*Special Issue: Laser Induced Breakdown Spectroscopy*

Applied Spectroscopy
www.s-a-s.org

# Detection and Classification of Bacterial Cells After Centrifugation and Filtration of Liquid Specimens Using Laser-Induced Breakdown Spectroscopy

Applied Spectroscopy
2022, Vol. 76(8) 894–904

DOI: 10.1177/0003702822092789
journals.sagepub.com/home/asp
SAGE

**Emma J. Blanchette[1], Sydney C. Sleiman[1], Haiqa Arain[1], Alayna Tieu[1], Chloe L. Clement[1], Griffin C. Howson[1], Emily A. Tracey[1], Hadia Malik[1], Jeremy C. Marvin[1] and Steven J. Rehse[1]**

## Abstract

Five species of bacteria including *Escherichia coli*, *Mycobacterium smegmatis*, *Pseudomonas aeruginosa*, *Staphylococcus epidermidis*, and *Enterobacter cloacae* were deposited from suspensions of various titers onto disposable nitrocellulose filter media for analysis by laser-induced breakdown spectroscopy (LIBS). Bacteria were concentrated and isolated in the center of the filter media during centrifugation using a simple and convenient sample preparation step. Summing all the single-shot LIBS spectra acquired from a given bacterial deposition provided perfectly sensitive and specific discrimination from sterile water control specimens in a partial least squares discriminant analysis (PLS-DA). Use of the single-shot spectra provided only a 0.87 and 0.72 sensitivity and specificity, respectively. To increase the statistical validity of chemometric analyses, a library of pseudodata was created by adding Gaussian noise to the measured intensity of every emission line in an averaged spectrum of each bacterium. The normally distributed pseudodata, consisting of 4995 spectra, were used to compare the performance of the PLS-DA with a discriminant function analysis (DFA) and an artificial neural network (ANN). For the highly similar bacterial data, no algorithm showed significantly superior performance, although the PLS-DA performed least accurately with a classification error of 0.21 compared to 0.16 and 0.17 for ANN and DFA, respectively. Single-shot LIBS spectra from all of the bacterial species were classified in a DFA model tested with a tenfold cross-validation. Classification errors ranging from 20% to 31% were measured due to repeatability limitations in the single-shot data.

## Keywords

Laser-induced breakdown spectroscopy, LIBS, bacteria, filtration medium, centrifugation, discriminant function analysis, artificial neural network, partial least squares discriminant analysis

Date received: 29 April 2021; accepted: 14 March 2022

## Introduction

The use of laser-induced breakdown spectroscopy (LIBS) for the rapid detection and classification of bacteria has been widely investigated.[1,2] Efforts to fully integrate the technology into a commercial clinical instrument have been impeded by a lack of consensus concerning the specimen preparation protocol. While LIBS is often touted as a "minimal-to-no sample preparation" analytic technique, generally the analysis of bacterial cells requires substantial preparation prior to ablation with the LIBS laser due to their small size and limited mass.

Examples of LIBS analyses of bacteria requiring truly minimal sample preparation have been offered by firing laser pulses directly onto "solid" bacterial colonies, which are as close to a bulk material as can be constructed in such a microbiological system.[3–8] While this is an effective means of generating sufficient quantities of cells to yield high signal-to-noise ratio spectra, the number of bacteria present in such specimens is clinically unrealistic without requiring additional culturing time which may take from 24 to 72 h. This delay

[1]Department of Physics, University of Windsor, Windsor, ON, Canada

**Corresponding author:**
Steven J. Rehse, Department of Physics, University of Windsor, 401 Sunset Ave, Windsor, ON N9B3P4, Canada.
Email: rehse@uwindsor.ca

obviates the advantage of speed and real-time diagnosis offered by the LIBS analysis.

A large number of bacterial cells are not intrinsically necessary for bacterial detection, as LIBS spectra have been obtained from a very low number of cells and even single cells present as bioaerosols in a dilute air stream.[9,10] This has also been demonstrated on single fungal spores and bacteria particles in an electro-dynamic balance (EDB) trap.[11] While these methods demonstrate single-cell limits of detection, the preparation methods required to isolate cells from a clinical specimen for inclusion in a dry gas stream or introduction into an EDB trap may preclude their adoption into a clinical environment.

Instead, many authors have investigated ways to easily detect the trace numbers of cells that would be present in a clinical specimen. Of these, serologically tagging cells with unique elemental nanoparticles has proven to be a sensitive technique.[12,13] One disadvantage to this technique is the necessity of knowing the identity of the organism being tested a priori to design the appropriate antibody tag. Alternately, an unknown specimen needs to be exposed to all available immunoassays and then every sample is tested in parallel, as in a multi-well immunoassay. As well, the selectivity and specificity of each immunoassay tag adds an additional area of concern to such approaches.

An alternate approach has been to deposit small numbers of cells on inorganic or abiotic substrates that do not contain the elements present in the bacterial cells responsible for identification. Such substrates include silicon wafers, aluminum disks, and steel disks.[14,15] A limitation of this technique is that due to the nonporosity of the inert surface, the liquid in the specimen is typically driven off by natural evaporation in a preparation step that regularly is conducted overnight, again slowing the process. Also, such techniques have only been shown to be effective for sampling very small volumes of liquids, on the order of 5–10 μL.

Another successfully demonstrated approach has been the use of disposable nitrocellulose filters through which significantly larger volumes of bacteria-containing liquid can be centrifuged in minutes.[16,17] The addition of a concentration cone to the centrifugation process concentrates the bacteria in only a small region of the filter while quickly eliminating the liquid.[18] This concentration lowered the limit of detection of *Escherichia coli* to approximately 1000 colony forming units (CFU) suspended in 1 mL of water, an improvement of a factor of 50 compared to nonconcentrated filters. The current study was conducted to test the ability to differentiate or classify several species of bacterial cells deposited in this manner and to investigate the effect that reducing the titer of the fluid specimens would have. In addition, the use of various chemometric algorithms was explored to determine any differences in effectiveness or accuracy when used to discriminate spectra obtained from the ablation of such bacterial concentrations.

## Experimental

### *Bacterial Deposition*

Liquid suspensions of *Escherichia coli*, *Mycobacterium smegmatis*, *Pseudomonas aeruginosa*, *Staphylococcus epidermidis*, and *Enterobacter cloacae* were prepared after culturing on solid growth media and repeated washing of the cells with deionized water. Cells were then suspended in pure deionized water and uniform concentrations were achieved by monitoring the fluid concentration using optical densitometry. Typically, many milliliters of these suspensions were fabricated with a nominal concentration of $10^8$–$10^9$ CFU per mL. Optical densitometry measurements were made on these "full concentration" suspensions. Serial dilutions were then made from these initial suspensions by removing known volumes of suspension with a micropipette and adding an amount of sterile deionized water to achieve the desired dilution. All dilutions were thus made relative to the full concentration initial suspensions and were made in concentrations of one-fifth (used in the majority of experiments), one-tenth, one-fiftieth, one-hundredth, and one-five hundredth. Concentrations below one-fifth were only made from *E. coli* and *M. smegmatis*. Once the appropriate concentrations were created, 1.0 mL aliquots of these suspensions were pipetted into a custom-fabricated centrifuge/cone-concentration apparatus that is described below. Using the full concentration suspension or even higher concentrations with a greater number of cells resulted in rapid clogging of the concentration cone hole, blocking the flow of cells through the cone during centrifugation.

