2023

Detection and Diagnosis of Bacterial Pathogens in Blood and Urine Using Laser-Induced Breakdown Spectroscopy

Emma J.M. Blanchette
University of Windsor

Detection and Diagnosis of Bacterial Pathogens in Blood and Urine Using Laser-Induced Breakdown Spectroscopy

By

Emma J. M. Blanchette

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Physics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2022

Detection and Diagnosis of Bacterial Pathogens in Blood and Urine Using Laser-Induced Breakdown Spectroscopy

By

Emma J. M. Blanchette

APPROVED BY:

_____

A. Hubberstey

Department of Biomedical Sciences

_____

T.J. Hammond

Department of Physics

_____

S. J. Rehse, Advisor

Department of Physics

December 13, 2022

Declaration of Co-Authorship/Previous Publication

I. Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

Chapters 4 and 5 incorporate material published with the following co-authors: Sydney Sleiman, Haiqa Arain, Alayna Tieu, Chloe Clement, Griffin Howson, Emily Tracey, Hadia Malik, Jeremy Marvin and Steven Rehse. In all cases, only my primary contributions towards these publications are included in this thesis, and the contribution of co-author Sydney Sleiman was primarily through data collection and analysis, Haiqa Arain and Alayna Tieu assisted primarily with data analysis, Chloe Clement assisted primarily through data collection, Griffin Howson assisted with data analysis, Emily Tracey was primarily through data collection, Hadia Malik produced figures, Jeremy Marvin provided assistance with experimental design and data collection, and Steven Rehse provided assistance with experimental design, and on feedback and editing of the manuscript. In all cases the key ideas, primary contributions, experimental design, data analysis, interpretation, and editing were performed by myself.

Chapters 5 and 6 incorporate unpublished material co-authored with Alayna Tieu, Grace Johnson, and August Baughan that describes the analysis of spectral data using an artificial neural network code implemented in Python. Alayna Tieu created the initial code upon which this analysis was based. Grace Johnson and August Baughan augmented that code by creating routines to automate the optimization of the parameters, and by introducing a principal component analysis preliminary analysis, written in collaboration with myself, into the Python code. In all cases presented describing the use of this code, the key ideas, experimental design, data analysis, and interpretation of results were performed by myself.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Previous Publication

This thesis includes 1 original paper that has been previously published/submitted to journals for publication, as follows:

Thesis Chapter	Publication title/full citation	Publication status*
Chapter 4	Blanchette, E. J., Sleiman, S. C., Arain, H., Tieu, A., Clement, C. L., Howson, G. C., Tracey, E. A., Malik, H., Marvin, J. C., Rehse, S. J. (2022). Detection and classification of bacterial cells after centrifugation and filtration of liquid specimens using laser-induced breakdown spectroscopy. *Applied Spectroscopy,* 76(8), 894- 904. https://doi.org/10.1177/00037028221092789.	*Published*
Chapter 5	Blanchette, E. J., Sleiman, S. C., Arain, H., Tieu, A., Clement, C. L., Howson, G. C., Tracey, E. A., Malik, H., Marvin, J. C., Rehse, S. J. (2022). Detection and classification of bacterial cells after centrifugation and filtration of liquid specimens using laser-induced breakdown spectroscopy. *Applied Spectroscopy,* 76(8), 894- 904. https://doi.org/10.1177/00037028221092789	*Published*

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

The aim of this thesis is to expand on and improve the existing techniques used for detecting and identifying bacterial pathogens in clinical specimens with laser-induced breakdown spectroscopy (LIBS). Specifically, the existing experimental procedures, including bacterial sample preparation and data acquisition, as well as the data analysis with chemometric algorithms were investigated. Substantial reductions in LIBS background signal were achieved by implementing rigorous cleaning steps and the introduction of the use of ultrapure water. Following this, a database of LIBS spectra was acquired from specimens of *E. coli, S. aureus, E. cloacae, M. smegmatis,* and *P. aeruginosa.* The use of both discriminant function analysis (DFA) and partial least squares discriminant analysis (PLSDA) were compared. A PLSDA model built using the sum of all spectra acquired from 21 filters of *E. coli* and deionized water resulted in a sensitivity and specificity of 100% and 100%, respectively, in an external validation. To optimize the classification accuracy of the single-shot spectra for *E. coli*, *E. cloacae*, and *S. aureus*, outlier rejection schemes and data pre-processing methods were investigated. Classification errors of 30% motivated the use of artificial neural network analysis with principle component analysis pre-processing (PCA-ANN). The average sensitivity and specificity obtained using a randomized 80:20 split validation of the data was 94.18% and 97.01%, respectively. External validation was done on 52 filters of *E. coli*, *E. cloacae*, and *S. aureus*, giving an average sensitivity of 65.7%, and on 49 filters of *E. coli, S. aureus*, and *M. smegmatis* giving an average sensitivity of 87.2%.

Samples of blood and urine were obtained from a hospital and spiked with the same species listed above. 98.9% sensitivity and 100% specificity were achieved for detection of bacteria in urine. 96.3% sensitivity and 98.6% specificity were achieved for detection of bacteria in blood. Discrimination using PCA-ANN on species in urine using an 80:20 split resulted in an average sensitivity and specificity of 97.2% and 98.6%, respectively. External validation on 16 filters gave an average sensitivity 77.5%. Applying PCA-ANN using an 80:20 split on species in blood resulted in 100% sensitivity and specificity. External validation on 19 filters gave an average sensitivity of 82.3%. These results indicate the potential usefulness of LIBS in the clinical setting.

Acknowledgements

This thesis would not have been possible without the guidance and support of my supervisor, my research group, friends, and family.

I would like to sincerely thank my supervisor Dr. Steven Rehse for his exemplary guidance and support whether it was related to research or not. I would also like to thank him for being a source of inspiration and motivation for me throughout my entire university career. None of this work would have been possible without him.

I would like to thank one of the previous masters students Allie Paulick, who laid the groundwork for my research and spent many patient hours teaching me how to use the equipment and properly design experiments. I would also like to thank previous PhD students whose work inspired me to investigate bacteria present in clinical fluids, namely Qassem Mohaidat. Thank you to the previous graduate student Jeremy Marvin, whose expertise on experimental design which was integral to Chapter 4. Thank you to the undergraduate members of my lab for their endless support and help, particularly Haiqa Arain, Alayna Tieu, Chloe Clement, and Emily Tracey. Their help and insights made Chapters 5 and 6 possible.

In addition to these students, I would also like to thank Grace Johnson, August Baughan, Hadia Malik, and Caroline Alionte for their work in data collection and analysis.

Finally, I would like to thank my family and friends for their endless support through the greatest and most difficult times over the course of my education. I would especially like to thank my parents Melissa and Ken, and my brother Eric.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

artificial neural networks (ANN)

cerebral spinal fluid (CSF)

colony forming unit (CFU)

discriminant function (DF)

discriminant function analysis (DFA)

enzyme linked immunosorbent assay (ELISA)

fluorescent in situ hybridization (FISH)

helium-neon (He-Ne)

intensified charge-coupled device (ICCD)

k-nearest neighbours (kNN)

laser-induced breakdown spectroscopy (LIBS)

laser-induced fluorescence (LIF)

laser-induced plasma (LIP)

latent variables (LV's)

lens-to-sample distance (LTSD)

lower respiratory infections (LRI)

mass spectrometry (MS)

matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS)

methicillin-resistant *S. aureus* (MRSA)

microchannel plate (MCP)

partial least squares discriminant analysis (PLSDA)

peptide mass fingerprint (PMF)

permutation feature importance (PFI)

polymerase chain reaction (PCR)

ratio model (RM)

standard cubic feet per hour (SCFH)

time-of-flight (TOF)

tryptic soy agar (TSA)

urinary tract infections (UTI)

World Health Organization (WHO)

List of Appendices

Appendix A:

Chapter 1: Introduction to Current Diagnostic Techniques for Bacterial Infections

Bacteria are omnipresent microorganisms found in the human body and the environment. Of the bacteria that inhabit the human body, many are harmless, and are in some cases helpful. However, some bacteria can cause infection leading to illness and mortality. While some of these infections could be treated in the past with broad spectrum antibiotics, new antibiotic resistant strains of bacteria are emerging making them harder to treat. A 1996 report published by the World Health Organization (WHO) stated that microbial disease is the leading cause of premature death worldwide.[1] A 2019 report by the Centers for Disease Control and Prevention stated that in the Unites States alone more than 2.8 million antibiotic-resistant infections causing 35000 deaths were recorded.[2] The number of antibiotic resistant pathogens continues to grow worldwide, causing many more deaths per year. Parallel to this rising number of antibiotic resistance agents is the economical strain on the health care system.[2] Antibiotic resistant infections are difficult to treat and often require longer hospital stays, follow up visits, and the use of drugs that may be more costly than the basic antibiotic. Antibiotic resistance also threatens to impair modern medicine; many surgeries and treatments rely on the availability of antibiotics to fight post-surgical infections.[2] Without the appropriate antibiotics, many life-saving medical procedures such as organ transplants cannot be offered. To combat this, several strategies are in use, including infection prevention through vaccination and sterilization, and the reduction of unnecessary use of antibiotics.[2] Rapid pathogen identification can result in quick and targeted treatment and can be introduced as another strategy to combat antibiotic resistance as it will reduce overuse of broad-spectrum antibiotics.

Aside from the effects of antibiotic resistance on patient outcomes, there are several other impacts of bacterial infections in general that cause a burden to the healthcare system and the population. For example, lower respiratory infections (LRI) caused by both viral and bacterial pathogens are a leading cause of death and infection for all ages.[3] LRI infections alone cause over 2 million deaths a year in all ages and disproportionately affect children under the age of 5 and countries with lower socio-economic status.[3] As well many infections can be acquired after invasive medical procedures causing prolonged hospital

stays and an overall greater burden on the healthcare system. These diseases that are readily curable still present a challenge and a burden to the healthcare system.[2]

In this thesis I investigate a new practical way to diagnose pathogenic bacteria in clinical specimens using a novel laser-based technique called laser-induced breakdown spectroscopy (LIBS). While the use of LIBS for this application is novel, there are many other techniques to diagnose pathogens that can be classified into four broad categories, these categories are: microbiological techniques, serological techniques, genetic-based techniques, and compositional techniques. By understanding the advantages and limitations of each of the existing well-developed techniques, the reader can better understand the role that a rapid LIBS-based diagnostic could play in the diagnosis of bacterial infections, lessening the burden of bacterial disease, and aiding in the prevention of antibiotic resistance. We can readily compare the LIBS technique to existing methods.

1.1 Microbiological Techniques of Diagnosis

The focus of microbiological techniques is to diagnose based on the physiological characteristics of the bacteria. This can be done by observing how they react to chemicals and stains, their membrane composition, and what media they can grow in. This section will review the techniques of culturing, Gram staining, and microscopy that can be used to observe these microbiological traits. For a more in-depth review, the reader is encouraged to review *Mechanisms of Microbial Disease (3rd Edition)*, Chapter 55, "Diagnostic Principles."

1.1.1 Culturing

Bacteria can be grown in liquid or solid media, referred to as broth and agar, respectively.[4] This process is known as culturing. Culturing is the most specific way to establish what bacteria is causing infection. The goal of culturing is to identify phenotypic traits characteristic of bacterial species, such as metabolic by-products, specific enzymes, utilization of various nutrients, and motility.[5] There are many culturing strategies that can be used based on where in the body the infection is coming from. If the infection is present in sterile body fluids such as cerebral spinal fluid, the culturing strategy is to recover any microorganisms that might be present in the fluid.[5] The bacteria will be cultured on a

variety of media to identify it. If the physician is looking for a specific pathogen in the sterile fluid, then specific media that the bacteria is known to grown on is used to confirm its presence.[5]

Infections present in other regions of the body in which non-pathogenic bacteria are known to inhabit, such as the colon, are more difficult to diagnose with culturing techniques. The strategy used for this is selective media. A selective medium contains specific nutrients that are meant to isolate certain species of bacteria. Using this method, pathogenic organisms can be discriminated from the normal gut flora.[4,5]

Culturing can identify the exact organism and can also establish a cell count, or colony forming unit (CFU) count, which will determine if enough bacteria were present to cause an infection.[6] Cell culturing can also help the physician ascertain susceptibility to antibiotics. There are however many drawbacks to this method. Not all bacterial pathogens can be cultured, and some culturing takes a day or more to complete. Patient outcomes and targeted treatment often depend on the time of diagnosis, therefore making culturing a poor defense against antibiotic resistance.[5] As well, interpreting a culture requires a skilled and experienced laboratory technician or physician, and a well-equipped microbiology laboratory with a large inventory of supplies and biochemicals – many of which are thermally sensitive or have limited shelf lives.[5] These drawbacks make the microbiology techniques unideal.

1.1.2 Microscopic Diagnosis Techniques

Microscopy-based techniques rely on knowledge of the pathogen's phenotypic properties. Under the microscope, features, movement patterns, and staining results can be used to accurately identify some bacterial species.[5] Most infections cannot be diagnosed through simple microscopy since their physical characteristics are too simple or are not distinguishing. A common infection diagnosed solely by microscopy is *Treponema pallidum,* the bacteria that causes syphilis, since it's spirochete shape and movement patterns are easily distinguishable from most other infectious or native bacteria.[5] Techniques employed with microscopy are Gram staining and fluoroscopy.

Gram staining is a method used often in conjunction with microscopy. It can potentially elucidate three clinically relevant pieces of information. The first piece of information is the presence of bacteria in a normally sterile bodily fluid, such as cerebral spinal fluid (CSF) or urine.[7] Second, the morphology and staining properties of the organism can further direct efforts of diagnosis. Finally, with clinical specimens that have distinctive morphologic features and are present in highly specific regions or fluids of the body a definitive diagnosis can be made.[5]

A major advantage to this technique is quick diagnosis; no culture is required to view bacteria under a microscope and Gram staining is a relatively quick procedure. As well, this technique can detect anaerobic organisms in many cases, such as *Heliobacter pylori,* without the need for culturing.[7] One of the drawbacks to Gram staining and subsequent analysis under the microscope is analyzing cultures containing both pathogens and species native to the body. Under these conditions, the Gram stain cannot distinguish a pathogen causing infection. As well, highly trained and experienced professionals are needed to interpret stains under a microscope.[5] More information on the procedure of Gram staining is given in chapter 3.

