets received at the node =  $D_{td}$
- No of data packets forwarded correctly by the node =  $D_{forw}$
- No of data packets dropped by the node =  $D_{drop}$
- No of data packets misrouted by the node =  $D_{mr}$
- No of data packets falsely injected by the node =  $D_{fi}$

Data Packets forwarded ratio, DFR is quantified as given in equation 2.

$$DFR = w_1 * (D_{forw} / D_{td}) + w_2 * (D_{drop} / D_{td}) + w_3 * (D_{mr} / D_{td}) + w_4 * (D_{fi} / D_{td}) \quad (2)$$

Where  $w_1, w_2, w_3, w_4$  are the proportionate weights assigned to the packet forwarding behaviour and can be altered according to the network conditions and

$$w_1 + w_2 + w_3 + w_4 = 1$$

*Control Packets Forwarded:*

Total no of Route Request packets received at the Node =  $R_{req}$

Total no of Route Reply packets received at the Node =  $R_{rep}$

Total no of Route error packets received at the Node =  $R_{terr}$

Total no of Route acknowledgment packets received at the Node =  $R_{tack}$

No of Route Request packets forwarded by the Node =  $R_{req}$

No of Route Reply packets forwarded by the Node =  $R_{rep}$

No of Route error packets forwarded by the Node =  $R_{err}$

No of Route acknowledgment packets forwarded by the Node =  $R_{ack}$

Control Packets Forwarded Ratio, CFR is computed as specified in equation 3.

$$CFR = w_1 * (R_{req}/R_{req}) + w_2 * (R_{rep}/R_{rep}) + w_3 * (R_{err}/R_{terr}) + w_4 * (R_{tack}/R_{ack}) \quad (3)$$

Where  $w_1, w_2, w_3, w_4$  are the proportionate weights assigned to the packet forwarding behaviour and can be changed according to the network conditions and

$$w_1 + w_2 + w_3 + w_4 = 1$$

Direct Trust TS is calculated using Data packets forwarded ratio DFR from equation 2 and Control packets forwarded ratio CFR from equation 3, as shown in equation 4

$$TS = w_1 * DFR + w_2 * CFR \quad (4)$$

Where  $w_1 + w_2 = 1$  and  $w_1, w_2$  are weights assigned to DFR & CFR based on the network environment.

### 3.3 Neighbour trust calculation

Neighbour Trust is collective trust evolution done by all neighbouring nodes which are located in 1 hop distance from the node whose trust is being calculated by quantifying above Data packet forwarding and Control packet forwarding parameters. Weights should be assigned to all the neighbour nodes depending on where they located and distance to the node specified in the network. Weight should be calculated using coordinates. Nearest Neighbour should be assigned with more weight value depending on the total number of neighbours at time  $t$  in the network.

Neighbour Trust, TO is evaluated using equation 5 given below

$$TO = \frac{(w_1 * NT_1 + w_2 * NT_2 + w_3 * NT_3 + w_4 * NT_4 \dots + w_n * NT_n)}{\text{(No of Neighbour Nodes in 1 Hop distance)}} \quad (5)$$

Where  $NT_1, NT_2, NT_3, NT_4 \dots NT_n$  are the trust values calculated by the neighbour nodes by their direct observations using above mentioned parameters. And  $w_1, w_2, w_3, w_4 \dots w_n$  represent the weights assigned to the neighbours depending on their distance in the network.

### 3.4 Historical trust calculation – nodes self appraisal

Nodes under trust evaluation can have their own trust values based upon their self appraisal subjected to their performance and involvement in the fair routing in the past and present. This can be computed as Self appraisal trust factor TH. Nodes Self appraisal is calculated based on its behaviour using the below parameters.

- Number of Packets correctly Forwarded (Good) =  $P_{for} = \alpha$
- Number of Packets dropped without forwarding (Bad) =  $P_{drop} = \beta$
- Number of Packets falsely injected (Bad) =  $P_{fi} = \beta$
- Historical Trust, Self Appraisal of node, TH is given by the equation 6.

$$TH = \alpha / (\alpha + \beta) \tag{6}$$

Final resultant, Node’s secure trust calculation is computed from three tier observations Direct Trust TS, Neighbour Trust TO and Self Appraisal TH using equations 4, 5, 6 respectively.

Secure Trust, T is evaluated as shown in equation 7.

$$\text{Secure Trust, } T = \alpha TS + \beta TO + \gamma TH \tag{7}$$

Where  $\alpha, \beta, \gamma$  are constants and assigned based on the weight factor given to the subsequent observation according to the network conditions.

The Secure trust T evaluated, falls in the range of 0 to 1.

$$0 \leq T \leq 1$$

### 3.5 Static threshold

Node’s secure trust is computed using the equation 7 and later it is matched with the static threshold in order to decide the nodes trustworthiness. Whether a node can be included as intermediate node for secure routing or not. Static Threshold is fixed based upon the network conditions. Various levels of static trust threshold fixed are given in Table 1.

**Table 1.** Levels and rankings for the trust value

Level	Resultant Secure Trust Value	Ranking
1	-1	Complete Distrust
2	0	New or Unknown
3	0.2	Very Low Trust
4	0.4	Low Trust
5	0.6	Average Trust
6	0.8	High Trust
7	1	Absolute Trust

The average static trust threshold value considered is

$$TH_{threshold} = 0.6 - Average\ Trust$$

Nodes trustworthiness is classified based on the average static threshold.

Node's Classification –  $T \geq TH_{threshold}$  = Trustworthy Node  
 $T < TH_{threshold}$  = Untrustworthy Node

### 3.6 Algorithm (Secure Trust Based Approach-STBA)

#### Procedure Direct Trust (TS, N1, N2, DFR, CFR)

//TS is the Direct Trust

//N1 is the node and N2 is its neighbor node

//Data packets forwarded ratio is DFR

//Control Packets Forwarded Ratio CFR

//Direct Trust TS

{

**Step 1:** if N1 initiates finding trustworthiness on N2 node  
 then the process starts.

**Step 2:** Data Packet Ratio is calculated as

$$DFR = w_1 * (D_{forw} / D_{td}) + w_2 * (D_{drop} / D_{td}) + w_3 * (D_{mr} / D_{td}) + w_4 * (D_{fi} / D_{td})$$

**Step 3:** Control Packet Ratio is calculated using

$$CFR = w_1 * (R_{req} / R_{treq}) + w_2 * (R_{rep} / R_{trep}) + w_3 * (R_{err} / R_{terr}) + w_4 * (R_{tack} / R_{ack})$$

**Step 4:** Then Direct Trust factor, TS is calculated from

$$TS = w_1 * DFR + w_2 * CFR$$

}

**end procedure**

#### Procedure Neighbor Trust (TO, N1, N2)

//TO is the Neighbor Trust

//N1 is the node of which trust to be evaluated and N2 is its neighbor node

//Neighbor Trust TO

{

**Step 1:** If all the neighbors in 1-hop distance initiates finding  
 the node trustworthiness of the node N1 under  
 consideration based on their direct observations. Then the process starts.

**Step 2:** Neighbor Trust, TO is calculated using

$$TO = (w_1 * NT_1 + w_2 * NT_2 + w_3 * NT_3 + w_4 * NT_4 + \dots + w_n * NT_n) / (\text{No of Neighbour Nodes in 1 Hop distance})$$

**Step 3:** Weights are assigned depending on location and the distance of the  
 neighbor nodes in the network.

}

**end procedure**

**Procedure Self Trust (TH, N1, PF, PD)**

```
//TH is the Historical/Self Appraisal Trust
//N1 is the node
//Data packets forwarded – PF
//Data packets dropped – PD
//Self Appraisal Trust TH
{
Step 1: Node N1 evaluates its own trust based upon the
        packets routing, process starts.
Step 2: Packets forwarded  $PF = PF + 1 - \alpha$ 
Step 3: Packets dropped  $PD = PD - 1 - \beta$ 
Step 4: Self Appraisal Trust, TH is calculated from
         $TH = \alpha / (\alpha + \beta)$ 
}
end procedure
```

**Procedure Secure Trust (T, TS, TO, TH, N1, N2)**

```
//TS is the Direct Trust
//TO is the Neighbor Trust
//TH is the Historical/Self Appraisal Trust
//T is resultant Secure Trust
{
Step 1: If node N1 gets TS, TO, TH on a node N2 whose trust is being evaluated
        then computes the resultant secure trust value.
        Secure Trust,  $T = \alpha TS + \beta TO + \gamma TH$  and  $0 \leq T \leq 1$ 
        else
Step 2: set final secure trust value, T to 0
        end if
}
end procedure
```

**Procedure Secure Routing (T,  $TH_{\text{threshold}}$ )**

```
//T is resultant Secure Trust
// $TH_{\text{threshold}}$  is the Static trust threshold
//Secure Routing
{
Step 1: Average static trust threshold is 0.6
Step 2: If  $T \geq TH_{\text{threshold}}$ , Node is classified as Trustworthy
        else
         $T \leq TH_{\text{threshold}}$ , Node is classified as malicious and
        isolated.
        end if
Step 3: Perform secure Routing involving only trustworthy
        nodes as intermediate nodes.
}
end procedure
```

## 4 Results and discussions

### 4.1 Simulation

Simulation is performed on Network Simulator NS2. The components in the different layers are: Wireless Physical layer followed through MAC 802.11 Data link layer and AODV for the network layer and finally, the UDP (User data gram protocol) for Transport layer. These settings are made in the Network Simulator 2 (NS2). The constant bit rate traffic is fixed with 512 bytes size for 200 and 100 packets per second i.e., packet rate. Simulation parameters considered can be seen in Table 2 and parameters for network configuration are given under Table 3 [9] [12].

**Table 2.** Parameters illustrating network configuration

Simulation tool	NS2
Total Number of Nodes used for Simulation	100
Malicious Nodes Inserted	15
Propagation Model used	Two ray ground
Malicious Nodes Declaration time	0t
Topography used	700*500(M)
Simulation Time	500s
Mobility(r)	5m/s

**Table 3.** Network configuration parameters

Parameter	Value
Simulation tool	NS2
Version	2.35 (base)
Operating System	Fedora 11
Channel	Wireless channel
Type of Network Interface	Wireless Physical
Medium Access Control Protocol	MAC 802.11
Type of Interface Queue	Drop Tail
Interface Queue Length	50
Type of Antenna	Omni Directional
Network Layer Protocol	AODV (Ad-hoc On-demand Distance Vector)
Random Motion	Disabled
Carrier Sense Threshold	4.21756e-11
Receiving threshold	4.4613e-10
Capture Threshold	75.0
Transmission Power	0.2818
Frequency	2.4e+9
Initial Energy	500u
Transmission Power	0.9u
Receiving Power	0.5u
Idle Power	0.45u
Sleep Power	0.05u

Simulations are performed for the four design goals in order to generate the performance of the proposed STBA method. The first case is the proposed STBA method where routing is performed involving the trustworthy nodes whose trust is calculated depending on three level observations. Second goal is existing NETM method and third goal is existing BETM method where in both cases routing is performed by involving the trustworthy nodes which are classified based on only direct & indirect trust computations. The last one is the simple AODV routing protocol where routing performed involving all the available nodes without any trust computation. Below are the performance parameters used to analyze the results and efficiency of the proposed work.

*Packet Delivery Ratio:* It is defined as total number of packets received at the destination divided by the total number of packets sent from the source in the network [18][19].

*Packet Drop Rate:* It is the ratio of the total number of dropped packets divided by the total number of sent packets by the source [20].

*Malicious Node Detection Ratio:* Malicious, bad behaviour Nodes detected out of total nodes present in the network [21].

*False Positive Detection:* The ratio defined as the total count of good behavior nodes wrongly designated as malicious one's to the total count of nodes present in the network is called as 'False positive detection' [22].

*Throughput:* It refers to how much data can be transferred in the network from source to destination within a given timeframe [23].

*Delay:* Time delay taken to transfer data packets from Source to Destination [24][25].

## 4.2 Results

After performing the simulations, results are analyzed. Node's Trustworthiness is evaluated based on the proposed method, secure trust based approach STBA. This method uses three tier observations for computation of trust factor. Direct, Neighbour and Self appraisal trusts are calculated using above mentioned equations. Results obtained and calculations carried out for secure trust computation from the simulation are tabulated in Table 4.

**Table 4.** Sample secure trust value computation

Node No.	Direct Trust Calculation	Neighbor Trust Calculation	Historical Trust Calculation	Node Secure Trust Value
N0	0.92	0.4323	0.83	0.84352
N1	0.71	0.3667	0.987	0.73265
N2	0.23	0.4591	0.89	0.21477
N3	0.31	0.5238	0.91	0.24725
N4	0.12	0.4791	0.60	0.28965
N5	0.22	0.3956	0.71	0.77864
N6	0.34	0.13274	0.89	0.79326
N7	0.05	0.725	0.79	0.3231
N8	0.62	0.5571	0.89	0.65326

(Continued)

**Table 4.** Sample secure trust value computation (Continued)

Node No.	Direct Trust Calculation	Neighbor Trust Calculation	Historical Trust Calculation	Node Secure Trust Value
N9	0.81	0.13674	0.889	0.532524
N10	0.69	0.513	0.99	0.74651
N11	0.56	0.3195	1	0.72689
N12	0.43	0.3535	0.95	0.73418
N13	0.61	0.3894	0.89	0.78865
N14	0.43	0.262	0.79	0.5982
N15	0.3	0.2238	0.73	0.64714
N16	0.62	0.2748	0.84	0.65444
N17	0.75	0.519	0.81	0.6057
N18	0.76	0.5815	0.79	0.63045
N19	0.42	0.528	0.75	0.6104
N20	0.28	0.375	0.88	0.6805

The proposed STBA method identifies and isolates the malicious nodes using Node’s Secure Trust computation as shown in Table 5.

**Table 5.** Sample malicious nodes identification and isolation

Node	Node Secure Trust	Static Trust Threshold	Decision
N0	0.84352	0.6	Trustworthy
N1	0.73265	0.6	Trustworthy
N2	0.21477	0.6	Malicious
N3	0.24725	0.6	Malicious
N4	0.28965	0.6	Malicious
N5	0.77864	0.6	Trustworthy
N6	0.79326	0.6	Trustworthy
N7	0.3231	0.6	Malicious
N8	0.65326	0.6	Malicious
N9	0.532524	0.6	Malicious
N10	0.74651	0.6	Trustworthy
N11	0.72689	0.6	Trustworthy
N12	0.73418	0.6	Trustworthy
N13	0.78865	0.6	Trustworthy
N14	0.5982	0.6	Malicious
N15	0.64714	0.6	Trustworthy
N16	0.65444	0.6	Trustworthy
N17	0.6057	0.6	Trustworthy
N18	0.63045	0.6	Trustworthy
N19	0.6104	0.6	Trustworthy
N20	0.6805	0.6	Trustworthy

The interpretations are made through the evaluations of the metrics. The efficiency of proposed STBA method is demonstrated using below performance parameters.

**Packet delivery ratio.** From the simulation results, it was noted that for 100pkts/s, 47012 packets received out of 50000 packets sent, so Packet Delivery Ratio is 94.9% for the proposed STBA method, 89.1% for NETM, 87.2% for the existing BTEM where routing involved with direct & indirect trust computation, 52.9% in case of fourth design goal where routing is performed without any prior trust computation. For 200pkts/s, in case of proposed STBA method 75016 packets received out of 100000 packets sent, Packet Delivery Ratio is 76.3%, in case of NETM and BTEM, it is 75.6% and 74.2% respectively and fourth case it is 31.2%. Figure 2 depicts packet delivery ratio for all the cases. It shows how the delivery of the packets is affected through the presence of malicious nodes.

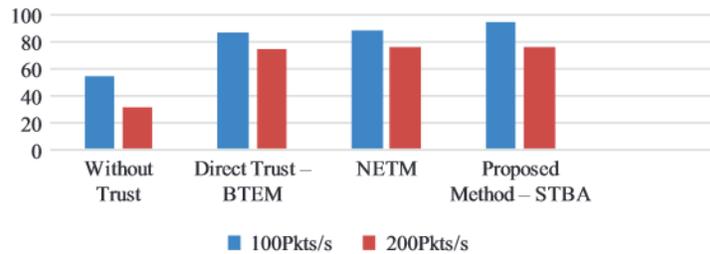


Fig. 2. Packet delivery ratio

**False positives detection ratio (FPD Ratio).** In case of proposed method for 100 packets, False Positive Detection Rate is 44%, whereas 36% for the second case NETM and 32% for third case BTEM where routing involved with direct & indirect trust. Figure 3 shows the comparison of False Positive Detection Rate of proposed method with NETM and BTEM method.

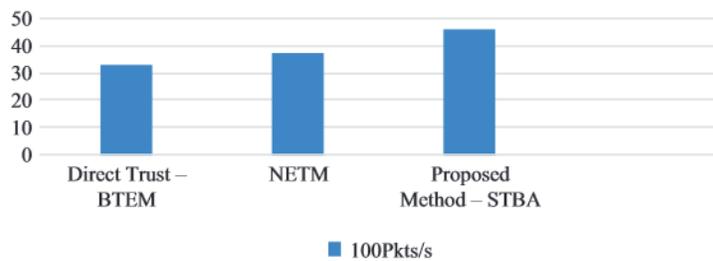


Fig. 3. False positive detection ratio of STBA

**Packet drop ratio.** Simulation results show that for 100pkts/s, 2990 packets lost out of 50000 packets sent, Packet Drop Ratio is 5.8% for the STBA method, it is 8.4% for NETM, 9.7% for the existing method BTEM where routing involved with Direct & Indirect trust computation, in case of fourth design goal, it is 43.3% where

routing is done without any trust calculation. For 200pkts/s, 24068 packets lost out of 100000 packets sent, Packet Drop Ratio is 21.124% in case of proposed STBA method, whereas for NETM it is 24.5%, BTEM it is 25.6%, and for fourth design goal it is 69.4%. Comparison of the above four cases in terms of the Packet drop ratio is shown in Figure 4.

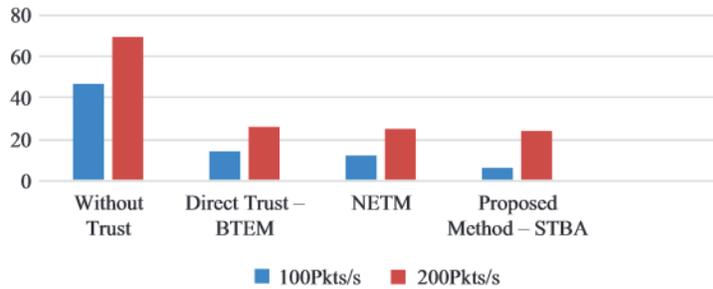


Fig. 4. Packet drop ratio

**Malicious node detection.** Malicious Node Detection rate for the proposed method is 26%, 24% for NETM and 22% for BTEM where routing is involved with direct & indirect trust. Figure 5 shows the comparison and efficiency of the proposed method in terms of malicious node detection.

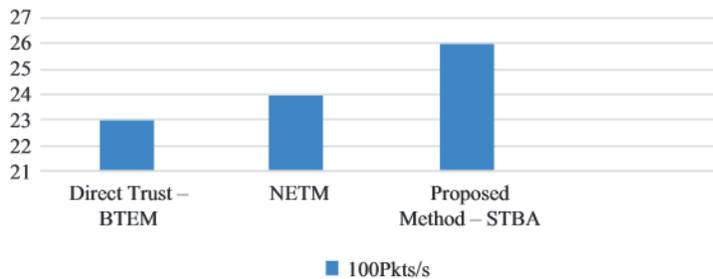


Fig. 5. Detection of malicious nodes

**Throughput.** Simulation results show that throughput for 100pkts/s is 389.1kbps for the proposed STBA method, 362.2kbps for the NETM, 356.3kbps for the existing BTEM, whereas it is 210.5kbps for the fourth design goal which involves routing without any trust calculation and in case of 200pkts/s, Throughput is 602.9kbps for proposed STBA method, whereas it is 591.4kbps, 587.6kbps, 228.3kbps for NETM, BTEM and fourth design goal respectively. Figure 6. Illustrates the throughput efficiency of the proposed method compared with other cases.

Basically, throughput shows the efficient delivery of packets. Hence, it can be interpreted that the proposed method performs very well in terms of throughput.

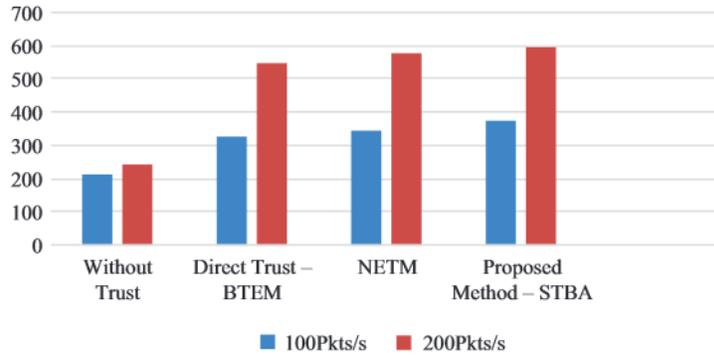


Fig. 6. Throughput comparison

**Delay.** From the results, in case of proposed method STBA, for 100pkts/s delay is noted as 192ms, 196ms for the second case NTEM, 198ms for the BTEM, where routing involved with direct & indirect trust, 221ms in fourth case where trust calculation is not done before routing and delay is 283ms for 200pkts/s in case of proposed STBA method, it is 291ms for NTEM, 293ms for BTEM and 298ms for fourth case. Efficiency of the proposed method in terms delay is shown in Figure 7.

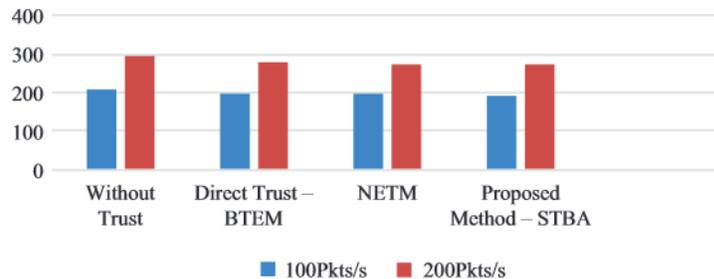


Fig. 7. Delay in milliseconds

**Discussions.** The proposed STBA method performs secure routing efficiently by evaluating trust factor for identifying the trustworthy nodes and isolating the malicious nodes. Secure trust value of a node is computed using factors, direct, indirect observations and self appraisal of the node. The proposed method STBA is compared with the existing NETM, BTEM mechanism and routing without trust calculation, simple AODV protocol. Simulation results prove the proposed STBA method is performing well. The comparison between the proposed and existing methods in terms of performance metrics are tabulated below in Table 6.

**Table 6.** Comparison of results – efficiency of the proposed method STBA

S. No	Performance Parameter	Proposed Method – STBA (%)		NETM (%)		BETM (%)		Without any Trust Calculation (AODY) (%)	
		100 Pkts/Sec	200 Pkts/Sec	100 Pkts/Sec	200 Pkts/Sec	100 Pkts/Sec	200 Pkts/Sec	100 Pkts/Sec	200 Pkts/Sec
1	Packet Delivery Ratio	94.9	76.3	89.1	75.6	87.2	74.2	52.9	31.2
2	Packet Drop Ratio	5.8	21.12	8.4	24.5	9.7	25.6	43.3	69.4
3	False Positive Detection Ratio	44		36		32		-	
4	Malicious Node Detection	24		24		22		-	
5	Throughout	389.1 kbps	602.9 kbps	362.2 kbps	591.4 kbps	356.3 kbps	587.6 kbps	210.5 kbps	228.3 kbps
6	Delay	192 ms	283 ms	196 ms	291 ms	198 ms	293 ms	221 ms	298 ms

## 5 Conclusion

From this paper, a quantitative model Secure Trust based Approach STBA is proposed to show the effective transfer of packets for communication in wireless networks using node's trustworthiness with three tier observations. The method successfully isolates the malicious nodes. This work is proved to be efficient when compared with other existing approaches like NETM and BETM where both uses hybrid observations for evolution of trustworthiness and isolation of malicious nodes. The aim is achieved through calculating the trust worthiness of the nodes and packet metrics. The appropriate results and evidences were pointed to show the effective combination of three tier observations for calculating node's trustworthiness and for secure transmission. This research can be extended in future by considering the factor of Adaptive trust threshold. The adaptive growth of the proposed model can be seen by implementing an adaptive threshold technique in place of static trust threshold factor to compare the secure trust calculated.

## 6 Acknowledgment

Special thanks to the members involved for support and knowledgeable efforts towards simulations directly and indirectly.

## 7 References

- [1] Wheeb, Ali H., and Nadia Adnan Shiltagh Al-Jamali. "Performance analysis of OLSR protocol in Mobile Ad Hoc networks." *iJIM* 16.01 (2022): 107. <https://doi.org/10.3991/ijim.v16i01.26663>
- [2] Almalkawi, Islam, et al. "A novel and Efficient priority-based cross-layer contextual unobservability scheme against global attacks for WMSNs." *iJIM* 15.03 (2021): 43–69. <https://doi.org/10.3991/ijim.v15i03.18327>
- [3] Sultan, Shahid, et al. "Collaborative-trust approach toward malicious node detection in vehicular ad hoc networks." *Environment, Development and Sustainability* (2021): 1–19. <https://doi.org/10.1007/s10668-021-01632-5>
- [4] Alnabhan, Mohammad. "Advanced GPSR in Mobile Ad-hoc Networks (MANETs)." *iJIM* 14.18 (2020): 107–131. <https://doi.org/10.3991/ijim.v14i18.16661>
- [5] Sheikh, Muhammad Sameer, Jun Liang, and Wensong Wang. "Security and privacy in vehicular Ad Hoc network and vehicle cloud computing: a survey." *Wireless Communications and Mobile Computing* 2020 (2020). <https://doi.org/10.1155/2020/5129620>
- [6] Srivastava, Vikas, et al. "Energy efficient optimized rate based congestion control routing in wireless sensor network." *Journal of Ambient Intelligence and Humanized Computing* 11.3 (2020): 1325–1338. <https://doi.org/10.1007/s12652-019-01449-1>
- [7] Dwivedi, Ashutosh Dhar, et al. "A decentralized privacy-preserving healthcare blockchain for IoT." *Sensors* 19.2 (2019): 326. <https://doi.org/10.3390/s19020326>
- [8] Elhoseny, Mohamed, and K. Shankar. "Reliable data transmission model for Mobile Ad Hoc network using signcryption technique." *IEEE Transactions on Reliability* 69.3 (2019): 1077–1086. <https://doi.org/10.1109/TR.2019.2915800>

- [9] Anwar, Raja Waseem, et al. "BTEM: Belief based trust evaluation mechanism for wireless sensor networks." *Future generation computer systems* 96 (2019): 605–616. <https://doi.org/10.1016/j.future.2019.02.004>
- [10] Syed, Salman Ali, and Ali Shahzad. "Enhanced dynamic source routing for verifying trust in Mobile Ad Hoc network for secure routing." *International Journal of Electrical and Computer Engineering* 12.1 (2022): 425. <https://doi.org/10.11591/ijece.v12i1.pp425-430>
- [11] Jamal, Tauseef, and Shariq Aziz Butt. "Malicious node analysis in MANETS." *International Journal of Information Technology* 11.4 (2019): 859–867. <https://doi.org/10.1007/s41870-018-0168-2>
- [12] Mukhedkar, Moresh Madhukar, and Uttam Kolekar. "Trust-based secure routing in Mobile Ad Hoc network using hybrid optimization algorithm." *The Computer Journal* 62.10 (2019): 1528–1545. <https://doi.org/10.1093/comjnl/bxz061>
- [13] Selvi, M., et al. "A rule based delay constrained energy efficient routing technique for wireless sensor networks." *Cluster Computing* 22.5 (2019): 10839–10848. <https://doi.org/10.1007/s10586-017-1191-y>
- [14] Abbas, Fakhar, and Pingzhi Fan. "Clustering-based reliable low-latency routing scheme using ACO method for vehicular networks." *Vehicular Communications* 12 (2018): 66–74. <https://doi.org/10.1016/j.vehcom.2018.02.004>
- [15] Ahmad, Farhan, Virginia NL Franqueira, and Asma Adnane. "TEAM: A trust evaluation and management framework in context-enabled vehicular ad-hoc networks." *IEEE Access* 6 (2018): 28643–28660. <https://doi.org/10.1109/ACCESS.2018.2837887>
- [16] Draz, Umar, et al. "Evaluation based analysis of packet delivery ratio for AODV and DSR under UDP and TCP environment." *2018 international conference on computing, mathematics and engineering technologies (iCoMET)*. IEEE, 2018. <https://doi.org/10.1109/ICOMET.2018.8346385>
- [17] Khan, Muhammad Asghar, et al. "Dynamic routing in flying ad-hoc networks using topology-based routing protocols." *Drones* 2.3 (2018): 27. <https://doi.org/10.3390/drones2030027>
- [18] Vaibhav, Akash, et al. "Security challenges, authentication, application and trust models for vehicular ad hoc network-a survey." *International Journal of Wireless and Microwave Technologies* 7.3 (2017): 36–48. <https://doi.org/10.5815/ijwmt.2017.03.04>
- [19] Arshdeep Kaur. "Vehicular Ad-hoc Network: A Survey." *International Journal of Computer Science Research* 5.2 (2017): 35–37.
- [20] Airehrour, David, Jairo Gutierrez, and Sayan Kumar Ray. "Secure routing for internet of things: A survey." *Journal of Network and Computer Applications* 66 (2016): 198–213. <https://doi.org/10.1016/j.jnca.2016.03.006>
- [21] A. Ouaddah, A. A. Elkalam, and A. A. Ouahman. "FairAccess: a new Blockchain-based access control framework for the Internet of Things." *Security and communication networks* 9, no. 18 (2016): 5943–5964. <https://doi.org/10.1002/sec.1748>
- [22] Li, Wenjia, and Houbing Song. "ART: An attack-resistant trust management scheme for securing vehicular ad hoc networks." *IEEE transactions on intelligent transportation systems* 17.4 (2015): 960–969. <https://doi.org/10.1109/TITS.2015.2494017>
- [23] Ho, Quang-Dung, et al. *Wireless communications networks for the smart grid*. Springer International Publishing, 2014.
- [24] Rahul Misra, and Prashant Sharma. "Challenges in Mobile Ad Hoc Network for Secure Data Transmission." *International Journal of Electrical & Electronics Research*, 1.1 (2013): 8–12. <https://doi.org/10.37391/IJEER.010103>
- [25] Bartoli, Andrea. "Security protocols suite for machine-to-machine systems." (2013).

## 8 Authors

**M Venkata Krishna Reddy** was born in Hyderabad, Telangana, India on 1982. He received his B.Tech in Computer Science and Engineering from JNTUH in 2005. He received his M.Tech in Computer Science and Engineering in 2009 from JNTUH. Currently, he is doing Ph.D. in the field of Mobile Adhoc Network from JNTUH. He is working as Asst, Professor in Computer Science Engineering Department, Chaitanya Bharathi Institute of Technology CBIT(A), Gandipet, Hyderabad, India. His researches focus on the trust computation for secure routing in MANETs.

**Dr. P.V.S. Srinivas** was born in Andhra Pradesh, India. He recieved his Ph.D in Computer Science and Engineering from JNTUH, Hyderabad,India. His research areas are Mobile Adhoc Networks, Machine Learning. Currently, he is working as Principal and Professor in CSE, Vignana Bharathi Institute of Technology(A), Hyderabad, India.

**Dr. M.Chandra Mohan** was born in Andhra Pradesh, India. He recieved his Ph.D in Computer Science and Engineering from JNTUH, Hyderabad, India. His research areas are Image Processing, Pattern Recognition and Software Engineering. Currently he is working as Director of Evaluation and Professor in CSE, Jawaharlal Nehru Technological University-JNTUH, Hyderabad, India.

Article submitted 2022-03-04. Resubmitted 2022-05-08. Final acceptance 2022-05-08. Final version published as submitted by the authors.



# Scalable data exchange using robust data exchange availability tactic on MQTT protocol

V. Tirupathi, K Sagar

**Keywords:** MQTT protocol, Scalable data exchange, MQTT broker, Subscriber, publisher client, Software Defined Network (SDN) controller.

## Abstract

The Internet of Things (IoT) is a network of interconnected, internet-connected objects that may gather and transmit data via a wireless network without the need for human participation. MQTT (Message Queuing Telemetry Transport) is a lightweight Internet of Things communication protocol with a publisher-subscriber messaging structure that allows for data flow across devices with complexities. Hence in this research, the technical complexities in achieving efficient data exchange in MQTT protocol is removed by using Scalable Data Exchange and Redirection of Clients in Large-scale Distributed IoT network thru MQTT protocol, in which the availability and scalability problem caused when the subscriber client could not obtain the data from the broker is removed by using Robust data exchange availability tactic which constructs a topic set table at each broker in the network. Moreover, the issue in handling the crashed MQTT broker while broadcasting the published data is eliminated by using Redirecting data flow control mechanism in which the brokers are controlled by the software-defined network (SDN) controller. Thus, the Scalable Data Exchange and Redirection of Clients in Large-scale Distributed IoT network thru MQTT protocol can be outperforming the other existing model with low delay and high throughput.

 PDF

Published

2022-12-03

Similarly, for a given item  $i$ , we can identify the top- $k$  similar items ( $neighbors(i)$ ) based on the cosine similarity between their latent factor representations:  $neighbors(i) = top - k$  items with highest similarity( $i, j$ )

Once we have identified the neighborhood of similar users or items, we can leverage their ratings or preferences to make recommendations for the target user or item.

The recommendations can be generated by considering the ratings of the neighborhood users or the preferences of the neighborhood items. Various techniques, such as weighted averaging or matrix factorization, can be used to generate accurate recommendations based on the neighborhood information.

By employing neighborhood-based collaborative filtering techniques, we can utilize the reduced-dimension  $U$  and  $V$  matrices to identify similar users or items and enhance recommendation accuracy in the presence of data sparsity challenges.

### 3.8 Evaluation Metrics

To assess the performance of the proposed semantic approach, several evaluation metrics can be used. The most commonly employed metrics in recommendation systems include recommendation accuracy, coverage, and diversity.

#### Recommendation Accuracy:

- Recommendation accuracy measures how well the proposed approach predicts or matches the actual user preferences. Common accuracy metrics include precision, recall, and mean average precision (MAP). These metrics can be calculated using the following mathematical notations:

- Precision:  $Precision = \frac{TP}{(TP + FP)}$ ,

where TP represents the true positive (correctly recommended items) and FP represents the false positive (incorrectly recommended items).

- Recall:  $Recall = \frac{TP}{(TP + FN)}$ , where FN represents the false negative (relevant items not recommended).

- Mean Average Precision (MAP):  $MAP = \left(\frac{1}{|U|}\right) * \sum(Precision(u) * Rel(u))$ , where  $|U|$  is the total number of users and  $Rel(u)$  represents the relevance of recommendations for user  $u$ .

- *Coverage*: Coverage measures the proportion of items that can be recommended. It ensures that the recommended items cover a wide range of preferences and cater to diverse user interests. Coverage can be calculated using the following mathematical notation:

- $Coverage = \frac{|Recommended\ Items|}{|Total\ Items|}$

- *Diversity*: Diversity measures the variety or dissimilarity among the recommended items. It ensures that the recommendations are not overly similar and provide a diverse set of options to users. Diversity can be quantified using metrics like the average intra-list dissimilarity or entropy.

2. *Comparison with Existing Methods*: To demonstrate the effectiveness of the proposed semantic approach in overcoming the identified challenges, a comparison with existing methods and techniques is essential. This comparison can be performed using appropriate statistical tests or performance metrics.

- *Statistical Tests*: Statistical tests, such as t-tests or ANOVA, can be used to compare the performance of the proposed approach with existing methods. These tests help determine if the differences in performance metrics are statistically significant.

- *Performance Metrics*: Performance metrics, such as precision, recall, F1 score, or RMSE (Root Mean Square Error), can be calculated for both the proposed approach and existing methods. These metrics provide a quantitative measure of how well each method addresses the challenges and helps in comparing their effectiveness.

Mathematical notations may vary depending on the specific metrics or statistical tests employed. It is crucial to select the appropriate evaluation metrics and conduct a comprehensive comparison to validate the superiority of the proposed semantic approach over existing methods in terms of recommendation accuracy, coverage, and diversity.

## 4. Result and analysis

In this study, we are solving the three major problems of recommendation systems – cold start, data sparsity and scalability by using Singular Value Decomposition.

For Cold start problem, as we discussed earlier, we employed two different methods-

Popularity based recommendations are used when there is no initial input from the user. This is an effective method which shows top rated movies by finding the mean of all the ratings of the movies.

```
movie_title
Close Shave, A (1995)          4.491071
Schindler's List (1993)       4.466443
Wrong Trousers, The (1993)    4.466102
Casablanca (1942)             4.456790
Wallace & Gromit: The Best of Aardman Animation (1996)  4.447761
...
His Girl Friday (1940)        4.000000
Babe (1995)                   3.995434
Cool Hand Luke (1967)         3.993902
Singin' in the Rain (1952)    3.992701
Patton (1970)                 3.992647
Name: rating, Length: 100, dtype: float64
```

Another method employed for solving the cold start problem is using the SVD itself. This is done by asking the user what movie he likes. According to the movie which he inputs, the SVD model recommends movies which he might like along with cosine distance value.

```
get_top_similarities('Star Wars (1977)', model)

vector cosine distance      movie title
0          0.000000          Star Wars (1977)
1          0.276419          Return of the Jedi (1983)
2          0.283642          Empire Strikes Back, The (1980)
3          0.415315          Raiders of the Lost Ark (1981)
```

Here we can see that all the movies which are recommended are Star Wars franchise movies.

For any particular movie, the 100 latent features that are calculated will remove the data sparsity problem from the system. All the movies which are not rated by a certain user is neutralized by SVD and based on singular values and the genre of the movie, the value is given for a certain movie for a certain user. Here is an example of the 100 latent feature values of the movie Toy Story.

```
model.qi[toy_story_row_idx]
array([-0.05850284,  0.10231776,  0.16862484,  0.04825799,  0.04861144,
        0.06971908,  0.03216021,  0.11469244, -0.03958442, -0.09248461,
        -0.18331583,  0.03482929, -0.17864314, -0.04088044,  0.06205576,
        -0.01882467,  0.07842878,  0.12704542, -0.10532431,  0.00241724,
        -0.17131561, -0.03630189,  0.19938462,  0.12494611,  0.01679418,
        -0.12538807, -0.18020934, -0.09805374, -0.12213447,  0.00758947,
        -0.08149168,  0.17087789,  0.04679968,  0.0649424 ,  0.00521205,
        -0.04925983,  0.05555483,  0.0960624 , -0.00108066, -0.0012403 ,
        0.09403904,  0.1449564 , -0.01077388,  0.09245831,  0.07104477,
        0.01195605,  0.05828126, -0.01175687,  0.09007754,  0.11822448,
        -0.09022341,  0.2842974 , -0.1489683 ,  0.07120596,  0.00793588,
        0.08649985,  0.02480679,  0.00702657,  0.11632491,  0.06321374,
        0.05808387, -0.00183687,  0.06547147, -0.16609839, -0.05657265,
        -0.02947632,  0.24531965,  0.08088937, -0.08294657, -0.0254696 ,
        -0.03298439, -0.05205582,  0.08526562,  0.08919324,  0.07242523,
        -0.09852793, -0.06986821, -0.00364756,  0.02618609, -0.06419099,
        -0.07037352,  0.11137158, -0.03926546, -0.04232707, -0.29685168,
        -0.00862732, -0.15178835, -0.15102777,  0.11506455,  0.1109844 ,
        0.04869982,  0.00617296,  0.06175039, -0.00363624, -0.0798074 ,
        -0.04573541, -0.10840233, -0.01023793,  0.15293672, -0.05391283])
```

The Scalability problem is solved by our model by using the SVD dimensionality reduction technique. SVD represents the user-item matrix as product of three matrices and each matrix has a certain functionality that is discussed earlier.

Finally, we calculated the root mean squared error and mean absolute error of our system and the values seem to be promising. The quality of recommendations are good according to the values of the RMSE and MAE

**RMSE: 0.9124**

**RMSE: 0.9123573080775754**

**MAE: 0.7161**

**MAE: 0.7161082532793506**

The amount of information is multiplying at an exponential rate because of the rapid development of Internet services. Most of the time, users have no idea how to get essential information more rapidly. Recommendation systems have proved to be useful in getting information and saving time searching.

In conclusion, the paper highlights the challenges faced by recommender systems in addressing the needs of users in the current digital landscape. Despite the use of collaborative filtering, problems such as cold start, data sparsity, and scalability continue to persist. The authors propose a solution that utilizes Singular Value Decomposition (SVD), a matrix factorization method that addresses the afore mentioned issues. SVD reduces the dimensionality of the data and allows the extraction of factors from the user-item-rating matrix. The proposed system offers promising results and has the potential to enhance the accuracy of recommendations in various domains.

The movie recommendation system built using SVD has several strengths and limitations. Here are some of the main ones:

#### 4.1 Strengths

1. Accuracy: SVD is known for its high accuracy in predicting user ratings. The model is capable of capturing complex patterns in the data, which leads to better recommendations.
2. Scalability: SVD can handle large and sparse datasets efficiently. This makes it suitable for recommendation systems with a large number of users and items.
3. Cold-start problem: SVD can handle the cold-start problem where new users or items have few or no ratings. The model can make predictions based on the characteristics of the user or item.

4. Interpretability: SVD provides interpretable factors that can be used to understand the relationship between users and items. This can be useful for domain experts who want to analyze the recommendations.
5. Flexibility: SVD can be customized to meet the specific needs of different recommendation systems. For example, the number of factors can be adjusted to optimize performance.

#### 4.2 Limitations

1. Data sparsity: SVD requires a significant amount of data to make accurate recommendations. When there is a high degree of data sparsity, the model's accuracy can be reduced.
2. Cold-start problem: While SVD can handle the cold-start problem to some extent, it still requires some initial data to make accurate recommendations. This can be a challenge for new systems that do not have any historical data.
3. Interpretability: While SVD provides interpretable factors, it can be challenging to interpret the meaning of each factor. The factors may not have a clear relationship with the user or item characteristics.
4. Limited feature representation: SVD is limited in its ability to represent complex user and item features. This can lead to suboptimal recommendations for systems with highly diverse user or item features.
5. Computational complexity: SVD can be computationally expensive, especially when dealing with large datasets. This can limit its scalability for some applications.

In summary, the movie recommendation system built using SVD has several strengths and limitations. While the model can provide accurate recommendations and handle the cold-start problem, it can be limited by data sparsity and computational complexity. The choice of recommendation algorithm should be based on the specific needs of the application and the trade-offs between different strengths and limitations.

Here are some potential areas for improvement of the movie recommendation system built using SVD:

1. Incorporating additional features: While SVD can handle some user and item features, incorporating additional features such as user demographics or movie genres can improve the model's performance. This can be achieved by using hybrid recommendation techniques that combine multiple algorithms.

2. Regularization: Regularization techniques can be used to prevent overfitting in SVD models. This can lead to more accurate and robust recommendations.
3. Hyperparameter tuning: The performance of the SVD model can be further optimized by tuning the hyperparameters such as the number of latent factors and regularization parameters. This can be done using cross-validation techniques.
4. Handling dynamic data: The movie recommendation system can be improved by handling dynamic data, such as new movie releases or changes in user preferences. This can be achieved by using online learning techniques that update the model in real-time.
5. Diversity of recommendations: While SVD provides accurate recommendations, it may not consider the diversity of recommendations. This can lead to a lack of serendipity in the recommendations. To address this, techniques such as diversity-based recommendation algorithms can be used.
6. Explanation of recommendations: Providing explanations for the recommended movies can improve user trust and engagement. This can be achieved by using techniques such as model interpretation or generating natural language explanations.

In summary, there are several areas for improvement of the movie recommendation system built using SVD. By incorporating additional features, regularizing the model, tuning hyperparameters, handling dynamic data, considering diversity, and providing explanations for the recommendations, the system can be further improved to provide better recommendations and user engagement.

#### 5. Conclusion

In conclusion, the implementation of Singular Value Decomposition (SVD) in the movie recommendation system has proven effective in addressing scalability, data sparsity, and cold-start problems. Through preprocessing techniques such as mean imputation and matrix factorization, we successfully handled challenges related to missing values and data sparsity. Evaluation metrics including RMSE and precision confirmed the accuracy and efficiency of the SVD model. While the approach showcased its ability to provide accurate recommendations, there are opportunities for further enhancement, such as incorporating additional features, regularization, and adapting to dynamic data. Overall, SVD offers significant implications for movie recommendation systems by addressing key challenges

and emphasizing the importance of preprocessing and evaluation metrics. Continued advancements in SVD-based recommendation systems have the potential to provide superior recommendations and enhance user engagement.

## References

- [1] Dina Fitria Murad; Rosilah Hassan; Bambang Dwi Wijanarko; Riyan Leandros; Silvia Ayunda Murad, 2022 “Evaluation of Hybrid Collaborative Filtering Approach with Context-Sensitive Recommendation System” arXiv:10.1109
- [2] Habeebunissa Begum, G.S.S Rao (2017). Associating Social Media to e-Merchandise - A Cold Start Commodity Recommendation. *International Journal of Computer Engineering In Research Trends*.4(10),378-382.
- [3] Mate Pocs, 2020 “Memory and Model based Collaborative Filtering techniques”
- [4] SongJie Gong, HongWu Ye, HengSong Tan, 2009 “Combining Memory-Based and Model-Based Collaborative Filtering in Recommender System ”
- [5] Milind M. Sutar. Tanveer I. Bagban (2017). Survey on: Prediction of Rating based on Social Sentiment. *International Journal of Computer Engineering In Research Trends*.4(11),533-538.
- [6] N.Satish Kumar, Sujana Babu Vadde (2015), Typicality Based Content-Boosted Collaborative Filtering Recommendation Framework. *International Journal of Computer Engineering In Research Trends*.2(1),809-813.
- [7] Sonule Prashika Abasaheb, Tanveer I. Bagban (2016). A Survey on Web Page Recommendation and Data Preprocessing. *International Journal of Computer Engineering In Research Trends*.3(4),204-209.
- [8] A.Avinash,N.Sujatha (2016). Location-aware and Personalized Collaborative Filtering for Web Service Recommendation. *International Journal of Computer Engineering In Research Trends*.3(5),356-360
- [9] Gladys T. Dimatacot , Katherine B. Parangat(2022). Effectiveness of Cooperative Learning On the Academic Performance in Mathematics of Junior High School Students in the Philippines. *International Journal of Computer Engineering In Research Trends*.9(2),51-58.
- [10] Kumar, P. ., Gupta, M. K. ., Rao, C. R. S. ., Bhavsingh, M. ., & Srilakshmi, M. (2023). A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(3s), 184–192. <https://doi.org/10.17762/ijritcc.v11i3s.6180>
- [11] Ramana, K. V. ., Muralidhar, A. ., Balusa, B. C. ., Bhavsingh, M., & Majeti, S. . (2023). An Approach for Mining Top-k High Utility Item Sets (HUI). *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(2s), 198–203. <https://doi.org/10.17762/ijritcc.v11i2s.6045>
- [12] Yu, K., Zhu, S., Lafferty, J., & Gong, Y. (2009, July). Fast nonparametric matrix factorization for large-scale collaborative filtering. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 211-218).
- [13] Salakhutdinov, R., Mnih, A., & Hinton, G. (2007, June). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning* (pp. 791-798).
- [14] Wang, H., Wang, N., & Yeung, D. Y. (2015, August). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1235-1244).
- [15] Xue, H. J., Dai, X., Zhang, J., Huang, S., & Chen, J. (2017, August). Deep matrix factorization models for recommender systems. In *IJCAI* (Vol. 17, pp. 3203-3209).
- [16] Zhang, S., Yao, L., & Xu, X. (2017, August). Autosvd++ an efficient hybrid collaborative filtering model via contractive auto-encoders. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 957-960).
- [17] Ouyang, Y., Liu, W., Rong, W., et al. (2014). Autoencoder-based collaborative filtering. In *Int. Conf. on Neural Information Processing* (pp. 284-291). Kuching, Malaysia.
- [18] Sedhain, S., Menon, A.K., Sanner, S., et al. (2015). Autorec: autoencoders meet collaborative filtering. In *Proc. 24th Int. Conf. on World Wide Web* (pp. 111-112). Florence, Italy.
- [19] Wu, Y., DuBois, C., Zheng, A.X., et al. (2016). Collaborative denoising auto-encoders for top-n recommender systems. In *Proc. of the Ninth ACM Int. Conf. on Web Search and Data Mining* (pp. 153-162). San Francisco, California, USA.
- [20] Yan, W., Wang, D., Cao, M., et al. (2019). Deep auto encoder model with convolutional text networks for video recommendation. *IEEE Access*, 7, 40333-40346.

- [21] Strub, F., Gaudel, R., Mary, J. (2016). Hybrid recommender system based on autoencoders. In Proc. of the 1st Workshop on Deep Learning for Recommender Systems (pp. 11-16). Boston, MA, USA.

**How to Cite:**

Kumar, G. K., Gopalachari, M. V., Mishra, R., Jayaram, D., Rakesh, S., & Rani, D. M. (2022). Enabling effective location-based services for road networks using spatial mining. *International Journal of Health Sciences*, 6(S4), 5174–5188.  
<https://doi.org/10.53730/ijhs.v6nS4.9313>

## **Enabling effective location-based services for road networks using spatial mining**

**G Kiran Kumar**

Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**M Venu Gopalachari**

Associate Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India  
Corresponding author email: [mvenugopalachari\\_it@cbit.ac.in](mailto:mvenugopalachari_it@cbit.ac.in)

**Rupesh Mishra**

Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**D Jayaram**

Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**S Rakesh**

Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**D. Malathi Rani**

Assistant Professor, ECE Department, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India

**Abstract**---A co-location pattern represents a subset of Boolean spatial attributes whose instances are located in a close geographic space. These patterns are important for location-based services. There are many methods for co-location pattern mining where the distance between the events in close geographic proximity is calculated using a straight-line distance called Euclidean distance. Since most of the real-time tasks are bounded to the road networks, the results computed using Euclidean distance is not appropriate. So to compute co-location patterns involving network we define a model where initially a network model is defined and the neighbourhood is obtained by using network distance. By comparing this approach with the previous Euclidean approach, the results

obtained for co-location patterns on a road network are accurate. Our experimental results for synthetic and real data show that the proposed approach is efficient and accurate for identifying co-location patterns involving network entities.

**Keywords**---spatial data, co-location patterns, road network distance, decision making, SVM, decision tree, random forest, Naïve Bayes.

## Introduction

Co-location Pattern for Mining Spatial data (CPMS) is used to identify close proximity objects based on the geographical data Wang (2018). Complex associations are generally present in geographical data and knowledge extraction from both spatial and non-spatial attributes can be performed through CPMS. Location-based service (LBS) is a mobile service which uses wireless internet. These are capable of providing geographic information based on the current location. The information about surroundings with respect to a given position has become important to solve many real-time issues.

Any data is said to be spatial if it contains location entities. Extracting the close proximity geographic objects is possible through CPMS. The closeness between the objects can be calculated using various distance metrics like Euclidean distance and path distance Yao (2018). Usually, the distance between the objects is calculated using straight line distance called Euclidean distance and is ill-suited because human activities are mainly constrained to road networks. So the distance measure used to calculate the road networks is the network distance.

Co-location rule mining has got many applications across domains like location-sensitive advertisements, genetics, transportation, tourism, ecology, weather prediction, and crime prediction Lu (2018). For example, a mobile service provider who has data of service patterns used by geographically nearby users and can provide location sensitive offers which boost his sales. Of late, a genetic field of biology is using the co-location mining for predicting the structures of DNA.

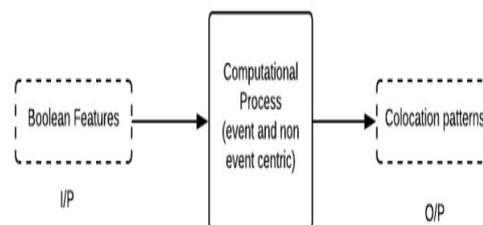


Figure 1 The high-level view of Spatial Mining

Other examples are the prediction of rainfall in a particular area based on climatic conditions of that area and other neighbouring areas of it and prediction of a possible crime in an area based on a series of events occurring in that area over a period of time. The high-level flow of CPMS is shown in Fig. 1.

The following terminology has been used throughout this paper.

- **Network:** A network is a weighted graph with a set of nodes  $N$  and a set of edges  $E$ . Represented as  $G=(N, E)$ .
- **Network Edge:** Any two nodes which have direct communication can be expressed as network edge. It can be created between the start node and the ending node. Multiple nodes can even be configured on the network edge based on the object distance from the starting node.
- **Network object:** A network object is expressed as  $(d, k)$  where  $d$  is the nearest edge and  $k$  is the distance from the starting node.
- **Network distance:** Network distance is defined as the shortest path distance from one object to another. For example, if we consider the objects which lie on the same edge then the distance between the two objects can be calculated and identified as the shortest distance between the objects.
- **Neighborhood:** All the objects within the defined threshold are treated as the neighborhood to a target object. This measure is required to find all the co-located objects in the geographical area. Dijkstra's algorithm is one of the most popular methods for finding the shortest path. This gives the optimal solution for the given dataset.

### **Related Work**

Based on the extensive literature, we prepared a taxonomy for identifying co-location patterns (Fig. 2.). Multiple researchers have worked on spatial co-location mining. Many of these past works are based on the Euclidean distance Yu (2017), Huang (2004), Cai (2018), Yoo (2004). However, most of the human activities depend on network distance Okabe (2009), Yamada (2007). The economic issues can be addressed using the network distance Lu (2017), Yu (2017).

In Agrawal (1994), the authors formulated the rules for generating advanced frequent patterns using the support and threshold. The authors in Koperski (1995) introduced the idea of spatial association mining in which they focused on certain spatial objects Flouvat (2015). But, these approaches may not contain the required reference feature. To address this, recent researches focused on CPMS.

CPMS is used for several applications especially in designing, natural investigations, business, and transportation Shekhar (2001). In another study, an object-centric model was proposed in Huang (2006) for co-location pattern mining which is based on clique instances rather than reference feature along with introducing a measure named participation index. But, this method has a lot of instances join operations, which increases the complexity.

To minimize this number of join operations, the authors in Yoo (2004) proposed different approaches. Firstly, they proposed an algorithm based on a plane dividing approach. Even though this approach performs better, still there are

noticeable instance join operations. In recent studies, this algorithm is further extended to make it more efficient Flouvat (2015), Philips (2012).

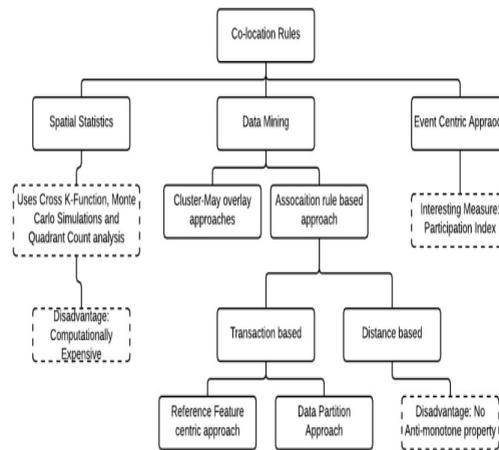


Figure 2. A proposed taxonomy for generating co-location patterns

In this paper, we propose to use a road network distance for CPMS. Instead of using the Dijkstra algorithm, we use A\* search algorithm for finding shortest paths which improve the performance. A\* algorithm is a widely used approach in path finding and graph traversal techniques.

The fundamental component of spatial data mining is based on how neighborhood relation is defined. Further, there are also many other approaches in the literature to discover the spatial co-location mining like buffer zones Appice (2003), topological connections Santos (2005), kNN algorithm Wan(2009). The utilization of network distance is not a new aspect of GIS. Many researchers Huang (2004), Cai (2018)Xie (2008), used network distance for CPMS but authors in Yu (2015) used a new interesting measure and their results are more convincing and accurate.

A format sheet with the margins and placement guides is available as both Word and PDF files as <format.docx> and <format.pdf>. It contains lines and boxes showing the margins and print areas. If you hold it and your printed page up to the light, you can easily check your margins to see if your print area fits within the space allowed.

### Proposed Approach for Identifying Co-Location Patterns

The proposed design for identifying co-location patterns is shown in Fig. 3. In the previous studies, the distance between the two objects is calculated using Euclidean distance. Since most of the human activities are constrained only to the road networks, the Euclidean model is ill-suited. To overcome this drawback, a new distance metric called network distance was used. This distance metric is used to calculate the distance between the two objects through a network. However, there are very few studies which focused on using road network

distance for CPMS. We use the existing framework of CPMS and customize it to use participation index as an interesting measure.

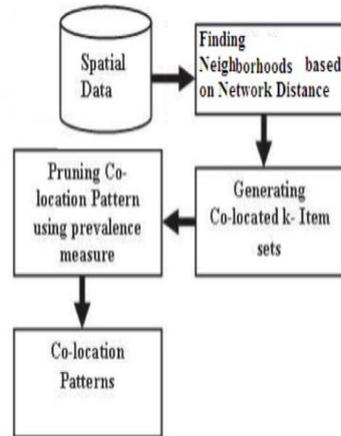


Figure 3 Proposed Architecture for CPMS (Co-location Patterns through Mining spatial data)

In this method, the shortest distance between the two objects located in close geographic proximity is considered. The algorithm used for calculating the shortest path is A\* search algorithm. In this context initially, the objects located in geographic proximity are identified. Later using the A\* search algorithm the shortest path from each object to another object is obtained. Based on the obtained distances a threshold is fixed.

The objects within the given threshold tend to form a pattern called the co-location pattern. Based on the objects, co-location patterns of various sizes will be obtained. This obtained co-location patterns will be useful for users in their trip planning, business decisions etc. Generally, the co-location pattern mining in spatial data follows these steps: Initially, a neighborhood relation graph is built based on the predefined neighborhood. Next, the clique instances (co-location instances) will be generated. Finally, the relevant patterns are obtained (i.e. pruning) by using a threshold. The major variation in our paper is that we focused on creating the neighborhood relation graph based on network distance.

### Experimental Results

Two different datasets have been considered in the experimentation, one is a synthetic data and other is a real-time data (Kukatpally area description). This approach analysed on a real data i.e., KPHB colony (Asia's largest colony) data to assess the performance and the details are as follows: In this, we considered all ATMs, shopping malls, hospitals, and banks. The data set consists of 107 banks, 238 ATMs, 49 Hospitals and 18 shopping malls (Fig. 4). We find latitude and longitude for each object and applied the concept of network distance where the equivalent weighted graph is constructed. We applied the A\* algorithm for identifying the shortest distance to find co-location patterns.

In urban areas, the services and location of the objects are tied to road networks. So to represent this, we use nodes, edges, network object, network distance, neighborhood, participation ratio, and participation index. Initially, the network distance from each event to another event is calculated and stored. Now threshold is fixed up to certain units. Based on the threshold the co-location patterns based on road networks will be obtained. Based on the threshold the co-location patterns of various objects will be obtained. The participation ratio and participation index are calculated for the obtained co-location patterns. A higher value of participation ratio indicates that the identified co-location patterns are strongly correlated. Fig. 5. illustrates the object representation in the spatial reference system.

If we consider the objects which lie on the same edge then the distance between them is considered as the shortest one. On the other hand, in Fig. 5., objects A1 and D3 lie on different edges the network distance between them can be calculated by using the Dijkstra's algorithm i.e. Network distance (A1, D3) = A1.pos + networkdistance(n7, n7) + D3.pos. Substituting the corresponding values we get 7.1 (3.8+0+3.3). Fig. 6 shows the co-location instances obtained, participation ration and participation index which are obtained after the complete execution of the code. Higher participation index indicates that the co-located patterns occur frequently within the threshold.

The following measures have been considered to access the performance of the system.

- Participation Ratio (PR): Participation ratio can be defined as the ratio of the number of distinct objects that are included in co-location instances to the total number of objects of the type (Equation (1)).  
Participation Ratio = (number of distinct objects) / (total number of objects of the type)
- Participation Index (PI): Participation index can be defined as the minimum values of participation ratio.

We tested the proposed approach for multiple threshold values. Taking 100 meters as the threshold, the results obtained are shown from Fig. 7 to Fig. 11. We extended our approach for other threshold values and the same are shown at the end of the section.

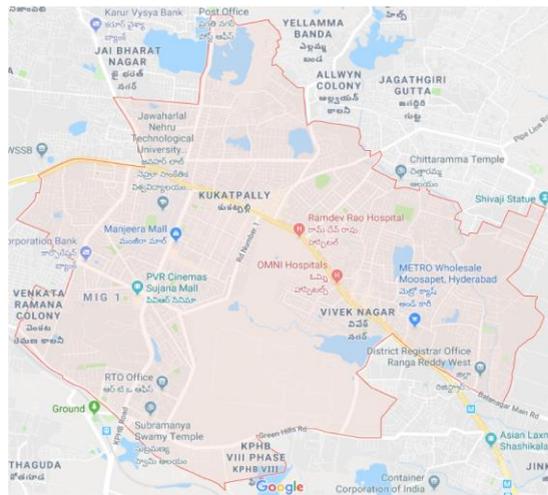


Figure 4 Map overview of the considered Real-Time Data Set

Applying the Proposed Approach on Synthetic Data set

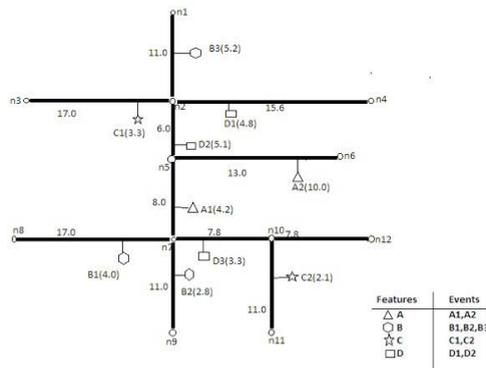


Figure 5 Linear referencing based Network Model

Fig 7 shows that the participation index for 'C, D' co-location patterns is very high. This indicates that the co-location patterns occur frequently in an area of 9 units. Applying the Proposed Approach on Real-Time Data Set Fig. 8 shows that the co-location patterns for ATMs and Banks is very high, which indicates that ATMs and Banks occur frequently within an area of 100m in Kukatpally area.

Co-location instances with distance threshold=9units					
Network co-location	{A,B}	{A,D}	{B,D}	{C,D}	{A,B,D}
Co-location instances	{A1,B1}	{A1,D2}	{B1,D3}	{C1,D1}	{A1,B1,D3}
	{A1,B2}	{A1,D3}	{B2,D3}	{C1,D2}	
				{C2,D3}	
Participation Ratio	1/2,2/3	1/2,2/3	2/3,1/3	1,1	1/2,1/3,1/3
Participation index	1/2	1/2	1/3	1	1/3

Figure 6 Co-location patterns for sample data set in Fig.1.

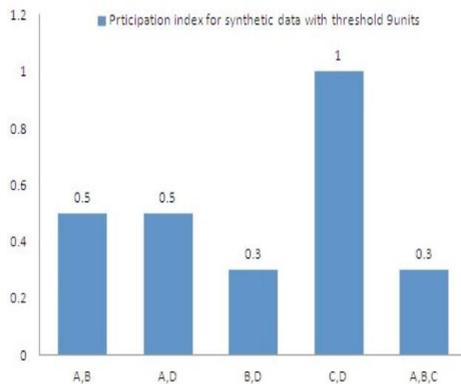


Figure 7 Participation index for co-location patterns in a synthetic dataset with threshold 9 units

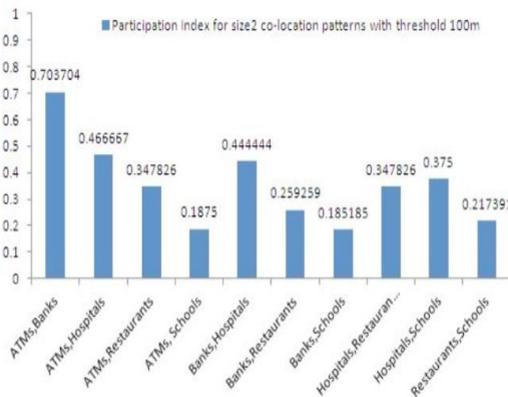


Figure 8 Participation index for size2 co-location patterns with threshold 100m

Fig. 9 shows that the co-location patterns for ATMs, Banks, and Hospitals are very high, which indicates that ATMs, Banks, and Hospitals occur frequently within an area of 100m in Kukatpally area.

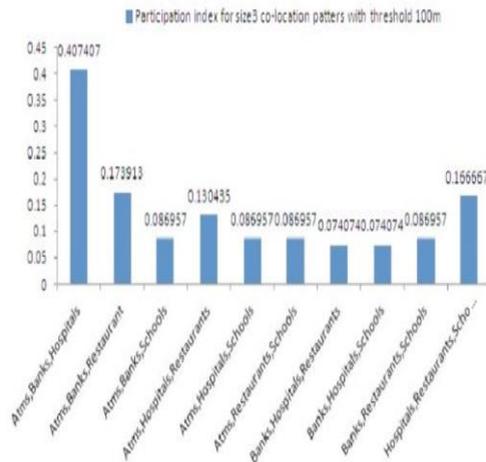


Figure 9 Participation index for size3 co-location patterns with threshold 100m

Fig. 10 shows that the co-location patterns for ATMs, Banks, Restaurants, and Schools is very high, which indicates that ATMs, Banks, Restaurants, and Schools occur frequently within an area of 100m in Kukatpally area. Fig. 11 shows that the co-location patterns for ATMs, Banks, Hospitals, Restaurants, and Schools are very high, which indicates that ATMs, Banks, Hospitals, Restaurants, and Schools occur frequently within an area of 100m in Kukatpally area.

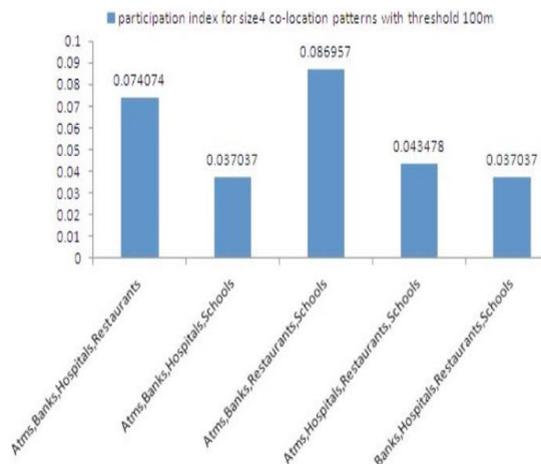


Figure 10 Participation index for size 4 co-location patterns with threshold 100m

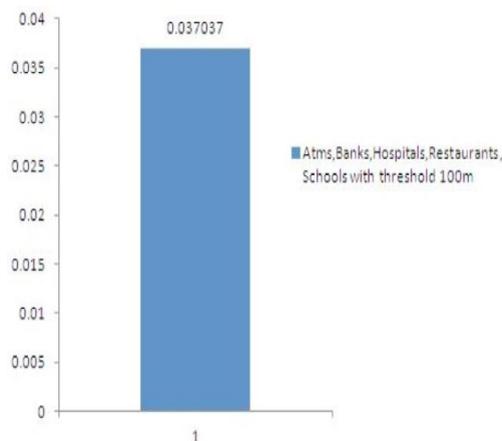


Figure 11 Participation index for size 5 co-location patterns with threshold 100m

We tested the proposed model using multiple spatial data sets with varying thresholds. We observed that when the threshold value is increasing, the probability of occurring the co-location patterns is increasing to a certain limit and later, it was almost constant (Table 1 and Table 2). The co-location patterns obtained for various threshold values are shown in Fig. 12.

Table 1 Co-location patterns with varying threshold values (Size=2)

Co-Location Patterns	Threshold=300m	Threshold = 500m
ATM, Banks	0.7407	0.777
ATM, Hospitals	0.8666	0.9333
ATM, Restaurants	0.8260	0.9565
ATM, Schools	0.562	0.8750
Banks, Hospitals	0.740	0.8518
Banks, Restaurants	0.5900	0.7777
Banks, Schools	0.5666	0.6255
Hospitals, Restaurants	0.800	0.933
Hospitals, Schools	0.8125	0.9375
Restaurants, Schools	0.500	0.625

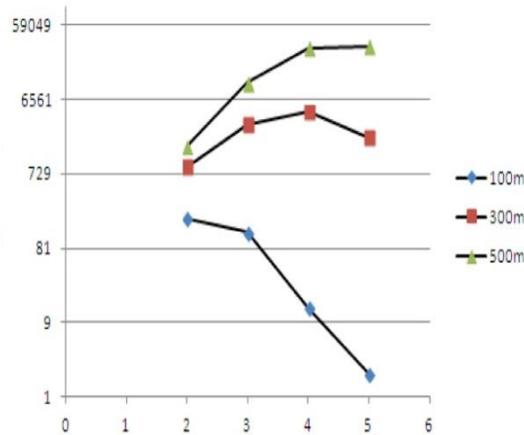


Figure 12 Various size co-location instances obtained for different threshold values

Table 2 Co-location patterns with varying threshold values (Size=3)

Co-Location Patterns	Threshold = 300m	Threshold = 500m
ATM, Bank, Hospitals	0.7077	0.7777
ATM, Bank, Restaurants	0.5925	0.7407
ATM, Bank, Shopping Malls	0.5625	0.625
ATM, Hospitals, Restaurants	0.7391	0.9333
ATM, Hospitals, Shopping Malls	0.5625	0.625
ATM, Restaurants, Shopping Malls	0.5	0.025
Bank, Hospitals, Restaurants	0.5925	0.7407
Bank, Hospitals, Shopping Malls	0.5625	0.625
Bank, Restaurants, Shopping Malls	0.500	0.625
Hospitals, Restaurants, Shopping Malls	0.500	0.625

We extend the proposed approach to identify the colocation patterns for each individual and aggregate them to predict the crowd in the target place. We consider each human activity on daily basis with respect to spatial and temporal aspects. We proposed an approach (Algorithm 1) to achieve the same.

Algorithm 1: Crowd prediction in the target locations from extracted individual spatiotemporal patterns

Input: User public activities recorded in the form of

$f(x,y,t)$ .

Output: Crowd prediction {High, Medium, Low} in the target location with respective to time limits.

1. For each user  $U_i$
2. Identify the source of tracking the user activities
3. Store and pre-process the collected data in the Time Range  $T_w$

4. Generate a graph with locations  $L_i[]$  as nodes and edges represent the sequence followed to visit locations in  $L_i[]$ .
5. End For
6. Aggregate the individual user activity graphs using spatiotemporal clustering techniques.
7. For Each path  $P_i$  in Aggregated Graph in TW
8. For Each item  $I_x$  in  $P_i$
9. Find the probability of  $I_x$  by considering  $I_{x-1}$  node for all  $P_j$
10. End For
11. If  $\text{Prob}(I_x) < \text{Thd}$
12. Purge that node from  $L_i[]$
13. End If
14. End For
15. Apply supervised machine learning algorithms for predicting the crowd at the Target location(s)

We initially identify the source of tracking the user activities like cab trajectory, self-agreed geographic and mobile tracking IAdamu (2018). We then record the sequence of locations that every user visits and all this information is aggregated using spatiotemporal clustering techniques. We purge the nodes from the aggregated graph which are not meeting the defined threshold. Finally, we apply supervised machine learning algorithms to predict the crowd at a particular location.

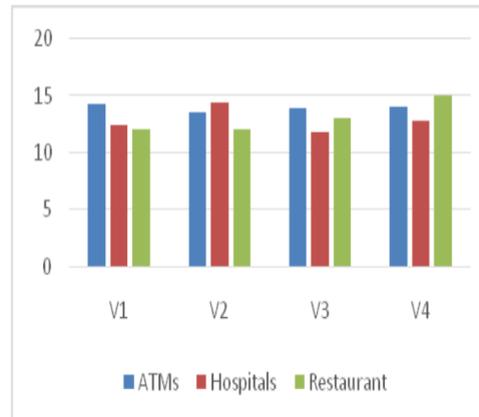


Figure 13 Identified Hotspots for multiple users across multiple visits

We tracked multiple user visits for several locations over a period of time and identified hotspot locations if they are meeting a certain threshold (Figure 13). Then, we identified multiple users visiting sequence containing the hotspots (Figure 14). We then have taken a new user sequence to predict the crowd at the next location that he/she is supposed to visit based on the hotspot visiting sequence. The classification accuracy after applying various supervised learning algorithms is shown in Figure 15. We further improved the classification accuracy by considering the various constraints like less congestion path, good road path and shortest path (Figure 16). We identified that the user visiting sequence

greatly varies based on the listed constraints by which crowd prediction accuracy at the target location is also improved.