The deposition of bacterial cells upon disposable filtration media prior to LIBS testing has been described in detail elsewhere.[17,18] Briefly, experiments were performed with 13 mm diameter nitrocellulose filters with a nominal 0.45 μm pore size (HAWP01300, Millipore-Sigma). The filter media were cut with a punch and die set to create 9.5 mm circular diameter filters which would fit within the diameter of custom-fabricated centrifuge tube inserts which needed to slide into standard centrifuge tubes. Once secured within the two-part centrifuge tube insert, a light-weight hollow aluminum cone which contained a 1 mm diameter hole through which liquid could pass was placed into the centrifuge tube insert. One milliliter of a water suspension containing bacterial cells was pipetted into the cone in the top of the centrifuge insert. When secured by the centrifuge tube cap, the apex of the metal cone pressed slightly into the surface of the filter medium, creating a light seal. This assembly was then placed in a standard centrifuge tube and centrifuged at 5000 rpm generating 2500 g of force for five minutes. During this centrifugation, the cone forced the liquid through the 1 mm diameter opening, concentrating all the bacterial cells onto an easily identifiable central region of the nitrocellulose filter. After centrifugation, the centrifuge tube insert was disassembled, exposing the filter which was then removed and

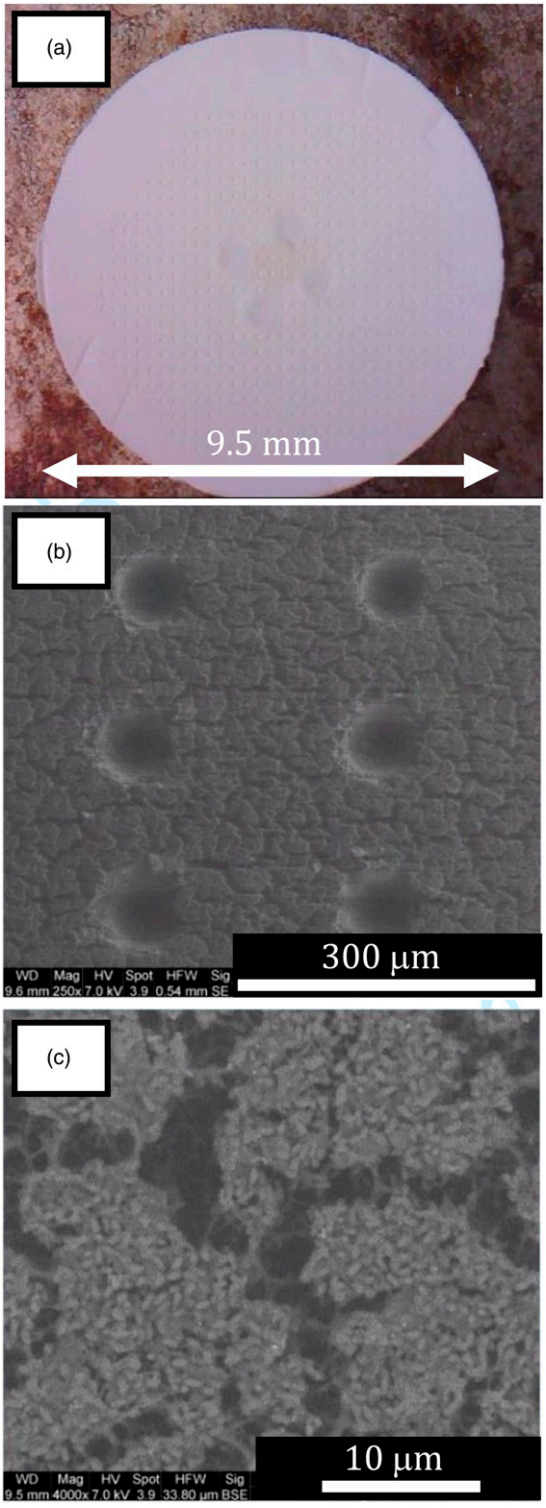


**Figure 1.** (a) Magnification of LIBS ablation craters on a nitrocellulose filter with a bacterial deposition in the center after centrifugation and concentration. A slight discoloration is evident as are four trapezoidal imprints from the centrifugation device used to localize the deposition. SEM micrographs of (b) LIBS ablation craters in a bacterial deposition of *S. epidermidis*, 250× magnification and (c) a 4000× magnification of the bacterial deposition in between the craters. The highly nonuniform nature of the bacterial deposition is evident in both (b) and (c).

mounted on a 25 mm × 25 mm piece of magnetized steel with a piece of double-sided sticky tape. The filter was allowed to dry in place to eliminate any residual water in the filter which tended to decrease overall LIBS intensity. Drying the filter on the double-sided sticky tape prevented the filter from curling up or developing wrinkles as it dried. Each filter was allowed to dry for at least 15 min prior to LIBS ablation. Beyond the 15-min minimum which ensured proper drying, no LIBS dependence on the duration of time between deposition and LIBS ablation was observed for times ranging from 15 min to 24 h. The majority of filters were ablated within six hours of bacterial deposition.

## *Laser-Induced Breakdown Spectroscopy*

The LIBS apparatus used in this investigation has been described in detail previously.[17–19] A 1064 nm neodymium-doped yttrium aluminum garnet (Nd:YAG) pulsed laser (Quanta Ray LAB-150-10, Spectra Physics) operating at 10 Hz produced 10 ns laser pulses that were attenuated in energy to produce 8 mJ pulses on the filter targets. A periscope assembly and a long working distance 5× AR-coated microscope objective focused the laser down to a diameter of 75 µm. All ablation was performed in an atmospheric pressure argon environment with an argon purge at a flow rate of 567 L/h.

All LIBS emission was collected by two matching off-axis parabolic aluminum mirrors which focused the emission into a 1 m steel-encased multi-modal optical fiber. The fiber directed the emission to an echelle spectrometer (ESA 3000, LLA Instruments, Inc.) with a spectral range from 200 to 780 nm. The dispersed light in the spectrometer was detected and recorded with an intensified charge-coupled device (ICCD) camera (Kodak KAF 1001). The timing of the laser, echelle spectrometer, and gating of the ICCD were controlled by a PC running the spectrometer ESAWIN v.3.20 software (LLA Instruments, Inc.). All bacterial spectra were acquired at a delay time of 2 µs after plasma formation using an ICCD gate window width of 20 µs. Amplification of the spectra by the ICCD was kept constant for all spectra. Data were acquired from a single laser shot and the sample was then manually translated 150 µm to acquire a spectrum from an undisturbed area of the target. Approximately 20–30 spectra were acquired from each circular deposition area, shown in Figure 1. More spectra could be acquired from a single filter if required by reducing the distance between laser shots. A representative spectrum obtained by averaging 20 single-shot LIBS spectra acquired from one filter prepared with a one-fifth *E. coli* deposition is shown in Figure 2. Figure 2 also shows a representative spectrum obtained by averaging 20 single-shot LIBS spectra acquired from one blank filter prepared with sterile deionized water.

