Spectrochimica Acta Part B: Atomic Spectroscopy 215 (2024) 106911

Contents lists available at ScienceDirect

Spectrochimica Acta Part B: Atomic Spectroscopy

journal homepage: www.elsevier.com/locate/sab

Detection and diagnosis of bacterial pathogens in blood using laser-induced breakdown spectroscopy

E.J. Blanchette [a], E.A. Tracey [a], A. Baughan [a], G.E. Johnson [a], H. Malik [a], C.N. Alionte [a], I.G. Arthur [b], M.E.S. Pontoni [a], S.J. Rehse [a,*]

[a] University of Windsor, Department of Physics, 401 Sunset Ave., Windsor, ON N9B 3P4, Canada
[b] University of Windsor, Department of Biomedical Sciences, 401 Sunset Ave., Windsor, ON N9B 3P4, Canada

ARTICLE INFO

Keywords:
Infection diagnosis
Bacteria
Laser-induced breakdown spectroscopy
Blood
Artificial neural network

ABSTRACT

The ability to rapidly and accurately detect and identify pathogenic bacteria in clinically-obtained blood specimens with laser-induced breakdown spectroscopy (LIBS) is evaluated. Samples of blood obtained from five patients in a local hospital were confirmed to be negative for the presence of bacteria by the pathology department and were then tested with LIBS. Specimens of blood were tested as obtained from the hospital with no other alteration as control samples and were also intentionally spiked with known aliquots of *Escherichia coli, Staphylococcus aureus, Enterobacter cloacae,* and *Pseudomonas aeruginosa* to simulate blood infections. LIBS spectra were acquired from blood deposited on nitrocellulose filters. The intensities of 15 emission lines measured in the spectra and 92 ratios of those line intensities were used as 107 independent variables in a partial least squares discriminant analysis (PLS-DA) to discriminate between sterile control samples and those spiked with bacteria. In addition, the entire LIBS spectrum from 200 nm – 590 nm was input into an artificial neural network analysis with principal component analysis pre-processing (PCA-ANN) to diagnose the bacterial species once detected.

The PLS-DA test possessed a 96.3% sensitivity and a 98.6% specificity for the detection of pathogenic bacteria in blood when 776 spectra from 26 filters were tested by removing one entire filter at a time from the model and testing each spectrum individually. When all the spectra obtained from a filter were averaged to enhance the signal to noise of the spectrum, 19 of 19 filters of infected blood tested positive and 7 of 7 filters with sterile blood tested negative, yielding 100% sensitivity and 100% specificity. The PCA-ANN test performed on the entire LIBS spectrum possessed a 100% sensitivity and 100% specificity when using 80% of the data to build a model and withholding 20% for cross-validation testing. The same PCA-ANN performed on each of the 19 filters individually, using the other 18 filters to build the model, possessed an average sensitivity of 85.5%, an average specificity of 95.0%, and a classification accuracy of 92.5%. These results indicate the potential usefulness of LIBS for detecting and diagnosing blood infections in a clinical setting.

1. Introduction

Bacteria are omnipresent microorganisms found in the human body and the environment. While many of the bacteria that inhabit the human body are harmless, indeed may even be helpful in some cases, some bacteria can cause infection leading to illnesses and mortality in humans. A 1996 report published by the World Health Organization (WHO) stated that microbial disease is the leading cause of premature death worldwide [1]. While many of these infections could be treated quickly and efficiently in the past with broad-spectrum antibiotics, new antibiotic resistant strains of bacteria are emerging, making them much more difficult to treat. A 2019 report by the Centers for Disease Control and Prevention stated that in the United States alone more than 2.8 million antibiotic-resistant infections causing 35,000 deaths were recorded [2] The increase in the number of antibiotic resistant pathogens continues to grow worldwide, causing many more deaths per year, and along with this increase in mortality is a concomitant increase in the economic strain on the health care system [2]. While several strategies are in use to combat this growing threat to human health, a rapid pathogen identification technology which could be used for quick and

* Corresponding author.
E-mail addresses: etracey4@uwo.ca (E.A. Tracey), baughana@uwindsor.ca (A. Baughan), johnso53@uwindsor.ca (G.E. Johnson), malik72@uwindsor.ca (H. Malik), alionte@uwindsor.ca (C.N. Alionte), arthuri@uwindsor.ca (I.G. Arthur), pontonim@uwindsor.ca (M.E.S. Pontoni), rehse@uwindsor.ca (S.J. Rehse).

https://doi.org/10.1016/j.sab.2024.106911
Received 28 December 2023; Received in revised form 28 March 2024; Accepted 1 April 2024
Available online 2 April 2024

targeted therapy could be introduced as another strategy to combat antibiotic resistance which arises due to the use and overuse of broad-spectrum antibiotics [2].

A highly relevant and dangerous nosocomial infection is sepsis, or bacterial blood infection. While sepsis is not always caused by bacterial infection, many cases are in fact bacterial in origin. A blood culture is required for the diagnosis of sepsis, but unfortunately, not all blood cultures will produce a positive test result. Antibiotics are the preferred treatment for sepsis, but even with the best treatment the mortality for patients who have reached septic shock is no better than 50% [3]. Complicating matters is the fact that the time it takes to initiate effective antimicrobial therapy is the single strongest predictor of patient outcome, with every hour of delay increasing patient mortality [4]. One compounding factor is that the standard diagnostic blood cultures are slow and labor intensive, usually requiring a two-step procedure consisting of an initial culture taking up to 120 h, followed by a 24–48 h subculture to identify the pathogens detected in a positive initial culture [5,6]. Clearly what is required is a much more rapid method for determining the specific pathogen responsible for a presentation of bacterial sepsis. It is believed that laser-induced breakdown spectroscopy (LIBS), with its inherent advantages of speed and need for minimal sample preparation, could fill this role as a nearly real-time bacterial pathogen diagnostic [7].

LIBS for blood analysis has been investigated earlier. Melikechi et al. used LIBS to characterize the spectrum obtained from frozen whole blood samples from mice in a helium environment [8]. Many studies have used LIBS specifically to detect and identify blood cancers or cancer biomarkers in human blood such as Markushin et al. who used tag femtosecond LIBS in 2015 to detect cancer biomarkers [9]. Chen et al. also used LIBS on whole blood to detect lymphoma, a cancer of the blood that affects the immune system, at early stages of the cancer in order to provide earlier diagnosis and treatment [10]. LIBS in combination with various chemometric analysis techniques has also been used to discriminate between lymphoma and multiple myeloma using whole blood samples [11]. In 2018, Gaudiuso et al. used femtosecond LIBS and appropriate chemometric algorithms on blood and tissue from mice to diagnose melanoma [12]. Blood serum (a component of whole blood) was tested by Emara et al. to determine electrolyte concentration to further characterize colorectal cancers [13]. Significant work is ongoing to develop LIBS as a method to characterize blood samples to detect or stage such cancers [14–16].