Fluoroscopy is also used in conjunction with microscopy. Fluoroscopy uses an antibody tagged with a fluorescent substance which allows the microscopist to visualize where the antibodies are binding. The specificity of this test depends on the specificity of the antibodies used and the choice of antibodies used depends on physician expertise.[5]

1.2 Serological Techniques of Diagnosis

Serological techniques rely on the detection of antigens of microorganisms and antibodies produced in bodily fluids through the antibody-antigen specific reaction.[5] Understanding the basics of the antibody-antigen specific reaction is required for understanding serological techniques of diagnosis. The antibody is a protein that is made up of two heavy chains and two light chains. Each light chain and heavy chain have constant and variable regions; the region we are concerned with here is the variable region. The sequences of amino acids in the variable region are different for each antibody and

correspond to one antigen. Therefore, the antibody can bind specifically to one antigen. These techniques are used to characterize the host's immune response.[8]

This section will summarize the use of the most prevalent serodiagnosis methods, including enzyme linked immunosorbent assay (ELISA) and Western blotting. A more detailed description of these methods is given in *Mechanisms of Microbial Disease (3rd Edition)*, Chapter 55, "Diagnostic Principles".

1.2.1 Western Blotting

Western blotting, also known as the immunoblot, is a highly specific test that allows for a protein to be identified in the midst of a complex mixture of other proteins due to the antibody-antigen interaction.[5] The technique first uses gel electrophoresis to separate the antigenic molecules from a pathogen. The separated macromolecules are then placed or 'blotted' onto a solid matrix. Patient serum can then be added to the matrix to determine if any of the patients' antibodies match with the antigens.[5]

A notable advantage of this technique is the high sensitivity and specificity; the highly specific nature of the antibody-antigen interaction dictates this. As well, this technique in theory would need only one antibody to correctly diagnose, making the limit of detection of this technique very low.[5] However, a major disadvantage is that it has limited use for any early diagnosis of acute infections. When the body responds to an acute infection, it takes time for the immune system to produce the appropriate antibodies needed to fight off infection.[4] As well, once these antibodies are produced, they stay in the body indefinitely, which indicates another disadvantage: unless further tests are performed, the physician cannot distinguish if the antibodies are part of an ongoing infection or exist because of a previous infection.[4] Another disadvantage of this technique is that immunocompromised patients have a weak serological response, rendering an antibody based test ineffective.[9]

1.2.2 Enzyme Linked Immunosorbent Assay

Enzyme linked immunosorbent assay (ELISA) is a technique that can detect and quantify proteins, antibodies, and other substances in solution.[10] ELISA is classified as an antigen detection test, which uses antibodies to capture microbial antigens.[5] First, the

antigen of interest is immobilized on a solid matrix support. Next, the antibody in patient serum is complexed with an enzyme that catalyzes the production of coloured compounds. The solid support is then incubated with this patient serum; if the antibodies that correspond to the antigens on the solid support are present in the serum, they will bind to the antigens. Any other material present in solution that does not bind is removed in a washing step, leaving behind only the antibodies bound to the antigens.[9] The purpose of the enzymes is for easier visual detection of the antibody-antigen complexes left behind after the washing step and improving the sensitivity of the technique.[5] This process is shown schematically in Figure 1.1

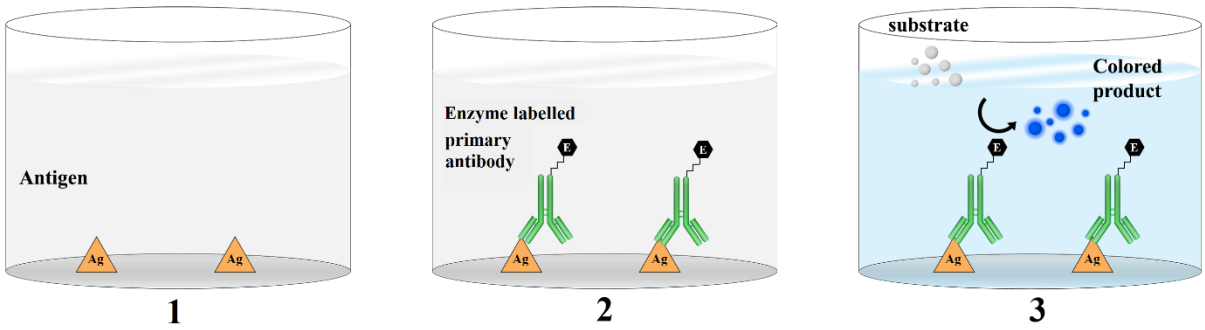

Figure 1.1: An overview of ELISA. (1) The antigen is adhered to a solid matrix support. (2) The antigen is inoculated with antibody and attached enzyme. (3) Substrate is added to mixture and enzyme converts it to coloured product. Adapted from ref [10].

The advantages and disadvantages of ELISA are similar to Western blotting. The advantages are the high sensitivity and specificity, as well as low numbers of antibodies required for detection.[10] The disadvantages are the extremely low numbers of antibodies present at the beginning of infection and the additional number of tests required to determine if the infection is ongoing or occurred earlier.[4] Additional disadvantages to this technique are that the presence of an enzyme on the antibody may decrease the affinity between the antibody and antigen and labelling antibodies with enzymes can be time consuming and expensive.[10] As well, the presence of antibodies in immunocompromised patients is often too weak to provide a definitive test response.[9]

1.3 Genetic Probes

Genetic based diagnosis relies on isolating and identifying DNA sequences that are specific to a certain pathogen. DNA is composed of two separate, but complementary

strands made of nucleic acids that are held together by hydrogen bonds. Because hydrogen bonds are relatively weak, they can be separated easily using heat in a process called denaturing. Putting the strands back together is called annealing; it can be done by simply cooling the DNA. Only complementary strands will be able to bind to one another once they are cooled.[11]

In this section, the polymerase chain reaction (PCR) to create more copies of DNA will be described, as well as the diagnostic test of fluorescent in situ hybridization (FISH). There are some inherent advantages and disadvantages that these nucleic acid techniques share. Advantages include speed of diagnosis over cultivation-based techniques, which can be accomplished within hours and with high sensitivity. Disadvantages include inability to further study the organism for antibiotic resistance and strain type.[8] As well, the techniques are sensitive, but a large amount of DNA material is needed to get a positive test. Therefore, before the probe is applied, PCR must be done for DNA amplification.[8] For a more detailed description of these processes, see *Mechanisms of Microbial Disease (3rd Edition)*, Chapter 55, "Diagnostic Principles." and *Bacteria Pathogenesis: A Molecular Approach*, Chapter 3, "Molecular Approaches to Diagnosis and Characterization of Bacterial Infections."

1.3.1 Fluorescent In Situ Hybridization

Fluorescence in situ hybridization (FISH) is a technique that relies heavily on the ideas of denaturing and annealing of DNA strands. FISH takes advantage of this process by using a probe sequence of nucleic acid with a fluorescent molecule attached to it; this is the first step of the process. The second step is to heat both the probe genes and the genes isolated from the microbial organism to denature them. The single stranded probes are then added to the mixture of single stranded DNA. Cooling of the new mixture takes place to anneal the probe to the complementary target gene sequence if it is present. The fluorescent molecule shows the site of hybridization between the two strands. The mixture is typically analyzed under the microscope.[12] This process is shown schematically in Figure 1.2.

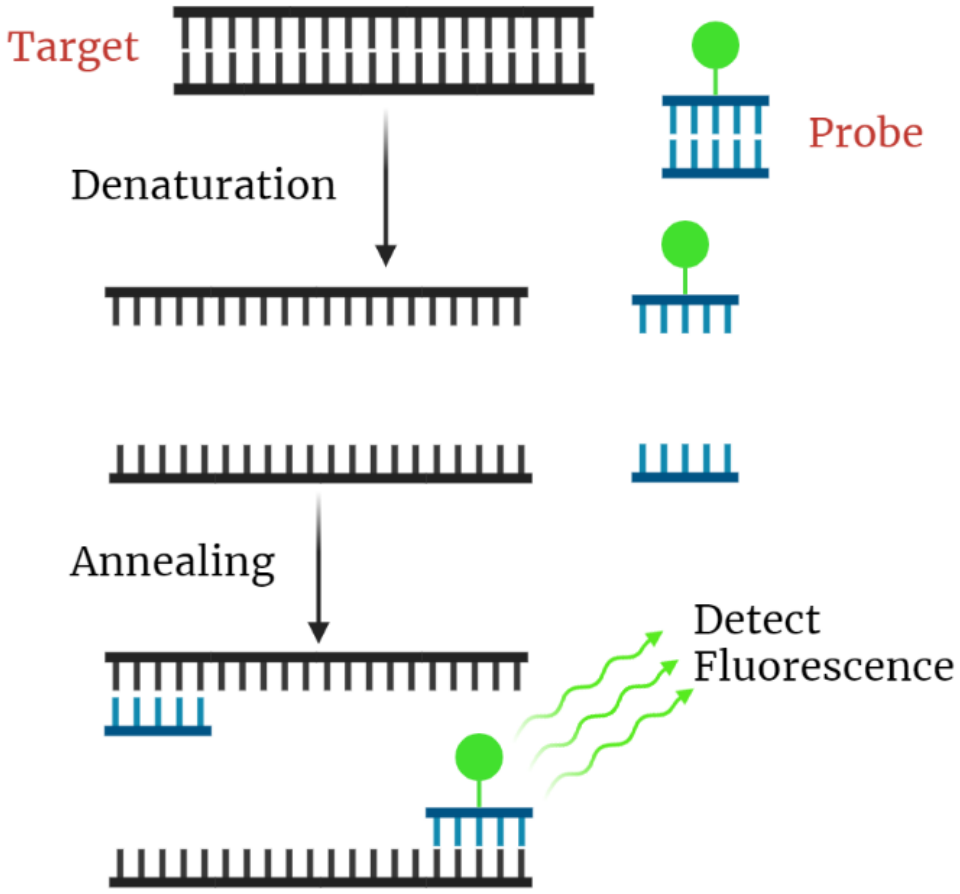

Figure 1.2: FISH process. Target DNA and fluorescence tagged probe sequence DNA is first denatured. Annealing of the target and probe occurs and fluorescence is detected. Figure created with bioRender.

The types of probes used have evolved over the decades as the technique became more prominent.[13] Originally, probes for the DNA or RNA of a specific bacterium were used, but this posed several challenges. One of these challenges was getting the length of the nucleotide sequence big enough to be more specific, small probes would be present in too many target molecules, large probes decrease selectivity towards the target. This method typically used sequences about 15-30 bases long and since each bacterium contains only one genome, the intensity of the fluorescence was not high.[13] Modern day approaches focus on using rRNA sequences which has several advantages over the previous method. First, most bacteria contain a few hundred to several hundred thousand ribosomes per cell

which greatly increases intensity of fluorescence. Second, one probe sequence can identify several types of bacteria.[14]

A notable advantage of FISH is that it can diagnose species that are notoriously difficult to culture because they are obligate or facultative anaerobes. Prominent anaerobic pathogenic bacteria that FISH can detect include *Coxellia burnetii* and *Tropheryma whipplei*, which cause infectious endocarditis. As well, *Yersinia pestis*, infectious agent of the bubonic, septicaemic, and pneumonic plagues can be detected with FISH.[14] These diseases can be highly fatal if not detected early,[2] making FISH and other non-cultivation-based techniques crucial. This technique also has speed and a low operating cost, making it more economical. As well, high sensitivity can be achieved since only one cell is required.[14] A notable disadvantage includes expertise of the operator when examining the final mixture; an operator with considerable education and training is required. As well, since microscopic analysis occurs, there is no standardization, and many false-positive results occur.[14]

1.3.2 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique that makes several copies of the target DNA sample.[5] First, the target gene is denatured, and a primer is annealed to each strand. Extension of the DNA occurs by adding a DNA polymerase and additional nucleic acids to the solution. The DNA polymerase synthesizes a new strand complementary to the template strand in the 5' to 3' direction. Once the DNA polymerase has completed catalyzing all the DNA synthesis, the solution is once again denatured so that all original strands and all newly synthesized strands become the template for the next cycle of polymerization.[5] The overall process can then be summarized in three steps as follows: denaturation (heating), annealing of primers, and DNA synthesis (cooling). With each cycle, the number of DNA copies in solution doubles. Once a large amount of DNA is obtained from the reaction, the strands are detected using gel electrophoresis.[15] This process is shown schematically in Figure 1.3.

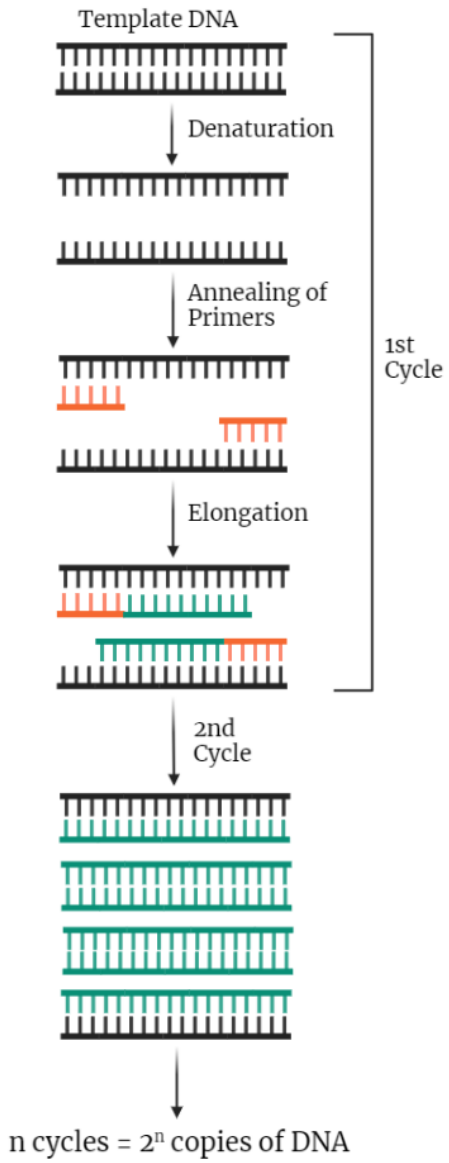

Figure 1.3: The process of polymerase chain reaction. Denaturation of the target strand (black) occurs first, followed by the annealing of a primer (red) to the target strand, and elongation of the new strand (green) creates a new copy. By the end of the process, there will be 2^n copies, where n is the number of cycles. Figure created with bioRender.