Each spectrum was analyzed by measuring the integrated emission intensity of 19 neutral and singly ionized lines from

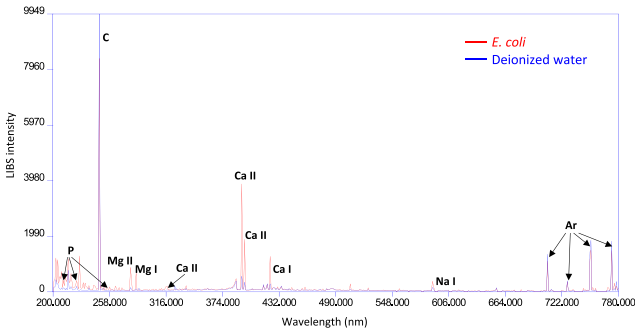


**Figure 2.** A representative spectrum obtained by averaging 20 single-shot LIBS spectra obtained from one filter prepared with a one-fifth *E. coli* deposition (red) and a representative spectrum obtained by averaging 20 single-shot LIBS spectra obtained from one blank filter prepared with sterile deionized water (blue). All spectra were obtained in an argon over-pressure environment at a delay time of 2 μs. The same elemental emission lines were observed in both spectra, but the intensity of the lines was larger in the bacteria spectrum. The carbon line at 247 nm is due primarily to ablation of nitrocellulose filter substrate and its intensity was the same (within uncertainty) in both spectra.

calcium, magnesium, phosphorus, sodium, and carbon, visible in Figure 2. The identification of the 19 specific emission lines is given in Table S1 (Supplemental Material). The ESAWIN software calculated the area under the curve of the background-subtracted spectrum in a region 60 pixels wide centered on the emission line. Due to the dependence of the spectrometer resolution on wavelength, the spectral width of that 60-pixel region varied from approximately 0.31 nm in the vicinity of the shortest wavelength line to approximately 0.82 in the region of the longest wavelength line. The 19 measured intensities were then summed together, and each intensity was divided by the sum to normalize the spectrum for small shot-to-shot variations. Those 19 normalized intensities were used to create 145 ratios yielding a total of 164 independent predictor variables to enhance discrimination accuracy when used in a chemometric analysis as has been described in detail elsewhere.[16,20,21] The identification of the 145 ratios used as predictor variables is given in Table S2 in the Supplemental Material. It has been established that the use of summed intensity ratios as predictor variables in a chemometric analysis reduces the effects of the shot-to-shot variability inherent to LIBS and provides nonlinear variables that can provide additional sample discrimination.[15,22]

### *Chemometric Analysis*

Chemometric analyses were performed using two software packages. Discriminant function analysis (DFA) was performed using SPSS Statistics v.27 (IBM, Inc., 2020). Partial least squares discriminant analysis (PLS-DA) and artificial neural network (ANN) analysis were both performed using PLS_Toolbox Solo v8.8.1 (Eigenvector Research, Inc., 2020). All calculations were performed using standard desktop computing resources.

Data was standardized by the PLS-DA algorithm prior to classification using an included preprocessing routine. Standardization involves both mean centering the data for a given predictor variable by subtracting the mean of that variable from all values to set the mean to zero and also dividing all the values by the standard deviation to set the standard deviation for that variable to unity. In the implementation of PLS-DA, typically the number of latent variables (LVs) used was the number suggested by the PLS_Toolbox, which arrives at this value by calculating the calibration classification error and the cross-validation classification error for one to 20 LVs every time a model is constructed, and then suggesting the number of LVs that minimized those errors without overfitting the data.

For the *E. coli* versus water discrimination tests described in the Detection of *E. coli*/Discrimination from Sterile Specimens section, five LVs were utilized in the construction of every model. For the Gaussian pseudodata tests described in the Comparison of Algorithms section, one or two LVs was sufficient, probably because each predictor variable was already normally distributed about a mean value. For the real bacterial data analysis described in the Classification of Bacteria by Species section, typically three to five LVs were required.

The ANN test had 164 input nodes, utilized only one hidden layer which contained 15 nodes, and there were five output nodes. These parameters for both PLS-DA and ANN are fairly standard for the chemometric analysis of LIBS data.[23,24]

## Results

### *Concentrations of Five Species*

LIBS data were collected from "blank" filters, which were never exposed to any water or microbiological specimens, "sterile water" filters, through which only sterile deionized water was centrifuged and concentrated, and the various concentrations of bacterial specimens from the five bacterial species. 20–30 single-shot LIBS spectra were collected from each filter, with the data acquisition spanning approximately two years. The experimental apparatus and the sample preparation protocol were unchanged during that time period. A dedicated stainless steel calibration sample was analyzed every day that data was acquired before acquiring bacterial LIBS data to monitor and track the performance of the apparatus and to ensure there were no overall changes in absolute LIBS intensity.

To quantify the strength of the LIBS emission in a way which would maximize the differences in LIBS spectra acquired from blank, sterile water, and bacterial filters, several methods of analyzing the data were investigated. The most

sensitive metric was found by calculating the sum of the 18 noncarbon lines in the normalized spectrum and dividing that sum by the normalized carbon intensity. In the blank filter spectra, the only spectral feature of any significant intensity was the carbon 247 nm line from the ablated nitrocellulose filter, minimizing the numerator while maximizing the denominator, resulting in a "noncarbon to carbon" ratio close to zero. In the bacterial spectra, strong spectra would contain intense lines of calcium, magnesium, sodium, and phosphorus (absent in the spectra from nonbiological targets), maximizing the numerator and minimizing the denominator, resulting in "noncarbon to carbon" ratios that ranged from zero (nearly empty spectra) to six (the strongest spectra). Spectra obtained from filters that had only been centrifuged with the sterile deionized water contained small emission lines from calcium, magnesium, and sodium as can be seen in Figure 2. At no time were any lines observed in the water spectra that were not observed in the bacterial spectra and the same 19 lines listed in Table S1 were used to analyze the water, although many of these emission lines were measured to be zero. The water spectra typically possessed a "noncarbon to carbon" ratio below one.