The use of LIBS for the detection and identification of bacteria has been described extensively [7,17]. However, the body of work describing LIBS for the detection and diagnosis of pathogens or pathology present in the blood is smaller than the previously mentioned body of work on detecting cancer biomarkers in blood. Al-Jeffery et al. used laser-induced fluorescence (LIF) and LIBS to detect trace rubidium in the blood to rapidly identify performance enhancing drugs [18]. Gaudiuso et al. investigated rapid diagnosis of gulf war illness using LIBS on blood samples [19]. Zhao et al. studied the use of chemometric methods to diagnose inflammation in the blood [20]. Initial work performed on the detection of pathogens in blood includes Wayua et al. who focused on using LIBS on blood for the detection of malaria and Multari et al. who investigated the feasibility of detecting bacterial, viral, and parasite pathogens in the blood using LIBS [21,22]. The intent of the work described here is to investigate the ability of LIBS to detect bacterial pathogens in unadulterated and unprocessed whole human blood and to diagnose the bacterial pathogen once detected utilizing appropriate machine learning techniques. Several common human pathogens were chosen for testing as representative microorganisms.

2. Materials and methods

2.1. Blood

Sterile blood was provided by the pathology lab at the Windsor Regional Hospital for LIBS testing. All studies were done in accordance with the regulations of the Research Ethics Boards at both the University of Windsor and the Windsor Regional Hospital. Specifically, all specimens were completely anonymized prior to transfer and no information concerning patient identity, background, or demographics was provided. Such studies are classified as specimen transfers and are exempted from requiring a Research Ethics Board certificate according to both institutions. Only blood specimens that had already been tested negative for bacterial infection were provided in these preliminary experiments. Five different blood samples from different patients were provided to account for the difference between patient physiology.

As provided, the blood specimens contained the anticoagulant sodium polyanetholesulfonate (SPS) to prevent the blood from clotting so it can be sampled and stored for extended periods of time. All the blood LIBS spectra collected were affected by the presence of the anticoagulant, most likely by enhancing the observed sodium emission due to the presence of sodium in the SPS. All blood specimens had typically been held by the hospital for at least five days before transfer for LIBS testing to allow for a negative bacterial culture result. At the time of LIBS testing the blood was at least a week old and had likely already undergone some hemolysis. Since the goal of a clinical application is to sample the blood immediately after it has been drawn, these specimens do not exactly mimic those that would be collected directly from a patient. No attempt was made in this preliminary study to immediately sample whole blood obtained from human participants.

2.2. Bacterial LIBS preparation

Colonies of *Escherichia coli, Staphylococcus aureus, Enterobacter cloacae,* and *Pseudomonas aeruginosa* were cultured and harvested from appropriate growth media. Bacterial colonies were suspended in ultrapure megohmic water to reduce any background elemental contributions. Samples were assayed using optical densitometry to ensure that approximately equal concentrations of 10^8 cells per cubic centimeter of water were prepared for each species of bacteria. In most of the experiments described below, dilutions of these standard suspensions were used to reduce the concentration to approximately 2×10^7 cells/mL.

The method for concentrating any bacteria-containing fluid onto the center of a disposable 0.45 μm filter has been described previously, but will briefly be recounted here [23]. Fig. 1(a) shows the disassembled centrifuge insert assembly in an exploded view. The aluminum cone with a 1 mm diameter aperture in the apex used to concentrate the bacteria on the disposable filter media is shown in Fig. 1(b). The entire sealed assembly is shown prior to centrifugation in Fig. 1(c). Figs. 1(d) and 1(e) show the filter medium after bacterial deposition. Discoloration due to the presence of a bacterial film can be seen in the center of the filter and ablation craters from LIBS sampling can be seen. In Fig. 1(e), approximately ten spectra were obtained from within the bacteria-containing central region, but in the tests described here, 30 spectra were acquired per deposition by decreasing the spacing between ablation sites to approximately 150 μm. The centrifuge insert was made to fit inside a standard 10 mL centrifuge tube with a hinged plastic cap, being small enough to allow the centrifuge tube to still be closed as shown in Fig. 1(c). The length of the assembled insert is 40 mm, and it can hold a total of 1.5 mL of fluid. For all tests, 0.45 μm pore size nitrocellulose membrane filters 9.5 mm in diameter (HAWP04700, Millipore Inc.) were used.

Simulated blood infection specimens were creating by 'spiking' the sterile blood with a small amount of bacterial suspension. One hundred microliters of blood was pipetted into the cone and centrifuge insert, followed by 100 μL of the diluted bacterial suspension. The sample was centrifuged (PowerSpin BX C884, Unico) at 5000 rpm (2500 g's of force) for 5 min. After centrifugation, all of the bacterial cells in the initial liquid suspension were deposited in a uniform layer inside of a circular area with diameter of 1 mm, achieving a cell surface density of 2.5×10^8 cells/cm^2. The filter was removed from the centrifuge insert and

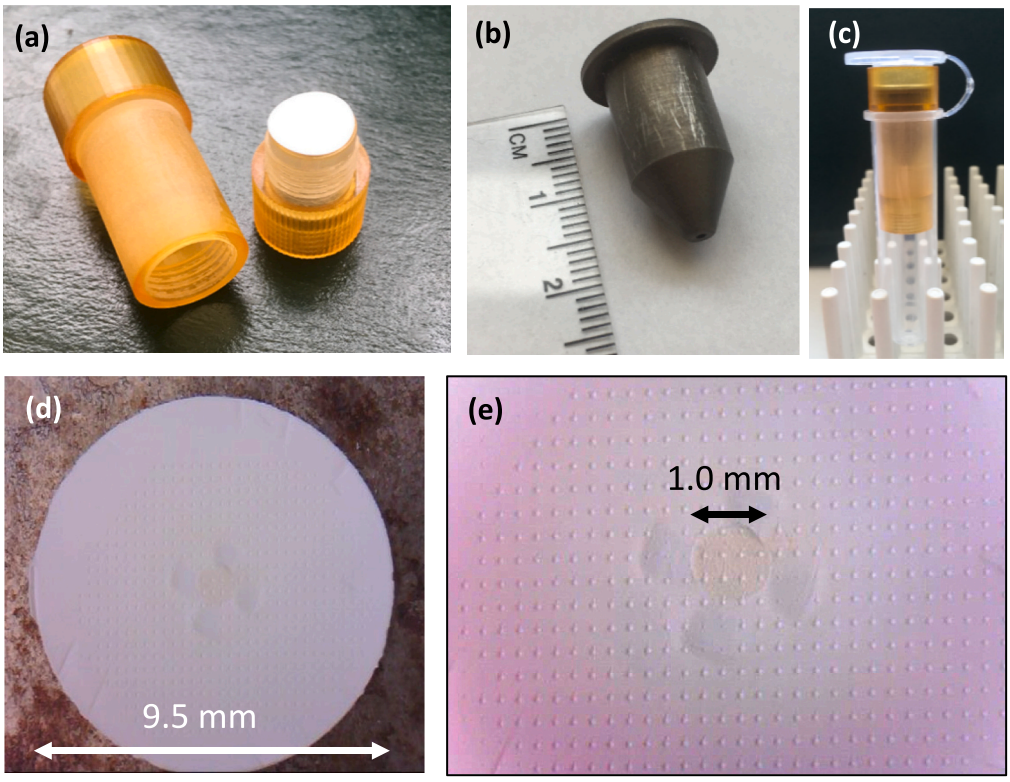

Fig. 1. The disassembled 3D printed centrifuge insert used for bacterial separation fitted with a 9.5 mm diameter disposal filter (a). The aluminum cone with a 1 mm diameter aperture in the apex used to concentrate the bacteria in the center of the filter (b) and the assembled apparatus inside a sealed centrifuge tube (c). Optical images of the filter medium with a bacterial deposition in the center and a rastered array of laser ablation sites (d) and (e).