There are several advantages to using PCR, one being that PCR is very economical. Unlike culturing, PCR requires only one reagent or primer to be used for every bacterial species. This is because PCR is performed on bacterial rRNA genes which contain nearly universal sequences across all species of bacteria. As well, PCR kits are relatively easy to use; they require less operator training because they come with simple instructions.[15] The disadvantage with this technique is the high potential for cross-contamination due to its high sensitivity and therefore false positives are commonly observed.[5]

1.4 Compositional Spectroscopic Methods

The most recent advancements for diagnosis of bacterial pathogens depend on analyzing the composition of the bacteria of interest. Two principle methods that are used in composition analysis are matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and Raman spectroscopy. Both rely on the specific molecular or protein composition of bacterial cells. MALDI-TOF MS uses the proteins of bacterial cells as a 'fingerprint' to identify each species.[16] Raman spectroscopy uses electromagnetic radiation to probe the molecular composition of the outer membranes of cells to produce a characteristic scattering spectrum.[17] These techniques have only been introduced and popularized over the last decade.[18] For a more in-depth review on MALDI-TOF MS the reader is encouraged to consult the textbook, *MALDI-TOF and Tandem MS For Clinical Microbiology*.

1.4.1 Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

MALDI-TOF MS uses mass spectrometry (MS), which ionizes chemical compounds into charged particles and measures the mass to charge ratio. Alone, MS can typically only analyze smaller molecules, but used in conjunction with matrix assisted laser desorption ionization (MALDI), it can be used on larger biomolecules such as proteins.[18] General sample preparation for MALDI-TOF MS begins by placing the sample in an organic matrix solution. Upon drying, the matrix and sample crystalize. Typically, samples are prepared in wells on a MALDI target plate. Applying a laser to the sample well ionizes the samples.[18] After ionization, the charged samples are accelerated across a fixed potential causing them to separate based on the mass-to-charge ratio. Measurement of the time-of-flight (TOF) of the protein fragments through a flight tube provides a characteristic spectrum called the peptide mass fingerprint (PMF). The majority of molecules detected in MALDI spectra are ribosomal proteins.[16] Diagnosis of the organism is done by comparing the PMF received to known PMF spectra in a database. The genus of the organism can be established using this method, and often the strain level.[16]

There are several advantages of this technique. It is relatively inexpensive to run samples, and results are available at the species and strain level in minutes.[16] One of the

major disadvantages is cost; there is a large initial investment to obtain a machine and continued investment is required for maintenance and servicing. This makes MALDI-TOF MS not readily accessible in smaller hospitals and clinics. As well, there are some closely related bacterial species or strains that are often confused for one another when using MALDI-TOF MS; an example being *Streptococcus* species.[16] MALDI-TOF MS requires enough bacteria to make a diagnosis, so culturing before analysis is often required. This of course limits the technique to identification of organisms that can be cultured. Depending on the species, this can add several hours to the time before a diagnosis can be made.[16]

1.4.2 Raman Spectroscopy

Raman spectroscopy relies on the inelastic scattering of laser light from molecules. The frequency of the scattered photons is shifted either up or down relative to the frequency used to probe the molecule. This is called the Raman shift and it provides information about the molecule's vibrational modes.[19] Analysis and classification of the resulting spectra is typically carried out by chemometric algorithms.[19]

Several advantages exist for the use of Raman spectroscopy on diagnosis of pathogenic organisms. This technique is culture independent and does not depend on growth media and growth phase. However, unlike MALDI-TOF, this is a non-destructive technique. High specificity can also be achieved using Raman spectroscopy at a relatively low cost.[20] Disadvantages of the technique include lower signal due to high background fluorescence; amino acids and nucleic acids make up this interfering background signal. Raman spectroscopy is also an emerging technology that is still under investigation for bacterial diagnosis, and therefore no commercial database exists for diagnosis.[19]

1.5 Scope of Thesis

The goal of our group's work is to use LIBS to develop a rapid medical diagnostic tool for bacterial infections. The focus of my thesis specifically is to develop a quick sample preparation method for bacteria in bodily fluids. My thesis will also describe our efforts to improve the detection of bacteria by reducing the background signal through more rigorous cleaning methods. My thesis will present the results of detection and diagnosis of bacteria in sterile clinical fluids, including blood and urine.

The first chapter of my thesis describes the current state of the antibiotic resistance crisis as well as notable challenges for diagnosing and treating infections in the clinical setting. LIBS is introduced as a robust, cost-effective technique that addresses many of the issues currently faced in the clinic because it will quickly diagnose deadly bacterial diseases and improve treatment timelines, thereby reducing the rate of antibiotic resistance. The first chapter of this thesis also provides the reader with a comprehensive overview of the current diagnostic tests used in the clinic, including the methodologies of each test and the advantages and disadvantages of each test, thus allowing the reader to better understand the relative advantages and disadvantages of the LIBS technique as it is used here.

The second chapter of my thesis provides an overview of the work that has been done in the field of LIBS with special emphasis given to LIBS done on biological specimens. This chapter describes the progression of the field from the initial studies conducted in 2003 focused on single-cell detection of *Bacillus anthracis,* to the first applications of chemometric algorithms, and finally to the study of more complex systems of cells. Chapter 3 of my thesis will introduce and discuss the LIBS theory, experimental setup, and methodologies used throughout the experiment. It will also contain an introduction to the chemometric algorithms used in this work. Chapter 4 will discuss reducing the background signal of our bacterial spectra by investigating more rigorous cleaning procedures, differing ablation substrate, and the water used for bacterial storage and preparation. Outlier rejection techniques are also investigated with the goal of improving discrimination. Chapter 5 discusses classification of bacteria by species using chemometric algorithms. Several preprocessing methods are applied to the data[21] before classification for both 5-class and 3-class tests. A new chemometric algorithm was developed and applied to the data which improved results. Chapter 6 describes detection and diagnosis of bacteria in sterile blood and urine. Detection of bacteria in sterile blood and urine is done by using chemometrics to compare spectra of sterile blood and urine to those contaminated with bacteria. Diagnosis between bacterial species present in either blood or urine is demonstrated in 3-class and 4-class tests using the same techniques used in Chapter 5. High accuracy of detection in clinical fluids is achieved, as well as in discrimination between species. Finally, chapter 7 focuses on studying the efficacy of dual centrifugation.

However, while studying dual-stage centrifugation, it was found that the deposition of cells on filter was not occurring as consistently as originally thought. Through further investigation, evidence suggested that the issues with our deposition procedure were the poor seal between the cone and filter, and cells going around the filter instead of landing on it. Suggestions for future studies are made in this chapter to prevent this flaw in deposition and improve sample reproducibility.

References

[1] The World Health Report 1996: Fighting disease, fostering development; report of the director-general. (1997). *Population and Development Review*, *23*(1), 203. https://doi.org/10.2307/2137484.

[2] Center for Disease Control and Prevention. (2019). *Antibiotic Resistant Threats in the United States, 2019*. U.S. Department of Health and Human Services. https://doi.org/10.15620/cdc:82532.

[3] Troeger, C. et al., (2017). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases*, *17*(11), 1133–1161. https://doi.org/10.1016/s1473-3099(17)30396-1.

[4] Washington, J. A. (1996). Principles of Diagnosis. In S. Baron (Ed.), *Medical microbiology* (4th ed.). School of Medicine, University of Texas Medical Branch at Galveston.

[5] Engleberg, N. C. (2007). Diagnostic Principles. In N. C. Engleberg, V. J. DiRita, T. Dermody, & M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 503–512). Lippincott Williams & Wilkins.

[6] Rajapaksha, P., Elbourne, A., Gangadoo, S., Brown, R., Cozzolino, D., & Chapman, J. (2019). A review of methods for the detection of pathogenic microorganisms. *The Analyst*, *144*(2), 396–411. https://doi.org/10.1039/c8an01488d.

[7] Boyanova, L. (2017). Direct gram staining and its various benefits in the diagnosis of bacterial infections. *Postgraduate Medicine*, *130*(1), 105–110. https://doi.org/10.1080/00325481.2018.1398049.

[8] Salyers, A. A., Whitt, D. D. (2002).The Second Line of Defense: Antibodies and Cytotoxic T Cells. In *Bacterial Pathogenesis: a Molecular Approach* (2nd ed., pp. 84-100). ASM Press.

[9] Vainionpää, R., Waris, M., & Leinikki, P. (2015). Diagnostic Techniques: Serological and Molecular Approaches. *Reference Module in Biomedical Sciences*, B978-0-12-801238-3.02558-7. https://doi.org/10.1016/B978-0-12-801238-3.02558-7.

[10] *Overview of elisa*. Thermo Fisher Scientific - US. (n.d.). Retrieved December 2021, from https://www.thermofisher.com/ca/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html.

[11] McPherson, M. J., & Møller, S. G. (2014). Understanding PCR. In *PCR* (2nd ed., pp. 9–20). Taylor and Francis.

[12] Ratan, Z. A., Zaman, S. B., Mehta, V., Haidere, M. F., Runa, N. J., & Akter, N. (2017). Application of Fluorescence In Situ Hybridization (FISH) Technique for the Detection of

Genetic Aberration in Medical Science. *Cureus*, *9*(6), e1325. https://doi.org/10.7759/cureus.1325.

[13] O'Connor, C. (2008). *Fluorescence In Situ Hybridization (FISH)*. Nature news. Retrieved December 2021, from https://www.nature.com/scitable/topicpage/fluorescence-in-situ-hybridization-fish-327/.

[14] Prudent, E., & Raoult, D. (2018). Fluorescence in situ hybridization, a complementary molecular tool for the clinical diagnosis of infectious diseases by intracellular and fastidious bacteria. *FEMS Microbiology Reviews*, *43*(1), 88–107. https://doi.org/10.1093/femsre/fuy040.

[15] McPherson, M. J., & Møller, S. G. (2014). Understanding PCR. In *PCR* (2nd ed., pp. 9–20). Taylor and Francis.

[16] Váradi, L., Luo, J. L., Hibbs, D. E., Perry, J. D., Anderson, R. J., Orenga, S., & Groundwater, P. W. (2017). Methods for the detection and identification of pathogenic bacteria: Past, present, and future. *Chemical Society Reviews*, *46*(16), 4818–4832. https://doi.org/10.1039/c6cs00693k.

[17] Rebrosova, K., Samek, O., Kizovsky, M., Bernatova, S., Hola, V., & Ruzicka, F. (2022). Raman spectroscopy—a novel method for identification and characterization of microbes on a single-cell level in clinical settings. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.866463.

[18] Kostrzewa, M., & Maier, T. (2017). Criteria for Development of MALDI-TOF Mass Spectral Database. In H. N. Shah & S. E. Gharbia (Eds.), *MALDI-TOF and Tandem MS for Clinical Microbiology* (1st ed., pp. 39–51). Wiley.

[19] Lussier, F., Thibault, V., Charron, B., Wallace, G. Q., & Masson, J.-F. (2020). Deep Learning and Artificial Intelligence Methods for Raman and surface-enhanced Raman scattering. *TrAC Trends in Analytical Chemistry*, *124*. https://doi.org/10.1016/j.trac.2019.115796.

[20] McGoverin, C., Steed, C., Esan, A., Robertson, J., Swift, S., & Vanholsbeeck, F. (2021). Optical methods for bacterial detection and characterization. *APL Photonics*, *6*(8). https://doi.org/10.1063/5.0057787.

Chapter 2: A Review of LIBS on Bacteria

Despite the many advancements of modern medicine in both diagnosis and treatment of pathogenic microorganisms, they are still a great threat to the human population. It is recognized that there is a need for a diagnostic tool that can quickly and accurately detect bacterial pathogens for more targeted treatment.[22] As discussed in the previous chapter, there are several existing techniques that provide great sensitivity and specificity, but typically at the cost of time, money, or both. Culturing is highly specific and provides a wealth of information about the pathogen, but often takes days to perform and is limited; not all bacteria can be cultured. Methods that require culturing before analysis, such as MALDI-TOF MS, suffer from the same shortcomings. A common disadvantage across many of the techniques, including culturing, FISH, and Gram staining is that they require operator expertise in performing the tests and interpreting results. Serological methods are easy and cost efficient to perform, but they often require some expertise in interpreting the result due to high cross-contamination potential and the presence of non-pathogenic bacteria. Finally, MALDI-TOF MS and Raman require little expertise to run but require huge capital investments and maintenance.[23]

LIBS is a solution proposed for many of these drawbacks in diagnosing bacterial pathogens and a significant effort has been made by several groups to make LIBS a viable option for pathogen diagnosis. This chapter will expand upon the previous efforts made by other groups investigating LIBS on pathogens. In this thesis, the proposed solution of LIBS is expanded upon by improving detection of bacteria and investigating LIBS on bodily fluids such as blood and urine to develop a point of care diagnostic tool. For a greater in-depth review on applications of LIBS on tissues, cancers, dentistry, surgery and other medical implementations, the reader is directed to *Laser-Induced Breakdown Spectroscopy: Theory and Applications*, Chapter 17, "Biomedical Applications of LIBS", as well as "A Review of the Use of Laser-Induced Breakdown Spectroscopy for Bacterial Classification, Quantification, and Identification," by Steven J. Rehse.