This analysis is shown in Figure 3, which plots the calculated "noncarbon to carbon" ratio versus the spectrum number for 1051 single-shot LIBS spectra obtained from blank filters, filters that water was centrifuged through, and filters with bacterial depositions. In this figure, blank filter data are shown in black, water data in blue, and the bacteria data by the colors indicated. Each separate filter is plotted using a different symbol and those data were acquired on different days. The concentrations of the bacterial suspensions used to make the depositions are given. The average (solid line) and the one-sigma standard deviations from the average (dashed lines) of our blank filter data (in black) and the deionized water filter data (in blue) are indicated to provide a measurement of our background signal, the intensity of the emission lines in the LIBS spectrum when only sterile water is present, and the repeatability of the measurement.

In Figure 3, the blank filter data possess an average value of 0.207 (solid black line) with a standard deviation of 0.038 (dashed black line), yielding a relative standard deviation of 0.183. It should be pointed out that a significant portion of that RSD is due to only one near-outlier point, the first data point, which has a value approximately twice as large as any other point. Without inclusion of this point, the RSD is 0.133, a value much more in line with what was obtained for small emission lines measured in the daily calibration steel sample mentioned earlier. The first data point is well within three sigma, so the point is retained. Nonetheless, the blank filter data all possess very small "noncarbon to carbon ratios" and, again neglecting the first data point, only one bacterial spectrum out of the 862 measured possessed a ratio value smaller than the largest

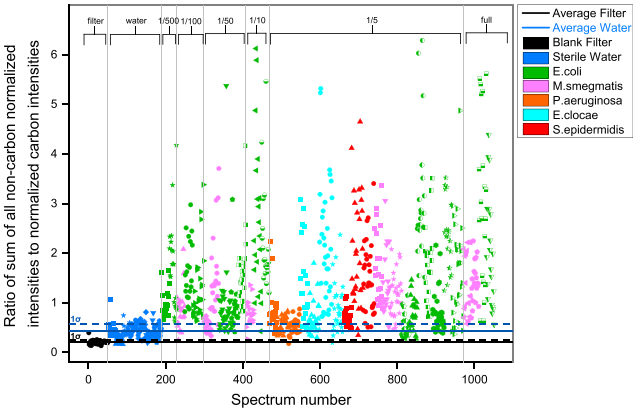


**Figure 3.** Analysis of the relative intensities of emission lines measured in 1051 single-shot LIBS spectra acquired from nitrocellulose filters. The majority of these filters had bacteria concentrated upon them, each color representing a different bacterial species. Spectra from filters through which sterile deionized water was centrifuged were analyzed (dark blue symbols at left), as were spectra from "blank" filters that were not prepared with water or bacteria (black symbols at far left). Various concentrations of bacterial suspensions were prepared prior to centrifugation as indicated at the top of the figure and the bacterial concentration increases from left to right in this figure. Data plotted with the same icon were obtained from the same filter (usually 20 or 30 spectra per filter), whereas a different icon indicates data acquired from a completely different deposition on a different day.

value measured for the filter, providing good separation between filter data and bacterial data.

The "noncarbon to carbon" ratios (shown in blue) calculated from water spectra which were acquired over a month from seven separate filters possessed an average of 0.422 (solid blue line) with a standard deviation of 0.142 (dashed blue line), yielding an RSD of 0.336. This RSD value was larger than was measured for other targets ablated with this apparatus, including the blank filters, indicating this was not an intrinsic property of the filter itself nor of the apparatus. Of the 139 water data points, 133 possessed "noncarbon to carbon" ratios that exceeded the one-sigma standard deviation of the blank filter, while only six fell within one-sigma of the blank filter average. Therefore, while the scatter of these single-shot measurements was high, the values of the ratios provided a good indicator of the presence of water.

The bacteria data possessed much greater variability and as such no average or RSD was calculated. The large scatter in repeated single-shot bacterial measurements has been demonstrated multiple times and is attributed to the stochastic nature of LIBS ablation exaggerated by the nonuniform and nonsolid bacterial surface which can be seen in the SEM micrographs in Figure 1. The bacterial depositions possessed characteristics more like a foam than a solid film, being filled

with air bubbles, gaps, and cracks that formed as the deposition film dried. For these reasons, as well as the thermal properties of the bacterial film itself, these surfaces ablated poorly.

In addition, the dependence of this ratio on the concentrations of the suspension did not follow a consistent trend. This was attributed to the fact that these were truly suspensions not solutions, and so they cannot be expected to be well characterized by a linear curve of growth.[25,26] The bacteria are frequently mucus-like. The cells are clumpy or stringy forming a heterogeneous suspension, not uniform homogenous suspension of isolated individual cells, making a true dilution difficult or impossible. Efforts to reduce the stickiness of the bacteria cells with standard microbiological detergents such as Tween 20 are ongoing, but have not reduced the scatter in subsequent shots obtained from a single filter in any way.[27] The nonhomogenous nature of the suspensions was also evidenced by some filters occasionally possessing twenty single-shot spectra where the ratio of all shots was less than or equal to the average for water, indicating that essentially no bacterial cells were transferred into the centrifuge insert prior to deposition. These filters were not included in this analysis. Finally, all of the *P. aeruginosa* data were extremely low intensity, indicating an inability to create proper suspensions of this particular bacterium. This was attributed to the peculiar nature of this species and its proclivity to form mucoid biofilms. Nonetheless, these filters were left in the analysis for completeness.

## *Detection of* E. coli*/Discrimination from Sterile Specimens*

Partial least square discriminant analysis was performed on all of the *E. coli* data at all concentrations and all of the water data. In total, 14 *E. coli* filters comprising 320 data shots and seven water filters comprising 139 data shots were analyzed.

First, a PLS-DA model was constructed from the single-shot LIBS spectra. The accuracy of a diagnostic test is defined by both the sensitivity and the specificity. The sensitivity is defined as the rate of true positives; in the case of this test, it was the fraction of test spectra correctly identified as *E. coli* out of all the *E. coli* spectra tested. The specificity is defined as the rate of true negatives; in the case of this test, it was the fraction of test spectra that correctly identified as not *E. coli* out of all the water spectra tested. Using five latent variables, the model was constructed with an *E. coli* sensitivity of 0.93 and specificity of 0.95 as tested with a 10-split venetian blinds cross-validation. Next, entire filters were withheld from the model one at a time and then tested against that model to allow 21 externally validated tests of the model. The sensitivity and the specificity of the tests were calculated for each filter based on the number of single-shot spectra on that filter that classified as *E. coli* or water. The average sensitivity of the single-shot spectra from the 14 *E. coli* filters was 0.87. Seven of those 14 filters had 100% of the shots on them (20–30 shots per filter) classified correctly as *E. coli* (sensitivity of 1.0). The worst performing filter had only 33% of the shots correctly classified as *E. coli*. The average specificity of the single-shot spectra from the seven water filters was 0.72. Two of the seven filters had 100% of the shots correctly classified as water, not *E. coli* (specificity of 1.0). The worst performing water filter had only 35% of the shots correctly classified as not belonging to the *E. coli* class. An example of one such external validation is shown in Figure 4 which shows a model constructed from the water (red data, nominal predictor score of 1) and the *E. coli* (green data, nominal predictor score of 0). The data points from one filter of *E. coli*, samples numbered 211 through 230, were not involved in constructing that model and instead were withheld for testing against that model. These unclassified points are represented by the gray points to the right of sample number 460. 100% of these data points correctly classified as *E. coli*, as can be seen by all the spectra possessing a predictor score below the Bayesian threshold, consistent with the *E. coli* spectra in that model.