mounted on a piece of steel with double-sided sticky tape. This was then transferred to the LIBS apparatus for analysis.

Five filters of *S. aureus* were tested, with 30 single-shot spectra obtained from each filter (150 spectra); five filters of *E. coli* were tested, with 30 single-shot spectra obtained from each filter (150 spectra); five filters of *E. cloacae* were tested, with 30 single-shot spectra obtained from each filter (150 spectra); and four filters of *P. aeruginosa* were tested, with 30 single-shot spectra obtained from each filter (120 spectra) for a total of 570 spectra obtained from blood spiked with bacteria. Seven filters of unspiked blood were tested, with 30 single shot spectra obtained from each of six filters and 26 spectra obtained from the seventh for a total of 206 spectra obtained from sterile blood.

2.3. LIBS apparatus

Samples were ablated in an argon gas environment and the argon flow was set to 567 L/h. The laser used for ablation was an Nd:YAG 1064 nm (Quanta Ray LAB-150-10, Spectra Physics) with 10 Hz repetition rate and 10 ns pulse duration. The laser was focused to a circular spot approximately 75 μm in diameter using an antireflection-coated microscope objective, creating an ablation area of 4.4×10^{-5} cm^2. The light from the plasma was collected by two matching parabolic aluminum mirrors (f = 5.08 cm, φ = 3.81 cm) and directed into a 1 m steel-encased multimodal optical fibre (numerical aperture = 0.22, core diameter = 600 μm). The mirrors focused the light onto the optical fibre and increased the amount of light collected by the fibre. The light was then dispersed by an echelle spectrometer (ESA 3000, LLA Instruments, GmbH) and detected by an intensified charge-coupled device (ICCD) camera. The grating used in this experiment allowed a broadband spectrum from 200 to 840 nm to be collected with approximately 12 pm resolution. The ICCD utilized a Kodak camera with a 2.54 cm × 2.54 cm CCD chip (1064 pixels by 1064 pixels, pixel size of 24 μm^2). The spectrometer data acquisition and the firing of the laser Q-switch were controlled with nanosecond timing control by the ESAWIN v3.20 software to set the delay time for data acquisition relative to the laser pulse (called the gate delay, τ_d) and the duration of spectral data acquisition (which is the camera exposure time, called the gate width, τ_w). The gate delay (τ_d) was set to 2 μs after the firing of the laser pulse and the gate width (τ_w) was set to 20 μs for all blood LIBS experiments.

In all experiments, only single-shot spectra were acquired, because multiple laser shots in a single location would ablate through the filter. All of the bacterial cells in the focal region were ablated by a single laser pulse as determined by scanning electron microscopy of the filters performed subsequent to LIBS analysis. The filter was translated 150 μm between laser shots to provide a fresh bacterial target for each laser pulse, allowing the acquisition of between 20 and 30 single-shot spectra from a single centrifugation deposition. Knowing the bacterial cell density on the filter after centrifugation as described in section 2.2 and the laser spot size allows a calculation of approximately 11,000 cells ablated per laser shot, which is a low and clinically relevant number of cells.

2.4. Computerized data analysis: PLS-DA

Fifteen emission line intensities from five elements were measured by the ESAWIN software. These 15 lines are shown in Table 1 and represent the elemental composition of the blood, bacterial cells, and the nitrocellulose filter medium.

Ninety-two ratios of those line intensities were used with the 15 line intensities to create 107 independent variables for use in a partial least squares discriminant analysis (PLS-DA) to discriminate between sterile control blood samples and those spiked with bacteria. The identification of these 107 ratios is provided in the Supplementary Material Tables S1 and S2. PLS-DA is a variation of partial least squares regression (PLS). PLS regression constructs a linear regression model to predict a dependent variable from a set of known independent variables. In the case of

Table 1
Identification and wavelengths of the 15 emission lines observed in the spectra of blood and blood/bacteria.

Species[a]	Wavelength (nm)
C I	247.856
P I	213.618
P I	214.914
P I	253.56
Mg II	279.079
Mg II	279.553
Mg II	279. 806
Mg II	280.271
Mg I	277.983
Ca II	317.933
Ca II	393.366
Ca II	396.847
Ca I	422.673
Na I	588.995
Na I	589.593

[a] I denotes a neutral species; II denotes a singly-ionized species.

PLS-DA the dependent variable is the class (or identity) of the test data set. PLS-DA constructs a smaller number of latent variables (LVs) from the 107 independent variables. The LVs are used as predictor variables to perform a classification prediction on unidentified test data sets. The PLS-DA program used for this work was PLS_toolbox v.8.7.1 running under Matlab 2016b v.9.1 (Eigenvector Research, Inc.) In the PLS_toolbox, the number of LVs can be adjusted manually or can be assigned automatically by the program using a statistical procedure known as cross-validation to optimize the classification accuracy of the model while avoiding overfitting of the data. For the bacterial discrimination of single-shot spectra, seven or eight LVs accounting for approximately 85% of the variance in the data were typically suggested by the software and were sufficient to allow efficient discrimination. The datasets of 107 independent variables were always pre-processed by mean-centering each column of variable data and setting the standard deviation of each column of variables to one, a standard pre-processing step the PLS_toolbox refers to as "auto-scaling."

PLS-DA was also performed on data sets where all the spectra on a given filter were averaged to increase the signal to noise ratio. This averaging was done in two ways: after the ESAWIN software had measured and exported the intensities of the 15 emission lines and by averaging the CCD camera images in the ESAWIN software prior to the measurement of the 15 line intensities. In these analyses, the 107 independent variables were again created from the averaged spectrum. These tests necessarily had far fewer data, with only 19 bacterial spectra and 7 control blood spectra. Due to the improved quality of the data and the reduced number of instances in each class used for model construction and validation, in these analyses only from one to three LVs were required for complete discrimination of the blood with bacteria from the sterile blood, accounting for 80%–90% of the variance in the data.