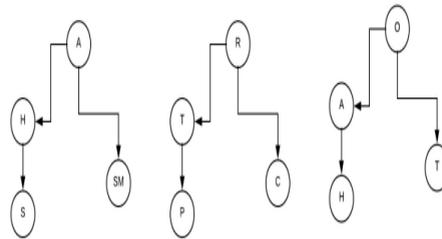


Figure 14 Identified user visiting sequence-based on the hot spots

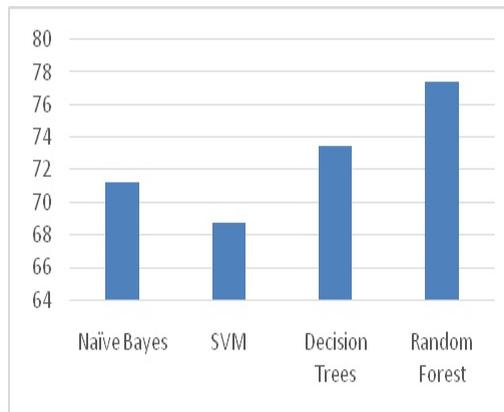


Figure 15 Crowd Prediction accuracy without constraints

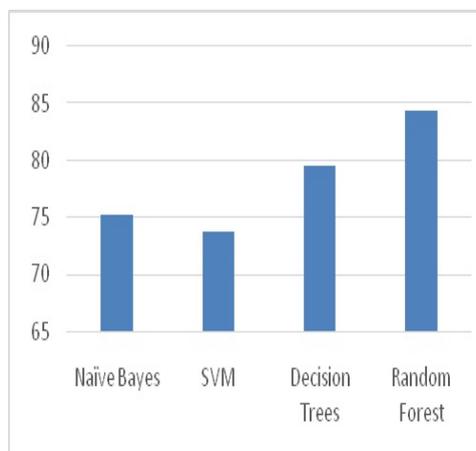


Figure 16 Crowd Prediction at the target location with constraints

## Conclusion

Spatial Co-Location Mining is an emerging area and has various real-time applications. We identified that the traditional distance measures cannot be directly used for identifying frequent co-location instances in the target geographical data. We evaluate the effect of network distance to extract the co-location patterns for the spatial data. We customized the existing CPMS model and added a new interesting measure to it. We tested the approach on multiple data sets (synthetic and real) and observed that the observed results are accurate for various co-location patterns. Since most of the human activities are constrained to road networks, this model will be useful. We extended the proposed approach to predict the crowd in target location based on the users visiting sequence based on time. As future work, we can apply the proposed approach for larger areas and test the accuracy of it.

## References

- [1] Agrawal, R., and Srikant, R. (1994). Fast algorithms for mining association rules. In Proceedings of the 20th International conference on very large databases (VLDB'94), 635 (pp. 487–499).
- [2] Appice, A., Ceci, M., Lanza, A., Lisi, F. A., and Malerba, D. (2003). Discovery of spatial association rules in geo-referenced census data: a relational mining approach. In Proceedings of the intelligent data analysis (pp. 541–566).
- [3] Cai, Jiannan, et al. Adaptive detection of statistically significant regional spatial co-location patterns. *Computers, Environment and Urban Systems* 68 (2018): 53-63.
- [4] Flouvat, F., Soc, J. N. V., Desmier, E., and Selmaoui-Folcher, N. (2015). Domain-driven co location mining. *Geoinformatica*, 19(1), 147–183.
- [5] Huang, Y., Shekhar, S., and Xiong, H. (2004). Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(12), 1472–1485.
- [6] Zuang, Y., Pei, J., and Xiong, H. (2006). Mining co-location patterns with rare events from spatial data sets. *Geoinformatica*, 10(3), 239–260.
- [7] Adamu, Joshua T. Jegede, Hilary I. Okagbue, and Pelumi E. Oguntunde. Shortest Path Planning Algorithm - A Particle Swarm Optimization (PSO) Approach Patience Proceedings of The World Congress on Engineering 2018, pp19-24.
- [8] Koperski, K., and Han, J. (1995). Discovery of spatial association rules in geographic information databases. In Proceedings of 4th international symposium on large spatial databases (pp. 47–66). Portland, Maine Springer
- [9] Lu, Junli, et al. "Mining strong symbiotic patterns hidden in spatial prevalent co-location patterns." *Knowledge-Based Systems* (2018): 190-202.
- [10] Okabe, A., Boots, B., Sugihara K., and Chiu, S. N. (2009). *Spatial tessellations: concepts and applications of Voronoi diagrams*. Chichester, UK: Wiley.
- [11] Phillips, P., and Lee, I. (2012). Mining co-distribution patterns for large crime datasets. *Expert Systems with Applications*, 39(14), 11556–11563.
- [12] Wang, Lizhen, et al. "Effective lossless condensed representation and discovery of spatial co-location patterns." *Information Sciences* 436 (2018): 197-213.

- [13] Wan, Y., and Zhou, J. (2009). KNFCOM-T: A k-nearest features-based co-location pattern mining algorithm for large spatial data sets by using t-trees. *International Journal of Business Intelligence and Data Mining*, 3(4), 375–389.
- [14] Xie, Z., and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 35(5), 396–406.
- [15] Yao, Xiaojing, et al. "A spatial co-location mining algorithm that includes adaptive proximity improvements and distant instance references." *International Journal of Geographical Information Science* 32.5 (2018): 980-1005.
- [16] Yu, Wenhao, et al. Spatial co-location pattern mining of facility points-of-interest improved by network neighborhood and distance decay effects. *International Journal of Geographical Information Science* 31.2 (2017): 280-296.
- [17] Yamada, I., and Thill, J. C. (2007). Local indicators of network-constrained clusters in spatial point patterns. *Geographical Analysis*, 39(3), 268–292.
- [18] Yu, W., Ai, T., and Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography*, 45, 32–47.
- [19] YYoo, J. S., and Shekhar, S. (2004). A partial join approach for mining co-location patterns. 712 In *Proceedings of the 12th annual ACM international workshop on geographic information systems* (pp. 241–249).
- [20] Yao, Xiaojing, et al. "Efficiently mining maximal co-locations in a spatial continuous field under directed road networks." *Information Sciences* 542 (2021): 357-379.
- [21] Wang, Lizhen, Yuan Fang, and Lihua Zhou. "Dominant Spatial Co-location Patterns." *Preference-based Spatial Co-location Pattern Mining*. Springer, Singapore, 2022. 167-199.
- [22] Dong, Shangjia, et al. "Modest flooding can trigger catastrophic road network collapse due to compound failure." *Communications Earth & Environment* 3.1 (2022): 1-10.
- [23] Rinaritha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
- [24] Rinaritha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
- [25] Arianggara, A. W., Baso, Y. S., Ramadany, S., Manapa, E. S., & Usman, A. N. (2021). Web-based competency test model for midwifery students. *International Journal of Health & Medical Sciences*, 4(1), 1-7. <https://doi.org/10.31295/ijhms.v4n1.380>

**How to Cite:**

Jayaram, D., Gopalachari, M. V., Rakesh, S., Sai, J. S., & Kumar, G. K. (2022). Fake face image detection using feature network. *International Journal of Health Sciences*, 6(S5), 3027–3039. <https://doi.org/10.53730/ijhs.v6nS5.9310>

# Fake face image detection using feature network

**D Jayaram**

Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**M Venu Gopalachari**

Associate Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

Corresponding author's Email: [mvenugopalachari\\_it@cbit.ac.in](mailto:mvenugopalachari_it@cbit.ac.in)

**S. Rakesh**

Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**J Shiva Sai**

Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**G Kiran Kumar**

Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

**Abstract**--In the recent times, the image data in social networks such as Instagram, Whatsapp, Facebook, Snapchat, twitter etc has an exponential growth in terms of volume, variety due to the velocity of the data stream. On the other hand, the advancements in the image and video processing led to increase in the fake images relatively in huge volumes. Due to the involvement in spreading fake news and leading mob incitements fake images became major concern to handle that demands an efficient a fake image detector is of at most concern to entire social networking organizations. In this paper, a deep learning framework is proposed that differentiates fabricated parts of the image from the real image using supervised learning strategies. Also a modified neural network structure called the Fake Feature Network is proposed in this work which consists of advanced convolution networks. In order to make model effective, the proposed methodology has a two major steps in learning which combines a modified neural structure that uses classifier and pairwise learning for the fake image detection. The performance of

the model is enhanced the contrastive loss to learn the common fake features using pairwise learning, which is achieved by incorporating Siamese neural network.

**Keywords**---Fake face image detection, Feature Network, Social Network Data, Deep learning, Pairwise learning.

## 1. Introduction

In this technological era, social media networks playing good role in human's life in all aspects. Over the years we have seen social media take different forms than it was created for. One of the problematic forms of social media is spreading fake information and misinformation. People started using social media as a tool to spread fake information to exploit others or damage their reputation. One of the main ways of doing this was creating fake images or videos. The major contributors of creating fake images are Generative Adversarial Networks aka GANs.

GANs are deep learning based generative models that have been widely used to create partial or whole realistic images or videos. These neural networks have been progressing every year creating more sophisticated networks like BigGAN, CycleGAN and PGGAN, which create highly realistic images. They also have the capability to create synthesized speech videos, which was used in the past to sabotage several politician's credibility.

Current forensic techniques require an expert to analyze the integrity of an image. Since the fake images created are highly sophisticated and almost impossible to detect a system is needed which is on par to the latest GANs models to detect them. A model using customized advanced neural networks to determine fake images has been implemented. A modified neural network structure called Fake Feature Network (FFN) has been proposed, which consists of Convolution Neural Networks, DenseNETs and Siamese Neural Networks. The model created will be trained using pairwise learning strategy to obtain an efficient model.

## 2. Related Work

There exists different kind of tools and techniques to handle digital image forensics in the literature [10,11]. In order to justify the image authenticity majority of the approaches focus on the format, Meta data of the image data. Though the advancements in this category of tools and techniques targeting fake images, detecting the fake images remains as a challenge due to the availability of the same advancements of the image processing techniques for the attacker. Few reputed social networking organizations such as facebook and Microsoft are sponsoring the fake detection challenge program in recent time to tackle this challenge [12].

One of the traditional approaches for fake image detection is frequency domain analysis that focuses on the compression data of the image that reveals different values for manipulated images. The fake images of faces usually tend to contain

anomalies when compared to real ones in frequency domain and the experiments in the literature with deep fake image detector had exhibit good results on the image spectrum [13]. Due to the existence of smooth and sophisticated edges in the image the frequency domain based fake image detectors could not perform upto the mark. To address these limitations in this category, few techniques such as JPEG Ghost proposed in the literature [14], in which the JPEG quality of different regions on the same image identified and verified. If the JPEG quality of all regions is not same, then the image assumed to be fake as it was recreated by incorporating another image region. But if the forged image maintains the same JPEG quality on all regions of the image, JPEG Ghost like techniques will not perform well.

On the other hand few researchers focused on the texture features of the image to differentiate fake from real images. However, designing a fake image detector based on the texture information seems critical as the sole global texture information could not differentiate fake images from real images [10]. As advancement in this approach gram matrix is introduced to capture texture information combined with Resnet [15]. If the images manipulated are of format JPGE, which is a lossy format, one can easily recognize the regions manipulated using Error Level Analysis. JPGE images maintains stable areas throughout the image which have minimum error. Any modification done to a picture will result in altering of stable areas. ELA tries identifies the areas manipulated since they are no longer at their minimum error level. By analysing the patter of the images after ELA once can determine which parts of the image are possibly faked. One can integrate machine learning to detect the anomalies in the error level analysed images too. If the different regions of the image are manipulated then this approach suits better whereas error level analysis is not efficient for GAN images.

Deep learning is also another latest approach in handling fake face detection techniques. Convolution Neural networks is the popular deep learning technique spread over various domain also can be used for face forgery methods such as FaceForensics[3], in which convolution and max pooling are the two major steps driving the supervised learning process. In this category of methods, self-attention generative adversarial networks are other technique which generates fake face data set which is refined [4].

In [6] and [7] authors proposed LSGAN (Least Square Generative Adversarial Networks) which improves the training model performance for fake image detection by supplying additional features such as noise features which are generated by steganalysis, as input to the CNN model. However, if the images have smooth edges then due to the challenges in extracting the noise feature, this kind of approaches cannot perform well for fake image detection. Few approaches in the literature such as MMC image forensic tool uses JPEG quality and metadata features [9]. Fotoforensics is the other tool that uses SNGAN (Spectral Normalization for Generative Adversial networks for fake image detection [8]. But the challenge lies with the images provide by GAN technique as well if the meta data itself is modified or hidden by the attacker, these kind of tools fail to identify fake images.

In [5], a self-consistency based learning method on social network data is proposed in which it justifies whether a single imaging pipeline could generate the content or not that requires the EXIF meta data. However this approach is also sensitive to the attacks on meta data of the fake images. The model proposed by “Fake image detection using machine learning” [2] To detect GANs generated images uses shallow networks instead of deep neural networks. They also use an adversary model to delete the entire metadata. The main limitation of their model is that they manually created fake face images to create the model, which is quite time consuming.

### 3. Proposed FFN based Model

The data is first preprocessed where the images are reshaped to 64 x 64 pixels resolution. Pairs of images from these reshaped images are made which are necessary for pairwise learning in Siamese networks. The Siamese network is built using the Fake Feature Network as its sister networks. The Siamese network is trained using contrastive loss function. Finally testing and validation is done using various validation methods. The working has been represented as shown in Fig. 1.

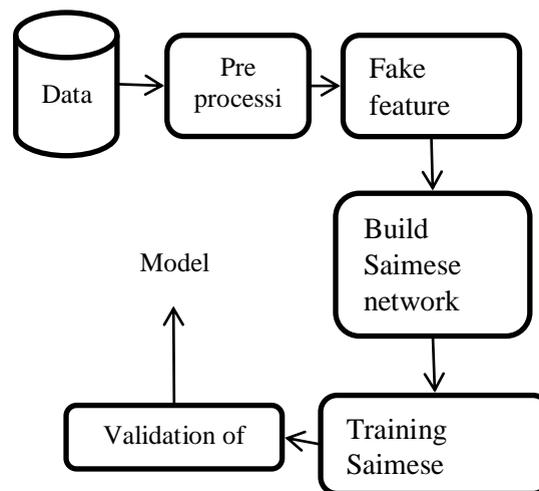


Figure 1 The architecture of the proposed model

#### A. Processing data

The images from the dataset were loaded to the model. Using an inbuilt function which images are loaded from the directory. For this the directory was organized in such a way that it consisted of two folders which contain images of each class respectively i.e., a folder with all real images and a folder with all fake images. The images loaded were reshaped to 64x64 pixels and the number of images loaded once per call, i.e., batch size was 16.

## B. Building Fake Feature Network

The architecture of the sister network is shown in Fig 2. The sister network initially consists of a convolution layer with  $7 \times 7$  kernel size and 48 channels with stride value being 4. The output of this layer is sent to dense blocks. DenseNets with different number of blocks in each unit are used to construct the Fake Feature Network Model (FFN). A total of two dense units containing 2 and 3 blocks, respectively are created. Each unit has 24 and 30 channels, respectively.

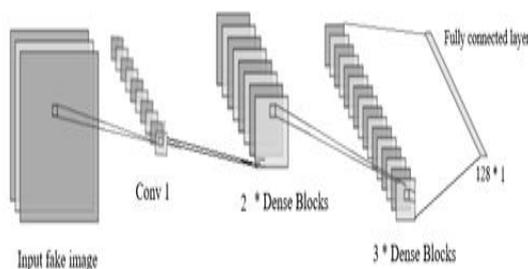


Figure 2 Design of Fake Feature Network

A translation layer is added in between every dense block, where the input tensor is halved. The dense block consists of batch normalization, ReLu activation and convolutions. Convolution with  $k * 4$  filters and  $1 \times 1$  kernel size are used, where  $k$  is the growth rate. Consecutively another batch normalization, ReLu and convolution is done. This time convolutions are done with  $k$  number of filters and  $3 \times 3$  kernel size. Two sister networks are trained using discriminative feature learning. Following these is a fully connected layer of 128 neurons i.e, the output of the first dense unit is concatenated with the final output of the last dense unit and is sent to the fully connected layer.

## C. Building Siamese neural network using FFN

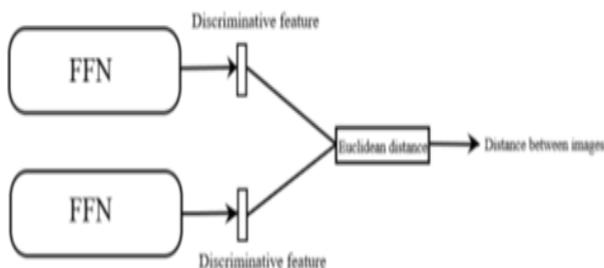


Figure 3 Siamese network architecture

Fig. 3 depicts the architecture of Siamese network. Sister networks have same architecture and parameters. Both networks work and train simultaneously to enable pairwise learning. A pair of images is given to the model where each sister network takes one image from the pair as input. The sister networks then give

extract discriminative features as output, which is passed to a function that calculates Euclidean distance between the features. If the distance between two images is less than 0.5, they belong to the same class else they belong to different classes.

#### D. Discriminative Feature Learning

To enhance the performance of the system, contrastive loss was introduced to learn the common fake features. Contrastive loss is extensively used in Siamese neural networks to calculate the distance between two images in vector space. The loss is low if the images are closer and high when images are farther apart. To achieve this, we incorporate contrastive loss in energy function of traditional loss function. The energy function defined as follows:

$$E_w(x_1, x_2) = ||f_{FFN}(x_1) - f_{FFN}(x_2)||^2$$

Where  $(x_1, x_2)$  is the face image pair.

After incorporating contrastive loss into the function, it will be defined as:

$$L(W, (P, x_1, x_2)) = 0.5 \times (y_{ij} E_w) + (1 - y_{ij}) \times \max(0, (m - E_w))^2$$

Where  $y$  indicated pairwise label, that  $y=0$  indicates an imposter pair and  $y=1$  indicates a genuine pair.  $m$  denoted the predefined threshold. When input is genuine, the cost function minimizes the energy by making  $(1 - y_{ij}) \times \max(0, (m - E_w))^2$  equate to zero. When the input pair is an imposter pair the contrastive loss will maximize energy by minimizing function  $\max(0, (m - E_w))$ . In this way by iteratively training the network using contrastive loss, common fake features of collected GANs can be learned.

#### 4. Results and Validating FFN

N-way one shot learning is used to validate the Siamese neural network. The algorithm works in such a way that same character is compared to  $n$  different characters out of which only one of them matches the original character. The model must give minimum distance to the one that matches the parameter when compared to the rest. If so, it is treated as a correct prediction. This procedure is repeated  $k$  times, and the percentage of correct predictions is calculated as follows.

$$\text{percent\_correct} = (100 \times n\_correct) / k$$

where  $k$  represents the total number of trials and  $n\_correct$  represents the number of correct predictions out of  $k$  trials.

The proposed model has been tested using N-way one shot learning technique. Where  $n$  number of pairs will be sent to the model out of which only one pair of images belong to the same class. The model is tested  $k$  number of times to check how accurately it predicts that one pair. To test this model 2,4,16 and 32 pairs of images have been tested 100 times for both real and fake images. The results are shown in Table 1.

Table 1 N-way one shot results

Class	Number of pairs	Repetitions	Accuracy
Real	2	100	100
Real	8	100	100
Real	4	100	100
Real	16	100	100
Real	32	100	99
Fake	2	100	100
Fake	4	100	100
Fake	8	100	100
Fake	16	100	98
Fake	32	100	98

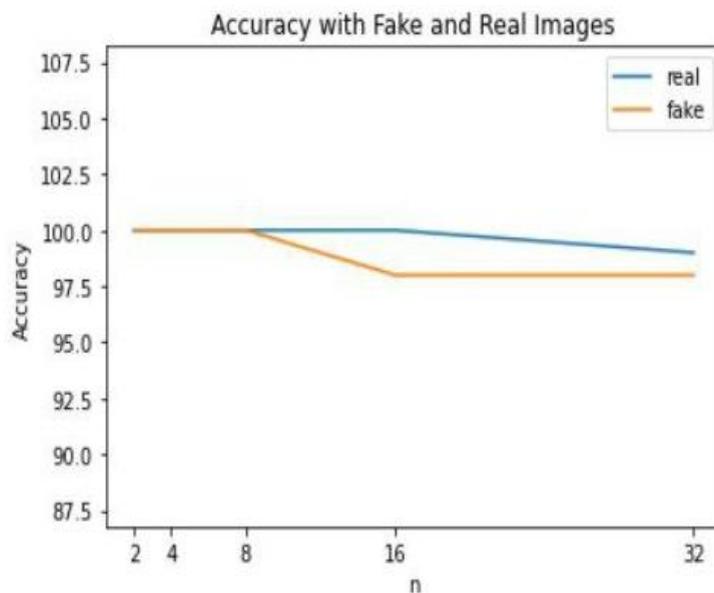


Figure 4 Accuracy curve

The model has also been tested by changing the batch size of input given to the model. Time taken in training has also been added as a constraint to test the performance of the model. The results are shown in Table 2.

Table 2 Test results of proposed model

Batch size	Time taken per epoch	Loss
8	6 hrs	0.0017
16	3.5 hrs	0.0017
32	2 hrs	0.0018
64	1.5 hrs	0.0020

To get the model with low training time per epoch and loss, models with 8,16,32 and 64 batch sizes were test and are show in Table 2. Based on our testing

process, it was found that modified neural model with 8 batch sizes performed better when compared to the rest.

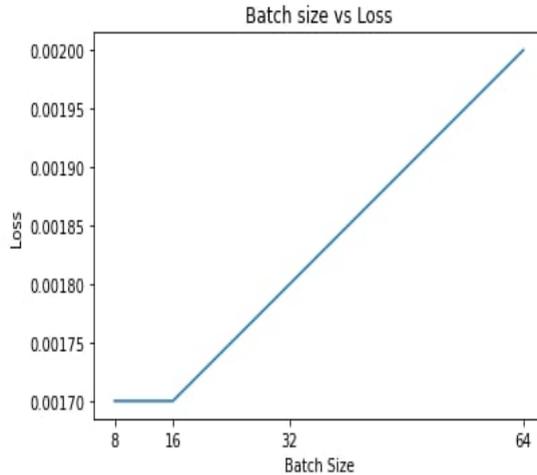


Figure 5 Batch size vs loss curve

On evaluating the model built over the validation data, the accuracy and loss obtained were 99.5% and 0.0017, respectively.

```
Epoch 1/2
103/2812 [>.....] - ETA: 3:22:45 - loss: 3.2688
```

Figure 6 Output when training

### Random Model

Random model is a common technique which is used while testing neural network incorporated with Siamese neural networks. Random model working is similar to that of N-way one shot learning. It classifies the image given as input by comparing it to 16 random images. The set of images it uses to compare the test image is called Support Set. The model then generates n random numbers between 0 and 1. n random numbers denote the similarity score between the test image and one of the images in the support set. While testing, if the model gives a maximum similarity score to the one it is like, then it is considered as correct prediction.

The model created was also tested with a Random model. A random model was created and tested using 2,00,000 images using different n values. 2, 4 8,16 and 32 n values were used to test the model, where each n value prediction was repeated 100 times.

Table 3 Test results for random model

N Value	Number of trials	Accuracy
2	100	44
4	100	25
8	100	16
16	100	7
32	100	5

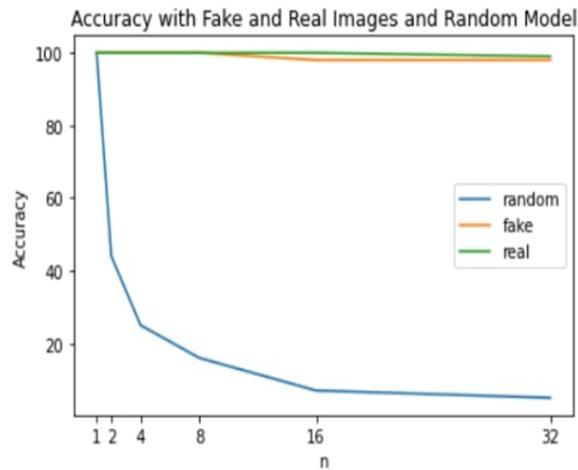


Figure 7 Random model vs FFN

The model was also tested with a random model. Table 3 represents the results of the random model. The results were compared to the model proposed and the results are shown by Fig 7. Random model's performance deteriorated as the number of numbers it generated increased. Whereas the proposed model's performance was almost linear when batch size increased. After comparing the results of both models, the FFN model proved to be more efficient.

### Executable model

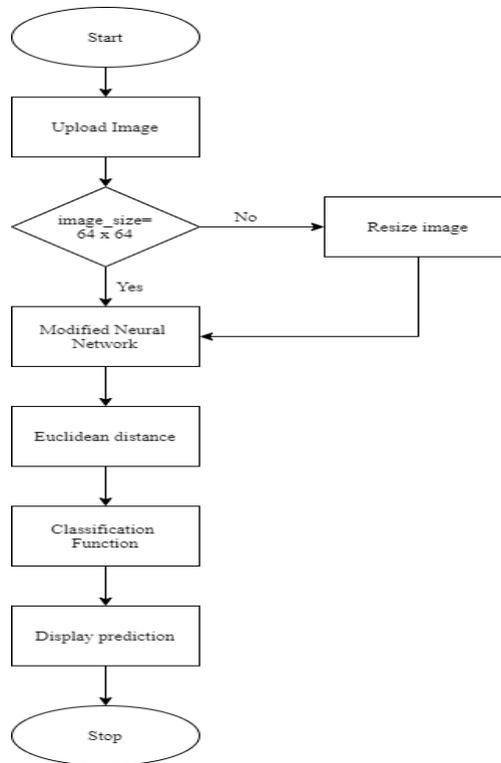


Figure 8 Flow chart of execution

### Upload Image

Once the user executes the application, a Graphical User Interface pops up on the screen. The GUI prompts the user to upload an image which he wants to classify. The dialogs and buttons on the GUI were constructed using Tkinter library in python.



Figure 9 Screenshot of user Interface

Upon clicking the upload image button, the user is directed to another window which allows him to choose his picture. The selected image's path is returned and passed to the modified neural network.



Figure 10 Screenshot of uploading image

### **Resize Image**

Dimension of the image uploaded by the user are first checked before sending it to the modified neural network. The input image is passed as it is if its is 64 x 64, else the image is resized using cv2.resize function which is a part of python OpenCV library. The resized image is then passed to the neural network.

### **Modified Neural Network**

The images given by the user are sent to the FFN. Along with the input image a set of n fake and real images are randomly picked and sent to the FFN. FFN extracts discriminative features from the input image and the set of n images given to it. The extracted features are stored in a feature vector. Feature vector is then sent to the lambda layer.

### **Euclidean distance**

The lambda layer calculates Euclidean distance between the input image and the set of n images it randomly chose, using the feature vectors given by the user. The Euclidean distance represents the distance between two images in a vector space. If the two images belong to the same class their Euclidean distance is less than 0.5 and more than 0.5 when they belong to different classes.

### **Classification Function**

The classification used works with a simple logic and uses simple python language. When the Euclidean distance is passed to the function it tries to check how close it is to the random picture it is being compared to. If the distance between them is less than 0.5 the random image label is checked, and the input image is classified in the same class as the random image. If the distance between

them is higher than 0.5 the input image is classified in a different class than that of the random image's.

### Display Prediction

Once the image is classified the class name is shown on the GUI through create\_text function present in Tkinter library of python.



Figure 11 Screenshot of prediction

## 5. Conclusion

A novel modified neural network for fake face detection is proposed in this paper. Since the model has only been trained using single face images it gave satisfactory results with images with multiple imaged were given as input. The ability of this system can be extended to detect fake videos and images with more than one face. One can detect fake news or videos and assures the credibility of the data. Incorporating face detection will help in case of images with multiple faces. The model can used face detection to identify the faces in the image and classify them individually.

## References

- [1] Chih-Chung Hsu, Yi-Xiu Zuhang and Chia-Yen Lee. "Deep Fake Image Detection Based on Pairwise Learning", Published in MDPI, Jan 3 2020
- [2] Muhammed Asfal Villain, Johns Paul, Kuncheria Kuruvilla and Eldo P Elias. "Fake Image Detection using Machine Learning" In proceedings of IEE, Mar-April 2017.
- [3] Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J." Is object localization for free?-weakly-supervised learning with convolutional neural networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
- [4] Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. "Self-Attention generative adversarial networks". In Proceedings of the 36th International Conference on Machine Learning; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR: Long Beach, CA, USA, 2019

- [5] Marra, F.; Gragnaniello, D.; Cozzolino, D.; Verdoliva, L. "Detection of GAN-Generated Fake Images over Social Networks." In Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval, Miami, FL, USA, 10–12 April 2018.
- [6] Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. "Improved training of wasserstein gans." In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
- [7] Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Smolley, S.P. "Least squares generative adversarial networks". In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
- [8] Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y." Spectral normalization for generative adversarial networks." In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
- [9] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. "ImageNet large scale visual recognition challenge." *Int. J. Comput. Vis. (IJCV)* 2015, 115, 211–252.
- [10] Zhengzhe Liu, Xiaojuan Qi, Jiaya Jia and P. Torr, "Global texture enhancement for fake face detection in the wild", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8057-8066, 2020.
- [11] Xiaodan Li, Yining Lang, Yuefeng Chen, Xiaofeng Mao, Yuan He, Shuhui Wang, et al., "Sharp multiple instance learning for deepfake video detection", Proceedings of the 28th ACM International Conference on Multimedia, Oct 2020.
- [12] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram and Cristian Canton Ferrer, "The deepfake detection challenge (DFDC) preview dataset", arXiv preprint arXiv:1910.08854 [cs.CV], 2019.
- [13] [Luca Guarnera, Oliver Giudice, Cristina Nastasi and Sebastiano Battiato, "Preliminary forensics analysis of deepfake images", arXiv preprint arXiv:2004.12626 [cs.CV], 2020.
- [14] Rinaritha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
- [15] S. Azarian-Pour, M. Babaie-Zadeh and A. R. Sadri, "An automatic JPEG ghost detection approach for digital image forensics," 2016 24th Iranian Conference on Electrical Engineering (ICEE), 2016, pp. 1645-1649.
- [16] Humeau-Heurtier, "Texture feature extraction methods: A survey", *IEEE Access*, vol. 7, pp. 8975-9000, 2019.

## DESIGN OF A NOVEL ENSEMBLE INTRUSION DETECTION FRAMEWORK USING THE CICIDS 2017 DATASET

**A.Prashanthi**

Research Scholar, Department of CSE, Osmania University,Hyd.  
aindala.prashanthi@gmail.com

**Dr. R. Ravinder Reddy**

Associate Professor, Dept. of CSE, Chaitanya Bharathi Institute of Technology, TS, India  
rravinderreddy\_cse@cbit.ac.in

**Abstract**— Complexity and diversity of today's cyber assaults make it challenging to create a multi-attack categorization intrusion detection system. Intrusion Detection Systems need efficient classification to counteract hackers' advanced strategies. A single classifier can't properly detect several sorts of attacks, which is another problem. We suggested an unique ensemble architecture named Leader Class and Confidence Decision Ensemble Technique (LC&CDET) that accurately detects threats. The recommended strategy for spotting attacks involves ranking the detection capabilities of different base classifiers. The voting technique utilises a majority of the classifiers, regardless of whether the algorithm can detect the assault. It chooses the best ML model from three advanced methods (XGBoost, LightGBM, and CatBoost) for each attack category. Class leader models with confidence values are used to assess cyberattack detection. The proposed LC&CDE successfully detects intrusions using publicly accessible CICIDS2017 dataset with an accuracy and F1-score of 99.813 and 99.811%.

**Keywords**— Intrusion Detection System, CAN Bus, LightGBM, XGBoost, Ensemble Learning CICIDS2017;

### I. INTRODUCTION

Cybercrime and other risks to computer and internet security are on the rise as the digital world develops. As a result, classic Intrusion Detection Systems (IDS) are rapidly becoming defunct. Previous IDS-based security solutions depended on signatures concepts that had already been specified [1], making them unable to spot newly generated abnormalities and attack variations [2]. The main issue was that the signature database wasn't being updated and expanded quickly enough to keep up with the rapidly changing nature of threats [2]. With the increasing complexity of modern assaults, researchers are constantly developing new methods for anomaly-based threat detection and protection [3]. In order to equip systems with robust intrusion detection tactics for the future, researchers are utilizing cutting-edge methods based on machine learning (ML) and deep learning (DL). However, not all modern-day threats can be detected using a single machine learning approach.

Numerous studies have been suggested in this area, with authors claiming that their IDS systems can accurately classify attacks even when presented with just historical data. The success of an Intrusion Detection model in the real world, however, hinges on how well it can

spot sophisticated attacks as they happen in real time. Most of the research in this field is conducted using outdated data sets [4]. In Table 1 are listed the most frequently employed datasets [5] in IDS studies. Attack detection using models built on historical data is ineffective with today's traffic.

Machine learning is used to spot invaders. Random Forest, Support Vector Machine, Decision Trees, and k-Nearest Neighbor (kNN). Complex IDS demands can't be covered by a single machine learning algorithm [7], [4], [8], [9]. IDS attacks have gotten more diversified and sophisticated as traffic has increased [6]. Every ML algorithm has pros and cons. Some algorithms may be effective against some attacks, but not others. When ML models are utilised in IDS systems, cyber-attack prediction performance might vary widely. In this research, we present Leader-Class and Confidence-Decision Ensemble Technique (LC&CDET), which incorporates three gradient-boosting ML algorithms: XGBoost [10], LightGBM [11], and CatBoost [12]. LC&CDET maximizes model performance by identifying which base ML model has the most confidence for each class.

The primary contributions of this study are as follows:

- 1) It recommends a new ensemble framework called LC&CDET for efficient intrusion detection by combining class leader and confidence decision procedures with gradient-boosting ML algorithms.
- 2) The framework is evaluated with CICIDS2017 datasets, which represent IVN and external network data, and public security datasets.
- 3) It evaluates the given model in relation to other, more advanced techniques.

The following is how the paper is organized.

- a. Section II presents an overview of recent research on intrusion detection using machine learning and ensemble models.
- b. Part III describes the proposed LC&CDET framework.
- c. Part IV presents and discusses the findings of the experiments.
- d. The conclusion is the last portion of the paper.

## II. RELATED WORK

Deep learning was used by W. Wang [1] to efficiently and automatically extract essential feature representations for use in detection. They provide a strong stacked contractive auto encoder (SCAE) approach for unsupervised feature extraction. With a high degree of resilience, the SCAE approach enables the automatic learning of superior low-dimensional properties from unprocessed network information. The studies show that the proposed SCAE+SVM methodology outperforms three current state-of-the-art approaches on the KDD 99 and NSL-KDD intrusion detection assessment datasets.

L. Qi [2] recommends utilising MDS AD. It's a combination of PCA, isolation forest, and locality-sensitive hashing (LSH). MDS AD has the following characteristics. 1) The new LSH can handle data from several angles. Second, MDS AD has a great degree of accuracy in detecting anomalies in experimental groups. Finally, the principal component analysis may be used to minimise the dimension of correlations between numerous attributes. Finally, MDS

AD is a streaming approach that can continually analyse data and update models while consuming little memory and processing power. Several studies are planned and carried out utilising the UNSW-NB15 dataset to validate the effectiveness of MDS AD. MDS AD outperforms the best current alternatives in tests, as proved.

R. Conde Camillo da Silva [3] used the CSIC 2010 dataset for both training and evaluation in his work. J48, Naive Bayes, OneR, Random Forest, and IBM Watson LGBM algorithms were compared. The measurements used were T-rate, accuracy, recall, and f-measure. The Watson tool's algorithm (LGBM) fared the best across the board when compared against other algorithms in the literature.

A two-stage pipeline was proposed by P. Barnard [4] for effective network intrusion detection. As a starting point, we employ an XGBoost model for supervised intrusion detection while developing model justifications with the SHapley Additive exPlanation (SHAP) framework. Using these explanations in the second stage teaches an auto-encoder to distinguish between known and unknown attacks. While the overall performance of our approach is comparable to many cutting-edge attempts in the cybersecurity literature, trials on the NSL-KDD dataset show that it can consistently identify new threats discovered during testing.

R. Zhao [5] used the CFS-DE technique to choose the best feature subset to minimize the size of the features. After that, a weighted Stacking method is presented, which raises the levels of base classifiers with excellent training results and lowers those with poor training results to enhance classification performance. As a result, the model improves classification efficiency while also generating higher accuracy. The NSL-KDD and CSE-CIC-IDS2018 data sets were used for all experiments throughout this study.

L. Wang [6] developed an efficient method for filtering only a few examples of exceptionally long round-trip durations from Internet packets. Experiments corroborate our findings that our proposed SSI detection method has a detection efficiency of more than 85.7% when used over the internet.

Z. Wu [7] proposes a dynamic ensemble evolutionary algorithms for intrusion detection approach (DEIL-RVM) and implements a dynamically adaptable ensemble intrusion detection model. The faulty ensemble model base component is eliminated and substituted with an alternate one that utilizes the fresh overall misclassification probability weighting factor (OMPW) produced from an incremental set or data chunk in this procedure. They are able to create a solid balance between accuracy and robustness by using the RVM's high feature space as the foundation.

S. Subbiah [8] proposes the Boruta feature extraction with grid search random forest (BFS-GSRF) approach in this study as a unique framework for IDS that can assist address these issues. BFS-efficiency GSRFs are compared to other ML approaches such as linear discrimination analysis (LDA) and regression tree and classification methods (CART), among

others. The recommended work was carried out and assessed using data from of the Knowledge on Discoveries dataset there at Network Security Laboratory (NSL-KDD). The experimental results show that the suggested model BFS-GSRF detects attacks more accurately than LDA, CART, as well as other existing algorithms, with such an accuracy of 99%.

M. A. Siddiqi [9] proposed an image-processing-based NIDS architecture based on best practises. The framework includes an image improvement and modification mechanism with an improved feature selection flow. In order to maximise efficiency, the proposed framework first reduces the number of features. Data that did not previously have any visual display is then transformed into visuals. To show the effectiveness of the suggested architecture, a comparison with numerous recent researches on vision processing systems for network intrusion detection is done.

The primary focus of this study is X. Zhou's [10], who created a Graph neural net (GNN)-based intrusion detection in IoT systems, as well as an unique hierarchical adversarial attack (HAA) generation approach is suggested for implementing the level-aware black-box adversarial assault approach. Using the publicly accessible data set UNSW-SOSR2019, the proposed HAA generating approach is compared against three reference techniques. When compared to new state GNN algorithms for NIDS in IoT environments, the results demonstrate that it can lower their classification accuracy by more than 30%.

By comparing features from of the UNSWNB15 and Bot-IoT data-sets based upon flow and Transmission Control Protocol, M. Zeeshan et al proposed 's of a Protocol Driven Deep Intrusion Detection (PB-DID) framework was made in this work (TCP). To appropriately classify regular, DoS, and DDoS traffic, we address problems such imbalanced and over-fitting. The deep learning (DL) technique allowed us to increase the classification accuracy to 96.3%.

In order to identify assaults without any prior information, G. Pu [12] introduced an unsupervised anomaly detection method that utilizes Sub-Space Clustering (SSC) and A Class Support Vector Machine (OCSVM). Another well NSL-KDD dataset is used to assess the suggested method. The experimental findings show that our strategy outperforms several of the currently used methods.

J. Yang [13] proposed an image-processing-based NIDS architecture based on best practises. The model integrates an image improvement and modification mechanism with an improved feature selection flow. A deep-learning classifier is used to help in the detection of anomalies in the pictures. The suggested technique is tested using three distinct intrusion detection benchmark datasets. To show the effectiveness of the suggested architecture, a comparison with numerous recent research on vision processing systems for network intrusion detection is done.

Y. Xie [14], proposed hybrid method that considers the anomaly degree of the entire provenance graph as well as the anomaly degree of a single provenance path. Additionally, it

encodes objects in the rule database that are duplicates, and it filters out noise that does not carry any incursion information. Effectiveness and efficiency have been demonstrated experimentally across a wide range of practical applications.

### III. PROPOSED FRAMEWORK

#### A. An Overview of the System

This study seeks to create an ensemble IDS architecture capable of identifying network threats on a consistent basis. Figure 1 depicts the two main components of the proposed system: model training and model prediction.

The CICIDS-2017 dataset is used to train three cutting-edge ML algorithms, XGBoost, LightGBM, and CatBoost, during model training. To identify assaults, model prediction employs class leader models and prediction confidences. This section outlines the algorithm.

#### B. Base Machine Learning Models

XGBoost, LightGBM, and CatBoost are all popular gradient boosting frameworks that can be used for various machine learning tasks, including anomaly detection. These frameworks are known for their speed, scalability, and accuracy, making them suitable for handling large datasets and complex feature sets.

In addition to these frameworks, there is a technique called Leader Class and Confidence Decision Ensemble (LC-CDE) that can be used for anomaly detection. This technique involves building multiple models, where each model is trained on a subset of the data, and then combining the results to make a final decision.

To implement this technique for anomaly detection, we first split our data into training and validation sets. Then, we train multiple models using XGBoost, LightGBM, and CatBoost, where each model is trained on a different subset of the training data. Once the models are trained, we use them to make predictions on the validation set and compute the anomaly score for each data point.

The anomaly score can be computed using a variety of methods, such as the Mahalanobis distance or the isolation forest algorithm. Once we have the anomaly scores for each data point, you can combine them using the LC-CDET technique to make a final decision on whether a data point is anomalous or not.

The LC-CDET technique involves assigning a leader class to each model based on its performance on the validation set. The leader class is the model that performs the best on the validation set. Then, for each data point, we can compute the confidence of each model's prediction and assign a confidence score based on its distance from the leader class prediction. The confidence scores can be combined using a weighted average to make the final decision. Overall, using XGBoost, LightGBM, and CatBoost with the LC-CDE technique can be an effective approach for anomaly detection, as it combines the strengths of multiple models and can handle complex datasets and feature sets. However, it's important to carefully tune the hyperparameters of the models and the LC-CDE technique to achieve the best performance.

#### C. LC&CDE: Proposed Ensemble Algorithm

Anomaly detection using ensemble techniques can be done using XGBoost, LightGBM, and CatBoost. The Leader Class and Confidence Decision Ensemble technique can be used to improve the performance of the algorithm. Here's an outline of the algorithm:

1. Data preprocessing: The first step is to preprocess the data. This includes data cleaning, data normalization, and feature selection.
2. Ensemble training: Ensemble training involves training multiple models using different algorithms such as XGBoost, LightGBM, and CatBoost. Each model is trained on a different subset of the data using different hyperparameters.
3. Model selection: Once the models are trained, they are evaluated on a validation set. The best-performing models are selected for the final ensemble.
4. Leader class selection: In the leader class selection step, the best-performing model is selected as the leader. This model is used as the primary classifier in the final ensemble.
5. Confidence threshold selection: The confidence threshold is the threshold at which the model classifies a sample as an anomaly. The confidence threshold is selected based on the validation set's performance.
6. Confidence decision ensemble: In the confidence decision ensemble step, the other models in the ensemble are used to provide a confidence score for each sample. If a sample is classified as an anomaly by the leader model and has a low confidence score, it is rejected.
7. Anomaly detection: Finally, the ensemble of models is used to classify samples as anomalous or non-anomalous. Samples that are classified as anomalous by the leader model and have a high confidence score are considered to be true anomalies.

This algorithm can be further optimized by tuning hyperparameters using cross-validation and selecting the optimal threshold using the ROC curve.

---

**Input:**  
 $D_{test}$ : the test set,  
 $M = \{M_1, M_2, M_3\}$ : the trained base ML model list, including  $M_1 =$  LightGBM,  $M_2 =$  XGBoost,  $M_3 =$  CatBoost,  
 $c = 1, 2, \dots, n$ : the class list for  $n$  different classes.

**Output:**  
 $L_{test}$ : the prediction classes for all test samples in  $D_{test}$ .

```

1 for each data sample  $x_i \in D_{test}$  do // For each test sample
2    $L_{i1}, p_{i1} \leftarrow Prediction(M_1, x_i)$ ; // Use the trained LightGBM
   model to predict the sample, and save the predicted class & confidence
3    $L_{i2}, p_{i2} \leftarrow Prediction(M_2, x_i)$ ; // Use XGBoost to predict
4    $L_{i3}, p_{i3} \leftarrow Prediction(M_3, x_i)$ ; // Use CatBoost to predict
5   if  $L_{i1} == L_{i2} == L_{i3}$  then // If the predicted classes of all the
   three models are the same
6      $L_i \leftarrow L_{i1}$ ; // Use this predicted class as the final predicted class
7   else if  $L_{i1} \neq L_{i2} \neq L_{i3}$  then // If the predicted classes of all the
   three models are different
8     for  $j = 1, 2, 3$  do // For each prediction model
9       if  $M_j == LM_{r_{i,j}}$  then // Check if the predicted class's
   original ML model is the same as its leader model
10       $L\_list_i \leftarrow L\_list_i \cup \{L_{i,j}\}$ ; // Save the predicted
   class
11       $p\_list_i \leftarrow p\_list_i \cup \{p_{i,j}\}$ ; // Save the confidence
12    end
13  end
14  if  $Len(L\_list_i) == 1$  then // If only one pair of the original
   model and the leader model for each predicted class is the same
15     $L_j \leftarrow L\_list_i[0]$ ; // Use the predicted class of the leader
   model as the final prediction class
16  else // If no pair or multiple pairs of the original prediction model
   and the leader model for each predicted class are the same
17    if  $Len(L\_list_i) == 0$  then
18       $p\_list_i \leftarrow \{p_{i1}, p_{i2}, p_{i3}\}$ ; // Avoid empty probability
   list
19    end
20     $p\_max_i \leftarrow \max(p\_list_i)$ ; // Find the highest confidence
21    if  $p\_max_i == p_{i1}$  then // Use the predicted class with the
   highest confidence as the final prediction class
22       $L_i \leftarrow L_{i1}$ ;
23    else if  $p\_max_i == p_{i2}$  then
24       $L_i \leftarrow L_{i2}$ ;
25    else
26       $L_i \leftarrow L_{i3}$ ;
27    end
28  end
29  else // If two predicted classes are the same and the other one is different
30     $n \leftarrow mode(L_{i1}, L_{i2}, L_{i3})$ ; // Find the predicted class with the
   majority vote
31     $L_i \leftarrow Prediction(M_n, x_i)$ ; // Use the predicted class of the
   leader model as the final prediction class
32  end
33   $L_{test} \leftarrow L_{test} \cup \{L_i\}$ ; // Save the predicted classes for all tested
   samples;
34 end

```

---

**Algorithm 1:** Leader Class and Confidence Decision Ensemble (LCCDE) - Model Training

---

**Input:**  
 $D_{train}$ : the training set,  
 $M = \{M_1, M_2, M_3\}$ : the base ML model list, including  $M_1 =$  LightGBM,  $M_2 =$  XGBoost,  $M_3 =$  CatBoost,  
 $c = 1, 2, \dots, n$ : the class list for  $n$  different classes.

**Output:**  
 $M = \{M_1, M_2, M_3\}$ : the trained base model list,  
 $LM = \{LM_1, LM_2, \dots, LM_n\}$ : the leader model list for all classes.

- 1  $M_1 \leftarrow Training(M_1, D_{train});$  // Train the LightGBM model
- 2  $M_2 \leftarrow Training(M_2, D_{train});$  // Train the XGBoost model
- 3  $M_3 \leftarrow Training(M_3, D_{train});$  // Train the CatBoost model
- 4 **for**  $c = 1, 2, \dots, n$  **do** // For each class (normal or a type of attack), find the leader model
  - 5  $Mlist_c \leftarrow BestPerforming(M_1, M_2, M_3, c);$  // Find the best-performing model for each class (e.g., has the highest F1-score)
  - 6 **if**  $Len(Mlist_c) == 1$  **then** // If only one model has the highest F1
    - 7  $LM_c \leftarrow Mlist_c[0];$  // Save this model as the leader model for the class  $c$
  - 8 **else** // If multiple ML models have the same highest F1-score
    - 9  $LM_c \leftarrow MostEfficient(Mlist_c);$  // Save the fastest or most efficient model as the leader model for the class  $c$
  - 10 **end**
  - 11  $LM \leftarrow LM \cup \{LM_c\};$  // Collect the leader model for each class
  - 12 **end**

---

## IV. PERFORMANCE EVALUATION

### A. Experimental Setup

Python Scikit-learn, Xgboost, Lightgbm, and Catboost packages were used to develop the recommended IDS. The experiments carried out on Google colab. Anomaly detection using ensemble techniques can be done using XGBoost, LightGBM, and CatBoost. The Leader Class and Confidence Decision Ensemble technique can be used to improve the performance of the algorithm. Here's an outline of the algorithm:

#### . # Load the data

```
df = pd.read_csv('data.csv')
```

#### # Split the data into training and testing sets

```
X_train, X_test, y_train, y_test = train_test_split(df.drop('target', axis=1), df['target'],
test_size=0.2, random_state=42)
```

#### # Scale the features

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

#### # Train the XGBoost model

```
xgb_model = xgb.XGBClassifier()
xgb_model.fit(X_train, y_train)
```

#### # Train the LightGBM model

```
lgb_model = lgb.LGBMClassifier()
```

```
lgb_model.fit(X_train, y_train)
```

#### # Train the CatBoost model

```
cb_model = cb.CatBoostClassifier()
cb_model.fit(X_train, y_train)
```

#### # Use the ensemble technique to combine the results of the three models

```
def ensemble_predict(X):
    xgb_pred = xgb_model.predict_proba(X)
    lgb_pred = lgb_model.predict_proba(X)
    cb_pred = cb_model.predict_proba(X)
    preds = np.vstack([xgb_pred[:,1], lgb_pred[:,1], cb_pred[:,1]])
    confs = np.vstack([xgb_pred[:,1] > 0.5, lgb_pred[:,1] > 0.5, cb_pred[:,1] > 0.5])
    leader = np.argmax(np.sum(confs, axis=0))
    return preds[leader,:]
```

#### # Use the ensemble model to predict anomalies in the testing dataset

```
y_pred = ensemble_predict(X_test)
y_pred_binary = y_pred > 0.5
tn, fp, fn, tp = confusion_matrix(y_test, y_pred_binary).ravel()
precision = tp / (tp + fp)
recall = tp / (tp + fn)
f1_score = 2 * (precision * recall) / (precision + recall)
print('Precision: {:.4f}'.format(precision))
print('Recall: {:.4f}'.format(recall))
print('F1 Score: {:.4f}'.format(f1_score))
```

## B. Results and Discussion from the Experiments

Tables I and II compare the LC&CDE model against LightGBM, XGBoost, and CatBoost on the CICIDS2017 datasets. Table I shows the F1-scores for detecting each type of attack in the two datasets. Different basic ML models' F1-scores for assault detection vary. Using the CICIDS2017 dataset, LightGBM achieves the highest F1-score among the three base learners for normal samples, DoS, sniffing, webattacks, botnets, and infiltration attacks, while XGBoost beats LightGBM for brute-force attack detection.

Table I shows that the proposed LC&CDE ensemble model outperforms state-of-the-art approaches in all classes. The recommended model has the highest F1-scores among the four ML models employed on the two datasets (Tables II and III). The LC&CDE model boosted the CICIDS2017 F1-score from 99.792% to 99.811%. Using the top-performing base models for each class as the LC&CDE ensemble model's starting point provides apparent advantages.

Tables II and III compare suggested method's performance on two datasets to state-of-the-art techniques [14]. The recommended LC&CDE model enhances state-of-the-art techniques by 0.09% and 0.11% on CICIDS2017 datasets. The LC&CDE model takes a bit longer to run than the three gradient-boosting models, but it's significantly faster than other base learners line

KNN and SVM provided in the literature. The recommended ensemble model is built with low-complexity ML models that can run in parallel and benefit from GPU support. The recommended model has the lowest execution time and highest F1-scores across both benchmark datasets.

TABLE I

MODEL PERFORMANCE COMPARISON FOR EACH CLASS IN DATASET

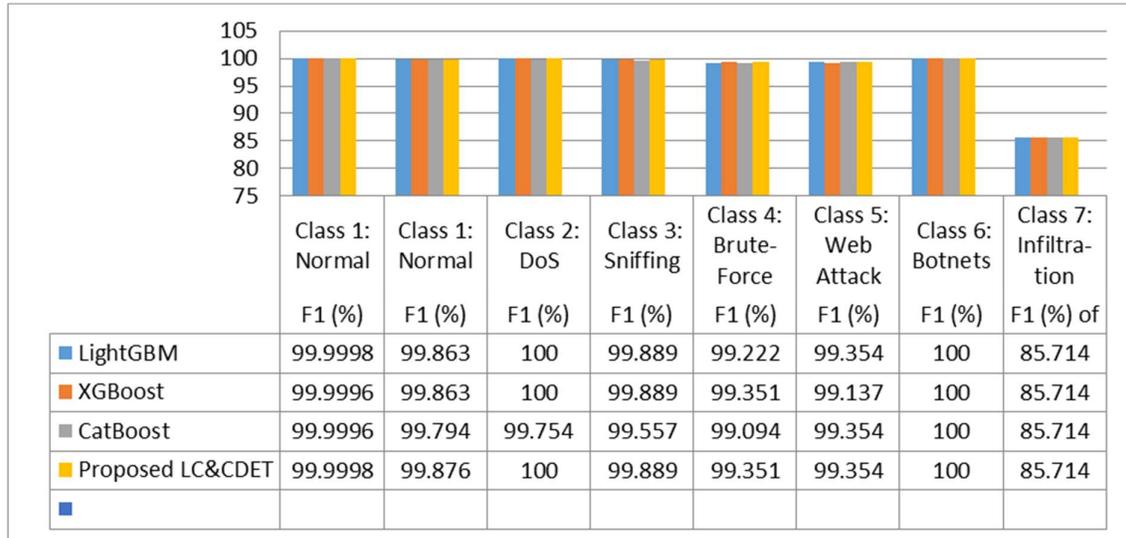
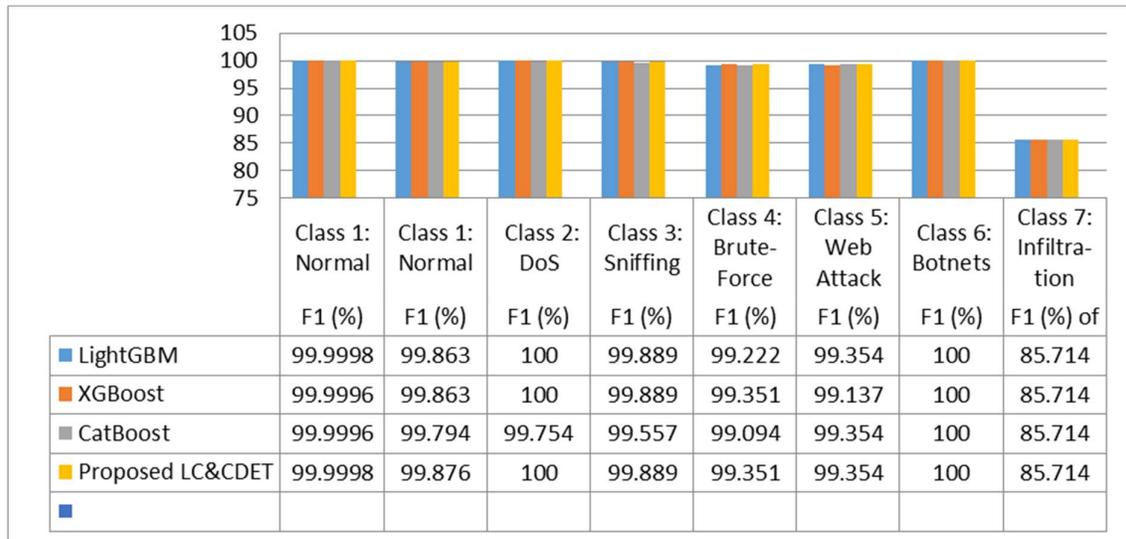


TABLE II

PERFORMANCE EVALUATION OF MODELS ON CICIDS2017



V. CONCLUSION

Methods of machine learning (ML) are used to detect attacks and improve security. The effectiveness of ML models relies on the type of assault analysed. In this study, we offer LC&CDET for top network attack performance on all attack types. It assesses which ML models better detect certain risks and employs them as "leading class models" in order to create

a credible ensemble model. To decrease the likelihood of outcomes and develop prediction classes, data on the predictor's degree of confidence is employed. Three gradient boosting machine learning algorithms—XGBoost, LightGBM, and CatBoost—are used in the proposed LC&CDET ensemble model. The suggested IDS framework can get 99.811% F1- scores on the CICIDS2017 datasets, according to experiments. The suggested model outperforms previous ML methods in terms of F1-scores for recognising all attacks. This demonstrates the viability of the leader-class-based Technique.

## REFERENCES

- [1] W. Wang, X. Du, D. Shan, R. Qin and N. Wang, "Cloud Intrusion Detection Method Based on Stacked Contractive Auto-Encoder and Support Vector Machine," in *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 1634-1646, 1 July-Sept. 2022, doi: 10.1109/TCC.2020.3001017.
- [2] L. Qi, Y. Yang, X. Zhou, W. Rafique and J. Ma, "Fast Anomaly Identification Based on Multiaspect Data Streams for Intelligent Intrusion Detection Toward Secure Industry 4.0," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 6503-6511, Sept. 2022, doi: 10.1109/TII.2021.3139363.
- [3] R. Conde Camillo da Silva, M. P. Oliveira Camargo, M. Sanches Quessada, A. Claiton Lopes, J. Diassala Monteiro Ernesto and K. A. Pontara da Costa, "An Intrusion Detection System for Web-Based Attacks Using IBM Watson," in *IEEE Latin America Transactions*, vol. 20, no. 2, pp. 191- 197, Feb. 2022, doi: 10.1109/TLA.2022.9661457.
- [4] P. Barnard, N. Marchetti and L. A. DaSilva, "Robust Network Intrusion Detection Through Explainable Artificial Intelligence (XAI)," in *IEEE Networking Letters*, vol. 4, no. 3, pp. 167- 171, Sept. 2022, doi: 10.1109/LNET.2022.3186589.
- [5] R. Zhao, Y. Mu, L. Zou and X. Wen, "A Hybrid Intrusion Detection System Based on Feature Selection and Weighted Stacking Classifier," in *IEEE Access*, vol. 10, pp. 71414-71426, 2022, doi: 10.1109/ACCESS.2022.3186975.
- [6] L. Wang, J. Yang, M. Workman and P. Wan, "Effective algorithms to detect stepping-stone intrusion by removing outliers of packet RTTs," in *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 432-442, April 2022, doi: 10.26599/TST.2021.9010041.
- [7] Z. Wu, P. Gao, L. Cui and J. Chen, "An Incremental Learning Method Based on Dynamic Ensemble RVM for Intrusion Detection," in *IEEE Transactions on Network and Service Management*, vol. 19, no. 1, pp. 671-685, March 2022, doi: 10.1109/TNSM.2021.3102388.
- [8] S. Subbiah, K. S. M. Anbananthen, S. Thangaraj, S. Kannan and D. Chelliah, "Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm," in *Journal of Communications and Networks*, vol. 24, no. 2, pp. 264-273, April 2022, doi: 10.23919/JCN.2022.000002.
- [9] M. A. Siddiqi and W. Pak, "Tier-Based Optimization for Synthesized Network Intrusion Detection System," in *IEEE Access*, vol. 10, pp. 108530-108544, 2022, doi: 10.1109/ACCESS.2022.3213937.

- [10] X. Zhou, W. Liang, W. Li, K. Yan, S. Shimizu and K. I. -K. Wang, "Hierarchical Adversarial Attacks Against Graph-Neural-Network-Based IoT Network Intrusion Detection System," in *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9310-9319, 15 June 2022, doi: 10.1109/JIOT.2021.3130434.
- [11] M. Zeeshan et al., "Protocol-Based Deep Intrusion Detection for DoS and DDoS Attacks Using UNSW-NB15 and Bot-IoT Data-Sets," in *IEEE Access*, vol. 10, pp. 2269-2283, 2022, doi: 10.1109/ACCESS.2021.3137201.
- [12] G. Pu, L. Wang, J. Shen and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," in *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 146-153, April 2021, doi: 10.26599/TST.2019.9010051.
- [13] J. Yang, X. Chen, S. Chen, X. Jiang and X. Tan, "Conditional Variational Auto-Encoder and Extreme Value Theory Aided Two-Stage Learning Approach for Intelligent Fine-Grained Known/Unknown Intrusion Detection," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3538-3553, 2021, doi: 10.1109/TIFS.2021.3083422.
- [14] Y. Xie, D. Feng, Y. Hu, Y. Li, S. Sample and D. Long, "Pagoda: A Hybrid Approach to Enable Efficient Real-Time Provenance Based Intrusion Detection in Big Data Environments," in *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 6, pp. 1283-1296, 1 Nov.-Dec. 2020, doi: 10.1109/TDSC.2018.2867595.



**2nd International Conference  
on  
Cognitive and Intelligent Computing (ICIC)**

27-28 December 2022, Vasavi College of Engineering, Hyderabad, India

**Certificate**

**Congratulations & Gratitude to**

**Kavita Goura**

*from*

**PhD Scholar, PDA college of Engineering**

**For presenting the paper 254 ICIC 2022 on**

**A Systematic Review on Multiple Sclerosis**



**General Co-Chair  
Dr. Amit Kumar**



**General Chair  
Dr. T. Adilakshmi**



**2nd International Conference  
on  
Cognitive and Intelligent Computing (ICIC)**

27-28 December 2022, Vasavi College of Engineering, Hyderabad, India

**Certificate**

**Congratulations & Gratitude to**

**Kavita Goura**

*from*

**Chaitanya Bharathi Institute of Technology**

*For presenting the paper 255 ICIC2B on*

**Smart Health Monitoring System Using Raspberry Pi**



**General Co-Chair  
Dr. Amit Kumar**



**General Chair  
Dr. T. Adilakshmi**

[Home](#) [Multimedia Systems](#) [Article](#)

# A new design of multimedia big data retrieval enabled by deep feature learning and Adaptive Semantic Similarity Function

Regular Paper Published: 05 February 2022

Volume 28, pages 1039–1058, (2022) [Cite this article](#)[Download PDF](#) ↓[Multimedia Systems](#)[Aims and scope](#)[Submit manuscript](#)[D. Sujatha](#) [✉](#), [M. Subramaniam](#) & [Chinnanadar Ramachandran Rene Robin](#)[📄](#) 2369 Accesses [📄](#) 3 Citations [Explore all metrics](#) →

## Abstract

Nowadays, multimedia big data have grown exponentially in diverse applications like social networks, transportation, health, and e-commerce, etc. Accessing preferred data in large-scale datasets needs efficient and sophisticated retrieval approaches. Multimedia big data consists of the most significant features with different types of data. Even though the multimedia supports various data formats with corresponding storage frameworks, similar semantic information is expressed by the multimedia. The overlap of semantic features is most efficient for theory and research related to semantic memory.

Correspondingly, in recent years, deep multimodal hashing gets more attention owing to the efficient performance of huge-scale multimedia retrieval applications. On the other

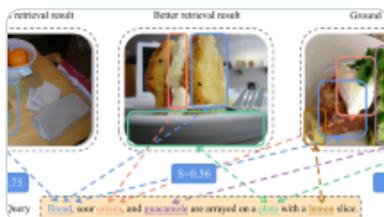
hand, the deep multimodal hashing has limited efforts for exploring the complex multilevel semantic structure. The main intention of this proposal is to develop enhanced deep multimedia big data retrieval with the Adaptive Semantic Similarity Function (A-SSF). The proposed model of this research covers several phases “(a) Data collection, (b) deep feature extraction, (c) semantic feature selection and (d) adaptive similarity function for retrieval. The two main processes of multimedia big data retrieval are training and testing. Once after collecting the dataset involved with video, text, images, and audio, the training phase starts. Here, the deep semantic feature extraction is performed by the Convolutional Neural Network (CNN), which is again subjected to the semantic feature selection process by the new hybrid algorithm termed Spider Monkey-Deer Hunting Optimization Algorithm (SM-DHOA). The final optimal semantic features are stored in the feature library. During testing, selected semantic features are added to the map-reduce framework in the Hadoop environment for handling the big data, thus ensuring the proper big data distribution. Here, the main contribution termed A-SSF is introduced to compute the correlation between the multimedia semantics of the testing data and training data, thus retrieving the data with minimum similarity. Extensive experiments on benchmark multimodal datasets demonstrate that the proposed method can outperform the state-of-the-art performance for all types of data.

## Similar content being viewed by others



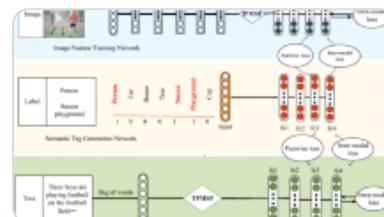
### Label-Based Deep Semantic Hashing for Cross-Modal Retrieval

Chapter | © 2019



### CLIP-based fusion-modal reconstructing hashing for large-scale unsupervised cross-...

Article | 22 February 2023



### Semantics-preserving hashing based on multi-scale fusion for cross-modal retrieval

Article | 02 November 2020

[Use our pre-submission checklist →](#)

Avoid common mistakes on your manuscript.



# 1 Introduction

---

In recent years, the emerging growth in multimedia data has been focused on different applications such as social networks, transportation, health, e-commerce, etc. Data diversity has become one of the major components of multimedia big data [9], where multimedia documents suffer from variations in the storage structures and data formats to show the equivalent semantic information [10]. Thus, retrieving and managing multimedia documents reflect the intent of users in heterogeneous big data platforms, which is considered a major problem. The efficient and sophisticated retrieval approaches need accessing of specific data in huge scale datasets [11]. Moreover, the major component for showing the large-scale data produced in a faster manner is big data owing to the advancements in cloud computing algorithms and technologies, storage, communication, and sensing [12]. However, this gigantic data has more number of problems in industries and businesses. Consequently, it offers more possibilities for eminent future growth by considering the efficient use of data for evaluation [13]. Effective and reliable access is offered for suitable data in huge scale image repositories due to the complexities in their contents that have been analyzed in recent years [18].

Multimedia big data retrieval has more benefits inclusive of switching a distributed pattern for storing and processing the enormous multimedia contents [14]. Even though multimedia computing solves the computational burden and maintenance issues, it suffers from the storage and processing of multimedia big data. In a big data environment, a huge amount of commodity computers requires more storage capacity and enormous computation power while generating multimedia content [2]. As more applications and services are provided, processed, edited, and retrieved some multimedia documents, the significant multimedia contents with their storage structures and data formats demonstrate similar semantic information [15]. Hence, multimedia big data storage and retrieval are affected by the most common limitation called heterogeneity. Another complexity in the multimedia retrieval process is users' intention. In most of the existing methods, the retrieval process is often restricted to a similar category of multimedia content like images [16]. It minimizes the Quality of Service (QoS) of multimedia retrieval due to the failure in identifying the intention of users' search owing to the lack of type diversity [17].

The semantic feature can be overlapped that plays a crucial role in the research field in terms of semantic memory [24]. In recent years, more research works is carried out in the

similarity measurement domain [19]. Different applications employ Euclidean distance [20] as a significant measure. Some of the specific scenarios use Mahalanobis distance [21] as a similarity measurement metric. Though, retrieval systems are impacted by the query conditions through inappropriate expressions. These problems are solved by adopting some deep learning methods, in which fuzzy matching techniques [22] are employed for reflecting the user's intent. On the other hand, these retrieval models do not precisely reflect the users' intentions concerning the incomplete information [23]. Hence, this model proposes a new multimedia big data retrieval model along with the semantic similarity function.

The major contributions of the developed model are given here as follows:

- To present a new multimedia big data retrieval model with deep semantic feature extraction technique, optimal semantic feature selection, and A-SSF-based retrieval under the map-reduce framework with the aid of a new hybrid meta-heuristic-based algorithm.
- To choose the most significant semantic features of multimedia big data using a new hybrid algorithm called SM-DHOA with the objective of correlation maximization among the features. Therefore, this process gets unique and most representative features for further processes.
- To develop a new A-SSF-based retrieval with SM-DHOA under the map-reduce framework for the efficient retrieval process. It aims to maximize the correlation and F1-score among the features for retrieving the most suitable data according to the user query.
- To introduce a new algorithm called SM-DHOA with the integration of Deer Hunting Optimization Algorithm (DHOA) and Spider Monkey Optimization (SMO) algorithms for selecting the optimal semantic features and to optimize the weight function of features in map-reduce framework to maximize the retrieval performance with high precision and F1-score.
- To validate the efficiency of the suggested A-SSF-based retrieval model with different existing approaches in terms of standard performance measures along with statistical analysis.

The remaining section of the developed model is given here. The existing works are reviewed in Sect. 2. The proposed multimedia data retrieval using adaptive semantic similarity and deep learning is depicted in Sect. 3. The optimal semantic feature selection and map reducer-based multimedia data retrieval are explained in Sect. 4. The proposed SM-DHOA for enhanced multimedia data retrieval is discussed in Sect. 5. The results are evaluated in Sect. 6. Finally, the concluding statements are given in Sect. 7.

## 2 Literature survey

---

### 2.1 Related work

In 2015, Guo et al. [1] have developed an economical and effective structure called “Semantic-based Heterogeneous Multimedia Retrieval (SHMR)” to use the low cost for retrieving and storing the semantic information. Initially, this model has addressed individuality in big data environments with heterogeneous multimedia retrieval. Then, a new method was proposed for extracting and representing the semantic information. Moreover, this model has offered the multimedia data in semantic storage using a NoSQL-based technique, which was processed in a parallel way among the distributed nodes. Last, the developed model has achieved good user experience and high retrieval precision through a user feedback-driven approach and MapReduce-based retrieval technique. The suggested model has shown better efficiency in terms of economic efficiency and retrieval performance.

In 2018, Ahmad et al. [2] have presented a more effective approach named Bi-Directional Fast Fourier Transform (BD-FFT) to get the condensed binary codes from the high dimensional deep features. It was used for saving the memory and simplified the computation for efficient retrieval. The transformed codes were considered as the hash codes to access the images in the huge scale datasets through Approximate Nearest Neighbor (ANN) search methods. It has achieved better retrieval accuracy based on the shorter length codes, which were experimented on Convolutional Neural Network (CNN) features. The experimentations were carried out on seven different huge scale datasets to demonstrate the efficiency of the suggested retrieval model.

In 2020, Xia et al. [3] have developed a deep correlation mining method for developing the multimedia retrieval model using “Levinberg-Marquard algorithm-based Deep Typical Correlation Analysis method (LM-DCCA)”. This model has intended to train the diverse medial features based on the suggested algorithm. Then, the correlation among the

trained features was attained to compare the features of diverse multimedia data. This method has addressed the problem of local optima problems with the Levenberg–Marquart method. The experimentation was conducted on different datasets to show the efficiency of cross-media retrieval. This model has shown superior retrieval approaches while comparing with other progressed multimedia retrieval approaches.

In 2020, Sun et al. [4] have implemented a new graph classification-based semantic similarity model by introducing the feature reduction technique. Here, the suggested technique has initially learned the representations of vectors through neural language systems with subtree patterns. Further, similar subtree patterns were merged semantically to get the novel features. This model has offered a novel feature discrimination score for selecting the most discriminative features. The experimental analysis was conducted on real datasets to show the efficiency of the suggested retrieval method using the FRS-KELM graph classifier.

In 2018, Beltran et al. [5] have developed a new multimedia retrieval model using probabilistic Latent Semantic Analysis (pLSA) for solving the problems in big data along with data reduction strategy. This model has aimed to enhance the extraction procedure for Content-Based Multimedia Retrieval (CBMR). The developed system has focused on discussing the result of document reduction than other existing approaches, which were performed using pLSA and Latent Dirichlet Allocation (LDA) in CBMR. The suggested model has used the retrieval performance in CBMR.