2.1 An Introduction to Laser-Induced Breakdown Spectroscopy

Laser-induced breakdown spectroscopy (LIBS) is a rapid minimally-destructive spectroscopic technique used to determine the elemental composition of a sample.[24] A typical LIBS experiment will focus a pulsed laser with pulse duration of a nanosecond, picosecond, or femtosecond onto a target material. The focused pulse ablates the target and ejects particles into the air directly above the target. The energy from the rest of the incoming pulse is absorbed by this newly formed cloud, creating a plasma. As the plasma cools, it gives off light that is characteristic of the elements in the plasma, which is collected by an optical fibre connected to a spectrometer, which results in a time-resolved optical spectrum. The spectrum can then be used to quantify the concentration of the elements present or used as a fingerprint to qualitatively identify the material ablated.[24] A summary of the process of LIBS is shown in Figure 2.1.[25]

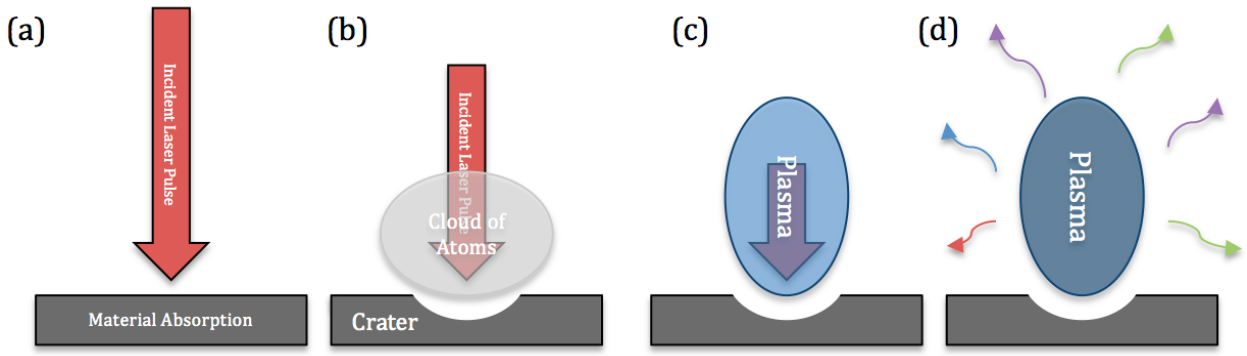

Figure 2.1: A summary of the LIBS process: (a) the incoming laser pulse, (b) laser pulse ablates a portion of the tagret forming a cloud of atoms above the analyte, (c) the remainder of the pulse is absorbed by the cloud, creating a plasma, and (d) the plasma gives off light as it cools. Adapted from ref [25].

The technique of LIBS is highly flexible, it can be used on solid, liquid, and gas phases and can detect very small amounts of material, typically parts per million or lower. It is a versatile technique that can detect a wide range of elements using either UV, visible, or infrared light. The technique is extremely fast; the time from the start of the pulse to displaying the data occurs in less than 1 second. With handheld LIBS devices, little expertise is needed because of the ease of use, and with chemometric algorithms for data analysis, virtually no expertise is needed for interpretation of results.[26]

Because of the high flexibility and high sensitivity, LIBS has found applications in several fields of study.[27] In the category of environment quality control, it has been used to study soil composition and quality extensively. A major application of soil study is agriculture where soil is tested with LIBS to determine which crops it will sustain best.[28] LIBS is used for geochemical fingerprinting to determine which geological processes were involved in the formation of a given rock or mineral.[29] LIBS at a distance (called "remote LIBS" or "stand-off LIBS") is used for the detection and identification of trace explosives.[30] LIBS on metals and quantifying metal alloys is another lucrative field of application.[31]

In the food industry and health sector, LIBS has been investigated for rapid pathogen detection and diagnosis. Rapid, real-time, and portable detection of pathogens is sorely needed in the food industry, as food-borne illnesses account for approximately 300,000 hospitalizations and several thousand deaths per year.[32] As well, the annual cost due to strain on the health care system and lost productivity is estimated between $2-$6 billion.[33] Currently the methods used to detect bacteria in food are ELISA and culturing, which are time consuming.[34] Work was done by Barnett et al. specifically for the food industry to detect *Salmonella enterica* in milk, chicken broth, and brain heart fusion using LIBS. Efficient identification was achieved with a fourth harmonic (266 nm) nanosecond LIBS system.[35]

In the health sector, proof of concepts involving the discrimination between pollen spores and *Bacillus anthracis* spores, as well as the detection of yeast and fungal spores have led to several advancements being made in diagnosing bacterial infections.[36]

2.2 A Review of the Progress Made in LIBS on Bacteria

Studies using LIBS on bacteria first appeared in 2003, in part motivated by the bioterrorism attacks with *Bacillus anthracis*, or 'anthrax' in 2001. These studies consisted mostly of proof-of-concept and focused heavily on the ability to achieve appreciable signal.[36] Morel et al. tested 6 different types of bacteria and showed that compressed bacterial pellets had a homogenous composition.[37] Hybl et al. demonstrated single cell detection capability when bacteria was delivered through dense aerosol streams. They also demonstrated discrimination using principal component analysis (PCA) between bacteria

B. globigii spores, pollen spores, and fungal spores; this is shown in Figure 2.2.[38] These early works laid the foundation and demonstrated proof-of-concept for continued studies of LIBS on bacteria. These studies however had some inherent flaws; many of the testing substrates involved very high numbers of bacteria compressed into unrealistic freeze-dried pellets. As well, no method of discriminating between species that produced highly similar spectra was investigated. However, it was demonstrated early on that spectra from other interferents such as pollen spores could be distinguished from bacterial spores, as shown by Samuels et al.[36]

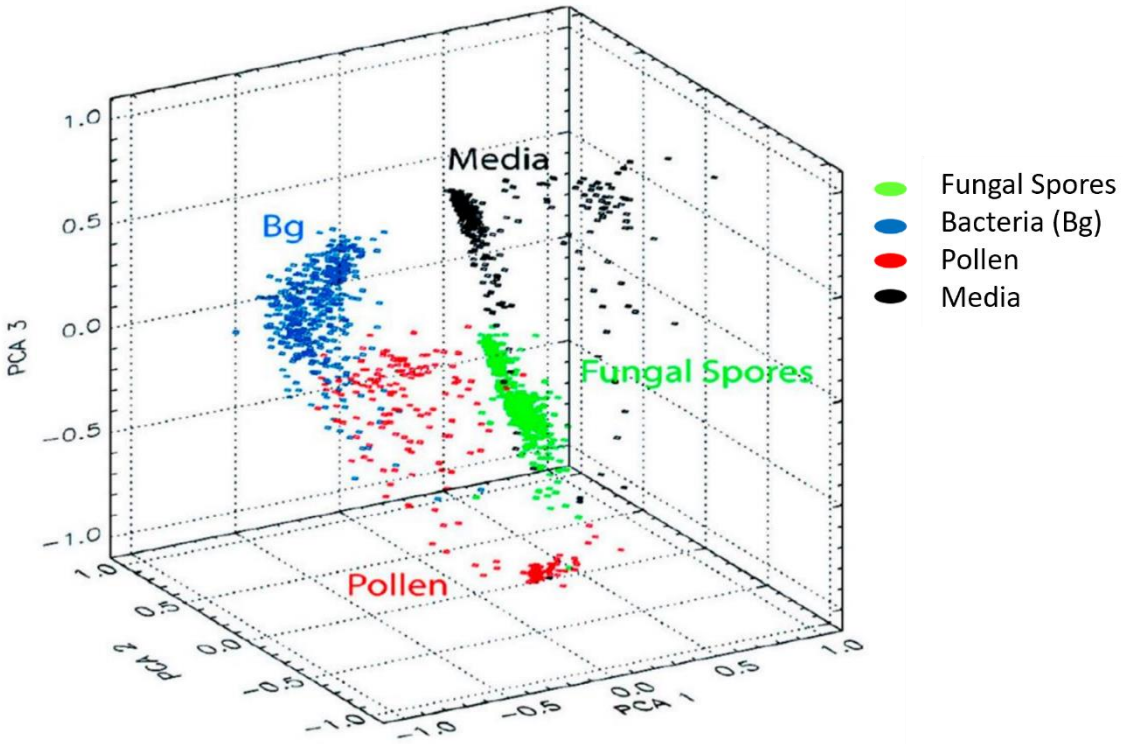

Figure 2.2: Principal component analysis (PCA) was used to generate a 3-dimensional plot to show the discrimination between media, pollen, fungus, and bacteria (Bg). Figure adapted from ref [38].

It was recognized early on that chemometric algorithms would play a large role in discriminating between bacterial species. A detailed description of how these chemometric algorithms are specifically used to discriminate bacterial LIBS spectra is provided in Chapter Three. In 2007, Merdes et al. began using principal component analysis (PCA) to distinguish between *B. anthracis* and other biological interferents, namely pollen spores, egg albumin, molds, and starch. Using the MATLAB PCA program and spectra with 2048 channels they were able to demonstrate successful identification of *B. anthracis* with a false

negative rate of 3% and false positive rate of 1%.[39] Development of other chemometric algorithms by the United States Army Research Laboratory followed this, with focus mostly on developing and testing artificial neural networks (ANN) and combining these with linear regression models. They focused on identifying the pathogenic bacteria *Bacillus atrophaeus* on unknown background substrates such as road dust or egg albumin.[40] Further progression in chemometrics was accomplished by Cisewski et al. who investigated using linear models to pre-process data before classification using a support vector machine (SVM). The goal of this approach was to categorize the analyte as a *Bacillus* spore. This method demonstrated that pre-processing of data could work and performed well with a classification error of 3%.[41]

Elucidating the appropriate parameters and how to use the chemometric algorithms followed shortly after development and testing with work done by Gottfried et al. and Munson et al. on down-selecting the number of variables used for testing. The reason for this was to lessen computing requirements and to ensure that the model was not being overfitted.[42,43] Munson et al. compared the chemometric algorithms PCA and soft independent modelling of class analogy (SIMCA) using *Bacillus* organisms and common interferents. They reduced the number of variables used in the algorithms by selecting only the lines in spectra that were pertinent to their analysis and disregarding any lines that gave them no information. Removing lines as opposed to a full-spectrum analysis is still being investigated today.[42] Gottfried et al. used this idea of variable down selection and partial least squares discriminant analysis (PLSDA) to discriminate between *Staphylococcus aureus, Escherichia coli, Bacillus atrophaeus,* MS-2 bacteriophage, and α-Hemolysin. By carefully choosing variables to down-select, they achieved excellent discrimination of the 5 aforementioned targets.[43]

Efforts made in testing the limits of chemometric algorithms in the interest of pursuing clinical diagnostic applications were occurring in parallel. Using discriminant function analysis (DFA) and down-selecting to 19 independent variables, Rehse et al. attempted to classify 3 strains of *E. coli,* environmental mold, and *Candida albicans* yeast. This was the first strain level discrimination between bacteria.[44] The effect of culturing in a different media on discriminatory ability was tested using the same species previously mentioned

and adding pathogenic enterohemorrhagic *E. coli* to the analysis. 100% discrimination was observed despite the different culturing media.[45] Finally, the same technique was used to determine the effect of MacConkey agar which contains bile salts that alter bacterial membranes. Two strains of *E. coli* were used along with *Pseudomonas aeruginosa.* The analysis demonstrated that MacConkey agar media was the only media that affected discrimination.[46]

Building on this, Rehse et al. purposely altered the membrane chemistry of Gram-negative bacterial species by using different culturing media. They found that LIBS could monitor the changing concentrations of the cations Ca, Mg, and Na and concluded that LIBS was possibly a serological or surface antigen-based detection technique.[47] Early work was also done by Baudelet et al. to compare nanosecond LIBS to femtosecond LIBS. Genus level discrimination was accomplished between *Acinetobacter baylyi, Bacillus subtilis, Erwinia chrysanthemi, Escherischia coli,* and *Shewanella oneidensis* using the elements Na, Mg, Ca, P, K, and Fe.[48] The advantages of femtosecond LIBS were also noted, namely the intense molecular CN band emission.[49] These last few experiments represent an important result: LIBS relies on membrane chemistry of bacteria for classification.

The effect of sterilization and metabolic state on classification of bacterial LIBS targets was also investigated by Rehse et al. A strain of *E. coli* (Strain C) was split into two groups, one was exposed to ultraviolet radiation and the other was autoclaved. Exposing bacteria to UV light or autoclaving them sterilizes them and makes them completely safe. LIBS measurements were made for each case, and it was demonstrated that though the bacteria were sterilized, no loss of signal was observed after exposure to UV light. The acquired spectra were then plugged into a chemometric algorithm with another *E. coli* strain (ATCCC 25922) and *M. smegmatis* where 100% correct classification was observed for the sterilized strains, showing that before working with bacteria clinically they can be sterilized and rendered completely safe without sacrificing accuracy of diagnosis.[50] This is shown in Figure 2.3.

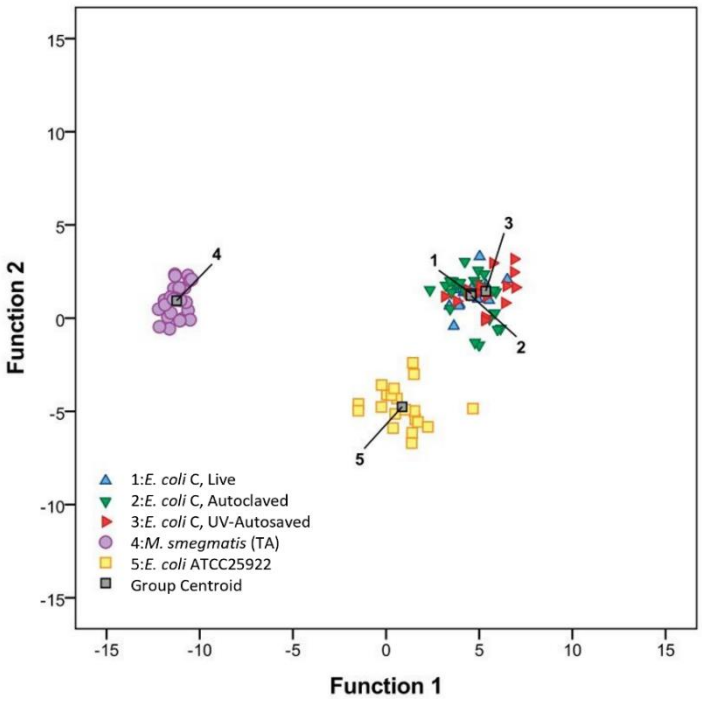

Figure 2.3: A DFA discrimination between UV-exposed, autoclaved, and live E. coli, along with M. smegmatis and a second strain of E. coli. The robust discrimination between E. coli (UV, autoclaved, live, ATCC25922) and M. smegmatis shows that sterilized bacteria produce the same result as live bacteria. Figure adapted from ref [50].