2.5. Computerized data analysis: PCA-ANN

To discriminate the four bacterial species present in the blood, an artificial neural network (ANN) was trained on a subset of the entire LIBS spectrum from 200 nm – 590 nm. The ANN was developed in Python with the libraries *Pandas*, *Numpy*, *Tensorflow*, *keras*, and *Scikit-Learn* and was run on a standard desktop personal computer using an Intel i9 CPU. To reduce the dimensionality of the data, the 42,000 variable spectra were first pre-processed with an unsupervised principal component analysis (PCA). Others have reported improved results when conducting a principal component analysis on LIBS data first, and then using the principal component scores as independent variables in an ANN [24,25]. The number of principle component scores (PC scores) retained from the PCA becomes the new number of input nodes in the ANN algorithm, significantly reducing training time. A PCA algorithm was developed in python using the libraries *sklearn.decomposition*, *pandas*, *numpy*, and *mpl_toolkits*. Ten PC scores which captured 99.58% of the variance were kept for use in the ANN. Initially the data was mean-centered prior to PCA by calculating the mean of the data across each wavelength and subtracting the mean from each data point of the same wavelength. While mean-centering provided acceptable results, they were no better than the results obtained without the use of pre-processing. Since no gain in classification accuracy was observed when using pre-processed data, for all the results provided below the spectral data were not pre-processed prior to PCA. This lack of improvement in classification accuracy is mostly likely due to the fact that in the full-spectrum LIBS data (over 42,000 independent variables) the majority of channels (variables) are background and noise, and are thus already scattered near zero with similar standard deviations.

An ANN optimization algorithm was developed for use with the ANN to optimize certain parameters utilized during algorithm training including the patience and number of hidden nodes. The hidden nodes are the loci where weighted inputs from the first neural network layer, which are the ten PC scores, are mapped onto output variables, in this case the bacterial class to which the specimen belongs. The patience is a parameter that assists in the convergence of the algorithm by determining when the iterations of the model construction should cease if no improvement of the model loss can be achieved. This algorithm calculated the average sensitivity and specificity for each value of patience and number of hidden nodes, allowing the algorithm to be tailored to the bacterial spectral data. In all analyses only one hidden layer was used in the ANN. Before the data was classified with ANN, the program randomly split all the data into a training set and a testing set. The testing set was created by removing 20% of the spectra from the data set. Typically in ANN, the larger the training set the better the performance, thus the use of this 80:20 cross-validation scheme. To represent a more realistic test, the data were also tested in an "external validation" scheme, where all of the shots on a single filter were withheld entirely from the model and used as test data. Nineteen individual tests, one for each filter withheld, were then performed to obtain the classification accuracy for each species of bacteria. In addition, each test was repeated 10 times in a row to account for variations in the convergence of the ANN. Variations in repeated results were observed due to the sensitivity of the ANN model construction to its initialization, and this is described in section 3.4. The sensitivities and specificities for each filter and species were constructed by averaging the results of the ten repetitions for each filter. Thus 190 PCA-ANN tests were performed on the data. Lastly, to demonstrate that the ANN was not fitting noise, experiments were performed where the spectral identities were randomized after PCA but before the PC scores were input to the ANN. During these tests no classification of the spectra was possible due to the lack of any variance between the now-randomized groups.

3. Results and discussion

Fig. 2 shows a spectrum of blood overlaid with a spectrum of blood spiked with *E. coli* bacteria. These spectra are consistent with spectra observed in the studies on whole blood referenced earlier, although in several of those studies, trace elements like Fe and K were observed and measured. Those elements were not reliably detected in the specimens tested in this study. As well, molecular emissions were observed (e.g. the CN violet system from 388 nm to 380 nm which can be seen in Fig. 2) but the intensities of molecular emissions were not measured. Hβ emission at 486 nm was observed but also not measured due to the extreme width of the Stark-broadened emission line. Hα emission at 656 nm was not observed due to the presence of spectral gaps in the ESA3000 echellogram at that wavelength. It is presumed that the observed hydrogen and molecular emission would play a role in the PCA-ANN analysis which