In 2014, Yilmaz et al. [6] have implemented a “RELIEF-based modality weighting approach termed RELIEF-MM”. This method was the modified version of challenges like handling unbalanced datasets, challenges with multi-labeled data and noise, class-specific feature selection, and utilizing the technique through machine learning identifications. It has focused on enhanced weight estimation function using RELIEF-MM to show the reliability and representation abilities of modalities with their discrimination capacity at lesser computational complexity. The experimental analysis was carried out on “TRECVID 2007, TRECVID 2008, and CCV datasets” that have shown the robustness and accuracy of the modality weighting on multimedia data retrieval model.

In 2018, Guo et al. [7] have suggested a new heterogeneous multimedia data retrieval model using Semantic Ontology Retrieval (SOR) for storing and retrieving the ontologies

by processing the big data tools. Initially, the environment of ontology representation and semantic extraction were solved on multimedia big data. Secondly, SOR was used as the retrieval technique for model definition. Third, MapReduce was used for parallel processing of SOR to propose a new retrieval framework based on distributed nodes. Last, a user feedback method was presented for achieving a better user experience and high retrieval precision. The numerical results have shown that the designed SOR was appropriate for heterogeneous multimedia big data and semantic-based retrieval.

In 2019, Liu et al. [8] have implemented the “Deep Hashing with Multilevel Similarity Learning (DHMSL)” to learn the discriminative and compact hash codes to explore the “multilevel semantic similarity correlations of multimedia data.” Initially, the unified binary hash codes were learned for exploring the multilevel similarity correlation through the exploitation of semantic label information and the local structure. Consequently, the unified hash codes were made compact by taking the constraints like quantization and bit balance. This model has combined the Deep Neural Networks (DNN) for learning the feature representations using learned unified binary codes. The developed model has minimized the prediction errors among the outputs of the networks and unified binary codes. The simulation results were applied on two extensively used multimodal datasets to show the performance on both text and image-query-image tasks.

In 2018, Alghamdi et al. [30] have employed two types of transfer learning techniques to retrain the pre-trained VGG-Net (Fine-tuning and VGG-Net as fixed feature extractor) and obtained two new networks VGG-MI1 and VGG-MI2. In the VGG-MI1 model, the last layer of the VGG-Net model was replaced with a specific layer according to our requirements, and various functions were optimized to reduce overfitting.

In 2018, Sedik et al. [31] have suggested two data-augmentation models to enhance learnability of the Convolutional Neural Network (CNN) and the Convolutional Long Short-Term Memory (ConvLSTM)-based deep learning models (DADLMs) and, by doing so, boost the accuracy of COVID-19 detection.

In 2018, Sedik et al. [32] have implemented a deep learning model based on CNN and a hybrid model that combined CNN with ConvLSTM to compute a three-tier probability that a biometric has been tempered. Simulation-based experiments have indicated that

the alteration detection accuracy matches those recorded in advanced methods with superior performance in terms of detecting central rotation alteration to fingerprints.

## 2.2 Problem statement

Multimedia big data are dependable to retrieve semantic information concerning video, text, image, or audio from data sources. The extraction of such information is a major and challenging problem owing to their growth in media. Thus, there is a huge demand for fast multimedia retrieval models with the smallest utility gap. These challenges have occurred with the size and diversity of media in the corresponding field. Numerous approaches are developed in multimedia retrieval that has diverse features and challenges, as given in Table 1. MapReduce [1] offers better economic efficiency and also reduces investments in hardware, though this model was not able to adapt to a real-time Internet environment and also suffers from slow retrieval speed. BD-FFT [2] increases the retrieval performance and overall efficiency and also improves the retrieval accuracy with deep features. However, this model does not perform on sparse features. LM-DCCA [3] has enhanced the training with a better convergence effect and also improves the overall performance. This model does not permit users for selecting the counter samples based on the user's feedback strategy. FRS-KELM graph classifier [4] attains better classification accuracy and selected the more discriminative features. On the other hand, it is not suitable for applying the developed feature reduction approach in Convolutional Neural Networks (CNN). Prior-based probabilistic Semantic Analysis (PpLSA) [5] improves the efficiency of information reduction and also increases the number of parameters required for extracting the information. Conversely, the developed approach does not support an incremental retrieval. RELIEF-MM [6] guarantees a higher accuracy rate with better performance and also a computationally efficient one. However, this model will not use some caching strategies that affect efficiency. MapReduce [7] achieves a good user experience and high retrieval precision and it cannot be applied in real-time applications. Multimodal hashing [8] is efficiently performed on multimedia retrieval. If the indiscriminative hash bits control the hash code, then the retrieval performance can be decreased. These challenges can be further used for developing new multimedia big data retrieval along with semantic feature similarity.

---

**Table 1 Features and challenges of existing multimedia big data retrieval**

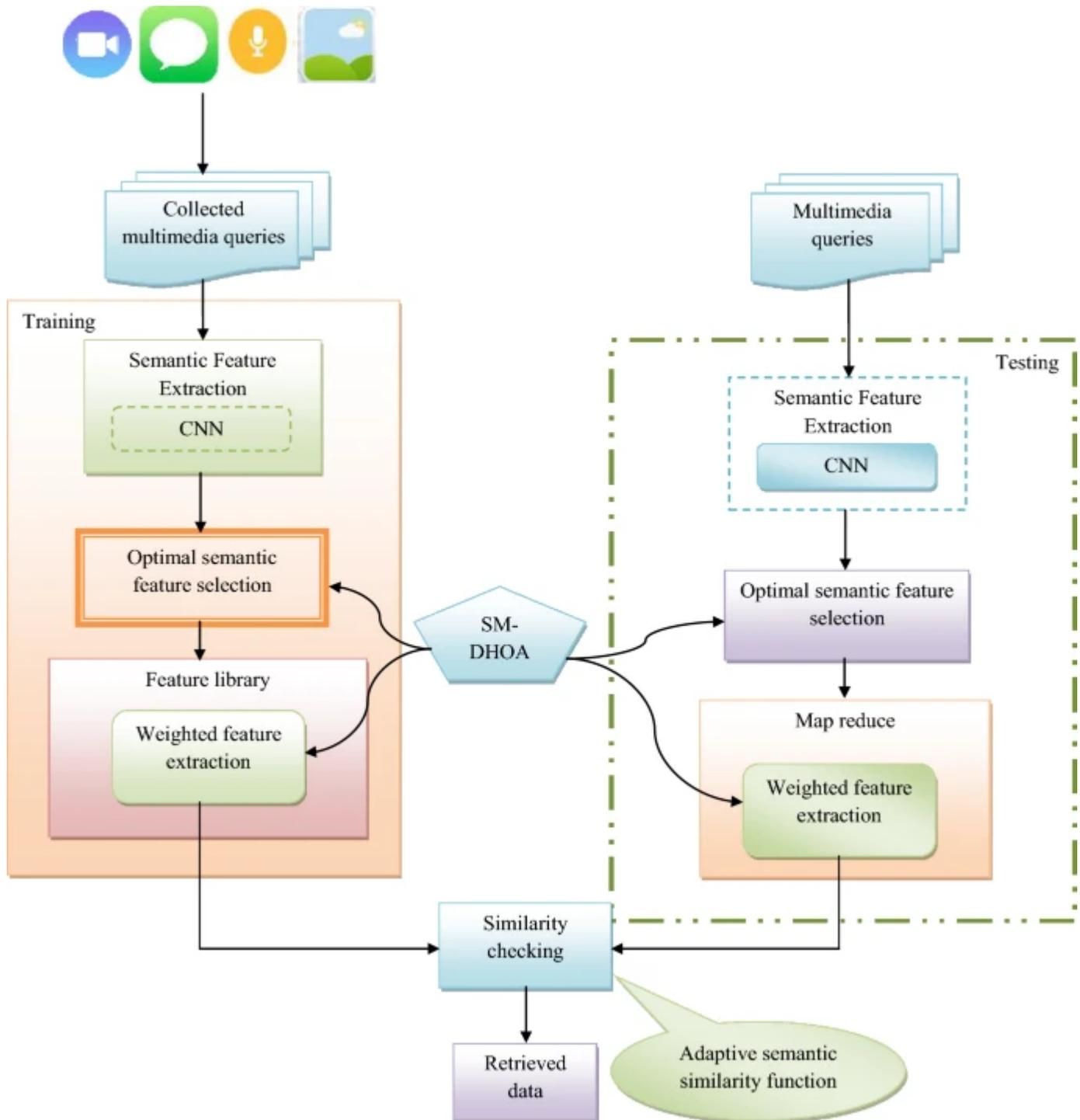
---

# 3 Proposed multimedia data retrieval using adaptive semantic similarity and deep learning

## 3.1 Architectural representation

Multimedia big data have emerged in recent years, which offers different mobile technologies and online services. Hence, more researches have been focusing on the multimedia era to target diverse constraints of big data analytics like capturing, storing, mining, indexing, and retrieval of multimedia big data. In recent days, the widespread and fast utilization of multimedia data consists of text, video, audio, and image with easy access and availability for managing the multimedia systems. The major problem in big data analytics is how to minimize the storage capacity and computational time when maintaining accurate outcomes even for small datasets. Therefore, to get semantic information about the text, video, audio, and image from data sources, the multimedia retrieval system is necessary. However, the extraction of this multimedia information is more complex owing to the growth in media. Moreover, there is a high demand for faster multimedia retrieval with the least utility gap. Conversely, it also suffers from the size of media and diversity. As the major requirement of multimedia is to match the content and other information based on the user's necessity, it has also offered best-matched data retrieval. On the other hand, there are some utility gaps among the delivered data and expected data, which happen due to the different conditions like failures in retrieval models for reaching each media data, false content description and poor query generation. These challenges must be considered while implementing a new retrieval model that assists in searching for entire items saved on the internet by any other platform through social networking sites or direct uploads. Moreover, a good multimedia retrieval model must have completeness and soundness. However, achieving the complete goodness in the retrieval model is challenging due to the retrieved irrelevant media. Likewise, completeness refers to the retrieval of entire significant multimedia data according to the user query, which is not properly attained on the internet. The multimedia big data suffer from the challenges like transmission, storage, additional compression, and analysis problems related to ensuring computing and scalability efficiency, dealing with understanding and cognition complexity, organization of heterogeneous and unstructured data, and solving the QoS and real-time requirements. Here, the technical and corresponding scientific issues promote new multimedia big data retrieval models with intelligent approaches, as depicted in Fig. 1.

Fig. 1



Multimedia data retrieval using adaptive semantic similarity and deep learning

The proposed multimedia data retrieval model consists of two stages: training and testing. Initially, the multimedia data including text, video, audio, and image as multimedia queries are collected from benchmark datasets, which are first processed to the feature extraction stage for getting the significant information. This suggested model

has used CNN for getting the semantic features for describing the semantic contents from multimedia data. The CNN can extract the most useful features through the pooling layer, which extracts the most important features for enhancing the retrieval performance by reducing the redundant data or dimensionality of data. The main advantages for the convolutional layer consist of many filters which apply convolution operation to the input to capture some special features and pass the result to the next layer. Due to these advantages, CNN is properly trained even using different images from the same class. Optimal feature selection reduces the falsely selected features. Removing the irrelevant data improves learning accuracy and reduces the computation time. The extracted semantic features are given to the feature selection process for getting the optimal semantic features by applying a newly proposed SM-DHOA technique. This process is essential for the multimedia retrieval model to get the necessary information. When the huge set of features is processed, the retrieval model suffers from computation and time complexity. Therefore, feature selection is necessary for getting the most representative features. Weighted feature extraction trial-and-error for determining the appropriate number of extracted features can be avoided. A weight-based feature extraction approach to reduce the number of features for text classification. In the training stage, the weighted features are stored in the feature library, which is used for the retrieval process in a testing stage. Moreover, in the testing stage, all the above-mentioned processes are carried out for multimedia queries. The optimally selected semantic features using the SM-DHOA are given to the map-reduce framework. Here, it consists of two functions like mapper and reducer for retrieving the appropriate multimedia data. The optimal features are mapped into the mapper and generate the intermediate data, which is further given to the reducer. The reducer performs the retrieval performance using the A-SSF. For computing this similarity function, the optimal features of both trained data and query data are multiplied by a weight function. This weight function is optimized by the SM-DHOA and using the weighted features, the semantic similarity function is computed, and hence it is termed as A-SSF. The DHO algorithm can highly balance the exploration and exploitation phases and improve the accuracy of the proposed model. SM is a new type of swarm intelligence-based algorithm. It is used for finding the best-fit solution and also applied to solve complex optimization problems. This aims to maximize the precision and F1-score for attaining an efficient retrieval process.

### 3.2 Dataset description

This developed multimedia big data retrieval model uses two datasets for experimentation, as follows.

Dataset 1: It is collected from “<http://press.liacs.nl/mirflickr/mirdownload.html>: Access Date: 2021-05-20”. It consists of 25,000 files like multimedia data like images, video, text, and audio files with different directories.

Dataset 2: It is gathered from the “[https://github.com/satishrdd/Fast\\_image\\_retrieval](https://github.com/satishrdd/Fast_image_retrieval): Access Date: 2021-05-20”. It consists of images from both query and database with different lists, where the database includes 1,000 files and the query consists of 54 files.

The input multimedia queries collected from both datasets are termed as  $\{Q_n\}$ , where  $(n = 1, 2, \dots, N)$  and the total number of multimedia data in the dataset is indicated as  $(N)$ .

### 3.3 CNN-based semantic feature extraction

It is the initial stage of the proposed multimedia big data retrieval model, which is carried out by applying CNN. The feature extraction is the essential step for developing a new retrieval model. It is the process of extracting the necessary features from multimedia data. It is required for extracting every feature of multimedia queries, which helps in retrieving the relevant multimedia data according to the queries. This stage gets the high dimensional features from the pooling layer of CNN.

CNN [27] is used in this developed multimedia big data retrieval model because of its automatic and efficient feature extraction capabilities without human intervention, which is also a computationally efficient approach. Moreover, this architecture is less dependent on pre-processing, and thus, CNN is more efficient in this model. Moreover, it is simple and easy for processing to offer higher accuracy in the retrieval process. CNN is a feedforward network, which consists of different layers like “convolutional and pooling (or subsampling)” layers that are grouped into modules. The multimedia query is given as the input for CNN, which is processed into different stages “convolutional and pooling” for attaining the specific representations, and finally, the output classes are attained from the output layer. This paper extracts the necessary semantic features from the pooling layers.

The input multimedia queries  $\{Q_n\}$  are given to the input layer of CNN, which is served as the feature extractors [30] for learning the feature representations from the  $\{Q_n\}$ . The convolution layer consisted of different neurons that are organized into

feature maps, in which each neuron has its receptive field. Each neuron is correlated with a neighborhood of neurons in the earlier layer through a set of trainable weights, namely filter bank. Here, a new feature map is computed by convolving the inputs through the learned weights. Moreover, the convolved results are forwarded with a non-linear activation function. Here, the  $m$ th output feature map  $(FE_{\{m\}})$  is computed in Eq. (1).

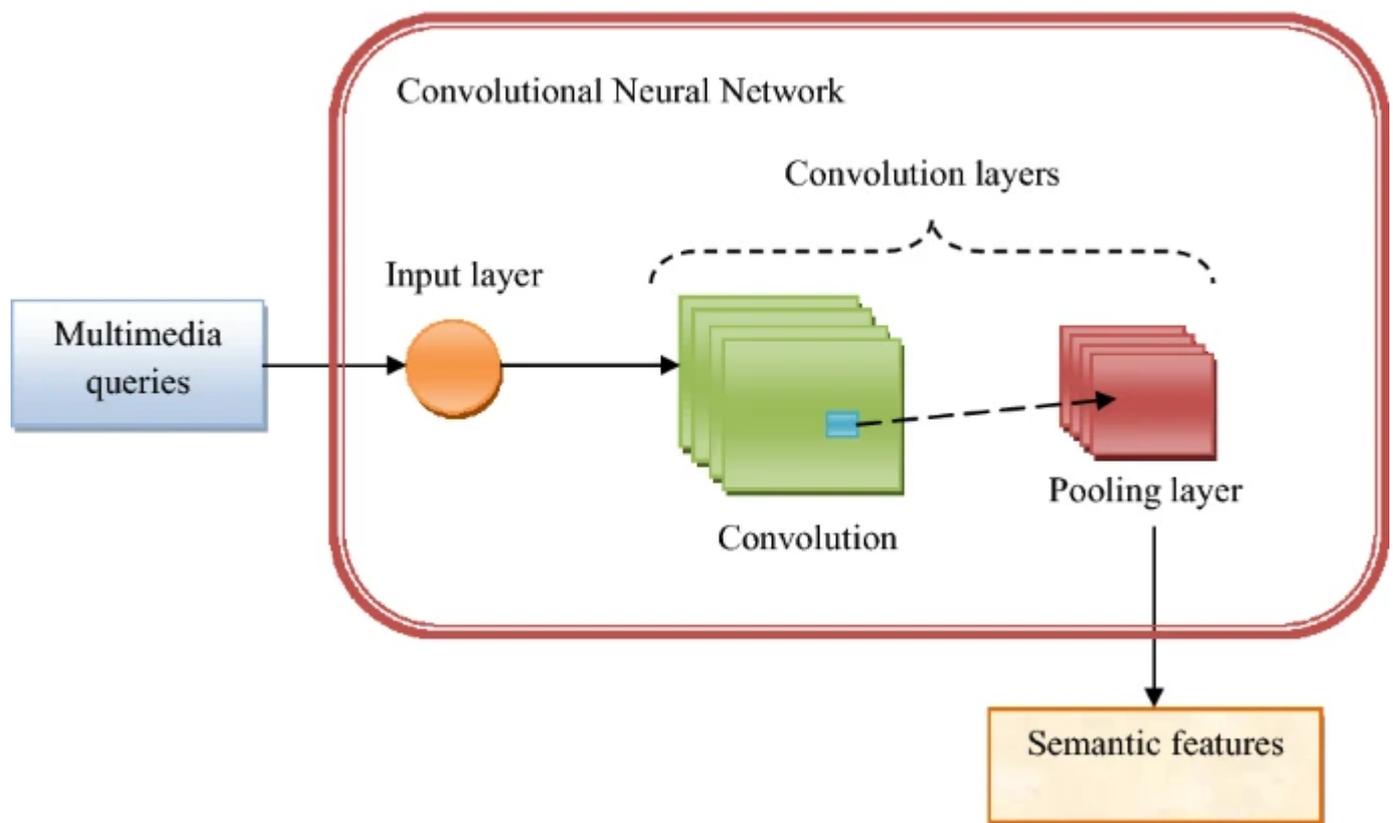
$$FE_{\{m\}} = \zeta \left( R_{\{m\}} * Q_{\{n\}} \right) \quad (1)$$

Here, the term  $(R_{\{m\}})$  denotes the convolutional filter concerning with  $m$ th feature map,  $(\zeta)$  represents the non-linear activation function, and the 2D convolutional operator is denoted as  $(\zeta \left( * \right))$  that is responsible for extracting the non-linear features from the input  $(Q_{\{n\}})$ . The pooling layer of CNN [31, 32] is used for minimizing the spatial resolution of the feature maps to get the spatial invariance for input translations and distortions. Most of the model employs average pooling aggregation layers for propagating the average of entire input values. The max-pooling layers choose the largest element in the corresponding receptive field as formulated in Eq. (2).

$$FE_{\{mps\}} = \mathop{\max} \limits_{\left\{ \left( \{u,v\} \right) \in \delta_{\{ps\}} \right\}} Q_{\{nmuv\}} \quad (2)$$

In Eq. (2), the element at location  $(\left( \{u,v\} \right))$  included through the pooling region  $(\delta_{\{ps\}})$  is termed as  $(Q_{\{nmuv\}})$  that is used for embodying a receptive field about the position  $(\left( \{p,s\} \right))$ , and the output of the pooling operation is denoted as  $(\text{FE}_{\{mps\}})$  that is correlated through the  $m$ th feature map. Therefore, the final semantic features  $(\text{FE}_{\{m\}})$  are attained using the pooling layer of CNN, where  $(m = 1, 2, \dots, M)$  and  $(M)$  indicate the total number of semantic features and it is obtained as 784 features, which is further given to the retrieval process. This CNN-based semantic feature extraction is diagrammatically represented in Fig. 2.

**Fig. 2**



CNN-based semantic feature extraction for multimedia retrieval process

## 4 Optimal semantic feature selection and map reducer-based multimedia data retrieval

### 4.1 Optimal semantic feature selection

The major contribution of this developed multimedia data retrieval model is to select the optimal semantic features from the extracted CNN-based semantic features  $\{\text{FE}\}_m$ . The feature selection is the process of reducing the noisy, excess, and extraneous multimedia feature subset to select the optimal features. Finally, the most significant and optimal features are selected using the SM-DHOA from the extracted semantic feature subsets. The optimal feature selection is performed using the SM-DHOA technique to reduce the retrieval time. The optimally selected semantic features using SM-DHOA are termed as  $\{\text{FS}\}_{m^*}$ , where  $(m^* = 1, 2, \dots, \{\text{OM}\})$  and  $\{\text{OM}\}$  represents the total number of optimal semantic features. The objective model here is minimizing the correlation between features.

The correlation  $\{\text{CT}\}$  among the optimal semantic features  $\{\text{FS}\}_{m^*}$  is found for attaining the minimum correlated features, which is the primary objective of the developed model as formulated in Eq. (3).

$$FF_1 = \mathop{\arg \max} \limits_{\left\{ \left\{ FS_{m^*} \right\} \right\}} \left( \frac{1}{CT} \right),$$

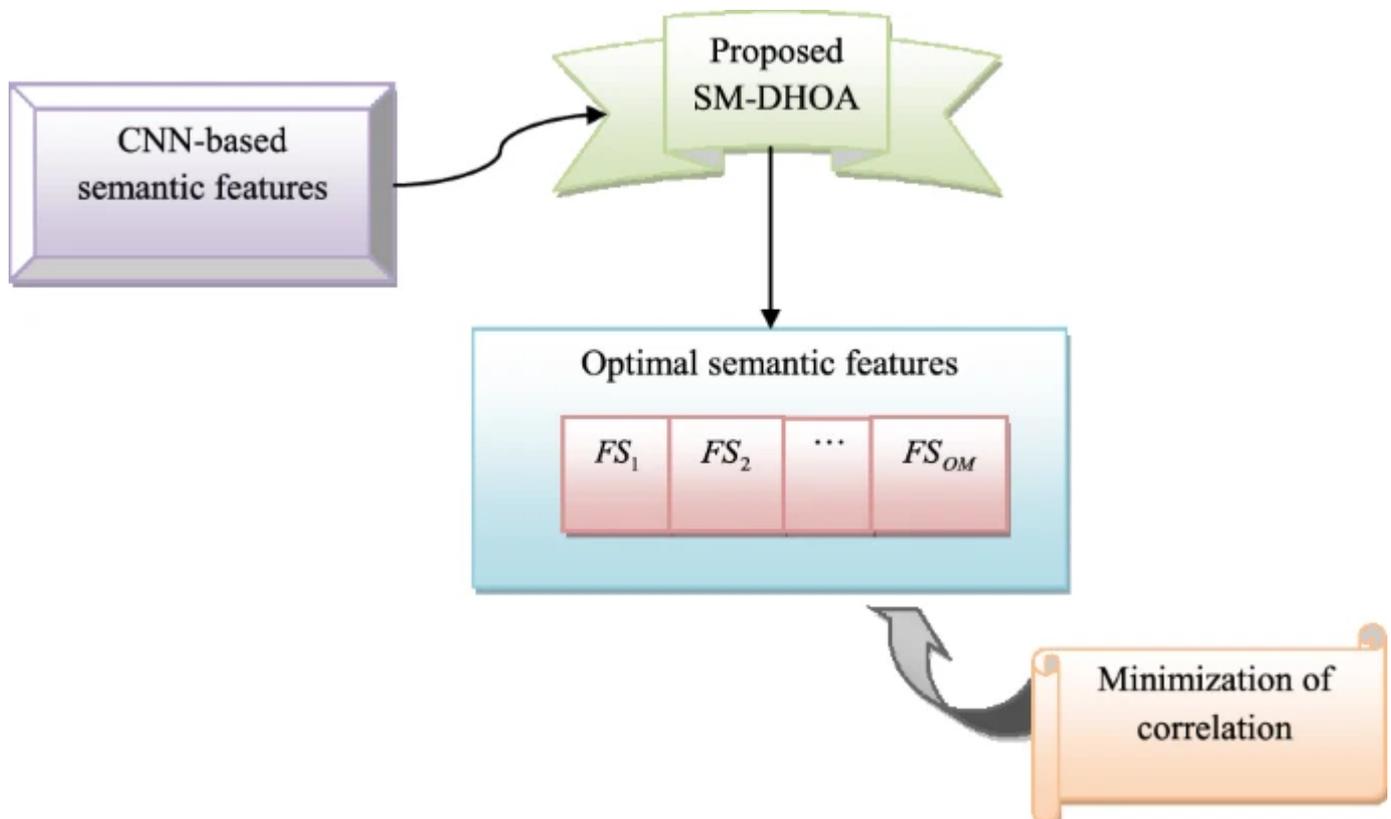
(3)

$$\text{CT} = \frac{\left\{ \left\{ \text{OM} \right\} \sum \{fs\} - \sum \{fs \sum \{fh\} \} \right\} \sqrt{\left( \left\{ \left\{ \text{OM} \right\} \sum \{fs^2\} - \left( \sum \{fs \} \right)^2 \right\} \right) - \left( \left\{ \left\{ \text{OM} \right\} \sum \{fh^2\} - \left( \sum \{fh \} \right)^2 \right\} \right)}}{}$$

(4)

Here, the two features in the optimal semantic feature set are represented as  $(fs)$  and  $(fh)$ , and the total number of extracted optimal semantic features is indicated as  $(\text{OM})$ . It is attained as 10. Hence, the proposed multimedia retrieval model aims to get unique semantic features with minimum correlation by the SM-DHOA. This optimal semantic feature selection enhances multimedia retrieval efficiency. The optimal semantic feature selection is shown in Fig. 3.

Fig. 3



## 4.2 Adaptive semantic similarity in map–reduce framework

This phase is used for retrieving the multimedia data using the map–reduce framework for handling the big data. The input to the map reduce–based retrieval process uses the optimal semantic features  $(\{\text{FS}\}_{m^*})$ . The map reduce model consist of different phases such as mapper function and reducer function. Here, the mapper function is utilized for mapping the optimal features to provide a transitional values and the reducer function is performed for processing the transitional values corresponding with the equal transitional features. In the map–reduce model, the input and output are considered as the query id and returned list, respectively. The optimally selected semantic features are given to the mapper, where these features are mapped together for getting the intermediate results. Further, the reducer finds the similarity function for efficient retrieval, where the weight function is multiplied with the features of the reducer and feature library. This weight is optimized using the SM–DHOA for developing the A–SSF. The weighted feature extraction of the feature library  $(\{\text{FE}\}_{tr}^{\{\{\text{TR}\}\}})$  is formulated based on Eq. (5).

$$\{\text{FE}\}_{tr}^{\{\{\text{TR}\}\}} \left( \{\text{new}\} \right) = \{\text{FE}\}_{tr}^{\{\{\text{TR}\}\}} \times \text{wg}_{tr}^{\{\{\text{TR}\}\}} \quad (5)$$

Here, the new semantic feature set in the feature library after multiplying the weight function is represented as  $(\{\text{FE}\}_{tr}^{\{\{\text{TR}\}\}} \left( \{\text{new}\} \right))$  and the weight function used for the feature library is termed as  $(\text{wg}_{tr}^{\{\{\text{TR}\}\}})$ , where  $(\{\text{FE}\}_{tr}^{\{\{\text{TR}\}\}})$  denotes the features present in the feature library. Similarly, the features attained from the reducer are considered as  $(\{\text{FS}\}_{tr}^{\{\{\text{Rd}\}\}})$ , which is multiplied with the weight function for getting the better retrieval results that are derived in Eq. (6).

$$\{\text{FS}\}_{tr}^{\{\{\text{Rd}\}\}} \left( \{\text{new}\} \right) = \{\text{FS}\}_{tr}^{\{\{\text{Rd}\}\}} \times \text{wg}_{tr}^{\{\{\text{Rd}\}\}} \quad (6)$$

Here, the new semantic weighted feature set is denoted as  $\{\text{FS}\}_{tr}^{\{\{\text{Rd}\}\left(\{\{\text{new}\}\}\right)\}}$ , and the weight function employed for the reducer is represented as  $(\text{wg}_{tr}^{\{\{\text{Rd}\}\}})$ . Therefore, the weighted features of both feature library and reducer are checked with adaptive similarity function, where the weights  $(\text{wg}_{tr}^{\{\{\text{Rd}\}\}})$  and  $(\text{wg}_{tr}^{\{\{\text{TR}\}\}})$  are optimized using the SM-DHOA. Therefore, the adaptive similarity function is formulated in the reducer as given in Eq. (7).

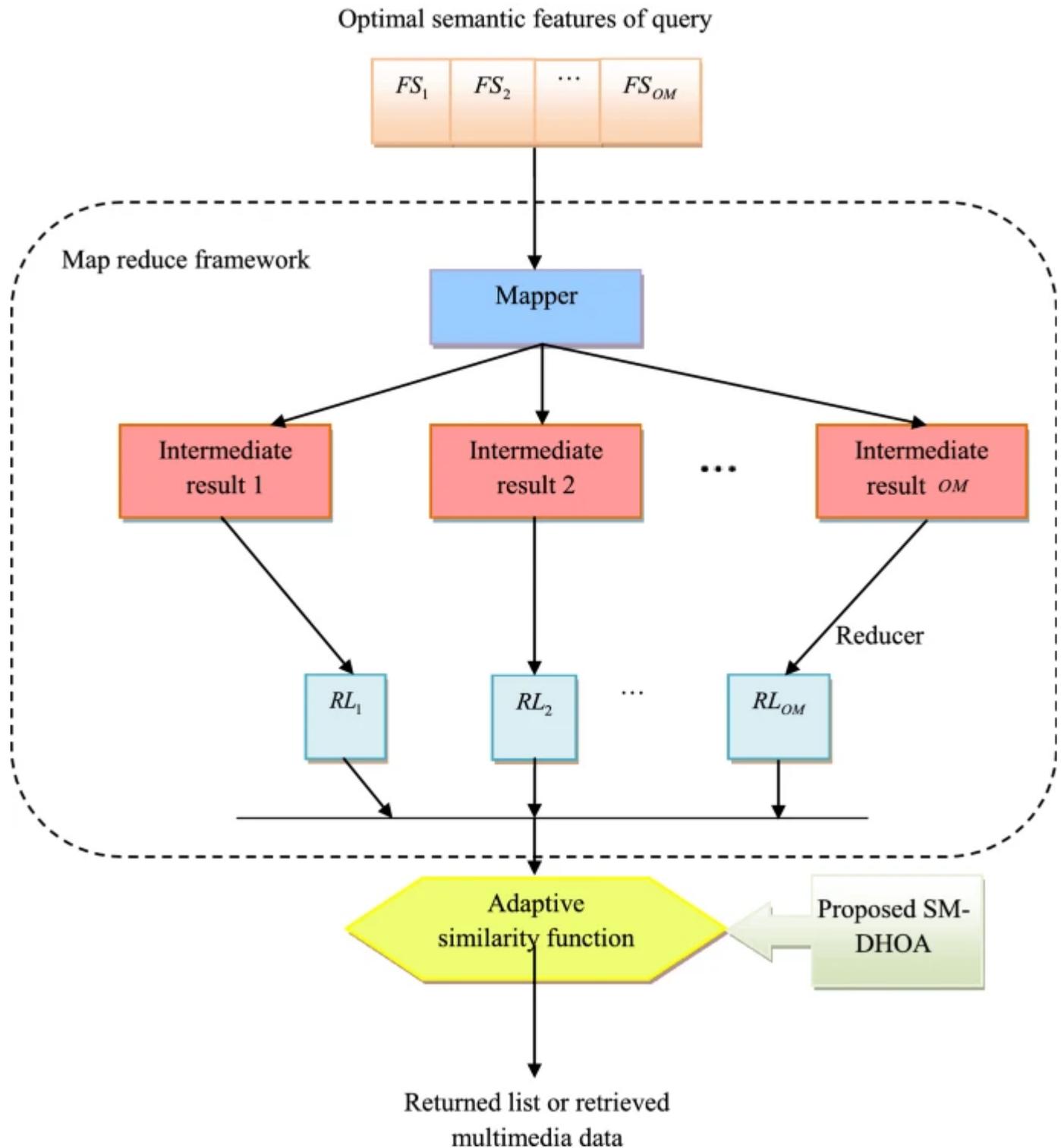
$$\begin{aligned} & \{\text{SF}\}\left(\{\{\text{FE}\}\}_{tr}^{\{\{\text{TR}\}\}\left(\{\{\text{new}\}\}\right)\}}, \{\text{FS}\}_{tr}^{\{\{\text{Rd}\}\}\left(\{\{\text{new}\}\}\right)\}} \right) \\ & \left( \frac{1}{E_{\{\{\text{FE}\}\}_{tr}^{\{\{\text{TR}\}\}\left(\{\{\text{new}\}\}\right)\}}} + E_{\{\{\text{FS}\}\}_{tr}^{\{\{\text{Rd}\}\}\left(\{\{\text{new}\}\}\right)\}} \right) \\ & \left( \left( \{\{\text{FE}\}\}_{tr}^{\{\{\text{TR}\}\}\left(\{\{\text{new}\}\}\right)\}} \right) \cap \left( \{\{\text{FS}\}\}_{tr}^{\{\{\text{Rd}\}\}\left(\{\{\text{new}\}\}\right)\}} \right) \right) \end{aligned}$$

(7)

Here, the total number of weighted features of  $(\{\{\text{FE}\}\}_{tr}^{\{\{\text{TR}\}\}\left(\{\{\text{new}\}\}\right)\})$  and  $(\{\{\text{FS}\}\}_{tr}^{\{\{\text{Rd}\}\}\left(\{\{\text{new}\}\}\right)\})$  are represented as  $(E_{\{\{\text{FE}\}\}_{tr}^{\{\{\text{TR}\}\}\left(\{\{\text{new}\}\}\right)\}})$  and  $(E_{\{\{\text{FS}\}\}_{tr}^{\{\{\text{Rd}\}\}\left(\{\{\text{new}\}\}\right)\}})$ , respectively.

To summarize, the mapper function considers the pairs of record values and attained the intermediate results. Then, each pair is processed to get the multimedia retrieval process using the adaptive similarity function by the reducer. Finally, the returned list is attained as the multimedia retrieval content. The adaptive semantic similarity in the map-reduce framework is shown in Fig. 4.

**Fig. 4**



Adaptive semantic similarity in map reduce framework

## 5 Proposed SM-DHOA for enhanced multimedia data retrieval

### 5.1 Developed SM-DHOA

This proposed multimedia data retrieval model uses SM-DHOA for selecting the optimal features for getting the significant features. The suggested model has offered a new adaptive semantic similarity in map-reduce framework using the SM-DHOA for

optimizing the weight function in the reducer model. This algorithm aims to enhance the multi-media data retrieval for big data.

DHOA [25] and SMO [26] algorithms are selected because of their features, in which DHOA offers efficient balancing among exploitation and exploration stages with superior searching capabilities, which has determined the optimal position effectively. It provides excellent sensitivity and solves real-time optimization problems owing to the robustness and ability of faster convergence, though DHOA suffers from premature convergence. Thus, there is a need of adopting a new SM-DHOA with the SMO algorithm. It offers convergence more quickly and requires fewer resources. Hence, SMO is adopted with DHOA for improving multi-media data retrieval.

In this proposed SM-DHOA, the propagation through the position angle of DHOA is replaced by the global leader phase of the SMO algorithm. Hence, if  $(f < 1)$ , then the solutions are updated based on the DHOA with leader and successor positions, or else the solutions are updated based on SMO algorithm with global leader phase.

The DHOA is motivated by considering the nature of the humans to hunt the deer with the mobility of hunters to get the best location with “leader and successor” solutions. The deer positions are updated until finding the deer. First, the deer is enclosed by the hunters with some constraints like “deer position and wind angle”. This algorithm has special characteristics like efficient teamwork among hunters for finding the deer, and better-attacking procedure to achieve the optimal location of prey through the position of leader and successor. Moreover, the prey or deer has diverse special characters as given in [25], which are studied for an efficient hunting process. DHOA is derived by initializing the population as given in Eq. (8).

$$X = \{X_1, X_2, \dots, X_p\}; \quad 1 < k \leq p$$

(8)

Here, the number of hunters in the Xth population is indicated as  $(p)$ . Moreover, the constraints like deer position and wind angle are derived in Eqs. (9) and (10), respectively.

$$\phi_k = 2\pi \cdot r$$

(9)

$$\omega_k = \varphi + \pi r,$$

(10)

In the aforementioned equation, wind angle  $(\omega)$  is necessary due to the circular search space. Here, the wind angle is termed as  $(\varphi)$ , the recent iteration is represented as  $(k)$ ,  $(\varphi)$  indicates the deer's position angle and  $(r)$  specifies the random number in the range of  $[0, 1]$ . The positions are propagated based on diverse constraints. Initially, the position of the optimal search area is anonymous and so, the solutions are randomly considered that is near to the optimal space through formulating the fitness function. As mentioned above, the solutions are considered as leader position and successor position, which are represented as  $(X^{ldr})$  and  $(X^{Sur})$ . These are considered as the first best position and successor position of the hunters, respectively.

(i) "Propagation through Leader's position": At the initial stage, the best positions are formulated, and then every individual in the population is tried to get the best position. The enclosing behavior to find deer is formulated in Eq. (11).

$$X_{k+1} = X^{ldr} - Y \cdot f \cdot |B \cdot X^{ldr} - X_k|$$

(11)

In Eq. (24), the position at the next iteration is denoted as  $(X_{k+1})$ , the random number of wind speed is represented as  $(f)$  that varies among  $[0, 2]$ , the position of the current iteration is termed as  $(X_k)$  and coefficient vectors are represented as  $(Y)$  and  $(B)$  as formulated in Eqs. (12) and (13), respectively.

$$Y = \frac{1}{4} \log \left( k + \frac{1}{k_{\max}} \right) p,$$

(12)

$$B = 2 \cdot \text{rnd}$$

(13)

Here, the term  $(p)$  is a parameter with the range of  $[-1, 1]$ , the maximum iteration is considered as  $(j_{\max})$ , and  $(rnd)$  specifies the random number in the bounding range of  $[-1, 1]$ .

(ii) “Propagation through position angle” based on global leader position of SMO algorithm: SMO is a population-based technique inspired by the social interactions among the spider monkeys, which has been considered the fission–fusion social strategy for intelligent foraging behavior. The monkeys have different groups of leaders like local and global leaders and corresponding group members. This hierarchy is used for maintaining defensive boundaries and social bonds. It has different phases like “local leader phase, global leader phase, local leader learning phase, global leader learning phase, local leader decision phase, and global leader decision phase”. In this developed SM-DHOA algorithm global leader phase is considered for updating the position. Once, the local leader phase is performed for finding the food, the entire spider monkeys reevaluate their positions based on the experience of members of the local group and global leader. This process helps in identifying the most suitable positions through Eq. (14).

$$X_{\text{new}kj} = X_{kj} + rn \left[ 0,1 \right] \left( GL_j - X_{kj} \right) + rn \left[ -1,1 \right] \left( X_{sj} - X_{kj} \right).$$

(14)

In Eq. (14), the term  $(X_{\text{new}kj})$  expresses the new position update based on the global leader, the  $k$ th spider monkey at  $j$ th dimension is formulated as  $(X_{kj})$ , the random number is denoted as  $(rn)$ , the global leader position at  $j$ th dimension is formulated as  $(GL_j)$  and the  $(s^{\text{th}})$  spider monkey at  $j$ th dimension is indicated as  $(X_{sj})$ , where an arbitrarily chosen index is given as  $(j \in \{1, 2, \dots, J\})$ .

(iii) “Propagation through Successor’s position”: The exploration process of DHOA is formulated by altering the vector  $(B)$  with the enclosing behavior. When  $(B < 1)$  the random search is initialized. Thus, the position is updated with the position of the successor by considering the obtained first-best solution, which permits a global search that is derived in Eq. (15).

$$X_{k+1} = X^{\text{Sur}} - Y \cdot f \cdot \left| B \cdot X^{\text{Sur}} - X_j \right|.$$

(15)

Here, the successor position of the search agent is represented as  $(X^{\text{Sur}})$  in the recent population.

Hence, the designed SM-DHOA is formulated based on the random parameter  $(f)$ , which decides the position updating through the SMO and DHOA. The pseudo-code of the proposed algorithm is given in Algorithm 1.

### Algorithm 1: Proposed SM-DHOA

Initialization of population  $X$

Obtain first and second-best **solutions** like  $X^{ldr}$  and  $X^{Sur}$ , respectively.

Start

While ( $k < k_{max}$ )

for every solution in  $X$

Compute the fitness function of every solution

Determine the random parameters

if( $f < 1$ )

if( $|B| \geq 1$ )

Position update based on DHOA

Update the search agents based on leader position by Eq. (12)

Else

Update the search agents based on successor position by Eq. (15)

End if

Else

Position update based on SMO

Update the search agents based on **the** global leader phase of SMO using Eq. (14)

End if

End for

Compute fitness function of every solution

Update  $X^{ldr}$  and  $X^{Sur}$

$k = k + 1$

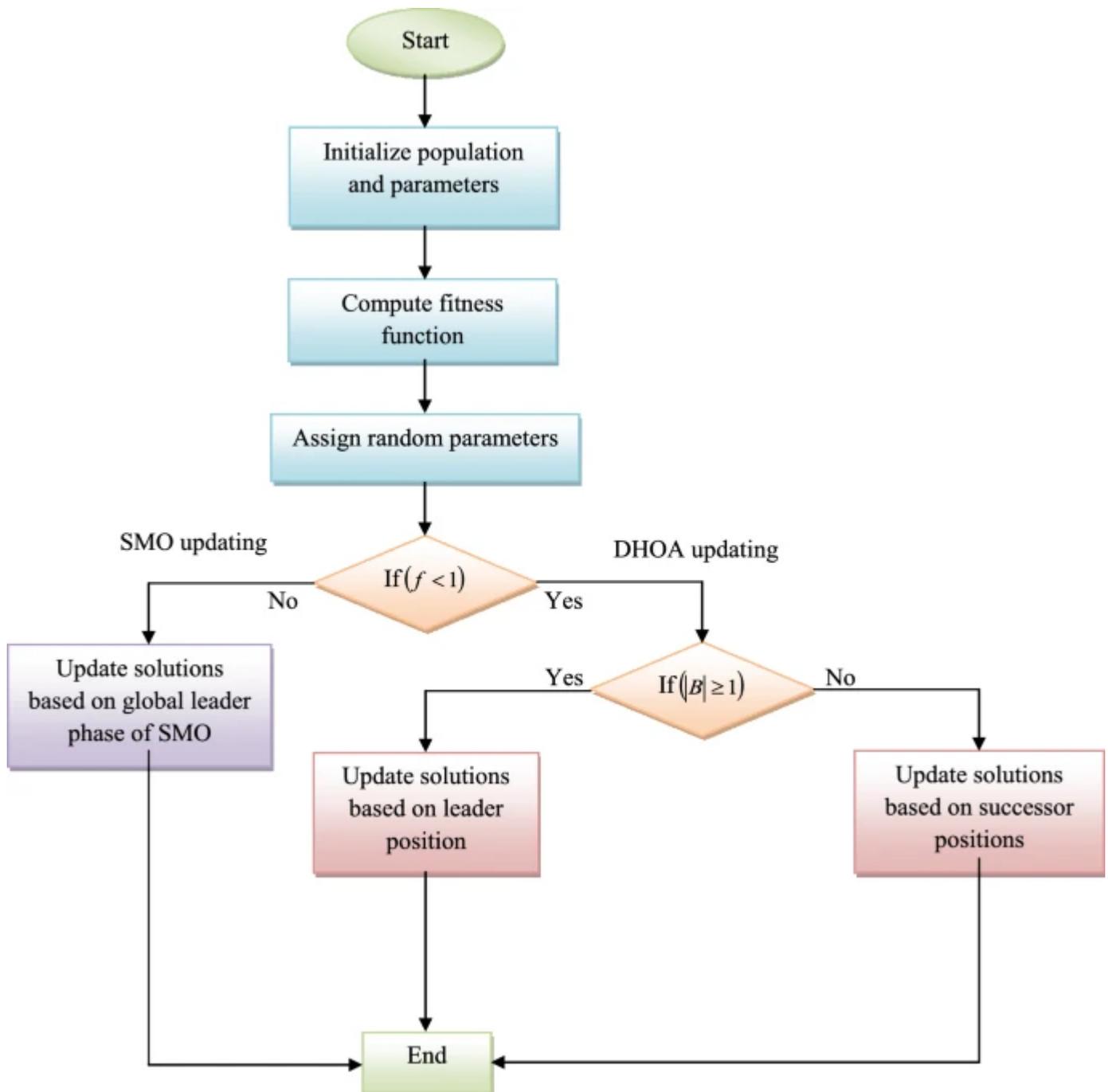
End while

Return  $X^{ldr}$

Terminate

This developed SM-DHOA is inspired by the hybridization concept, which offers efficient convergence behavior for the proposed multi-media data retrieval model. The flowchart of the designed algorithm is given in Fig. 5.

Fig. 5



Flowchart of the designed SM-DHOA algorithm

## 5.2 Objective model for a-ssf in data retrieval

The proposed multimedia data retrieval model focuses on efficient retrieval using a new SM-DHOA with optimal semantic feature selection with A-SSF. The weighted feature extraction is performed to implement the A-SSF computation and the optimization [33] strategy is used to tune the weight function, thus calling the similarity function A-SSF. Here, the map-reduce framework is used for retrieval by optimizing the weight function of features in the reducer for increasing the performance in terms of precision and F1-score. The major multi-objective function  $(FF_{2})$  for A-SSF using SM-DHOA is derived in Eq. (16).

$$FF_{2} = \mathop{\arg \min}_{\{wg_{tr}^{Rd}, wg_{tr}^{TR}\}} \left( \frac{1}{prn + F1 - \text{score}} \right). \quad (16)$$

Here, precision is “the ratio of positive observations that are predicted exactly to the total number of observations that are positively predicted” as formulated in Eq. (17).

$$prn = \frac{te^{ps}}{te^{ps} + fe^{ps}}. \quad (17)$$

The F1-score is the “harmonic mean between precision and recall. It is used as a statistical measure to rate performance” as derived in Eq. (18).

$$F1 - \text{score} = \frac{2te^{ps}}{2te^{ps} + fe^{ps} + fe^{ng}}. \quad (18)$$

Here, terms  $(te^{ps})$ ,  $(te^{ng})$ ,  $(fe^{ps})$  and  $(fe^{ng})$  refer to the “true positives, true negatives, false positives, and false negatives,” respectively. The multi-objective function focuses on improving the retrieval performance through adaptive semantic similarity in the map-reduce framework using SM-DHOA.

## 6 Results and discussions

### 6.1 Experimental setup

The proposed multimedia big data retrieval was developed in MATLAB 2020a, and the experimental analysis was carried out. From the datasets, video to video and audio to an image is analyzed using dataset 1 and also text to image and image to image is analyzed using dataset 2. Here, the performance of the proposed model was compared over the conventional methods like PSO [28], GWO [29], DHOA [25], and SMO [26] and existing models like BD-FFT [2], FRS-KELM [4], and RELIEF-MM [6]. The experimental analysis was performed by considering the number of iteration as 10 and the number of populations as 10.

## 6.2 Performance measures

Various performance metrics are used for evaluating the performance, which is described below.

(a) MCC: “correlation coefficient computed by four values”.

$$\text{MCC} = \frac{(t_{ps} \times t_{ng} - f_{ps} \times f_{ng})}{\sqrt{(t_{ps} + f_{ps})(t_{ps} + f_{ng})(t_{ng} + f_{ps})(t_{ng} + f_{ng})}}$$

(19)

(b) NPV: “probability that subjects with a negative screening test truly don't have the disease”.

$$\text{NPV} = \frac{t_{ng}}{f_{ng} + t_{ng}}$$

(20)

(c) FDR: “the number of false positives in all of the rejected hypotheses”.

$$\text{FDR} = \frac{f_{ps}}{f_{ps} + t_{ps}}$$

(21)

(d) FPR: “the ratio of the count of false-positive predictions to the entire count of negative predictions”.

$$\text{FPR} = \frac{f_{ps}}{f_{ps} + t_{ng}}$$

(22)

(e) FNR: “the proportion of positives which yield negative test outcomes with the test”.

$$\text{FNR} = \frac{fe^{ng}}{te^{ng} + te^{ps}}$$

(23)

(f) Sensitivity: “the number of true positives, which are recognized exactly”.

$$\text{Se} = \frac{te^{ps}}{te^{ps} + fe^{ng}}$$

(24)

(g) Specificity: “the number of true negatives, which are determined precisely”.

$$\text{Sp} = \frac{te^{ng}}{te^{ng} + fe^{ps}}$$

(25)

(h) Accuracy: It is a “ratio of the observation of exactly predicted to the whole observations”.

$$\text{Ac} = \frac{\left( te^{ps} + te^{ng} \right)}{\left( te^{ps} + te^{ng} + fe^{ps} + fe^{ng} \right)}$$

(26)

(i) Recall: It is measured “as the number of true positives divided by the total number of true positives and false negatives.”

$$\text{Re} = \frac{te^{ps}}{te^{ps} + fe^{ng}}$$

(27)

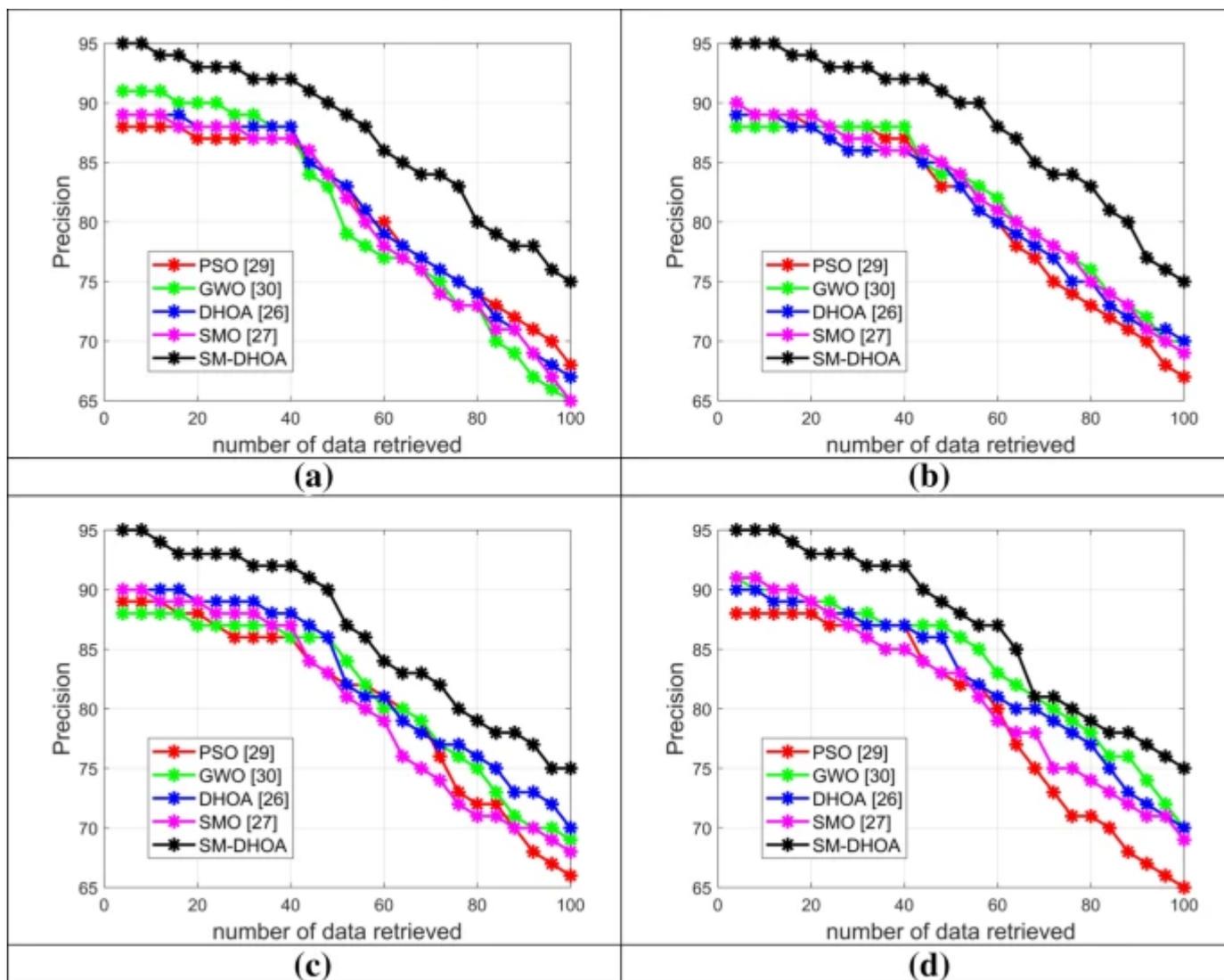
## 6.3 Analysis of precision

The performance of the designed multimedia big data retrieval model using SM-DHOA with A-SSF is analyzed in terms of precision by varying the number of multi-media data

retrieved, which is depicted in Figs. [6](#) and [7](#) for meta-heuristic-based algorithms and conventional models, respectively. When considering the number of the multi-media data retrieved as 20 for text to image retrieval, the suggested SM-DHOA gets 7.9%, 4.3%, 6.7%, 6%, 4.4%, 6.7%, and 5.5% higher precision rates than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. For an image to image retrieval, the proposed SM-DHOA attains 5.7%, 4.5%, 6.97%, 5%, 6.9%, 6.9%, and 6.9% maximum precision rate than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while taking the number of multi-media data retrieved as 40. Similarly, the performance of the developed SM-DHOA is analyzed in terms of precision with the number of multimedia data retrieved as 60, which is 3.7%, 5%, 2.4%, 7.6%, 2.4%, 1.2%, and 9% superior to PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for a video to video retrieval. Likewise, the efficiency of the recommended SM-DHOA for audio to image retrieval is 11%, 1.28%, 2.5%, 6.7%, 8%, 6.75%, and 5.3% progressed than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while taking the number of multi-media data retrieved as 80. Finally, the proposed model shows better performance by comparing with the existing approaches.

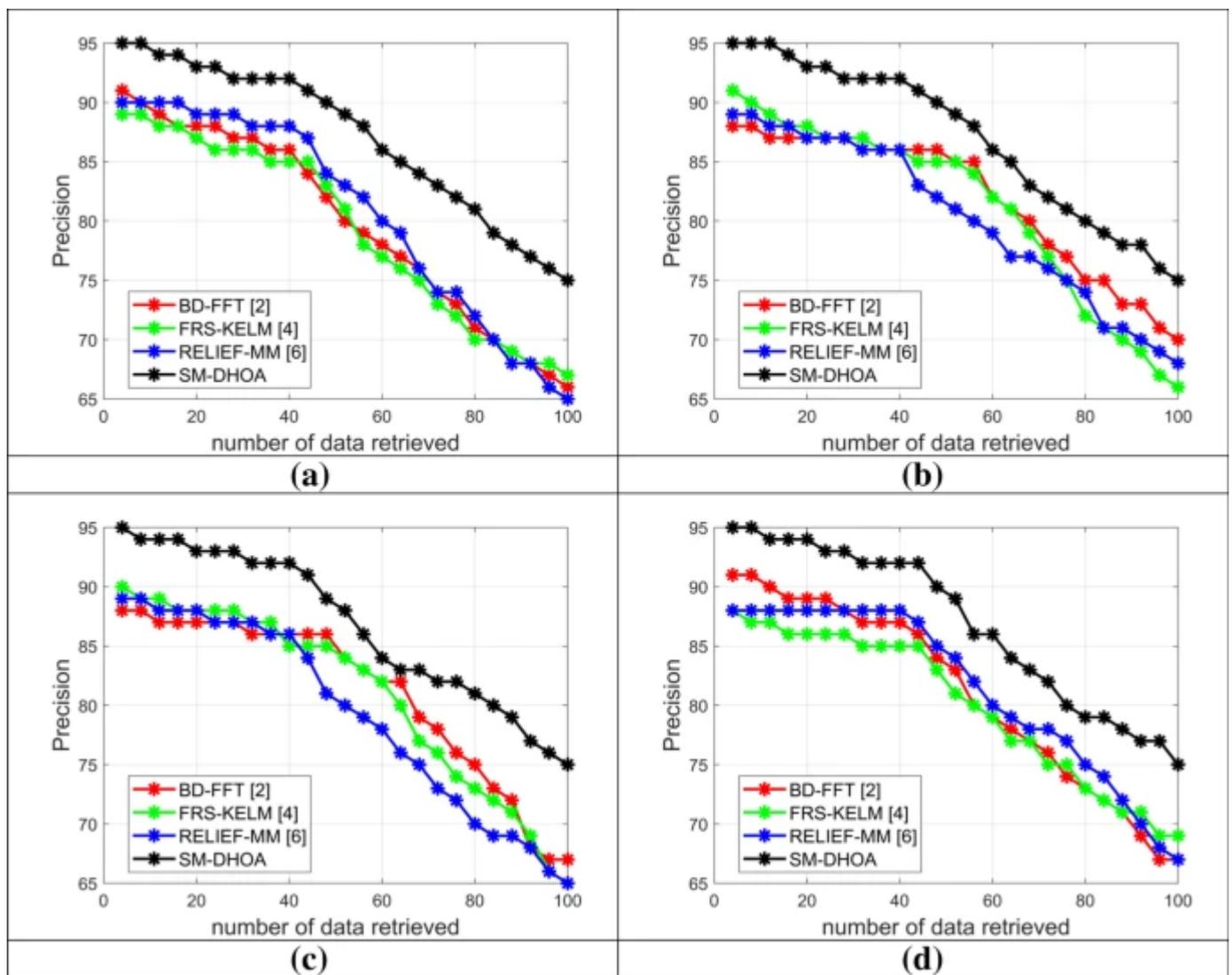
---

## Fig. 6



Performance analysis on the precision of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for a text to image retrieval, b image to image retrieval, c video to video retrieval, and d audio to image retrieval

Fig. 7



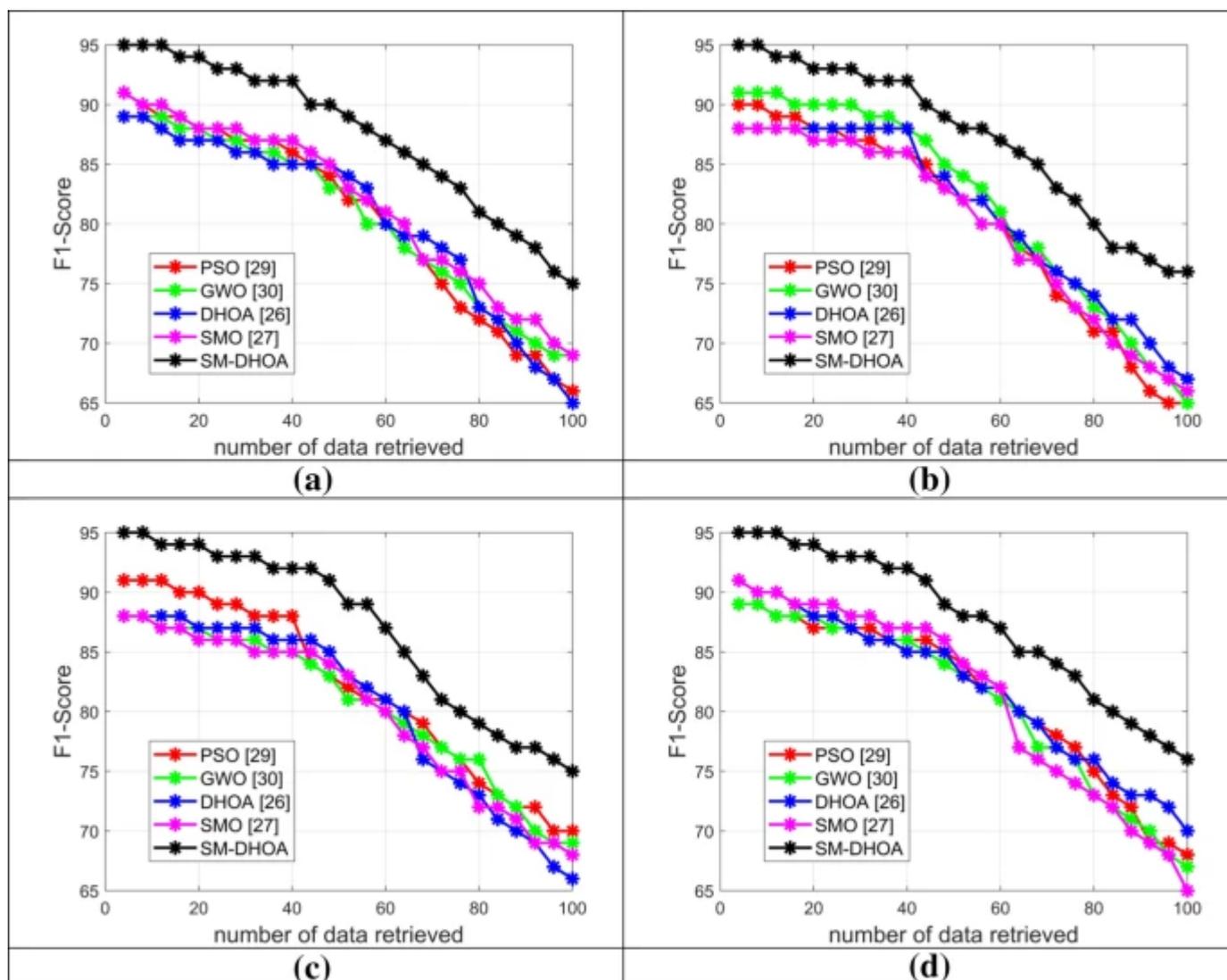
Performance analysis on the precision of the proposed Multimedia Big Data Retrieval model with different machine learning-based algorithms for a text to image retrieval, b image to image retrieval, c video to video retrieval, and d audio to image retrieval

## 6.4 Analysis of F1-score

For the statistical analysis, the number of the multi-media data retrieval was taken as  $R@100$  for image retrieval. In our proposed model, image 1 of the actual and predicted images are compared in terms of recall which is regarded as  $R@1$ , further the set of 2 of the images are compared in terms of recall which is regarded as  $R@2$ . Similarly, the comparison is made for the  $R@100$  images. The proposed multimedia big data retrieval is evaluated in terms of F1-score with different optimization-based algorithms and classifiers as given in Figs. 8 and 9, respectively. The F1-score of the suggested SM-DHOA shows the better performance, which is 8%, 9.4%, 10.7%, 6.8%, 8%, 6.8%, and 9.4% advanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, by considering the number of multi-media data retrieved as 40 for text to

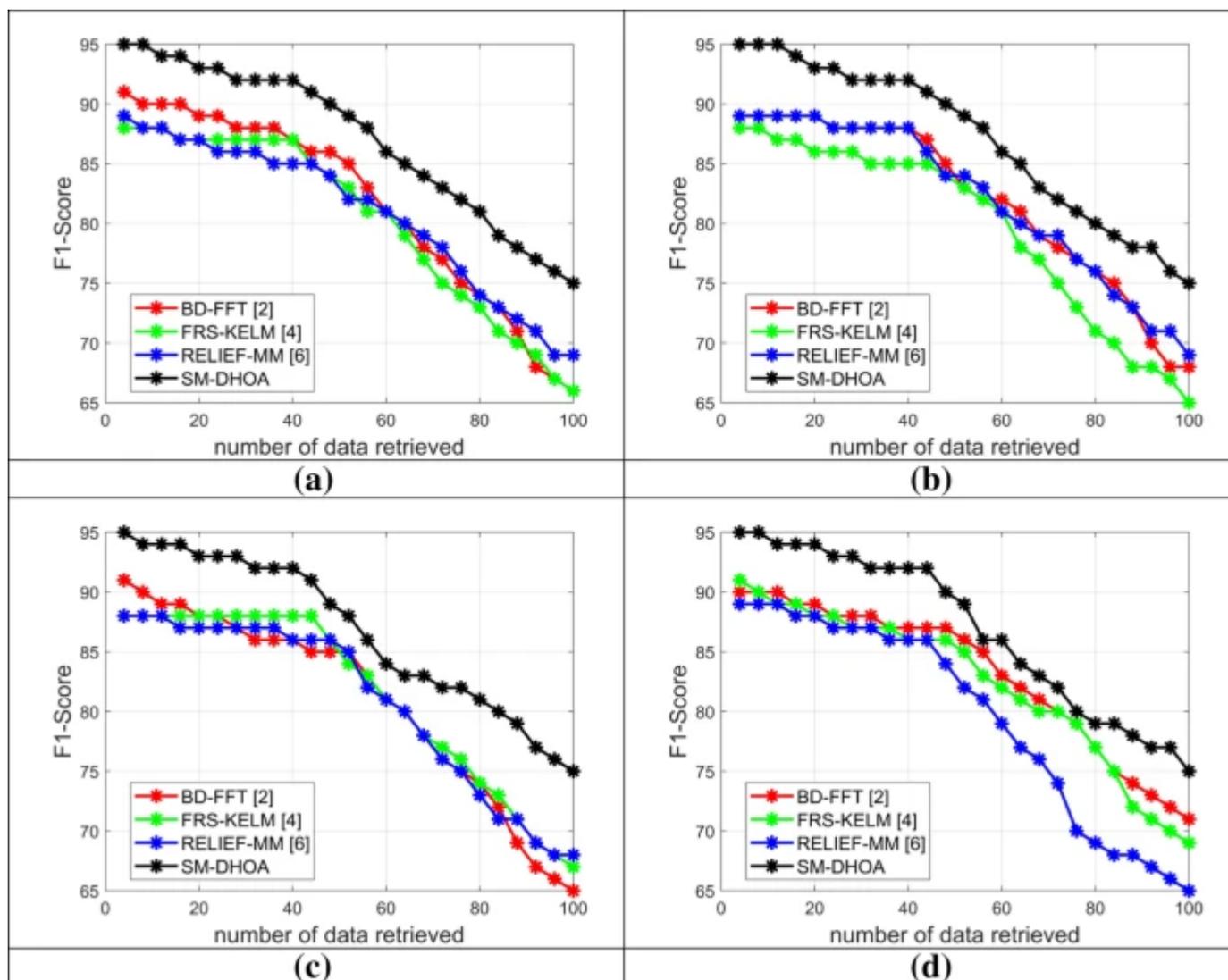
image retrieval. While taking the image to image retrieval, the performance of the recommended SM-DHOA is 8.75% advanced than PSO and GWO, 8.75% advanced than DHOA, 7.4% advanced than SMO, 4.8% advanced than BD-FFT, 6% advanced than FRS-KELM, and 6% advanced than RELIEF-MM when evaluating the number of multi-media data retrieved as 60. Likewise, when considering the video to video retrieval with the number of multi-media data retrieved as 80, the developed SM-DHOA is 6.75%, 3.9%, 8.2%, 9.7%, 9.4%, 10.9%, and 8.7% better than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. Finally, when taking the number of multi-media data retrieved as 100, the implemented SM-DHOA is 11.7%, 13.4%, 8.5%, 16.9%, 5.6%, 8.69%, and 15.3% enhanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for audio to image retrieval. Further, the designed multimedia big data retrieval model using SM-DHOA with A-SSF establishes superior performance in terms of F1-score when compared to other approaches.

Fig. 8



Performance analysis on F1-score of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for a text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

**Fig. 9**



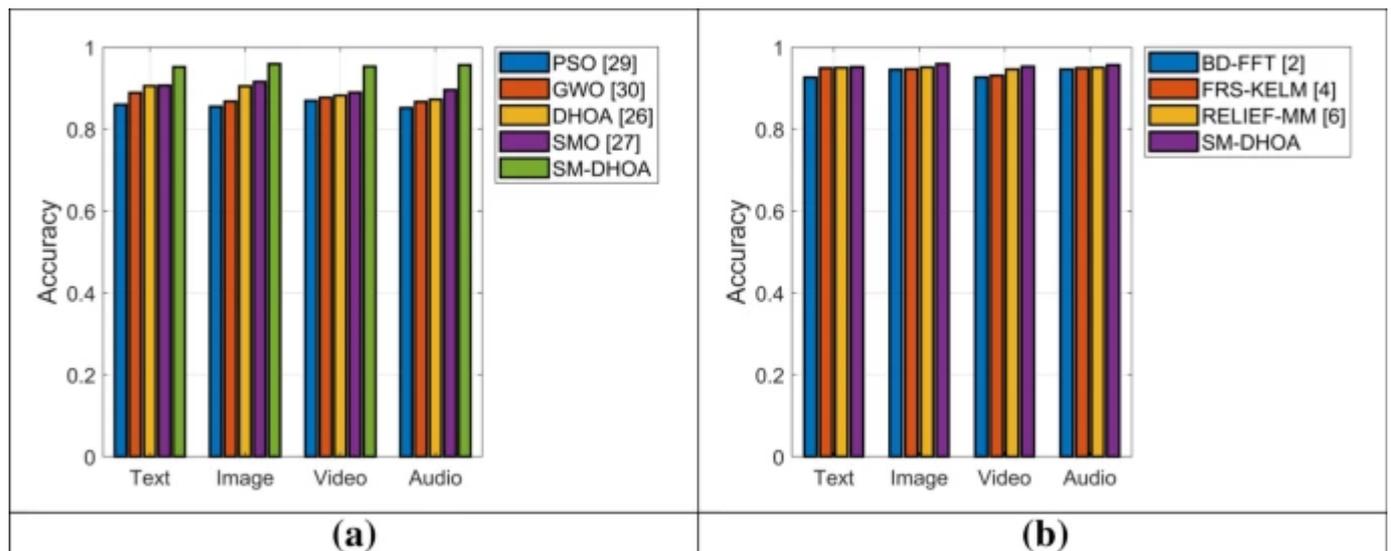
Performance analysis on F1-score of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for a text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

## 6.5 Overall performance analysis on the accuracy

The overall performance of the designed multimedia big data retrieval model is analyzed in terms of accuracy with different optimization-based algorithms and classifiers as

given in Fig. 10. While considering the text to image retrieval, the performance of the designed SM-DHOA is 9.3%, 6.8%, 4.44%, and 4.44% superior to PSO, GWO, DHOA, and SMO, respectively. The performance of the designed SM-DHOA is 4.3%, 3.2%, and 1% enhanced than BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the video to video retrieval. Hence, the performance of the designed model with A-SSF using SM-DHOA gets a more accurate retrieval rate than other conventional models.

**Fig. 10**



Performance analysis on Accuracy for the proposed multimedia big data retrieval model for four datasets with **a** different optimization-based algorithms and **b** classifiers

## 6.6 Statistical analysis on precision

The statistical analysis on precision for the developed multimedia data retrieval model is given in Tables 2 and 3, respectively, by considering different query data. “The mean is the average value of the best and worst values and the median is referred to as the center point of the best and worst values, whereas the standard deviation is represented as the degree of deviation between each execution”. While considering the text to image retrieval, the accuracy of the SM-DHOA is 7.8%, 8.3%, 7.8%, 8.5%, 9.2%, 9.6%, and 7.69% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. Likewise, the statistical analysis on precision for all the datasets shows superior performance while comparing with the other existing approaches.

---

## **Table 2 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of precision**

---

---

## **Table 3 Statistical analysis of the proposed Multimedia Big Data Retrieval model with conventional models for four datasets in terms of precision**

---

### **6.7 Statistical analysis on F1-score**

The statistical analysis on F1-score of the proposed multimedia data retrieval model with A-SSF is given in Tables [4](#) and [5](#). The performance of the SM-DHOA is 7.5%, 8%, 6.7%, 7.6%, 4.4%, 5.5%, and 9.11% advanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while considering the audio to image retrieval. The median of the suggested SM-DHOA is 5.9%, 7.22%, 5.9%, 5.3%, 3.4%, 5.2%, and 7.8% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for audio to image retrieval. Similarly, superior performance is attained by the proposed model than existing approaches.