Next, Rehse et al. investigated LIBS on non-culturable bacteria. They deposited *S. viridians* and *E. coli* (Strain C) on agar media containing no nutrients and left the bacteria to grow at 21 °C. Because of the lack of nutrients and lower temperatures, the bacteria went into a non-reproducing or metabolically inactive state. The bacteria therefore were not culturable. Spectra were acquired after several days, and it was shown that the spectra of the non-culturable specimens were not altered compared to their previous spectra. As well, discrimination was achieved with high accuracy. This result demonstrates that LIBS is capable of detecting and diagnosing non-culturable bacteria, which is highly applicable in clinical laboratories.[50]

Important and relevant clinical questions have also been investigated, namely the efficacy of the technique in mixed cultures and the efficacy of LIBS when other solutes or fluids are present. Rehse et al. investigated the effectiveness of LIBS on mixed cultures, or cultures containing two or more bacterial species.[51] This is clinically relevant because the body contains several species of non-pathogenic microorganisms that help the body, particularly in the GI tract and on the skin. In these environments, detecting a pathogen

amongst other species could prove to be a challenge.[52] Two-component bacterial mixtures were prepared containing *E. coli* and *M. smegmatis*. The ratio of these mixtures was altered to determine where the threshold of accurate detection was. It was found that when the mixture consisted of 80% or more of the pathogen of interest, classification was accomplished with good accuracy. For any concentration below 80%, the discrimination capability dropped rapidly, reaching 50% for 50:50 mixtures. This result was duplicated with mixtures of *E. coli* and *Enterobacter cloacae*.[51] While this seems discouraging, the reader must take into account the fact that infecting microorganisms must compete with the native fauna for resources to establish an infection, so these higher ratios may be more realistic. Another important result that followed from this was the investigation of serial dilutions of bacterial species. Rehse et al. showed that dilutions of *M. smegmatis* could still be accurately discriminated from other strains of *M. smegmatis,* indicating that multiple concentrations of bacteria could be identified and more importantly, small concentrations of bacteria could be identified.[51]

Rehse et al. has also shown that the presence of solutes and minerals present in urine do not affect the classification of pathogens. *S. epidermis* was harvested from sterile urine and tested against *S. aureus, S. saprophyticus,* and *S. epidermis* harvested from water. DFA was able to classify the *S. epidermis* from urine with the *S. epidermis* from water 100% of the time.[53] This addresses the question of the efficacy of LIBS in the presence of other bodily fluids and shows that point-of-care diagnosis of pathogens present in urine and potentially other sterile bodily fluids such as blood and cerebral spinal fluid is possible.

Finally, excellent genus and strain level identification using chemometric algorithms has been shown. Rehse et al. showed genus level discrimination between spectra from 5 different species that span 13 strains, including *Escherichia, Enterobacter, Staphylococcus, Streptococcus,* and *Mycobacterium.* Using DFA, the sensitivity and specificity for classification was obtained both for the 5-class test and the 13-class test. As well, external validation was used to construct truth tables.[54] External validation means that the spectrum being used as a test, and those collected at the same time as it, were withheld or not used in the library that they are being tested against. The results of this are shown in Figure 2.4; sensitivities were approximately 85% and specificities of 95% and above were

observed. This model also showed that LIBS and chemometrics were capable of differentiating between two species that were highly similar. For *Escherichia,* 89.97% of the 299 spectra classified correctly, and 7.02% were classified incorrectly as *Enterobacter*, while the remaining spectra were incorrectly classified as the remaining species.[53] Multari et al. also demonstrated strain level identification with *E. coli,* 3 similar strains of methicillin resistant *Staphylococcus aureus* (MRSA), and one unrelated strain of MRSA. They used partial least squares (PLS) algorithms with a flowchart architecture which put the spectra through a series of 'yes' or 'no' tests that tested spectra with progressively less variation than the last step. This methodology resulted in 100% classification accuracy. [55]

LIBS on viruses has been investigated previously, but little work has been done in this field. LIBS is a mass-dependent technique, and viral particles are approximately 10^9 times smaller (in volume) than the bacterial cell. As well, they lack the inorganic metals Mg, Ca, Na, and K that bacteria possess and that chemometric algorithms use for discrimination. The presence of these metals is what has provided strong signals from bacteria.[56] However, some work was done to investigate LIBS detection for the MS-2 bacteriophage. A significant result from Multari et al. showed the discrimination of 4 strains of *hantavirus* that are responsible for a variety of infections.[56]

There has been impressive progress in the field of LIBS on bacteria for both food industry and clinical use. LIBS can address the issue of speed of diagnosis as it takes only a few seconds to perform. LIBS can achieve accurate discrimination, regardless of the metabolic state or sterility of bacteria, making it more universally applicable and the handling of specimens safer.[52] LIBS can also detect and diagnose non-culturable strains, which is a huge improvement to other techniques such as MALDI-TOF MS, culturing, and FISH that rely on culturing.[50] LIBS could detect bacteria present in mixtures and also present in bodily fluids, which allows for better point-of-care diagnostics.[32,34] LIBS systems and particularly LIBS handheld systems are relatively less costly than other techniques, and they will likely require less operator training and expertise since the results from chemometrics are readily interpretable. In summary, LIBS addresses many of the drawbacks of the other techniques used today and could represent a fast, cost effective, and accurate way to diagnose bacterial infections.

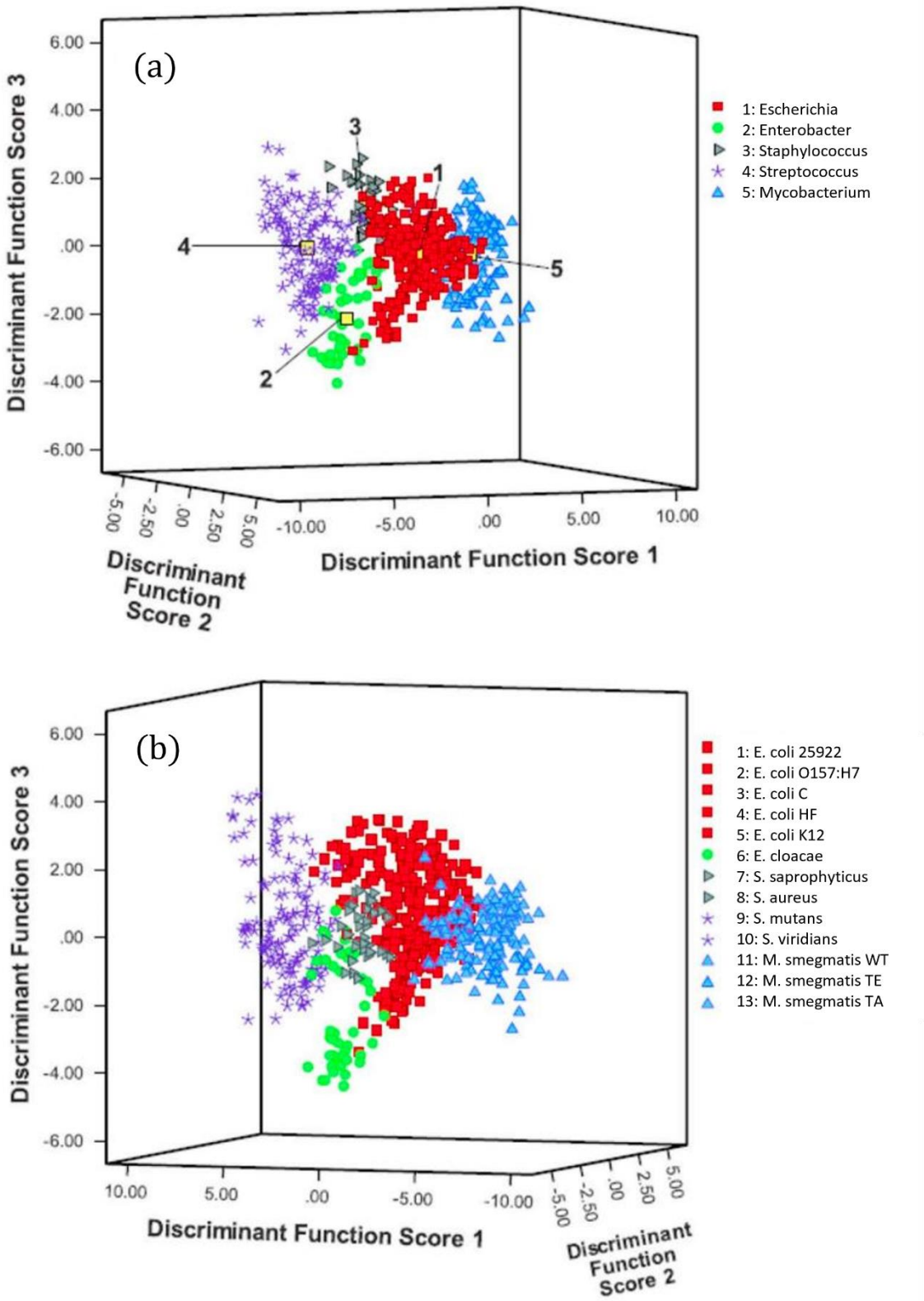

Figure 2.4: Discriminant function analysis (DFA) done on a 5-class genus level discrimination (a) and 13 class strain level (b). 699 spectra were used and plotted using the first 3 discriminant function scores. Figure adapted from ref [54].

References

[22] Center for Disease Control and Prevention. (2019). *Antibiotic Resistant Threats in the United States, 2019*. U.S. Department of Health and Human Services. https://doi.org/10.15620/cdc:82532.

[23] Engleberg, N. C. (2007). Diagnostic Principles. In N. C. Engleberg, V. J. DiRita, T. Dermody, & M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 503–512). Lippincott Williams & Wilkins.

[24] Cremers, D., Radziemski, L.J. (2013). Introduction, *Handbook of Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 1-18). Wiley.

[25] Paulick, A. (2018). *Development of Laser-Induced Breakdown Spectroscopy as a Rapid Diagnostic Tool for Bacterial Infection.* [Master's thesis, University of Windsor].

[26] Thakur, S. N., Singh, J.P. (2020). Fundamentals of LIBS and Recent Developments. In Thakur, S. N., & Singh, J.P. (Eds.), *Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 3-19). Elsevier Science.

[27] Lazcka, O., Del Campo F.J., & Muño F.X. (2007). Pathogen detection: a perspective of traditional methods and biosensors, *Biosensensors and Bioelectronics 22*(7), 1205–1217. https:// doi.org/10.1016/j.bios.2006.06.036.

[28] Gondal, M. A., Dastageer M.A. (2016). Elemental Analysis of Soils by Laser Induced Breakdown Spectroscopy. In Musazzi, S., Perini, U. (Eds.), *Laser Induced Breakdown Spectroscopy: Theory and Applications* (Vol. 182., pp. 293-306). Springer Berlin.

[29] Hark, R.R., Harmon, R. S. (2016). Geochemical Fingerprinting Using LIBS. In Musazzi, S., Perini, U. (Eds.), *Laser Induced Breakdown Spectroscopy: Theory and Applications* (Vol. 182, pp. 309-344). Springer Berlin.

[30] Moros, J., Fortes, F. J., Vadillo, J.M., & laserna, J. (2016). LIBS Detection of Explosives in Traces. In *Laser-induced breakdown spectroscopy theory and applications* (Vol. 182, pp. 349-375). Springer Berlin.

[31] Musazzi, S., Perini, U., Legnaioli, S., Lorenzetti, G., Pardini, L., Cavalcanti, G. H., & Palleschi, V. (2016). Applications of LIBS to the Analysis of Metals. In *Laser-induced breakdown spectroscopy theory and applications* (Vol. 182, pp. 169–191). Springer Berlin.

[32] Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M., & Tauxe, R. V. (1999). Food-Related Illness and Death in the United States. *Emerging Infectious Diseases*, *5*(5), 607–625. https://doi.org/10.3201/eid0505.990502.

[33] Alocilja, E. C., & Radke, S. M. (2003). Market analysis of biosensors for food safety. *Biosensors and Bioelectronics*, *18*(5–6), 841–846. https://doi.org/10.1016/s0956-5663(03)00009-5.

[34] Cohn, G. E. Systems and technologies for clinical diagnostics and drug discovery (San Jose CA, 26-27 January 1998). In *SPIE proceedings series*. SPIE.

[35] Barnett, C., Bell, C., Vig, K., Akpovo, A. C., Johnson, L., Pillai, S., & Singh, S. (2011). Development of a LIBS assay for the detection of Salmonella enterica serovar Typhimurium from food. *Analytical and Bioanalytical Chemistry*, *400*(10), 3323–3330. https://doi.org/10.1007/s00216-011-4844-3.

[36] Samuels, A. C., DeLucia, F. C., McNesby, K. L., & Miziolek, A. W. (2003). Laser-induced breakdown spectroscopy of bacterial spores, molds, pollens, and protein: initial studies of discrimination potential. *Applied Optics*, *42*(30), 6205–6209. https://doi.org/10.1364/ao.42.006205.

[37] Morel, S., Leone, N., Adam, P., & Amouroux, J. (2003). Detection of bacteria by time-resolved laser-induced breakdown spectroscopy. *Applied Optics*, *42*(30), 6184–6191. https://doi.org/10.1364/ao.42.006184.