The variation in the LIBS intensities of single-shot spectra on all the filters containing bacteria, shown in Figure 3, suggested that it was better to sum all of the spectra acquired on a given filter rather than rely on a single-shot spectrum. Ultimately, when applied as a bacterial diagnostic there will be no need for 20 or 30 individual LIBS spectra obtained from one specimen. Only one diagnosis per test specimen is desired, with perhaps an additional independent measurement made for confirmation purposes and/or redundancy. To

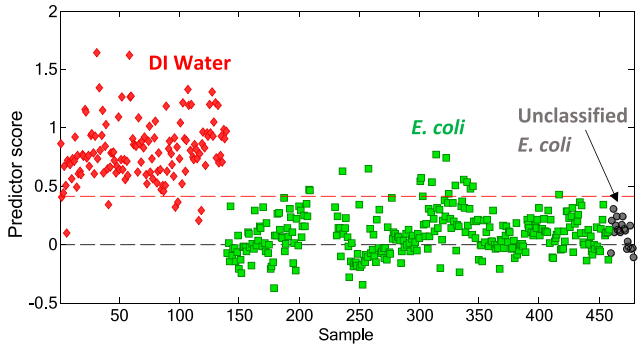


**Figure 4.** A PLS-DA test of 139 single-shot LIBS spectra from seven filters exposed to only sterile DI water (red data, nominal predictor score of 1) and 320 single-shot LIBS spectra from 14 filters upon which various concentrations of *E. coli* were deposited (green data, nominal predictor score of 0). Data were acquired over the span of approximately a year. In the test shown, all spectra from the fourth filter of *E. coli* (samples 211 through 230) were removed from the model, which can be seen by the gap in the *E. coli* data. These unclassified data were then used to validate that model (the gray data to the right of sample 460). 100% of these data were correctly classified as *E. coli*. Every filter of *E. coli* and water was withheld one at a time and tested in this way to determine the diagnostic sensitivity and specificity.

accomplish this, the "add all" feature in the ESAWIN software was used which added up the raw pixel intensities from the ICCD for each individual spectrum prior to CCD readout to produce a single spectrum. This effectively averages the 20 spectra, reducing noise and allowing smaller spectral features to be measured with more accuracy. This is referred to as an "add-all" spectrum.

A PLS-DA model was constructed from the 21 add-all spectra, one spectrum per filter, and again an external validation was performed by sequentially withholding each spectrum from a model and testing it against the model constructed from the others. In this test, 13 of the 14 *E. coli* filters correctly classified as *E. coli* and six of the seven water filters correctly classified as water, yielding a sensitivity of 0.93 and a specificity of 0.86.

Finally, an averaged spectrum for all the single-shot spectra on a filter was generated. These spectra were created by extracting the 19 emission intensities for the 20 or 30 spectra on each filter as described earlier. The measured peak intensities for each variable were then summed together in Excel and the resulting summed spectrum was normalized (as was always done for single-shot spectra). The ratios were then calculated from this summed and normalized spectrum.

In principle, this summation processing is similar, but not identical, to the "add-all" processing because that is performed on data before intensities have been measured and extracted from the raw camera data, while the summation is done after individual peak intensities were extracted from the raw data. Extensive studies of these two methods for averaging multiple shots from a single filter have shown that the two methods yield similar, but not necessarily identical, results. Using these summed spectra, a PLS-DA was performed. In this test, 14 of 14 *E. coli* filters correctly classified as *E. coli* and seven of seven water filters correctly classified as water, yielding a sensitivity of 1.00 and a specificity of 1.00. The results of all three analyses are tabulated in Table I.

## *Comparison of Algorithms*

Prior to attempting to classify all of the highly noisy single-shot bacteria spectra from all five species, tests were conducted to determine which algorithm and which parameters maximized diagnostic accuracy. To conduct these tests on a very large, robust, and well-behaved dataset, a library of pseudodata was generated. This was done by taking all the single-shot spectra from a single species and averaging the values for each of the 19 measured emission line intensities. This average spectrum was then normalized by the sum of all intensities, as described previously. This was done for each of the five types of bacteria to generate five seed spectra, one for each species.

Each seed spectrum was used to generate 998 replicates of each species by applying a Gaussian noise filter to each of the 19 channels. The Gaussian noise filter had a variable width that ranged from 1, which generated highly similar copies of the seed spectrum (low noise), to 2500, which generated highly noisy very dissimilar spectra. This width is unitless and is denoted by $\sigma$ since it defines the typical width of the Gaussian distribution which is normally distributed about the mean of the distribution. The use of a Gaussian filter ensured that for a given emission line intensity (predictor variable) the values of that intensity in the 999 spectra were normally distributed about a mean and possessed a well-defined standard deviation. This was done for every emission line and the $\sigma$ used was the same for all emission lines in a given pseudodataset. To simulate random noise on each individual emission line and not merely fluctuations of the entire spectral intensity, the noise filter was applied independently to each of the 19 lines. This resulted in all of the noise in each channel of the simulated pseudodata being uncorrelated. This was perhaps an overestimation of the noise experienced in the true data.

These pseudodata were discriminated using a DFA, PLS-DA, and ANN. All of the $\sigma$ were tested, but only the $\sigma = 250$ data is presented here. This data was chosen because the performance of the classifications resembled the performance of the actual LIBS data. Also, it was desirable to use pseudodata that possessed lower sensitivities and specificities as the goal of the study was to determine if any one algorithm was superior in discriminating very difficult to classify noisy data. Pseudodata with very small $\sigma$ (below 5) generally were discriminated with 100% sensitivity and specificity for all algorithms, providing no basis for comparison.

For each of the algorithms, 800 pseudodata points were chosen at random to construct the training model, and the 199 remaining points were withheld for validation of the model. This 80:20 split is also common for such studies.[23] Confusion matrices were constructed from the results of each validation and the sensitivity and specificity for each species were calculated. Lastly, the classification error was calculated, which is a way to combine the sensitivity and specificity. The classification error is the average of the false positive rate and the false negative rate for a species and is equal to

$$1 - \left[\frac{(\text{sensitivity} + \text{specificity})}{2}\right] \quad (1)$$

**Table I.** Sensitivity and specificity of PLS-DA tests of LIBS spectra to detect *E. coli* in DI water specimens.

| | Single shots (459 spectra) | "Add-all" filters (21 filters) | "Summed" filters (21 filters) |
|---|---|---|---|
| Sensitivity | 87% | 93% | 100% |
| Specificity | 72% | 86% | 100% |

where the sensitivity and specificity have values as defined earlier. A perfectly sensitive and specific test would possess a classification error of 0 (indicating no errors). The results of the ANN, PLS-DA, and DFA for these data are shown in Figure S1 (Supplemental Material). The classification errors for each of the five species were averaged to obtain a total classification error for the algorithms used. The classification error for the discrimination of the five species was calculated to be 0.16, 0.21, and 0.17 for ANN, PLS-DA, and DFA, respectively. Although no one algorithm was clearly optimal at performing the classifications, the PLS-DA was clearly the worst performing algorithm with these pseudodata, while the ANN and DFA performed similarly (a difference in classification of 0.01 is not a significant difference in this context). These results show that such performance is a limitation of the data, not an intrinsic limitation of the algorithm.