---

## **Table 4 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of F1-score**

---

---

## **Table 5 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different conventional models for four datasets in terms of F1-score**

---

### **6.8 Statistical analysis on recall**

The performance of the implemented SM-DHOA for the multimedia data retrieval is validated in Tables [6](#) and [7](#). The efficiency of the recommended SM-DHOA is 7%, 6%, 7%,

7.3%, 7.7%, 6.9%, and 8.3% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for an image to image retrieval. The median of the proposed SM-DHOA is 7.3%, 2.3%, 3.5%, 3.5%, 7.8%, 7%, and 7.8% enhanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the image to image retrieval. Therefore, the performance of the designed model is validated in terms of recall while compared with the other approaches.

---

### **Table 6 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of recall**

---

### **Table 7 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different conventional models for four datasets in terms of recall**

---

## **7 Conclusion**

---

This paper has developed a new deep multimedia big data retrieval with A-SSF based on hybrid SM-DHOA. The gathered multimedia data were given to the deep CNN-based semantic feature extraction for getting the significant features. Then, the optimal semantic features were selected using the SM-DHOA to minimize the correlation among the optimal features. Furthermore, the map-reduce framework with A-SSF was developed using SM-DHOA for an efficient retrieval process. Through the performance analysis, while considering dataset 1, the accuracy of the developed SM-DHOA was 4.3%, 3.2%, and 1% enhanced than BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the video to video retrieval. Hence, superior performance was observed by the proposed model while compared to the conventional algorithms. In our proposed SM-DHOA model, results were found only for text-to-image, image-to-image, video-to-video, and audio-to-image. In future works, the proposed model extends towards the utilization of audio-to-video or audio-to-image-to-image processing by utilizing intelligent approaches like the ensemble learning method.

## **References**

---

1. Guo, K., Pan, W., Mingming, Lu., Zhou, X., Ma, J.: An effective and economical architecture for semantic-based heterogeneous multimedia big data retrieval. *J. Syst. Softw.* **102**, 207–216 (2015)

[Article](#) [Google Scholar](#)

2. Ahmad, J., Muhammad, K., Lloret, J., Baik, S.W.: Efficient conversion of deep features to compact binary codes using Fourier decomposition for multimedia big data. *IEEE Trans. Industr. Inf.* **14**(7), 3205–3215 (2018)

[Article](#) [Google Scholar](#)

3. Xia, D., Miao, L., Fan, A.: A cross-modal multimedia retrieval method using depth correlation mining in big data environment. *Multimedia Tools Appl.* **79**, 1339–1354 (2020)

[Article](#) [Google Scholar](#)

4. Sun, Z., Huo, H., Huan, J., Vitter, J.S.: Feature reduction based on semantic similarity for graph classification. *Neurocomputing* **397**, 114–126 (2020)

[Article](#) [Google Scholar](#)

5. Fernandez-Beltran, R., Pla, F.: Prior-based probabilistic latent semantic analysis for multimedia retrieval. *Multimedia Tools Appl.* **77**, 16771–16793 (2018)

[Article](#) [Google Scholar](#)

6. Yilmaz, T., Yazici, A., Kitsuregawa, M.: RELIEF-MM: effective modality weighting for multimedia information retrieval. *Multimedia Syst.* **20**, 389–413 (2014)

[Article](#) [Google Scholar](#)

7. Guo, K., Liang, Z., Tang, Y., Chi, T.: SOR: an optimized semantic ontology retrieval algorithm for heterogeneous multimedia big data. *J. Comput. Sci.* **28**, 455–465 (2018)

[Article](#) [MathSciNet](#) [Google Scholar](#)

8. Liu, Q., Jin, L., Li, Z., Tang, J.: Multimedia retrieval by deep hashing with multilevel similarity learning. *J. Vis. Commun. Image Represent.* **59**, 150–158 (2019)

[Article](#) [Google Scholar](#)

9. Uma, R., Muneeswaran, K.: OMIR: ontology-based multimedia information retrieval system for web usage mining. *Cybernet. Syst.* **48**(4), 393–414 (2017)

[Article](#) [Google Scholar](#)

10. Giangreco, I., Schuldt, H.: ADAM pro: database support for big multimedia retrieval. *Datenbank Spektrum* **16**, 17–26 (2016)

[Article](#) [Google Scholar](#)

11. Zhou, P., Wang, K., Xu, J., Wu, D.: Differentially-private and trustworthy online social multimedia big data retrieval in edge computing. *IEEE Trans. Multimedia* **21**(3), 539–554 (2019)

[Article](#) [Google Scholar](#)

12. Wason, R., Jain, V., Narula, G.S., Balyan, A.: Deep understanding of 3-D multimedia information retrieval on social media: implications and challenges. *Iran J. Comput. Sci.* **2**, 101–111 (2019)

[Article](#) [Google Scholar](#)

13. Lee, C., Min, G., Chen, W.: Data mining and machine learning technologies for multimedia information retrieval and recommendation. *Multimedia Tools Appl.* **75**, 4845–4849 (2016)

[Article](#) [Google Scholar](#)

14. Cheruiyot, W., Tan, G.-Z., Musau, F., Mushi, J.C.: Query quality refinement in singular value decomposition to improve genetic algorithms for multimedia data retrieval. *Multimedia Syst.* **17**, 507–521 (2011)

[Article](#) [Google Scholar](#)

15. Maki, W.S., Krinsky, M., Muñoz, S.: An efficient method for estimating semantic similarity based on feature overlap: reliability and validity of semantic feature ratings. *Behav. Res. Methods* **38**, 153–157 (2006)

[Article](#) [Google Scholar](#)

16. Baâzaoui, A., Abderrahim, M., Barhoumi, W.: Dynamic distance learning for joint assessment of visual and semantic similarities within the framework of medical image retrieval. *Comput. Biol. Med.* **122**, 103833 (2020)

[Article](#) [Google Scholar](#)

17. Sun, Z., et al.: Feature reduction based on semantic similarity for graph classification. *Neurocomputing* **397**, 114–126 (2020)

[Article](#) [Google Scholar](#)

18. Ashraf, R., Ahmed, M., Ahmad, U., et al.: MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features. *Multimedia Tools Appl.* **79**, 8553–8579 (2020)

[Article](#) [Google Scholar](#)

19. Hu, C., Xu, Z., Liu, Y., Mei, L., Chen, L., Luo, X.: Semantic link network-based model for organizing multimedia big data. *IEEE Trans. Emerg. Top. Comput.* **2**(3), 376–387 (2014)

[Article](#) [Google Scholar](#)

20. Jayasena, K.P.N., Li, L., Xie, Q.: Multi-modal multimedia big data analyzing architecture and resource allocation on cloud platform. *Neurocomputing* **253**, 135–143 (2017)

[Article](#) [Google Scholar](#)

21. Zhou, Z., Zhao, L.: Cloud computing model for big data processing and performance optimization of multimedia communication. *Comput. Commun.* **160**, 326–332 (2020)

[Article](#) [Google Scholar](#)

22. Tang, J., Lin, J., Li, Z., Yang, J.: Discriminative deep quantization hashing for face image retrieval. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 6154–6162 (2018)

[Article](#) [Google Scholar](#)

23. Jin, L., Li, K., Hu, H., Qi, G.-J., Tang, J.: Semantic neighbor graph hashing for multimodal retrieval. *IEEE Trans. Image Process.* **27**, 1405–1417 (2018)

[Article](#) [MathSciNet](#) [Google Scholar](#)

24. Jin, L., Li, K., Li, Z., Xiao, F., Qi, G.-J., Tang, J.: Deep semantic-preserving ordinal hashing for cross-modal similarity search. *IEEE Trans. Neural Netw. Learn. Syst.* **30**(5), 1429–1440 (2019)

[Article](#) [MathSciNet](#) [Google Scholar](#)

25. Brammya, G., Praveena, S., Ninu Preetha, N.S., Ramya, R., Rajakumar, B. R. and Binu, D.: Deer Hunting Optimization Algorithm: a new nature-inspired meta-heuristic paradigm. *Comput. J.* (2019)
26. Agrawal, V., Rastogi, R., Tiwari, D.C.: Spider Monkey Optimization: a survey. *Int. J. Syst. Assur. Eng. Manage.* **9**(4), 929–941 (2018)

[Article](#) [Google Scholar](#)

27. Putzu, L., Piras, L., Giacinto, G.: Convolutional neural networks for relevance feedback in content based image retrieval. *Multimedia Tools Appl.* **79**, 26995–27021 (2020)

[Article](#) [Google Scholar](#)

28. Alloui, T., Boussebough, I., Chaoui, A.: A Particle Swarm Optimization Algorithm for web information retrieval: a novel approach. *Int. J. Intell. Inf. Technol.* **11**(3), 1200–1216 (2015)

[Article](#) [Google Scholar](#)

29. Benyl Renita, D., Seldev Christopher, C.: Novel real time content based medical image retrieval scheme with GWO-SVM. *Multimedia Tools Appl.* **79**, 17227–17243 (2020)

[Article](#) [Google Scholar](#)

30. Alghamdi, A., Hammad, M., Ugail, H., et al.: Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimed Tools Appl.* (2020). <https://doi.org/10.1007/s11042-020-08769-x>

[Article](#) [Google Scholar](#)

31. Sedik, A., Iliyasu, A.M., El-Rahiem, B.A., Samea, M.E.A., Abdel-Raheem, A., Hammad, M., Peng, J., El-Samie, F.E.A., El-Latif, A.A.A.: Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. *Viruses* 12(7), 769 (2020)

[Article](#) [Google Scholar](#)

32. Sedik, A., Tawalbeh, L., Hammad, M., El-Latif, A.A.A., El-Banby, G.M., Khalaf, A.A.M., El-Samie, F.E.A.: Deep learning modalities for biometric alteration detection in 5G networks-based secure smart cities. *IEEE Access* 9, 94780–94788 (2021)

[Article](#) [Google Scholar](#)

33. Malipatil, S., Maheshwari, V., and Chandra, M.B.: Area optimization of CMOS full adder design using 3T XOR. In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp. 192–194 (2020)

## Author information

---

### Authors and Affiliations

Information Technology, St. Peter's College of Engineering and Technology, Chennai, India

D. Sujatha

Dept. of Information Technology, SreeVidyanikethan Engineering College, Tirupati, 517 102, India

M. Subramaniam

Dept. of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, 602109, India

Chinnanadar Ramachandran Rene Robin

### Corresponding author

Correspondence to [D. Sujatha](#).

## Additional information

---

Communicated by A. Liu.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Rights and permissions

---

[Reprints and permissions](#)

## About this article

---

### Cite this article

Sujatha, D., Subramaniam, M. & Rene Robin, C.R. A new design of multimedia big data retrieval enabled by deep feature learning and Adaptive Semantic Similarity Function. *Multimedia Systems* 28, 1039–1058 (2022). <https://doi.org/10.1007/s00530-022-00897-8>

Received

16 August 2021

Accepted

15 January 2022

Published

05 February 2022

Issue Date

June 2022

DOI

<https://doi.org/10.1007/s00530-022-00897-8>

### Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

## Keywords

[Multimedia big data retrieval](#)

[Adaptive Semantic Similarity Function](#)

[CNN-based semantic feature extraction](#)

[Optimal feature selection](#)

[Spider Monkey-Deer Hunting Optimization Algorithm](#)

[Map-reduce framework](#)

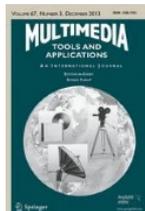
[Home](#) [Multimedia Tools and Applications](#) [Article](#)

# A proficient video recommendation framework using hybrid fuzzy C means clustering and Kullback–Leibler divergence algorithms

Published: 07 February 2023

Volume 82, pages 20989–21004, (2023) [Cite this article](#)[Download PDF](#) ↓

Access provided by CBIT–Library &amp; Information Centre Hyderabad

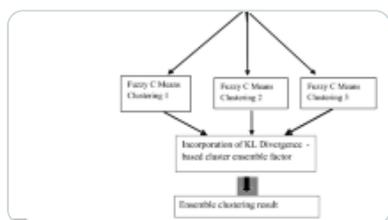
**Multimedia Tools and Applications**[Aims and scope](#)[Submit manuscript](#)[H. Anwar Basha](#) , [S. K. B Sangeetha](#), [S. Sasikumar](#), [J. Arunnehr](#) & [M. Subramaniam](#) 173 Accesses  2 Citations  1 Altmetric [Explore all metrics](#) →

## Abstract

A video recommendation framework for e-commerce clients is proposed using the collaborative filtering (CF) process. One of the most important features of the CF algorithm is its scalability. To avoid the issue, a hybrid model-based collaborative filtering approach is proposed. KL Divergence was developed to address the CF technique's scalability problem. The clustering with enhanced sqrt-cosine similarity Recommender scheme is proposed. For successful clustering, Kullback–Leibler Divergence-based Fuzzy C-Means clustering is suggested, with the aim of focusing on greater accuracy during

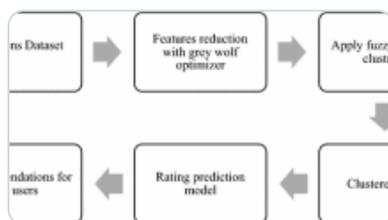
movie recommendation. The proposed scheme is viewed as a trustworthy contribution that significantly improves the ability of movie recommendation by virtue of the KL divergence-based Fuzzy C-Means clustering mechanism and enhanced sqrt-cosine similarity. The proposed scheme highlighted and addressed the critical role of the KL divergence-based cluster ensemble factor in improving clustering stability and robustness. For prediction, the enhanced sqrt-cosine similarity was used to calculate successful related neighbor users. The performance of Recommendation is improved when KLD-FCM is combined with improved sqrt-cosine similarity. The proposed scheme's empirical work on the Movielens dataset in terms of MAE, RMSE, SD, and Recall were found to be superior in recommendation accuracy compared to traditional approaches and some non-clustering based methods recommended for study. With the specified number of clusters, it is capable of providing accurate and customized movie recommendation systems.

## Similar content being viewed by others



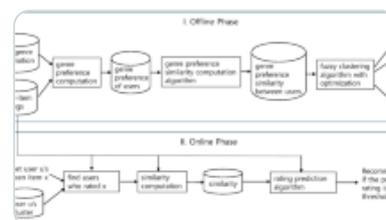
**A Kullback–Leibler divergence-based fuzzy C-means clustering for enhancing the potenti...**

Article | 10 June 2019



**Recommender system with grey wolf optimizer and FCM**

Article | 27 December 2016



**Fuzzy clustering with optimization for collaborative filtering-based recommender...**

Article | 03 November 2021

[Use our pre-submission checklist →](#)

Avoid common mistakes on your manuscript.



## 1 Introduction

A Recommendation System (RS) is a filtering program that enables consumers to evaluate product recommendations for online purchases and provide information into products that they are interested in. In recent years, the extensive advancement of science and

technology has resulted in vast amounts of digital knowledge being accessible via the internet. As a result of the overabundance of data, users are unable to obtain accurate taste information. This problem of information overload can be solved with RS, which filters out irrelevant information and only recommends related things to users. RS's knowledge filtering system assists users in making decisions in difficult situations by scanning vast data for items of interest. It also offers a personalized proposal based on the user's desires and preferences. RS is widely used in e-learning, e-shopping, e-tourism, e-business, e-government, and social networking sites like Facebook. Articles from research, books, music, news, films, DVDs/CDs, and other e-shopping products recommended by RS for their clients [1, 2].

To create a working recommendation system, a large amount of data must be collected. RS accepts a variety of inputs, both explicit and implicit. The explicit ratings of 1 to 5 given by users for their preferences in the items they purchase make up the covert reviews. User actions such as accessing and navigating past websites, click and search logs provide implicit input. Demographic data is another addition to RS. This information index was developed for each client who visits the site. Following this point, the data is filtered to obtain sufficient data for customer/user suggestions. Content-based filtering, collaborative filtering, and hybrid filtering technology are examples of filtering algorithms that would be more suitable for the recommending engine [3, 4]. The collection of data and application of recommendation filtering methods yield a set of recommendations that comply with the procedures to be considered in a recommendation system's calculation. In general, two types of performance are predicted and suggested. The ranking items for which the target consumer has not been rated are forecasted by prediction. On the basis of the forecasted ratings, the recommendation recommends the top-n recommendations to the target consumer, where each item does not include an evaluation value for these top-n recommendations. The consumer should be ecstatic with the performance of the recommender [5].

A framework that offers good and helpful recommendations for its own users requires the use of appropriate and reliable recommendation techniques. The content-based technology employs a domain-based algorithm that focuses on analyzing the characteristics of predictive posts. When documents such as blogs, magazines, and news are recommended, it is the most effective material filtration technique. The user profile recommendation in the Content-Based Filtering (CBF) approach is based on characteristics extracted from the content of items checked by the user. Objects that are

mostly associated with positive items are recommended to the customer. CBF employs a variety of models to identify similarities between documents in order to generate useful recommendations. To form the relationship between different documents within the Corpus, it could use a Vector Space Model or a probabilistic model like the Naive Bayes Classification method, Decision Trees, or Neural Networks. It's also possible to use the Vector Classification method. These approaches make recommendations based on the underlying model's statistical analysis or machine learning techniques [6].

Other users' profiles aren't needed for content-based filtering because they don't affect the recommendation. Furthermore, as the user profile changes, CBF's approach adjusts its recommendations in a very short time. The key drawback of this method is that it necessitates thoroughly informing and explaining the characteristics of the objects in the profile. CF is a domain-agnostic content prediction technique that can't be easily or reliably classified by metadata like film or music. Filtering by working together creates a database of user preferences (user-item matrix). It then matches individuals with relevant interests and preferences to make recommendations. This group of people is creating a social network. A consumer receives recommendations for items that he hasn't yet rated but that other users in his field have given high marks to. CF may make recommendations in the form of forecasts or recommendations. Collaborative filtering has a range of benefits over CBF, including the ability to be used in fields where object content is scarce and computer system content is difficult to assess (such as opinions and ideals). CF technology should provide persuasion recommendations so that things that are useful to the user can be recommended, even though the user profile does not include the content [9].

Hybrid filtration strategies combine various recommendation approaches to boost device optimization and prevent some of the drawbacks and issues that come with pure recommending systems. Since one algorithm can overcome its drawbacks with another algorithm, a combination of algorithms is built to make recommendations more accurately and efficiently than a single algorithm. Using various recommendation models, a combined model will eradicate the flaws of a single process. Separate algorithms can be applied and the results merged, content-based co-filters can be used, content-based collaborative filters can be used, and a single recommendation system can be developed to bring all methods together [10].

The most promising products from which consumers can choose are included in the recommendation question. Some well-known approaches for solving the problem of scalability under model-based collaboration filters are clustering-based approaches. Predominantly, the majority of CF-based clustering strategies have relied on K-means and Fuzzy C-means clustering, which lack the ability to pick a relevant clustering core, lowering the predictive efficiency. As a result, there are trade-offs between scalability and predictive efficiency. Improved clustering techniques were suggested in the study to recognize agreed problems due to scalability. Many previous studies have shown that clustering-based CF systems (CF combined with clustering algorithms) are a promising schema for providing accurate personal recommendations and solving large-scale problems [8]. Fuzzy C-Means is a soft clustering approach that allows each individual data to be allocated to multiple clusters based on different membership degrees. They also concluded that good clustering-based CF performance is dependent on appropriate clustering techniques as well as the dataset's design. The analysis shows that the Clustering algorithm's stability and robustness need to be improved in order to achieve critical accuracy in the process of movie recommendation to target users. The following are the limitations of the current method:

- Fuzzy-C refers to a lack of ability to choose the initial cluster Center point, which can lead to a local optimum solution and affect clustering accuracy. In certain cases, the obtained clusters can be impractical, affecting the Recommendation outcomes.
- To get a good grouping of data, most current clustering algorithms require the configuration of some parameters.
- The disadvantage of FCM is that it has a higher error rate and needs further iterations to obtain well-framed clusters.
- Because of the prejudices and assumptions that each clustering algorithm contains, applying a single clustering method generally results in inconsistent results.

The Fuzzy C Means clustering with KL divergence is suggested to solve the aforementioned limitations. The research work's contributions have the following features.

- To prevent the drawbacks of a bad initialization, Ensemble FCM clustering is used to divide users into separate groups.

- The Ensemble Fuzzy C-Means clustering methods use a KL divergence-based cluster ensemble factor to improve the stability and accuracy of the clustering process, resulting in successful clustering with the goal of focusing on better performance results during movie recommendation.
- A better approach is to treat the membership vector as a discrete probability function, with the statistical distance, such as KL divergence, serving as the similarity metric.
- For active users, the enhanced sqrt-cosine similarity is often used to find the most powerful nearest neighbors.
- Reduce the problem of scalability.

The latest analysis of RS methods is summarized in Section [2](#). In addition, work on various RS is discussed in this chapter. The proposed framework model for the hybrid video recommender is defined in Section [3](#). The quality of predictions as measured by the assessment metrics is also stated. Section [4](#) brings the analysis to a close by highlighting the algorithms that aided in the achievement of the objectives and promoted the desired outcome. The study's limitations have been established, and potential research directions have been summarized.

## 2 Related study

---

Recommend systems use data mining and predictive algorithms to predict user preferences among the vast array of images, goods, and services available. The rapid growth of knowledge on the Internet, as well as the number of visits to websites, are posing significant challenges to system recommendations. The development of precise recommendations, the successful management of a large number of recommendations, and the large number of system members are all examples of these challenges. As a result, new system recommendation technologies are needed that can produce high-quality recommendations quickly, even for large data sets. There are numerous methods and algorithms for data filtering and recommendation. This section provides a brief overview of recent system-related studies in the literature.

Collaborative filtering is a widely used and relevant technology that makes predictions and suggestions based on the ratings and actions of other system users. The key premise of this strategy is that the user can pick and consolidate views from other users in order to

understand his choice for the active user. Memory-based CF algorithms generate a forecast for the entire database or a subset of the user-item database. Each consumer is a member of a group of people who share similar interests. A prediction of a new user's tastes for new objects can be rendered by identifying related neighbors (or active users). Memory-based collective filtering has a number of drawbacks, the most significant of which is that it must use the entire database every time it predicts something, making it extremely sluggish in memory. If the rating matrix is so broad that many people use it, the issue becomes serious. Computing resources are depleted, and device performance suffers as a result, making it impossible for the system to respond quickly to user requests [11].

The model-based approach learns a model to boost collaborative filtering technology efficiency using prior scores. The model could be built using machine learning or data mining techniques. Since these approaches rely on pre-computed models, they can quickly suggest a large number of items and have been shown to yield results that are comparable to neighborhood-based recommendation techniques. Dimension reduction, for example, includes techniques including singular decomposition, matrix completion, latent semantic approaches, regression, and clustering. Content-based filters recommend items on a user's item profile and user profile. When an account is created and the framework is first used, these types of profiles are created. As a result of the user's interaction with the system, a better user profile is developed. If a user likes an object in the past, the CBF scheme assumes that the user would like similar things in the future [7]. The most powerful filtering technique is used in information documents such as web sites, journals, and news. To produce meaningful recommendations, CBF employs a variety of models to detect correlations between documents. Model-based vector space models can be used to represent the relationships between various documents within a corpus, such as reverse frequency or probabilistic models like the Naive Bayes Classification, Decision Trees, or Networks. Object metadata is used in the filtration mechanisms. Before users can get a recommendation, they'll need a large collection of items and a well-organized user profile. As a result, the effectiveness of CBF is dependent on the availability of descriptive data. Over-specialization is another major problem with the CBF methodology. Users can only get suggestions that are close to their own [12].

In certain applications, hybrids of various types outperformed individual algorithms. When the algorithms in question cover a wide range of use cases or aspects of the data set, hybrids can be particularly useful. The suggestion has been suggested to be implemented

using a range of approaches, including material-, collaborative-, knowledge-based, and other techniques. Every form of recommendation has its own set of strengths and weaknesses. In order to improve efficiency, these strategies were often combined into hybrid recommenders. The hybrid recommender method has a higher level of complexity and implementation costs [14].

The Knowledge Base (KB) suggestion suggests things based on user experience, artifacts, and/or user relationships. In most cases, KB recommendations maintain a knowledge base describing how a particular item serves the needs of a specific person, which can be carried out based on inferences about the relationship between a user's need and a potential recommendation. The semantic similarity between objects can be calculated using the domain ontology. Social recommendation services are an integral part of everyday life on social media. Every minute, users on social media exchange details. Social advisor programs are designed to help people understand what they really want by reducing the amount of information available on social media. They want to help people on social networking sites like Facebook, Twitter, YouTube, Flickr, and Weibo by providing them with tweets and profiles that meet their needs [13].

Systems that are recommended are those that can analyze past user habits and make suggestions for current issues. Simply put, information on similar behavior, remarks, and users will be used in the RSs to try to define the user's thought style in order to assess and suggest the user's taste as the most appropriate and near object. Many of the methods in the RSs are used to make recommendations that are as accurate as the users need. As a result, several models for RSs clustering existing data have been presented in order to efficiently process data with large volumes. Given the goals, special characteristics, and relationships between data in the RSs, using an effective data cluster approach to efficiently process data in subsequent steps and produce more reliable suggestions has always been seen as an important research area for developing these systems [15].

Collaborative filtering recommendation systems are extremely useful for a wide range of online activities, including e-commerce. However, there are significant issues, especially in scalable and dynamic scenario implementations where new users, objects, and ratings are added frequently. Scalability refers to a system's, network's, or process's ability to control or expand its capacity to accommodate development. For example, a device is called scalable if it can increase its total power under increased load as resources (typically hardware) are added. Dimension reduction-based approaches address

scalability problems such as SVD, MF, clustering, and so on. In short, cluster-based approaches retain the advantages of low computation cost (for searching candidates) over memory-based approaches as models for dimension reduction [[16](#), [17](#)].

To improve the performance of the recommendation, related users are grouped together based on their interests. Clustering was used to quickly locate a user's neighbors. By reducing the size of the original data to more manageable partitions, clustering systems can react quickly. Clustering, in particular, boosts the scalability and accuracy of recommendation systems. Despite the advantages of low machine costs (for searching candidates) over memory-based and SVD methods, MF methods remain cluster-based methods. One of the most serious problems with a recommending method is scarring, and data sparsing has an effect on the accuracy of the recommendation. In general, machine data like Movie Lentsils is interpreted as a user-item matrix made up of films, which increases matrix dimensions and sparsity since the user and items are no longer used. Most users do not rate most items, and there are few available ratings. The key explanation for this is a lack of knowledge. In order to condense the user's object matrix, the reduction of dimension addresses the issue of scarcity by excluding non-representative or insignificant users or objects. However, during the reduction process, potentially valuable information is lost. However, some potentially valuable information can be lost during the reduction process [[18](#), [19](#)].

Collaborative filters create this problem because they depend on the rating matrix in most cases. Many researchers have attempted to address this problem, but more research is still required in this area. Most recommendation systems on the major electronic commerce platforms have been influenced by the long tail effect in some way. Since accuracy-focused recommender systems tend to recommend common goods, recommending items with few ratings is critical (long tail items). Popular products that are likely to be less helpful to users can be easily recommended using detailed recommendation algorithms. To assess the ability of systems to recommend unpopular goods, the assessment metrics diversity and innovation have been added. Recommending long tail artifacts can result in the recommendation's precise results being lost. As a result, a recommendation process must be developed that recommends controversial products while minimizing accuracy loss. Several guidelines have recently been proposed to strike a balance between precision, diversity, and novelty [[20](#),[21](#),[22](#)].

Existing CF clustering method algorithms are ineffective at improving RS efficiency and addressing scalability issues. The recommended performance has an effect on the efficiency of the clustering procedures [23, 24]. As a result, there is a lack of precision and coverage, which makes clustering-based approaches in recommender systems difficult to use in practice. To improve the recommendation's performance, better methods for optimizing the above problem are needed.

## 3 Proposed system model

### 3.1 KL divergence based ensemble fuzzy C means clustering

Cluster-based CF has been shown to address scalability issues while also improving the consistency of recommendation outcomes in recent years. The aim of clustering algorithms is to group objects into clusters with the shortest distance between them in order to find objects that are identical. Clustering strategies will typically group a large number of users into various clusters based on their rating similarity in order to locate "like-minded" neighbors. One of the most widely used clustering methods is fuzzy clustering. To get a decent grouping of data, most current FCM algorithms require the specification of some parameters. As a result, the Fuzzy cluster ensemble solution usually prevents the drawbacks of a bad initialization.

#### 3.1.1 Fuzzy C means clustering algorithm

1. Initialize Membership matrix  $M$  with random data points.
2. Fuzzy cluster center is calculated  $C$
3. Calculate the objective function  $F = M * (X_i - C_j)$
4. For every iteration fuzzy Membership is updated by using  $M = \sum C_k$  Where  $k$  is the number of clusters
5. The iteration will stop when  $(k+1) - (k) < \text{termination criterion}$

Ensemble clustering blends a dataset's various simple partitions into a more stable and robust one. The basic concept behind the Ensemble Fuzzy C Means cluster method is to apply the clustering method to the data several times (rather than only once) and then merge the results into a single partition. Ensemble clustering takes a collection of data partitions as input. The cluster ensembles are divided into two parts. The ensemble clustering generator is one, and the consensus function is another. The first section

focuses on generating more diverse clustering results, while the second section focuses on seeking a good consensus feature to increase the results' accuracy. Homogeneous ensemble FCM clustering is used in the first component of cluster ensembles. The term "homogeneous ensemble" refers to the use of multiple runs of a single clustering algorithm (fuzzy c-means algorithm) with different initializations and fuzzy parameter values. Several soft partitions of the data are obtained at the end of the first stage of the ensemble method as a result of several runs of the algorithm(s). The aim of this is to improve the accuracy and consistency of fuzzy cluster analysis procedures. Soft ensembles are characterized by the concatenation of membership probability distributions in the second part of cluster ensembles. The obtained partition will be combined in the second stage using a KL divergence-based objective function to produce a single final partition. The Kullback–Leibler (KL) divergence was then used to describe a distance measure between two instances. The similarity of a membership vector to a cluster center is measured by FCM using squared Euclidean distance. This is ineffective in situations where a data's membership in all clusters normally equals one. A better approach is to treat the membership vector as a discrete probability function, with the statistical distance, such as KL divergence, serving as the similarity metric. This algorithm is identical to fuzzy c-means, with the exception that it employs the KL divergence to treat memberships as discrete probabilities.

The data from the User Item is first categorized using homogeneous fuzzy clustering methods. After that, a fuzzy KL divergence-based objective function aggregates the soft clustering effects.

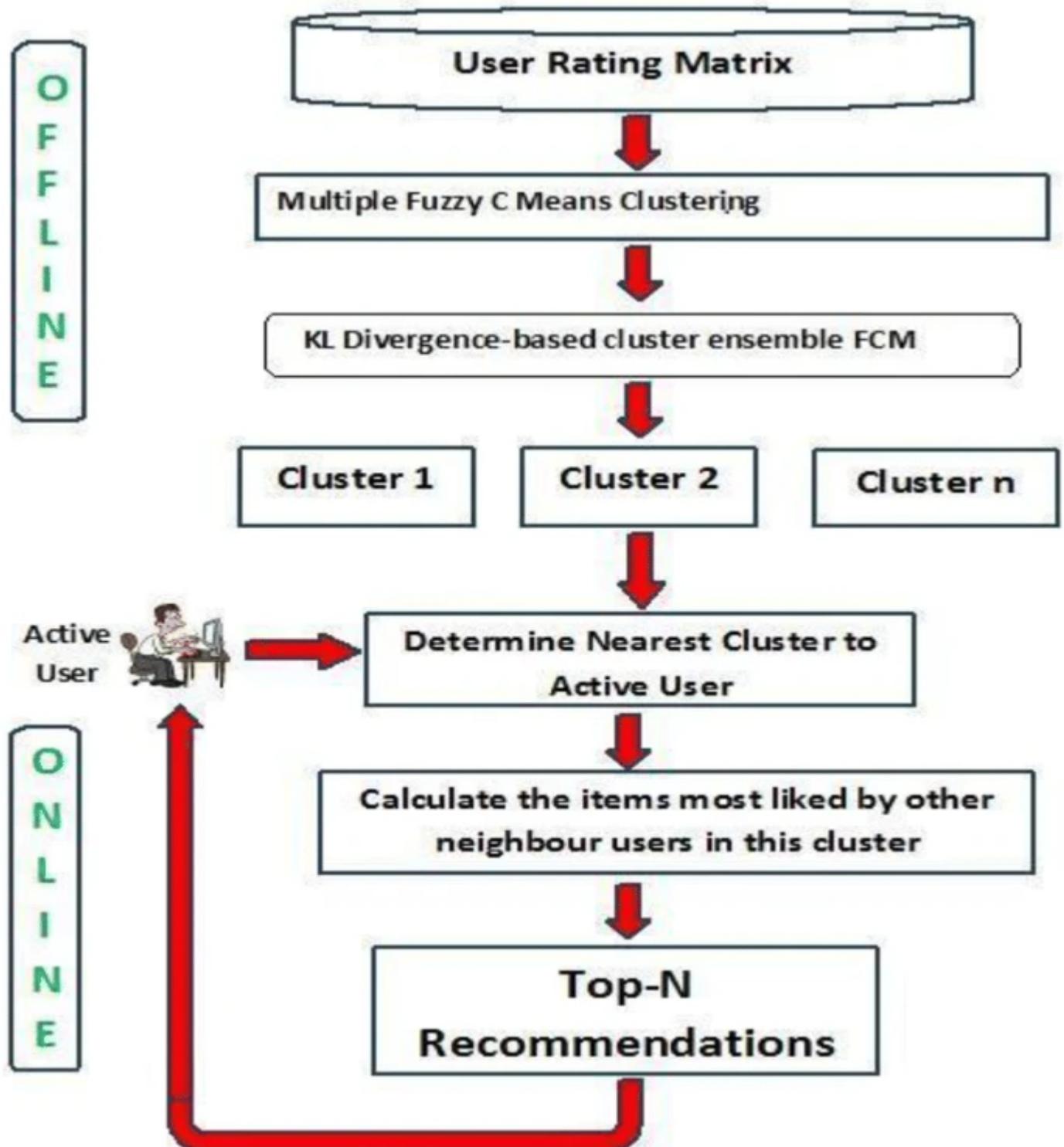
### 3.2 Improved Sqrt-cosine similarity

Improved Sqrt-Cosine Similarity (ISC) is a modern similarity measure that uses Hellinger distance and is based on sqrt cosine similarity. Hellinger distance (L1 norm) is a much better metric for high-dimensional data mining applications than Euclidean distance (L2 norm). In terms of implementation, the ISC is very similar to cosine similarity, and it outperforms other similarity measures in high-dimensional results. The enhanced sqrt-cosine similarity determines how close two users are. For High Dimensional data, the Hellinger distance-based Similarity is more accurate. The KLD-FCM with enhanced sqrt-cosine similarity outperforms current systems in terms of recommendation performance.

### 3.3 Proposed KLD-FCM based movie recommendation scheme

For improving movie recommendation methods (KLD FCM-RS), a kullback–leibler divergence-based fuzzy c-means clustering is proposed, and the steps involved in KLD FCM-RS are discussed in detail. The aim of the proposed method is to develop a Collaborative Movie Recommender framework that can solve scalability problems while also improving prediction accuracy in terms of MAE, Precision, Recall, and Speed. The proposed KLD-FCM based Movie Recommendation Scheme is architecturally depicted in Fig. 1 as follows.

Fig. 1



## Architectural view of the proposed KLD-FCM Scheme

---

It is divided into two phases: offline and online. The User Item Rating Matrix derived from the used Movie Lens Dataset is used as a possible input during the offline process. Then, over the extracted user Item scores, a method of different homogeneous Fuzzy C means clustering is applied to divide the users into different classes. Furthermore, KL Divergence-based cluster ensemble FCM is used to combine the various FCM clustering findings in order to generate efficient single User clusters. The nearest cluster estimation for Active consumer is computed using the Euclidean distance method in the online process. The active user's nearest neighbors in his or her nearest cluster are then found using an enhanced sqrt-cosine similarity tool. Finally, the top list of recommended movie items is calculated in an online mode by determining the movie items that are most frequently recommended by the context's neighborhood users. Leibler-Kullback For the Recommender method, divergence-based Fuzzy C-Means clustering with enhanced sqrt-cosine similarity worked well. Three possible steps are included in this proposed KLD-FCM:

### 3.3.1 Procedure

#### Step 1 (FCM-KLD)

FCM-KLD is divided into two phases. Multiple Fuzzy C Means Clustering is the first step. The user data ratings from the Movie Lens data set are used as input in this phase of the clustering process. Over the User Item scores, apply three different homogeneous Fuzzy C Means clustering methods with different initializations. By executing the FCM several times for each initialization with different fuzzy parameter values, homogeneous FCM clustering is used to create several partitions with different random initializations (here 1.5, 2 and 2.5 are used). **Input:** User Rating Matrix **Output:** Three Clustering results.

Table 1 shows a snapshot of the User object rating matrix from the Movie Lens dataset for 5 users on 5 movies, where U1-U5 are users and M1-M5 are objects (movies). The value 1-5 represents the user's likelihood rating for a specific film. The value '0' denotes that the consumer has not rated (or seen) the film. The recommender framework identifies unrated values and suggests the top N films to the consumer Tables 2, 3, 4, and 5.

## Table 1 Example of rating matrix from Movie Lens dataset

---



---

## Table 2 MAE comparison analysis

---



---

## Table 3 RMSE comparison analysis

---



---

## Table 4 Recall analysis

---



---

## Table 5 Accuracy analysis

---

## Step 2 (determine the nearest cluster to active user)

After clustering the users into various clusters, the Euclidean distance approach is used to determine the nearest cluster to Active User.

$$d_i = \sqrt{\sum_{j=1}^d (\text{Cent}_{i,j} - U_j)^2} \quad (1)$$

(1)

$\text{Cent}_i$ : is the centroid of 'i' th cluster,  $U$  is the Active User Profile.

$d$ : is the dimension of data (Number of Attribute).

$Cent_{i,j}$ : is the  $j$ th attribute of centroid profile in cluster  $i$ .

### Step 3 (using improved sqrt-cosine similarity, determine the top N recommended movies)

Improved sqrt-cosine similarity is used to calculate Active User's nearest neighbors. The following formula is used to determine how close users  $u_1$  and  $u_2$  are.

$$\text{sim}(u_1, u_2) = \frac{\sum_{i=1}^m \sqrt{R_{u_1, i} R_{u_2, i}}}{\sqrt{\left(\sum_{i=1}^m R_{u_1, i}\right)} \sqrt{\left(\sum_{i=1}^m R_{u_2, i}\right)}}$$

(2)

$m$ : Set of common items rated by user  $u_1$  and user  $u_2$ .

$R_{u_1, i}$ : is the rating given to item 'i' by user  $u_1$ .

$R_{u_2, i}$ : is the rating given to item 'i' by user  $u_2$ .

The Hellinger distance is used to compute the similarity between two vectors in the enhanced sqrt-cosine similarity. This phase is essential for comparing each individual user's rating to the ratings of other users in the clusters. Finally, the top list of recommended movie items that could be suggested to an active user at any time is calculated based on the movie items that are most often recommended by the context's neighborhood users. The movies are recommended to target users who are most likely used by other neighbor users and are not used by him/her during the Recommendation process. The weighted average of the ratings of items in the same cluster neighbor's is used to predict the ranking of unrated items for active users, and then the top-N suggestions list is sent to the active user. The rating of unrated movie (item) 'i' for an active user 'a' is predicted by  $P_a(i)$

$$P_a(i) = \frac{\sum_{N \in \text{C}_x} \text{sim}(a, N)}{\sum_{N \in \text{C}_x} \left( \left| \text{sim}(a, N) \right| \right)}$$

(3)

Where,

$a$ : Active User.

$\langle \{\underset{\_}{R}\}_a \rangle$ : Average of active user  $a$ .

$C_x$ : set of nearest neighbors of active user  $a$  belonging to one common cluster.

$\langle \{\underset{\_}{R}\}_N \rangle$ : Average rating score given by active user's neighbor  $N$ .

$Sim(a, N)$ : similarity between active user  $a$  and Neighbor.

## 4 Experimental design

---

Experiments using the publicly accessible Movielens dataset are used to equate the performance of the proposed KLD-FCM with that of baseline recommendation system. The Movielens data set used to equate the proposed KLD-FCM scheme to the compared COA, FCM-BAT, FCM, and ICF contains 10,00,000 ratings, with 850 users theoretically rating them. This Movielens data set contains reviews ranging from approximately 1000 to 1513 movies, each scored on a scale of 1 to 5. By partitioning the entire Movielens data set using the k-cross validation process, the performance of the proposed KLD-FCM approach is investigated. The findings were evaluated using 5-fold cross validation. The original dataset is divided into five equal subsets. One is used as a test set (20%), while the other is used as a training set (80%). The procedure is repeated five times, with a different test set selected each time, and the average results recorded. In terms of MAE and RMSE, the proposed method was compared to non-clustering methods such as Basic CF (BCF), User-Based CF (UBCF), SVDM (a variant of Single Value Decomposition (SVD) that uses batch learning with a learning momentum), and RSVD (Regularized SVD model). Various collaborative approaches are selected from the literature to verify the proposed method's role in comparison to other cluster-based techniques.

### ICF(integrated Collaborative Framework)

The merits of the item k-NN algorithm were used to propose an Integrated Collaborative Framework (ICF). This ICF also provided classification restrictions, ensuring that only

potential rules are used during the collaborative filter-based categorization of user ratings.

## UPCC (user based CF Pearson correlation coefficient based CF)

For the Collaborative filtering recommender scheme, the Pearson Correlation Coefficient test is used to determine how closely two users are related.

## FCM (fuzzy C means)

The performance of User-Based Collaborative Filtering with Fuzzy C Means is compared to that of other clustering methods such as K-means and Self-Organizing Maps (SOM).

## FCMBAT

The Fuzzy C-Means and BAT-based Movie Recommendation Scheme (FCM-BAT) is an integrated Fuzzy C-Means and BAT-based Movie Recommendation Scheme for promoting efficient and collaborative recommendation to the target users. This FCM-BAT was proposed to address scalability issues and improve the clustering process, with the aim of improving the consistency of the recommendation process.

## COA(cuckoo Optimization Algorithms)

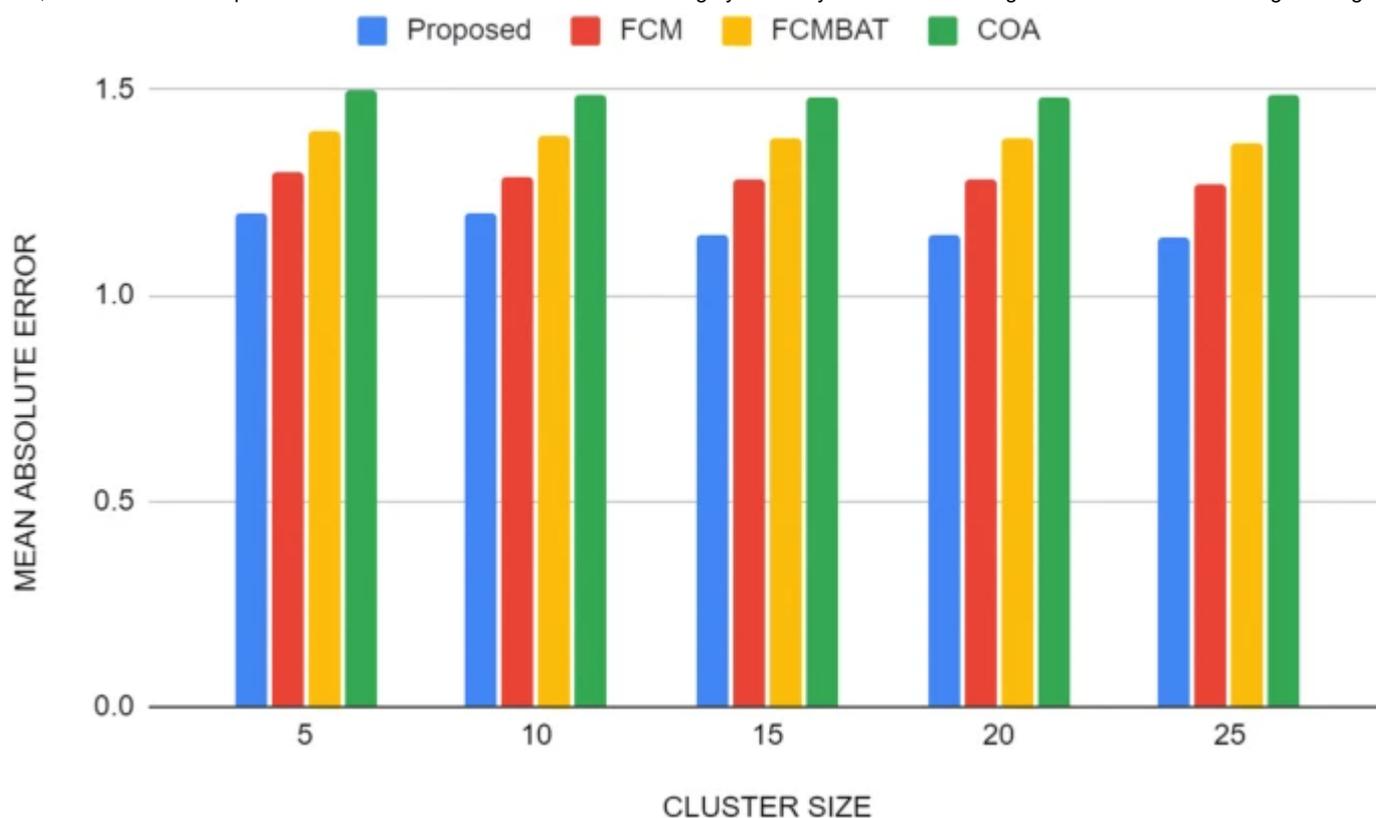
Furthermore, a possible movie recommendation system based on k-means and COA is proposed in order to improve the rate of recommendation accuracy when using the Movielens dataset.

## 4.1 Mean absolute error

The suggested method used to measure MAE can be seen in Fig. 2 for different numbers of neighbors.

---

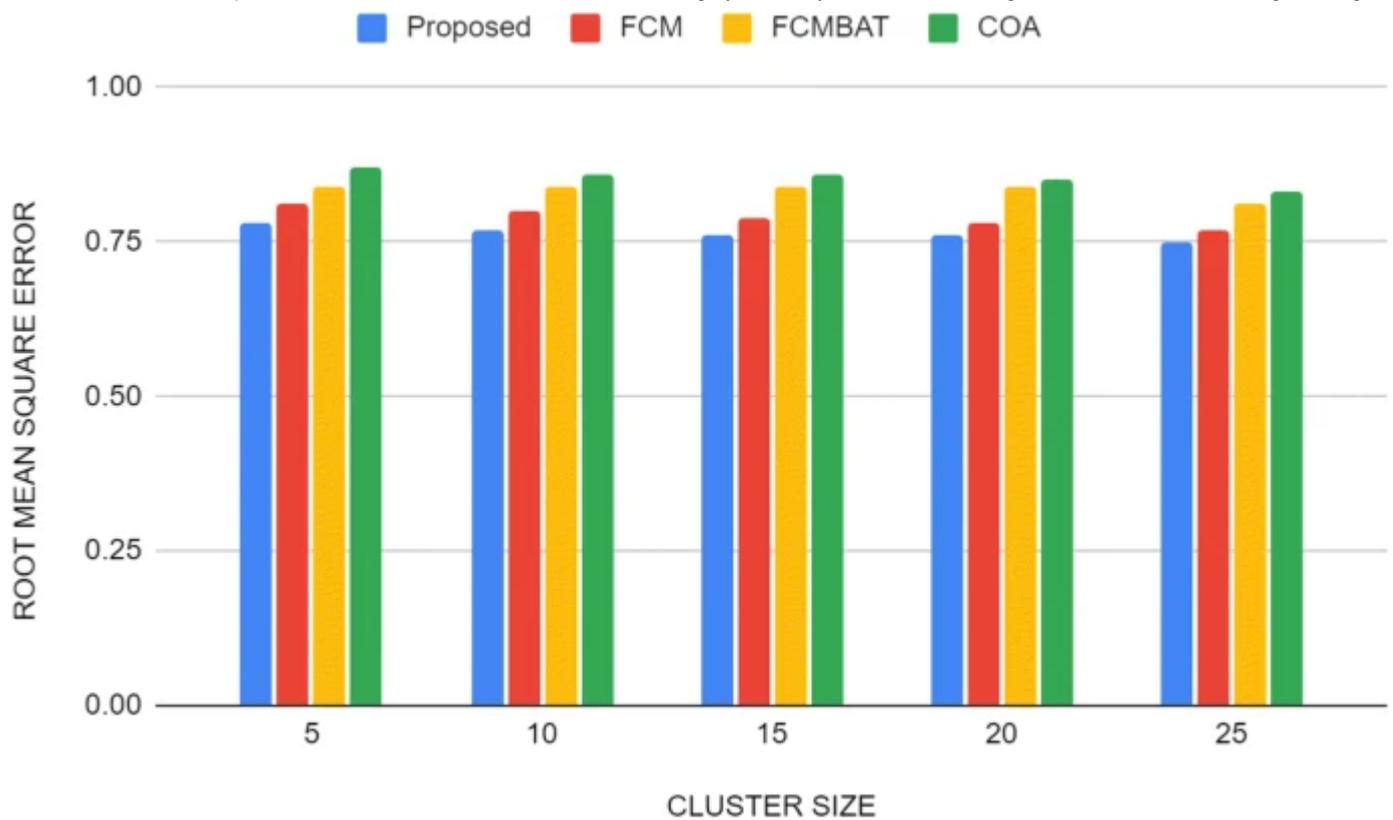
**Fig. 2**



### MAE comparison analysis

Figures 2 and 3 show the contrast of the proposed FCM KLD with current Collaborative Recommender framework methods in terms of MAE and RMSE for different numbers of clusters. The proposed scheme's MAE and RMSE are proving to be significantly lower than those of current schemes. As a result, the proposed scheme's MAE is lower than the COA, FCM-BAT, FCM, and ICF approaches. The proposed scheme's RMSE is also tested to be substantially lower than the baseline schemes under consideration.

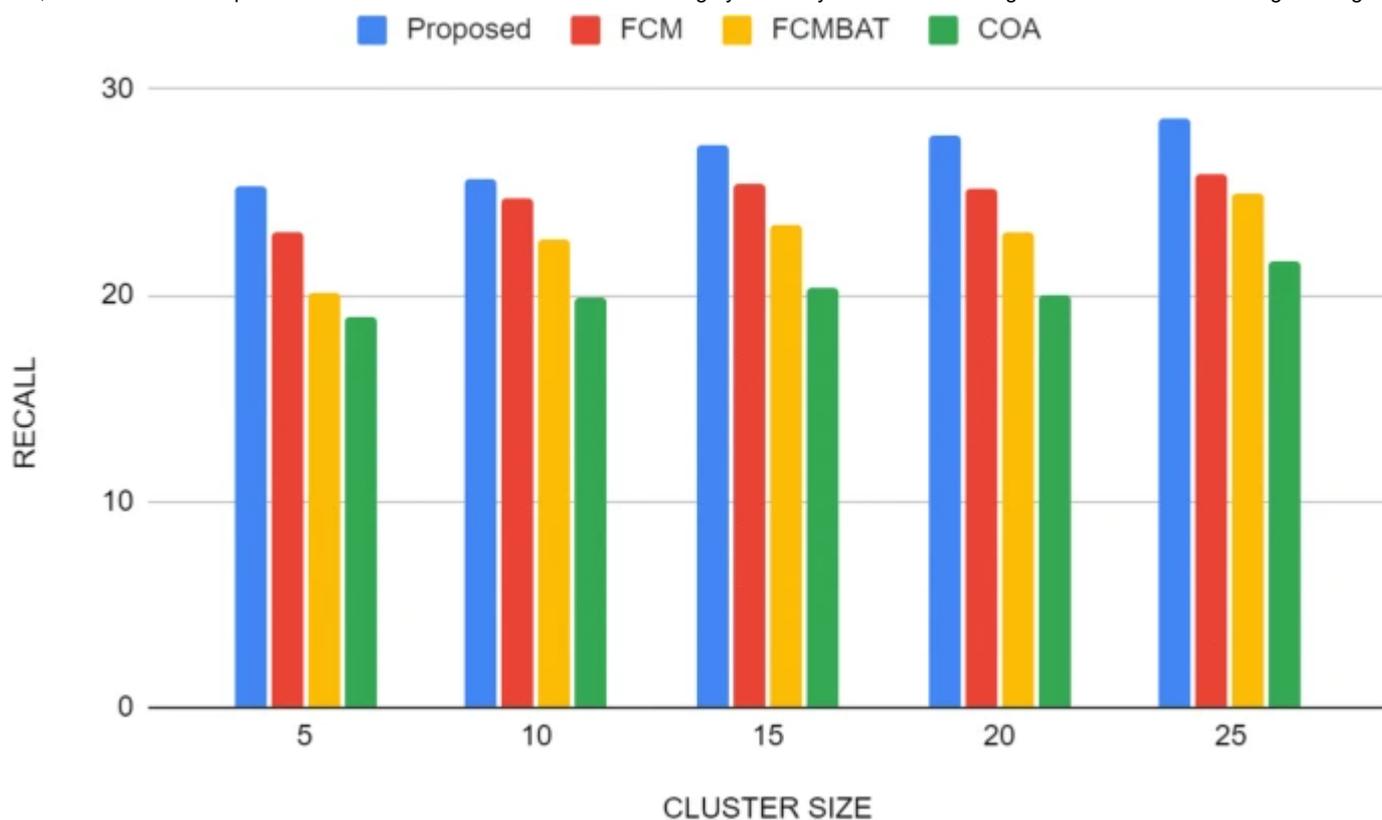
**Fig. 3**



#### RMSE comparison analysis

Figure 4 graphically depicts the contrast of the proposed KLD-FCM with current Collaborative Recommender framework approaches in terms of recall. It highlights the efficiency of the proposed KLD-FCM scheme as measured by Recall for a variety of cluster sizes. The proposed KLD-FCM scheme's Recall value is determined to be excellent as compared to baseline methods, since the KL divergence factor's guidance in the clustering phase is responsible for the majority of progress.

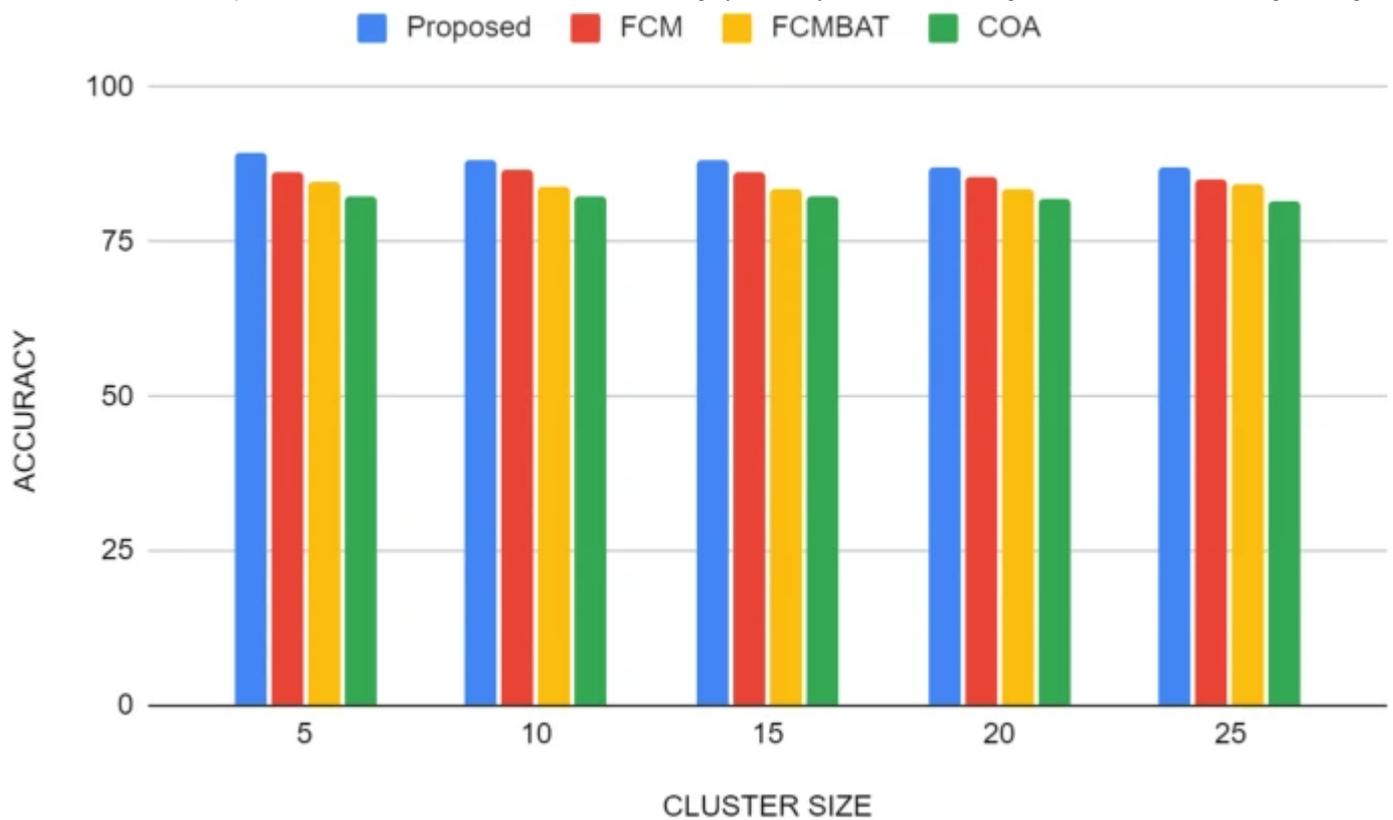
**Fig. 4**



#### Recall comparison analysis

Figure 5 depicts a graphic comparison of the proposed KLD-FCM with current Collaborative Recommender framework approaches in terms of accuracy. The performance of the proposed method is significantly better than that of current methods. Since it takes advantage of the advantages of KL Divergence Fuzzy C, Clustering with enhanced sqrt-cosine similarity is possible.

**Fig. 5**



Accuracy analysis

## 5 Conclusion

The KL Divergence Fuzzy C means Clustering with improved sqrt-cosine similarity Recommender framework (KLD-FCM) is proposed to solve the CF technique's scalability problem. For successful clustering, Kullback-Leibler Divergence-based Fuzzy C-Means clustering is suggested, with the aim of focusing on greater accuracy during movie recommendation. The proposed KLD-FCM scheme is described as a trustworthy contribution that significantly improves the ability of movie recommendation by virtue of the KL divergence dependent Fuzzy C-Means clustering mechanism and enhanced sqrt-cosine similarity. The proposed scheme emphasized and presented the critical role of the KL divergence-based cluster ensemble factor in improving clustering stability and robustness. For prediction, the enhanced sqrt-cosine similarity was used to calculate successful related neighbor users. The performance of Recommendation is improved when KLD-FCM is combined with improved sqrt-cosine similarity. The proposed KLD-FCM scheme was found to be superior in recommendation Accuracy compared to the COA, FCM-BAT, FCM and ICF approaches, as well as some non-clustering based methods considered for study, when tested on the Movielens dataset in terms of MAE, RMSE, SD, and Recall. With the specified number of clusters, it is capable of providing accurate and

customized movie recommendation systems. In future work, the proposed design has to be tested with different datasets.

## References

---

1. Antony Vijay J, Anwar Basha H, Arun Nehru J (2021) A dynamic approach for detecting the fake news using random forest classifier and NLP. In: In computational methods and data engineering. Springer, Singapore, pp 331–341
2. Asadi E & Charkari N (2012). Video summarization using fuzzy c-means clustering. ICEE 2012 - 20th Iranian Conference on Electrical Engineering. 690–694.  
<https://doi.org/10.1109/IranianCEE.2012.6292442>
3. Basha SM, Rajput DS (2019). Survey on evaluating the performance of machine learning algorithms: past contributions and future roadmap. In deep learning and parallel computing environment for bioengineering systems. Academic Press, Cambridge, pp 153–164
4. Clement J. (2020). Impact of recommendation engine on video-sharing platform - YouTube. <https://doi.org/10.13140/RG.2.2.15746.50882>
5. Cui L & Dong L & Fu X & Wen Z & Lu N & Zhang G. (2016). A video recommendation algorithm based on the combination of video content and social network: CONTENT AND SOCIAL NETWORK BASED VIDEO RECOMMENDATION. Concurrency and Computation: Practice and Experience. 29: <https://doi.org/10.1002/cpe.3900>.
6. Davidson J, Liebald B, Liu J, Nandy P, Vleet T, Gargi U, Gupta S, He Y, Lambert M, Livingston B, Sampath D (2010). The YouTube video recommendation system. 293–296. <https://doi.org/10.1145/1864708.1864770>

[Chapter](#) [Google Scholar](#)

7. De Vriendt J, Degrande N, Verhoeyen M (2011) Video Content Recommendation: An Overview and Discussion on Technologies and Business Models. *Bell Labs Tech J* 16:235–250. <https://doi.org/10.1002/bltj.20513>

[Article](#) [Google Scholar](#)

8. Deldjoo Y. (2019). Enhancing video recommendation using multimedia content. [https://doi.org/10.1007/978-3-030-32094-2\\_6](https://doi.org/10.1007/978-3-030-32094-2_6)

9. Deldjoo Y, Elahi M, Quadrana M, Cremonesi P (2015). Toward Building a Content-Based Video Recommendation System Based on Low-Level Features. <https://doi.org/10.1007/978-3-319-27729-5>

10. Deldjoo Y, Elahi M, Cremonesi P, Garzotto F, Piazzolla P, Quadrana M (2016) Content-Based Video Recommendation System Based on Stylistic Visual Features. *J Data Semant* 5:1–15. <https://doi.org/10.1007/s13740-016-0060-9>

[Article](#) [Google Scholar](#)

11. Deldjoo Y, Schedl M, Cremonesi P, Pasi G (2020) Recommender Systems Leveraging Multimedia Content. *Comput Surv* 53:1–38. <https://doi.org/10.1145/3407190>

[Article](#) [Google Scholar](#)

12. Gupta M, Thakkar A, Gupta V, Rathore DP (2021). Movie Recommender System Using Collaborative Filtering. 978–979

13. Homann L, Martins D, Vossen G, Kraume K (2018) Enhancing traditional recommender systems via social communities. *Vietnam J Comput Sci* 6. <https://doi.org/10.1142/S2196888819500040>

14. Huang Y, Cui B, Jiang J, Hong K, Zhang W, Xie Y (2016). Real-time Video Recommendation Exploration. 35–46. <https://doi.org/10.1145/2882903.2903743>.
15. Kamran M, Shah SS, Baig MN, Khan RH (2020). A movie recommender system by combining both content based and collaborative filtering algorithms
16. Khadse VP, Basha SM, Iyengar N, Caytiles R (2018) Recommendation engine for predicting best rated movies. Int J Adv Sci Technol 110:65–76

[Article](#) [Google Scholar](#)

17. Lu W & Chung FL (2016). Computational Creativity Based Video Recommendation. 793–796. <https://doi.org/10.1145/2911451.2914707>.
18. Mercanoglu O & Yildirim Z (2017). Video Recommendation System Using Collaborative Filtering
19. Mohamed A, Sherif A, Osama F, Roshdy Y, Hassan MA, El Ashmawi WH (2020). A new challenge on video recommendation by content. <https://doi.org/10.1109/ICCES48960.2019.9068169>.
20. Patil, Lalit. (2016). Fuzzy C means clustering MATLAB code. <https://doi.org/10.13140/RG.2.1.3924.9046>.
21. Ramezani M, Yaghmaee F (2016) A novel video recommendation system based on efficient retrieval of human actions. Physica A: Statistical Mechanics and its Applications 457. <https://doi.org/10.1016/j.physa.2016.03.101>
22. Shah P, Sanghvi S (2020) Video Recommender System

23. Tohidi N, Dadkhah C (2020) Improving the performance of video Collaborative Filtering Recommender Systems using Optimization Algorithm. *Int J Nonlinear Anal Appl (IJNAA)* 11:283–295. <https://doi.org/10.22075/IJNAA.2020.19127.2058>

[Article](#) [MATH](#) [Google Scholar](#)

24. Zhou X, Chen L, Zhang Y, Cao L, Huang G, Wang C (2015). Online Video Recommendation in Sharing Community. 1645–1656. <https://doi.org/10.1145/2723372.2749444>.

## Author information

---

### Authors and Affiliations

School of Computer Science and Engineering, REVA University, Bengaluru, India

H. Anwar Basha

Department of Computer Science and Engineering, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India

S. K. B Sangeetha & J. Arunnehru

Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India

S. Sasikumar

Department of CSE , Chaitanya Bharathi Institute of Technology, Hyderabad, India

M. Subramaniam

### Corresponding author

Correspondence to [H. Anwar Basha](#).

### Ethics declarations

---

### Conflict of interest

The author(s) propose a clear no conflict of interest involved in this research work in form of publication in this Journal *Multimedia Tools and Applications*.

## Additional information

---

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Rights and permissions

---

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

[Reprints and permissions](#)

## About this article

---

### Cite this article

Basha, H.A., Sangeetha, S.K.B., Sasikumar, S. *et al.* A proficient video recommendation framework using hybrid fuzzy C means clustering and Kullback–Leibler divergence algorithms. *Multimed Tools Appl* **82**, 20989–21004 (2023).

<https://doi.org/10.1007/s11042-023-14460-8>

Received

14 April 2021

Revised

09 February 2022

Accepted

31 January 2023

Published

07 February 2023

Issue Date

June 2023

DOI

<https://doi.org/10.1007/s11042-023-14460-8>

## Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Provided by the Springer Nature SharedIt content-sharing initiative

## Keywords

[Recommendation systems](#)

[Content-based filtering](#)

[Knowledge base](#)

[Fuzzy C-means algorithm](#)

# Direct current magnetron sputtered Ni<sub>3</sub>Al thin films with electron transport behaviour for superior electromagnetic shielding

Published: 03 April 2023

Volume 129, article number 313, (2023) Cite this article



Applied Physics A

Aims and scope

Submit manuscript

[Santhosh Kumar Adpa](#) , [S. Shanmukharao Samatham](#), [Radhamanohar Aepuru](#), [Kalyani Date](#), [Ravi Prakash Magisetty](#), [Suwarna Datar](#), [S. N. Kale](#), [Rodrigo Espinoza González & Vijaya Bhaskara Rao Bhaviripudi](#) 

 203 Accesses  1 Citation [Explore all metrics](#) →

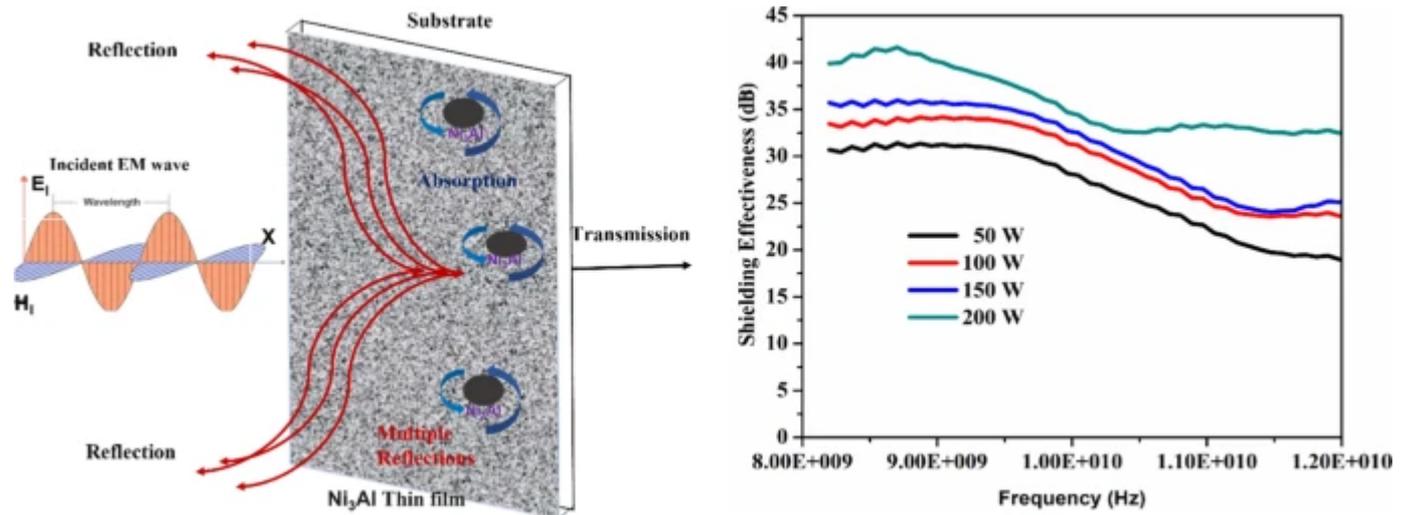
## Abstract

Nanocrystalline Ni<sub>3</sub>Al thin films were deposited by direct current (DC) magnetron sputtering and studied the electron transport mechanism and electromagnetic shielding properties for the first time. The mechanism of electron transport and barrier heights were studied using scanning tunneling microscope (STM). The barrier height of Ni<sub>3</sub>Al thin films were decreased from 1.51 eV to 1.33 eV with sputtering power and the entire tunneling current in these films appeared to come from Schottky emission.

Electromagnetic shielding properties of Ni<sub>3</sub>Al thin films have been studied in X-band (8.2–12.4 GHz) region and the shielding effectiveness (SE) increases with sputtering

power and achieved the highest SE value 36 dB (99.96% attenuation) at 200 W sputter deposited Ni<sub>3</sub>Al films with 396 nm. The sputtering power alters the thickness of Ni<sub>3</sub>Al films thereby enhancing the electromagnetic shielding performance due to the increase of conductivity, the negative dielectric permittivity and magnetic relaxation arising from high nickel (Ni) concentration in Ni<sub>3</sub>Al. The results suggested that as prepared ultra-thin Ni<sub>3</sub>Al films can be an effective and new type of electromagnetic shielding materials.

## Graphical abstract



**i** This is a preview of subscription content, [log in via an institution](#) to check access.

### Access this article

[Log in via an institution](#)

[Buy article PDF 39,95 €](#)

Price includes VAT (India)

Instant access to the full article PDF.

Rent this article via [DeepDyve](#)

[Institutional subscriptions](#) →

## Data availability

---

Data will be available on request.