[38] Hybl, J. D., Lithgow, G. A., & Buckley, S. G. (2003). Laser-induced breakdown spectroscopy detection and classification of biological aerosols. *Applied Spectroscopy*, *57*(10), 1207–1215. https://doi.org/10.1366/000370203769699054.

[39] Merdes, D. W., Suhan, J. M., J.M.K., D.M.H., & Bradley, W. R. (2007). The investigation of laser-induced breakdown spectroscopy for detection of biological contaminants on surfaces. *Spectroscopy*, *22*, 28–38.

[40] Snyder, E. G., Munson, C. A., Gottfried, J. L., de Lucia, Jr., F. C., Gullett, B., & Miziolek, A. (2008). Laser-induced breakdown spectroscopy for the classification of unknown powders. *Applied Optics*, *47*(31), G80–G87. https://doi.org/10.1364/ao.47.000g80.

[41] Cisewski, J., Snyder, E., Hannig, J., & Oudejans, L. (2012). Support vector machine classification of suspect powders using laser-induced breakdown spectroscopy (LIBS) spectral data. *Journal of Chemometrics*, *26*(5), 143–149. https://doi.org/10.1002/cem.2422.

[42] Munson, C. A., de Lucia, F. C., Piehler, T., McNesby, K. L., & Miziolek, A. W. (2005). Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent simulants. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *60*(7–8), 1217–1224. https://doi.org/10.1016/j.sab.2005.05.017.

[43] Gottfried, J. L. (2011). Discrimination of biological and chemical threat simulants in residue mixtures on multiple substrates. *Analytical and Bioanalytical Chemistry*, *400*(10), 3289–3301. https://doi.org/10.1007/s00216-011-4746-4.

[44] Diedrich, J., Rehse, S. J., & Palchaudhuri, S. (2007). *Escherichia coli* identification and strain discrimination using nanosecond laser-induced breakdown spectroscopy. *Applied Physics Letters*, *90*(16), 163901. https://doi.org/10.1063/1.2723659.

[45] Diedrich, J., Rehse, S. J., & Palchaudhuri, S. (2007b). Pathogenic *Escherichia coli* strain discrimination using laser-induced breakdown spectroscopy. *Journal of Applied Physics*, *102*(1), 014702. https://doi.org/10.1063/1.2752784.

[46] Rehse, S. J., Diedrich, J., & Palchaudhuri, S. (2007). Identification and discrimination of Pseudomonas aeruginosa bacteria grown in blood and bile by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *62*(10), 1169–1176. https://doi.org/10.1016/j.sab.2007.07.008.

[47] Rehse, S. J., Jeyasingham, N., Diedrich, J., & Palchaudhuri, S. (2009). A membrane basis for bacterial identification and discrimination using laser-induced breakdown spectroscopy. *Journal of Applied Physics*, *105*(10), 102034. https://doi.org/10.1063/1.3116141.

[48] Baudelet, M., Yu, J., Bossu, M., Jovelet, J., Wolf, J. P., Amodeo, T., Fréjafon, E., & Laloi, P. (2006). Discrimination of microbiological samples using femtosecond laser-induced breakdown spectroscopy. *Applied Physics Letters*, *89*(16), 163903. https://doi.org/10.1063/1.2361270.

[49] Baudelet, M., Guyon, L., Yu, J., Wolf, J. P., Amodeo, T., Fréjafon, E., & Laloi, P. (2006). Spectral signature of native CN bonds for bacterium detection and identification using femtosecond laser-induced breakdown spectroscopy. *Applied Physics Letters*, *88*(6), 063901. https://doi.org/10.1063/1.2170437.

[50] Mohaidat, Q., Palchaudhuri, S., & Rehse, S. J. (2011a). The effect of bacterial environmental and metabolic stresses on a laser-induced breakdown spectroscopy (LIBS) based identification of Escherichia coli and Streptococcus viridans. *Applied Spectroscopy*, *65*(4), 386–392. https://doi.org/10.1366/10-06178.

[51] Rehse, S. J., Mohaidat, Q. I., & Palchaudhuri, S. (2010). Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: The effect of mixed cultures and sample dilution on bacterial identification. *Applied Optics*, *49*(13), C27–C35. https://doi.org/10.1364/ao.49.000c27.

[52] Engleberg, N. C. (2007). Diagnostic Principles. In N. C. Engleberg, V. J. DiRita, T. Dermody, & M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 503–512). Chapter, Lippincott Williams & Wilkins.

[53] Mohaidat, Q. I., Sheikh, K., Palchaudhuri, S., & Rehse, S. J. (2012). Pathogen identification with laser-induced breakdown spectroscopy: The effect of bacterial and biofluid specimen contamination. *Applied Optics*, *51*(7), B99–B107. https://doi.org/10.1364/ao.51.000b99.

[54] Rehse, S. J. (2019). A review of the use of laser-induced breakdown spectroscopy for bacterial classification, quantification, and identification. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *154*, 50–69. https://doi.org/10.1016/j.sab.2019.02.005.

[55] Multari, R. A., Cremers, D. A., Dupre, J. M., & Gustafson, J. E. (2010). The use of laser-induced breakdown spectroscopy for distinguishing between bacterial pathogen species and strains. *Applied Spectroscopy*, *64*(7), 750–759. https://doi.org/10.1366/000370210791666183.

[56] Multari, R. A., Cremers, D. A., & Bostian, M. L. (2012). Use of laser-induced breakdown spectroscopy for the differentiation of pathogens and viruses on substrates. *Applied Optics*, *51*(7), B57–B64. https://doi.org/10.1364/ao.51.000b57.

Chapter 3: Laser-Induced Breakdown Spectroscopy; Apparatus and Experimental Procedures

3.1 Theory of LIBS

Shortly after the development of the laser by Theodore Maiman in 1960, the first observation of laser-induced plasmas (LIP) was reported.[57] LIPs are created when a laser is absorbed by a solid target causing vaporization and ionization of the material in a process known as ablation.[58,59] This process will be described in more detail later. After the first observation of the LIP, the ablation and subsequent plasma formation was used by Debras-Guedon and Liodec for spectrochemical analysis of surfaces in the first laser-induced breakdown spectroscopy (LIBS) experiments.[57] The LIBS technique uses a pulsed laser with an intensity on the order of 10^9 W/cm^2 to ablate a sample which can be solid, liquid, or gas. Intensities this high are required to form a plasma.[58] The plasma consists of high-temperature atoms, ions, neutral species, and molecules which emit light at characteristic frequencies. The light is collected by a spectrometer and analyzed to determine the sample's elemental composition.[57] An example of a spectrum collected in our laboratory is shown below in Figure 3.1. The spectrum is constructed by dispersing light from the optical emissions of atoms, ions, neutrals, and molecules.

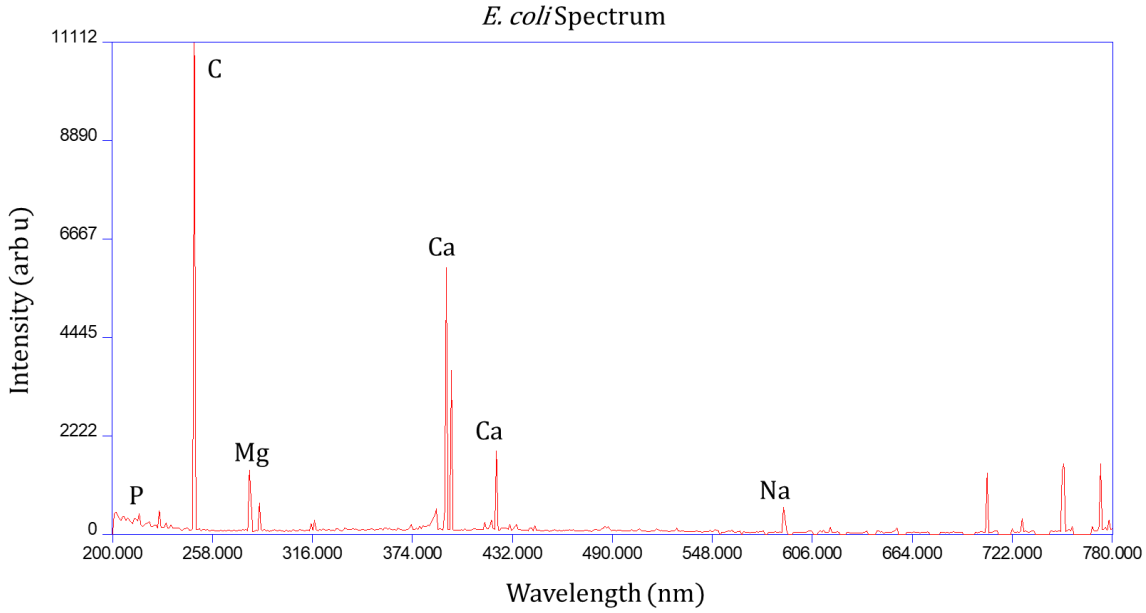

Figure 3.1: A typical E. coli spectrum collected using our LIBS apparatus.

The following sections of this chapter will describe the theory of LIBS and LIBS plasmas, including atomic transitions, optical emission, mechanism of plasma formation and collection, and plasma parameters. This chapter will also describe the experimental apparatus utilized in all the experiments including the laser, spectrometers and relevant optics. Bacteria physiology will be discussed as well as the protocols used to collect and prepare bacterial samples. Finally, a discussion on the chemometric techniques used to classify bacterial species will be provided.

3.2 Atomic Transitions

A basic understanding of LIBS requires an understanding of the atomic transitions that produce optical emission. This section will introduce the basic processes responsible for radiative emission in a LIBS plasma. There are three types of atomic transitions that can occur to produce optical emission in isolated atoms: stimulated emission, stimulated absorption, and spontaneous emission.[60] The latter process makes up most radiated photons in the LIBS plasma.

Each atom has quantized energy levels that its electrons occupy. Valence electrons can either become excited and move to a higher energy orbital, or de-excited and drop to a lower energy orbital. The processes responsible for radiative excitation and de-excitation are absorption and emission of a photon, respectively. Other methods of energy transfer do occur within a plasma, such as collisional energy transfer, but these methods will not be considered here.

Spontaneous emission occurs when a valence electron decays to a lower energy state by releasing a photon. The energy of the photon released is found by considering the energy difference between the two energy levels. Denoting the upper state j, with energy E_j, and the lower state i, with energy E_i, the spontaneous emission event between states j and i releases a photon with energy $\Delta E = E_j - E_i = h\nu_{ji}$. The transition probability of spontaneous emission from state j to i per unit time is given by the Einstein coefficient, A_{ji}.[60]

A spectral line is the result of a transition between the discrete energy levels within a specific atom or ion. The energy difference between discrete states is different for every

element, so each atom emits photons at specific and unique energies and wavelengths (since $E = \frac{hc}{\lambda}$). The measurement of spectral lines at specific wavelengths is therefore indicative of the presence of that element in the ablated sample.

Along with the atoms and ions a plasma contains unbound electrons which can freely exist in an energy continuum. These free electrons contribute to other highly relevant forms of emission within the LIBS plasma, including the transition of electrons within the energy continuum (also known as free-free transitions), or between the continuous energy state and a bound or discrete state. Transition from a continuous energy state to a bound energy state is called recombination and occurs when an atom captures a free electron and releases a photon (also known as free-bound radiation). The free-free transitions within the continuum gives rise to Bremsstrahlung radiation. Bremsstrahlung occurs when a fast-moving particle is decelerated by another charged particle, thereby losing kinetic energy. A photon with energy equivalent to the loss in kinetic energy is emitted in response.[58]

Free-free and free-bound emission is not wavelength specific, therefore it gives no information about the composition of the analyte and is useless in a LIBS measurement. In fact, it can hinder elemental identification by obscuring small element-specific transitions. To minimize the amount of continuum emission, LIBS measurements are delayed in time until the plasma has cooled, and fewer free electrons are present.[58]

3.3 Plasma Formation and Plasma Parameters

This section will describe the basic steps of LIBS plasma formation and the parameters that effect plasma formation, including temperature and electron density. For a more complete understanding of plasma formation and parameters the chapter, "Physics and Dynamics of Plasma in Laser-Induced Breakdown Spectroscopy," from the book, *Laser-Induced Breakdown Spectroscopy* is suggested. As well, additional information can be found in the chapter, "Basics of LIBS Plasma," in the *Handbook of Laser-Induced Breakdown Spectroscopy*.

3.3.1 Plasma Formation

To understand plasma formation, first the phenomenon of laser-induced breakdown needs to be discussed. The case of laser-induced breakdown in gasses will be presented first followed by laser-induced breakdown in solids.