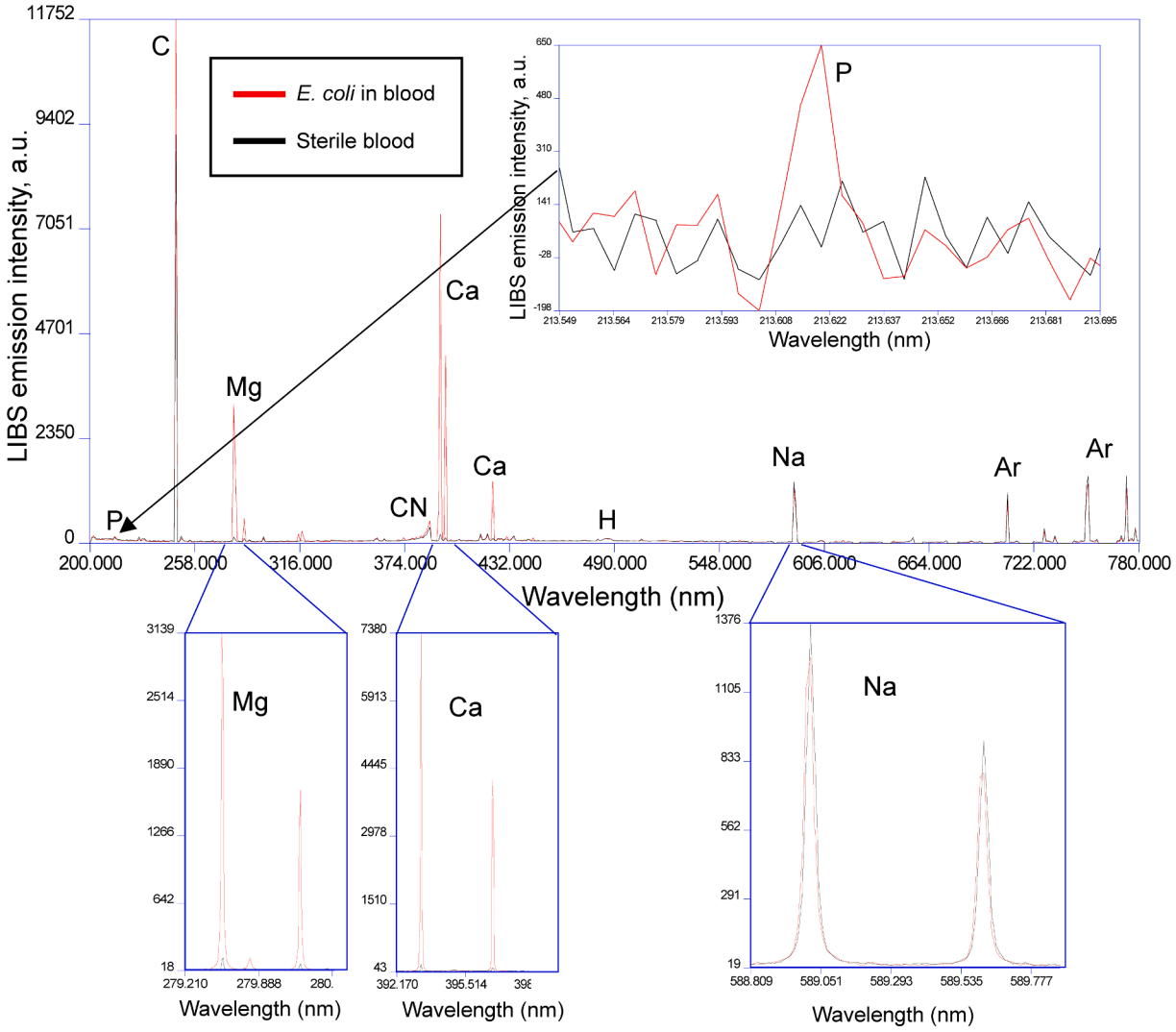

Fig. 2. Overlaid spectra of *E. coli* bacteria mixed with blood (red) and sterile blood (black). The spectrum is dominated by the emission from calcium, magnesium, sodium, and carbon. The calcium, magnesium, and phosphorus line intensities were consistently higher in the bacteria mixed with blood than in blood alone, as can be seen in the zoomed in spectral regions around the strongest lines from those elements. Phosphorus emission was typically quite small in the bacterial spectra (as shown in the inset) but was never observed in the absence of bacteria, making it an indicator of the presence of bacterial cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

utilized the entire LIBS spectrum. This is one explanation for the improved performance of that particular analysis.

In Fig. 2 each spectrum is the average of thirty laser shots taken across an entire filter to increase the signal to noise for visualization purposes only. In the analysis described here, the emission intensity of a spectrum acquired from approximately 11,000 cells in the focal region of a single laser shot was sufficient to allow classification in both the PLS-DA and the PCA-ANN. As shown in the insets in Fig. 2, a careful comparison between the spectra showed that the intensities of calcium, magnesium, and phosphorus lines in bacteria mixed with blood were reliably and consistently higher than the lines from those elements in blood alone. The sodium lines were less consistent, possibly due to the anticoagulant in the specimens. In the inset in Fig. 2, the sodium lines in the spectra from the bacteria mixed with blood were slightly smaller than in the blood alone, although this was within the noise of the measurement and was not consistently observed. Most notably, the phosphorus lines in bacteria mixed with blood, although relatively weak, were always higher than in blood alone. Phosphorus emission (as shown in the inset) was never observed in the absence of bacteria, making it a reliable indicator of the presence of bacterial cells. It is apparent from Fig. 2 that the spectra from bacteria mixed with blood and the spectra from blood alone, although similar, are qualitatively and reproducibly different.

3.1. PLS-DA on single-shot spectra

The PLS-DA was performed as described in Section 2.4 on the five filters of *S. aureus*, five filters of *E. coli*, five filters of *E. cloacae*, four filters of *P. aeruginosa*, and seven filters of unspiked blood. A total of 26 tests were conducted. In each test all the spectra obtained from a single filter were withheld and tested against a model constructed from the remaining 25 filters classified as either “bacteria” or “no bacteria.” After recording the classification results of the 30 spectra, this filter was then reinserted into the dataset, a different filter was withheld for testing, and a new model constructed. This process was repeated sequentially until each filter had been tested individually against a new model. The results of these tests are show in in Table 2.

For the 19 filters with bacteria-containing blood depositions, the fraction of the 30 single-shot LIBS spectra that classified as bacteria were used to calculate a sensitivity for that filter. The sensitivity is the pro-

Table 2
The results of 26 PLS-DA tests performed on filters of bacteria-containing blood and sterile blood.

Filter Identity	Sensitivity %	Specificity %	Averaged Result (Excel)[a]	Averaged Result (ESA)[b]
E. cloacae + Blood Filter #1	96.67	–	Bacteria	Bacteria
E. cloacae + Blood Filter #2	100	–	Bacteria	Bacteria
E. cloacae + Blood Filter #3	100	–	Bacteria	Bacteria
E. cloacae + Blood Filter #4	100	–	Bacteria	Bacteria
E. cloacae + Blood Filter #5	100	–	Bacteria	Bacteria
E. coli + Blood Filter #1	100	–	Bacteria	Bacteria
E. coli + Blood Filter #2	100	–	Bacteria	Bacteria
E. coli + Blood Filter #3	100	–	Bacteria	Bacteria
E. coli + Blood Filter #4	80.00	–	Bacteria	Bacteria
E. coli + Blood Filter #5	100	–	Bacteria	Bacteria
S. aureus + Blood Filter #1	90.00	–	Bacteria	Bacteria
S. aureus + Blood Filter #2	100	–	Bacteria	Bacteria
S. aureus + Blood Filter #3	100	–	Bacteria	Bacteria
S. aureus + Blood Filter #4	96.67	–	Bacteria	Bacteria
S. aureus + Blood Filter #5	100	–	Bacteria	Bacteria
P. aeruginosa + Blood Filter #1	90.00	–	Bacteria	Bacteria
P. aeruginosa + Blood Filter #2	83.33	–	Bacteria	Bacteria
P. aeruginosa + Blood Filter #3	93.33	–	Bacteria	Bacteria
P. aeruginosa + Blood Filter #4	100	–	Bacteria	Bacteria
Sterile Blood Filter #1	–	93.33	No Bacteria	No Bacteria
Sterile Blood Filter #2	–	96.67	No Bacteria	No Bacteria
Sterile Blood Filter #3	–	100	No Bacteria	No Bacteria
Sterile Blood Filter #4	–	100	No Bacteria	No Bacteria
Sterile Blood Filter #5	–	100	No Bacteria	No Bacteria
Sterile Blood Filter #6	–	100	No Bacteria	No Bacteria
Sterile Blood Filter #7	–	100	No Bacteria	No Bacteria

[a] refers to the averaging of spectral line intensities in MS Excel after measurement and extraction by the ESAWIN software.
[b] refers to the averaging of spectral line intensities on the CCD chip by the ESAWIN software prior to the measurement of emission line intensities.

portion of spectra that are classified by the model as a positive test result and are genuinely positive. On these bacterial-deposition filters all of the spectra were genuinely positive. The sensitivity also depends on the number of false negatives, which occur when the test returns a negative result when it is actually positive. Reported as a percentage, the formula for the sensitivity is as follows: $sensitivity = \frac{TP}{TP+FN} \times 100\%$, where TP is true positive and FN is false negative. The optimal medical test would have a sensitivity, or true positive rate, of 100%.