## References

---

1. R. Kumaran, M. Alagar, S.D. Kumar, V. Subramanian, K. Dinakaran, *Appl. Phys. Lett.* **107**, 113107 (2015)

[Article](#) [ADS](#) [Google Scholar](#)

2. J. Ling, W. Zhai, W. Feng, B. Shen, J. Zhang, W. Zheng, *A.C.S. Appl. Mater. Interfaces* **5**, 2677 (2013)

[Article](#) [Google Scholar](#)

3. J.L. Roland, A.N. Richard, *J. Commun. Health* **17**, 291 (1992)

[Article](#) [Google Scholar](#)

4. S. Stankovich, D.A. Dikin, G.H.B. Dommett, K.M. Kohlhaas, E.J. Zimney, E.A. Stach, R.D. Piner, S.T. Nguyen, R.S. Ruoff, *Nature* **442**, 282 (2006)

[Article](#) [ADS](#) [Google Scholar](#)

5. Y.J. Choi, S.C. Gong, D.C. Johnson, S. Golledge, G.Y. Yeom, H.H. Park, *Appl. Surf. Sci.* **269**, 92 (2013)

[Article](#) [ADS](#) [Google Scholar](#)

6. F.S. Hung, F.Y. Hung, C.M. Chiang, Appl. Surf. Sci. **257**, 3733 (2011)

[Article](#) [ADS](#) [Google Scholar](#)

7. Y. Huang, M.J. Aziz, J.W. Hutchinson, A.G. Evans, R. Saha, W.D. Nix, Acta Mater. **49**, 2853 (2001)

[Article](#) [ADS](#) [Google Scholar](#)

8. A.C. Abhyankar, S.N. Kaul, J. Phys. Condens. Matter. **20**, 445228 (2008)

[Article](#) [Google Scholar](#)

9. A.C. Abhyankar, S.N. Kaul, Appl. Phys. Lett. **88**, 193125 (2006)

[Article](#) [ADS](#) [Google Scholar](#)

10. G.B. Thompson, R. Banerjee, X.D. Zhang, P.M. Anderson, H.L. Fraser, Acta Mater. **50**, 643 (2002)

[Article](#) [ADS](#) [Google Scholar](#)

11. R. Banerjee, H.L. Fraser, Thin Solid Films **441**, 255 (2003)

[Article](#) [ADS](#) [Google Scholar](#)

12. M.T. Le, Y.U. Sohn, J.W. Lim, G.S. Choi, Mater. Trans. **5**, 116 (2009)

[Google Scholar](#)

13. J. Jahanmir, P.E. West, T.N. Rhodin, Appl. Phys. Lett. **52**(24), 2086 (1988)

[Article](#) [ADS](#) [Google Scholar](#)

14. S. Kumar, G. Datt, A.S. Kumar, A.C. Abhyankar, J. Appl. Phys. **120**, 164901 (2016)

[Article](#) [ADS](#) [Google Scholar](#)

15. R.B. Schulz, V.C. Plantz, D.R. Brush, IEEE Trans. Electromag. Compat. **30**, 187 (1988)

[Article](#) [ADS](#) [Google Scholar](#)

16. P. Saini, V. Choudhary, B.P. Singh, R.B. Mathur, S.K. Dhawan Mater. Chem. Phys. **113**, 919 (2009).

17. H. Huang, F.-H. Xue, B. Lu, F. Wang, X.-L. Dong, W.-J. Prog. Electromagn Res. B, **34**, 31 (2011).

18. M. Chen, L. Zhang, S. Duan, S. Jing, H. Jiang, M. Luo, C. Li, Nanoscale **6**, 3796 (2014)

[Article](#) [ADS](#) [Google Scholar](#)

19. B. V. Bhaskara Rao, et al, Phys. Chem. Chem. Phys., **17**, 18363 (2015).

20. X. Shui, D.D.L. Chung, J. Electron. Mater **26**, 928 (1997)

[Article](#) [ADS](#) [Google Scholar](#)

21. O. A. Testov, A. E. Komlev, K. G. Gareev, I. K. Khmelnitskiy, V. V. Luchinin, E. N. Sevost'yanov, I. O. Testov, Coatings, **11**, 1455 (2021)

22. P. Jozwik, W. Polkowski, Z. Bojar, Materials (Basel) **8**(5), 2537–2568 (2015)

[Article](#) [ADS](#) [Google Scholar](#)

## Acknowledgements

---

Vijaya Bhaskara Rao Bhaviripudi acknowledges FONDECYT Project No: 3200452, ANID, Chile, for funding and Rodrigo Espinoza-González for laboratory support. Santhosh Kumar and Shanmukharao S acknowledges Science and Engineering Research Board, India for the financial support through Core Research Grant (CRG/2022/0007993).

## Author information

---

### Authors and Affiliations

Department of Physics, Chaitanya Bharathi Institute of Technology, Hyderabad, India  
Santhosh Kumar Adpa & S. Shanmukharao Samatham

Departamento de Ingeniería Química, Biotecnología y Materiales, FCFM, Universidad de Chile, Santiago, Chile

Rodrigo Espinoza González & Vijaya Bhaskara Rao Bhaviripudi

Departamento de Ingeniería Mecánica, Facultad de Ingeniería, Universidad Tecnológica Metropolitana, Santiago, Chile

Radhamanohar Aepuru

Department of Applied Physics, Defence Institute of Advanced Technology, Pune, India  
Kalyani Date, Suwarna Datar & S. N. Kale

Pohang University of Science and Technology, Pohang, South Korea

Ravi Prakash Magisetty

### Contributions

SKA and VBRB: proof-of-concept, investigation, methodology, data curation, formal analysis, writing-original draft, writing-review and editing. KD: investigation, data curation, SSS: writing-original draft, review and editing, RA: formal analysis, writing-original draft, review & editing, RPM: formal analysis, review and editing, SD: proof-of-concept, supervision, funding resources, writing-review and editing, SNK: proof-of-concept, supervision, funding resources, writing-review and editing, REG: formal analysis, writing-original draft, review and editing.

### Corresponding authors

Correspondence to [Santhosh Kumar Adpa](#) or [Vijaya Bhaskara Rao Bhaviripudi](#).

# Ethics declarations

---

## Conflict of interest

Authors declare that they have known competing financial interest or personal relationships that could have appeared to influence the work reported in this manuscript. Authors also declare that the present manuscript designates our original research work and not submitted or published elsewhere in any form consideration for publication. There is no conflict of interest exists. If the article accepted, it will not be published elsewhere in the same form, in any language, without the written consent of the publisher.

## Additional information

---

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary Information

---

Below is the link to the electronic supplementary material.

[Supplementary file1 \(DOCX 951 KB\)](#)

## Rights and permissions

---

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

[Reprints and permissions](#)

## About this article

---

### Cite this article

Adpa, S.K., Samatham, S.S., Aepuru, R. *et al.* Direct current magnetron sputtered Ni<sub>3</sub>Al thin films with electron transport behaviour for superior electromagnetic shielding. *Appl. Phys. A* **129**, 313 (2023). <https://doi.org/10.1007/s00339-023-06553-w>

Received

28 November 2022

Accepted

05 March 2023

Published

03 April 2023

DOI

<https://doi.org/10.1007/s00339-023-06553-w>

### Keywords

[Intermetallic Ni<sub>3</sub>Al](#)

[Sputtering power](#)

[Shielding effectiveness](#)

[Electron-transport](#)



All



ADVANCED SEARCH

PDF

Help

Conferences > 2023 IEEE International Magne... ?

# Electronic Structure and Exchange Bias of a Compensated Ferrimagnet Mn<sub>2</sub>PtAl

Publisher: IEEE

[Cite This](#)

PDF

Alexey V. Lukoyanov ; S. Shanmukharao Samatham ; Akhilesh Kumar Patel ; P. D. Babu ; K. G. Suresh **All Authors** ...



35 Full Text Views

## Alerts

Manage Content Alerts  
Add to Citation Alerts

### Abstract



Downl PDF

#### Document Sections

- I. Introduction
- II. Theoretical Calculations
- III. Experimental results

**Abstract:**Mn<sub>2</sub> PtAl alloy has been investigated for its electronic structure properties along with structural, magnetic and exchange bias properties. It crystallizes in hexagonal s... [View more](#)

#### Metadata

##### Abstract:

Mn<sub>2</sub> PtAl alloy has been investigated for its electronic structure properties along with structural, magnetic and exchange bias properties. It crystallizes in hexagonal structure with lattice constants of  $a = b = 4.34 \text{ \AA}$ , and  $c = 5.47 \text{ \AA}$  at room temperature. Theoretical calculations reveal a ferrimagnetic arrangement of the magnetic moments of Mn with a small value of total magnetic moment in good agreement with the experimental magnetization measurements. Magnetization isothermal curves in the ambient of a cooling field  $H_{cool} = 70 \text{ kOe}$ , show asymmetric hysteresis with a negative shift along H-direction. The estimated exchange bias in  $H_{cool} = 70 \text{ kOe}$  is giant (2.73 kOe) at 3 K which is attributed to the unidirectional anisotropy associated with ferromagnetic clusters.

**Published in:** 2023 IEEE International Magnetic Conference - Short Papers (INTERMAG Short Papers)

**Date of Conference:** 15-19 May 2023

**DOI:**

**Date Added**

10.1109/3.10228261

**ISBN Info**

**Funding A**

**Access to this document requires a subscription.**

IEEE offers both personal and institutional subscriptions. Whether you are an academic, a practitioner, or a student, IEEE offers a range of individual and institutional subscription options that can meet your needs.

[LEARN MORE](#)

[Close](#)



**I. Introduction**

THE INVESTIGATION of magnetic materials for multifunctional applications has been fascinating with a ray of hope to achieve an efficient multifunctional material for technological applications. For the past one decade, one of the topical interests of condensed matter physics researchers has been to discover materials (without multilayer interface) with large exchange bias for spintronics applications [1]. Some of the recent investigations suggest Heusler alloys are prominent materials for spintronics, magnetic refrigeration and magnetoresistance. Sometimes, the mixed or coexisting contrasting magnetic phases (such as ferromagnetic-antiferromagnetic, ferromagnetic-ferrimagnetic, ferrimagnetic-antiferromagnetic etc.) results in an exchange bias (EB). EB is useful for effective magnetic storage. EB in single uniform antiferromagnets and compensated ferrimagnets is advantageous for the spintronics application.

Authors	PDF
Figures	Help
References	▼
Keywords	▼
Metrics	▼

**More Like This**

Enhancement of saturation magnetization in sputtered FeCoPd alloy and [FeCo-Pd]<sub>n</sub>/super-lattice films at room temperature  
 IEEE Transactions on Magnetics  
 Published: 2005

High field and temperature dependent magnetisation measurements on some RE<sub>2</sub>/Co<sub>17</sub> single crystals  
 IEEE Transactions on Magnetics  
 Published: 1988

Show More

PDF

Help

**IEEE Personal Account**

CHANGE USERNAME/PASSWORD

**Purchase Details**

PAYMENT OPTIONS  
VIEW PURCHASED DOCUMENTS

**Profile Information**

COMMUNICATIONS PREFERENCES  
PROFESSION AND EDUCATION  
TECHNICAL INTERESTS

**Need Help?**

US & CANADA: +1 800 678 4333  
WORLDWIDE: +1 732 981 0060  
CONTACT & SUPPORT

**Follow**



[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#) | [Sitemap](#) | [IEEE Privacy Policy](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2024 IEEE - All rights reserved.

**IEEE Account**

- » Change Username/Password
- » Update Address

**Purchase Details**

- » Payment Options
- » Order History
- » View Purchased Documents

**Profile Information**

- » Communications Preferences
- » Profession and Education
- » Technical Interests

**Need Help?**

- » **US & Canada:** +1 800 678 4333
- » **Worldwide:** +1 732 981 0060

[» Contact & Support](#)

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [Sitemap](#) | [Privacy & Opting Out of Cookies](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.  
© Copyright 2024 IEEE - All rights reserved. Use of this web site signifies your agreement to the terms and conditions.

PDF

Help

# SMART SHOPPING CART USING IOT

**Varshitha Koppula, Sushma polisetty, Trisha karhale, Dr. B.Veera Jyoti,**

Varshitha Koppula B-TECH, Dept of Information Technology, Chaitanya Bharathi Institute of Technology,  
Osman Sagar Rd, Kokapet, Gandipet, Telangana 500075.

Sushma polisetty B-TECH, Dept of Information Technology, Chaitanya Bharathi Institute of Technology,  
Osman Sagar Rd, Kokapet, Gandipet, Telangana 500075.

Trisha karhale B-TECH, Dept of Information Technology, Chaitanya Bharathi Institute of Technology, Osman  
Sagar Rd, Kokapet, Gandipet, Telangana 500075.

Dr. B.Veera Jyoti Assistant Professor, Information Technology, Chaitanya Bharathi Institute of Technology,  
Osman Sagar Rd, Kokapet, Gandipet, Telangana 500075

## Abstract

The ultramodern age of technology in which utmost of the customers needs to stay in the supermarket for shopping because it's a largely time-consuming process. A huge crowd in the supermarket makes trouble to stay by long ranges because of a bar-law grounded billing process. In this regard, an Internet of effects (IoT) grounded smart shopping wain using RFID detectors, an Arduino, a Bluetooth module is proposed as a result to this issue. RFID detectors works on wireless communication. RFID markers are attached to each product and RFID anthology reads the product information efficiently. After this, each product information is showed in the web page. The bill will be generated automatically via WiFi module in web page, which helps customers to get bill conveniently and payment can be done through card at the wain. Since the entire process of billing is automated it reduces the possibility of mortal error mainly. Also the system has a point to cancel the scrutinized products by client to further optimize the shopping experience and it shows offers grounded on our scrutinized products. This is an IOT grounded system which is stoner interactive, ensures security, also time saving which avoids long ranges for billing. This will exclude time-consuming shopping process and quality of services issues.

## 1. INTRODUCTION

A new revolution against industrial, economic, and environmental institutions has been sparked by the Internet of Things. IoT refers to a network of devices, including cars, household appliances, and other goods, that are connected to one another and

may exchange data. These devices are integrated with electronics, software, sensors, actuators, and connections. Interactions between objects are now a reality in this Internet of Things era. With the aid of the Internet, everything in our world is about to be connected together. People purchase everything they need on a daily basis from stores, including food items, clothing, electrical goods, and so on. Thousands of large and small stores are currently up all over the world due to rising consumer demand and expenditure. Customers occasionally complain about incomplete sales data for the goods and unnecessary waiting periods at the billing counters. The traditional billing system needs constant improvement to give customers a better shopping experience. This is an RFID BASED SHOPPING CART to overcome such problems and improve the current technology. This may be accomplished by actually identifying the products with RFID tags and using an RFID reader with a shopping cart. This device makes it possible for the customer to understand the product data, including the rate for everything that is scanned through the RFID Reader.

Many people have repeatedly demonstrated their desire to revolutionised the way we shop. The barcode generating applications for mobile devices, Arduino microcontroller, RFID, and wi-fi sensors are widely used in supermarkets. In many supermarkets nowadays, barcode generation is used as well as in operation. A barcode is a



# SMART STICK FOR VISUALLY IMPAIRED USING IOT

**Dr. B. Veerajyothi<sup>1</sup>, Anusha Bandaru<sup>2</sup>, Anvitha Namasani<sup>3</sup> and Sahithi Chiluveru<sup>4</sup>**

<sup>1</sup>Assistant Professor, Department of Information Technology, CBIT, Hyderabad 500061, India

<sup>2</sup> Department of Information Technology, CBIT, Hyderabad 500061, India

<sup>3</sup> Department of Information Technology, CBIT, Hyderabad 500061, India

<sup>4</sup> Department of Information Technology, CBIT, Hyderabad 500061, India

## ABSTRACT

The visually impaired community experiences significant difficulties in navigating through daily life due to the limitations of their vision. Traditional methods of navigation such as trained dogs and simple white sticks are not efficient in breaking the barrier of blindness and do not provide the necessary level of comfort and independence for individuals with visual impairments. To address this issue, a project has been developed with the main objective of creating a useful smart stick that empowers visually impaired individuals to live independently without relying on other animals or humans for assistance. The smart stick is equipped with sensors that detect fire/smoke, a GPS tracker, voice-oriented devices, pits detecting capabilities, and an Arduino microcontroller. By incorporating these advanced technologies into the smart stick, visually impaired individuals can navigate their surroundings with greater ease and safety. The smart stick not only detects potential hazards but also provides voice guidance and navigation assistance, allowing for a more comfortable and independent way of life. The roles of each are as follows: the voice-oriented devices give alert notification either reached destination or directions or location of a lost stick, the sensors detect the cigarettes smoke or fire and gives an alert notification, whenever there is an obstacle present in the walkway it vibrates. Pit detecting plays the important role as there are many pits across the paths which cannot be known to the blind, with the smart stick they can know that there is a pit, and it gives an alert about it with help of ultrasonic sensor. The proposed approach not only overcomes the disadvantages of the existing methods, but it is also optimal, cost efficient and easier to carry not only for the blind people but also for the elderly ones.

Index terms Arduino kit, ultrasonic sensor, GPS tracker, fire and smoke sensors.

## INTRODUCTION

The visually impaired individuals face a great challenge in navigating unfamiliar environments due to their limited eyesight. In India, the latest survey reports that approximately 1.5 million people are visually challenged and find it difficult to perform their daily activities. This number is part of the 170 million

# Revolutionizing Water Quality Monitoring: The Intersection of Machine Learning and IoT for Enhanced Detection

**Dr.Katta Sugamya**

*Assistant Professor: Department of Information Technology  
Chaitanya Bharathi Institute of Technology  
Hyderabad, India*

**Anusha Bandaru**

*Undergraduate: Department of Information Technology  
Chaitanya Bharathi Institute of Technology  
Hyderabad, India*

**Charitha Gajarla**

*Undergraduate: Department of Information Technology  
Chaitanya Bharathi Institute of Technology  
Hyderabad, India*

**Manasa Bedadha**

*Undergraduate: Department of Information Technology  
Chaitanya Bharathi Institute of Technology  
Hyderabad, India*

**Abstract**—Water quality prediction is an important task in ensuring the safety and sustainability of water resources. With the increasing population and industrialization, the quality of water is becoming a major concern worldwide. Water quality prediction models have been developed to predict the concentration of pollutants in water bodies based on various factors such as temperature, pH, dissolved oxygen, and other chemical and physical parameters. These models use statistical and machine learning techniques to analyze historical data and forecast future water quality. The accuracy of water quality prediction models can be improved by incorporating real-time monitoring data, remote sensing data, and meteorological data. In this project Machine learning algorithms such as random forest and decision tree algorithms are used to find the potability. The use of pH, turbidity and TDS sensors help in real time sensing of data. In this project cloud plays a major role to collect and provide the data for analysis and prediction.

## I. INTRODUCTION

When it comes to pollution, global warming, and other environmental disasters, there is no safe drinking water for the world's pollutants in the 21st century. There are many obstacles to real-time water quality monitoring today, including limited water resources, an ever-increasing population, and so on and so forth. As a result, new methods are needed to monitor the water quality in real time. Some of the parameters determining the water quality includes:

- **pH:** The parameters determining the purity of the water in order to determine the amount of hydrogen ions present in a solution, the pH scale must be used. It shows if the water's pH is basic or acidic. Alkaline water has a pH greater than or equal to 7, whereas acidic water has a pH lower or equal to 7. The pH scale value varies in between 0 and 14. The permissible level for pH ranges between 6.5 and 8.5 for drinking water.

- **Turbidity:** A measure of how many invisible particles are in a given volume of water is known as Turbidity. As turbidity increases, the risk of diarrhea increases. Water is clean when the turbidity is low.
- **TDS:** This refers to total dissolved solids in water. TDS is the total amount of inorganic and organic compounds dissolved in water. TDS levels in drinking water are tolerable up to 500 ppm.
- **Sulphate:** It is a naturally occurring chemical in water, although excessive concentrations can have a laxative effect. The maximum allowable amount for drinking water is 250 ppm.
- **Chloramines:** Chloramines are a disinfectant that is used in water treatment to destroy dangerous microorganisms. Up to 4 ppm is considered appropriate for drinking water.
- **Hardness:** The presence of minerals in water, such as calcium and magnesium, is referred to as hardness. Scaling and appliance damage can result from high amounts of hardness.
- **Organic Carbon:** The total amount of organic matter present in water that can alter both the taste and smell of water is determined by its organic carbon content. Up to 2 ppm is the acceptable range for drinking water.
- **Conductivity:** Water's conductivity, especially is determined by the amount of dissolved minerals, provides a factor in how well it conducts electricity. Poor water quality may be proven by high conductivity levels.
- **Trihalomethanes (THMs):** When chlorine is utilized to disinfect water, THMs are potential side effects of the process. The health of an individual can be harmed by high levels of THMs

Due to rapid Urbanisation and Industrialisation water quality

# Text Extraction from Hoardings using Deep Learning Techniques

**D Jayaram**

IT Department, Chaitanya Bharathi Institute of Technology (A), Hyderabad, India

**CRK Reddy**

CSE Department, MGIT (A), Hyderabad, India

**V Kamakshi Prasad**

CSE Department, JNTUH, India

## Abstract

Still text detection and extraction from hoardings is a challenging task because of its complex backgrounds and large variations of text pattern. Scene text detection and extraction methods based on neural networks have emerged recently and have shown promising results. Previous methods failed in text detection and extraction from complex and large images. Deep learning techniques work effectively on complex and larger images. Convolutional neural network (CNN) is the backbone of the proposed method. Extracted text can be converted into speech then it is helpful to the tourist's in understanding the local languages and also useful to the visually challenged and illiterate persons.

**Key words:** Deep learning, CNN, region score, affinity score, neural networks etc.,

## 1. Introduction

Deep learning techniques are useful to extract the text from the images which contains complex background. In the past few decades, Deep learning has proved to be a very powerful tool because of its ability to handle large amounts of data. The interest to use hidden layers has surpassed traditional techniques, especially in pattern recognition. One of the most popular deep neural networks is Convolutional Neural Networks. The



# Finding the Optimal Path using anytime Repairing A\* Algorithm in Dynamic Environment

P. Kiranmaie | A. Niharika | T. Harini | D. Shruthi

Department of Information Technology, Chaitanya Bharathi Institute of Technology

## To Cite this Article

P. Kiranmaie, A. Niharika, T. Harini and D. Shruthi. Finding the Optimal Path using anytime Repairing A\* Algorithm in Dynamic Environment. International Journal for Modern Trends in Science and Technology 2023, 9(04), pp. 58-63. <https://doi.org/10.46501/IJMTST0904010>

## Article Info

Received: 02 March 2023; Accepted: 28 March 2023; Published: 30 March 2023.

## ABSTRACT

*This research builds the heuristic function utilized in the conventional A-star method to introduce a path-planning algorithm. The reduction in the number of nodes investigated is the primary justification for switching from the Dijkstra algorithm to the A-star method. This project's major goal is to minimize the number of nodes examined and the turning angles. The features are established using the grid technique. Instead of maintaining a search graph, this approach maintains a search tree. Hence, nodes that have already been studied along a different path can be avoided via Anytime Repairing A-star (ARA). In order to do this, a method known as "incremental heuristic search" is used, in which the heuristic function is modified throughout the search to include information about the cost of travelling to the target from the current node. This enables the algorithm to proceed in the desired direction even if the cost of travelling between nodes fluctuates, or in the presence of dynamic impediments.. By utilizing Octile distance as a heuristic function, the number of nodes studied is reduced by around 56% to 81% and the number of turns is decreased by 50% to 75%. Additionally to deal with dynamic environments in which the obstacles itself has dynamic behavior may cause more colliding perspectives and increase the percentage of colliding with the mobile robot, to overcome that the dynamic window approach (DWA) is majorly introduced with two major approach within it, which prevents collision and also to reach its goal destination with optimal solution. As this DWA is proposed to be a collaboration with A-star which was introduced earlier, it basically take the path which is generated previously by A-star and accomplishes it into DWA as reference path in a dynamic environment.*

**KEYWORDS:** Anytime repairing A\* (ARA), Octile Distance, DWA, trajectories.

## 1. INTRODUCTION

Anytime repairing A\* (ARA), Chebyshev Distance, Octile Distance, DWA, optimal path finding, trajectories, mobile robot. reduction in time to respond to any deductions without impact to payment for the invoice and unpredictability in working capital planning.

Anytime repairing A\* (ARA), Chebyshev Distance,

Octile Distance, DWA, optimal path finding, trajectories, mobile robot.

It is possible to use both local and global route planning techniques in the process of path planning. Using projected positions of both static and moving obstacles, DWV developed options for obstacle-avoidable paths. Simulators and tests were used to show how the suggested strategy worked. The

# Identification of Lung Cancer Using Ensemble Methods Based on Gene Expression Data

K. Mary Sudha Rani<sup>1\*</sup>, Dr. V. Kamakshi Prasad<sup>2</sup>

Submitted: 28/05/2023

Revised: 06/07/2023

Accepted: 25/07/2023

**Abstract:** Lung cancer is consistently classified as the most dangerous form of the disease since the beginning of recorded history. Patients with lung cancer who receive appropriate medical care, such as a low-dose CT scan, have a far better chance of survival since the disease is detected and diagnosed early. Nonetheless, there are certain drawbacks to this attempt. The gene expression level in hundreds of genes or cells within each tissue may now be determined because of developments in DNA microarray technology. Even though machine learning (ML) is rapidly being used in the medical field for lung cancer detection, the shortage of interpretability of these models remains a significant hurdle. Machine learning can be used to analyze gene expression data (DNA microarray) to predict whether or not a patient has lung cancer. The Collective Random Forest and Adaptive Boosting were employed to determine who was responsible for the harm. KPCA, or Kernel principal component analysis, was used for the feature reduction procedure. We calculated the correlation between each feature and the target using the statistical parameters provided by KPCA. Determining the proportion of the correct predictions for a given data set is one way to calculate the accuracy of a classification model. We tested the validity of the proposed technique in this work using a dataset including information about lung cancer. The dataset includes GSE4115 from the Gene Expression Omnibus (GEO) database, as well as the expression profiles it contains. The findings demonstrate the Identification of Lung Cancer (IOLC) model's potential to detect lung cancer in terms of accuracy, precision, recall, F-Measure, and error rate, with results indicating an accuracy of 81%, the precision of 81.2%, recall of 78.9%, F-Measure of 77.7%, and error rate of 0.29%, respectively.

**Keywords:** Gene Expression, Lung cancer, Ensemble machine learning Random forest, AdaBoost

## 1. Introduction

Cancer is a disease that causes cell destruction in the body. Cells develop and increase in a controlled manner; nevertheless, this control may fail if an error occurs in the cell's genetic blueprints. A variety of factors can cause this mistake. Lung cancer is the most common and lethal malignant tumor seen worldwide. In 2012, around 1,800,000 new lung cancer cases were detected, with 1,600,000 people dying due to the condition. Lung cancer is more common in women and is the leading cause of cancer death. Although smoking is the primary cause of lung cancer, around 15% of male and 53% of female lung cancer patients did not smoke. Furthermore, it is estimated that 25% of lung cancer patients worldwide did not get the disease due to smoking. Previously, the primary resource in biology was gene networks GNs [1], commonly depicted as graphs with nodes and rods, with nodes representing genes and rods signifying gene interactions. These rods may be assigned a numerical number or weight based on the strength of the relationships between the parties involved. As a result, GNs can uncover genes linked with biological processes and their interactions, providing a complete picture of the processes under inquiry. GNs are widely utilized in many

fields, inquiry. GNs are widely utilized in many fields, including but not limited to biology, healthcare, and bioinformatics.

Furthermore, when it comes to non-smokers have a different carcinogenic pathway, clinic pathological features, epidemiology, and natural history than smokers. Lung cancer is the most common type diagnosed worldwide and the leading cause of cancer fatalities wheezing, hoarseness, chest tightness, coughing, and spitting up blood are all indications of lung cancer. Indications and symptoms include chest discomfort, shortness of breath, and wheezing [2]. To avoid this dreadful situation, we require machine learning algorithms to aid in the early detection and prevention of lung cancer. Treatments can be more effective and less likely to recur if started early in lung cancer [3]. As a result, preventative lung cancer screening and detection may be therapeutically beneficial, particularly for patients with undiagnosed lung disease. Experiments have uncovered the genes responsible for lung cancer mutagenicity and pathogenesis, albeit most genes have only a tenuous link to the disease. To determine whether a gene is linked to lung cancer, one must run several trials, which would necessitate a considerable financial commitment.

On the other hand, machine learning (ML) algorithms can prioritize disease-triggering genes where their significant

<sup>1\*</sup>Research scholar, Dept. of CSE, JNTUH, Assistant Professor, CSE Dept., Chaitanya Bharathi of Technology Hyderabad, Telangana, India  
Email: kmarysudha\_cse@cbit.ac.in

<sup>2</sup>Professor, Dept. of CSE, JNTUH, Telangana, India

studies offer the ability to uncover the association amid cancer identification genes. These findings can potentially be used in the early detection of cancer. In particular, successful ML approaches are described works. These algorithms include artificial neural network-based computer-aided diagnosis [4, 5], ensemble approaches [6, 7], and hybrid methods.

Because of its extensive prevalence, it has been examined for cancer biomarkers that can predict a disease's prognosis. To be more exact, lung carcinoma is one of the most common types of cancer, with tobacco use accounting for over 85 percent of cases. Regrettably, the vast majority of instances are fatal. This is due, in part, to a delayed diagnosis, which necessitates specialized medical procedures such as bronchoscopy. As a result, lung cancer biomarkers are seen as critical in the disease's early identification; as a response, numerous initiatives have investigated non-invasive approaches for testing these biomarkers [8]. Ensemble learning could be effective in our investigation because it can increase a model's robustness and accuracy by merging multiple imperfect classifiers. Ensemble learning, which includes bagging and boosting, can be viewed as a general bagging technique for enhancing cancer classification. Random forest is another popular bagging technique. Gene microarray (GMA) recently appeared as a promising cancer detection and classification technique. Statistical analysis and machine learning methods were used to uncover accurate gene characteristics that can be used as inputs for cancer classification models. The lung cancer data's limited sample size makes the interpretation and training of microarray data problematic. The presence of noise in the samples can have a negative impact on the training models' performance. Furthermore, the random forest was utilized to search the classifiers at random, and in the training stage, a better judgment was generated.

One technique like self-paced learning, while another develops a novel formulation to uncover samples [10]. As the SPL regularizer's penalty steadily increases during optimization, more samples during the training phase are selected modes. Adoption has been quick, especially in multi-task learning [11], image categorization [12], and molecular descriptor selection [13, 14]. So, in the current article, we extract high-quality samples using this strategy. The following contributions were made by this paper, which is given below.

- We initially proposed using the IOLC to train a cancer detection and classification model using DNA microarray technology. The samples' degrees of confidence ranged from high to low.
- We built a machine learning prediction model using the Ensemble Methods Random Forest, and AdaBoost was the second stage of this project.

- The suggested approach's accuracy, F1-score, and recall are much greater than previously utilized classifiers.
- Furthermore, the proposed technique chooses a small number of genes (less than 1%) that are extremely important in predicting the disease's early prognosis.

The following are the organization of the paper: We describe the related work of lung cancer identification models based on genes data in section 2. Section 3 proposed work. The section 4 covers the results and discussion. Finally, in Section 5, concludes the paper.

## 2. Background and Related work

In [15], the authors investigated the link between socioeconomic status and the prevalence of lung cancer in several locations of the world, using educational degrees as a proxy for socioeconomic class. This study's data came from 18 prospective cohorts dispersed over 15 countries, including the United States, Europe, Asia, and Australia. They examined the link between educational level and the incidence of lung cancer in people who had never smoked and those who now or had previously smoked using Cox proportional hazards models. The International Standard Classification of Education was used to harmonise education data, which was then modeled as an ordinal variable divided into four categories. The models were modified to consider age, gender, whether the individuals smoked currently or previously, as well as smoking duration, quantity of cigarettes per day, and time since leaving.

In [16], the authors examined various methods, including machine learning, Ensemble learning, deep learning approaches, and numerous ways based on image processing techniques and text information that contribute significantly to determining cancer malignancy degree. Lung cancer has been listed as one of the most deadly diseases humans have faced since the species' inception. It is even one of the malignancies that causes the most continuous fatalities and contributes significantly to the overall mortality rate. The number of persons diagnosed with lung cancer continues to rise. In India, roughly 70.0 thousand cases are reported each year. It is impossible to identify early because the disease is often asymptomatic in its early stages. As a result, discovering cancer earlier is beneficial to save lives. Learning about a patient's illness as soon as possible will improve their chances of rehabilitation and recovery. Cancer diagnosis frequently relies substantially on technical breakthroughs. They intended to use this to combine or bring together Ensemble learning techniques such as stacking, blinding, Max voting, boosting, and XGBoost to provide a comprehensive methodology for evaluating and investigating the outcomes. Compared to other strategies, the Blinding ensemble learning methodology emerges as

the most successful way based on performance criteria such as accuracy, F1 score, precision, and recall.

In recent years, computer technology has been used to resolve various diagnostic concerns. To accurately predict the lung cancer severity level, these newly designed systems include several deep learning and machine learning tactics and specific image processing methods. As a result, this methodology aims to provide a new and unique approach to lung cancer diagnostics. The initial stage in data collection is to download two benchmark datasets. These datasets contain attribute information extracted from the medical records of a range of individuals. To extract features, the techniques of "Principal Component Analysis (PCA)" and "t-Distributed Stochastic Neighbour embedding (t-SNE)" were used. Furthermore, the deep characteristics are derived from what is known as "the pooling layer of Convolutional Neural Network (CNN)." The Best Fitness-based Squirrel Search Algorithm (BF-SSA), also known as optimal feature selection, is used to pick the features themselves in addition to the important features. This is referred to as feature selection. This hybrid optimization strategy is advantageous in many industries because it more efficiently explores the search space and performs better using feature selection.

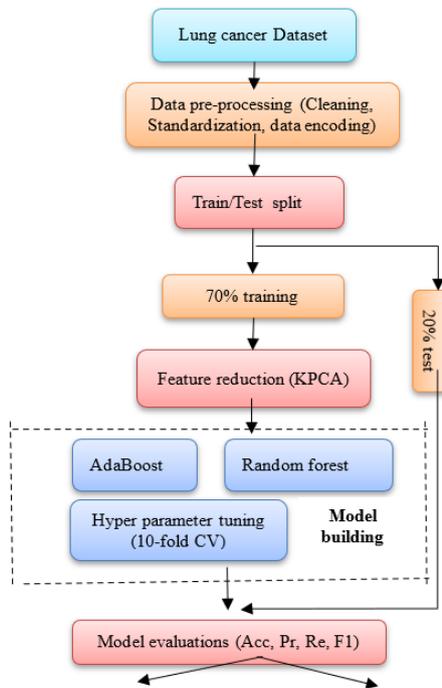
In [18], the authors evaluated relevant surveys, underlining the need for a further study focusing on Ensemble Classifiers (ECs) utilized for cancer diagnosis and prognosis. By integrating several types of input data, learning methods, or characteristics, ensemble approaches strive to increase performance. They are being used in cancer detection and prognosis, among other things. Nonetheless, the scientific community needs to catch up in this technological sphere. A systematic evaluation of ensemble methodologies used in cancer prognosis and diagnosis, coupled with a taxonomy of such methods, can help the scientific community keep up with technology and, if comprehensive enough, even lead the trend. The following stage will thoroughly review the possible methodologies, both classical and deep learning-based. In addition to identifying the well-studied cancer kinds, the

best ensemble methods used for the linked purposes, the most common input data types, the most common decision-making strategies, and the most common assessment methodology, the review creates a taxonomy. All of this happens as a result of the evaluation. Furthermore, they recommend future directions for scholars who want to continue existing research trends or concentrate on areas of the subject that have yet to receive less attention.

In [19], the authors developed Neural Ensemble-based detection for automatic disease detection (NED). An artificial neural network ensemble detects lung cancer cells in patient needle samples. They used Neural Ensemble-based Detection (NED) to achieve autonomous anomalous detection. The ensemble was divided into two levels. Because each network has two outputs, normal and malignant, the first-level ensemble can accurately detect normal cells. There are five types of lung cancer cells produced by each network: adenocarcinoma, squamous cell carcinoma, small cell carcinoma, prominent cell carcinoma, and normal. The second-level ensemble deals with cancer cells discovered by the first. To include network predictions, plurality voting is employed. NED has a high detection rate and a low false negative rate, which occurs when cancer cells are misidentified as normal. This reduces the number of undiagnosed cancer patients, hence saving lives.

### 3. Proposed Model

This section comprehensively explains the methods and materials used in this work, beginning with the architecture of the proposed Identification of Lung Cancer (IOLC) model. Figure 1 depicts the various processes involved in implementing and utilizing the model in the form of major blocks. The following subsections provide a complete overview of the architecture's fundamental building blocks, which include data set collection (genomic data from mutant and normal genes), data preparation (label encoding), feature extraction, and classification.



**Fig 1:** Working procedure of Lung cancer identification model

**Description of the dataset:**

The dataset for discussion here is comparable to one used in a prior study at the Boston University Medical Centre [20, 21]. In these investigations, a microarray was used to examine the level of gene expression found in epithelial cells originating in smokers' respiratory tracts. This dataset was used to extract the levels of expression of 22284 genes collected from 192 smoking subjects. Tissue samples were gathered and isolated from tissue samples. Patients were separated into 3 groups: those who had already been diagnosed with lung cancer (97), those who had not yet been diagnosed with lung cancer (90), and those who were thought to be at a high risk of developing cancer (5). This dataset was chosen specifically for its ability to perform in-depth research into the underlying genetic abnormality in smokers who acquire lung cancer. Although it was created on an older platform (the Affymetrix U133A array), it was chosen carefully. The

dataset, known as GDS2771 and associated with the reference number GSE4115, is freely available for download from the NCBI's Gene Expression Omnibus (GEO) database [22]. The Affymetrix Human Genome U133A Array (HG-U133A) was used as the screening platform to gather this data. This array provided the information on the probesets.

The working procedure of the Lung cancer identification model is repeated for each dataset containing gene expression data. After extracting the probe annotations, it gets represented to a respective gene, and those that do not match any of the genes in the dataset were excluded from further consideration. If the gene has more than one probe, the gene's expression was calculated by taking the mean of all probe expressions. If the gene does not have many probes, the value was calculated by taking the mean of only one probe's expression. The Lung Cancer gene expression dataset is listed in Table 1.

**Table 1** Lung Cancer gene expression dataset

GEO accession number	Disease	Number of samples (Disease/Control)	No. of Genes	Micro array Platform	Platform
GSE 4115	Lung Cancer	187 (97/90)	22,215	Affymetrix Human Genome U133A Array	GPL96 (HG-U133A)

**Data Preprocessing:**

Before using the data for training, 163 samples with complete clinical features were removed from the sample, including 85 smokers who did not have lung cancer and

78 smokers with lung cancer. Clinical data were gathered for future research, including patients' ages, genders, smoking histories, smoking indices, tumor diameters, and the presence or absence of lymphadenopathy. The Affymetrix Human Genome U133A Array platform was

used for the expression profile analysis. Each probe ID was matched to the symbol of a matching gene based on the information saved on the platform (GPL96-15653.txt). Because multiple probes could be associated with the equivalent gene sample, the domino effect was combined and averaged. The Z-score was used to normalize all gene expression values, which was calculated using the standard deviation (SD and mean of every gene symbol and then correcting the X value. This was done to mitigate the effect of variances in the quantities of intrinsic expression found in different genes. The new equation produces the value X, which is the mean/standard deviation ratio. The expression levels of all genes in each dataset were normalized using the methods described below.

$$z_{ij} = \frac{g_{ij} - \text{mean}(g_i)}{\text{std}(g_i)} \quad (1)$$

where  $g_{ij}$  represents the expression value of gene  $i$  in sample  $j$ , and  $\text{mean}(g_i)$  and  $\text{std}(g_i)$  respectively represents mean and standard deviation of the expression vector for gene  $i$  across all samples.

#### Feature reduction:

**Kernel Principal Component analysis:** PCA is widely used when one wants to minimize the dimensionality of a dataset while retaining as much information as possible. The entire dataset (with  $m$  dimensions) is mapped onto a new subspace (with  $j$  dimensions).  $j$  is smaller than  $x$ . This projection approach is useful for reducing both computing costs and the errors that can occur while estimating parameters ("the curse of dimensionality"). Suppose the data cannot be separated linearly. In that case, a nonlinear technique must be used to reduce the dimensionality of the dataset with KPCA, or Kernel Principal Component Analysis, which is a method for analyzing linearly inseparable data. PCA improves output by generating a feature subspace which reduces variance and normalizes the dataset to a unit scale (with mean = 0 and variance = 1). This is required for a wide range of ML methods to perform correctly. The main task is to transform the  $m$ -dimensional dataset (represented by A) into a new sample set (represented by B) with a lower dimension ( $k$  less than  $m$ ). In this situation, B will stand in for the most important part of A, designated by A.

$$B = PC(A) \quad (2)$$

With  $X$  comprises of  $n$  vectors  $(x_1, x_2, \dots, x_n)$ , each  $x_i$  signifies dataset instance, so:

$$\sum_{j=1}^x \delta(a_j) = 0 \quad (3)$$

To compute covariance matrix (CM) we take

$$C = \frac{1}{x} \sum_{j=1}^x \delta(a_j) \delta(a_j)^T \quad (4)$$

The eigenvectors are:

$$Cu_k = \alpha_k u_k, k = 1, \dots, M \quad (5)$$

After constructing the Eigen space from the covariance matrix and removing the less relevant regions, the original data will have a better chance of being accurate. To avoid access to the feature area and instead concentrate on kernels, which are as follows:

$$J(a_l, a_p) = \delta(a_l)^T \delta(a_p) \quad (6)$$

#### Ensemble Learning:

Ensemble learning enhances generalizability and resilience over a single model by aggregating multiple models, like the J-node regression tree. This is achieved by combining the predictions of several simple models or base learners. Bagging and boosting are two typical tactics in group music.

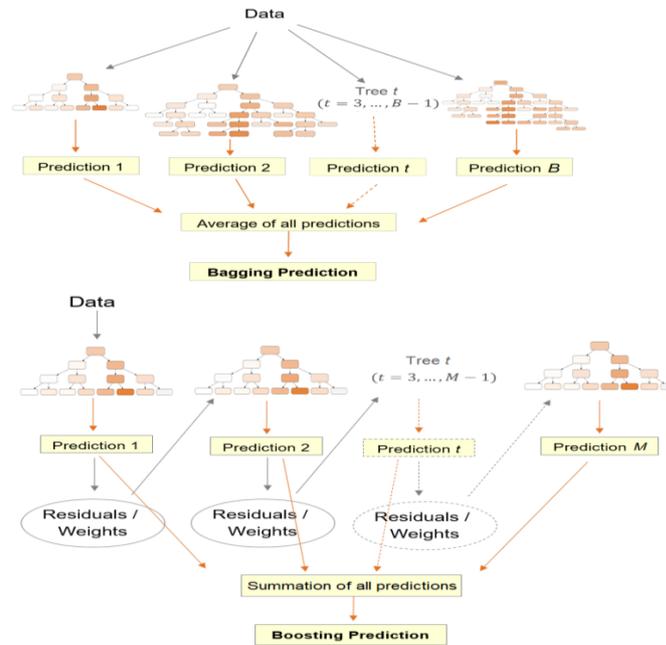
#### Random Forest classifier

This was built using a modified version that generates a huge sample of uncorrelated trees and takes the average of those trees. This enabled the formation of a Random Forest [23]. This was formed as a result of this change. During the tree generation process, choose  $p$  input variables in the random subset from the entire set of  $v$  input variables to be examined for their appropriateness as a candidate for a split. This is referred to as the tree generation process. Because time series forecasting is being performed, each new training set is generated without replacing previous data. As a result, a single regression tree named  $T_b$  is produced by iteratively repeating the steps described below for each node in the tree until the smallest possible node size is obtained. This technique is repeated several times until the tree achieves the appropriate level of precision.

This procedure is performed on every  $B$  tree. The function can be expressed as an average of all  $B$  trees.

$$\hat{F}(x_i) = \frac{1}{B} \sum_{b=1}^B T(x_i; \theta_b) \quad (7)$$

where  $\theta_b$  denotes the tree for node split. In [23], the findings show that using randomness and diversity in tree construction results in a lower generalization error and an overall better model with less variance.



**Fig 2:** Tree structures of Bagging and Boosting

The boosting method entails sequentially developing the trees by combining the knowledge gained from previously formed trees with modified training data (Figure 2). This can be stated as follows:

$$F_m(x_i) = F_{m-1}(x_i) + \sum_{y=1}^{I_m} \gamma_{jm}(x_i \in R_{jm}) \quad (8)$$

So, the updated model will put in the form as

$$\begin{aligned} \hat{F}(x_i) = F_M(x_i) &= \sum_{m=1}^M T(x_i; \Theta_m) \\ &= F_{m-1}(x_i) + \sum_{y=1}^{I_m} \gamma_{jm}(x_i \in R_{jm}) \quad (9) \end{aligned}$$

$\Theta_m = \{R_{jm}, \gamma_{jm}\}_1^m, F_{m-1}(x_i)$  signifies the prior model, while  $\hat{F}(x_i) = F_M(x_i)$  signifies the current tree.

### AdaBoost classifier

The classifier in [24] updates by attaching weights  $\{w_1, w_2, \dots, w_N\}$  for every training instance  $(x_i, y_i)$  (Figure 2). As a result, a total of  $N$  weights will be used. At the start of the procedure, each weight is assigned the value  $w_i = 1/N$ , indicating that the data is being trained in the usual manner. This is known as the learning period. The weighted observations training approach will be continued until all stages have been completed at each subsequent step ( $m = 2, 3$ , etc.). This will be repeated until all phases have been accomplished. The weights of the various components are changed at each of these steps. To be more exact, the weights for the observations that were mistakenly predicted in the previous step are given higher

priority in step  $m$ , whereas the weights for the observations that were correctly predicted are given lower priority. This arises because the weights for the erroneously predicted observations are more likely to contain errors. As a result, as the iterations advance, the findings that are hardest to predict gain increasing emphasis. Finally, as shown in Equation (9), the final prediction is formed by combining the weighted predictions from each tree. This yields the final projection. Gradient boosting, a technique that may be applied to any arbitrary differentiable objective function, can be used to extend boost. Initial training data are used to instruct a tree in the first step of the procedure. As a result, the gradient may be determined to be [25] for all  $i$  values ranging from 1 to  $N$  inclusive.

$$-g_{im} = - \left[ \frac{\partial L}{\partial F(x_i)} \right] F = F_{m-1} \quad (10)$$

For the squared error loss, the negative gradient signifies the residual  $-g_{im} = y_i - F_{m-1}(x_i)$ .

### Model Selection and Validation

An optimization strategy is employed during the learning phase to forecast the values of various parameters based on the collected data. These parameters contain the splitting variable and the splitting point value. On the other hand, each learning algorithm includes a set of hyperparameters that are not learned and must instead be tailored to the unique modeling task at hand. The hyperparameters govern both the model's architecture and its level of complexity. The data and the problem at hand decide their ideal values. However, the training data residual sum of squares cannot be calculated because

doing so weakens a model's capability to generalize to new data. This is because doing so would reduce the size of the training set. As a result, three distinct data sets were used: the training set, the validation set, and the test set. The training set was used to train the model, while the validation set was used to evaluate and fine-tune the model's parameters and hyperparameters. Ultimately, the test set was only used to estimate the generalization error. As a result of this, we were able to select machine learning models with hyperparameter values.

### Performance evaluation

Several validation metrics were discovered during our inquiry. Accuracy (Acc), F1-score (F1), Precision (Pr), and Recall or sensitivity (Re) were among them. The formula for each validation parameter is presented in equations (11) through (14). The abbreviations TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative outcomes, respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (11)$$

$$Re = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

$$Pr = \frac{TP}{TP + FP} \times 100\% \quad (13)$$

$$F1 = \frac{2 \times (Pr \times Re)}{Pr + Re} \times 100\% \quad (14)$$

### 4. Results and Discussion

In this section, the first step is to divide the data into two categories: training (70%) and testing (30%). Several machine learning algorithms, such as feature scaling, KPCA, ROS, and hyperparameter tuning, are utilized to determine the optimum model that delivers the highest level of accuracy. Good classification was picked by

combining all the ML algorithms used in this study. This experiment necessitates the use of specified resources. The suggested system's environment configuration includes an Intel® Core™ i-3-1005G1 CPU running at 1.20GHz, 8GB of RAM, the Anaconda tool, and the Python programming language, which was utilized to construct the model for this study. As shown in Figure 3, Cancer and Non-cancer data of the lung cancer dataset used in this study.

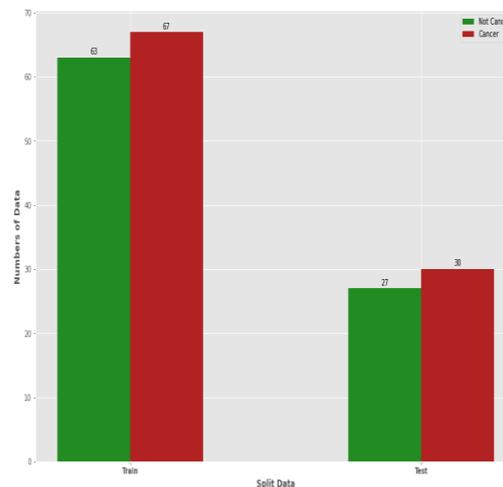


Fig 3: Cancer and Non-cancer data

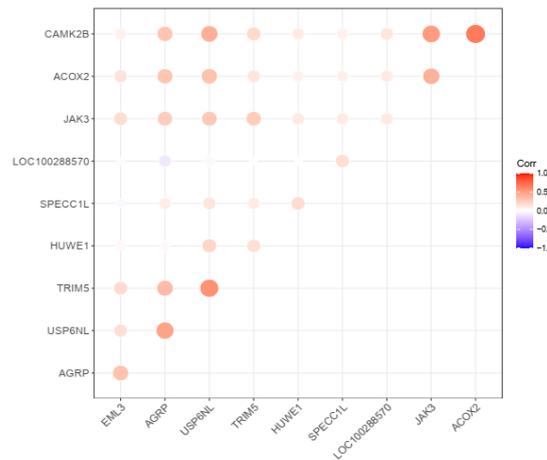
In this paper, we use stable LASSO operation to provide more proof of the efficacy of our technique in computer-assisted diagnostics. Table 1 shows the ten genes that received the highest rankings after being submitted to stable LASSO analysis across all datasets. Most stability ratings are close to one, indicating that the genes chosen are tough. Furthermore, obtain the important p-values that are statistically best for this study. Much research is being done on the functional analysis of gene expression. USP6NL, is one such protein that acts as a GTPase activator for RAB5A.

Table 1: Best 10 genes from GSE 4115 dataset

Gene Name	Gene Symbol	Stabl e Score	p-Value
USP6 N-terminal like	(USP6NL)	1	<0.01
acyl-CoA oxidase 2	(ACOX2)	0.98	<0.01
agouti related neuropeptide	(AGRP)	0.53	<0.01
HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	(HUWE1)	0.99	<0.01
calcium/calmodulin dependent protein kinase II beta	(CAMK2B)	1	<0.01
tripartite motif containing 5	(TRIM5)	1	<0.01
Janus kinase 3	(JAK3)	1	<0.01
sperm antigen with calponin homology and coiled-coil domains 1 like	(SPECC1L)	0.96	<0.01
sperm antigen with calponin homology and coiled-coil domains 1 like	(EML3)	1	<0.01
glycosylphosphatidylinositol anchor attachment protein 1 homolog (yeast) pseudogene	(LOC100288570)	1	<0.01

Figure 4 shows the heat map correlation discovered between the genes in the meantime. Red is used when there is a positive correlation, and violet is used when there is a negative correlation. The stronger the

correlation, the greater the degree of resemblance. As seen in Figure 4, most identified genes have a positive relationship.



**Fig 4:** Graph showing heat map

Several well-known matrices, such as accuracy, recall (also known as sensitivity), precision, and F1-score, are used to assess the classification algorithms' performance.

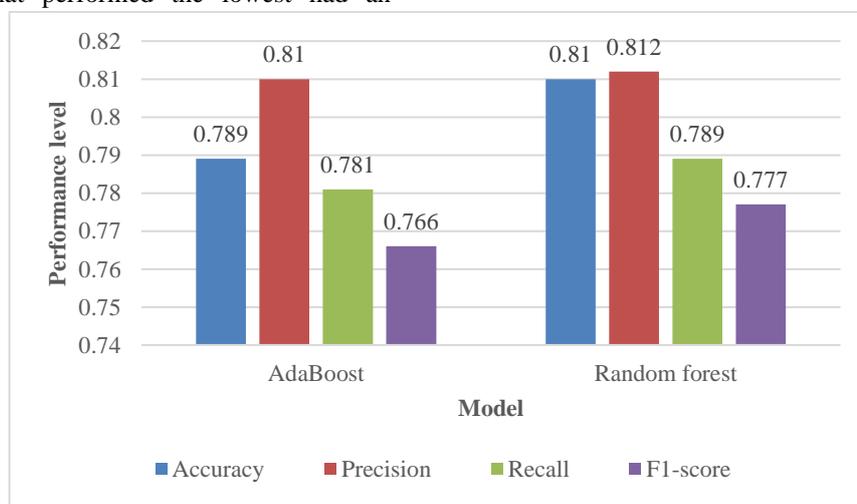
Table 2 shows the performance of RF and AdaBoost in terms of various evaluation metrics.

**Table 2:** Performance analysis of RF and AdaBoost models

Model	Accuracy	Precision	Recall	F1-score
AdaBoost	0.789	0.810	0.781	0.766
Random forest	0.810	0.812	0.789	0.777

In the case of the GSE4115 example presented in Figure 5, the best model obtained is a Random forest, with an accuracy of 0.810 and an F-1 score of 0.777, respectively. This demonstrates that ML is a suitable strategy for working with the dataset. Meanwhile, we noticed that the AdaBoost model that performed the lowest had an

accuracy of 0.789 and an F-1 score of 0.766. Meanwhile, the best recall score in MI for the Random forest classification approach is 0.789, implying that all models can reliably predict genuine positives while avoiding false pessimistic predictions.

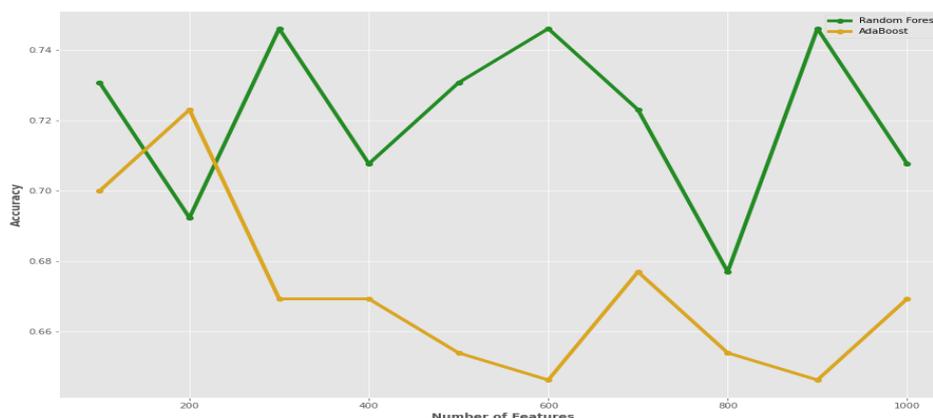


**Fig 5:** Performance comparison of AdaBoost and Random forest on GSE 4115 dataset

## Results on Feature Selection operation:

We used 5-fold cross-validation to investigate the effect of feature number on overall model performance. This enabled us to reduce the overall amount of features. The ideal number of attributes was determined by examining a range of values from 2 to 10. Figure 6 shows that the

variation of the scores produced using AdaBoost and Random Forest is substantially more considerable than that achieved using any other approaches for GSE4115. This indicates that the AdaBoost approach's performance highly depends on the number of features utilized. In an unexpected turn of events, the Random forest approach revealed a significant decline in a feature's overall score.



**Fig 6:** Performance comparison of AdaBoost and Random forest after applying feature selection on GSE 4115 dataset

## 5. Conclusion

In this article, we propose a novel method for detecting lung cancer by building an ensemble classifier and comparing its findings to the RF classifier. In the Ensemble-Classifier, we used two machine learning models: AdaBoost and Random Forest. We begin by extracting features from the dataset, then divide it into 70:30 proportions for training and testing. We classified cancer as Tumor or Normal using the confusion matrix and then provided a classification report that contained accuracy, precision, recall, and F1-score. The feature selection procedure involved calculating the correlation between the feature and the target using statistical parameters, also known as KPCA. Deep learning techniques, such as CNN, may one day aid in diagnosing lung cancer. Images from many scanning modalities, including MRI, CT, PET, and X-ray, can be considered. This can increase precision, allowing the medical sector to provide rapid prevention at a minimal cost. In addition to categorized information, continuous information can be used.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

- [1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2018. *CA: a cancer journal for clinicians*, 68(1):7–30, 2018.
- [2] Lindsey A. Torre, Rebecca L. Siegel, and Ahmedin Jemal. *Lung Cancer Statistics*. Springer International Publishing, 2016.
- [3] Howard Lee and Yi Ping Phoebe Chen. Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications* 42(12):5356–5365, 2015.
- [4] Azian Azamimi Abdullah and Syamimi Mardiah Shaharum. Lung cancer cell classification method using artificial neural network. *Information engineering letters*, 2(1), 2012.
- [5] Z. Cai, D. Xu, Q. Zhang, J. Zhang, S. M. Ngai, and J. Shao. Classification of lung cancer using ensemble-based feature selection and machine learning methods. *Molecular Biosystems*, 11(3):791–800, 2015.
- [6] Maciej Zięba, Jakub M Tomczak, Marek Lubicz, and Jerzy Świątek. Boosted svm for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied soft computing*, 14:99–108, 2014.
- [7] Golrokh Mirzaei, Anahita Adeli, and Hojjat Adeli. Imaging and machine learning techniques for diagnosis of alzheimer’s disease. *Reviews in the Neurosciences*, 27(8):857–870, 2016.
- [8] Aboul Ella Hassanien, Hossam M Moftah, Ahmad Taher Azar, and Mahmoud Shoman. Mri breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. *Applied Soft Computing Journal*, 14(1):62–71, 2014.
- [9] Qingyong Wang, Liang-Yong Xia, Hua Chai, and Yun Zhou. Semi-supervised learning with ensemble self-training for cancer classification. In *2018 IEEE*

Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud& Big Data Computing, Internet of People and Smart City Innovation (Smart World/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), pages 796–803. IEEE, 2018.

- [10] M Pawan Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems*, pages 1189–1197, 2010.
- [11] Changsheng Li, Junchi Yan, Fan Wei, Weishan Dong, Qingshan Liu, and Hongyuan Zha. Self-paced multi-task learning. In *AAAI*, pages 2175–2181, 2017.
- [12] Ye Tang, Yu Bin Yang, and Yang Gao. Self-paced dictionary learning for image classification. In *ACM International Conference on Multimedia*, pages 833–836, 2012.
- [13] Liang-Yong Xia, Qing-Yong Wang, Zehong Cao, and Yong Liang. Descriptor selection improvements for quantitative structure-activity relationships. *International Journal of Neural Systems*, pages 1–16, 2019.
- [14] Abiezer, Otniel & Nhita, Fhira & Kurniawan, Isman. (2022). Identification of Lung Cancer in Smoker Person Using Ensemble Methods Based on Gene Expression Data. 89-93. 10.1109/IC2IE56416.2022.9970035.
- [15] Onwuka, Justina & Zahed, Hana & Feng, Xiaoshuang & Alcalá, Karine & Johansson, Mattias & Robbins, Hilary & Consortium, Lung. (2023). Abstract 1950: Socioeconomic status and lung cancer incidence: An analysis of data from 15 countries in the Lung Cancer Cohort Consortium. *Cancer Research*. 83. 1950-1950. 10.1158/1538-7445.AM2023-1950.
- [16] Fatima, Fayeza Sifat & Jaiswal, Arunima & Sachdeva, Nitin. (2023). Lung Cancer Detection Using Ensemble Learning. 10.1007/978-3-031-23724-9\_15.
- [17] Zolfaghari, Behrouz & Mirsadeghi, Leila & Bibak, Khodakhast & Kavousi, Kaveh. (2023). Cancer Prognosis and Diagnosis Methods Based on Ensemble Learning. *ACM Computing Surveys*. 55. 10.1145/3580218.
- [18] Pradhan, Kanchan & Chawla, Priyanka & Tiwari, Rajeev. (2022). HRDEL: High Ranking Deep Ensemble Learning-based Lung Cancer Diagnosis Model. *Expert Systems with Applications*. 213. 118956. 10.1016/j.eswa.2022.118956.
- [19] Zhou, Zhi & Yang, Yu & Chen, Shi. (2002). Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles. *Artificial intelligence in medicine*. 24. 25-36. 10.1016/S0933-3657(01)00094-X.
- [20] Spira, A.; Beane, J.E.; Shah, V.; Steiling, K.; Liu, G.; Schembri, F.; Gilman, S.; Dumas, Y.M.; Calner, P.; Sebastiani, P.; et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* 2007, 13, 361.
- [21] Gustafson, A.M.; Soldi, R.; Anderlind, C.; Scholand, M.B.; Qian, J.; Zhang, X.; Cooper, K.; Walker, D.; Mc Williams, A.; Liu, G.; et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci. Transl. Med.* 2010, 2, 26ra25–26ra25.
- [22] Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002, 30, 207–210.
- [23] Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32.
- [24] Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* 1997, 55, 119–139.
- [25] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer: Berlin, Germany, 2009.
- [26] Leila Abadi, Amira Khalid, Predictive Maintenance in Renewable Energy Systems using Machine Learning , *Machine Learning Applications Conference Proceedings*, Vol 3 2023.
- [27] Martin, S., Wood, T., Hernandez, M., González, F., & Rodríguez, D. Machine Learning for Personalized Advertising and Recommendation. *Kuwait Journal of Machine Learning*, 1(4). Retrieved from <http://kuwaitjournals.com/index.php/kjml/article/view/156>
- [28] Raghavendra, S., Dhabliya, D., Mondal, D., Omarov, B., Sankaran, K. S., Dhabliya, A., . . . Shabaz, M. (2022). Development of intrusion detection system using machine learning for the analytics of internet of things enabled enterprises. *IET Communications*, doi:10.1049/cmu2.12530

# Exploring Machine Learning in Lung Cancer: Predictive Modelling, Gene Associations, and Challenges

<sup>1</sup>K. Mary Sudha Rani<sup>2</sup> Dr. V. Kamakshi Prasad

**Submitted:** 13/02/2023

**Revised:** 11/04/2023

**Accepted:** 09/05/2023

**Abstract:** Lung cancer is a disease with a high mortality rate and widespread occurrence. Therefore, developing accurate prediction methods and practical gene association analyses is crucial. The utilization of high-throughput genomic data to reveal significant genetic factors has seen an increase in the application of machine-learning techniques. This document presents a thorough examination of machine learning methodologies that are presently utilized to forecast lung cancer and scrutinize gene correlations. The analysis examines different data types, such as gene expression profiles, genomic variants, and clinical data. The primary focus is on integrating multi-omics data for a more comprehensive understanding. Our study comprehensively examines a variety of machine learning algorithms, including traditional methods such as support vector machines and random forests, advanced deep learning architectures, and network-based methodologies. The following discourse explores the pragmatic utilization of the methods above in predictive modeling, biomarker identification, and drug discovery routes. The article addresses common obstacles in the field, such as interpretability and validation, and proposes potential avenues for future research, such as incorporating multi-omics data and implementing personalized medicine. This survey provides a detailed analysis of the recent advancements in machine learning techniques for lung cancer research. It aims to establish a strong basis for future improvements in diagnosis, prognosis, and treatment strategies.

**Keywords:** Lung cancer, prediction, gene association, machine learning, high-throughput genomic data, multi-omics data, support vector machines, random forests, deep understanding, network-based methodologies, predictive modeling, biomarker identification

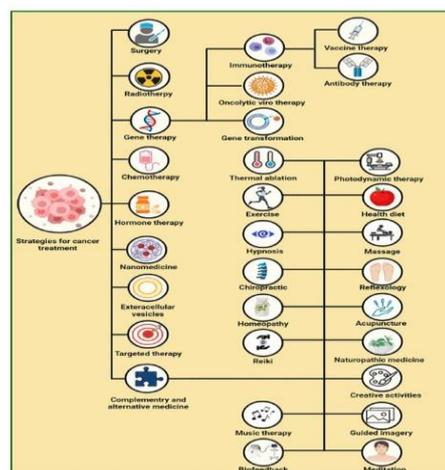
## 1. Introduction:

Cancer is a complex disease that manifests through abnormal molecular pathways. The intricate biological mechanisms and operational effects of gene sets linked to cancer are fundamental to understanding disease progression, identifying therapeutic targets, and developing personalized treatment plans. As such, contemporary cancer research leans heavily on gene-set analysis to interpret high-throughput genomic data within cancer-related biological processes [1].

Lung cancer, in particular, remains a significant global concern. With an estimated 1.8 million new cases annually, it presents a pressing public health issue [2]. It

is reported that 85% of lung cancers are non-small cell lung cancers (NSCLC) [3]. Precise prediction methods and meticulous gene correlation studies are indispensable to augment lung cancer detection, diagnosis, prognosis, and treatment.

Integrating algorithms, machine learning, and multi-omics data can enhance our understanding and prediction of lung cancer. Recent advances in the field have incorporated machine-learning techniques in lung cancer research to analyze and interpret high-throughput genomic data [4]. These methodologies pave the way for identifying lung cancer-related genes through genetic and molecular data analysis.



<sup>1</sup>Research Scholar,CSE dept., JNTUH Hyderabad, Assistant Professor,AIML Dept., Chaitanya Bharathi Institute of Technology, Hyderabad.

kmarysudha\_cseaiml@cbit.ac.in

<sup>2</sup> Professor,CSE dept., JNTUH Hyderabad

kamakshiPrasad@jntuh.ac.in

Such techniques facilitate the analysis of varied data types, including gene expression profiles, genomic variants, and clinical data. They aid in identifying biomarkers, disease subtypes, and fundamental molecular mechanisms. With its capacity to unveil intricate patterns and relationships in genomic data, machine learning is poised to revolutionize our approach to cancer research.

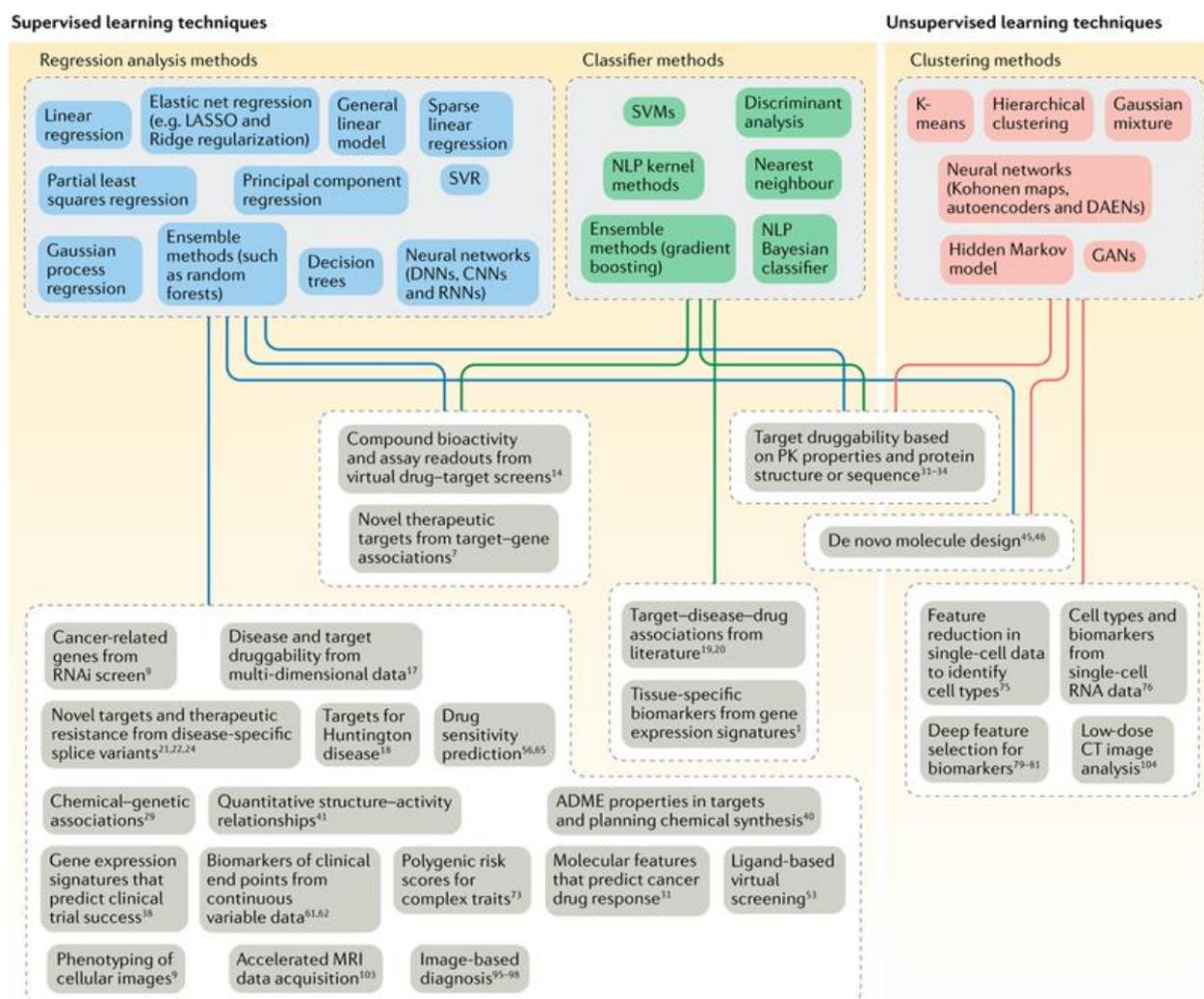
One of the promising advancements in this arena is the integration of multi-omics data, given its potential to enhance lung cancer analysis [5]. By leveraging genomic, transcriptomic, epigenomic, and other omics data, researchers are offered a more comprehensive understanding of the molecular foundation of lung cancer. This, in turn, fosters the discovery of novel diagnostic and therapeutic targets.

Machine learning methods encompass a wide range of tools, including support vector machines (SVM), random forests (RF), and logistic regression, as well as deep learning architectures like artificial neural networks (ANN), convolutional neural networks (CNN), and

recurrent neural networks (RNN) [7]. These algorithms are proficient in executing lung cancer research feature selection, classification, clustering, and prediction tasks.

Gene interaction networks provide insights into gene-disease connections and may facilitate the identification of novel NSCLC genes. Graph-based algorithms such as Graph Convolutional Networks (GCN) and deep walk can analyze these networks effectively. Furthermore, network-based methodologies have emerged as promising tools for studying gene associations in lung cancer [8].

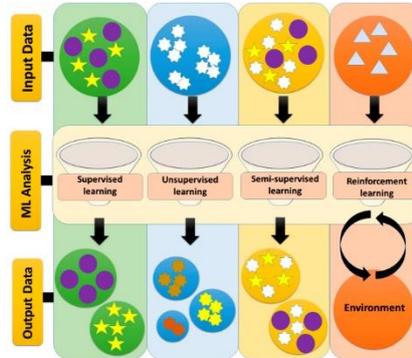
Machine learning has already led to significant strides in lung cancer research. Predictive models employing machine learning have shown promise in identifying high-risk individuals, aiding early detection and customized treatment [9]. Moreover, machine learning has unveiled prognostic biomarkers and implicated molecular pathways in the causation of lung cancer. These insights can optimize drug discovery pathways, identify therapeutic targets, and improve treatment outcomes [10].



## Machine learning tools and their drug discovery applications.

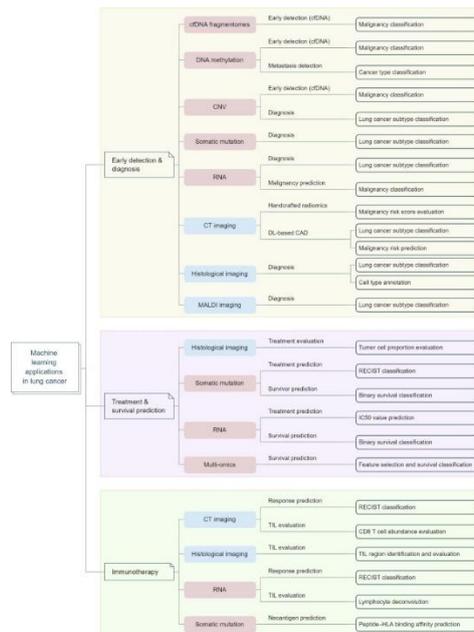
Despite this progress, several challenges persist in applying machine learning for lung cancer research. The necessity for transparency and interpretability in complex algorithms presents a hurdle, especially considering the

need for decision-making tools in clinical settings [11]. The heterogeneity of data sources and the requirement for large and diverse datasets for robust analysis adds further complexities to the validation and reproducibility of studies. Furthermore, harmonization, normalization, and multi-omics data integration are prerequisites for accurate analysis [12].