There are two key steps to cause LIBS in gasses; free electrons must be present in the focal volume of the beam, and for ns pulses, laser irradiances must be between 10^{10}-10^{12} W/cm^2 to generate sufficient electron and ion densities in gas plasmas. Free electrons can be generated in the focal volume by multiphoton effects caused by the first few photons of the pulse. Generation of sufficient electron or ion densities can be achieved through multiphoton ionization at high laser irradiances. In general, high-powered lasers are used in conjunction with focusing optics to achieve high laser irradiances large enough to cause breakdown and ablation. Multiphoton ionization occurs when multiple photons are absorbed simultaneously to liberate an electron:

$$A + nhv \rightarrow A^+ + e^- \quad (1)$$

where A is the atom, n represents the number of photons, and e^- is the electron. The free electrons are then accelerated by the electric field and collide with other atoms to ionize them, thereby creating more free electrons. Free electrons can also acquire energy through inverse bremsstrahlung. Free electrons can collide with atoms causing ionization if the free electrons contain more energy than the ionization energy of the atom, causing cascade ionization and electron multiplication.[60,61] Electron multiplication occurs during the laser pulse which causes ionization of the gas and eventually breakdown. [62]

In solids, the process of breakdown is similar. Laser irradiances of 10^8-10^{10} W/cm^2 are used. The leading edge of the laser pulse will rapidly heat, melt, then vaporize the target material, creating a cloud of atoms above the surface; this is known as ablation. Some of the remaining laser energy then heats the cloud of atoms and initiates plasma production through multiphoton ionizations and inverse bremsstrahlung.[61] Once the plasma becomes weakly ionized, part of the laser beam will continue through to the target surface, while some of the laser beam is absorbed. As more of the laser beam is absorbed, the plasma

becomes opaque to the remainder of the beam causing plasma shielding. At this point the beam is no longer absorbed by the surface and ablation ceases. The plasma front will continue to grow along the laser beam. Plasma opacity and shielding occurs when a critical density of electrons is reached, given by

$$n_c \sim (10^{21}/\lambda^2)/cm^3 \quad (2)$$

where λ is the wavelength of the laser. After the remainder of the laser pulse is absorbed, the plasma will lose energy and decay.[62] LIBS is shown in Figure 3.2.[63]

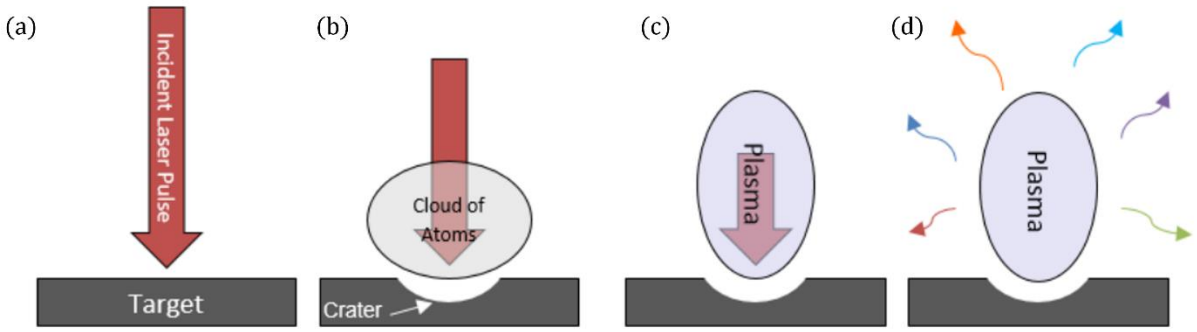

Figure 3.2: Formation of a LIBS plasma. (a) Laser pulse is incident on target and absorbed by analyte. (b) Heating then vaporizes target casuing a cloud of atoms to be ejected above analyte. (c) The cloud of atoms absorbs the lagging end of the laser pulse, ionizing the ejected atoms and forming the LIP. (d) photons representative of the vaporized elements in the analyte are emitted as the plasma cools. Adapted from ref [63].

Although LIBS can be performed in a vacuum or within a liquid environment, in all the experiments described in this thesis, the plasma expands into an argon gas environment after the ablation ceases. Argon gas provides an ambient environment for expansion and increases the temperature and electron density of the plasma.[64] Expansion of the plasma continues until a pressure equilibrium between the argon gas and the plasma is met. As the plasma expands into the surrounding argon its constituents evolve with time. At the earliest observable period, the ratio of electrons to other species is less than 10%. Having a ratio of electrons to ions being 10% or less is typical of a LIBS plasma.[65] As the plasma begins to cool, recombination events occur reducing the number of ionized species. Neutral species are more abundant towards the outer edges of the plasma, whereas multiply ionized species are more abundant at the center. In our plasmas at typical observation times, only singly ionized species are observed.

After approximately 1 μs of time has passed, continuum emission has decayed substantially, and recombination causes more neutral species to form. At this point in time emission lines from ions and neutrals dominate the spectrum, and continuum emission contributes to the noise. Typically, measurement begins at approximately 1 μs since the signal to noise ratio for elemental emission lines is high.[65] The waiting time between the ablation event and observation of light emitted by the plasma is called the gate delay (τ_d). The amount of time that the plasma is observed for is called the gate width (τ_w). The end of plasma lifetime is marked by the formation of molecules, which are rarely indicative of the molecules present in the analyte. The time evolution of a ns laser induced plasma is shown schematically in Figure 3.3 and is typically characterized by exponential decay at times later then 100-500 ns. The time progression of spectra from a plasma undergoing continuum emission to the emission of neutral lines is shown in Figure 3.4.

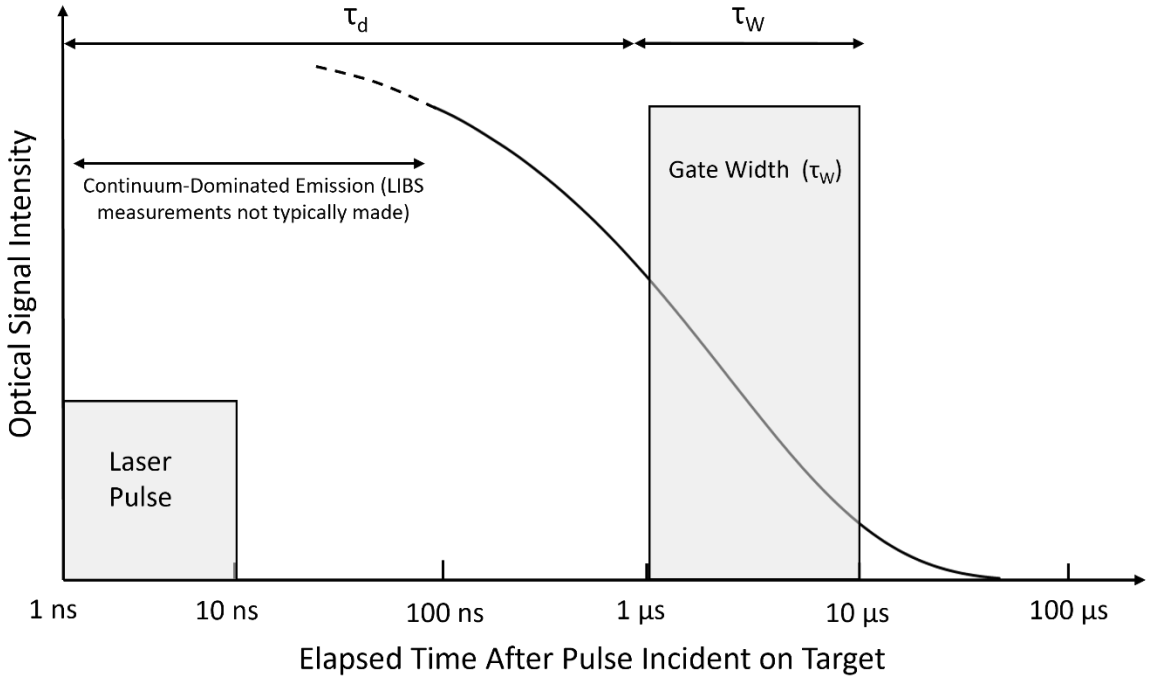

Figure 3.3: Timing diagram of LIBS plasma evolution. Plasma observation occurs during the gate width (τ_w).

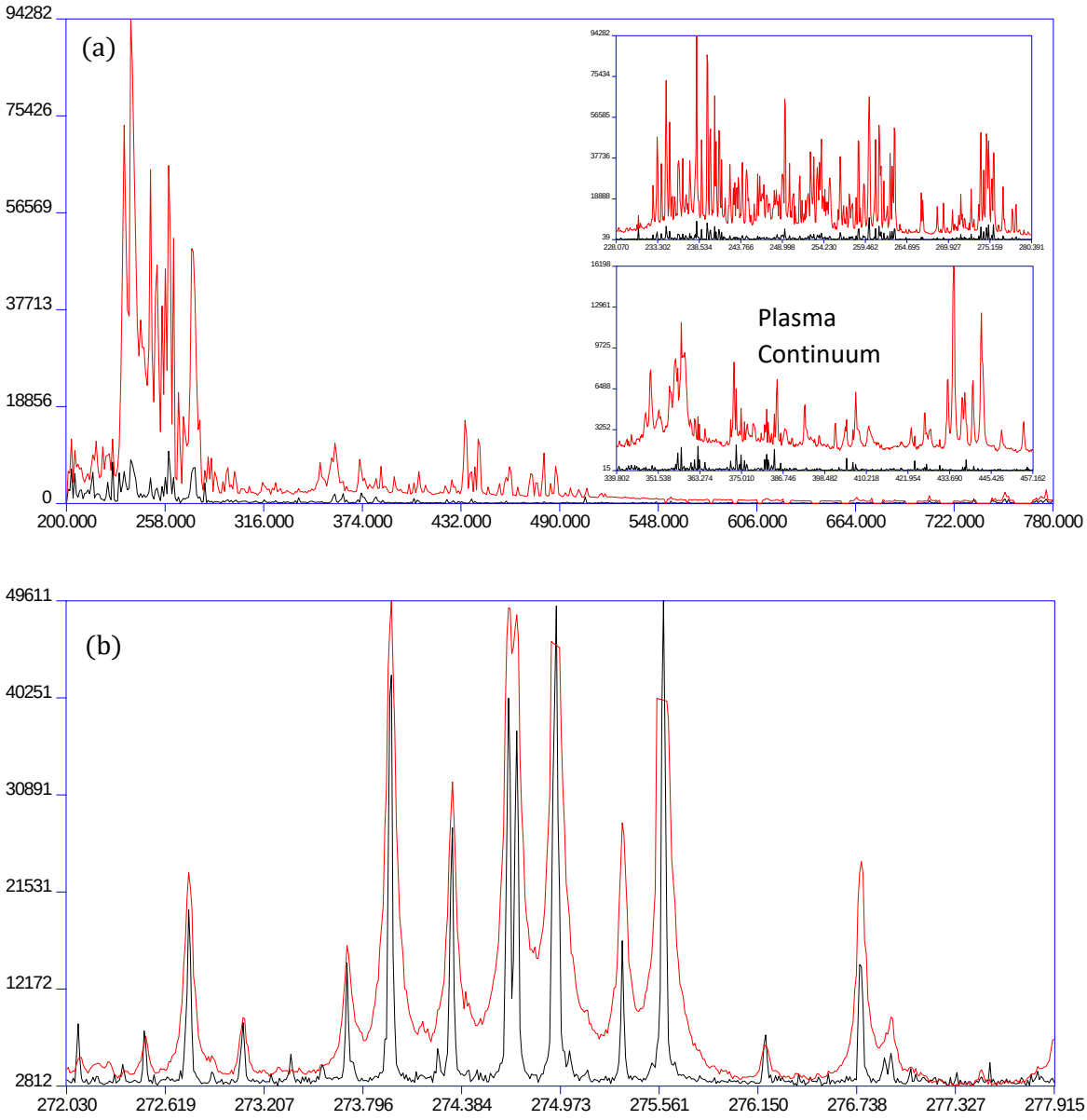

Figure 3.4: (a) Steel spectra taken at different times overlaid. The red spectrum was taken at a τ_d of 300 ns, the black spectrum was taken at a τ_d of 8 µs. The insets in the spectrum show a zoomed in view of the decaying ion lines and reduced background amplitude. (b) Steel spectrum taken at τ_d = 300 ns and τ_d = 8 µs overlaid, this view shows the spectra without absolute scaling. The lines at later times are thinner and more highly resolved, while the lines at earlier times are broader.

3.3.2 Plasma Parameters

The two parameters used to characterize a plasma are temperature and electron density. Temperature and electron density can only be determined when the plasma is in local thermodynamic equilibrium (LTE), and when it is optically thin. Local thermodynamic

equilibrium means that all species in the plasma must be at the same temperature. An optically thin plasma is one in which the emitted photons are not likely to be reabsorbed by the plasma.[65]

The temperature of the plasma can be determined with a sufficient number of spectral lines by using the Boltzmann plot method.[65] The emissivity I_{ji} of the spectral line resulting in a transition from an upper-level j to lower level i can be given by:

$$I_{ji} = \frac{hc}{4\pi\lambda_{ji}} A_{ji} L \frac{N}{Z} g_j e^{-\frac{E_j}{k_B T}} \tag{3}$$

where λ_{ji} is the wavelength, A_{ji} is the transition probability, L is the length of the plasma, N is the total number density of species in the plasma, Z is the partition function of the species, g_j and E_j are the statistical weight and energy of the upper level, k_B is the Boltzmann constant, and T is the temperature of the plasma. Rearranging equation (1) can give an expression for the temperature

$$\ln\left(\frac{I_{ji}\lambda_{ji}}{g_j A_{ji}}\right) = -\frac{E_j}{k_B T} + \ln\left(\frac{hcLN}{4\pi Z}\right) \tag{4}$$

Upon inspection, this equation has the form $y = mx + b$. Plotting equation (4) as a function of E_j yields a linear Boltzmann plot with slope $m = -\frac{1}{k_B T}$ and intercept $\ln\left(\frac{hcLN}{4\pi Z}\right)$.[65] The Boltzmann plot requires a large range of line intensities corresponding to different upper energy levels E_j from the same species to perform a linear regression but does not require the value of the intercept to extract the temperature. The temperature is not directly calculated in our LIBS bacteria experiments because there are not enough lines in the bacteria spectra that originate from different upper energy levels.

The electron density of the plasma can be determined with the Saha-Boltzmann distribution or Stark broadened lines. The Saha-Boltzmann equation can be used to calculate the value of plasma electron density using the ratios of line intensities given by ionization states for a specific element. The Saha-Boltzmann equation is given as

$$n_e = \frac{2(2\pi m_e k_B T)^{\frac{3}{2}}}{h^3}\left(\frac{I^I_{nm} A_{ji} g^{II}_j \lambda_{nm}}{I^{II}_{ji} A_{nm} g^I_n \lambda_{ji}}\right) e^{-\frac{E_{ion}+E^{II}_j - E^I_n}{k_B T}} \quad (5)$$

where *I* represents the lower ionization state of a line and *II* represents the higher ionization state of the same line. The rest mass of the electron is m_e, E_{ion} is ionization potential of the elemental species, E_j and E_n represent two different upper energy levels j and n, respectively.[65,66] The two different lower levels of the atom are represented by E_i and E_m, respectively.[67] Note that for this method the temperature is required, thus the electron density is not calculated by the Saha-Boltzmann method in this work.