The DFA discrimination (performed by the SPSS program) could build a model from the 4000 "known" pseudodata spectra and test the 955 unclassified spectra (each spectrum composed of 164 independent predictor variables) in a few seconds on a desktop PC. The ANN discrimination (performed by the PLS_Toolbox) took between 12 and 36 h on the same desktop PC to build the model before testing of the unclassified spectra could be performed. Based on this difference in computational requirements, it was determined to perform DFA on the real LIBS datasets until a more optimized ANN analysis could be performed.

### *Classification of Bacteria by Species*

Presented in Figure 3 are the 862 individual spectra from the five bacterial species that were classified using a DFA model. All data were standardized prior to analysis. A tenfold cross-validation was performed where the spectrum order was randomized, and the dataset was separated into 10 groups: nine groups were used to train the discrimination model and the tenth was used for validation.[28–30] The cross-validation was performed a total of 10 times until all 10 groups were used for validation and the overall performance was calculated. The imbalance in the number of spectra present in each class and the relativity low numbers of spectra (in some cases) relative to the number of predictor variables degraded classification accuracy. The sensitivity, specificity, classification error and the number of spectra per species are shown in Table II.

## Discussion

The use of a ratio of the sum of normalized noncarbon emission lines to the carbon line emission intensity was found to provide a very useful metric for quickly quantifying the relative intensity of the bacterial LIBS spectra. This ratio has not yet proven robust enough to calculate a concentration curve which would allow a calculation of a limit of detection but was found to be very useful in identifying spectra with very few or no bacterial cells. This ratio was also useful for identifying the presence of nonzero emission line intensities from the important elements of sodium, magnesium, and calcium in the LIBS spectra obtained from sterile water. These nonzero line intensities in the absence of any bacterial cells were not due to the filter media but were due to the presence of dissolved elements in the water used to suspend the bacteria as well as improper cleaning of all components of the apparatus prior to use. Primarily, this was attributed to the bacterial suspensions of previous tests leaving behind trace amounts of those elements which were always strongly observed in the bacterial LIBS spectrum. An aggressive cleaning protocol consisting of ultrasonication in acetone and methanol were finally adopted to minimize the LIBS intensity acquired from pure deionized water specimens. No quantity of cleaning was effective at reducing the water spectral intensity to that of a blank filter indicating the presence of these elements in the water. Future experiments will attempt to utilize ultrapure 18.2 MΩ water in place of the deionized water previously used to further lower this background.

In addition, attempts to reduce the intrinsic clumpiness or stickiness of the bacterial cells utilizing decreasing concentrations of microbial detergents as well as mechanical disruption will be investigated. It is hoped that this will address some of the high degree of variability caused by the nonuniformity of the deposited bacterial film as well as reduce the uncertainty on the concentrations of the suspensions being tested. Culturing the bacterial samples in liquid growth media rather than from solid agar plates may also prove to be effective at reducing this clumping. However, such culturing is followed by subsequent centrifugation and pelletization, which

**Table II.** Sensitivity, specificity, and classification error of a DFA tenfold cross-validation of LIBS spectra from five species of bacteria.

| | *E. coli* | *S. epidermidis* | *E. cloacae* | *P. aeruginosa* | *M. smegmatis* |
|---|---|---|---|---|---|
| No. of spectra | 400 | 80 | 113 | 80 | 189 |
| Sensitivity | 60% | 64% | 50% | 66% | 65% |
| Specificity | 79% | 91% | 91% | 94% | 82% |
| Classification error | 31% | 23% | 29% | 20% | 27% |

may lead to an unavoidable reclumping of the cells, effectively negating the advantages gained by this method of culturing.

Moving forward, it is expected that the summation or averaging of individual single-shot spectra from a single filter will be utilized to generate significantly less noisy and more reproducible data. The major drawback of this is the reduction in the number of spectra that can be tested to achieve good statistics, but the benefit is improved classification accuracy. In the *E. coli* versus water study, the sensitivity and the specificity obtained by analyzing single-shot spectra were 0.87 and 0.73, respectively, while analysis of the "add all" spectra improved these numbers to 0.93 and 0.86, respectively, albeit with fewer samples to ensure accurate statistics. This result suggests that the summation of spectra, which theoretically allows the use of all available cells on a filter to construct a single spectrum (or perhaps two for redundancy), will ultimately provide the best diagnostic accuracy. This result may be further improved in the future by acquiring more spectra per filter by moving the laser shots closer together. Significantly more data needs to be acquired to demonstrate this, but it is time consuming with our current setup to generate data from a statistically large number of different filters.

The classification accuracy of the data collected was lower than expected due to the high variability of the data, as can be seen in Figure 3. Unlike the degree of reproducibility observed in spectra from blank filters obtained in our studies (approximately 18%, but more likely closer to 13%) or from serum deposited and then dried on a filter (with reproducibility values of less than 1% for spectra from a given filter and reproducibility across larger datasets comprised of many filters at approximately 6%), the mechanical properties of the bacterial films caused a loss of shot-to-shot reproducibility.[28] A DFA performed on single-shot spectra from five species (presented in Table II) reflected this loss of reproducibility in the calculated classification errors. In this analysis, classification errors ranged from 20% to 31% with the two species of *E. coli* and *E. cloacae* being the hardest species to discriminate, with classification errors of 31% and 29%, respectively. A plausible explanation for this is that *E. coli* and *E. cloacae* both belong to the same group of Gram-negative rods and thus possess a similar phenotype, making them biochemically similar. They are the two most similar microorganisms out of the five tested in this study.

The use of add-all or summed data was shown in this work to increase accuracy when bacterial spectra were differentiated from sterile water spectra, but caused a loss of statistical certainty, as this decreased the number of datasets by at least a factor of 20 per specimen. For several of the species in this study, only three or four filters were tested, which is not adequate for statistical certainty. In most statistical chemometric analyses, a conservative guideline is that the number of samples should be a factor of 10 greater than the number of independent predictor variables to avoid overfitting the data. This is seldom, if ever, achieved in LIBS studies utilizing chemometrics. In our case, this would necessitate over a thousand filters per species (approximately 20 000 laser shots) which is not feasible. Nonetheless, data from a greater number of filters are being accumulated to allow a true external validation of the data, where the summed spectrum from each filter is withheld one at a time and tested against a model that contains at least 10–20 other spectra from separate filters.