Because the 19 bacteria-containing filters were all positive for bacteria, no rates of false positives (specificity) could be calculated from these data. The results of the seven sterile blood filters were used to calculate specificity. For the seven filters with sterile blood depositions, the fraction of the 30 single-shot LIBS spectra that classified as blood (not bacteria) were used to calculate a specificity for that filter. The specificity is the true negative rate of the test and is the proportion of the spectra that are classified by the model as a negative test result and are genuinely negative. On these sterile blood-deposition filters all of the spectra were genuinely negative. Reported as a percentage, the formula for the specificity is as follows: $specificity = \frac{TN}{TN+FP} \times 100\%$, where TN is true negative and FP is false positive. The optimal medical test would have a specificity of 100%, meaning that no uninfected samples would be diagnosed as infected.

Averaging the measured sensitivities for the 19 bacteria filters yielded an overall sensitivity of 96.3% for this blood test and averaging the specificities for the seven blood filters yielded an overall specificity of 98.6% when 11,000 bacterial cells were ablated per laser shot.

To determine which of the 107 independent variables contributed most strongly to the discrimination of unspiked blood from specimens spiked with bacteria, the loadings of the latent variables were studied. This is shown in Supplemental Material Fig. S1. The first three latent variables accounted for 30.08%, 14.22%, and 18.79% of the variance in the data, respectively. The next latent variable accounted for only 5.4% of the variance of the data and subsequent latent variables even less. Inspection of the loadings revealed that as was seen before, the relative abundances of calcium and magnesium were very important, as were the intensities of the phosphorus lines. Of particular importance were the ratios constructed from the measured emission intensities that are identified in Supplemental Material Table S2. Careful inspection of these loadings did not reveal any obvious trend as to which independent variables were most significant overall, as the magnitudes of the most significant loadings for each latent variable were found to be similar. This analysis confirms that it is not the intensity or absence of emission from any one element that is significant. Rather it is the multivariate nature of the algorithm that provides such efficient differentiation.

3.2. PLS-DA on averaged spectra

Table 2 also shows the results of PLS-DA when all the spectra on a given filter were averaged together to increase signal to noise as described in Section 2.4. Both methods of averaging the spectra yielded a sensitivity of 100% and a specificity of 100%, with no filters being misclassified, when approximately 330,000 total bacterial cells were ablated. This is still a clinically relevant number. Although there were far fewer data in the model and thus far fewer results due to the averaging, it is believed that this way of analyzing the data will ultimately be of clinical utility, as the specimen contributed by a single patient must yield a single diagnosis to the physician, not thirty, to allow the initiation of appropriate therapy. The best method for interpreting and relaying the results of tests such as those presented in Table 2 is still being studied, but it is encouraging that both single-shot spectra and spectra averaged from multiple acquisitions can provide high sensitivities and specificities. It is also encouraging that the particular method for obtaining an average spectrum from all of the spectra acquired from an individual patient was found to be statistically insignificant.

Table 3
Full-spectrum PCA-ANN results using the spectra from 19 filters of bacteria-containing blood tested using an 80:20 cross-validation.

	Bacterial Identity			
	S. aureus	*E. coli*	*E. cloacae*	*P. aeruginosa*
Sensitivity	100%	100%	100%	100%
Specificity	100%	100%	100%	100%
Classification Accuracy	100%	100%	100%	100%

3.3. PCA-ANN on full-Spectrum data: 80:20 cross-validation

Table 3 shows the results of the full-spectrum PCA-ANN performed on all of the spectra obtained from the 19 filters with bacteria-containing blood utilizing an 80:20 cross-validation as described in Section 2.5. No sterile blood spectra were included in this model. In this table, sensitivity measures the percentage of bacteria spectra correctly classified, (e.g. an *E. coli* sensitivity of 100% means that all individual *E. coli* spectra were correctly classified as belonging to the *E. coli* class, a true positive) while specificity measures the percentage of spectra that did not belong to that class that were correctly classified as not belonging to the class, (e.g. an *E. coli* specificity of 100% means that none of the spectra from any of the other three classes were incorrectly classified as belonging to the *E. coli* class, which would have been a false positive.)

For an optimal medical test, the sensitivity and specificity should be optimized to the highest values they can be, without compromising the accuracy of one value in favor of the other. Ideally, the sensitivity and specificity for a medical test should be 100%, but typically there are no realistic medical tests that achieve this level of accuracy. Rather, the focus is on achieving a balance between the two values. To summarize the overall performance of a medical test, classification accuracy is the metric used, which is defined as the fraction of predictions a model or test predicted correctly out of the total number of predictions. This can be summarized as: $classification\ accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$. These values are shown in Table 3. The overall average sensitivity, specificity, and classification accuracy for this model tested in this way were all 100%.

3.4. PCA-ANN on full-Spectrum data: External validation

To provide a more realistic test, 190 PCA-ANN tests were run by performing 10 repetitions of 19 models. Each model was constructed by withholding all the spectra acquired on a single unique filter and then constructing a PCA-ANN model from all the spectra from the other 18 filters. The spectra from the withheld filter were then classified by this model, as described in Section 2.5. Table 4 shows the results of this full-spectrum PCA-ANN tested in this more realistic way. In this test, the sensitivities and the specificities were acquired by averaging over each of the ten repetitions. The results for each test are provided in the Supplementary Material in Tables S3 through S6. Table 4 also provides the overall test sensitivity, specificity, and classification accuracy obtained by averaging the values obtained for each of the four species.

As expected, the results shown in Table 4 are lower than the results presented in Table 3, and this is a result of the more realistic testing of the model. In the 80:20 cross-validation, since the model and test data were all randomized it is expected that 80% of the spectra from any given filter were included in the model, making it significantly easier to identify unclassified spectra from the remaining 20% of the spectra from that filter. The external validation is much more realistic because it incorporates possible variations in sample preparation, day-to-day variations in the LIBS apparatus, and true differences in the blood from different patients used for the tests on different days. When the 30 spectra from a given filter were tested, there were no other spectra from that specimen included in the model construction. This type of testing was done to simulate all of these unknown variations which could be potentially encountered when a new specimen is to be tested.

Table 4
Full-spectrum PCA-ANN results using the spectra from 19 filters of bacteria-containing blood tested with an external validation (entire filters withheld from the model sequentially).

	Bacterial Identity			
	S. aureus	*E. coli*	*E. cloacae*	*P. aeruginosa*
Sensitivity	61.7%	93.7%	91.6%	95.0%
Specificity	96.2%	97.4%	90.2%	96.1%
Classification Accuracy	87.1%	96.4%	90.6%	95.9%
Overall Test Sensitivity	85.5%			
Overall Test Specificity	95.0%			
Overall Test Classification Accuracy	92.5%			

The overall test sensitivity, specificity, and classification accuracy were calculated by averaging the values obtained from each species. Final values of 85.5%, 95.0% and 92.5% respectively were obtained.