This paper comprehensively surveys machine-learning techniques for predicting lung cancer and analyzing gene associations. Our exploration extends across the various types of data, challenges faced in the field, and potential

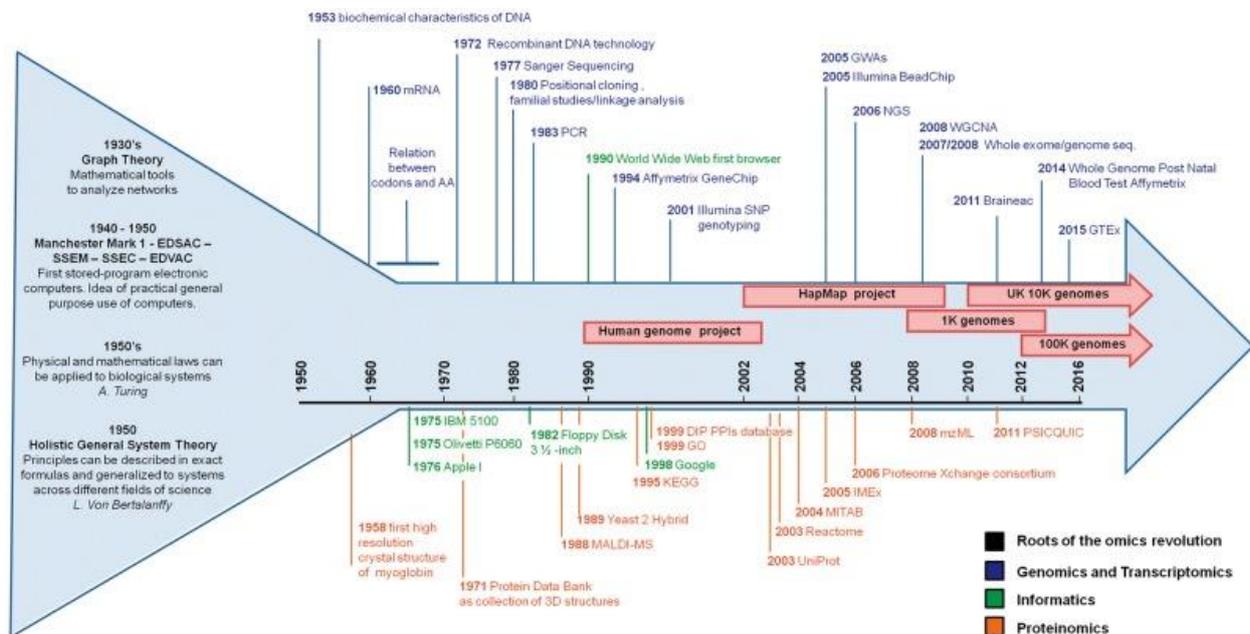
future directions, ultimately serving as a robust foundation for future progress in diagnosis, prognosis, and treatment strategies.



## 2. Background:

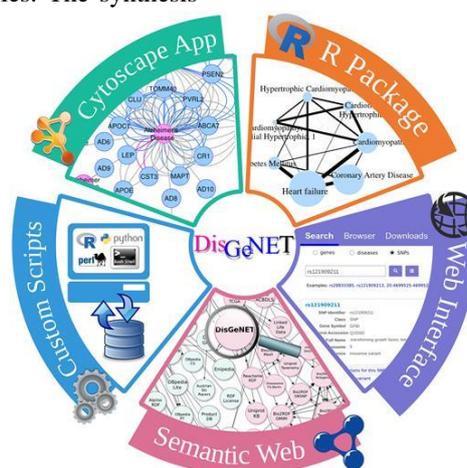
The exponential growth and complexity of biomedical data present both challenges and opportunities. High-throughput technologies such as genomics,

transcriptomics, proteomics, and metabolomics have propelled a surge in biological data production, thus providing deeper insights into health and disease states [13].



Translational bioinformatics seeks to bridge the gap between basic research and medical applications, leveraging biomedical data to understand diseases and develop personalized medicine strategies. The synthesis

of computational and statistical methods with biological and clinical data yields practical insights with potential real-world impact [14].



The DisGeNET platform emerges as a valuable tool for studying and interpreting human disease genes. DisGeNET serves up gene-disease associations that can be examined in the context of numerous diseases. Drawing from curated databases, scientific literature, and various online resources, it offers a comprehensive knowledge base to researchers and clinicians [15].

The platform encompasses genetic variants, gene-disease associations, disease pathways, and pharmacogenomics data. It integrates information from PubMed, ClinVar, UniProt, and GWAS Catalog. Standardized vocabularies like the Unified Medical Language System (UMLS) are used to ensure consistency and comprehensibility. What sets DisGeNET apart is its extensive catalog of gene-disease links, spanning both well-established and preliminary associations. This capability opens up new avenues of exploration and potential research subjects for

investigators. DisGeNET also offers user-friendly search and filtering options, enabling users to search for genes, diseases, and associations based on criteria like evidence level, data source, or disease category. The resulting information includes supporting evidence, genetic variants, linked phenotypes, and pertinent references for gene-disease associations. DisGeNET integrates with other bioinformatics tools and workflows via APIs. This functionality facilitates the integration of DisGeNET data with other relevant datasets for comprehensive analysis.

Overall, DisGeNET significantly simplifies the process of gene-disease research and its implications for human health. It proves invaluable to professionals in genetics, genomics, precision medicine, and drug discovery, as well as clinicians and researchers. As a centralized repository of gene-disease associations, DisGeNET supports identifying therapeutic targets, discovering novel disease

mechanisms, and developing personalized treatment strategies.

### Genome-Wide Association Studies

Genome-Wide Association Studies (GWAS) have profoundly influenced gene-disease research by identifying genetic variations associated with diseases, including lung cancer. GWAS examines numerous genetic variations across the genome to find correlations between genetic variants and disease risk. They specifically target common genetic variations, usually Single Nucleotide Polymorphisms (SNPs), that could affect the risk of specific diseases [16].

The GWAS process involves two main steps: genotyping and association analysis. Genotyping identifies genetic variations in large cases (individuals with the disease) and controls (healthy individuals), using high-throughput genotyping technologies to analyze hundreds of thousands or even millions of SNPs across the genome. The association analysis then identifies genetic variants associated with a disease phenotype, correcting for population stratification and genetic ancestry.

GWAS has made significant discoveries in lung cancer genetics. Hung et al. (2008) found a common genetic variant in the nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNA4) that increased lung cancer risk. Other studies, such as those by Hu et al. (2012) and Wang et al. (2020), identified susceptibility loci for lung adenocarcinoma and linked lung cancer risk to MUC4 and PRSS8 gene variants, respectively.

While GWAS findings offer valuable insights into the genetic makeup of lung cancer and identify genes and pathways for further study, they primarily reveal correlations, not causality. To establish mechanisms and basis, functional analyses and validation are necessary.

GWAS also has some limitations. They focus on common genetic variations with moderate to large effect sizes, potentially overlooking rare or minor variations. They are also often conducted on populations with specific ancestries, limiting their applicability to others. Furthermore, while GWAS can identify genetic variations linked to disease susceptibility, they do not explain the underlying biological mechanisms.

To overcome these constraints, several methods can be employed. These include meta-analysis to increase sample size and statistical power, integrating GWAS data with functional genomics and gene expression data to rank candidate genes, and fine-mapping to reduce genomic regions of interest.

In conclusion, GWAS have significantly contributed to our understanding of the genetics of lung cancer and other complex diseases, identifying potential target genes and

pathways for further study. However, interpreting GWAS findings requires caution and additional functional studies to clarify biological mechanisms and establish causality [14].

### Next-Generation Sequencing (NGS)

Next-Generation Sequencing (NGS) technologies have revolutionized genetic research by enabling fast and cost-effective sequencing of large amounts of DNA or RNA, providing researchers with an unprecedented volume of genomic data. Specifically, Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) have significantly contributed to the identification of disease-related genes, including those linked to lung cancer [17].

WES focuses on sequencing the protein-coding regions of the genome, known as the exome, while WGS sequences the entire genome, including both protein-coding and non-coding regions. These NGS technologies have the advantage of detecting rare genetic variants that traditional genotyping methods may miss. These rare variants can provide insights into disease susceptibility and help uncover genes and pathways associated with lung cancer.

The application of NGS in cancer research has led to the discovery of driver mutations and novel genes associated with lung cancer. For example, using NGS, the Cancer Genome Atlas (TCGA) project has identified recurrent somatic mutations in genes like TP53, EGFR, KRAS, and ALK in lung adenocarcinoma. NGS can also detect somatic copy number alterations, structural variations, and gene fusions, revealing complex mutational patterns in lung cancer. These molecular alterations can aid in the classification of lung cancer subtypes and guide treatment decisions.

NGS technologies also present several challenges. The high volume of sequencing data produced necessitates robust computational and bioinformatics infrastructure for data storage, processing, and analysis. Advanced bioinformatics algorithms and tools are required for reading alignment, variant calling, and annotation.

Interpreting the vast number of genetic variants identified by NGS can be challenging. Distinguishing between pathogenic and benign variants is complex and often requires functional studies to confirm their effects. Additionally, the cost of WGS, which involves sequencing the entire genome, can be prohibitive, though technological advances are gradually making NGS more affordable and accessible.

In conclusion, NGS technologies, specifically WES and WGS, have substantially advanced our understanding of the genetic underpinnings of lung cancer and other diseases. They offer a powerful tool for genomic profiling,

detecting rare variants, and discovering disease-related genes and pathways. Yet, to fully harness the potential of NGS in lung cancer genetics, issues concerning data analysis, interpretation, and cost need to be addressed.

### Machine learning and deep learning

Machine learning and deep learning have been increasingly used in bioinformatics to predict disease-gene associations. These methods leverage computational algorithms and statistical models to analyze large volumes of genomic and biomedical data, aiding the identification of disease-related genes and contributing to a better understanding of disease mechanisms [18].

Commonly utilized algorithms in disease-gene association studies include Support Vector Machines (SVM), Random Forest (RF), and logistic regression. These algorithms are capable of handling high-dimensional data and capturing complex genetic-disease relationships. They facilitate the identification of disease-related genes by performing feature selection, classification, and prediction.

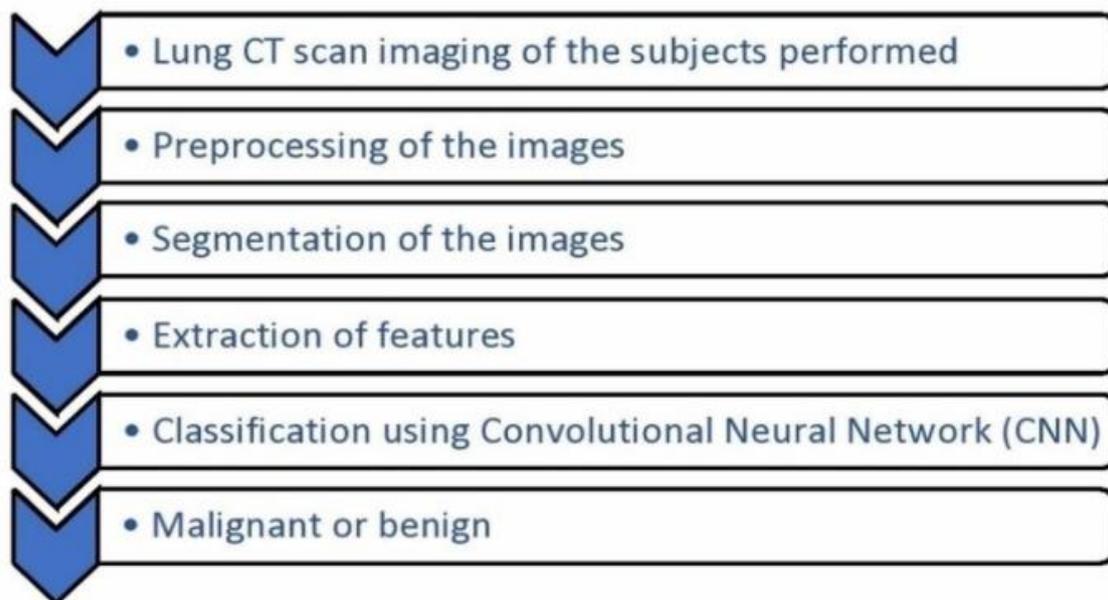
Deep learning, mainly Artificial Neural Networks (ANNs), has recently gained popularity in bioinformatics. ANNs, by mimicking the human brain's function, can learn complex patterns and extract meaningful representations from high-dimensional data. In genomics and disease-gene association studies, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are widely used types of ANNs.

For instance, Zhu et al. developed a novel framework for predicting disease-gene associations using embedding

graph representation and Graph Convolutional Networks (GCN). Graph embedding methods represent gene interaction networks in low-dimensional vector spaces. GCN, a graph-based deep learning approach, leverages network structure to uncover gene-disease relationships. Their model outperformed other state-of-the-art methods in predicting disease-gene associations on multiple benchmark datasets, illustrating the potential of deep learning approaches in elucidating complex gene-disease relationships.

Despite their significant contributions, machine learning and deep learning methods have limitations in bioinformatics. Deep learning models, often described as 'black boxes,' can be challenging to interpret, complicating understanding the underlying biological mechanisms and drawing meaningful conclusions. Furthermore, the performance of these models heavily depends on the quality and representativeness of training data, appropriate feature selection, and careful hyperparameter tuning.

In conclusion, machine learning and deep learning methods have ushered in transformative changes in bioinformatics, notably in predicting disease-gene associations. These methods facilitate the analysis of genomic data, expose complex gene-disease relationships, and pinpoint potential disease markers. Zhu et al.'s application of graph embedding representation and GCN is a prime example of how deep learning approaches can accelerate the discovery of new disease-related genes and pathways.



## Steps in a deep learning model of lung nodule detection

### GCN and DeepWalk

DeepWalk and Graph Convolutional Networks (GCNs) are two well-known methods for analyzing and modeling complex biological networks and have become increasingly popular in bioinformatics [19].

DeepWalk uses a strategy of embedding nodes into low-dimensional vector spaces, capturing the network structure. It employs a Skip-Gram model to learn node embeddings from sequences of nodes generated by random walks. By mapping nodes to continuous vector representations, DeepWalk facilitates downstream analysis tasks.

On the other hand, GCN is a deep learning architecture explicitly designed for graph-structured data. GCN predicts node-level attributes by aggregating information from neighboring nodes. By harnessing spectral graph theory and convolutional operations, GCN learns node representations that capture the graph's structure locally and globally. GCN has shown exceptional performance in predicting protein-protein interactions, analyzing gene expression, and predicting disease-gene associations.

In the context of disease-gene association prediction, a combined DeepWalk-GCN framework has been proposed. In this approach, DeepWalk is used to generate gene node embeddings from the gene interaction network, capturing the structure of the gene network. The GCN model then utilizes these embeddings to predict disease-gene associations. Notably, this integrated model was able to predict network connectivity patterns accurately.

The study demonstrates the utility of DeepWalk and GCN in leveraging network structure and node embeddings to predict disease-gene associations. DeepWalk extracts informative features from the gene interaction network, which GCN then utilizes to capture complex gene-disease relationships. The synergy of these two methods improves predictive performance and deepens our understanding of gene-disease associations.

### Deep Neural Networks (DNNs)

Deep Neural Networks (DNNs) are potent machine learning models that derive their structure and function from the human brain. These models consist of multiple layers of artificial neurons, nodes, or units, which process and transform input data to generate a prediction. The strength of DNNs lies in their ability to automatically learn hierarchical representations of data, thereby capturing complex patterns and relationships[20].

DNNs have found extensive applications in fields like computer vision, natural language processing (NLP), and bioinformatics. In bioinformatics, DNNs are utilized for

gene expression analysis, protein structure prediction, and disease classification, among other tasks. The deep and hierarchical structure of DNNs allows them to model complex biological systems effectively.

The abstract presents a case where a DNN was used to predict genes related to Non-Small Cell Lung Cancer (NSCLC). The DNN model was trained using a large dataset comprising gene expression profiles and clinical data from NSCLC patients. This enabled the model to learn intricate relationships between genes and NSCLC.

Employing its deep architecture, the DNN could automatically learn informative representations of the gene expression data, capturing local and global dependencies between genes and NSCLC. Further, by training on a sizable dataset, the DNN could generalize and accurately predict unseen data. The genes predicted to be related to NSCLC by the DNN provided insights into disease mechanisms and potential therapeutic targets.

In conclusion, DNNs, with their capability to process high-dimensional genomic data and identify meaningful patterns, have shown promise in predicting NSCLC-related genes. Their ability to model complex gene-disease relationships, courtesy of their deep architecture, can aid researchers in identifying new gene candidates and deepening their understanding of the genetics of NSCLC.

### DisGeNET and other Databases

DisGeNET is a comprehensive database widely used in genomics for its extensive compilation of gene-disease associations. It amalgamates data from diverse sources, such as scientific literature, curated databases, and genomic datasets, providing insights into the genetic underpinnings of various diseases[21].

Several other databases, such as OMIM (Online Mendelian Inheritance in Man), GAD (Genetic Association Database), and HGMD (Human Gene Mutation Database), also serve as valuable repositories for gene-disease association data. These databases aggregate datasets from genetic studies, clinical reports, and experimental data, enriching our understanding of gene-disease relationships.

The key advantages of gene-disease association databases include the following:

**Comprehensive Coverage:** They collate data from numerous sources, encompassing various diseases and genetic variants.

**Centralized Repository:** These databases provide easy access to a vast collection of gene-disease association data for analysis.

**Curation and Quality Control:** Many databases employ rigorous curation processes and quality checks to ensure data accuracy and reliability, with experts reviewing and annotating data from literature and experimental studies.

**Limitations to the databases:**

**Data Heterogeneity:** The data in these databases originate from various studies that use different methodologies and data formats. Integrating and standardizing such heterogeneous data is challenging.

**Incompleteness:** Publication bias, limitations of available data, and research constraints may lead to specific gene-disease associations needing to be more represented and included.

**Data Quality and Reliability:** Despite stringent quality checks, inaccuracies can occur. Researchers need to exercise caution and verify the original data sources.

**Lack of Functional Information:** While these databases highlight gene-disease associations, they often must elucidate the biological mechanisms driving these relationships.

In conclusion, databases like DisGeNET, OMIM, GAD, and HGMD provide a wealth of information on gene-disease associations. They enhance data coverage, centralization, curation, and integration from diverse sources. However, researchers should be mindful of data heterogeneity, potential gaps in the data, potential quality issues, and the limited functional information available [1,2,3,4].

### **Future Directions**

Lung cancer research and gene-disease association analysis continue to evolve, promising to enhance disease comprehension and improve clinical outcomes. This evolution is primarily driven by developments in several areas, including integrating multi-omics data, exploring novel technologies and methodologies, and the advent of personalized medicine.

**Integration of Multi-Omics Data:** Merging data from diverse domains of the 'omics' field – genomics, transcriptomics, epigenomics, proteomics, and metabolomics – can offer a more comprehensive understanding of lung cancer at the molecular level. By combining these datasets, researchers can uncover complex molecular interactions, identify novel biomarkers, and elucidate disease mechanisms, thereby significantly enhancing gene-disease association analyses.

**Exploration of Emerging Technologies and Methodologies:** Rapid technological advancements profoundly impact lung cancer research. Single-cell sequencing elucidates tumor cellular heterogeneity,

providing insights into the tumor microenvironment and cellular interactions. Spatial transcriptomics and imaging techniques contribute to identifying spatially regulated genes and therapeutic targets. Furthermore, AI and machine learning algorithms can detect complex patterns in large genomic datasets, underscoring their potential utility in this field.

**Implementation of Personalized Medicine:** Personalized medicine, with treatments tailored to an individual's genetic profile, disease characteristics, and clinical factors, is a rising paradigm. Targeted therapies and immunotherapies have already revolutionized cancer treatment. Utilizing genomic data in conjunction with patient demographics, treatment history, and lifestyle factors may aid in the formulation of personalized treatment plans. Machine learning algorithms can facilitate treatment selection, prognostic assessment, and patient stratification.

**Functional Annotation and Validation:** The volume of data generated by gene-disease association studies necessitates functional annotation and validation of the identified genes. Experimental models (in vitro and in vivo), functional genomics, and high-throughput screening can elucidate the biological functions of lung cancer genes. These studies can establish causality, clarify molecular mechanisms, and identify druggable targets.

**Big Data and Collaborative Efforts:** The integration and analysis of big data, sourced from electronic health records, population-based cohorts, and multi-center studies, can unveil novel gene-disease associations and population-specific patterns. Collaboration and data sharing foster large-scale analyses, enhancing statistical power and producing more robust findings.

In conclusion, the future of lung cancer research is bright, with advancements in multi-omics data integration, emerging technologies, personalized medicine, functional annotation, and collaboration promising to catalyze discoveries. These developments could improve diagnostics, treatments, and patient outcomes.

### **3. Conclusion**

This survey paper highlights the significant contributions of gene-disease association analysis in lung cancer research, driven by machine learning and deep learning techniques. These methodologies have revolutionized the interpretation of vast genomic data, enabling the identification of biomarkers, disease subtypes, and therapeutic targets. Challenges remain in identifying causal variants and validating their functional relevance. Next-generation sequencing technologies have facilitated comprehensive genomic profiling but require robust bioinformatics pipelines and data management solutions.

Machine learning and deep learning have transformed bioinformatics, paving the way for personalized medicine predictive models. Gene-disease association databases play a crucial role, necessitating data quality, coverage, and biases considerations. Our understanding of the molecular basis of lung cancer has significantly improved through the integration of multi-omics data, advanced computational methods, and innovative technologies. These advancements hold great promise for targeted therapies, personalized medicine, and improved patient outcomes in lung cancer diagnosis, prognosis, and treatment.

## References:

- [1] Knox, S. S. (2010, April 26). *From "omics" to complex disease: a systems biology approach to gene-environment interactions in cancer*. PubMed Central (PMC). <https://doi.org/10.1186/1475-2867-10-11>
- [2] Thandra, K. C., Barsouk, A., Saginala, K., Aluru, J. S., & Barsouk, A. (2021, February 23). *Epidemiology of lung cancer*. PubMed Central (PMC). <https://doi.org/10.5114/wo.2021.103829>
- [3] Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (n.d.). *Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship*. PubMed Central (PMC). <https://doi.org/10.4065/83.5.584>
- [4] *Machine learning for multi-omics data integration in cancer*. (2022, January 22). Machine Learning for Multi-omics Data Integration in Cancer - ScienceDirect. <https://doi.org/10.1016/j.isci.2022.103798>
- [5] Ruan, X., Ye, Y., Cheng, W., Xu, L., Huang, M., Chen, Y., Zhu, J., Lu, X., & Yan, F. (2022, May 6). *Multi-Omics Integrative Analysis of Lung Adenocarcinoma: An in silico Profiling for Precise Medicine*. Frontiers. <https://doi.org/10.3389/fmed.2022.894338>
- [6] Cano-Gamez, E., & Trynka, G. (2020, April 6). *From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases*. Frontiers. <https://doi.org/10.3389/fgene.2020.00424>
- [7] Pai, A. (2020, February 17). *CNN vs. RNN vs. ANN - Analyzing 3 Types of Neural Networks in Deep Learning*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/>
- [8] *A novel candidate disease gene prioritization method using deep graph convolutional networks and semi-supervised learning - PubMed*. (2022, October 14). PubMed. <https://doi.org/10.1186/s12859-022-04954-x>
- [9] Mathew, C. J., David, A. M., & Joy Mathew, C. M. (2020, December 11). *Artificial Intelligence and its future potential in lung cancer screening*. PubMed Central (PMC). <https://doi.org/10.17179/excli2020-3095>
- [10] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (n.d.). *Applications of machine learning in drug discovery and development*. PubMed Central (PMC). <https://doi.org/10.1038/s41573-019-0024-5>
- [11] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (n.d.). *Applications of machine learning in drug discovery and development*. PubMed Central (PMC). <https://doi.org/10.1038/s41573-019-0024-5>
- [12] *Genome, transcriptome, and proteome: the rise of omics data and their integration in biomedical sciences - PubMed*. (2018, March 1). PubMed. <https://doi.org/10.1093/bib/bbw114>
- [13] Krassowski, M., Das, V., Sahu, S. K., & Misra, B. B. (2020, November 20). *State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing*. Frontiers. <https://doi.org/10.3389/fgene.2020.610798>
- [14] *Translational bioinformatics and healthcare informatics: computational and ethical challenges - PubMed*. (2009, September 16). PubMed. <https://pubmed.ncbi.nlm.nih.gov/20169020/>
- [15] <https://academic.oup.com/nar/article/45/D1/D833/2290909>. (n.d.). <https://academic.oup.com/nar/article/45/D1/D833/2290909>
- [16] Witte, J. S. (n.d.). *Genome-Wide Association Studies and Beyond*. PubMed Central (PMC). <https://doi.org/10.1146/annurev.publhealth.012809.103723>
- [17] *Next Generation Sequencing - an overview | ScienceDirect Topics*. (n.d.). Next Generation Sequencing - an Overview | ScienceDirect Topics. <https://doi.org/10.1016/B978-0-12-801418-9.00002-0>
- [18] Abass, Y. A., & Adeshina, S. A. (n.d.). *Deep Learning Methodologies for Genomic Data Prediction: Review | Atlantis Press*. Deep Learning Methodologies for Genomic Data Prediction: Review | Atlantis Press. <https://doi.org/10.2991/jaims.d.210512.001>
- [19] Muzio, G., O'Bray, L., & Borgwardt, K. (2021, March 1). *Biological network analysis with deep*

- learning*. OUP Academic.  
<https://doi.org/10.1093/bib/bbaa257>
- [20] *State-of-the-art in artificial neural network applications: A survey*. (2018, November 23). State-of-the-art in Artificial Neural Network Applications: A Survey - ScienceDirect.  
<https://doi.org/10.1016/j.heliyon.2018.e00938>
- [21] Piñero, J., Bravo, L., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., & Furlong, L. I. (2016, October 19). *DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants*. PubMed Central (PMC). <https://doi.org/10.1093/nar/gkw943>

link.springer.com/chapter/10.1007/978-3-031-21385-4\_4

ICAIDS 2021: **Artificial Intelligence and Data Science** pp 41–51 | [Cite as](#)

Home > [Artificial Intelligence and Data Science](#) > Conference paper

## A Systematic Review on Autonomous Vehicle: Traffic Sign Detection and Drowsiness Detection

[Panna Lal Boda](#) & [Y. Ramadevi](#) 

Conference paper | [First Online: 14 December 2022](#)

**318** Accesses

Part of the [Communications in Computer and Information Science](#) book series (CCIS, volume 1673)

### Abstract

The automatic detection of traffic signs is necessary for assisted driving, autonomous driving, and driving safety. Traffic sign play a significant task for advanced driver assistance systems (ADAS) also for autonomous driving vehicles and also driver drowsiness detection are an important part. Due to fatigue and drowsiness of the drivers, each day more number of fatalities and deaths are massively increases. In order to avoid these problems, developed a traffic signs detection and drowsiness detection based on machine learning and deep learning techniques. Histogram of Oriented Gradients (HOG), Adaptive Momentum Estimation (ADAM) optimizer features, Random Forest (RF), Region-based Convolutional Neural Network (R-CNN), Long Short Term Memory (LSTM), and Support Vector Machine (SVM) method are used. German Traffic Sign Detection Benchmarks (GTSDB) dataset is used for classification and

**Access via your institution** →

Chapter EUR 29.95  
Price includes VAT (India)

- Available as PDF
- Read on any device
- Instant download
- Own it forever

**Buy Chapter**

eBook EUR 74.89

Softcover Book EUR 89.99

Tax calculation will be finalised at checkout

Purchases are for personal use only  
[Learn about institutional subscriptions](#)

**Sections** **References**

[Abstract](#)

[References](#)

33°C Mostly sunny

Search

ENG IN

16:40 19-02-2024

PHYSICAL REVIEW B (/PRB/)  
covering condensed matter and materials physics

[Highlights \(/prb/highlights\)](#)  
 [Recent \(/prb/recent\)](#)  
 [Accepted \(/prb/accepted\)](#)  
 [Collections \(/prb/collections\)](#)  
 [Authors \(/prb/authors\)](#)  
[Referees \(/prb/referees\)](#)  
[Search \(/search\)](#)  
[Press \(/press\)](#)  
[About \(/prb/about\)](#)  
[Editorial Team \(/prb/staff\)](#)  
[RSS \(/feeds\)](#)

Editors' Suggestion   Letter

## FeRhCrSi: Spin semimetal with spin-valve behavior at room temperature

Y. Venkateswara, Jadupati Nag, S. Shanmukharao Samatham, Akhilesh Kumar Patel, P. D. Babu, Manoj Raama Varma, Jayita Nayak, K. G. Suresh, and Aftab Alam  
Phys. Rev. B **107**, L100401 – Published 2 March 2023



[https://www.altmetric.com/details.php?domain=journals.aps.org&citation\\_id=143063347](https://www.altmetric.com/details.php?domain=journals.aps.org&citation_id=143063347)

More

Article

[PDF \(/prb/pdf/10.1103/PhysRevB.107.L100401\)](#)  
[HTML \(/prb/abstract/10.1103/PhysRevB.107.L100401#fulltext\)](#)

[Export Citation \(/prb/export/10.1103/PhysRevB.107.L100401\)](#)



**ABSTRACT**

**AUTHORS**

**ARTICLE TEXT**

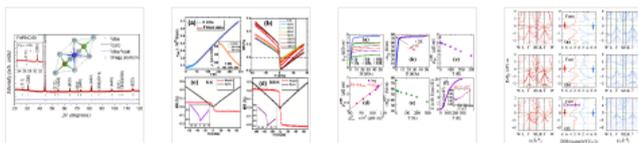
**ACKNOWLEDGMENTS**

**SUPPLEMENTAL MATERIAL**

**REFERENCES**

### ABSTRACT

Spin semimetals are a recently discovered new class of spintronic materials, which exhibit a band gap in one spin channel while exhibiting a semimetallic feature in the other, thus allowing for tunable spin transport. Here, we present experimental verification of spin semimetallic behavior in FeRhCrSi, a quaternary Heusler alloy with saturation moment  $2 \mu_B$  and Curie temperature  $> 400$  K. It crystallizes in the  $L2_1$  structure with 50% antisite disorder between Fe and Rh. Below 300 K, it shows a weakly temperature-dependent electrical resistivity with a negative temperature coefficient, indicating normal semimetal or spin semimetal behavior. Anomalous magnetoresistance data reveal the dominant contribution from the asymmetric part, a clear signature of the spin-valve nature, which is retained even at room temperature. The asymmetric part of the magnetoresistance shows an unusual increase with increasing temperature. Hall measurements confirm the anomalous nature of the conductivity originating from the intrinsic Berry curvature, with holes being the majority carriers. *Ab initio* simulation confirms a unique long-range ferrimagnetic ordering to be the ground state, explaining the origin behind the unexpected low saturation moment. The ferrimagnetic disordered structure confirms the spin semimetallic feature of FeRhCrSi, as observed experimentally.



Received 20 November 2022   Accepted 13 February 2023

DOI: <https://doi.org/10.1103/PhysRevB.107.L100401>

©2023 American Physical Society

Physics Subject Headings (PhySH)

Research Areas

This site uses cookies. To find out more, read our [Privacy Policy \(https://www.aps.org/about/privacy.cfm\)](https://www.aps.org/about/privacy.cfm).

I Agree ()





Reuse & Permissions (<https://powerxeditor.aptaracorp.com/sciprisaps/RnPRErequest/submit?ArticleTitle=FeRhCrSi%3A+Spin+semimetal+with+spin-valve+behavior+at+room+temperature&AuthorName=Y.+Venkateswara+et+al.&JournalCode=PRB&contentid=10.1103%2FPhysRevB.107.L100401>)

### Access Options

[Buy Article » \(/cart/add/10.1103/PhysRevB.107.L100401\)](#)

[Log in with individual APS Journal Account » \(https://journals.aps.org/login\)](#)

[Log in with a username/password provided by your institution » \(/login\\_inst\\_user?rt=https%3A%2F%2Fjournals.aps.org%2Fprb%2Fabstract%2F10.1103%2FPhysRevB.107.L100401\)](#)

[Get access through a U.S. public or high school library » \(/free-access-for-us-public-and-high-school-libraries\)](#)

PRX LIFE 

First content now online

[\(/prxlife/?\)](https://prxlife/)

[utm\\_source=prb&utm\\_medium=web&utm\\_campaign=prxlife\)](#)

PRX ENERGY 

Learn More

[\(/prxenergy/?utm\\_source=prb&utm\\_medium=web&utm\\_campaign=prxenergy\)](https://prxenergy/?utm_source=prb&utm_medium=web&utm_campaign=prxenergy)

Translate to high-quality English

Edit to perfection

Tell your story to the world

Trust our highly experienced editors and science communicators to deliver a great product, always!

APS powered by Editage

Get Quote

[\(/https://authorservices.aps.org/?utm\\_source=physicalreviewjournals&utm\\_medium=referral\)](https://authorservices.aps.org/?utm_source=physicalreviewjournals&utm_medium=referral)

This site uses cookies. To find out more, read our [Privacy Policy \(https://www.aps.org/about/privacy.cfm\)](https://www.aps.org/about/privacy.cfm).

I Agree ()

Sign up to receive regular email alerts from *Physical Review B*

**Sign up** (<https://info.aps.org/journals-emails>)

[APS \(https://www.aps.org/\)](https://www.aps.org/) | [Earlier Issues \(prb/issues\)](https://prb.aps.org/earlier-issues) | [News & Announcements \(prb/edannounce\)](https://prb.aps.org/news-announcements) | [About this Journal \(prb/about\)](https://prb.aps.org/about) | [Editorial Team \(prb/staff\)](https://prb.aps.org/staff) | [About the Journals \(/about\)](https://www.facebook.com/apsphysics) | [Join APS \(https://www.facebook.com/apsphysics\)](https://www.facebook.com/apsphysics) | [Join APS \(https://www.aps.org/membership/join.cfm\)](https://www.facebook.com/apsphysics) | [https://www.facebook.com/apsphysics](https://twitter.com/APSphysics) | <https://twitter.com/APSphysics>

**AUTHORS**

[General Information \(prb/authors\)](https://prb.aps.org/authors)  
[Submit a Manuscript \(https://authors.aps.org/submissions/\)](https://authors.aps.org/submissions)  
[Publication Rights \(pub\\_rights.html\)](https://prb.aps.org/publication-rights)  
[Open Access \(open\\_access.html\)](https://prb.aps.org/open-access)  
[Policies & Practices \(authors/editorial-policies\)](https://prb.aps.org/policies-practices)  
[Tips for Authors \(authors/tips-authors-physical-review-physical-review-letters\)](https://prb.aps.org/tips-for-authors)  
[Professional Conduct \(authors/professional-conduct-ethics\)](https://prb.aps.org/professional-conduct)

**REFEREES**

[General Information \(prb/referees\)](https://prb.aps.org/referees)  
[Submit a Report \(http://referees.aps.org/\)](http://referees.aps.org/)  
[Update Your Information \(http://referees.aps.org/\)](http://referees.aps.org/)  
[Policies & Practices \(authors/editorial-policies\)](https://prb.aps.org/policies-practices)  
[Referee FAQ \(referees/faq.html\)](https://prb.aps.org/referee-faq)  
[Guidelines for Referees \(prb/referees/advice-referees-physical-review\)](https://prb.aps.org/referees-advice)  
[Outstanding Referees \(OutstandingReferees\)](https://prb.aps.org/outstanding-referees)

**LIBRARIANS**

[General Information \(https://librarians.aps.org/\)](https://librarians.aps.org/)  
[Subscriptions \(https://librarians.aps.org/subscriptions\)](https://librarians.aps.org/subscriptions)  
[Online License Agreement \(https://librarians.aps.org/sitelicense.pdf\)](https://librarians.aps.org/sitelicense.pdf)  
[Usage Statistics \(https://librarians.aps.org/login\)](https://librarians.aps.org/login)  
[Your Account \(https://librarians.aps.org/account\)](https://librarians.aps.org/account)

**STUDENTS**

[Physics \(https://physics.aps.org/\)](https://physics.aps.org/)  
[PhysicsCentral \(http://www.physicscentral.com/\)](http://www.physicscentral.com/)  
[Student Membership \(https://www.aps.org/membership/student.cfm\)](https://www.aps.org/membership/student.cfm)

**APS MEMBERS**

[Subscriptions \(https://www.aps.org/membership/aps-publications.cfm\)](https://www.aps.org/membership/aps-publications.cfm)  
[Article Packs \(https://journals.aps.org/article-packs\)](https://journals.aps.org/article-packs)  
[Membership \(https://www.aps.org/membership/index.cfm\)](https://www.aps.org/membership/index.cfm)  
[FAQ \(https://www.aps.org/membership/faq.cfm\)](https://www.aps.org/membership/faq.cfm)  
[APS News \(https://www.aps.org/publications/apsnews/index.cfm\)](https://www.aps.org/publications/apsnews/index.cfm)  
[Meetings & Events \(https://www.aps.org/meetings/index.cfm\)](https://www.aps.org/meetings/index.cfm)

[Privacy \(https://www.aps.org/about/webpolicies.cfm#privacy\)](https://www.aps.org/about/webpolicies.cfm#privacy) | [Policies \(/policies\)](https://www.aps.org/policies) | [Contact Information \(/contact.html\)](https://www.aps.org/contact.html) | [Feedback \(mailto:feedback@aps.org\)](mailto:feedback@aps.org)

ISSN 2469-9969 (online), 2469-9950 (print). ©2024 American Physical Society (<https://www.aps.org/>) All rights reserved. *Physical Review B*<sup>™</sup> is a trademark of the American Physical Society, registered in the United States, Canada, European Union, and Japan. The *APS Physics logo* and *Physics logo* are trademarks of the American Physical Society. Information about registration may be found [here \(/legal\)](#). Use of the American Physical Society websites and journals implies that the user has read and agrees to our [Terms and Conditions \(/info/terms.html\)](https://www.aps.org/info/terms.html), and any applicable [Subscription Agreement \(https://librarians.aps.org/sitelicense.pdf\)](https://librarians.aps.org/sitelicense.pdf).

This site uses cookies. To find out more, read our [Privacy Policy \(https://www.aps.org/about/privacy.cfm\)](https://www.aps.org/about/privacy.cfm).

I Agree ()



# Substitution driven ground states of Fe<sub>1-x</sub>Cr<sub>x</sub>Si: A resistivity study

Sankararao Yadam<sup>a,e</sup>  , S. Shanmukharao Samatham<sup>b,1</sup>, Raghavendra Kulkarni<sup>a</sup>, D. Venkateshwarlu<sup>c</sup>, V. Ganesan<sup>d,e</sup>

Show more 

 Share  Cite

<https://doi.org/10.1016/j.cryogenics.2023.103683> 

[Get rights and content](#) 

## Highlights

- Composition-temperature ( $x$ - $T$ ) phase diagram represents the manifestation of the ground states of Fe<sub>1-x</sub>Cr<sub>x</sub>Si.
- Kondo-insulating behavior at low temperatures and activated behavior at high temperature are noticed for the compositions below the critical concentration  $x_{c,MIT} \sim 0.02$ .
- Gradual reduction of low temperature insulating phase is evident from the enhancement of metallicity for  $x > x_{c,MIT}$ .
- The low temperature upturn of the resistivity in these alloys is attributed to the localization effects assisted magnetic correlations. Nevertheless, the alloys with  $x \geq 0.5$  exhibit a pseudo gap semimetal behavior.

## Abstract

We report the electrical resistivity of Fe<sub>1-x</sub>Cr<sub>x</sub>Si with nominal compositions of  $x = 0.0, 0.005, 0.025, 0.075, 0.1, 0.125, 0.15, 0.2, 0.5, 0.75$  and 1. Composition-temperature ( $x$ - $T$ ) phase diagram is constructed based on the variation of electrical properties with substitution and temperature. The compositions below the critical concentration of metal-to-insulator transition (i.e.,  $x_c \sim 0.02$ ) are noticed to exhibit Kondo behaviour at low temperatures and semiconducting behaviour (due to activated charge carriers) at high temperature. Above  $x_c$ , a gradual reduction of low temperature insulating phase is compensated by the enhancement of metallic phase. Additionally, the high temperature semiconducting phase is noticed to reduce with increasing  $x$ . For the compositions  $x_c \leq x < 0.5$ , the low temperature rise of electrical resistivity with the reduction of temperature is attributed to the magnetic correlations aided by localization effects. The two-band (metallic and semiconducting parallel conductive channels) analysis reveals the formation of impurity states and reduction of energy gap. Nevertheless, the compositions  $x \geq 0.5$  exhibit pseudo-gap behavior. The energy band gap of CrSi is found to be  $0.320 \pm 0.019$  meV with a residual resistivity of  $4.773 \mu\Omega\text{-cm}$ .

## Introduction

The investigation of the consequences of electronic and magnetic interactions among different degrees of freedom has been gaining interest in the recent years to predict the various electrical and magnetic ground state properties of materials. Belonging to this category, metal-to-insulator (MI) phenomenon in materials has become fascinating because of its various physical origins. There have been explanations for the insulators such as Mott, Mott-Hubbard, Anderson, charge-transfer-insulators etc. [1]. In the recent past, the research on narrow band gap insulators/semiconductors, with an energy band gap of about few tens of meV, has been intensified due to ease of tuning the ground state properties by the effect of non-thermal external parameters. Generally, the physical and magnetic properties of compositions with MI transition are sensitive to the nature of the substituting element (electron/hole concentration and non-magnetic/magnetic). FeSi, with a band gap of about 50–60meV [2], [3], is one such attention-grabbing alloy whose electrical and magnetic properties are strongly electron-correlation dependent. It has been reported to exhibit unusual magnetic properties [4], [5] and high thermoelectric power [6], [7], [8], [9]. Recently, the low energy electronic structural studies of FeSi revealed nonexistence of surface Fermi arcs joining the many fold degenerate bulk band crossing points near the Fermi level in spite it is being chiral topological system [9].

The magnetic and electrical resistivity of FeSi were examined under the influence of magnetic field  $H$ . A negative magnetoresistance at reduced temperatures was reported in this weakly disordered system FeSi, arising out of weak localization [2]. However, a partial replacement of Fe by other transition metals and Si by Ge/Al was substantially affected the electrical and magnetic properties of FeSi. For example, substitution of non-magnetic metallic element Al resulted in a Kondo insulator to heavy Fermion metal transition in FeSi<sub>1-x</sub>Al<sub>x</sub> [6], [10], the substitution of non-magnetic semiconducting element Ge resulted in a ferromagnetic metallic state in FeSi<sub>1-x</sub>Ge<sub>x</sub> [11], [12]. In particular, MI transition in Al-doped FeSi has been reported to arise due to impurity-band-conduction [6]. In FeSi, the substitution and magnetic field-tuned exotic ground states such as field-induced heavy Fermion, Fermi liquid and ferromagnetism have been reported for the concentrations above MI transition (hereafter MIT concentration is referred to as  $x_c$ ) [6], [10], [13], [14], [15], [16].

For cobalt substituted FeSi, i.e. Fe<sub>1-x</sub>Co<sub>x</sub>Si, the onset of helimagnetic state was reported for the compositions  $x > x_{c,mag}$  (critical concentration for nonmagnetic to long-range magnetic state), in addition to MI transition at  $x_c \sim 0.018$ . Further, the positive magnetoresistance below the magnetic transition temperature ( $T_c$ ) in the magnetic compositions of Fe<sub>1-x</sub>Co<sub>x</sub>Si has created a renewed interest. Under the influence of external hydrostatic pressure, although FeSi exhibits unusual metallic state which is neither Fermi-liquid nor non-Fermi-liquid [17], substituted-FeSi has been reported to show quantum critical phase transition [18]. Our previous studies of Fe<sub>1-x</sub>Cr<sub>x</sub>Si revealed the emergence of low temperature-metallic behavior and reduction of energy band gap with increasing  $x$  while causing MI transition at  $x_c \sim 0.02$  [19]. The low temperature metallic and high temperature semiconducting natures of Fe<sub>1-x</sub>Cr<sub>x</sub>Si were explained by the impurity band and excitation of charge carriers respectively using two-band conduction mechanism. Though doping dependent caloric and electrical properties of Fe<sub>1-x</sub>Cr<sub>x</sub>Si (up to a substitution level of  $x = 0.15$ ) have been studied, the actual growth of the metallic phase at higher concentrations (up to  $x = 1$ ) has not been addressed. Recently, it was reported theoretically that metallic states found in the band gap of CrSi gives rise to the topological surfaces [20]. It has been also reported that CrSi exhibits Dzyaloshinskii–Moriya interaction and large positive magnetoresistance qualifying CrSi as a potential candidate for both thermoelectric and spintronic applications [21]. Principally, FeSi ( $x=0$ ) is a Kondo insulator with correlated unconventional semiconducting behavior and CrSi ( $x=1$ ) is an unconventional metal with electronic topological transition. Nevertheless, the manifestation of physical properties, the electrical resistivity in particular, of the intermediate compositions is not comprehensively studied.

In this paper, we report a comprehensive study of the electrical ground state properties of Fe<sub>1-x</sub>Cr<sub>x</sub>Si by taking inferences from the composition variation of electrical resistivity in the absence of magnetic fields. A phase diagram of composition versus temperature ( $x$ - $T$ ) is established in which the changes in the electrical ground state properties of Fe<sub>1-x</sub>Cr<sub>x</sub>Si are assigned in the respective  $x$ - $T$  segments.

## Section snippets

### Experimental methods

The polycrystalline alloys of Fe<sub>1-x</sub>Cr<sub>x</sub>Si were synthesized by arc-melting method. The constituent elements (Fe, Cr and Si) with high purity were taken in the required stoichiometric ratio and melted congruently. The as-cast ingots were vacuum sealed at 10<sup>-5</sup> mbar and annealed at 980 °C for a week. For the present study, we have performed measurements on Fe<sub>1-x</sub>Cr<sub>x</sub>Si with nominal compositions of  $x=0.20, 0.50, 0.75$  and 1. In addition, we have used the data of our previously reported samples  $x=0.0, 0.2, 0.5, 0.75$  and 1.

## Results and discussion

X-ray diffraction pattern of  $x=0.20, 0.50, 0.75$  and 1 are shown in the Fig. 1(a). The patterns were refined using FullProf suite [22] and the samples were found to crystallize in B20 cubic structure with P2<sub>1</sub>3 space group. The x-ray diffraction pattern of  $x=0.0, 0.2, 0.5, 0.75$  and 1 along with the Rietveld refinement is shown in the Fig. 1(a). Lattice parameter and goodness of fit are shown in Table 1. The silicon, being a semiconductor, is expected to show large (of the order of MΩ-cm)...

## Summary

Summarizing, Fe<sub>1-x</sub>Cr<sub>x</sub>Si with nominal compositions of  $x=0$  to 1 has been studied using x-ray diffraction and electrical resistivity measurements. The lattice parameter increases linearly with Cr-substitution, following Vegard's law, by indicating that iron site is successfully replaced by chromium. The electrical resistivity of Fe<sub>1-x</sub>Cr<sub>x</sub>Si is investigated for  $x=0.0$  to 1 in view of unveiling the changes in the electrical ground states as a function of  $x$  and consequently a  $x$ - $T$  phase diagram is ...

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper...

## Acknowledgement

Er. P. Saravanan and his group from UGC-DAE CSR, Indore M. P. are thanked for their help in providing Liquid He and Nitrogen. The author Sankararao Yadam acknowledges CRS, TEQIP-III JNTUH with Procs No. JNTUH/TEQIP-III/CRS/2019/Physics/05 for the financial support. Dr. C. V. Raghava Reddy, Prof. G. Bhikshamaiah and Prof. Narasimhacharyulu of CVR College of Engineering thanked for their encouragement and support. The author SSS acknowledges Science and Engineering Research Board (SERB, Govt. of...

[Recommended articles](#)

---

## References (31)

- P. Samuely *et al.*  
Phys B (1996)
- S. Yadam *et al.*  
J Alloys Compounds (2016)
- J. Rodriguez-Carvajal  
Phys B (1993)
- Z. Schlesinger *et al.*  
Phys B Condens Matter (1997)
- M. Imada *et al.*  
Rev Mod Phys (1998)

S. Paschen *et al.*

Phys Rev B (1997)

B.C. Sales *et al.*

Phys Rev B (1994)

S. Krannich *et al.*

Nat Commun (2015)

L.S. Sharath Chandra *et al.*

J Phys Condens Matter (2008)

B. Buschinger *et al.*

Phys B (1997)



View more references

---

## Cited by (0)

---

1 ORCID ID: <https://orcid.org/0000-0002-1668-2304>

[View full text](#)

© 2023 Elsevier Ltd. All rights reserved.



All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

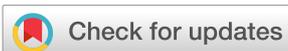


[Log in / register](#)

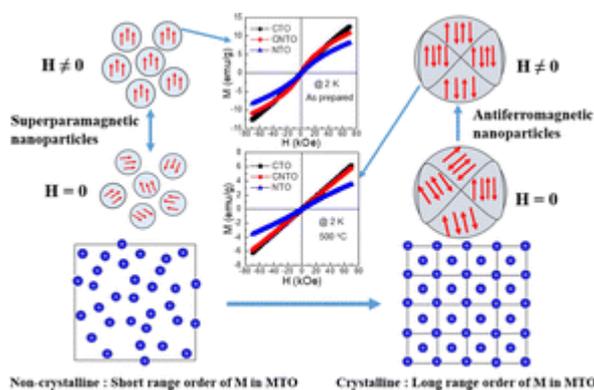
Issue 4, 2023

[Previous](#)[Next](#)

From the journal:

**Physical Chemistry Chemical Physics****Unveiling the correlation between structural and magnetic ordering in nano Co<sub>1-x</sub>Ni<sub>x</sub>TeO<sub>4</sub>**[Akhilesh Kumar Patel](#), <sup>a</sup> [S. Shanmukharao Samatham](#), <sup>b</sup> [Ekta Rani](#), <sup>c</sup> [K. G. Suresh](#) <sup>a</sup> and [Harishchandra Singh](#) <sup>\*ac</sup>[Author affiliations](#)**Abstract**

Nanomaterials with unique structures and exotic magnetic phenomena are always intriguing; however, the direct correlation of structural and magnetic ordering up to a few nanometers remains critical. We report structural and magnetic properties of sol-gel grown Co<sub>1-x</sub>Ni<sub>x</sub>TeO<sub>4</sub> ( $x = 0, 0.5$  and  $1$ ) nanoparticles. An increase in the calcination temperature leads to the enhancement of the particle size and structural ordering. This is accompanied by changes in the magnetic interactions as well. Calcination at lower temperatures retains the short-range non-crystalline structure and superparamagnetic behavior, while calcination at higher temperatures results in long-range ordering in both the crystal and magnetic structures. Superparamagnetic to antiferromagnetic ordering observed from temperature- and field-dependent magnetization is attributed to the changes in structural ordering. This study presents a new family of nanomaterials displaying stable magnetic order up to  $\sim 6$  nm, where the magnetic properties can be uniquely controlled by changing the structural ordering.



## Introduction

Technological advances for subsequent generation devices have escalated the research at the nanoscale as an unprecedented change in nanostructures' structural, electronic, and vibration properties has been evidenced due to the quantum confinement effect and/or increase in surface to volume ratio.<sup>1-3</sup> Depending on their properties, such nanostructures have been found to be suitable for different applications.<sup>2,4</sup> Interest in magnetic nanostructures has been boosted in the past few decades by virtue of their potential in diverse fields, such as ultrahigh-density recording,<sup>5</sup> medical diagnostics,<sup>6</sup> catalysis,<sup>7</sup> and sensors.<sup>8</sup> Apart from the properties mentioned above, the magnetic behavior of the nanostructures also differs from that of their bulk counterparts. This is because size reduction to a nanometric scale strongly rearranges the magnetic moment/domain. Thus, exploring magnetic materials of varying sizes is essential and can reveal novel promising applications.

The manifestation of magnetic properties in nanostructures has been reported for several systems, such as GdAl<sub>2</sub>,<sup>9</sup> Ni(OH)<sub>2</sub>,<sup>10</sup> La<sub>1-x</sub>Sr<sub>x</sub>CoO<sub>3</sub>,<sup>11</sup> and perovskite manganite R<sub>1-x</sub>A<sub>x</sub>MnO<sub>3</sub> (R = La, Pr, Sm *etc.*, and A = Ca, Sr, Ba, and Pb).<sup>12,13</sup> For example, GdAl<sub>2</sub> nanocrystallites displayed broadening in the ferromagnetic (FM) transition, rather than shift, with a decrease in nanocrystallite size along with superparamagnetic behavior for smaller nanocrystallites (~18 nm) below 30 K.<sup>9</sup> Ni(OH)<sub>2</sub> nanoparticles of average size ~8 nm, on the other hand, displayed paramagnetic behavior at higher temperatures and a paramagnet to ferromagnet transition with a decrease in temperature.<sup>10</sup> Moreover, for one of the most studied and applied compounds, Fe<sub>2</sub>O<sub>3</sub>, both coercivity and remanence were found to approach zero at a particle size of ~19.3 nm.<sup>3</sup> Though the size-dependent magnetic interaction has already been investigated, a direct correlation among particle size, crystallographic structural ordering, and magnetic properties is lacking.

Recently, nanostructured transition metal tellurates/tellurides have come into the limelight as potential candidates for energy applications.<sup>14,15</sup> Moreover, cobalt-based tellurates/tellurides display intriguing magnetic behavior, such that CoTeO<sub>4</sub>,<sup>14</sup> CoTe,<sup>16</sup> Co<sub>2</sub>Te<sub>3</sub>O<sub>8</sub>,<sup>17</sup> and Co<sub>3</sub>TeO<sub>6</sub><sup>18,19</sup> either show weak FM or antiferromagnetic (AFM) transition in polycrystalline form. However, the effect of

calcination temperature on the crystallization and the particle size-dependent magnetic properties of telluride nanoparticles has not yet been studied. Therefore, from the application point of view, it is interesting to take up a system, such as CoTeO<sub>4</sub>, and study the effects of substitutions and preparation conditions, such as calcination temperature, on the particle size and magnetic field properties. In the present work, sol-gel-grown Co<sub>1-x</sub>Ni<sub>x</sub>TeO<sub>4</sub> ( $x = 0, 0.5, \text{ and } 1$ ) nanoparticles with different sizes are studied to investigate their structural and magnetic properties. Following detailed microstructural and structural analysis, the correlation of structural ordering and magnetic interactions is explored with varying calcination temperatures.

## Experimental

### Synthesis

Co<sub>1-x</sub>Ni<sub>x</sub>TeO<sub>4</sub> ( $x = 0, 0.5 \text{ and } 1$ ) compounds were prepared by a sol-gel method using the initial compounds of Co(NO<sub>3</sub>)<sub>2</sub>·6H<sub>2</sub>O, Ni(NO<sub>3</sub>)<sub>2</sub>·6H<sub>2</sub>O, and H<sub>2</sub>O<sub>4</sub>Te·2H<sub>2</sub>O. The constituent compounds were taken and mixed in distilled water using a magnetic stirrer for 3 h. Post stirring, each solution was dried on a hot plate at 80 °C. The dried powder of each composition was divided into three parts and kept for calcination at different temperatures for different durations (Table S1, ESI †). Hereafter, CoTeO<sub>4</sub>, Co<sub>0.5</sub>Ni<sub>0.5</sub>TeO<sub>4</sub>, and NiTeO<sub>4</sub>, (MTO; M = Co, Co/Ni, and Ni), are referred to as CTO, CNTO, and NTO, respectively.

### Instrumentation

The room-temperature X-ray diffraction (XRD) patterns were recorded on well-ground specimens using a PANalytical X-ray diffractometer. The scanning electron microscopy (SEM) and energy dispersive spectroscopy (EDS) data were collected using a Field Emission Gun FEG-SEM (JEOL JSM-7600F FEG-SEM). Transmission electron microscopy (TEM) measurements were carried out using a JEOL JEM2200FS EFTEM/STEM instrument. Selected area electron diffraction (SAED) was performed by FEG-TEM. Magnetic moments *versus* temperature and magnetic field were measured using a SQUID vibrating sample magnetometer (VSM). The magnetic measurements were carried out in zero-field cooling (ZFC) and field-cooling warming (FCW) processes. In the ZFC process, the specimen was cooled down to 2 K (from 300 K) in the absence of a magnetic field and then a field was applied. The data were collected during warming. In the FCW process, the specimen was cooled in the presence of a magnetic field and the data were collected during warming without switching off the applied magnetic field.

## Results and discussion

The room temperature XRD patterns of the pristine (as-prepared) MTO and those calcined at 400 and 500 °C are shown in [Fig. 1\(a–c\)](#). The non-crystalline nature of the pristine and low-temperature calcined (400 °C) MTOs is evident from the broad peaks of the patterns. However, crystallinity is found to enhance as the temperature is increased to 500 °C. This observation corroborates a recent study, wherein a single crystalline phase is obtained at a calcination temperature of 600 °C.<sup>14</sup> Thus, the transformation of short-range structural ordering (*i.e.*, non-crystalline structure) to long-range structural ordering (crystalline structure) is observed with the increase in calcination temperature. EDS spectra of each compound at different calcination temperatures (Fig. S1(a–i), ESI †) and corresponding atomic percentages (Table S2, ESI †) suggest the elemental homogeneity of the grown nanomaterials.<sup>20</sup> Fig. S2(a–i) (ESI †) shows the high-resolution SEM images for pristine  $\text{Co}_{1-x}\text{Ni}_x\text{TeO}_4$  ( $x = 0, 0.5$  and 1) compounds along with their calcined counterparts. The morphology of these nanostructures is found to be spherical in nature. An increase in nanoparticle size and its crystallinity with the increase in calcination temperature is evident. Moreover, the SEM images of the pristine compounds (Fig. S2(a, d and g), ESI †) show that nanoparticles are not well separated, whereas these nanoparticles start to separate with the calcination temperature (Fig. S2(c, f and i), ESI †). The TEM-EDS mapping ([Fig. 2](#)) for pristine and calcined CTO nanoparticles further shows the homogeneous distribution of Co, Te, and O elements. The calculated elemental atomic percentages also follow the elemental atomic percentages as found from SEM-EDS. The particle sizes of the CTO compounds, as inferred from TEM (Fig. S3, ESI †), show an average size of ~3 nm for the pristine compound, whereas an increase in average size from ~10 to ~25 nm with the increase in the calcination temperature from 400 °C to 500 °C is noted. The average particle sizes for all compounds are given in Table S3 (ESI †).



**Fig. 1** The room temperature X-ray diffraction patterns of pristine and calcined (a) CTO, (b) CNTO, and (c) NTO showing increases in the crystallinity as the calcination temperature is increased.



**Fig. 2** The TEM-EDS mapping of pristine and calcined (400 °C and 500 °C) CTO.

The magnetization as a function of temperature ( $M-T$ ) in 100 Oe is measured for the pristine nanoparticles and their calcined counterparts.  $M-T$  plots of CTO are shown in [Fig. 3\(a–c\)](#). [Fig. 3\(a\)](#) shows magnetization behavior of the pristine compound, wherein no signatures of magnetic ordering are found down to 5 K. To estimate the possible existing magnetic correlations, the inverse susceptibility of the same is fitted using the Curie–Weiss law,

$$\chi = \chi_0 + C/(T - \theta_{CW}) \quad (1)$$

where  $\chi_0$  is the independent magnetic susceptibility,  $C$  is the Curie constant, and  $\theta_{CW}$  is the Curie–Weiss temperature. The fit is shown in the inset of [Fig. 3\(a\)](#). The fit resulted in a paramagnetic Curie–Weiss temperature of  $-37.6$  K and an effective magnetic moment of  $4.1\mu_B$ . Although the pristine compound does not seem to order magnetically down to 5 K, a negative  $\theta_{CW}$  ( $= -37.6$  K) signifies AFM correlations of the spins. Despite magnetic correlations, the magnetic order might be prevented by the lack of atomic ordering in their respective atomic/Wyckoff positions due to the non-crystalline nature of the pristine compound. As the calcination temperature increases, the negative  $\theta_{CW}$  is found to increase from  $-36.7$  K (pristine) to  $-58.4$  K (for  $400^\circ\text{C}$ ). An increase in negative  $\theta_{CW}$  is understood to correlate with the crystallinity enhancement at elevated calcination temperatures. This indicates that the magnetic ordering is stabilized with the increase in crystallinity. The compound calcined at  $400^\circ\text{C}$  ([Fig. 3\(b\)](#)) exhibits a sharp peak at  $8.9$  K, at which the bifurcation of ZFC and FCW also onsets. Inferring from the negative paramagnetic Curie–Weiss temperature, the observed transition results from the AFM behavior with a Néel temperature of  $T_N = 8.9$  K. A further increase in calcination temperature has led to an enhanced  $T_N$  from  $8.9$  K ( $400^\circ\text{C}$ ) to  $43.5$  K ( $500^\circ\text{C}$ ). Such an enhancement is supported by the increased  $\theta_{CW}$ , indicating strong AFM interactions in the nanoparticles. For CNTO, the pristine compound exhibits a negative  $\theta_{CW}$  but does not order magnetically down to 5 K ([Fig. 3\(d\)](#)). However, it exhibits AFM order below  $39.1$  K when calcined at  $400^\circ\text{C}$  ([Fig. 3\(e\)](#)) and  $42.7$  K when calcined at  $500^\circ\text{C}$  ([Fig. 3\(f\)](#)). Similarly, the pristine NTO does not order magnetically down to 5 K ([Fig. 3\(g\)](#)), whereas for calcined specimens,  $T_N$  is found to be  $5.5$  K ( $400^\circ\text{C}$ : [Fig. 3\(h\)](#)) and  $6.1$  K ( $500^\circ\text{C}$ : [Fig. 3\(i\)](#)). The particulars of the calcination temperatures,  $T_N$  and  $\theta_{CW}$ , and the effective magnetic moments are detailed in [Table 1](#). The observed variation in the effective magnetic moments of each CTO compound could be related to the crystalline nature of these compounds.

**Fig. 3** (a) Temperature-dependent magnetization at 100 Oe for (a–c) CTO, (d–f) CNTO and (g–i) NTO for pristine and calcined ( $400^\circ\text{C}$  and  $500^\circ\text{C}$ ) compounds along with the corresponding modified Curie–Weiss fitting at 100 Oe ZFC (insets).

**Table 1** Magnetic transition temperature ( $T_N$ ), Curie–Weiss temperature ( $\theta_{CW}$ ), and effective magnetic moments ( $\mu_{\text{eff}}$ ) of CTO, CNTO and NTO compounds at different calcination temperatures

Calcination temperature	$T_N$ (K)	$\theta_{CW}$ (K)	$\mu_{\text{eff}}$ ( $\mu_B$ f.u. <sup>-1</sup> )
CTO			

Calcination temperature	$T_N$ (K)	$\theta_{CW}$ (K)	$\mu_{eff}$ ( $\mu_B$ f.u. <sup>-1</sup> )
<b>CTO</b>			
Pristine	—	$-37.6 \pm 0.1$	$4.13 \pm 0.01$
400 °C	8.9	$-58.4 \pm 0.1$	$4.91 \pm 0.01$
500 °C	43.5	$-74.1 \pm 0.2$	$5.02 \pm 0.01$
<b>CNTO</b>			
Pristine	—	$-32.5 \pm 0.1$	$3.21 \pm 0.01$
400 °C	39.1	$-55.3 \pm 0.5$	$4.24 \pm 0.02$
500 °C	42.7	$-61.8 \pm 0.6$	$4.02 \pm 0.02$
<b>NTO</b>			
Pristine	—	$-34.4 \pm 0.2$	$2.84 \pm 0.01$
400 °C	5.5	$-62.7 \pm 0.5$	$3.32 \pm 0.02$
500 °C	6.1	$-125.2 \pm 0.7$	$3.41 \pm 0.02$

[Fig. 4](#) shows the isothermal magnetization *versus* magnetic fields ( $M-H$ ) measured at 2, 100, and 300 K. The measurements were carried out in the ZFC process.  $M-H$  curves were recorded in five quadrants (0 → 70 → 0 → -70 → 0 → 70 kOe) by sweeping the magnetic field. For the sake of clarity and to be precise, we discuss the first quadrant of  $M-H$  data. [Fig. 4\(a-c\)](#) shows the curves for CTO. Pristine CTO exhibits an almost non-linear behavior with a non-saturating magnetic moment of about 12.5 emu g<sup>-1</sup> under the influence of 70 kOe. However, in the paramagnetic state, *i.e.*, at 100 and 300 K, the  $M-H$  curves are linear. Though the 2 K isotherm is still non-linear for a 400 °C calcined CTO, it is linear for a 500 °C calcined compound indicating the AFM behavior. A quasi-non-linear behavior with excess non-saturating magnetic moment is attributed to the non-crystalline nature of the pristine and 400 °C calcined compounds. A similar nature is observed in the CNTO and NTO compounds, as shown in [Fig. 4\(d-f\) and \(g-i\)](#), respectively. The  $M-H$  data of MTO commonly reveals the stable magnetic order with the improvement of the crystallinity as the calcination temperature is increased.

**Fig. 4** Field-dependent magnetization at 2, 100 and 300 K for (a-c) CTO, (d-f) CNTO, and (g-i) NTO for pristine, 400 °C and 500 °C calcined compounds.

[Fig. 5](#) shows the schematic diagram of the morphological, structural, and magnetization behavior of MTO with calcination temperature, summarizing the transition of the magnetic interaction from

superparamagnetic to AFM because of the change in short-range structural ordering to long-range structural ordering. It is evident that the pristine nanoparticles are non-crystalline in nature, while the calcination at high temperature induces the long-range crystalline order as indicated by the horizontal arrow in [Fig. 5](#). Under  $H = 0$ , these nanoparticles show superparamagnetic behavior and the presence of a magnetic field tries to bring an overall magnetic order by orienting the spins of the individual nanoparticles in the field's direction. As a result, soft FM-like behavior is observed in the magnetization vs. magnetic field isotherms of these non-crystalline pristine nanoparticles of MTO. The origin of net magnetic moment in these compounds could be due to the uncompensated  $\text{Co}^{2+}/\text{Ni}^{2+}$  ions at the surface of the particles. However, as shown in [Fig. 4\(a, d and g\)](#), a linear  $M-H$  is due to paramagnetic behavior. A weak FM behavior at low temperatures and paramagnetic behavior at high temperatures in these nano-sized pristine compounds can be understood in terms of blocked spins (in the superparamagnetic state) overcoming the energy barrier using the thermal energy at high temperatures ( $\text{NiO}$ ,<sup>21</sup>  $\text{CuO}$ <sup>22</sup> and  $\text{MnO}$ <sup>23</sup>). In contrast, the calcinated crystalline nanoparticles of MTO show AFM order under  $H = 0$ . The strong AFM coupling is evident from the linear magnetization isotherms at 2 K. Overall, a transition from superparamagnetic (pristine and non-crystalline) to AFM (calcinated and crystalline) is reported for these MTO nanoparticles.



**Fig. 5**  $M-H$  curves of pristine and 500 °C calcined MTO compounds along with a schematic diagram of the magnetic response with and without magnetic field and structural ordering.

This study particularly correlates the observed magnetic behavior with non-crystalline to crystalline structural transition as a function of particle size.<sup>24,25</sup> The key difference between the observed magnetic behavior is that no two atomic sites are equivalent in a non-crystalline structure, whereas many macroscopic directions may be equivalent in crystalline nanostructures. Thus, this new class of nanomaterials provides excellent control over magnetic behavior *via* particle size and crystallographic structural ordering, not recounted so far. Moreover, since applications of magnetic materials rely on the stable magnetic order, the magnetic anisotropy energy along certain directions becomes comparable to the thermal energy with the reduction in particle size, forcing nanoparticles to lose their stable magnetic order and become superparamagnetic.<sup>26</sup> In contrast, the current work presents a novel class of materials that display stable magnetic order in nanoparticles of size  $\sim 6$  nm (Table S3, ESI †).

## Conclusions

In summary,  $\text{Co}_{1-x}\text{Ni}_x\text{TeO}_4$  ( $x = 0, 0.5, \text{ and } 1$ ) nanomaterials with different particle sizes are prepared by a sol-gel method and their structural and magnetic interactions are investigated. Microstructural and structural data evidence the increase in particle size and long-range structural ordering with the

calcination temperature. These nanomaterials exhibit different magnetic interactions with different calcination temperatures. Calcination at lower temperatures shows the short-range non-crystalline structure and superparamagnetic behavior, while calcination at higher temperatures exhibits long-range ordering in both the crystal and magnetic structures. Noted direct correlation between structural ordering and magnetic interactions *via* temperature and field-dependent magnetization measurements presents a new family of nanomaterials with a unique structure and magnetic property control.

## Conflicts of interest

The authors declare no conflicts of interest.

## Acknowledgements

The authors acknowledge SAIF-IIT Bombay and NCPRE-IIT Bombay for providing experimental facilities. H. S. acknowledges the Science & Engineering Research Board, DST, India, for support under NPDF project PDF/2016/001159. E. R. and H. S. also acknowledge The Academy of Finland (grant #311934) for the financial supports.

## References

1. G.-H. Lee , H. Moon , H. Kim , G. H. Lee , W. Kwon , S. Yoo , D. Myung , S. H. Yun , Z. Bao and S. K. Hahn , Multifunctional materials for implantable and wearable photonic healthcare devices, *Nat. Rev. Mater.*, 2020, **5** , 149 —165 [CrossRef](#) .
2. S. Tong , C. A. Quinto , L. Zhang , P. Mohindra and G. Bao , Size-Dependent Heating of Magnetic Iron Oxide Nanoparticles, *ACS Nano*, 2017, **11** , 6808 —6816 [CrossRef](#) [PubMed](#) .
3. Z. Cheng , Q. Fu , H. Duan , Z. Cui , Y. Xue and W. Zhang , Size-Dependent Thermodynamics of Structural Transition and Magnetic Properties of Nano-Fe<sub>2</sub>O<sub>3</sub>, *Ind. Eng. Chem. Res.*, 2019, **58** , 8418 —8425 [CrossRef](#) .
4. K. Pathakoti , M. Manubolu and H. M. Hwang , Nanostructures: Current uses and future applications in food science, *J. Food Drug Anal.*, 2017, **25** , 245 —253 [CrossRef](#) [PubMed](#) .
5. R. Dronskowski The little maghemite story: A classic functional material, *Adv. Funct. Mater.*, 2001, **11** , 27 —29 [CrossRef](#) .
6. Q. Zhang , P. Wang , X. Li , Y. Yang , X. Liu , F. Zhang , Y. Ling and Y. Zhou , Preparation of highly dispersed  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> and GdPO<sub>4</sub> co-functionalized mesoporous carbon spheres for dual-mode MR imaging and anti-cancer drug carrying, *J. Mater. Chem. B*, 2017, **5** , 3765 —3770 [RSC](#) .
7. S. Yang , C. Liu , H. Chang , L. Ma , Z. Qu , N. Yan , C. Wang and J. Li , Improvement of the activity of  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> for the selective catalytic reduction of NO with NH<sub>3</sub> at high temperatures: NO reduction versus NH<sub>3</sub> oxidization, *Ind. Eng. Chem. Res.*, 2013, **52** , 5601 —5610 [CrossRef](#) .