The creation of pseudodata by applying Gaussian noise to the real spectral data allowed for the creation of an arbitrary number of datasets unconstrained by this limitation. These studies, which agreed broadly with the tenfold cross-validation studies of the real data, also showed a significant decrease in classification accuracy compared to what had previously been observed when the bacteria were ablated on nutrient-free agar, which provided essentially background-free spectra.[20] Work is ongoing to improve the method for constructing pseudodata so that it more accurately represents the scatter in the real bacterial LIBS data, allowing a more rapid development of effective chemometric strategies by using an arbitrarily large library of pseudodata to obtain the true statistics of the algorithm performance. Future efforts will focus on strategies for monitoring individual laser shots to perform real-time outlier rejection of spectra comprised of anomalously weak emission lines such that only spectra that exceed a threshold intensity are recorded. Experiments are underway to deposit the bacteria from a heterogeneous suspension onto the ablation substrate in a much more controlled and uniform way. Potentially, the use of more realistic clinical specimens will solve this problem, as in most clinical specimens, bacteria are not present in the form of stringy or mucoid aggregates.

## Conclusion

One-thousand fifty-one single-shot LIBS spectra were acquired from sterile nitrocellulose filter media prepared in three ways. Spectra from blank media were acquired as were spectra from media through which sterile deionized water had been centrifuged and from media upon which suspensions of bacteria of various concentrations had been deposited via centrifugation. The spectra were acquired over the course of approximately two years utilizing a consistent apparatus and specimen preparation protocol. Five species of bacteria were analyzed in this way.

*Escherichia coli* of all concentrations was efficiently discriminated from sterile water specimens in a five latent variable PLS-DA. The sensitivity and specificity of the externally validated test were 87% and 72%, respectively when the single-shot spectra were used but rose to 93% and 86% when the spectra were summed in the raw data, and 100% and 100% when the intensities of measured spectra peaks were summed after extraction from the raw data.

A large library of spectral pseudodata was created by applying a Gaussian noise filter to an averaged bacterial spectrum of each species. In this way, 998 replicates of each

species were constructed to compare the performances of DFA, PLS-DA, and ANN with a model constructed from 4000 spectra. For the five bacterial species, sensitivities ranged from approximately 50%–100%, with *S. epidermidis* and *E. cloacae* being the hardest species to discriminate, performing poorly in all of the algorithms. The overall classification error for all species was 0.16 for ANN, 0.21 for PLS-DA and 0.17 for DFA. No clear advantage was gained by using a specific algorithm with these data, pointing to limitations in the data itself. The utilization of ANN when performed on 5000 spectra, each composed of 164 independent predictor variables, was found to be prohibitively time and computationally intensive when a desktop PC was used (over 12 h to construct a model with the processor being used at 100% for the calculation).

The sensitivity, specificity, and classification error for a DFA performed on the single-shot spectra from the five species in a tenfold cross-validation were measured. Classification errors ranging from 20% to 30% were obtained. The single-shot spectra obtained in this way were found to possess a significant amount of variation in their spectral intensities, leading to this decrease in classification accuracy from what has been previously observed. It was shown that the addition or averaging of single-shot data enhanced the accuracy, although not enough data was accumulated for all five species to demonstrate this broadly and this was accompanied by an unavoidable decrease in statistical certainty due to a severe reduction in the number of datasets. Nonetheless, this method of spectral addition will continue to be investigated in the future.

## Acknowledgments

The authors are grateful to Ingrid Churchill who supplied the initial stabs for all bacterial cultures and provided advice and assistance with microbiological questions and Sharon Lackie who performed the measurements with the scanning electron microscope.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant award number RGPIN/05842-2017.

## ORCID iD

Steven J. Rehse https://orcid.org/0000-0003-1173-773X

## Supplemental Material

All supplemental material mentioned in the text is available in the online version of the journal.

## References

1. S.J. Rehse. "A Review of the Use of Laser-Induced Breakdown Spectroscopy for Bacterial Classification, Quantification, and Identification". Spectrochim. Acta, Part B. 2019. 154: 50-69. doi:10.1016/j.sab.2019.02.005
2. V.K. Singh, J. Sharma, A.K. Pathak, C.T. Ghany, M.A. Gondal. "Laser-Induced Breakdown Spectroscopy (LIBS): A Novel Technology for Identifying Microbes Causing Infectious Diseases". Biophys. Rev. 2018. 10: 1221-1239. doi:10.1007/s12551-018-0465-9
3. D. Prochazka, M. Mazura, O. Samek, K. Rebrošová, et al. "Combination of Laser-Induced Breakdown Spectroscopy and Raman Spectroscopy for Multivariate Classification of Bacteria". Spectrochim. Acta, Part B. 2018. 139: 6-12. doi:10.1016/j.sab.2017.11.004
4. S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J.A. Ayala, et al. "Rapid Identification and Discrimination of Bacterial Strains by Laser-Induced Breakdown Spectroscopy and Neural Networks". Talanta. 2014. 121: 65-70. doi:10.1016/j.talanta.2013.12.057
5. R.A. Multari, D.A. Cremers, M.L. Bostian, J.M. Dupre, J.E. Gustafson. "Proof of Principle for a Real-Time Pathogen Isolation Media Diagnostic: The Use of Laser-Induced Breakdown Spectroscopy to Discriminate Bacterial Pathogens and Antimicrobial-Resistant *Staphylococcus aureus* Strains Grown on Blood Agar". J. Pathog. 2013. 2013: 898106. doi:10.1155/2013/898106
6. D. Marcos-Martinez, J.A. Ayala, R.C. Izquierdo-Hornillos, F.J. Manuel de Villena, J.O. Caceres. "Identification and Discrimination of Bacterial Strains by Laser-Induced Breakdown Spectroscopy and Neural Networks". Talanta. 2011. 84(3): 730-737. doi:10.1016/j.talanta.2011.01.069
7. J. Diedrich, S.J. Rehse, S. Palchaudhuri. "Escherichia coli Identification and Strain Discrimination Using Nanosecond Laser-Induced Breakdown Spectroscopy". Appl. Phys. Lett. 2007. 90: 163901. doi:10.1063/1.2723659
8. T. Kim, Z.G. Specht, P.S. Vary, C.T. Lin. "Spectral Fingerprints of Bacterial Strains by Laser-Induced Breakdown Spectroscopy". J. Phys. Chem. B. 2004. 108(17): 5477-5482. doi:10.1021/jp031269i
9. J.D. Hybl, G.A. Lithgow, S.G. Buckley. "Laser-Induced Breakdown Spectroscopy Detection and Classification of Biological Aerosols". Appl. Spectrosc. 2003. 57(10): 1207-1215. doi:10.1366/000370203769699054
10. P.B. Dixon, D.W. Hahn. "Feasibility of Detection and Identification of Individual Bioaerosols Using Laser-Induced Breakdown Spectroscopy". Anal. Chem. 2005. 77(2): 631-638. doi:10.1021/ac048838i
11. S. Saari, S. Järvinen, T. Reponen, J. Mensah-Attipoe, et al. "Identification of Single Microbial Particles Using Electro-Dynamic Balance Assisted Laser-Induced Breakdown and Fluorescence Spectroscopy". Aerosol Sci. Technol. 2016. 50(2): 126-132. doi:10.1080/02786826.2015.1134764
12. Y. Markushin, P. Sivakumar, D. Connolly, N. Melikechi. "Tag-Femtosecond Laser-Induced Breakdown Spectroscopy for the