All of the tests performed well, except for two of the five tests performed on *S. aureus*. When this species was tested, three of the five filters had sensitivities of 95% or greater. In fact, 25 of the 30 tests run on these three filters (ten repetitions of each) possessed perfect sensitivities, with the lowest test of the 30 having a sensitivity of 67%. But two of the five filters performed extremely poorly. Both of these filters were prepared in an identical manner at the same time and were tested on the same day. This can be seen in the results shown in the Supplemental Material, Table S3. Of these two, one of them had only one spectrum out of 300 correctly classify as *S. aureus*. The other filter exhibited greater variation in the results of the ten repetitions of PCA-ANN, with two of them possessing a sensitivity of 100% and 67% and the other eight possessing a sensitivity of 0%. An examination of the spectra from these filters revealed that they did present as slightly anomalous from the typical *S. aureus* spectra, having noticeable differences in the intensities and ratios of the observed intensities. Cleary when the data were very poor and inconsistent with the other data used to build the model for the assigned class, the model not only had a very difficult time classifying unknown spectra, but was also capable of exhibiting significant variations run-to-run due to the stochastic nature of the ANN model building. These two anomalous filters, which by themselves out of the total of 19 filters tested were almost completely responsible for the reported sensitivity and specificity being below 95%, were kept in the analysis as there was no valid reason for exclusion. Further tests will be performed to see if the nature of the variation observed in those two filters can be identified or mitigated. It is not believed that there is anything inherent to *S. aureus* which would inhibit its ability to be discriminated from these other pathogens, rather it was some systematic variation of the experimental setup or sample preparation which occurred on that particular day.

Nonetheless, the conclusions drawn from this PCA-ANN analysis is that the technique can be accurate when high quality data is used to construct the model and is also used for testing. Such data was able to be acquired across four different species of bacteria and multiple patient blood specimens, indicating the overall robustness of the technique in general. Although this study was conducted using a bacterial titer resulting in the ablation of approximately 11,000 bacterial cells per LIBS spectrum, currently studies are under way to determine the decrease in accuracy that will inevitably occur when the titer of the bacteria added to the sterile blood specimens is reduced. A new design of the concentration cone/centrifugation insert has recently been produced which forms a much better seal with the nitrocellulose filter medium, allowing for almost no loss of cells from the 1 mm circular deposition region. This loss of cells when the insert was not adequately assembled was responsible for a significant reduction in the amount of LIBS emission obtained from the bacteria. This increase in concentration ability should allow us to further reduce the bacterial titer to even lower concentrations. In addition, we are actively performing similar tests on sterile urine specimens to simulate a rapid test for the presence of urinary tract infections, which typically would possess orders of magnitude more bacteria than a sepsis infection.

No attempt was made in this study to create or test mixtures of different bacterial species, representing a “mixed culture.” While such mixed cultures do exist in humans (i.e. in the gut or in the mouth, throat, or oral cavity) such mixtures do not typically present clinically. Infections of nominally sterile bodily fluids (blood, urine, or spinal fluid) are almost always caused by a single microorganism only, and thus the ability to identify a single species of microbe in the background of the

fluid environment is all that is necessary. Even in the case of a non-sterile environment (e.g. the throat), an infection is typically caused by the growth and spread of a single microorganism that has achieved numerical dominance over the normal human flora. So again, equal "mixtures" would not be presented clinically in such infections.

4. Conclusions

An investigation into the use of laser-induced breakdown spectroscopy (LIBS) in combination with appropriate machine learning techniques to analyze blood specimens for the purpose of detecting bacterial pathogens and identifying the species of those pathogens has been reported. Clinical specimens of blood were tested "as is" (with an anticoagulant additive still present and no other chemical processing) by spiking blood from various patients with known aliquots of four species of bacteria. A PLS-DA test using seven latent variables was found to be adequate for discriminating sterile blood from the blood spiked with the known bacteria when single-shot spectra were acquired. Each spectrum was acquired from approximately 11,000 bacterial cells in this case. The PLS-DA test possessed a 96.3% sensitivity and a 98.6% specificity for the detection of pathogenic bacteria in blood when 776 spectra from 26 filters were tested by removing one entire filter at a time from the model and testing each spectrum individually. In addition, this test was repeated with all the spectra obtained from a single filter averaged to enhance the signal to noise of the overall spectrum. In this case, 19 of 19 filters of infected blood tested positive and 7 of 7 filters with sterile blood tested negative, yielding 100% sensitivity and 100% specificity. Due to the reduced size of the validation and test data sets after averaging, the PLS-DA models utilized in this test of averaged spectra required no more than three latent variables.

An artificial neural network with one hidden layer was constructed to identify the pathogens present in the spiked blood samples. A principal component analysis was performed on the entire LIBS spectrum to reduce the dimensionality of the data from 42,000 independent variables down to ten. The first ten principal component scores captured more than 99% of the variance in the data and were used as the input data to the ANN implemented with python on a standard desktop personal computer. A typical cross-validation was done by performing an 80:20 split of the data, testing 20% of all available data (chosen randomly) against the remaining 80% of the data which was used to construct the model. Spectra tested in this way demonstrated a sensitivity and specificity of 100%, with no errors in identification being made.

The model was also tested by withholding one filter at a time from the model construction and then testing the spectra from that filter ten times sequentially to examine the variance in the results of the ANN performance. When the spectra were tested in this way, the overall test sensitivity, specificity, and classification accuracy were calculated by averaging the values obtained from each of the four bacterial species and final values of 85.5%, 95.0% and 92.5% respectively were obtained. Two very poorly performing filters of *S. aureus* were found to be responsible for a significant decrease in the overall sensitivity and classification accuracy of the test, but were retained in the analysis for completeness. Investigations of the same methodology for the testing of urine specimens to develop a rapid diagnostic technology for treating urinary tract infections and an investigation of the effect on decreasing the number of bacteria spiked into the blood are ongoing.

CRediT authorship contribution statement

E.J. Blanchette: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **E.A. Tracey:** Validation, Investigation, Formal analysis. **A. Baughan:** Validation, Software, Investigation, Formal analysis. **G.E. Johnson:** Validation, Software, Investigation, Formal analysis. **H. Malik:** Validation, Investigation, Formal analysis. **C.N. Alionte:** Validation, Investigation, Formal analysis. **I.G. Arthur:** Visualization, Investigation, Formal analysis. **M.E.S. Pontoni:** Validation, Software, Formal analysis. **S.J. Rehse:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

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.

Data availability

Data can be made available on request.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant award number RGPIN/05842-2017. We are grateful to the Windsor Regional Hospital, specifically Lucy DiPietro and Dr. Mohamed El-Fakharany for their assistance in obtaining the clinical blood specimens; to Ingrid Churchill who supplied the initial stabs for all bacterial cultures and provided advice and assistance with microbiological questions; and to Sharon Lackie who performed the measurements with the scanning electron microscope.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sab.2024.106911.