8. N. M. Li , K. M. Li , S. Wang , K. Q. Yang , L. J. Zhang , Q. Chen and W. M. Zhang , Gold embedded maghemite hybrid nanowires and their gas sensing properties, *ACS Appl. Mater. Interfaces*, 2015, **7** , 10534 —10540 [CrossRef](#) [PubMed](#) .
9. V. G. De Paula , L. M. Da Silva , A. O. Dos Santos , R. Lang , L. Otubo , A. A. Coelho and L. P. Cardoso , Magnetocaloric effect and evidence of superparamagnetism in GdAl<sub>2</sub> nanocrystallites: A magnetic-structural correlation, *Phys. Rev. B*, 2016, **93** , 1 —9 [CrossRef](#) .
10. S. D. Tiwari and K. P. Rajeev , Paramagnetic to ferromagnetic transition and superparamagnetic blocking in Ni (OH)<sub>2</sub> nanoparticles, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **77** , 6 —11 [Search PubMed](#) .
11. R. Caciuffo , D. Rinaldi , G. Barucca , J. Mira , J. Rivas , M. A. Señarís-Rodríguez , P. G. Radaelli , D. Fiorani and J. B. Goodenough , Structural details and magnetic order of La<sub>1-x</sub>Sr<sub>x</sub>CoO<sub>3</sub> ( $x \leq 0.3$ ), *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59** , 1068 —1078 [CrossRef](#) .
12. W. Xia , H. Wu , P. Xue and X. Zhu , Microstructural, Magnetic, and Optical Properties of Pr-Doped Perovskite Synthesized via Sol-Gel Process, *Nanoscale Res. Lett.*, 2018, **13** , 135 [CrossRef](#) [PubMed](#) .
13. T. Zhang , X. P. Wang , Q. F. Fang and X. G. Li , Magnetic and charge ordering in nanosized manganites, *Appl. Phys. Rev.*, 2014, **1** , 031302 [Search PubMed](#) .
14. A. K. Patel , M. R. Panda , E. Rani , H. Singh , S. S. Samatham , A. Nagendra , S. N. Jha , D. Bhattacharyya , K. G. Suresh and S. Mitra , Unique Structure-Induced Magnetic and Electrochemical Activity in Nanostructured Transition Metal Tellurates Co<sub>1-x</sub>Ni<sub>x</sub>TeO<sub>4</sub> ( $x = 0, 0.5, \text{ and } 1$ ), *ACS Appl. Energy Mater.*, 2020, **3** , 9436 —9448 [CrossRef](#) .
15. M. R. Panda , A. Raj K , A. Ghosh , A. Kumar , D. Muthuraj , S. Sau , W. Yu , Y. Zhang , A. K. Sinha , M. Weyland , Q. Bao and S. Mitra , Blocks of molybdenum ditelluride: A high rate anode for sodium-ion battery and full cell prototype study, *Nano Energy*, 2019, **64** , 103951 [CrossRef](#) .
16. B. R. Dahal , R. P. Dulal , I. L. Pegg and J. Philip , Electrical transport and magnetic properties of cobalt telluride nanostructures, *J. Vac. Sci. Technol., B: Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom.*, 2016, **34** , 051801 [Search PubMed](#) .
17. C. R. Feger , G. L. Schimek and J. W. Kolis , Hydrothermal Synthesis and Characterization of M<sub>2</sub>Te<sub>3</sub>O<sub>8</sub> (M = Mn, Co, Ni, Cu, Zn): A Series of Compounds with the Spiroffite Structure, *J. Solid State Chem.*, 1999, **143** , 246 —253 [CrossRef](#) .
18. H. Singh , H. Ghosh , T. V. Chandrasekhar Rao , G. Sharma , J. Saha and S. Patnaik , Short range ferromagnetic, magneto-electric, and magneto-dielectric effect in ceramic Co<sub>3</sub>TeO<sub>6</sub>, *J. Appl. Phys.*, 2016, **119** , 044104 [CrossRef](#) .
19. H. Singh , H. Ghosh , C. L. Prajapat and M. R. Singh , Griffiths like robust ferromagnetism in Co<sub>3-x</sub>Mn<sub>x</sub>TeO<sub>6</sub>; ( $x = 0.5, 1, 2$ ), *Mater. Res. Bull.*, 2016, **80** , 273 —279 [CrossRef](#) [CAS](#) .
20. S.-C. Zhu , G.-W. Chen , D. Zhang , L. Xu , Z.-P. Liu , H. Mao and Q. Hu , Topological Ordering of Memory Glass on Extended Length Scales, *J. Am. Chem. Soc.*, 2022, **144** , 7414 —7421 [CrossRef](#) [CAS](#) .

21. T. Iimori , Y. Imamoto , N. Uchida , Y. Kikuchi , K. Honda , T. Iwahashi and Y. Ouchi , Magnetic moment distribution in nanosized antiferromagnetic NiO, *J. Appl. Phys.*, 2020, **127** , 023902 [CrossRef](#) [CAS](#) .
22. T. Sorop Superparamagnetic behaviour of antiferromagnetic DyPO<sub>4</sub> nanoparticles, *J. Magn. Magn. Mater.*, 2004, **272–276** , 1573 —1574 [CrossRef](#) [CAS](#) .
23. S. Sako , Y. Umemura , K. Ohshima , M. Sakai and S. Bandow , Magnetic Property of Antiferromagnetic MnO Ultrafine-Particle, *J. Phys. Soc. Jpn.*, 1996, **65** , 280 —284 [CrossRef](#) [CAS](#) .
24. Q. Li , C. W. Kartikowati , S. Horie , T. Ogi , T. Iwaki and K. Okuyama , Correlation between particle size/domain structure and magnetic properties of highly crystalline Fe<sub>3</sub>O<sub>4</sub> nanoparticles, *Sci. Rep.*, 2017, **7** , 9894 [CrossRef](#) [PubMed](#) .
25. N. V. Rama Rao , R. Gopalan , M. Manivel Raja , V. Chandrasekaran and K. G. Suresh , Mössbauer studies on structural ordering and magnetic properties of melt-spun Ni–Fe–Ga ribbons, *Appl. Phys. Lett.*, 2008, **93** , 202503 [CrossRef](#) .
26. V. Skumryev , S. Stoyanov , Y. Zhang , G. Hadjipanayis , D. Givord and J. Nogués , Beating the superparamagnetic limit with exchange bias, *Nature*, 2003, **423** , 850 —853 [CrossRef](#) [PubMed](#) .

## Footnote

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2cp05592a>

This journal is © the Owner Societies 2023

About

Cited by

Related

### Download this article

PDF format

Article HTML

## Supplementary files

### Supplementary information

PDF (871K)

## Article information

<https://doi.org/10.1039/D2CP05592A>

### Article type

Paper

### Submitted

30 Nov 2022

### Accepted

15 Dec 2022

### First published

16 Dec 2022



This article is Open Access



### Citation

*Phys. Chem. Chem. Phys.*, 2023, **25**, 3144-3150

BibTex



Go

### Permissions

[Request permissions](#)

### Social activity



Tweet

Share

### Search articles by author

- Akhilesh Kumar Patel
- S. Shanmukharao Samatham
- Ekta Rani
- K. G. Suresh
- Harishchandra Singh

Go

Spotlight

Advertisements

[Journals, books & databases](#)[Home](#)[About us](#)[Membership & professional community](#)[Campaigning & outreach](#)[Journals, books & databases](#)[Teaching & learning](#)[News & events](#)[Locations & contacts](#)[Careers](#)[Awards & funding](#)[Advertise](#)[Help & legal](#)[Privacy policy](#)[Terms & conditions](#)

© Royal Society of Chemistry 2024

Registered charity number: 207890

# Scalable data exchange using redirecting data flow control mechanism on mqtt protocol

V Tirupathi, Dr K Sagar

**Keywords:** MQTT protocol, Scalable data exchange, MQTT broker, Subscriber, publisher client, Software Defined Network (SDN) controller.

## Abstract

The Internet of Things (IoT) is a network of interconnected, internet-connected objects that may gather and transmit data via a wireless network without the need for human participation. MQTT (Message Queuing Telemetry Transport) is a lightweight Internet of Things communication protocol with a publisher-subscriber messaging structure that allows for data flow across devices with complexities. Hence in this research, the technical complexities in achieving efficient data exchange in MQTT protocol is removed by using Scalable Data Exchange and Redirection of Clients in Large-scale Distributed IoT network thru MQTT protocol, in which the availability and scalability problem caused when the subscriber client could not obtain the data from the broker is removed by using Robust data exchange availability tactic which constructs a topic set table at each broker in the network. Moreover, the issue in handling the crashed MQTT broker while broadcasting the published data is eliminated by using Redirecting data flow control mechanism in which the brokers are controlled by the software-defined network (SDN) controller. Thus, the Scalable Data Exchange and Redirection of Clients in Large-scale Distributed IoT network thru MQTT protocol can be outperforming the other existing model with low delay and high throughput.

 PDF

Published

2022-12-23

## CREATING A BRIDGE OF BROKERS FOR MESSAGE QUEUE TELEMETRY TRANSPORT (MQTT) PROTOCOL

V. Thirupathi<sup>1,2, a</sup> Dr K. Sagar<sup>3</sup>

### Author Affiliations

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering, SR University,  
Warangal, Telangana

<sup>2</sup>Research Scholar Osmania University, Hyderabad, Telangana

<sup>3</sup>Professor, Department of Computer Science and Engineering, CBIT, Hyderabad, Telangana

### Author Emails

<sup>a)</sup> Corresponding author: [v.thirupathi@sru.edu.in](mailto:v.thirupathi@sru.edu.in)

**Abstract.** As we all know that, all soft computing devices are able to connect with each other through internet. This is calling as internet of things (IoT) [4],[8]. IoT uses different protocols to exchange information among various devices. Different protocols are available i.e., constrained application protocol (CoAP), message queue telemetry transport (MQTT), Advanced message queuing protocol (AMQP) and so on. These protocols are intended for constrained devices i.e., devices with low RAM size and processor. MQTT is a light weight protocol with publish/subscribe communication model. MQTT uses a broker to provide communication among devices which are subscribed to the broker. Various brokers are available in the market like mosquito, RabbitMQ and so on. These brokers allow to receive and send information, only the devices which have subscribed to a particular broker. The devices which have subscribed to other broker are unable establish communication. If devices would like make communication with the devices connected with other brokers is required to make a group or bridge of brokers.

### Introduction

Internet of Things (IoT) [7],[8] is a way of connecting everything to the internet. All these things/devices are connected with each other through internet and exchange information. These devices are constrained in terms of resources like less RAM size and lower processing speeds. All these devices have low computational power, so they require special communication protocols to establish communication. Constrained application protocol (CoAP), Message queue telemetry transport (MQTT), Advanced message queuing protocol (AMQP) and DDS. These protocols are being used at application layer level. Among these protocols some may follow client/server communication model or publish/subscribe model. For example, CoAP follows client/server and MQTT follows publish/subscribe communication model.

Message queue telemetry transport (MQTT) [4] is a light weight communication protocol, it follows publish/subscribe communication model. MQTT protocols requires intermediary

application to create communication among various devices. Various brokers are existing in the market like mosquitto, RabbitMQ and emqx.

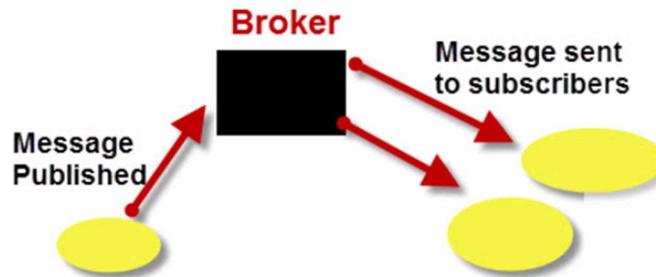


FIGURE 1: MQTT Publish Subscribe Model

MQTT [4] works as TV broad casting mechanism. Each program played on a specific channel by the TV broadcasters, users are required to tune a particular channel so that they can view the required program. Various topics are need to be created on broker. Clients who send messages are called as publishers and who receives messages are calling as subscribers. Subscribers are need to subscribe to a particular topic so that they can receive messages which are published on the same topic by the publishers. MQTT protocol uses TCP hand shake protocol to establish connection among clients. Transmission control protocol (TCP) is reliable since it is connection oriented.

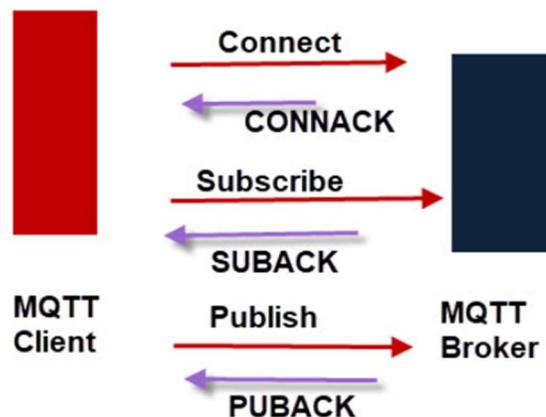


FIGURE 2. MQTT Message flow

As TCP is connection oriented, clients will receive acknowledgements after each communication. Every client has a unique name or client id. If any new connection will try to make a connection with the existing client name, connection will be lost. All clients need to connect to the broker then only they can able to communicate with each other. If clients of different brokers need to communicate with other broker's clients, then brokers must be communicated with other.

## Related Study

Authors have explained remote patient monitoring in [1], This system monitors pulse oxymeter and electrocardiogram (ECG) readings. Message queue telemetry transport (MQTT) protocol being used as a communication protocol to read sensor readings. The clients who have subscribed to a particular topic, only those clients receive the information. Others who have not subscribed to that topic will not receive messages from the sensors.

Paper [2] explains the real time remote monitoring. It reads electrocardiogram (ECG) recordings remotely. To read this information authors have used MQTT (Message queue telemetry transport) protocol. It has tested in LAN (Local Area Network) and WAN (Wide Area Networks) [5], [7]. They have created a topic on MQTT broker. All the clients who wish to receive information need to subscribe to that topic.

In [3] authors explained how to read blood pressure (BP), blood oxygen levels, body weight and body temperature with different sensors. All the data will reach to the central point from these sensors. For communication authors have used (RFID) radio frequency identification and (MQTT) message queue telemetry transport protocol. Through MQTT protocols clients will receive information from the sensors.

[4] paper has explained home automation system using MQTT protocol. Authors have explained how to control home appliances remotely using MQTT protocol. Home appliances and control unit must have connected to the same topic on the MQTT broker. They have used mosquito as the MQTT broker. But if others would like to control the home appliance, they are unable to do it.

## Proposed Model

MQTT bridge allows to increase scalability. If there are two brokers in the network i.e., brokers B1 and B2, clients of B1 are unable subscribe the topics on B2 and clients of B2 are unable to subscribe to topics on B1. If it needs to be happened, we have to create a bridge of brokers.

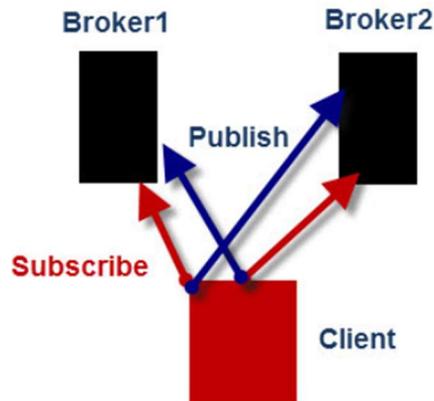


FIGURE 3: MQTT Bridge

When we group a bridge of brokers, one of the broker acts as bridge or server. Other brokers act as a normal client. Figure 3 depicts MQTT [9] bridge, it has two brokers and all the clients of two brokers can able to communicate. We need to configure any one of the brokers as a server. While configuring the server we must mention the IP address and port number of server broker. It also needs to mention the client's name and topic names which are going to be publish message or subscribe to. We must mention all the details as follows

topic topic-pattern direction qos local/remote prefix

it uses in, out and both as directions. Out means, writes messages on to subscribers. In means receive messages from publishers. Both means, they can able to receive and send messages. QOS indicates quality of service levels. MQTT provides three types of quality of services i.e QOS1, QOS2 and QOS3.

```
#connection <name>

connection bridge-01
address 192.168.1.184:1883

topic # out 0
topic # in 0
```

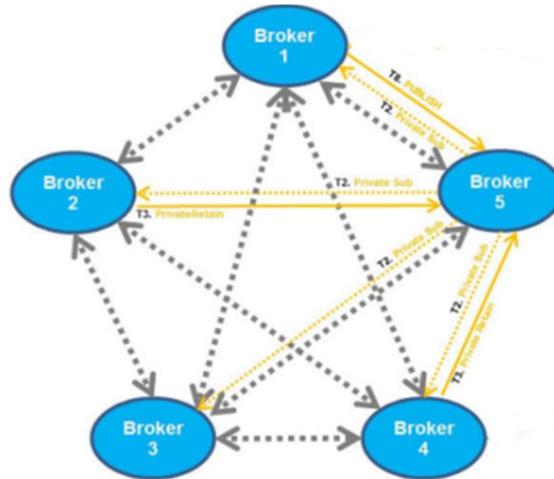


FIGURE 4: MQTT broker cluster

The above figure is MQTT [4] broker cluster, which has many numbers of brokers so that it increases scalability. When we use a cluster architecture, we must need to concentrate on adjusting the clients to other brokers if any one of the brokers misfunctions. We must also identify a good technique to monitor the total network. As the number of brokers increasing in the network, it increases the complexity.

### The Result Analysis

We have done some experiments. Where we sent different number of messages through different brokers. n=10k means 10000 messages to send per client, c=100 means 100 clients to start. MsgSize=1000bytes, QoS=2.

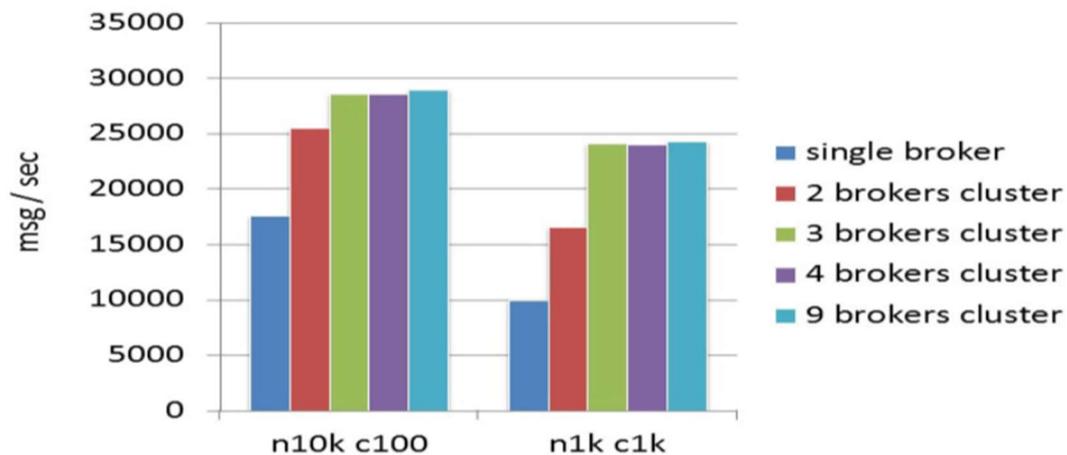


FIGURE 5: Mean bandwidth

## Conclusion and Future Scope

We can able to create a cluster of brokers with many brokers. But there is a problem with cluster is, there is no way of monitoring the network. If any one of the brokers has failed, the total network might get down. We can solve this problem with the help of software defined network (SDN) [9], [11]. With the help of SDN we can monitor the network, if any broker failed it adjust the clients to other brokers.

## References

1. Timothy Malche, Sumegh Tharewal, Pradeep Kumar Tiwari, Mohamed Yaseen Jabarulla, Abeer Ali Alnuaim, Wesam Atef Hatamleh, Mohammad Aman Ullah, "Artificial Intelligence of Things- (AIoT-) Based Patient Activity Tracking System for Remote Patient Monitoring", *Journal of Healthcare Engineering*, vol. 2022, Article ID 8732213, 15 pages, 2022. <https://doi.org/10.1155/2022/8732213>
2. H. T. Yew, M. F. Ng, S. Z. Ping, S. K. Chung, A. Chekima and J. A. Dargham, "IoT Based Real-Time Remote Patient Monitoring System," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), 2020, pp. 176-179, doi: 10.1109/CSPA48992.2020.9068699.
3. Yung-Chung Tsao, Fu-Jen Cheng, Yi-Hua Li, Lun-De Liao, "An IoT-Based Smart System with an MQTT Broker for Individual Patient Vital Sign Monitoring in Potential Emergency or Prehospital Applications", *Emergency Medicine International*, vol. 2022, Article ID 7245650, 13 pages, 2022. <https://doi.org/10.1155/2022/7245650>
4. Thirupathi V and Sagar K 2018 Implementation of home automation system using mqtt protocol and esp32 *International Journal of Engineering and Advanced Technology* 8(2C2) 111–113.
5. Pramod Kumar P, Naresh Kumar S., Thirupathi V. and Sandeep C, (2019), QOS and security problems in 4G networks and QOS mechanisms offered by 4G. *International Journal of Advanced Science and Technology* 28(20) 600-606.
6. Kothandaraman D, Sheshikala M, Seena Naik K, Chanti Y, Vijaykumar B, (2019), Design of an optimized multicast routing algorithm for internet of things *Int J Recent Technol Eng* vol 8(2) pp 4048-4053
7. Chythanya KR, Kumar KS, Rajesh M, Tharun Reddy S, (2020), Sensor Cloud: A Breakdown information on the Utilization of Wireless Sensor Network by Means of Cloud Computing *Test Eng Manage* vol 82 pp13945-13954
8. Sallauddin Mohmmad , G. Sunil , Ranganath Kanakam, (2017), A Survey On New Approaches Of Internet Of Things Data Mining *International Journal of Advanced Research in Computer Science* vol 8(8) pp 123-135

9. Sallauddin Mohmmad, Dr.M.Sheshikala, Shabana, (2018), Software Defined Security (SDSec):Reliable centralized security system to decentralized applications in SDN and their challenges Jour of Adv Research in Dynamical & Control Systems Vol 10(10) pp 147-152
10. Pramod Kumar P, Naresh Kumar S., Thirupathi V. and Sandeep C, (2019), QOS and security problems in 4G networks and QOS mechanisms offered by 4G. International Journal of Advanced Science and Technology 28(20) 600-606.
11. Thirupathi V, Sandeep C, Naresh Kumar S and Pramod Kumar P 2019 A comprehensive review on sdn architecture, applications and major benifits of SDN International Journal of Advanced Science and Technology 28(20) 607–614.



# Energy Aware Priority Based Event Routing Protocol Using TDMA Communication for Internet of Things

Vijaya Krishna Akula<sup>1</sup> · I. Ravi Prakash Reddy<sup>1</sup> · A. Anny Leema<sup>2</sup> · Ramana Kadiyala<sup>3,6</sup>  · Raman Dugyala<sup>4</sup> · K. Prasanna<sup>5</sup>

Accepted: 17 May 2023 / Published online: 6 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

The Internet of Things (IoT) is a well-known platform for analysing data generated by environmental sensors and instruments. Transporting data from IoT sensor nodes to the cloud is time-consuming due to congestion and energy consumption in IoT networks. As a consequence, in this paper the Priority-Based Event Routing Protocol using TDMA channel to forward the packets in different network traffic constraints. In our proposed work we categorize the normal and priority packet by two properties such as transmission rate and priority then we can use the Time slot based mechanism TDMA protocol to forward the packet in the network based on their time intervals to avoid the congestion and increase the packet forwarding ratio. This approach will increase the network robustness and reduces the congestion. In our work we uses the Qualnet Simulator, the experiments showed that the proposed technique performs more and enhance the performance in IoT networks with saving energy consumption by 19.696%, reduces routing overhead by 77%, and reduces end-to-end delay by 50.6% with respect to existing ones.

**Keywords** Internet of things · Routing protocols · Energy · Congestion · Communications

## 1 Introduction

The most recent development in communication technology is IoT, where it is collection of RFID Tags or Sensor nodes [1]. The communication between the sensor nodes are carried through the Wifi, LTE, Wimax and Bluetooth [2-4]. The sensor nodes are having limited storage, computation and battery power and they can communicate with in shorter distances. IoT devices need to be smart in handling, communicating and transmitting about the occurred event in the environment [5, 6]. The event manager in the IoT environment handles the middleware to publish or subscribe events and it acts as the coordinator between the sensor nodes. Many routing protocols are developed to address the communication in IoT [7]. Most of them are concentrated on IEEE 802.15.4 and extended IPV6 standards and it is not preferable for small sensor networks. In the recent studies [8-10], the researchers are concentrated on developing the clustering techniques for effective routing in IoT nodes. The most common issue in the IoT network is event based routing. The events in the IoT

Extended author information available on the last page of the article

need to be delivered instantaneously to achieve the requirements of the applications. Event based routing allows the devices to concentrate on particular events. If any device found the event triggering, it notifies to the subscribers who are interested in particular event.

The following is a list of significant contributions made by the study:

- Creating an event-routing-first approach in IoT systems for node traffic, latency and energy consumption may be reduced
- In the RPL protocol, node selection for the purpose of a parent node to avoid excessive branching and save time and energy.
- Improving IoT network's optimal packet transmission speed efficiency.

The remainder of the work is organised in the following manner. The second portion focuses on recent research in priority routing and energy-aware routing. Section 3 goes into the energy notion of the IoT Network. The priority order that is suggested for an effective routing method is explained in 4. Segment 5 discusses the experimental analysis. Finally, Segment 6 brings the investigation to an end.

## 2 Literature Survey

This section deals with the related work regarding the energy efficient and priority based routing protocols in the IoT environment.

### 2.1 Energy Efficient Models

The authors of [11] created a method for calculating data centre electricity consumption in a cloud environment that is energy-conscious. The effectiveness of the suggested approach was evaluated using the linear programming model. [12] offers the ant colony technique for IoT routing. This method is useful for determining the shortest path between two places. The authors developed the energy-aware AODV protocol for vehicle ad-hoc networks in [13]. By joining the vehicles that are close to one another, this technique finds the fastest path. To cut down on data centres' energy usage, the authors of [14] developed an autonomous cloud architecture. With the goal of lowering network latency, eliminating duplicate packets, and extending network lifetime, the authors of [15] established the content-centric routing strategy for content information flow and data integration.

In [16], The multi-objective routing technique for the energy-aware IoT model was created by the authors. Node lifetime, distance, and energy consumption are three factors taken into account while evaluating its performance. The authors developed the C-means clustering approach to help the IoT network have fewer nodes. The process performed better than heuristic strategies, according to the test findings. In [17], the authors developed the fuzzy approach for clustering in IoT networks. This mechanism used the immune system model for routing to minimize the nodes energy consumption. This method proved that the jitter ratio and packet delivery ratio is less compared against the optimization algorithms. In [18], the authors proposed the markov model for routing in vehicular IoT networks. The probabilistic graph is used to help the communication between the nodes for minimizing the energy and delay.

## 2.2 Priority Based Routing Protocols

The routing protocols in WSNs are broadly classified as centralized and distributed. For instance, in [19], the authors developed the merger tree mechanism for sensor networks which is a centralized process. This mechanism improves the network lifetime of the WSN. In [20], the authors developed the data aggregation model for minimizing the energy consumption in the WSNs. It uses the distributed clustering process for reducing the packet overhead. Furthermore, DAG [6] or tree based model [21] require the specific routing mechanism to handle the network dynamically. It is due to that if any network failure or the discharge of crucial IoT nodes occurs, To avoid packet loss, the routing mechanism must update the information to the remaining IoT nodes.

In [22], the authors presented a priority-based and energy-efficient routing method. This technique is based on the routing protocol for low-power and lossy networks (RPL model), which establishes routing through contents. Each network slot uses timing patterns to deliver data to the destination while accounting for network traffic, audio, and image data. In this work, they are more concerned with traffic than with energy consumption.

In [23], the authors presented a Cluster-based Backpressure routing scheme was proposed. It employs an energy load-balancing strategy to improve data transmission reliability and network longevity. Furthermore, this technique will do data aggregation and prevent duplicate data packets. They select the closest cluster head and sink node to save the greatest energy in this work, however they do not focus on forwarding priority packets in emergency situations.

In [24], the authors proposed a Cluster-based Energy-aware Data Aggregation Routing (CEDAR) protocol in the IoT. To overcome these challenges, the fuzzy logic system and the Capuchin Search Algorithm (CapSA) were combined. The proposed hybrid routing algorithm involves two main phases: cluster formation and intra/extra cluster routing. They primarily organise sensor nodes into optimal clusters in the first phase. In the second step, identify the most effective paths between cluster heads and cluster members. In this work, they primarily concentrated on cluster construction and data transport within and between clusters. They fall short, however, in terms of handling congestion in low-powered devices and sending packets in emergency scenarios in the event of unusual events.

By this study we can analyze that there is some gaps and challenges in the Existed works. Many of the researchers are not resolved these issues. They are:

- Delay
- Priority Based Packets Routing
- Congestion
- Packet Loops

In our Proposed work we are working on this challenges to enhance the performance in IoT.

### 3 Energy Model for IOT

When designing a routing system for an IoT network, energy balance and economy are essential factors. The aim of this research was to develop a node-life-extending, energy-efficient routing method. We explored network heterogeneity to three different degrees based on node initial energy to minimize energy loss. The network’s nodes are all in static mode. Ordinary nodes have the least energy, transitional nodes have a medium amount, and super nodes have the most energy on the first level.

In our proposed work we are using the modified version of LEACH protocol. We are able to cluster the nodes using these techniques. Each cluster picks a cluster head to collect data from the member nodes and send it to the sink node. The cluster head is crucial for reducing energy consumption and extending node life as a result. We investigated network heterogeneity at three distinct levels based on node starting energy in order to restrict energy dissipation. The network’s nodes are all operating in static mode. The energy level of normal nodes is the lowest, that of intermediate nodes is the middle, and that of advanced nodes is the highest. In this method, we can save the energy of WSNs.

The aggregate energy of the ordinary, transitional, and super nodes is shown in Eqs. 1, 2, and 3.

$$\vartheta_{On} = k\vartheta_0(1 - a - b) \tag{1}$$

$$\vartheta_{Tn} = kb\vartheta_0(1 + B) \tag{2}$$

$$\vartheta_{Sn} = ka\vartheta_0(1 + A) \tag{3}$$

An ordinary node is shown with  $\vartheta_{On}$ , for Transitional nodes it has an energy of  $\vartheta_{Tn}$ , and for advanced nodes is  $\vartheta_{Sn}$ . The energy of the network’s Super nodes are denoted by the symbol a, and their energy leveling is given the letter A. In the network, b represents the Transitional nodes and their energy leveling is specified as  $B=A/2$ . k denotes the node size of each kind.

The overall energy usage of the entire network is listed.

$$\begin{aligned} \vartheta_{tot} &= k\vartheta_0(1 - a - b) + kb\vartheta_0(1 + B) + ka\vartheta_0(1 + A) \\ &= k\vartheta_0(1 + aA + bB) \end{aligned} \tag{4}$$

We investigate the SEP and LEACH protocol models for CH election. The likelihood of being chosen as the cluster head determines the threshold value for each type of node. Assume S1, S2, and S3 are sets of nodes that belong to each category but were not selected as cluster heads. Cluster heads should be chosen from the nodes below.

For  $O_n$ :

$$P_{On} = \frac{P}{1 + aA + bB} \tag{5}$$

$$T_{K_{On}} = \begin{cases} \frac{P_{On}}{1 - P_{On}^{(u \bmod \frac{1}{P_{On}})}}; & \text{if } k_{On} \in S_1 \\ 0; & \text{otherwise} \end{cases} \tag{6}$$

For  $T_n$ :

$$P_{Tn} = \frac{P(1 + B)}{1 + aA + bB} \tag{7}$$

$$T_{K_{Tn}} = \begin{cases} \frac{P_{Tn}}{1 - P_{Tn}^{(u \bmod 1/P_{Tn})}}; & \text{if } k_{Tn} \in S_2 \\ 0; & \text{otherwise} \end{cases} \tag{8}$$

For  $S_n$ :

$$P_{Sn} = \frac{P(1 + A)}{1 + aA + bB} \tag{9}$$

$$T_{K_{Sn}} = \begin{cases} \frac{P_{Sn}}{1 - P_{Sn}^{(u \bmod 1/P_{Sn})}}; & \text{if } k_{Sn} \in S_3 \\ 0; & \text{otherwise} \end{cases} \tag{10}$$

Expected probability of choosing the cluster heads on each epoch is deliberated using Eqs. 5, 7 and 9.

$$k(1 - a - b)P_{On} + kaP_{Sn} + kbP_{Tn} = kP \tag{11}$$

Equation 11 generates an average probability of selecting cluster heads in a diverse environment that is comparable to the value of the LEACH method.

The communication paradigm for heterogeneous IoT networks is depicted in Fig. 1. In a heterogeneous network, energy dissipation is computed using either a free space or a multipath fading model. The sensor node uses up energy to transfer each packet of 'n' bits, as shown in 12 and 13.

$$\vartheta_{ir}(N, D) = \vartheta_{ir\_ele}(N) + \vartheta_{ir\_amp}(N, D) \tag{12}$$

$$\vartheta_{ir}(N, D) = \begin{cases} \vartheta_{ele} \times N + \vartheta_{FS} \times N \times D^2, & D \leq D_0 \\ \vartheta_{ele} \times N + \vartheta_{MP} \times N \times D^4, & D > D_0 \end{cases} \tag{13}$$

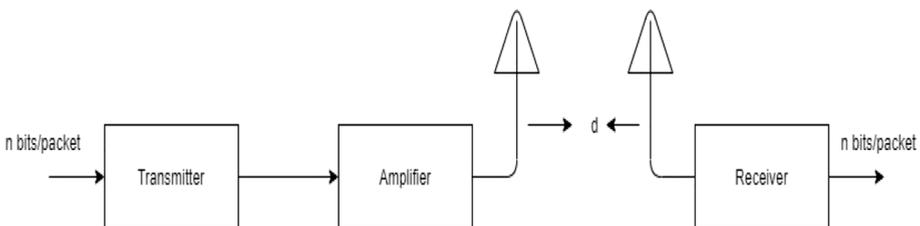


Fig. 1 IoT sensor node communication model

$\vartheta_{FS}$  and  $\vartheta_{MP}$  describes the fading channels with free and multi-path  $\vartheta_{ele}$  how much power the sensor node used throughout the packet delivery operation is represented by

$$\vartheta_{rc}(N) = \vartheta_{rc\_ele}(N) + N\vartheta_{ele} \tag{14}$$

The amount of energy consumed by a sensor nodes to receive the packets by total number of sensor nodes is represented in the Eq. 14. With this equation we can calculate how much amount of energy is consumed by an IoT sensor nodes for receiving the packets in the network.

The cluster head is chosen by keeping track of each round’s energy usage based on an energy threshold value. To increase network longevity, the suggested method establishes the energy dissipation threshold for each type of node.

### 4 Proposed Model

In IoT, gathering huge amounts of information from the different contents leads to the congestion in the system. To overcome this dispute, we propose the priority-based event routing model where it considers the trimming pattern in the network. This timing pattern considers the time and distance of each time slot’s packets that were transmitted to the receiver. This method helps to prevent network congestion and improves QoS in routing. [25].

Figure 2 explains about the framework for the priority-based event routing mechanism. The node must figure out the most effective way to send data to other nodes. Finding the best path is the challenging task in the IoT network. In IoT network, sensor nodes are suffering with resource restrictions, hence managing the resources is important in the node communication over the Internet. The proposed routing model finds the appropriate route based on the priority

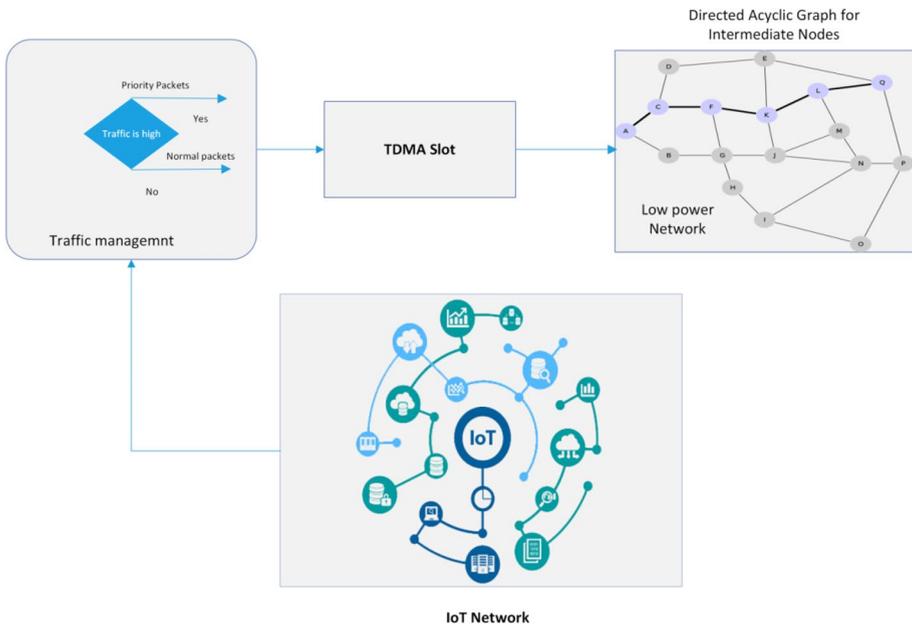


Fig. 2 Proposed priority based event routing framework

packets. The network traffic is considered as two types: low traffic and high traffic scenario. In the low traffic scenario, the normal packets with general procedure are adapted. In the high traffic scenario, the packets with highest priority are given preference to transmit in to the network.

In our work we are uses trimming mechanism based on that we can categorize the Packets and assigns the timeslots to forward the packet in TDMA channel. After that we uses RPL protocol to forward the packets into the cloud. For networks with various forms of traffic, this technology, which utilizes the same Zigbee standard, enables both one-to-one and many-to-one communication. It creates only one route in order to create Directed Acyclic Graphs (DAGs), an Objective Function (OF) that the user specifies is used. The process for choosing the best path among all sensing devices is described in the OF. In our work we supposed to forward the priority and regular packets by avoiding loops using RPL protocol to the cloud.

In our work we uses the TDMA channel allocation method. With TDMA, the spectrum is used more effectively because more users can share the same frequency depending on the time period. Using the TDMA approach, we can avoid channel contention for several users. Because TDMA networks have lower operational costs than typical FDMA networks, we can favour them in our work over alternative channel allocation techniques such as FDMA and CDMA. Voice, data, and video are all provided at different data speeds, and TDMA can handle all of them. The energy of an IoT sensor node can be enhanced thanks to TDMA's efficient resource use on a need-basis.

#### 4.1 Proposed Priority Based Event Routing Model

We concentrated on event based routing approach the nodes are assigned with two properties such as transmission rate and priority. Here we classify the priority and regular packets by labelling. The node priority is assigned with high and low using 0 and 1. The packets with highest priority are attached with 0 bit by the sender and forward to the receiver. The nodes with lowest priority are attached with 1 bit by the sender and forward to the receiver. When there is network congestion or in the emergency conditions, the highest priority packets are forwarded and in the normal load conditions the low priority packets are forwarded.

The TDMA channel is used for communication. This time slot used in TDMA channel synchronizes the sender and receiver for data forwarding and also it reduces the energy consumption. Before forwarding the packets in the TDMA channel, the network traffic is verified. If the network traffic is high, the high priorities packets are forwarded otherwise low priority packets are forwarded. The timeslot allocation for data transmission in IoT network is depending on the node transmission rate and priorities. The coordination is compulsory between the sender and receiver for deciding the time slots. If the two are more nodes having the same priority, then TDMA slot is selected based on the rate of data transmission.

---

 Algorithm 1: priority based Event routing (PBER)
 

---

Input:  $G \rightarrow$  {priority packets, Non-priority packets}

Output: optimal route based on the energy consumption and packet priority

Begin

1. TDMA time slot allocation
  2. NC: Network Congestion
  3. If  $NC=1$  then
  4.     Send the priority packets
  5.     Call **function send-priority (priority packets)**
  6.     Else
  7.         Send the non-priority packets
  8.     End if
  9.     Return
  10. **function send-priority (priority packets)**
  11. While(priority packets)
  12.     Divide data stream in to packets
  13.     Configure frame settings
  14.     Prioritize frames
  15.     Send the packets based on their priorities
  16.     If packets are received then
  17.         Verify the TDMA slot for ordering the packets
  18.     End if
  19.     Calculate the energy consumption using Eqs.13 and 14 to send and receive the packets
  20.     packets
  21.     End while
  22.     return
- 

In the above algorithm we can categorizes the packets into Priority and Regular packets by assigning the packet label with 0 or 1. Based on this the proposed mechanism identifies the regular and priority packet. After that we can apply trimming pattern mechanism. This trimming pattern mechanism is works on two properties such as Transmission rate and Priority. If there is a priority packets we can send the priority packets, otherwise we can forward the regular packets. While sending the priority packets we can divide the data stream into packets and configure the frame arding the

**Table 1** Parameters for experimental analysis

Parameters	Values
Size of the Network	1000 m×1000 m
Node radio range	250 m
Simulation time	200 s
Movement model of the Node	Random waypoint
Type of the Traffic	CBR
Node Energy	3 J
Packet Size	3000 bits
Transmission Power	20dBm

packets we can calculate the energy consumption and monitor the network. Settings and send the packets in assigned TDMA slots. In this we can avoid the congestion in the network. While forwarding the packets we can calculate the energy consumption and monitor the network.

## 5 Experimental Analyses

The QualNet simulator was used to test the proposed PBER technique. Table 1 shows the experimental assessment parameters. In our work we are comparing our proposed mechanism PBER against QoS-based routing (QRPL) [26] and PRAI [27] techniques in terms of end-to-end latency, routing overhead, and energy use. We chosen these existed algorithms for comparing our work with respect to modified version of RPL protocol with the existed ones QRPL and for priority event routing we can take the consideration of existed work PRAI.

Routing Overhead ( $R_{OH}$ ):  $P_{size}$  is the ratio of the total size of control packets to all packets received at the destination. Route request (Rreq) and route error (RErr) packets are two types of control packets. The routing overhead associated with the IoT network is depicted in Eq. 15.

$$R_{OH} = \frac{R_{req} + R_{Err}}{P_{size}} \quad (15)$$

End-to- End delay ( $E^2D$ ): it is defined with three parameters such as packet transfer time ( $d_{tr}$ ), time taken for identifying the packet priority in the TDMA slot ( $d_{pr}$ ), time taken for verifying the network traffic ( $d_{iff}$ ).

$$E^2D = \frac{\sum_{i=1}^n d_{tr} + d_{pr} + d_{iff}}{K} \quad (16)$$

$k$  represents the overall number of packets that were successfully delivered to the destination.

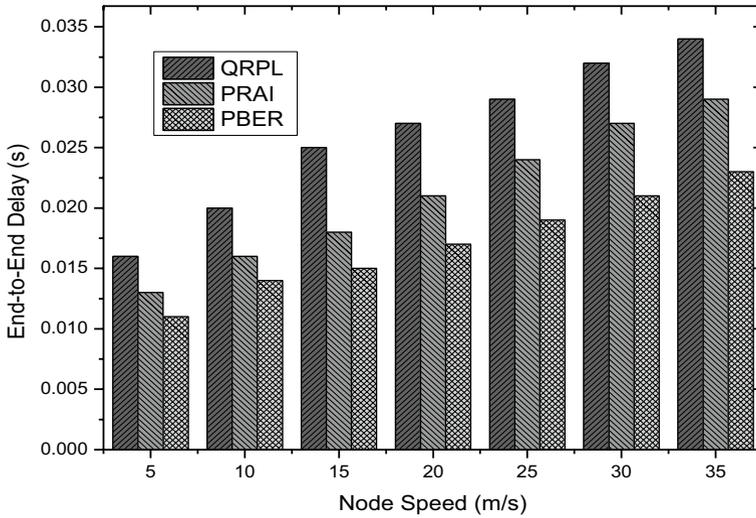


Fig. 3 End-to-end delay Vs node speed

### 5.1 Comparison of End-to-End Delay

In relation to node speed, Fig. 3 compares the end-to-end delays of the proposed and existing algorithms. The nodes’ speed is shown to be increasing, the delay of the algorithms are also increases. It happens due to the breakdown of the network paths in the network topology.

The End-to-End Delay with regard to Node Speed attained with the help of the QRPL, PRAI, and PBER algorithms is shown in Fig. 3. In QRPL and PRAI, the End-to-End Delay is respectively 54.6% and 72.2% of the Node speed. In PBER, the End-to-End Delay in relation to Node Speed is 45.7%. We may conclude from the findings that our suggested approach, End-to-End Delay Vs Node speed, is 26.5% less efficient than QRPL and 8.9% less efficient than PRAI. By doing this, even if node performance improves, we may decrease the End-to-End delay in our model.

Figure 4 illustrates the system’s end-to-end delay when node pause time is applied to the network. Observations indicate that when node rest periods are increased, the network becomes more dependable, reduces delays, and experiences fewer link failures. The proposed PBER algorithm gives messages a higher priority based on latency reduction.

Figure 4 show the End-to-End Delay with respect to Node Pause Time achieved using the QRPL, PRAI and PBER algorithms. The End-to-End Delay with respect to Node Pause Time in QRPL and PRAI is 58.3% and 50%. The End-to-End Delay with respect to Node Pause Time in PBER is 42%. With the results we can say that our proposed algorithm, End-to-End Delay Vs Node Pause Time is 16.3% less than as compared to QRPL and 8% less than PRAI. In this way we can reduce the End-to-End delay in our model even the node Pause Time increases also.

In Fig. 5, the end-to-end latency and node count of QRPL, PRAI, and PBER are all compared. The nodes’ maximum speed, energy consumption, and pause time are all set at 30 m/s, 3 J, and 2 s, respectively. End-to-end latency appears to have minimal influence as the number of nodes rises, as seen in the graph. Both the suggested and present

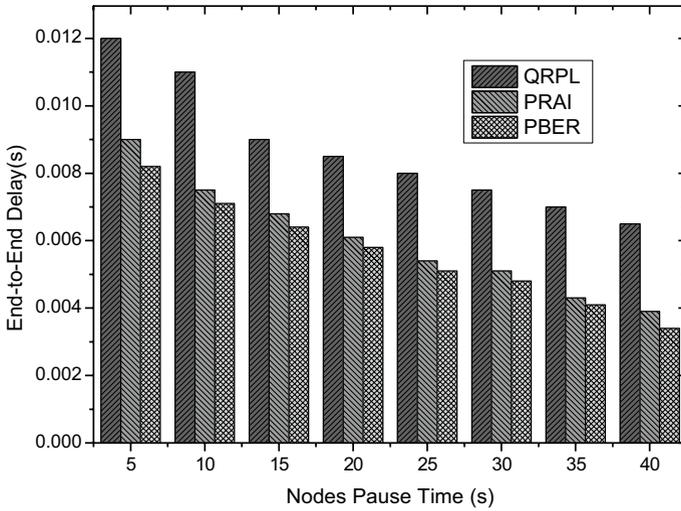


Fig. 4 End-to-end delay Vs node pause time

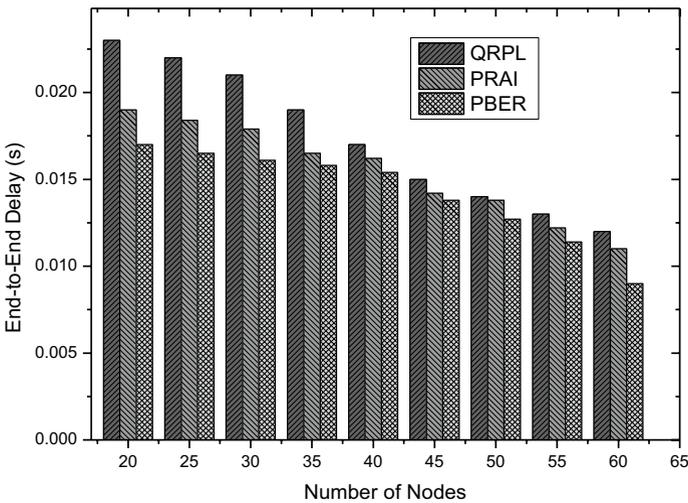
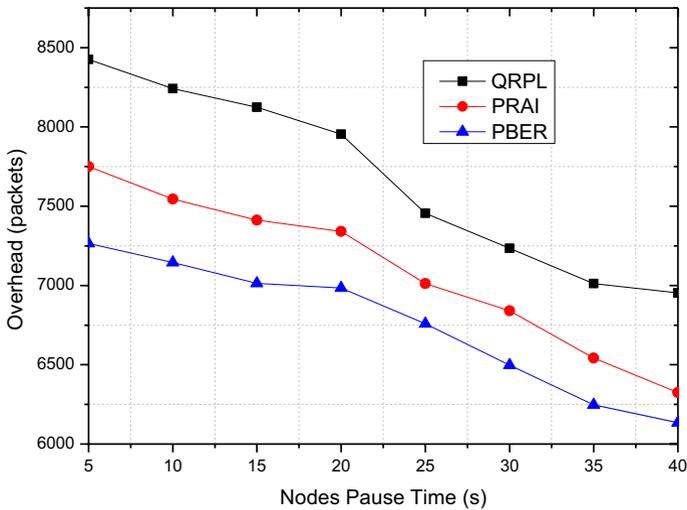


Fig. 5 End-to-end delay Vs number of nodes

techniques, on the other hand, showed a reduction in end-to-end latency. This is due to the fact that there are several pathways from the source to the destination, as well as the option of using auxiliary nodes.

Using the QRPL, PRAI, and PBER algorithms, the End-to-End Delay with respect to Number of Nodes is shown in Fig. 5. In QRPL and PRAI, the End-to-End Delay is 81.6% and 69.08%, respectively, depending on the number of nodes. According to the number of nodes in PBER, the end-to-end delay is 64.2%. We may conclude from the findings that our suggested approach, End-to-End Delay Vs Number of Nodes, is 17.4%



**Fig. 6** Routing overhead Vs nodes pause time

faster than QRPL and 4.88% faster than PRAI. By doing this, even as the number of nodes in our model grows, we can decrease the end-to-end delay.

## 5.2 Routing Overhead

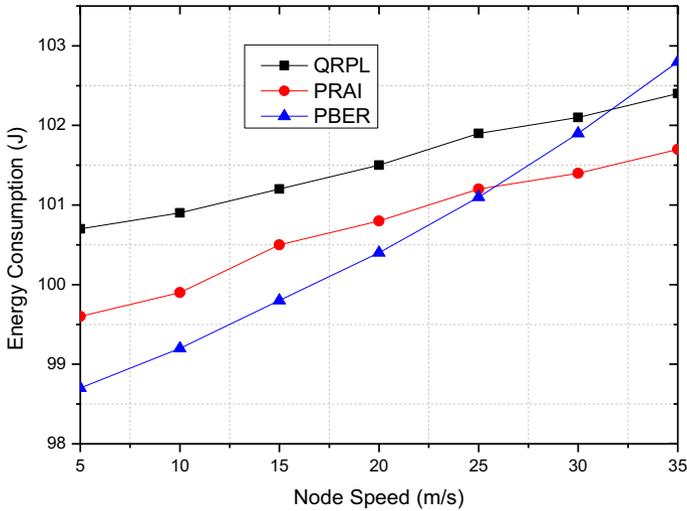
Figure 6 depicts the routing overhead for the QRPL, PRAI, and PBER algorithms. For the simulation, 100 nodes are chosen, with each node having a maximum speed of 30 m/s, spending 3 J of energy, and halting for 2 s. The PBER algorithm offers a smaller routing overhead than the QRPL and PRAI algorithms. It is due to that the node pause time increases will ultimately slows the mobility of the nodes. This process reduces the routing overhead. In the proposed PBER algorithms, the packet prioritization and path selection is easily carried when the routing overhead is less.

The Routing Overhead with reference to Node Pause time is displayed in Fig. 6 using the QRPL, PRAI, and PBER algorithms. The Routing Overhead in relation to Node Pause time in QRPL and PRAI is 92% and 84%, respectively. In PBER, the Routing Overhead to Node Pause Time ratio is 77%. According to the data, our proposed method has a 15% lower routing overhead than QRPL and a 7% lower routing overhead than PRAI when compared to node pause time. In this way we can reduce the routing overhead in our model even the Nodes Pause Time increases also.

## 5.3 Energy Consumption

Figure 7 shows how the PBER algorithm compares to the QRPL and PRAI algorithms in terms of energy usage. It has been observed that as node speeds increase, so does network energy consumption. The network dynamics alter when the node speed of the recommended model is raised, demanding extra routing and wasting substantial energy.

Figure 7 depicts the Energy Consumption with relation to Node Speed using the QRPL, PRAI, and PBER algorithms. QRPL and PRAI consume 30% and 22.2% of their energy in



**Fig. 7** Energy consumption Vs node speed

relation to node speed, respectively. PBER consumes 19.96% of its energy as a function of node speed. With the results we can say that our proposed algorithm, Energy Consumption Vs Node speed is 10.04% less than as compared to QRPL and 2.04% less than PRAI. In this way we can reduce the Energy Consumption in our model even the node speed increases also.

## 6 Conclusion

In IoT networks, many of the researchers are not focused on some issues like Priority Based Packets Routing, Congestion, Packet Loops But in our proposed work we are working on this challenges to enhance the performance in IoT. In this work, we proposed the priority-based event routing for IoT networks. The suggested technique takes into account the IoT's QoS characteristics and transmits packets via a TDMA slot. This time slot used in TDMA channel synchronizes the sender and receiver for data forwarding and also it reduces the energy consumption. Before forwarding the packets in the TDMA channel, the network traffic is verified. If the network traffic is high, the high priorities packets are forwarded otherwise low priority packets are forwarded. By using the suggested algorithm, you can cut your energy use by 19.96%, routing overhead by 77%, end-to-end delays based on node speed by 45.7%, end-to-end delays based on node pause time by 42%, and end-to-end delays based on node number by 64.2%. Meta heuristic frame transfer capabilities for the IoT will be included in later iterations of the suggested approach.

**Author Contribution** VKA: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Writing—original draft; RPRI: Data Curation, Investigation, Resources, Software; ALA: Supervision, Writing – review & editing, Project administration, Visualization; RK: Supervision, Writing – review & editing, Project administration, Visualization; GD: Writing – review & editing, Visualization, RD: Software, Project administration, Validation; PK: Supervision, Writing – review & editing, Visualization.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data and Material Availability** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to containing information that could compromise the privacy of research participants.

**Code Availability** The code that supports the findings of this study are available on request from the corresponding author. The code is not publicly available due to containing information that could compromise the privacy of research participants.

**Data Availability** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to containing information that could compromise the privacy of research participants.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical Approval** This research did not contain any studies involving animal or human participants, nor did it take place on any private or protected areas. No specific permissions were required for corresponding locations.

## References

1. Park, S., Crespi, N., Park, H., & Kim, S. H. (2014, March). IoT routing architecture with autonomous systems of things. In *2014 IEEE World Forum on Internet of Things (WF-IoT)* (pp. 442–445). IEEE.
2. Dhumane, A., Guja, S., Deo, S., & Prasad, R. (2018). Context awareness in IoT routing. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1–5). IEEE.
3. Yehia, K., et al. (2019). Cloud-based multi-agent cooperation for IoT devices using workflow-nets. *Journal of Grid Computing*, *17*, 625–650.
4. Al Ridhawi, I., Kotb, Y., Aloqaily, M., Jararweh, Y., & Baker, T. (2019). A profitable and energy-efficient cooperative fog solution for IoT services. *IEEE Transactions on Industrial Informatics*, *16*(5), 3578–3586.
5. Souri, A., Hussien, A., Hoseyninezhad, M., & Norouzi, M. (2022). A systematic review of IoT communication strategies for an efficient smart environment. *Transactions on Emerging Telecommunications Technologies*, *33*(3), e3736.
6. Kharrufa, H., Al-Kashoash, H. A., & Kemp, A. H. (2019). RPL-based routing protocols in IoT applications: A review. *IEEE Sensors Journal*, *19*(15), 5952–5967.
7. Zikria, Y. B., Afzal, M. K., Ishmanov, F., Kim, S. W., & Heejung, Y. (2018). A survey on routing protocols supported by the Contiki Internet of things operating system. *Future Generation Computer Systems*, *82*, 200–219. <https://doi.org/10.1016/j.future.2017.12.045>
8. Souri, A., & Norouzi, M. (2019). A state-of-the-art survey on formal verification of the internet of things applications. *Journal of Service Science Research*, *11*(1), 47–67. <https://doi.org/10.1007/s12927-019-0003-8>
9. Musaddiq, A., Zikria, Y. B., & Kim, S. W. (2018, November). Energy-aware adaptive trickle timer algorithm for RPL-based routing in the Internet of Things. In *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1–6). IEEE.
10. Souraya, Y., et al. (2018). An energy efficient and QoS aware routing protocol for wireless sensor and actuator networks. *AEU-International Journal of Electronics and Communications*, *83*, 193–203.
11. Baker, T., Al-Dawsari, B., Tawfik, H., Reid, D., & Ngoko, Y. (2015). GreeDi: An energy efficient routing algorithm for big data on cloud. *Ad Hoc Networks*, *35*, 83–96. <https://doi.org/10.1016/j.adhoc.2015.06.008>

12. Baker, T., García-Campos, J. M., Reina, D. G., Toral, S., Tawfik, H., Al-Jumeily, D., & Hussain, A. (2018). GreeAODV: An energy efficient routing protocol for vehicular ad hoc networks. In *Intelligent Computing Methodologies: 14th International Conference, ICIC 2018, Wuhan, China, Proceedings, Part III* 14 (pp. 670–681). Springer International Publishing.
13. Baker, T., Ngoko, Y., Tolosana-Calasanz, R., Rana, O. F., & Randles, M. (2013). Energy efficient cloud computing environment via autonomic meta-director framework. In *2013 Sixth International Conference on Developments in eSystems Engineering* (pp. 198–203). IEEE.
14. Cai, Y., Liu, J., Fan, X., Qiu, Y., & Tan, B. (2018). Software defined status aware routing in content-centric networking. In *2018 International Conference on Information Networking (ICOIN)* (pp. 283–288). IEEE.
15. Dhumane, A. V., & Prasad, R. S. (2019). Multi-objective fractional gravitational search algorithm for energy efficient routing in IoT. *Wireless networks*, 25, 399–413.
16. Preeth, S. S. L., Dhanalakshmi, R., Kumar, R., & Shakeel, P. M. (2018). An adaptive fuzzy rule based energy efficient clustering and immune-inspired routing protocol for WSN-assisted IoT system. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-018-1154-z>
17. Wang, Y., & Wang, Z. (2019). Routing algorithm of energy efficient wireless sensor network based on partial energy level. *Cluster computing*, 22, 8629–8638.
18. Yuan, P., Wang, Y., Su, M., Yang, Z., & Zhang, Q. (2019). Markov decision process-based routing algorithm in hybrid Satellites/UAVs disruption-tolerant sensing networks. *IET Communications*, 13(10), 1415–1424.
19. Ambigavathi, M., & Sridharan, D. (2018). Energy-aware data aggregation techniques in wireless sensor network. In A. Garg, A. K. Bhoi, S. PadmanabanSanjeevikumar, & K. Kamani (Eds.), *Advances in Power Systems and Energy Management: ETAEERE-2016* (pp. 165–173). Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-10-4394-9\\_17](https://doi.org/10.1007/978-981-10-4394-9_17)
20. Karkazis, P., Leligou, H. C., Sarakis, L., Zahariadis, T., Trakadas, P., Velivassaki, T. H., & Capsalis, C. (2012). Design of primary and composite routing metrics for RPL-compliant wireless sensor networks. In *2012 international conference on telecommunications and multimedia (TEMU)* (pp. 13–18). IEEE.
21. Han, Z., Li, Y., & Li, J. (2018). A novel routing algorithm for IoT cloud based on hash offset tree. *Future Generation Computer Systems*, 86, 456–463.
22. Safara, F., Soury, A., Baker, T., Al Ridhawi, I., & Aloqaily, M. (2020). PriNergy: A priority-based energy-efficient routing method for IoT systems. *The Journal of Supercomputing*, 76(11), 8609–8626.
23. Maheswar, R., Jayarajan, P., Sampathkumar, A., Kanagachidambaresan, G. R., Hindia, M. N., Tilwari, V., & Amiri, I. S. (2021). CBPR: A cluster-based backpressure routing for the internet of things. *Wireless Personal Communications*, 118, 3167–3185.
24. Mohseni, M., Amirghafouri, F., & Pourghebleh, B. (2023). CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic. *Peer-to-Peer Networking and Applications*, 16(1), 189–209.
25. Mohamed, B., & Mohamed, F. (2015). QoS routing RPL for low power and lossy networks. *International Journal of Distributed Sensor Networks*, 11(11), 971545.
26. Krishna, A. V., & Leema, A. A. (2022). Efficient routing algorithm for improving the network performance in internet of things. *International Journal of Internet Protocol Technology*, 15(2), 107–115.
27. Krishna, P. V., Obaidat, M. S., Nagaraju, D., & Saritha, V. (2017). CHSEO: an energy optimization approach for communication in the internet of things. In *GLOBECOM 2017-2017 IEEE Global Communications Conference* (pp. 1–6). IEEE.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



**Akula Vijaya Krishna** pursued his B.Tech from Mekapati RajaMohan Reddy Institute of Technology and Sciences, Udayagiri from JNTUA in 2009 and M.Tech from Annamacharya Institute of Technology and Sciences, Rajampet from JNTUA in 2011. He obtained his Ph.D from Vellore Institute of Technology, Vellore in 2022. He has over 11 years of teaching experience and has 10 research publications to his credit in both National and International Journals. He joined GNITS in 2022 in Computer Science & Technology. His research work focuses on Internet of Things, Wireless ad-hoc networks and Wireless Sensor Networks.



**I. Ravi Prakash Reddy** pursued his B.Tech from Vasavi College of Engineering, Hyderabad in 1994 and M.Tech from Andhra University, Vizagin 1997. He obtained his Ph.D from JNTU Hyderabad in 2011. He has over 25 years of teaching experience and has 30 research publications to his credit in both National and International Journals. Currently 3 researchscholars are pursuing Ph.D under his guidance. He joined GNITS in 2001 and took charge as Head of the department IT in 2011. He is a member of Editorial Board of Journal of Current Trends in Information Technology, and also a member of professional bodies such as CSI, ISTE. His research work focuses on Internet of Things, Data Science, Distributed Systems, Software Agents, and Operating Systems.



**A. Anny Leema** working as an Associate Professor in the department of Smart Computing, School of Information Technology & Engineering (SITE) at VIT University, Vellore. She completed her PhD in the year 2013 with good number of publications indexed by Scopus. She is having around 19 years of teaching experience. Her research areas are data mining, RFID, Wireless sensor networks, E-learning, Machine Learning and IoT. She has edited the book Machine Learning Approaches for Improvising Modern Learning Systems and written five chapters in the research areas e-learning and RFID. Published more than 30 research papers in the International Journals and forty two papers in National and International Conferences. She is a reviewer of several International Journals and Program Committee member of various International Conferences. Life time of IAENG and Indian Science Congress.



**Ramana Kadiyala** is currently working as Associate Professor in Department of Information Technology, Chaitanya Bharathi Institute of Technology (Autonomous), Hyderabad. During 2007–2021, he was an Assistant Professor in the Department of Information Technology at Annamacharya Institute of Technology and Sciences (Autonomous), Rajampet. He has completed his B.Tech. in Information Technology in 2007 from Jawaharlal Nehru Technological University, Hyderabad and he has received his M.Tech. in Information Technology in 2011 from Sathyabhama University, Chennai. He has completed Doctor of Philosophy in November 2019 in SRM University, Chennai, on Load balancing in web server systems. His area of research is Distributed Systems, Parallel and Distributed Systems, Software Defined Networking, Cluster Computing, and Web Technologies. He has published more than 50 papers in reputed International Journals and also attended various international conferences. He is a reviewer of several International Journals and Program Committee member of various International Conferences. He was Life time member of IAENG, IE(I), ISTE and

IACSIT.



**Raman Dugyala** Received Bachelor's degree in Computer Science and Engineering from Jawaharlal Nehru Technological University Hyderabad. Then he obtained his Master's degree in Computer Science and PhD in Computer Science majoring in Web Security: Information Flow Models and Program Design Models, both from Jawaharlal Nehru Technological University Hyderabad. He has also obtained CCNA professional qualification. Currently, he is a Professor at the Faculty of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology. His specializations include Web Security, Networking, Malware Security, Vulnerability Signature generation Algorithms and Machine Learning Algorithms. His current research interests are Information Flow Models, Program Design Models, Network Security, Cloud Security, Vulnerability Detection Models, SCADA Security in Cyber Physical Systems and Smart Grid Computing security. Dr Raman is principal investigator for various research projects funded by DST worth of more than 60 Lakhs. In addition, he filed around fifteen patents in his area of research and is in possession of life

memberships in various leading technical societies. He is a senior member of the IEEE and a member of the ACM.



**K. Prasanna** received Ph.D. in CSE from JNTU-Hyderabad and currently working as Associate Professor in CSE, GNITS, Hyderabad. His research interests Data Mining, Artificial Intelligence, Machine Learning, and Data Science & Analytics. He is a recipient of the AICTE Grant and has guided 2 Ph.D. scholars. He has authored 2 books and published several papers.

## Authors and Affiliations

Vijaya Krishna Akula<sup>1</sup> · I. Ravi Prakash Reddy<sup>1</sup> · A. Anny Leema<sup>2</sup> ·  
Ramana Kadiyala<sup>3,6</sup>  · Raman Dugyala<sup>4</sup> · K. Prasanna<sup>5</sup>

✉ Ramana Kadiyala  
ramana.it01@gmail.com

Vijaya Krishna Akula  
vijayakrishnaakula2022@gnits.ac.in

I. Ravi Prakash Reddy  
irpreddy@gnits.ac.in

A. Anny Leema  
annyleema.a@vit.ac.in

Raman Dugyala  
raman.vsd@gmail.com

K. Prasanna  
prasanna.k642@gmail.com

<sup>1</sup> Department of Computer Science and Technology, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

<sup>2</sup> School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

<sup>3</sup> Department of Artificial Intelligence and Data Science, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

<sup>4</sup> Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

<sup>5</sup> Department of Computer Science and Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

<sup>6</sup> Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon



## **ESRRL – An Energy efficient Secure routing protocol based on Reinforcement learning for MANETs**

**Mrs. G. R Rama Devi**, Research Scholar, Department of Computer Science and Engineering, Osmania University, Hyderabad, Telangana : [ramadevigorree@gmail.com](mailto:ramadevigorree@gmail.com)

**Dr. M Swamy Das**, Professor, Department of Computer Science and Engineering, CBIT, Gandipet, Hyderabad, Telangana : [msdas@cbit.ac.in](mailto:msdas@cbit.ac.in)

**Prof. M. V. Ramana Murthy**, Department of Mathematics, Osmania University, Hyderabad, Telangana, India : [mv.rm50@gmail.com](mailto:mv.rm50@gmail.com)

### **Abstract**

The Mobile Ad Hoc Network, or MANET, is made up of a collection of wireless nodes that are able to offer users the benefits of communication despite the absence of access points or centralized control. When designing the routing protocols for a mobile ad hoc network (MANET), one of the most difficult challenges is to create a network that is both secure and energy efficient in order to achieve the highest possible levels of network performance and safety. In order to handle MANET applications, robust security development is required. Cryptography techniques are highly advantageous for security purposes. However, resource limitations in MANET, makes it difficult to apply any security scheme. Hence, to use cryptography scheme, energy-efficiency must be guaranteed in MANET. Therefore, a special energy efficient and secure routing protocol is of great importance in MANETs. In this paper, a new energy efficient secure routing protocol based on reinforcement learning (ESRRL) is proposed. In this work, for energy efficiency, the network is divided into clusters and each cluster is represented with CHs. The CHs are selected using Firefly algorithm which takes residual energy, distance, link lifetime parameters. A novel routing scheme based on reinforcement learning algorithm is introduced. Multi-objective parameters like delay cost, node distance, energy cost, node mobility, and node degree are considered for routing process. A light-weight novel hybrid key establishment scheme is proposed for optimal key management that requires less computation resources and minimum computation cost. The efficiency of the suggested routing protocol is demonstrated by comparing it to that of some other, more established routing methods. Performance research demonstrates that the suggested protocol has better energy consumption and secure key agreements for network data transmission.

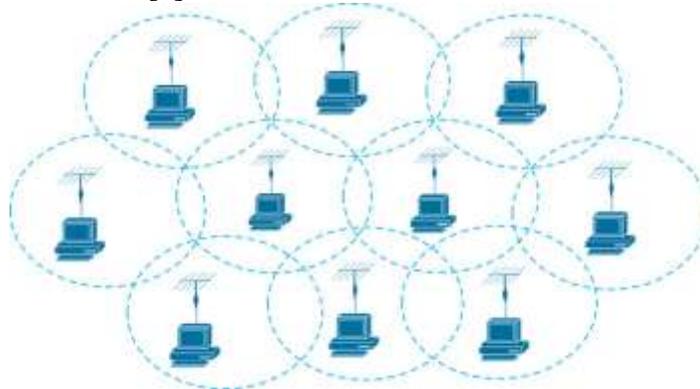
**Keywords:** MANET, Multi-objective, Key management, Energy consumption, Secure communication.

### **I. Introduction**

Mobile ad hoc networks, also known as MANETs, are the wireless networks of the future that will be used as the communication mechanism for all devices. MANETs are able to function despite the lack of any complex design that is often built into wired infrastructure networks [1]. These types of networks are able to connect and communicate with one another regardless of the device or time it was added to the network. Because the nodes in a typical wired network are fixed or stationary, it is not possible to communicate data using mobile nodes in this manner because the traditional wired network does not have mobile nodes. Researchers are focusing their efforts on developing new methods of communication, specifically wireless technologies for the transmission of data [2]. The nodes in a wireless ad hoc network are free to roam anywhere on the network while still being able to communicate with one another, in contrast to the limitations of wired networks. The implementation of the wireless network in a challenging environment is easy, and the deployment of this is carried out in a wasteful amount of time. The simplicity

with which nodes can move freely and rapidly deploy themselves inside the network is one reason why the MANET has found a wide variety of applications [3].

Figure 1 depicts a MANET, which consists of a collection of  $n$  mobile nodes that have been arbitrarily placed in an ad hoc manner, connecting to the wireless medium to build a dynamic network with wireless links in order to provide paths for data transmission within the network. Nodes in MANETs are free to pick their own transmission links, and both existing and new nodes can join or leave the network at any moment. The mobility of the nodes in MANETs necessitates effective routing because of the extensive coverage area of the network [4].



**Fig 1:** Infrastructure-less mobile ad hoc networks

In a mobile ad hoc network (MANET), the mobile nodes are dispersed at random, and each mobile node operates autonomously. As a result, MANET has found use in a wide variety of contexts. The most significant problem with the existing wireless network is the restriction on the amount of battery power, or the amount of energy that can be utilized by the nodes. Other issues include the restricted bandwidth and the restricted transmission range [5]. The mobile nodes are responsible for forwarding packets to their neighboring nodes. Additionally, while the mobile nodes are communicating with other nodes in the network, they use up energy from their batteries. This is due to the fact that the routes are multi-hop and the mobile nodes require energy for transmission and reception within the network. The variance in the energy levels of the nodes, the cost per packet, the maximum node cost, and the time it takes for the network to partition all have the potential to have an effect on the consumption of energy. The minimum energy route is the method by which the overall consumption of energy when delivering a packet is kept to a minimum. The amount of energy that is used in MANETs is primarily determined by [6], which is the amount of time it takes for the network to be partitioned. This means that the network should distribute energy consumption evenly to each of the node links in the network throughout the path providing the minimum amount of energy consumption path, and the path for data transmission should be chosen by the existing routing algorithm that is used with it. In the process of providing the minimum energy cost metrics for each node, the network may end up becoming unbalanced and disrupting its overall functionality. This can happen because some of the nodes have characteristics that try to burden the network's performance and consume more energy than is required.

Due to the sensitive information that MANET works with and monitors, the topic of security has become an increasingly important one in the field of research [7]. It is necessary to develop efficient strategies that guarantee safety and minimize energy waste. MANET needs to be protected from a wide variety of threats, which is why it requires a comprehensive security solution. On the other hand, the amount of MANET's energy resource is often very little. As a consequence of this, the many security protocols cannot be used in a straightforward manner to address the numerous security issues [8]. More resources are required for the most cutting-edge security protocols. Planetary security and cryptographic



protocols are primarily designed for use with traditional wired networks and ad hoc wireless networks. Yet, the most important problem is figuring out how we can implement these protocols on the MANET platform. The question of how much of these protocols should be used in order to guarantee the highest level of safety is a difficult one.

When a mobile node runs out of energy, it can't send packets to other network nodes. However, routing protocols only offer the shortest path, which can have a negative impact on network performance and raise network overhead [9]. MANET necessitates high-quality routing protocols that can simultaneously address a significant number of ad hoc networks. MANETs face a number of challenges and limitations despite their widespread distribution. However, they are impacted by a large number of serious security flaws that are challenging to discover within the network. In order to support sending data packets to the destination, a mobile node needs confidentiality. Security mechanisms based on cryptography are presented to verify the identities of nodes and guarantee the confidentiality and integrity of data [10]. Since a malicious entity without an authorized identity is no longer able to contribute to the network or launch malicious behaviors, it appears that security issues have been resolved. However, these cryptography-based mechanisms have a high energy consumption due to their high computational complexity, which results in increased overhead.

To overcome the above-mentioned issues, in this paper, we propose a new energy efficient secure routing protocol based on reinforcement learning called ESRRL. In this work, the network is divided into clusters and each cluster is represented with CHs. The Firefly algorithm is used for CH selection which takes residual energy, distance, link lifetime parameters. A novel routing scheme based on two hop relay using reinforcement learning algorithm is introduced that takes multi-objective parameters such as delay cost, node distance, energy cost, node mobility, and node degree. A light-weight novel hybrid key establishment scheme is proposed for optimal key management that requires less computation resources and minimum computation cost.