Sensitive Detection of Cancer Antigen 125 in Blood Plasma". Anal. Bioanal. Chem. 2015. 407: 1849-1855. doi:10.1007/s00216-014-8433-0

13. J. Wu, Y. Liu, Y. Cui, X. Zhao, D. Dong. "A Laser-Induced Breakdown Spectroscopy-Integrated Lateral Flow Strip (LIBS-LFS) Sensor for Rapid Detection of Pathogen". Biosens. Bioelectron. 2019. 142: 1115008. doi:10.1016/j.bios.2019.111508
14. W. Liao, Q. Lin, S. Xie, Y. He, et al. "A Novel Strategy for Rapid Detection of Bacteria in Water by the Combination of Three-Dimensional Surface-Enhanced Raman Scattering (3D SERS) and Laser Induced Breakdown Spectroscopy (LIBS)". Anal. Chim. Acta. 2018. 1043: 64-71. doi:10.1016/j.aca.2018.06.058
15. J.L. Gottfried. "Discrimination of Biological and Chemical Threat Simulants in Residue Mixtures on Multiple Substrates". Anal. Bioanal. Chem. 2011. 400: 3289-3301. doi:10.1007/s00216-011-4746-4
16. D.J. Malenfant, D.J. Gillies, S.J. Rehse. "Bacterial Suspensions Deposited on Microbiological Filter Material for Rapid Laser-Induced Breakdown Spectroscopy Identification". Appl. Spectrosc. 2016. 70(3): 485-493. doi:10.1177/0003702815626673
17. D.J. Malenfant, A.E. Paulick, S.J. Rehse. "A Simple and Efficient Centrifugation Filtration Method for Bacterial Concentration and Isolation Prior to Testing Liquid Specimens with Laser-Induced Breakdown Spectroscopy". Spectrochim. Acta, Part B. 2019. 158: 105629. doi:10.1016/j.sab.2019.05.018
18. A.E. Paulick, D.J. Malenfant, S.J. Rehse. "Concentration of Bacterial Specimens during Centrifugation Prior to Laser-Induced Breakdown Spectroscopy Analysis". Spectrochim. Acta, Part B. 2019. 157: 68-75. doi:10.1016/j.sab.2019.05.012
19. J.C. Marvin. Signal Optimization and Enhancement of Laser-Induced Breakdown Spectroscopy for Discrimination of Bacterial Organisms. [Master of Science in Physics Thesis]. Windsor, Ontario, Canada: University of Windsor, 2020. https://scholar.uwindsor.ca/etd/8525/
20. R.A. Putnam, Q.I. Mohaidat, A. Daabous, S.J. Rehse. "A Comparison of Multivariate Analysis Techniques and Variable Selection Strategies in a Laser-Induced Breakdown Spectroscopy Bacterial Classification". Spectrochim. Acta, Part B. 2013. 87: 161-167. doi:10.1016/j.sab.2013.05.014
21. D.J. Malenfant. Influences on the Emissions of Bacterial Plasmas Generated through Nanosecond Laser-Induced Breakdown Spectroscopy. [Master of Science in Physics Thesis]. Windsor, Canada: University of Windsor, 2016. https://scholar.uwindsor.ca/etd/5843.
22. J.L. Gottfried, F.C. De Lucia Jr., C.A. Munson, A.W. Miziolek. "Laser-Induced Breakdown Spectroscopy for Detection of Explosives Residues: A Review of Recent Advances, Challenges, and Future Prospects". Anal. Bioanal. Chem. 2009. 395: 283-300. doi:10.1007/s00216-009-2802-0
23. S. Manzoor, L. Ugena, J. Tornero-Lopéz, H. Martín, et al. "Laser-Induced Breakdown Spectroscopy for the Discrimination of Candida Strains". Talanta. 2016. 155: 101-106. doi:10.1016/j.talanta.2016.04.030
24. X. Chen, X. Li, X. Yu, D. Chen, A. Liu. "Diagnosis of Human Malignancies Using Laser-Induced Breakdown Spectroscopy in Combination with Chemometric Methods". Spectrochim. Acta, Part B. 2018. 139: 63-69. doi:10.1016/j.sab.2017.11.016
25. J.A. Aguilera, J. Bengoechea, C. Aragon. "Curves of Growth of Spectral Lines Emitted by a Laser-Induced Plasma: Influence of the Temporal Evolution and Spatial Inhomogeneity of the Plasma". Spectrochim. Acta, Part B. 2003. 58(2): 221-237. doi:10.1016/S0584-8547(02)00258-6
26. C. Aragon, F. Penalba, J.A. Aguilera. "Curves of Growth of Neutral Atom and Ion Lines Emitted by a Laser Induced Plasma". Spectrochim. Acta, Part B. 2005. 60(7-8): 879-887. doi:10.1016/j.sab.2005.05.015
27. A.E. Paulick. Development of Laser-Induced Breakdown Spectroscopy as a Rapid Diagnostic Tool for Bacterial Infection. [Master of Science in Physics Thesis]. Windsor, Canada: University of Windsor, 2018. https://scholar.uwindsor.ca/etd/7653.
28. X. Chen, X. Li, S. Yang, X. Yu, A. Liu. "Discrimination of Lymphoma Using Laser-Induced Breakdown Spectroscopy Conducted on Whole Blood Samples". Biomed. Opt. Exp. 2018. 9(3): 1057-1068. doi:10.1364/BOE.9.001057
29. R. Gaudiuso, E. Ewusi-Annan, N. Melikechi, X. Sun, et al. "Using LIBS to Diagnose Melanoma in Biomedical Fluids Deposited on Solid Substrates: Limits of Direct Spectral Analysis and Capability of Machine Learning". Spectrochim. Acta, Part B. 2018. 146: 106-114. doi:10.1016/j.sab.2018.05.010
30. X. Li, S. Yang, R. Fan, X. Yu, D. Chen. "Discrimination of Soft Tissues Using Laser-Induced Breakdown Spectroscopy in Combination with k Nearest Neighbors (kNN) and Support Vector Machine (SVM) Classifiers". Opt. Laser Tech. 2018. 102: 233-239. doi:10.1016/j.optlastec.2018.01.028