References

[1] World Health Organization, The World Health Report, Fighting disease, fostering development; report of the director-general, Popul. Dev. Rev. 23 (1997) (1996) 203–204, https://doi.org/10.2307/2137484.
[2] Center for Disease Control and Prevention, Antibiotic Resistant Threats in the United States, 2019, U.S. Department of Health and Human Services, 2019, https://doi.org/10.15620/cdc:82532.
[3] E. Whitnack, Sepsis, in: N.C. Engleberg, V.J. DiRita, T. Dermody, M. Schaechter (Eds.), Mechanisms of Microbial Disease, 3rd ed, Lippincott Williams & Wilkins, 2007, pp. 564–572.
[4] A. Kumar, D. Roberts, K.E. Wood, B. Light, J.E. Parrillo, et al., Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock, Crit. Care Med. 34 (2006) 1589–1596, https://doi.org/10.1097/01.CCM.0000217961.75225.E9.
[5] J.C. Lagier, S. Edouard, I. Pagnier, O. Mediannikov, M. Drancourt, D. Raoult, Current and past strategies for bacterial culture in clinical microbiology, Clin. Microbiol. Rev. 28 (2015) 208–236, https://doi.org/10.1128/cmr.00110-14.
[6] D.J. Shin, N. Andini, K. Hsieh, S. Yang, T.-H. Wang, Emerging analytical techniques for rapid pathogen identification and susceptibility testing, Ann. Rev. Anal. Chem. 12 (2019) 41–67, https://doi.org/10.1146/annurev-anchem-061318-115529.
[7] S.J. Rehse, A review of the use of laser-induced breakdown spectroscopy for bacterial classification, quantification, and identification, Spectrochim. Acta Part B. 154 (2019) 50–69, https://doi.org/10.1016/j.sab.2019.02.005.
[8] N. Melikechi, H. Ding, S. Rock, O. Marcano, D. Connolly, Laser-Induced Breakdown Spectroscopy of Whole Blood and Other Liquid Organic Compounds, Proc. SPIE 6863, Optical Diagnostics and Sensing VIII, 68630O, 2008, https://doi.org/10.1117/12.761901.
[9] 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. 407 (2015) 1849–1855, https://doi.org/10.1007/s00216-014-8433-0.
[10] 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. Express 9 (2018) 1057–1068, https://doi.org/10.1364/BOE.9.001057.
[11] X. Chen, X. Li, S. Yang, X. Yu, D. Chen, A. Liu, Diagnosis of human malignancies using laser-induced breakdown spectroscopy in combination with chemometric methods, Spectrochim. Acta Part B 139 (2018) 63–69, https://doi.org/10.1016/j.sab.2017.11.016.
[12] R. Gaudiuso, E. Ewusi-Annan, N. Melikechi, X. Sun, B. Liu, L. Felipe Campesato, 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 146 (2018) 106–114, https://doi.org/10.1016/j.sab.2018.05.010.
[13] E.M. Emara, H. Song, H. Imam, W.M. Elwekeel, X. Gao, M.M. Mohammed, et al., Detection of hypokalemia disorder and its relation with hypercalcemia in blood serum using LIBS technique for patients of colorectal cancer grade I and grade II, Lasers Med. Sci. 37 (2022) 1081–1093, https://doi.org/10.1007/s10103-021-03355-5.
[14] Y. Chu, F. Chen, Z. Sheng, D. Zhang, S. Zhang, W. Wang, et al., Blood cancer diagnosis using ensemble learning based on a random subspace method in laser-induced breakdown spectroscopy, biomed, Opt. Express 11 (2020) 4191–4202, https://doi.org/10.1364/BOE.395332.
[15] Z. Yue, C. Sun, F. Chen, Y. Zhang, W. Xu, S. Shabbir, et al., Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis, biomed, Opt. Express 12 (2021) 2559–2574, https://doi.org/10.1364/BOE.421961.
[16] B.S. Idrees, G. Teng, A. Israr, H. Zaib, Y. Jamil, M. Bilal, et al., Comparison of whole blood and serum samples of breast cancer based on laser-induced breakdown spectroscopy with machine learning, biomed, Opt. Express 14 (2023) 2492–2509, https://doi.org/10.1364/BOE.489513.
[17] S. Ghatak, G. Sharma, P.K. Rai, S. Yadav, G. Watal, Laser-induced breakdown spectroscopy for the identification of bacterial pathogens, in: V.K. Singh, D. K. Tripathi, Y. Deguchi, Z. Wang (Eds.), Laser Induced Breakdown Spectroscopy (LIBS): Concepts, Instrumentation, Data Analysis and Applications, John Wiley & Sons, Ltd, 2023, pp. 745–753, https://doi.org/10.1002/9781119758396.ch37.
[18] M. Al-Jeffery, H.H. Telle, LIBS and LIF for Rapid Detection of Rb Traces in Blood, Proc. SPIE 4613, Optical Biopsy IV, 2002, https://doi.org/10.1117/12.465241.
[19] R. Gaudiuso, S. Chen, E. Kokkotou, L. Conboy, E. Jacobson, E.B. Hanlon, et al., Diagnosis of gulf war illness using laser-induced breakdown spectra acquired from blood samples, Appl. Spectrosc. 76 (2021) 887–893, https://doi.org/10.1177/00037028211042049.
[20] Z. Zhao, W. Ma, G. Teng, X. Xu, K. Wei, G. Chen, et al., Accurate identification of inflammation in blood based on laser-induced breakdown spectroscopy using chemometric methods, Spectrochim. Acta Part B 202 (2023) 10644, https://doi.org/10.1016/j.sab.2023.106644.
[21] D.M. Wayua, H.K. Angeyo, A. Dehayem-Kamadjeu, K.A. Kaduki, Direct analysis of blood for diagnostic metals for malaria by peak-free laser-induced breakdown spectroscopy (LIBS) with artificial neural networks (ANN) and partial least squares (PLS), Anal. Lett. 55 (2022) 2669–2682, https://doi.org/10.1080/00032719.2022.2067862.
[22] R. Multari, D.A. Cremers, A. Nelson, Z. Karimi, S. Young, C. Fisher, et al., The use of laser-based diagnostics for the rapid identification of infectious agents in human blood, Appl. Microbiol. 126 (2019) 1606–1617, https://doi.org/10.1111/jam.14222.
[23] E.J. Blanchette, S.C. Sleiman, H. Arain, A. Tieu, C.L. Clement, G.C. Howson, et al., Detection and classification of bacterial cells after centrifugation and filtration of liquid specimens using laser-induced breakdown spectroscopy, Appl. Spectrosc. 76 (2022) 894–904, https://doi.org/10.1177/00037028221092789.
[24] Y. Zhao, Q. Wang, X. Cui, G. Teng, K. Wei, H. Liu, Discrimination of hazardous bacteria with combination laser-induced breakdown spectroscopy and statistical methods, Appl. Optics 59 (2020) 1329–1337, https://doi.org/10.1364/AO.379136.
[25] Y. Yang, C. Li, S. Liu, H. Min, C. Yan, M. Yang, et al., Classification and identification of brands of iron ores using laser-induced breakdown spectroscopy combined with principal component analysis and artificial neural networks, Anal. Meth. 12 (2020), https://doi.org/10.1039/C9AY02443C, 1316–132.