### Contributions

- An improved CH selection using firefly algorithm's cumulative fitness function selects stable CHs which improves data aggregation efficiency which further results in energy efficiency.
- The proposed routing scheme using reinforcement algorithm finds the best shortest route by multiple objectives that satisfies stability and reliability.
- The proposed scheme use both symmetric and asymmetric cryptography with minimized computation cost & overhead.

### Energy model

A routing protocol's primary objective is to keep the network functioning as far as possible, rather than setting up the best routes between the nodes [11]. By using minimum energy routing, the shortest aggregate energy usage for packet transmission is chosen, while by using max-min routing, the route with the greatest bottleneck residual node energy is chosen. The expression of this energy node at a particular moment is provided by

$$E_{en} = E_{res}(t_0) - E_{res}(t_1)$$

Here,  $E_{res}$  represents the energy node at the time 0 and 1, respectively.

## II. Related work

The Trust-aware Ad hoc Routing (T2AR) protocol was suggested by Dhananjayan et al. [12]. This protocol evaluates the trust level between the nodes in the MANET and carries out a secure data transmission between the nodes. It does this by basing its estimation of trustworthiness on factors such as energy, movement, and RSSI-based distance measurement. In order to report the success and failure rate



of packet transmission, this protocol gathers the information from neighbours and uses it. The estimated value of trust is derived from comparing the sequence ID of the packet to the log reports generated by the node. As a result, in the work that has been proposed, the trust value is computed by estimating the energy, the success rate of packet delivery, and mobility. With the utilization of a Bayesian framework and the monitoring of the nodes, direct calculation of trust value may be accomplished.

In their study on network security, Muthuramalingam et al. [13] established a method that calculates trust value based on both direct and indirect observations. This method is utilized to conduct research on network safety. For calculating the trust value based on direct observations, a full probability-based Bayesian interface is utilised; however, while calculating the trust value based on indirect observations, neighbour hop-based information is utilized. The Dempster-Shafter theory, which incorporates both direct and indirect trust values into its calculations, is utilized in order to arrive at the ultimate trust value. When determining the path that is the shortest between two nodes in a network, another approach that is utilised is Dijkstra's shortest path algorithm.

An intrusion detection and prevention system was proposed by Singh et al. [14], and it is designed to be utilized in MANETs for the purpose of locating hostile nodes. This trust-based system recognizes malicious nodes through the use of a behavior classifier, and it does so on the basis of circumstances that have been predefined as thresholds and risk factors. In order to determine the aggregate trust value, both direct and indirect trust values are factored in. These two aspects of trust are what go into the calculation of the risk factor.

Both Hmouda and Li [15] have experienced issues with dropped packets in the network. They have changed the Enhanced Adaptive Acknowledgement (EAACK) protocol so that it is based on hybrid cryptography techniques such as the Data Encryption Standard (DES) and RSA since these approaches can efficiently execute block encryption and manage key cyphers accordingly.

To handle authentication, access control, and security, Hurley-Smith et al. [16] created the Security Utilizing Pre-Existing Routing for MANets (SUPERMAN) framework. Although other protocols provide alternatives, this one concentrates on routing and network layer security issues.

The Modified EAP based Pre-authentication technique using Improved ElGamal (MEPIE) presented by M. Deva Priya et al. [17] enhances the current Enhanced EAP based preAuthentication scheme by employing an improved ElGamal digital signature and ElGamal encryption algorithm. MEPIE improves ElGamal by utilising more random variables to prevent DoS and replay attacks.

Chukka Niranjan Kumar and Nikitha Kukururu [18] proposed simulated Annealing technique. In this scenario, routing will be determined based on the path cost and link cost metric, which in turn is determined by the characteristics of the network and nodes, such as connection availability, network load, battery drain rate, and so on. As a result, the nodes that are chosen for the routing will be appropriate, and this will guarantee that the data are successfully delivered to the destination. As a result of the fact that this method takes into account each and every minute characteristic of the network and node, it is possible to be certain that the majority of the scenarios in which the network fails may be avoided, hence ensuring that the network will function effectively.

### III. System model

In this section, the proposed energy efficient secure routing protocol ESRRL is presented. In ESRRL, the network is divided into clusters and each cluster is represented with CHs. The Firefly algorithm selects CHs for each cluster based on residual energy, distance, link lifetime parameters. A novel routing scheme based on two hop relay using reinforcement learning algorithm is introduced. The multi-objective parameters such as delay cost, node distance, energy cost, node mobility, and node degree are utilized for



routing. A light-weight novel hybrid key establishment scheme is proposed for optimal key management that requires less computation resources and minimum computation cost.

## Clustering & CH selection

### Firefly algorithm (FA)

In the past ten years, there has been a lot of interest in meta-heuristic algorithms, which are algorithms that are inspired by nature. In particular, there has been a lot of interest in algorithms that are based on swarm intelligence. Late in 2007, Xin-She Yang was the one who first presented the firefly algorithm, also known as FA [19]. The actions of the algorithm are determined by optical patterns as well as the actions of fireflies.

There are two factors that are defined in the firefly algorithm. These parameters are the attraction and the variation of the light intensity. The brilliance of a firefly is the most important factor in determining its allure. In addition to this, brightness is defined according to an objective function.

During this stage of development, the network will be clustered with the help of the firefly method. In point of fact, it is often considered that the likelihood of a node in a group of nearby nodes being chosen as a CH is determined by each firefly in the group. The ignition of fireflies occurs at random. As a consequence, each node is thought of as a firefly, and it is possible to choose it as a CH. The Rand function is used to generate a random number that is used to determine the primary attraction level of fireflies  $\beta_0$ . To begin, every node in the network sends a message called "Hello" to the nodes that are immediately adjacent to it. This message includes the node's position information. When a node receives a Welcome message, it immediately updates the information in the table that is adjacent to it. The mobile nodes are able to devise a fitness function for themselves by using the Welcome messages that they have received as input. The definition of the fitness value, or  $F_i$ , of node  $i$  is determined by using three parameters. Where  $1 \leq i \leq N$ , and  $N$  is the total number of nodes in the network:

**Residual energy (RE):** All the time, nodes know exactly how much energy they have. Because CHs need to have adequate energy to send data to the base station, the nodes with the highest energy levels are prioritised for that role.

$$f1 = RE = \frac{1}{\sum_{i=1}^n \frac{E_R}{E_T}}$$

Here,  $E_R$  represents the node energy and  $E_T$  is total energy.

**The average distance (Dist<sub>i</sub>):** According to the coordinates included in Welcome messages, node  $i$  may figure out how far away it is from its neighbours using the Euclidean distance measure. How far away node  $i$  is from the cluster's epicentre is indicated by its Dist. Since cluster members use less energy transmitting data to nodes closer to the cluster's centre, these nodes are more likely to be chosen as CHs. It can be calculated as following:

$$f2 = Dist_i = \frac{\sum_{j=1}^k \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}}{k}$$

Where  $K$  is the number of the neighbors of node  $i$ ,  $(X_i, Y_i)$  denotes the coordinates of node  $i$ , and  $(X_j, Y_j)$  indicates the coordinates of the neighboring node  $j$ .

**Link lifetime:** - The parameter link lifetime (LT) takes into account a number of other parameters in order to make predictions regarding the lifetime of the link (LT). Within each cluster, there is a connection that connects every node to the node that serves as the head of the cluster. This link lifetime



(LT) provides the maximum amount of time that each cluster's data can be transferred. When the value of the link lifetime is increased, the overall lifetime of the network also increases. LT is formulated as follows,

$$f3 = LT = \frac{2*((E_x + E_n))}{T + P_s + P_r + R + \|D_x - D_n\|}$$

The energy levels of the normal node  $x$  and the cluster head  $n$  are denoted by  $E_x$  &  $E_n$ , respectively.  $P_s$  is the rate at which packets are sent out by the  $x^{th}$  node,  $P_r$  is the rate at which packets are received by the  $x^{th}$  node,  $R$  is the transmission range,  $T$  is the value that remains constant, and  $\|D_x - D_n\|$  is the distance between the normal node  $x$  and the cluster head node.

Finally, the fitness value of node  $i$  ( $F_i$ ) is derived as follows:

$$F = \varphi_1 f_1 + \varphi_2 f_2 + \varphi_3 f_3$$

$$\text{where } \sum_{i=1}^3 \varphi_i = 1, \varphi_i \in (0,1)$$

Updating firefly locations is necessary after calculating their fitness value. Hence, the firefly algorithm uses the following equations to determine the fireflies' attraction and their new location.

According to the equation that is provided below, the attraction of the firefly  $\beta$  is determined by the distance ' $d$ ' that is given.

$$\beta = \beta_0 e^{-\gamma d^2}$$

If the firefly  $i$  is drawn towards the firefly  $j$ , then the moment of the firefly can be found by solving the equation that follows.

$$K_i = K_i + \beta_0 e^{-\gamma d^2} (K_i - K_j) + r\omega$$

where  $K$  represents the instant at which the fireflies were observed,  $r$  stands for the randomised variable, and  $\omega$  denotes either the uniform or the Gaussian distribution of the time.

#### Algorithm

1. Do the initialization of the nodes using the parameters.
2. For all nodes  $i$ , where  $i \leq N$
3. Evaluate the fitness  $F_i$ .
4. Compare the nodes' fitness values
5. If  $F_j > F_i$ 
  - Establish the  $F_j$  as the top node at the moment.
  - Provide responsibility for all of the nodes to the  $F_j$
  - Assess the distance  $d$ , and make any necessary adjustments to the attractiveness
6. Continue performing the process until each node in the cluster has determined fitness  $F$
7. Determine which of the nodes will serve as the cluster head.
8. End for

After being designated as a CH, a node will notify its neighbors through broadcast message that it is now acting in that capacity. Upon receiving the advertisement message from member nodes join with the appropriate CHs.

#### Routing based on Reinforcement Learning algorithm

The routing algorithm's primary goal is to extend the lifetime of the network and reduce energy consumption without sacrificing performance. A node's energy depletion can be avoided by taking into account routing parameters. A multi-metric-based Reinforcement Learning strategy for making routing decisions based on the metrics' current values (states and actions) is described in this section. The main goal of the proposed routing scheme is to suggest energy-efficient routing scheme for better performance.



As a means of making informed routing decisions based on estimated path costs, the proposed routing scheme takes into account **delay cost, node distance, energy cost, and node degree as routing metrics**. The best possible way for the information to get to the destination with the fewest resources required is via the optimal route. The relay selection in the proposed work is carried out using the RL algorithm. In this work, each point-to-point communication is modeled as a Markov decision process (MDP), and routing policies that aim to maximize PDR are chosen using a RL approach. The MDP model contains the following set of elements for each node  $N_i$ , where  $i = 1, 2, \dots, n$

$$\{S_i, A_i, T_i, R_i\}$$

Where,  $S_i, A_i, T_i, R_i$  are denoted as states, actions, transition model & rewards.

**States ( $S_i$ ):** The states ( $S_i$ ) for each node,  $N_i$ , are determined by the residual energy, distance, and node mobility values.  $S_i$  here describes the range of potential environmental situations. The relay node is chosen to be one of the set  $N_i$ 's neighbouring nodes (next intermediate node).

- i. **Actions ( $A_i$ ):** This is referred to be a collection of activities or behaviours exhibited by the node during the given time  $t$ . It is possible to transition from one state to another by performing the action. One of the various actions that can be taken on any given node is to (i) choose any one of the available surrounding nodes to operate as a relay node for the transfer of data. (ii) The packet has been finished.

**Transition model ( $T_i$ ):** Here is a description of the state transition probability function in the action-dependent transition model, which shows the probabilities of transitioning from state  $S_i$  to state  $S_i + 1$ , k.

$$T_i = S_i \times A_i \times S_i \rightarrow [0,1]$$

**Reward functions ( $R_i$ ):** A reinforcement function is another name for this concept. The immediate effect of an action, or the change from one state to another, can be evaluated using the reward. It is possible to define the reward function as

$$R_i = S_i \times A_i \times S_i \rightarrow R$$

### Routing process

In order to find the best route, all the nodes need to calculate their probability score based on the following multi-objective parameters

**Energy cost:** During the transmission ( $E_{Tx}$ ) and reception ( $E_{Rx}$ ) of the time-critical event data traffic, node  $i$  incurs an energy cost, denoted by  $eCOST_i$ , as shown in the following equation.

$$E_{cost}(i) = E_{Tx}(l_i, d) + E_{Rx}(l_i)$$

The overall amount of energy required to transmit  $l$  -bits message over a distance  $d$  is given by the following equation

$$E_{Tx}((q_i + l_i), d) = \begin{cases} l_i E_{elec} + l_i \epsilon_{fs} d^2, & d \leq d_0 \\ l_i E_{elec} + l_i \epsilon_{mp} d^4, & d \geq d_0 \end{cases}$$

where  $E_{elec}$  is the amount of energy lost when operating the transmitter or receiver circuit,  $\epsilon_{fs}$  is the amount of energy used for amplification in the free space model,  $\epsilon_{mp}$  is the amount of energy used for amplification in the multi-path model, and  $d_0$  is the minimum transmission distance.

The total amount of power that is required for the radio to transmit ( $l$ ) -bits message is given by the following equation

$$E_{Rx}(q_i) = (q_i)E_{elec}$$

Thus, the energy cost function  $eCOST_i$  is defined as the energy consumed by the routing path from the source node  $i$  to the CH as given in equation



$$eCOST_{itoCH} = \sum e_{cost}(i)/x$$

Where,  $e_{cost}(i)$  is the energy cost of node (i) and  $x$  is the number of relay nodes in routing path.

**Delay cost function:** - The amount of delay that a packet encounters when travelling from node  $i$  to CH is referred to as the link delay cost ( $dCOST_i$ ) of node  $i$ . This cost is expressed as a percentage. A link delay is comprised of three components: the queuing delay  $Q_D$ , the transmission delay  $T_d$ , and the propagation delay  $P_d$ . These components are defined as follows in the equation:

$$dCOST_{itoCH} = Q_D(i, CH) + T_D(i, CH) + P_D(i, CH)$$

$$Q_D(i, CH) = \frac{1}{\alpha - \beta_{CH}}$$

Where  $\alpha$  represents the service rate and  $\beta$  represents the pace at which new packets are added to the CH node

$$T_D(i, CH) = \frac{l_i}{\gamma}$$

where  $l_i$  is the packet size (bits) of at node  $i$ , and  $\gamma$  is the link bandwidth available via connection  $i$  (bps).

$$P_D(i, CH) = \frac{\|i - CH\|}{S}$$

where  $\|i - CH\|$  – is the distance in metres between two points on a network  $i$  to CH, and  $S$  is the propagation speed (m/s).

**Node distance to CH:** - To what extent a communicating node drains its energy supply is determined by its distance from the other nodes. The less time it takes for a node to go from its location to the CH, the less energy it will need. Thus that the average distance between the nodes and CH can be decreased, the routing algorithms account for this variable.

$$Dist_{i2CH} = \sum_{i=1}^N \left( \frac{D_{(N-CH)}}{D_{AVG(N-CH)}} \right)$$

$D_{(N-CH)}$  reflects the node's distance, measured in terms of the CH, according to the Euclidean algorithm whereas,  $D_{AVG(N-CH)}$  indicates the standard deviation of the distance between the node and the CH.

**Node degree:** - When there is a huge amount of network area, the communication within the cluster becomes the predominant entity. If the selection of a node as a relay is carried out independently of the number of nodes that are immediately adjacent to the node in question, the outcome will be the selection of a node that is remote in relation to the other nodes. As a result, in order to prevent this kind of selection from happening, the number of nodes that are nearby is taken into account.

$$ND = \frac{D_{ij}}{NUM_c}$$

$D_{ij}$  represents the distance between the nodes  $i$  and  $j$ ;  $NUM_c$  represents the total number of member nodes in the cluster.

The probability score of the node  $i$  based on the above-mentioned multi-objective selection parameters can be defined as follows:

$$PS_i = eCOST_i + dCOST_i + dist_i + ND_i$$

Based on the calculated High probability score, the actions  $A_i$  are chosen and the state of the nodes is transitioned. The nodes with High probability score are selected as relay nodes.



The algorithm for relay selection based on the reinforcement learning is given in the following

### Algorithm

1. Data transmission phase
2. Initialize  $Q(S_i, A_i) = 0$  for all  $S_i$  &  $A_i$
3. If the destination is found  
Send the packet directly to destination  
Else
4. Calculate  $eCOST, dCOST_i, dist_i, ND_i$
5. Calculate probability score  $PS = \sum_{i=1}^N eCOST + dCOST + dist + ND$
6. Chose an action  $A_i$  based on the high probability score
7. Observe the transition probability
8. Select relay nodes
9. Forward the data packets

### Key management phase

A novel hybrid key management scheme is discussed in this section. The proposed key management scheme is light-weighted and consume less energy, resulting in lower computation costs. The proposed scheme involved asymmetric cryptography operations between the Cluster Head (CH) and the member nodes. The proposed approach is both useful and highly secure for critical environments. The proposed technique is depicted in the block diagram seen in Figure 2.

This section will provide an in-depth explanation of each of the three stages that make up the proposed plan. The initial step, known as key preloading, is when all of the nodes that will be participating in the network load their respective keys. This scheme's second phase, in which secure session keys are established, is the phase that is considered to be the most important. Finally, in the 3<sup>rd</sup> phase, the rekeying operation has been carried out in the event of an error or a change in the network.

**Keys Preloading:** The first phase of the proposed scheme is the key preloading. In this phase, each node  $i$  is loaded with a public and private key

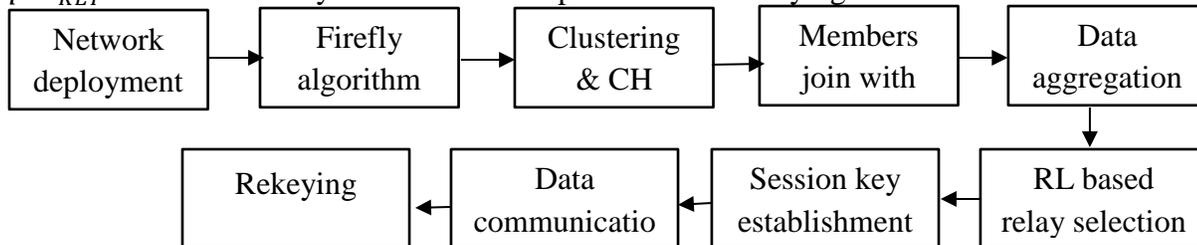
**Session Key Establishment Phase:** Session key establishment is the important phase of the proposed scheme. Following are the algorithmic steps taken for session key establishment in this phase:

1. Mobile nodes generates a unique random number  $Rand_i$  and it is considered as a pre-session key, i.e.,  $pre_{KEY}$
2. Member nodes then encrypts  $pre_{KEY}$  the public key that was used to generate the hash value.
3. CH decrypts the information after receiving it since it knows the private key  $PR_{CH}$ , i.e.,  $(pre_{KEY} || h(pre_{KEY}))$  here, the integrity of  $pre_{KEY}$  is also checked.
4. Following this, CH creates a session key as  $S_{KEY} = pre_{KEY} XOR RAND_i$ .
5. CH then encrypts this session key  $S_{KEY}$  and the hash of that session key  $h(S_{KEY})$  using the  $pre_{KEY}$  and sending that information to the member node.
6. Then the member node decrypts the generated session key using pre-session key  $pre_{KEY}$ , and the integrity of the session key has been checked.
7. Once the node gets the information about the session key, it encrypts the session key  $S_{KEY}$  combines it with a hash of session key  $h(S_{KEY})$  and sends them to all nodes in the cluster.



8. In order to ensure that the session key for the entire network is communicated in a safe manner, the information that has been received is decrypted and recovered by the node  $i$  using symmetric decryption with a key.

**Rekeying Phase:** Rekeying phase operations are comparable to session key setup phase operations. The main difference is that each node produces a fresh random number, like  $rand$ , during the rekeying step. The new randomly generated integer is therefore regarded as the new pre-session key,  $pre_{KEY}'$ . The session key establishment step and the full rekeying method are identical.



**Fig 2:** Block diagram of proposed method

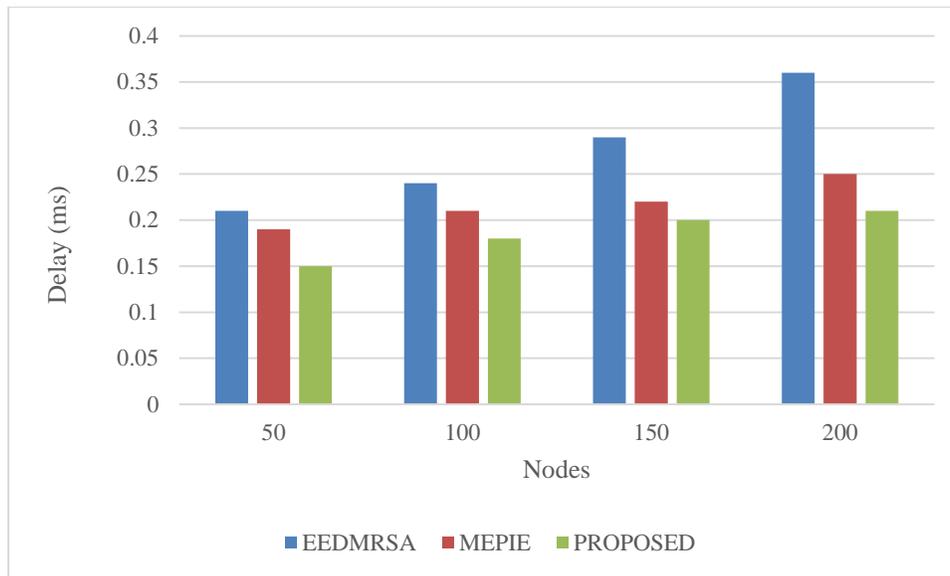
#### IV. Result and Discussion

##### Simulation analysis

The suggested approach is simulated in a 1000 m x 1000 m environment with 50–200 mobile nodes to assess efficiency. The experiment uses the Random Waypoint Mobility Model with pauses between direction and speed changes. After staying at its previous position until expiration, a mobile node chooses a random destination. A node travels over the area at a random speed distributed uniformly from  $v_0$  to  $v_{max}$ , the least and maximum node velocities. Repeat this random location at random velocity until the simulation is complete. Table 1 represent the simulation parameters of network.

**Table 1:** Parameters of Simulation

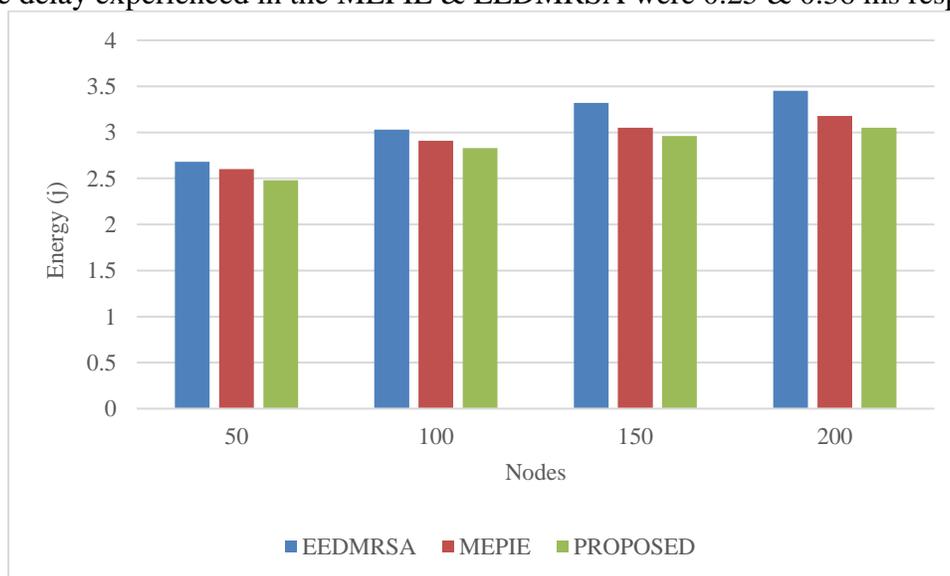
Parameter	Value
Network area	1000 m x 1000 m
Number of nodes	50 to 200
Transmission type	CBR
Initial energy	100j
Packet size	1024 bytes
Routing protocol	AODV
mobility pattern	Random way point model



**Fig 3: End to End Delay**

Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	0.15	0.19	0.21
2	100	0.18	0.21	0.24
3	150	0.20	0.22	0.29
4	200	0.21	0.25	0.36

The above figure 3 demonstrates the evaluation of end-to-end delay time for the data transmission. The minimum average delay experienced in the proposed network was 0.15ms whereas the maximum of 0.21ms delay was experienced when the network size was increased to 200 nodes. The relay selection using the proposed reinforcement algorithm & consideration of multi-objective parameters select the minimum hop path and to achieve the minimum time delay compared with the existing methods. The maximum time delay experienced in the MEPIE & EEDMRSA were 0.25 & 0.36 ms respectively.

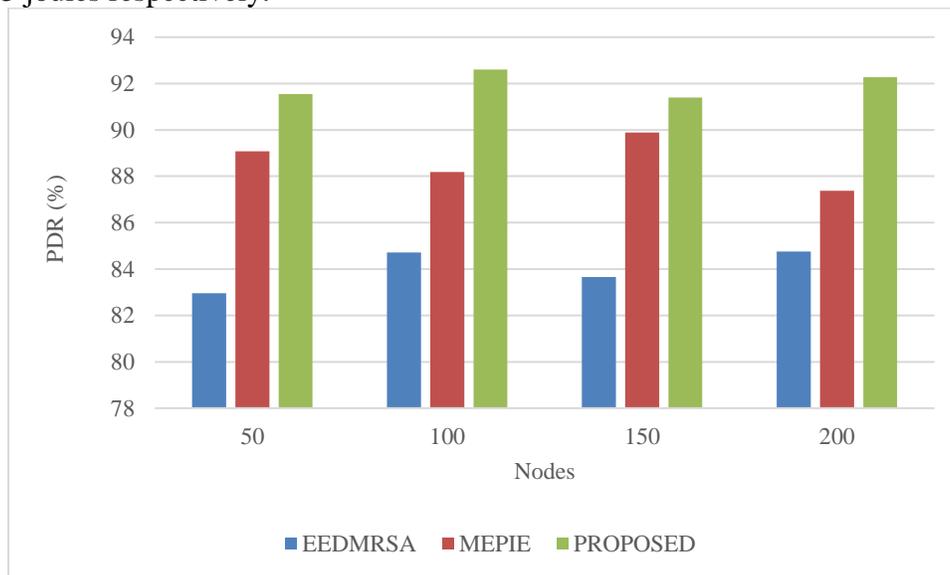


**Fig 4: Energy consumption**



Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	2.48	2.60	2.68
2	100	2.83	2.91	3.03
3	150	2.96	3.05	3.32
4	200	3.05	3.18	3.45

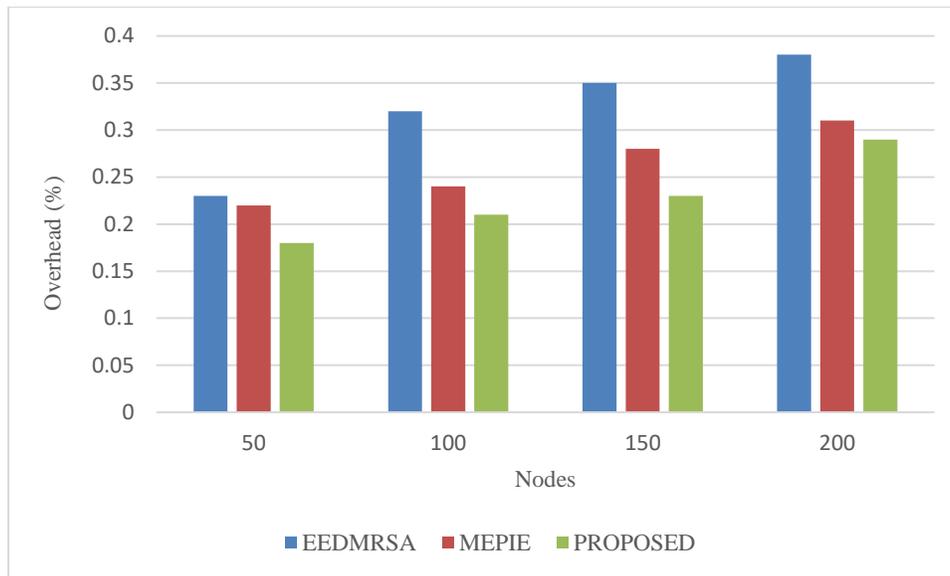
The above simulation results figure 4 show the energy consumption rate of the both existing & proposed methods. All the mobile nodes are homogeneous and configured with same energy (100joules) levels initially. Generally, energy is consumed for all network activities. In MANETs, path loss during node mobility is the major source of energy consumption. Also, the creation and distribution of cryptography keys consumes considerable amount of energy. In the proposed ESRRL method, optimal relay nodes are selected as well as light-weight cryptography technique was used which results in lesser energy consumption than the previous protocols. In the experimental results, average energy consumed in the proposed network was 2.830 joules. But in the MEPIE & EEDMRSA the average energy consumed was 3.032 & 3.1175 joules respectively.



**Fig 5: Packet delivery ratio**

Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	91.54	89.08	82.96
2	100	92.60	88.19	84.72
3	150	91.40	89.88	83.65
4	200	92.27	87.38	84.75

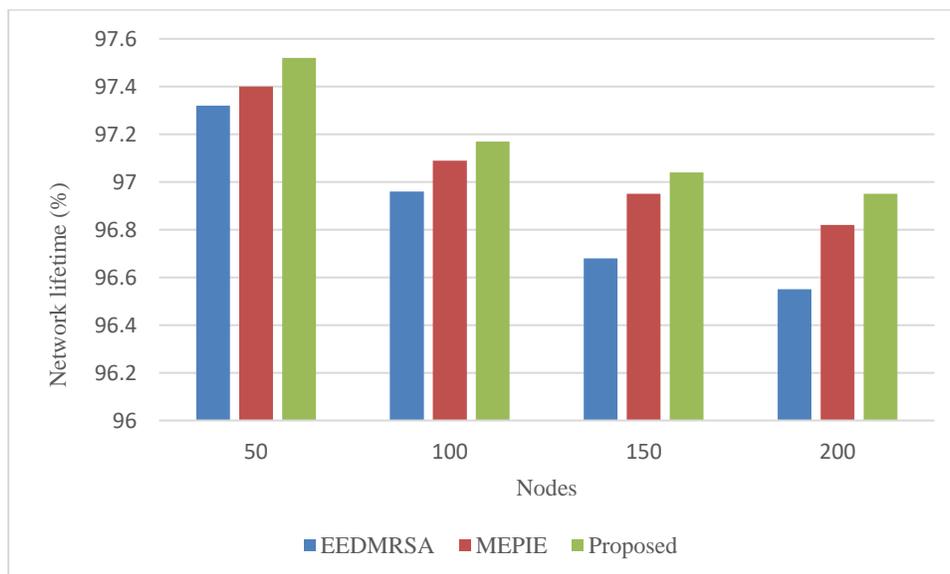
The above figure 5 demonstrates the simulation results of packet delivery ratio for the proposed & existing methods. The performance of the network in terms of data transmission is characterized by PDR. The proposed method achieved maximum PDR of 92.27% whereas the existing methods maintained the average PDR rate of 87 & 84% respectively. The proposed reinforcement algorithm-based relay selection using multi-objective parameters helped the proposed method to select efficient path for communication and helps to outperform than the existing methods.



**Fig 6: Routing overhead**

Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	0.18	0.22	0.23
2	100	0.21	0.24	0.32
3	150	0.23	0.28	0.35
4	200	0.29	0.31	0.38

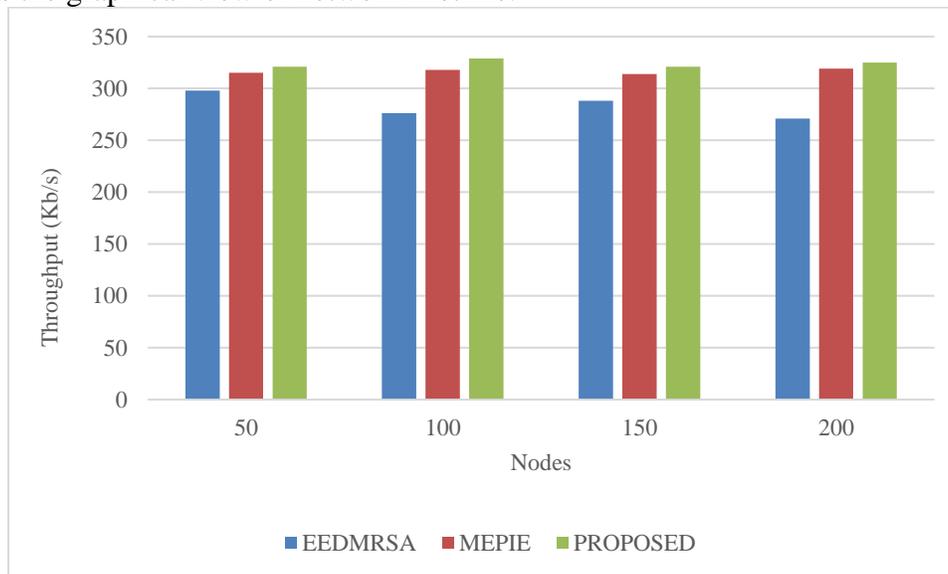
The simulation results of the network's overhead are shown in figure 6 above. The overhead depends on how many control packets are transmitted throughout the network to carry out network operations. Approximately 0.26% of overhead was recorded in the proposed method and MEPIE, EEDMRSA protocols experienced around 0.31 & 0.38 respectively. The proposed relay selection algorithm & the light-weight key management scheme reduces the network overhead. Also, consideration of multi-objective parameters like delay cost, node degree in relay selection reduces and chance of path failure and reducing frequent broadcasting of control packets. Hence, the suggested network maintains a low overhead.



**Fig 7: Network lifetime**

Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	97.52	97.40	97.32
2	100	97.17	97.09	96.96
3	150	97.04	96.95	96.68
4	200	96.95	96.82	96.55

Generally defined as the time during which the network is operational or the time until the first node or group of sensor nodes in the network runs out of energy. In our experiment the network lifetime is related with the remaining energy availability of the network after performing the network activities. More remaining energy means more lifetime to the network. The values in the following table demonstrate that, in comparison to the existing methods, the proposed strategy significantly increases network longevity. Figure 7 shows the graphical view of network lifetime.

**Fig 8: Throughput**

Sl.No	NODE	PROPOSED	MEPIE [17]	EEDMRSA [18]
1	50	321	315	298
2	100	329	318	276
3	150	321	314	288
4	200	325	319	271

The above results figure 8 show the simulation results of Throughput of both proposed & existing methods. Throughput is data transmission between sensor nodes. High throughput ensures the high amount of data delivery. The above listed table proves that the proposed method gives high throughput rate compared with existing methods. The proposed technique achieved an average throughput rate of 324kbps, while the MEPIE and EEDMRSA methods achieved 316 and 283kbps, respectively.

### Conclusion

A significant obstacle is in the provision of routing algorithms that are both secure and environmentally friendly for networks that have a dynamic topology and a distributed nature. MANETs make use of a



variety of routing strategies in order to enhance the performance of the network and lower the amount of energy that is consumed. Because of the constraints imposed by individual nodes, routing is actually one of the most difficult difficulties presented by MANETs. To tackle these issues, in this paper, a new reinforcement learning based energy-efficient and secure routing protocol (ESRRL) is proposed. For energy efficiency, the network is divided into clusters in this work, and CHs are used to represent each cluster. The Firefly algorithm, which considers link lifetime, residual energy, and distance, is used to select the CHs. An innovative reinforcement learning-based relay routing strategy is presented. The routing process takes into account multi-objective parameters like delay cost, node distance, energy cost, node mobility, and node degree. For optimal key management, a novel, lightweight hybrid key establishment scheme with minimal computation cost and use of compute resources is proposed. The proposed energy-efficient routing protocol is compared to other well-known routing protocols to validate its performance. The proposed protocol performs better in terms of energy consumption and secure key agreements for network data transmission, according to the performance analysis.

### References:

- [1] Laqtib, S., El Yassini, K. and Hasnaoui, M.L., 2020. A technical review and comparative analysis of machine learning techniques for intrusion detection systems in MANET. *International Journal of Electrical and Computer Engineering*, 10(3), p.2701.
- [2] Elmeadawy, S. and Shubair, R.M., 2020. Enabling technologies for 6G future wireless communications: Opportunities and challenges. *arXiv preprint arXiv:2002.06068*.
- [3] Agrawal, R., Faujdar, N., Romero, C.A.T., Sharma, O., Abdulsahib, G.M., Khalaf, O.I., Mansoor, R.F. and Ghoneim, O.A., 2022. Classification and comparison of ad hoc networks: A review. *Egyptian Informatics Journal*.
- [4] Al-Absi, M.A., Al-Absi, A.A., Sain, M. and Lee, H., 2021. Moving ad hoc networks—A comparative study. *Sustainability*, 13(11), p.6187.
- [5] Zhang, S., Kong, S., Chi, K. and Huang, L., 2021. Energy management for secure transmission in wireless powered communication networks. *IEEE Internet of Things Journal*, 9(2), pp.1171-1181.
- [6] Prasad P, R. and Shankar, S., 2020. Efficient Performance Analysis of Energy Aware on Demand Routing Protocol in Mobile Ad-Hoc Network. *Engineering Reports*, 2(3), p.e12116.
- [7] Sharma, M. and Rashid, M., 2020, April. Security Attacks In MANET—A Comprehensive Study. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.
- [8] Hasan, M.K., Islam, S., Sulaiman, R., Khan, S., Hashim, A.H.A., Habib, S., Islam, M., Alyahya, S., Ahmed, M.M., Kamil, S. and Hassan, M.A., 2021. Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access*, 9, pp.47731-47742.
- [9] Kumari, P. and Sahana, S.K., 2022. PSO-DQ: An Improved Routing Protocol Based on PSO using Dynamic Queue Mechanism for MANETs. *Journal of Information Science & Engineering*, 38(1).
- [10] Khashan, O.A. and Khafajah, N.M., 2023. Efficient hybrid centralized and blockchain-based authentication architecture for heterogeneous IoT systems. *Journal of King Saud University-Computer and Information Sciences*.
- [11] Srilakshmi, U., Veeraiyah, N., Alotaibi, Y., Alghamdi, S.A., Khalaf, O.I. and Subbayamma, B.V., 2021. An improved hybrid secure multipath routing protocol for MANET. *IEEE Access*, 9, pp.163043-163053.
- [12] Dhananjayan G, Subbiah J (2016) T2AR: trust-aware ad-hoc routing protocol for MANET. Springerplus 5(1):995.



- [13] Muthuramalingam S, Suba Nachiar T (2016) Enhancing the security for manet by identifying untrusted nodes using uncertainty rules. *Indian J Sci Technol* 9(4). <https://doi.org/10.17485/ijst/2016/v9i4/87043>.
- [14] Singh O, Singh J, Singh R (2017) An intelligent intrusion detection and prevention system for safeguard mobile adhoc networks against malicious nodes. *Indian J Sci Technol* 10(14). <https://doi.org/10.17485/ijst/2017/v10i14/110833>.
- [15] Larafa, S., & Laurent, M. (2010). Authentication protocol runtime evaluation in distributed AAA framework for mobile ad-hoc networks. In *Proceedings of the IEEE international conference on wireless communications, networking and information security* (pp. 277–281).
- [16] Larafa, S., & Laurent-Maknavicius, M. (2009). Protocols for distributed AAA framework in mobile ad-hoc networks. In *Proceedings of the workshop on mobile and wireless networks security* (pp. 75–86).
- [17] Deva Priya, M., Janakiraman, S., Sandhya, G. and Aishwaryalakshmi, G., 2021. Efficient pre-authentication scheme for inter-ASN handover in high mobility MANET. *Wireless networks*, 27(2), pp.893-907.
- [18] Kumar, C.N. and Kukunuru, N., 2021. Energy efficient disjoint multipath routing protocol using simulated annealing in MANET. *Wireless Personal Communications*, 120(2), pp.1027-1042.
- [19] Yang, X.S.: *Nature-inspired metaheuristic algorithms*. Luniver Press (2008).

# A Semantic Approach to Solve Scalability, Data Sparsity and Cold-Start Problems in Movie Recommendation Systems

<sup>1</sup>T. Suvarna Kumari, <sup>2</sup>K. Sagar

Submitted:13/02/2023

Revised:16/04/2023

Accepted:10/05/2023

**Abstract:** Recommender systems play a vital role in providing users with personalized information and enhancing their browsing experiences. However, despite the advancements in collaborative filtering techniques, several challenges persist in movie recommendation systems, including the cold start problem, scalability limitations, and data sparsity. The cold start problem arises when there is insufficient data to establish connections between users and items, resulting in inaccurate recommendations. Data sparsity further complicates the issue by making it difficult to identify reliable similar users due to the limited ratings provided by active users. Scalability poses yet another challenge, as real-time environments with a high number of users and extensive data processing requirements struggle to deliver efficient recommendations. To address these issues, this paper proposes a semantic approach that leverages singular value decomposition (SVD), a matrix factorization technique. By applying SVD, the system reduces the dimensionality of the data, overcoming the limitations of the cold start problem, scalability, and data sparsity. Experimental results demonstrate the effectiveness of the proposed system, showcasing improved recommendation accuracy and the ability to generate reliable suggestions even in situations with limited data. Moreover, the system showcases scalability by efficiently processing large volumes of data in real-time, ensuring seamless user experiences. Overall, this semantic approach offers a comprehensive solution to tackle the challenges of scalability, data sparsity, and the cold start problem in movie recommendation systems, potentially enhancing user satisfaction and recommendation quality.

**Keywords:** Scalability, Data sparsity, Cold-start problem, Singular Value Decomposition.

## 1. Introduction

The field of movie recommendation systems has garnered substantial attention in recent years. Exponentially, the internet is expanding due to the proliferation of available data. Amidst exploring websites and navigating the vast online landscape, users demand personalized and high-quality information more than ever before. By utilizing user preferences, search histories, and other factors, recommender systems are important in fulfilling this requirement. They offer personalized recommendations. To enhance the precision and effectiveness of these suggestions, collaborative filtering techniques [1] have been broadly adopted. Nevertheless, despite multiple methods and techniques, various obstacles remain in motion picture suggestion systems. Challenges comprise obstacles such as starting difficulties in low temperatures, scalability limitations, and a paucity of data.

The cold start problem [2] creates a fundamental challenge for recommendation systems. The lack of adequate data to establish relevant connections between users and items results in this phenomenon. By

introducing new users or items, there is insufficient information to produce accurate recommendations. Less personalized or relevant recommendations may be a consequence at first. The system may experience problems offering important suggestions when historical data about user preferences or item characteristics is not present. The outcome is a subpar user experience. The cold start problem is quite frequent in dynamic environments that continuously add users and items. Effectively handling this challenge demands the implementation of robust solutions. Movie recommendation systems encounter another significant issue: data sparsity. Acquiring dependable and comprehensive ratings is difficult because of the huge number of movies available and the vast user base. This includes all potential combinations of users and items. Sparse user-item matrices result from active users usually rating only a small portion of available items. Finding similar users with sufficient overlap in preferences is difficult for the system due to this sparsity. Dependable recommendations rely on this being in place. Movie recommendation systems' performance and effectiveness are undermined due to data sparsity.

Another important issue in recommendation systems is scalability. In situations where there is a high number of users and significant data processing requirements in real-time, this particularly stands out. The increasing demand for recommendations necessitates that the

<sup>1</sup>Research scholar at Osmania University,  
Assistant Professor, Computer Science and Engineering Department,  
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India.  
Email ID: [suvarnakumari\\_cse@cbit.ac.in](mailto:suvarnakumari_cse@cbit.ac.in)

<sup>2</sup>Professor and vice principal, Sreyas institute of engineering and  
technology, Email ID: [Sagar.k@sreyas.ac.in](mailto:Sagar.k@sreyas.ac.in)

system efficiently handle expanding user bases and data volumes. It is imperative that the system can conform and grow with evolving necessities. The requirement to quickly process and analyze huge amounts of data creates scalability challenges. Promptly and seamlessly delivering recommendations becomes challenging as the user base of the system expands. An efficient and productive movie recommendation system is imperative to tackle the challenges of the cold start problem, scalability limitations, and data sparsity. The system must be capable of offering precise suggestions despite limited or no user or item insights. Enhancing user satisfaction and improving the accuracy of recommendations are possible if these challenges are overcome. This could eventually increase user engagement and business growth.

Proposing a semantic approach could potentially advance the field of movie recommendation systems, making the significance of this research noteworthy. This approach handles the aforementioned obstacles. To reduce data dimensionality while mitigating scalability limitations, the cold start problem, and data sparsity issues, the proposed solution uses a matrix factorization approach relying on the singular value decomposition (SVD) technique. The system's performance is improved by utilizing this approach. Applying SVD can efficiently capture latent factors and patterns in user-item interactions. Even with limited or sparse data, this remains a truth. This approach can potentially improve recommendation accuracy. Even without abundant data, it can offer trustworthy recommendations to its users. This research benefits from having two important contributions. Presented initially is a comprehensive analysis and understanding of the challenges faced by movie recommendation systems. The main focus is on the cold start problem, scalability limits, and data scarcity. This research furthers the existing understanding by emphasizing these challenges. Those working in the field, including researchers and practitioners, may find its valuable reference helpful.

In addition, a distinct semantic method that leverages singular value decomposition (SVD) is presented to overcome the challenges highlighted in this research. SVD [6] facilitates the reduction of dimensionality and the extraction of latent factors. Accurate recommendations become easier when data is limited and user-item matrices are sparse. The proposed approach is evaluated for its effectiveness and performance through extensive experiments and comparisons with existing methods. The cold start problem, scalability limitations, and data sparsity are effectively addressed with this approach, as shown by empirical evidence.

Overall, this research aims to contribute to the field of movie recommendation systems by providing a thorough analysis of the challenges faced in the domain and proposing a novel semantic approach to address these challenges. The potential impact of this research includes improved recommendation accuracy, enhanced user satisfaction, and increased user engagement, which can benefit both users and businesses operating in the movie recommendation domain. By offering a solution to the cold start problem, scalability limitations, and data sparsity, this research has the potential to advance the state-of-the-art in movie recommendation systems and pave the way for further innovations in the field.

## 2. Related Work

Recent works relating to approaches for making recommendations will be discussed in this section and the purpose of these new methodologies is to tackle obstacles that earlier strategies - namely Bayesian inference [6], Latent Semantic Analysis (LSA) [7], Clustering Algorithm [8] Regression-Based Approaches and Matrix Factorization Methods (9)- had previously encountered. When it comes to collaborative filtering methods it is safe to say that the majority prefers using the MD algorithm[10] and this particular algorithm is capable of transforming both users and items into corresponding feature vectors with matching dimensions that capture their underlying properties. On-singular value decomposition but Singular Value Decomposition (SVD) [11], Probabilistic Matrix Factorization(PMF) [12] are among the representative works that employ this algorithm, however the effectiveness of MD algorithms can be limited by highly sparse rating matrices which makes it challenging for them to accurately learn appropriate latent vectors.

Recent advancements in deep learning approaches have led to impressive outcomes in the domain of collaborative filtering [13] and the restricted Boltzmann machine approach [13] was among the first to utilize deep learning. A collaborative deep learning framework that intertwines the functionality of stacked denoising autoencoder and PMF models is presented in the study by Wang et al and [14] In their study [15], Xue et al a deep mining of related characteristics necessitates the use of a multi-layer feedforward neural network and to gauge a predicted rating using specific lower-dimensional features, the recommendation system employs an inner-product computation by Zhang et al.'s [16] approach to improving recommendation accuracy involves using a combination of learned video data features from an autoencoder along with implicit feedback gathered via SVD++ and this technique is known as Auto-SVD++, and has been shown to be effective.

The researchers at Ouyang et al named their collaborative filtration technology - Auto Encoder-Based Collaborative Filterer or simply ACF during their experiments. Breaking down a user's rating score for an object into five distinct vectors lies at the core of this approach and predicting accurate integer scores through ACF method is feasible but its drawback lies with a lack of consideration towards sparse scoring matrices which may impact predictive capabilities. Sedhain and colleagues introduced a method called AutoRec [18] that aims at recreating the initial input dataset with high accuracy. Even though this technique solves computing decimal-free rating quantities problem but omits adding random perturbation in its inputs which could make it more resilient and lessen likelihood of overfitting. A predictive model called CDAE was introduced by Wu and colleagues [19], which utilizes a user's unexpressed feedback information about products to produce rankings and the input part of this model contains several perceptrons each corresponding to an individual item representing how much interested a user is on it ranging from values zero and one. Based on predictions made by output layer perceptron's from within a model, related product recommendations can be served up one at a time to users.

The issue of cold start is addressed in Yan et al's paper [20] where they report that both the ACF and AutoRec models suffer from this problem. The CFN framework was suggested by Strub et al. It joins together both content-based data and a score matrix in order to present precise prediction results and this improved model has exhibited higher predictive power than the former recommendation systems. As mentioned by Yan et al [21], CFN model faces limitations due to lack of complexity and sparsity of available data despite being renowned for its strength in applying a content-based approach for recommendations.

### 3. Methodology

In order to achieve the twofold contribution of this research, a comprehensive methodology is employed to address the challenges faced by movie recommendation systems, namely the cold start problem, scalability limitations, and data sparsity. This methodology encompasses the following steps:

#### 3.1 Data Collection and Preparation

The MovieLens 20M dataset is a widely used dataset in the field of movie recommendation systems. It is a rich and extensive dataset that contains a large collection of movie ratings and user-item interactions. The dataset consists of approximately 20 million ratings given by users to movies, hence the name "MovieLens 20M." These ratings are provided by a diverse group of users,

reflecting various tastes and preferences. The dataset also includes information about the movies themselves, such as genres, release year, and tags associated with each movie.

The MovieLens 20M dataset offers a representative sample of real-world movie recommendation scenarios, providing a valuable resource for researchers and practitioners in the field. The dataset captures a wide range of user behaviors and preferences, enabling the analysis and evaluation of different recommendation algorithms and approaches. In terms of data preparation, the MovieLens 20M dataset typically requires minimal preprocessing. However, it is important to ensure data consistency and integrity before conducting any analysis or implementing recommendation algorithms. This may involve checking for missing values, removing duplicates, and handling outliers or noisy data if necessary. Additionally, the dataset can be divided into training and testing sets to evaluate the performance of recommendation algorithms accurately. The training set is typically used to train the recommendation models, while the testing set is employed to assess the accuracy and effectiveness of the models in generating recommendations.

The MovieLens 20M dataset has been extensively utilized in research studies and serves as a benchmark for evaluating the performance of movie recommendation systems. Its large size and comprehensive nature allow researchers to explore various aspects of recommendation algorithms, including addressing the challenges of the cold start problem, scalability limitations, and data sparsity. Overall, the MovieLens 20M dataset provides a valuable resource for researchers and practitioners to conduct experiments, validate algorithms, and contribute to the advancement of movie recommendation systems. Its realistic and diverse nature makes it an ideal choice for analyzing the challenges faced by recommendation systems and proposing innovative solutions.

#### 3.2 Comprehensive Analysis

In order to present a comprehensive analysis and understanding of the challenges faced by movie recommendation systems, particularly the cold start problem, scalability limitations, and data sparsity, a mathematical model can be employed. This model helps in quantifying the challenges and evaluating their impact on recommendation accuracy. Let's delve into the details:

##### 1. Cold Start Problem:

The cold start problem occurs when there is insufficient data available to establish meaningful user-item associations. To address this problem, the model can define a measure of similarity between users or items

based on available attributes or metadata. Let's denote the similarity between user  $u$  and item  $i$  as  $Sim(u, i)$ . This similarity can be computed using various techniques such as content-based filtering, collaborative filtering, or hybrid methods. One way to quantify the cold start problem is by measuring the sparsity of the user-item matrix. Let's denote the user-item matrix as  $R$ , where each entry  $R[u, i]$  represents the rating given by user  $u$  to item  $i$ . The sparsity of the matrix can be calculated using the following equation:

$$Sparsity = \frac{(Number\ of\ Missing\ Entries)}{(Total\ Number\ of\ Entries)}$$

A higher sparsity value indicates a more severe cold start problem, as there are numerous missing ratings in the user-item matrix.

$$Cosine\ Similarity(u, i) = \frac{(\Sigma u * \Sigma i)}{(\sqrt{\Sigma u^2} * \sqrt{\Sigma i^2})}$$

Here,  $u$  represents the user's preference vector, and  $i$  represents the item's content vector. The cosine similarity measure calculates the cosine of the angle between the user preference vector and the item content vector, indicating their similarity.

## 2. Scalability Limitations:

Scalability limitations in movie recommendation systems arise due to the increasing number of users and the large volume of data that needs to be processed. One approach to address scalability is matrix factorization, which reduces the dimensionality of the user-item interaction matrix and captures latent factors. Singular value decomposition (SVD) is a commonly used matrix factorization technique. It decomposes the user-item interaction matrix into three matrices:  $U$ ,  $\Sigma$ , and  $V$ .

$$User\text{-}Item\ Matrix \approx U * \Sigma * V^T$$

Here,  $U$  represents the user matrix,  $\Sigma$  represents the diagonal matrix of singular values, and  $V$  represents the item matrix. By reducing the dimensionality of the matrix, scalability can be improved, allowing faster processing and recommendation generation.

To handle scalability, the model can employ techniques like singular value decomposition (SVD) or non-negative matrix factorization (NMF) to efficiently compute the lower-rank approximation of the user-item matrix. These techniques reduce the computational complexity and enable faster processing of recommendations in real-time environments.

## 3. Data Sparsity:

Data sparsity occurs when it becomes difficult to find reliable similar users or items due to the limited ratings provided by active users. To address data sparsity, the

model can utilize collaborative filtering techniques to find similarities between users or items and make recommendations based on these similarities.

One common approach is neighborhood-based collaborative filtering, where similar users or items are identified based on their ratings. Let's denote the set of similar users to user  $u$  as  $N(u)$ . The similarity between users can be computed using various metrics such as cosine similarity or Pearson correlation coefficient. The recommendation for user  $u$  can then be generated by aggregating the ratings of similar users:

$$\text{Predicted Rating for user } u \text{ and item } i = \frac{\sum (Sim(u, v) * R[v, i])}{\sum Sim(u, v)}$$

Here,  $Sim(u, v)$  represents the similarity between users  $u$  and  $v$ , and  $R[v, i]$  represents the rating of user  $v$  for item  $i$ .

$$Pearson\ Correlation(u, v) = \frac{\Sigma((Ru - \bar{Y}u) * (Rv - \bar{Y}v))}{(\sqrt{\Sigma(Ru - \bar{Y}u)^2} * \sqrt{\Sigma(Rv - \bar{Y}v)^2})}$$

Here,  $Ru$  represents the ratings given by user  $u$ ,  $\bar{Y}u$  represents the mean rating of user  $u$ ,  $Rv$  represents the ratings given by user  $v$ , and  $\bar{Y}v$  represents the mean rating of user  $v$ . The Pearson correlation coefficient measures the linear correlation between the ratings of users  $u$  and  $v$ .

By incorporating these mathematical models and equations, the comprehensive analysis and understanding of the challenges faced by movie recommendation systems, including the cold start problem, scalability limitations, and data sparsity, can be quantified and evaluated. These models provide a framework to assess the impact of these challenges on recommendation accuracy and guide the development of effective solutions.

By employing these mathematical models and equations, the methodology provides a comprehensive analysis of the challenges faced by movie recommendation systems. It addresses the cold start problem through content-based filtering, scalability limitations through matrix factorization, and data sparsity through neighborhood-based collaborative filtering. These mathematical formulations enable a deeper understanding of the challenges and lay the foundation for proposing innovative solutions to enhance recommendation accuracy and system performance in movie recommendation systems.

**Experimental Design:** A set of experiments is designed to evaluate the impact of the identified challenges on the

performance of existing movie recommendation systems. Multiple metrics, such as recommendation accuracy, coverage, and diversity, are selected to measure the effectiveness of different approaches in addressing the challenges.

### 3.3 Data Collection and Preprocessing:

Preprocessing the MovieLens 20M dataset involves handling missing values, removing duplicates, and addressing data inconsistencies or noise to ensure the integrity and accuracy of the data. Let's discuss each step in detail, along with mathematical models and equations where applicable:

#### 1. Handling Missing Values:

- Missing values in the dataset can occur when users have not provided ratings for certain movies. One approach to handle missing values is to impute them with appropriate values. One commonly used imputation technique is mean imputation, where missing ratings are replaced with the mean rating of the respective user or item.

Mathematical Model:

- Mean Imputation for User  $u$ : If a user  $u$  has missing ratings  $(R_{u,m})$  for movie  $m$ , the imputed rating  $(R_{imputed\_u,m})$  can be calculated as follows:  $R_{imputed\_u,m} = \text{Mean}(\text{Ratings}_u)$
- Mean Imputation for Item  $m$ : If an item  $m$  has missing ratings  $(R_{u,m})$  by user  $u$ , the imputed rating  $(R_{imputed\_u,m})$  can be calculated as follows:  $R_{imputed\_u,m} = \text{Mean}(\text{Ratings}_m)$

#### 2. Removing Duplicates:

- Duplicates in the dataset can occur when the same user rates the same movie multiple times or due to data entry errors. Removing duplicates ensures that each user-movie rating is unique and eliminates redundancy.

Mathematical Model:

- Duplicate Removal: For each user  $u$  and movie  $m$ , if there are multiple ratings  $(R_{u,m})$  available, remove the duplicate ratings and keep only the unique rating.

#### 3. Addressing Data Inconsistencies and Noise:

- Data inconsistencies or noise in the dataset can arise due to various factors, such as outliers, incorrect ratings, or data entry errors. It is important to identify and address these issues to ensure accurate analysis and reliable recommendations.
- Outlier Detection and Removal: Statistical techniques like z-score or interquartile range (IQR) can be

employed to identify outliers in the ratings. Outliers can be removed or adjusted to mitigate their impact on the analysis.

- Data Consistency Checks: Perform consistency checks to identify any inconsistent or incorrect ratings. For example, ratings that are outside the valid rating range or ratings provided by users who have not engaged with the system for an extended period can be flagged for further investigation or removal.

### Algorithm: Algorithm: Preprocess MovieLens 20M Dataset

**Input:** MovieLens 20M dataset **Output:** Cleaned and preprocessed dataset

#### 1. Handle Missing Values:

- For each user  $u$  and movie  $m$  in the dataset:
- If the rating  $R_{u,m}$  is missing (null or NaN):
- Calculate the mean rating for user  $u$ :  $\text{Mean}_u = \text{Mean}(\text{Ratings}_u)$
- Set the missing rating  $R_{imputed\_u,m}$  to  $\text{Mean}_u$

#### 2. Remove Duplicates:

- Create an empty set to store unique user-movie rating pairs:  $\text{Unique}_{\text{Ratings}_{Set}}$
- For each user  $u$  and movie  $m$  in the dataset:
  - If the rating  $R_{u,m}$  is not present in  $\text{Unique}_{\text{Ratings}_{Set}}$ :
  - Add the rating  $R_{u,m}$  to  $\text{Unique}_{\text{Ratings}_{Set}}$
  - Else, remove the duplicate rating  $R_{u,m}$  from the dataset

#### 3. Address Data Inconsistencies and Noise:

- For each user  $u$  and movie  $m$  in the dataset:
  - Check for data inconsistencies or noise in the rating  $R_{u,m}$ :
  - If  $R_{u,m}$  is an outlier or inconsistent:
  - Adjust or remove the rating  $R_{u,m}$  from the dataset

#### 4. Return the cleaned and preprocessed dataset

### Pseudo code: Preprocess MovieLens 20M Dataset

**function preprocessMovieLens20M(dataset):**

**// Step 1: Handle Missing Values**

*for each user u in dataset:*

*for each movie m in dataset:*

*if rating  $R_{u,m}$  is missing:*

*Mean\_u = calculateMean(Ratings\_u)*

*R\_imputed\_u,m = Mean\_u*

**// Step 2: Remove Duplicates**

*Unique\_Ratings\_Set = empty set*

*for each user u in dataset:*

*for each movie m in dataset:*

*if  $R_{u,m}$  not in Unique\_Ratings\_Set:*

*Add  $R_{u,m}$  to Unique\_Ratings\_Set*

*else:*

*Remove  $R_{u,m}$  from dataset*

**// Step 3: Address Data Inconsistencies and Noise**

*for each user u in dataset:*

*for each movie m in dataset:*

*if  $R_{u,m}$  is an outlier or inconsistent:*

*Adjust or remove  $R_{u,m}$  from dataset*

*return dataset*

### 3.4 Singular Value Decomposition (SVD) in the context of the MovieLens 20M dataset:

Singular Value Decomposition (SVD) is a matrix factorization technique that decomposes a matrix into three separate matrices, allowing us to reduce dimensionality and capture latent factors. In the case of the MovieLens 20M dataset, we can apply SVD to the user-item interaction matrix to extract meaningful information.

1. *User-Item Interaction Matrix:* Let's denote the user-item interaction matrix as  $M$ , which represents the ratings given by users to movies. The matrix  $M$  has dimensions of  $m \times n$ , where  $m$  is the number of users and  $n$  is the number of items. Each element  $M(i, j)$  in the matrix represents the rating given by user  $i$  to item  $j$ .
2. *Singular Value Decomposition (SVD):* SVD decomposes the user-item interaction matrix  $M$  into

three separate matrices:  $U$ ,  $\Sigma$ , and  $V$ , such that  $M = U\Sigma V^T$ .

- *User Matrix (U):* The user matrix  $U$  has dimensions  $m \times k$ , where  $k$  represents the desired reduced dimensionality. Each row  $U(i, :)$  in  $U$  represents the latent factors associated with user  $i$ . The columns of  $U$  contain the corresponding singular vectors.
- *Diagonal Matrix of Singular Values ( $\Sigma$ ):* The diagonal matrix  $\Sigma$  is a  $k \times k$  matrix, where  $k$  represents the reduced dimensionality. The diagonal elements of  $\Sigma$  contain the singular values in descending order. These singular values quantify the importance of each latent factor.
- *Item Matrix (V):* The item matrix  $V$  has dimensions  $n \times k$ , where each row  $V(j, :)$  represents the latent factors associated with item  $j$ . The columns of  $V$  contain the corresponding singular vectors.

Mathematically, the SVD can be represented as:  $M = U\Sigma V^T$

3. *Dimensionality Reduction:* To reduce the dimensionality of the SVD representation, we can select the top- $k$  singular values and corresponding singular vectors. This allows us to retain the most significant latent factors while discarding the less important ones.

Let's denote the reduced matrices as  $U'(m \times k')$ ,  $\Sigma'(k' \times k')$ , and  $V'(n \times k')$ , where  $k'$  represents the reduced dimensionality. To achieve dimensionality reduction, we keep only the top- $k'$  singular values and the corresponding columns in  $U$  and  $V$ .

The dimensionality-reduced representation can be expressed as:  $M \approx U'\Sigma'V'^T$

Here,  $U'$  has dimensions  $m \times k'$ ,  $\Sigma'$  is a  $k' \times k'$  diagonal matrix, and  $V'$  has dimensions  $n \times k'$ . These matrices represent the reduced-dimensional representations of the user-item interaction matrix.

By applying SVD and dimensionality reduction, we can effectively capture the latent factors associated with users and items in the MovieLens 20M dataset. This allows us to address challenges such as the cold start problem, scalability limitations, and data sparsity, leading to improved recommendation accuracy and efficiency in movie recommendation systems.

### 3.5 Cold start problem

Let's explain how to address the cold start problem using the reduced-dimension  $U$  and  $V$  matrices obtained from Singular Value Decomposition (SVD), along with incorporating content-based filtering techniques. We'll

present the explanation using mathematical models and equations:

1. Utilizing Reduced-Dimension  $U$  and  $V$  Matrices for Recommendations: After applying SVD to the user-item interaction matrix, we obtain the reduced-dimension  $U$  and  $V$  matrices, namely  $U'$  and  $V'$ , respectively. These matrices capture the latent factors associated with users and items.

To generate recommendations for new users or items, we can utilize these reduced-dimension matrices. The recommendation process involves calculating the predicted ratings based on the latent factors and similarities between users/items.

Let's denote the latent factor vectors for a new user as  $u_{new}$  and for a new item as  $v_{new}$ . The predicted rating for the new user  $u_{new}$  and new item  $v_{new}$  can be computed as the dot product between their latent factor vectors and the corresponding columns of the reduced-dimension matrices  $U'$  and  $V'$ :

$$\text{Predicted Rating}(u_{new}, v_{new}) = u_{new} * (V')^T * (U')^T * v_{new}$$

Here,  $(V')^T$  represents the transpose of the reduced-dimension item matrix  $V'$ , and  $(U')^T$  represents the transpose of the reduced-dimension user matrix  $U'$ . The dot product of  $u_{new}$  and  $(V')^T$  captures the similarity between the new user and existing items, while the dot product of  $v_{new}$  and  $(U')^T$  captures the similarity between the new item and existing users.

2. Incorporating Content-Based Filtering Techniques: To leverage the characteristics of items and user preferences, content-based filtering techniques can be incorporated into the recommendation process. Content-based filtering utilizes item features and user preferences to calculate similarity scores and enhance recommendation accuracy.

Let's denote the feature vector of item  $v$  as  $f_v$ , which represents the characteristics or attributes of the item (e.g., movie genre, actors, directors). Similarly, let's denote the preference vector of user  $u$  as  $p_u$ , which represents the user's preferences or profile.

To incorporate content-based filtering, we can calculate the similarity between the feature vector  $f_v$  of the item  $v$  and the preference vector  $p_u$  of the user  $u$ . One common similarity measure used is the cosine similarity:

$$\text{Cosine Similarity}(u, v) = \frac{(p_u * f_v)}{(|p_u| * |f_v|)}$$

Here,  $|p_u|$  and  $|f_v|$  represent the Euclidean norms of the preference vector and feature vector, respectively.

By combining the predicted rating based on the reduced-dimension  $U$  and  $V$  matrices with the content-based similarity scores, we can generate more accurate recommendations for new users or items.

The final recommendation score for the new user-item pair  $(u_{new}, v_{new})$  can be calculated as a weighted sum of the predicted rating and the content-based similarity score:

$$\begin{aligned} \text{Final Recommendation Score}(u_{new}, v_{new}) &= (1 - \alpha) \\ &* \text{Predicted Rating}(u_{new}, v_{new}) \\ &+ \alpha \\ &* \text{Cosine Similarity}(u_{new}, v_{new}) \end{aligned}$$

Here,  $\alpha$  is a weighting factor that determines the relative importance of the predicted rating and the content-based similarity score. By adjusting the value of  $\alpha$ , we can control the balance between collaborative filtering (based on the reduced-dimension  $U$  and  $V$  matrices) and content-based filtering.

By utilizing the reduced-dimension  $U$  and  $V$  matrices obtained from SVD and incorporating content-based filtering techniques, we can effectively address the cold start problem in movie recommendation systems. This approach leverages both collaborative and content-based information to generate accurate recommendations for new users or items, enhancing the overall recommendation quality.

### 3.6 Scalability Limitations:

To address scalability limitations in movie recommendation systems, one widely used technique is approximate matrix factorization. This technique aims to optimize the computation and recommendation generation in real-time environments with a high number of users. By utilizing the reduced-dimension  $U$  and  $V$  matrices obtained from Singular Value Decomposition (SVD), we can achieve efficient processing of large volumes of user-item interactions.

One of the most recommended approaches for optimizing computation and recommendation generation in real-time environments is using matrix factorization with gradient descent optimization. This technique iteratively updates the elements of the  $U$  and  $V$  matrices to minimize the prediction error between the actual ratings and the predicted ratings.

Here's an explanation of this technique in mathematical notation and equations:

1. Objective Function: The objective is to minimize the prediction error between the actual ratings  $(R_u, m)$  and the predicted ratings  $(\hat{R}_{u, m})$  obtained from the matrix

factorization. This can be achieved by minimizing the following objective function:

$$J = \sum(u, m) (R_{u, m} - \hat{R}_{u, m})^2 + \lambda (||U||^2 + ||V||^2)$$

Here,  $\sum(u, m)$  represents the sum over all user-item pairs,  $\lambda$  is a regularization parameter, and  $||U||^2$  and  $||V||^2$  represent the L2 regularization terms to prevent overfitting.

2. Update Rules: To minimize the objective function, we use gradient descent optimization to iteratively update the elements of the  $U$  and  $V$  matrices. The update rules for  $U$  and  $V$  are as follows:

$$U' = U + \alpha (\nabla U - \lambda U) \quad V' = V + \alpha (\nabla V - \lambda V)$$

Here,  $U'$  and  $V'$  represent the updated  $U$  and  $V$  matrices,  $\alpha$  is the learning rate,  $\nabla U$  and  $\nabla V$  represent the gradients of the objective function with respect to  $U$  and  $V$ , and  $\lambda U$  and  $\lambda V$  are the regularization terms.

3. Gradient Computation: The gradients of the objective function with respect to  $U$  and  $V$  can be computed as follows:

$$\begin{aligned} \nabla U &= -2 \sum(u, m) (R_{u, m} - \hat{R}_{u, m}) V \nabla V \\ &= -2 \sum(u, m) (R_{u, m} - \hat{R}_{u, m}) V \\ &\quad - \lambda U \end{aligned}$$

These gradients provide the direction in which the  $U$  and  $V$  matrices should be updated to minimize the prediction error.

By iteratively updating the  $U$  and  $V$  matrices using the gradient descent optimization technique, we can efficiently process large volumes of user-item interactions and generate recommendations in real-time. This approach allows us to overcome scalability limitations and provide timely and accurate recommendations in movie recommendation systems.

### 3.7 Datasparsity

Let's explain how the reduced-dimension  $U$  and  $V$  matrices can be utilized to mitigate data sparsity challenges and employ neighborhood-based collaborative filtering techniques to enhance recommendation accuracy in mathematical terms:

1. Utilizing Reduced-Dimension  $U$  and  $V$  Matrices for Data Sparsity: The reduced-dimension  $U$  and  $V$  matrices obtained from Singular Value Decomposition (SVD) capture the latent factors associated with users and items in a lower-dimensional space. This representation helps overcome data sparsity challenges by

identifying similar users or items based on their latent factor representations.

Mathematically, let's consider the reduced-dimension  $U'$  matrix ( $m \times k'$ ) and  $V'$  matrix ( $n \times k'$ ), where  $k'$  represents the reduced dimensionality. These matrices capture the latent factors associated with users and items, respectively.

To mitigate data sparsity challenges, we can measure the similarity between users or items based on their latent factor representations. One common similarity measure is the cosine similarity.

For two users  $u$  and  $v$ , the cosine similarity ( $similarity(u, v)$ ) can be calculated using the reduced-dimension  $U'$  matrix as follows:  $similarity(u, v) = U'(u, :) \cdot \frac{U'(v, :)}{(||U'(u, :)|| * ||U'(v, :)||)}$

Here,  $U'(u, :)$  and  $U'(v, :)$  represent the latent factor representations of users  $u$  and  $v$  in the reduced-dimension  $U'$  matrix. The dot product ( $\cdot$ ) calculates the similarity between the two user vectors, and the denominators normalize the similarity values.

Similarly, for two items  $i$  and  $j$ , the cosine similarity ( $similarity(i, j)$ ) can be calculated using the reduced-dimension  $V'$  matrix:  $similarity(i, j) = V'(i, :) \cdot \frac{V'(j, :)}{(||V'(i, :)|| * ||V'(j, :)||)}$

Here,  $V'(i, :)$  and  $V'(j, :)$  represent the latent factor representations of items  $i$  and  $j$  in the reduced-dimension  $V'$  matrix.

By calculating the cosine similarity between users or items, we can identify similar users or items with respect to their latent factors. This information can be utilized to improve recommendation accuracy and overcome data sparsity challenges.

2. Neighborhood-Based Collaborative Filtering: Neighborhood-based collaborative filtering is a technique that utilizes the similarity between users or items to generate recommendations. It identifies a set of similar users or items, called the neighborhood, and uses their ratings or preferences to make recommendations.

To employ neighborhood-based collaborative filtering, we need to define a threshold for similarity and select the top-k most similar users or items.

For a given user  $u$ , we can identify the top-k similar users ( $neighbors(u)$ ) based on the cosine similarity between their latent factor representations:  $neighbors(u) = \text{top-k users with highest similarity}(u, v)$