An alternative method for finding the electron density of the plasma is by analyzing Stark broadening. Stark broadening occurs when energy levels in an atom are perturbed by an electric field, caused by the presence of free electrons in a plasma. The effect of perturbation of energy levels is a shifting of the centre wavelength and a broadening of observed emission lines that are several times wider than the original peaks.[6610] Broad lines therefore indicate a large electron density, with wider lines indicating greater electron density.

The full width at half maximum (FWHM) of a Stark-broadened line can be found by using the below equation

$$\Delta\lambda_{FWHM} = \frac{2wn_e}{10^{16}}\left(1 + 1.75A\left(\frac{n_e}{10^{16}}\right)\right)\left(1 - \frac{3}{4}N_D^{-\frac{1}{3}}\right) \quad (6)$$

where $\Delta\lambda_{FWHM}$ is the FWHM. A and w are given in the literature as the ion broadening parameter and electron impact parameter, respectively. N_D refers to the number of particles in the Debye sphere and can be replaced with the constant $N_D = 1.72 \times 10^9\left(\frac{T_e^{3/2}}{n_e^{1/2}}\right)$. T_e is the temperature of the electron.[66] Equation 4 simplifies when the broadening due to ions is small:

$$\Delta\lambda_{\frac{1}{2}} = \frac{2wn_e}{10^{16}} \quad (7)$$

The FWHM can thus be used to directly estimate the electron density if w is known from tabulated values. The lines usually used in LIBS to examine electron density are hydrogen lines, hydrogen-like metal lines, and heavy metal species because they show the most pronounced effects.[66] Though we have calcium present in our spectra, we do not observe any Stark broadening due to the relatively low temperature and electron density of our plasmas at the time we observe them. It is also difficult to observe Stark broadening under low temperature and electron density conditions because the spectral resolution of the width of each line is limited by the spectrometer. Therefore, the electron density is not determined in this work.

3.4 LIBS Experimental Setup

Laser-induced breakdown spectroscopy requires a pulsed laser with sufficiently high energy, beam focusing optics, and a spectrometer to collect light from ablation. The first part of this section will describe each component of the experimental setup as well as its purpose. This section will also illustrate the path that the beam takes to reach the sample. The last part of this section discusses the collection of light by a spectrometer from sample ablation. The theory and structure of the echelle grating is examined, as well as the other components of the echelle spectrometer. The detection of light by an intensified charge coupled device and a brief overview on how these devices work is also discussed. Finally, a figure of the spectrometer output is provided and explained.

3.4.1 Overview of LIBS Apparatus

This work utilizes a 1064 nm Nd:YAG laser (Spectra Physics, LAB-150-10) with a pulse repetition rate of 10 Hz, a pulse duration of 10 ns, and a beam diameter of 9 mm. The beam is first directed into a half-wave plate and a polarizing beam-splitter cube to reduce the pulse energy to 180 mJ/pulse. Excess beam energy is directed into a beam dump. A 3x telescope beam expander then expands the beam diameter to three times its initial diameter. The telescope consists of an antireflection coated plano-convex (f = 18.5 cm, φ = 7.62 cm) and plano-concave (f = -5 cm, φ = 2.54 cm) lens. The beam is then directed into an iris with 9 mm diameter to reduce the beam diameter to its initial size and to keep only the central part of the Gaussian beam. A periscope mirror with a highly reflective di-electric

coating directs the beam downward where it is transmitted through a beam splitter, and into a high power Nd:YAG MicroSpot focusing objective (OFR, LMH-20X-1064) to focus the light onto the sample. The beam splitter is present to allow a CCD camera to view the target positioning and ablation area on a monitor. The final energy of the pulse at the target is 8 mJ. This setup is shown in figure 3.5b.

The sample to be ablated is mounted on a magnetic stage in a Plexiglas chamber that contains the microscope objective. The chamber sits on a translation stage so that it can be moved in the x, y, and z directions. The translation stage allows the sample to be translated through the laser focus for sampling in two dimensions and allows for adjustment of the lens-to-sample distance (LTSD). The LTSD dictates where the sample is relative to the focus of the beam, and therefore the LIBS spectrum is very dependent upon the LTSD and the resulting change in spot size and laser intensity.

Determination of the proper LTSD in the z direction is achieved by a helium-neon (He-Ne) laser directed to the target by aluminium mirrors. The He-Ne laser appears as a bright spot on the monitor and is used to set the LTSD, or the height of the sample within the chamber. All samples in this work were mounted on a steel piece using double sided sticky tape. A schematic of the laser beam and optical system are shown in Figure 3.5.

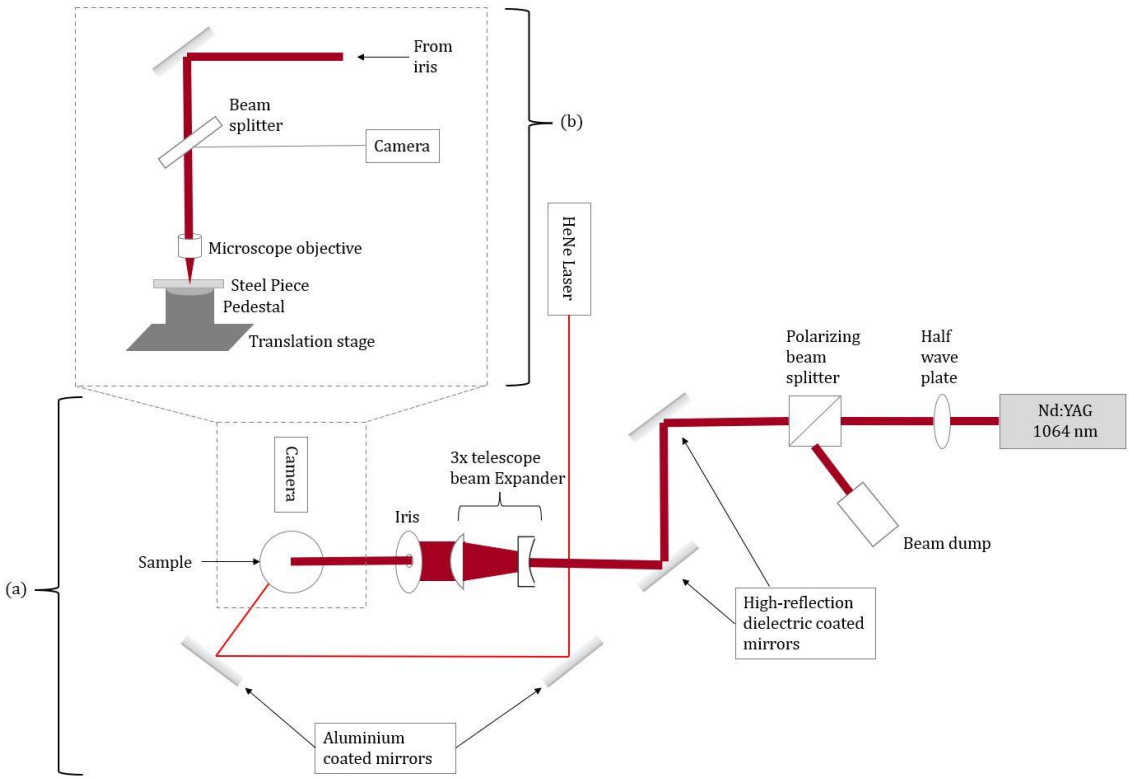

Figure 3.5: Schematic of LIBS experimental setup: (a) top view of entire setup, and (b) side view of laser incident on target.

3.4.2 Detection of Light From Plasma

The light from the plasma is collected by two matching parabolic aluminium mirrors (f = 5.08 cm, φ = 3.81 cm) and directed into a 1 m steel-encased multimodal optical fibre (NA = 0.22, core φ = 600 μm). The mirrors focus the light onto the optical fibre and increase the light that can be inputted into the fibre. The light is then dispersed by an echelle spectrometer (ESA 3000, LLA Instruments, GmbH) and detected by an intensified charge-coupled device (ICCD) camera. The spectrometer was controlled by ESAWIN v3.20 software to change the width and delay times of the ICCD light collection. This software also controlled the firing of the Nd:YAG laser Q-switch to provide nanosecond timing control.

The echelle spectrometer contains an echelle grating which disperses the plasma light spatially by wavelength. Each wavelength is diffracted at a different angle which is described by the equation

$$m\lambda = d(\sin\alpha + \sin\beta) \quad (8)$$

where m is an integer describing the diffraction order, λ is the wavelength of incident light, d is the groove spacing, α is the angle of incidence and β is the angle of diffraction. It can be seen from the above equation that the angular spacing of an incident wavelength decreases as the diffraction order increases. It can also be proven from the above equation that light at different wavelengths and diffraction orders will overlap. A first order line of wavelength 800 nm will overlap with a second order line of 400 nm. More generally, a first order line of wavelength λ will overlap with a second order line of wavelength λ/2. This trend continues for higher orders. To help spread the wavelengths and orders so that they are more easily distinguishable, a glass prism mounted in front of and perpendicular to the echelle grating is used to cross-disperse the light. The echelle grating separates the orders along a vertical axis and the wavelengths along a horizontal axis into a two-dimensional pattern. The grating used in this experiment disperses light into orders $m = 29$ up to $m = 119$ and has a spectral coverage of 200-840 nm. Figure 3.6 shows the schematic of the ESA 3000 spectrometer.[68]

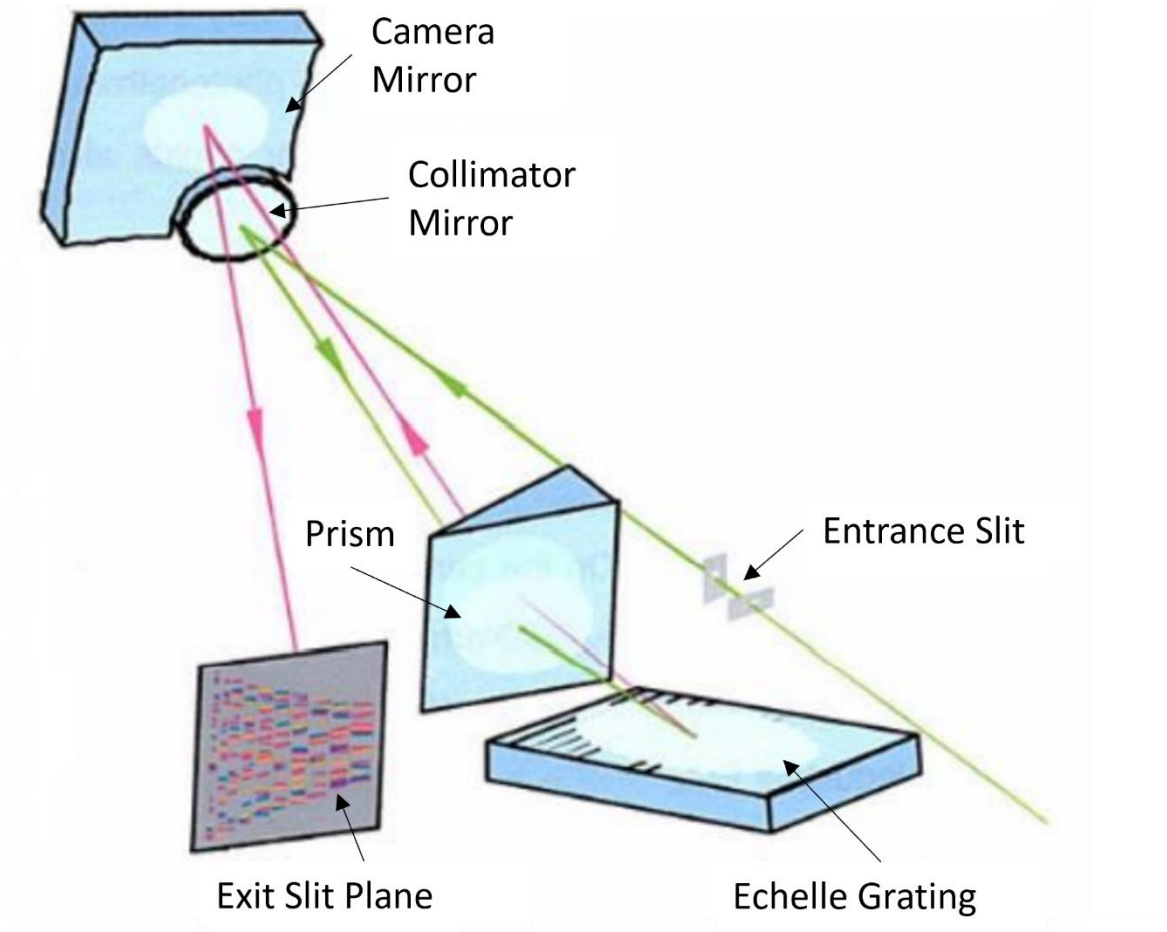

Figure 3.6: Cartoon schematic of the echelle spectrometer components. Adapted from ref [68].

The now two-dimensional diffraction pattern is reflected back through the prism and is detected and recorded on a ICCD Kodak camera. The CCD camera is a chip that is 1 inch by 1 inch (1064 pixels by 1064 pixels, pixel size of 24 μm^2). This two-dimensional diffraction pattern is known as an echellogram and the output of the CCD camera recording an echellogram from a steel LIBS spectrum is shown in Figure 3.7 below. The data shown is a false colour image. The yellow regions on the image show areas where no light was seen by the camera, whereas the dark spots indicate areas that the CCD observed and recorded light. The CCD camera records incoming light on light-sensitive elements, called pixels, that are arranged on a semi-conductor material. Incoming photons produce free electrons and thus electron-hole pairs in each pixel. The number of electron holes is linearly proportional to the number of photons each pixel observed so the charge of each pixel is measured to determine a photon intensity. The CCD camera also contains an image intensifier which is composed of a microchannel plate (MCP) which multiplies the number of photoelectrons

and a phosphor screen to convert electrons back into photons before being detected by the CCD. The image intensifier gain can be adjusted in the ESAWIN software to amplify weak signals. The use of the image intensifier with the CCD camera is what makes this an ICCD.

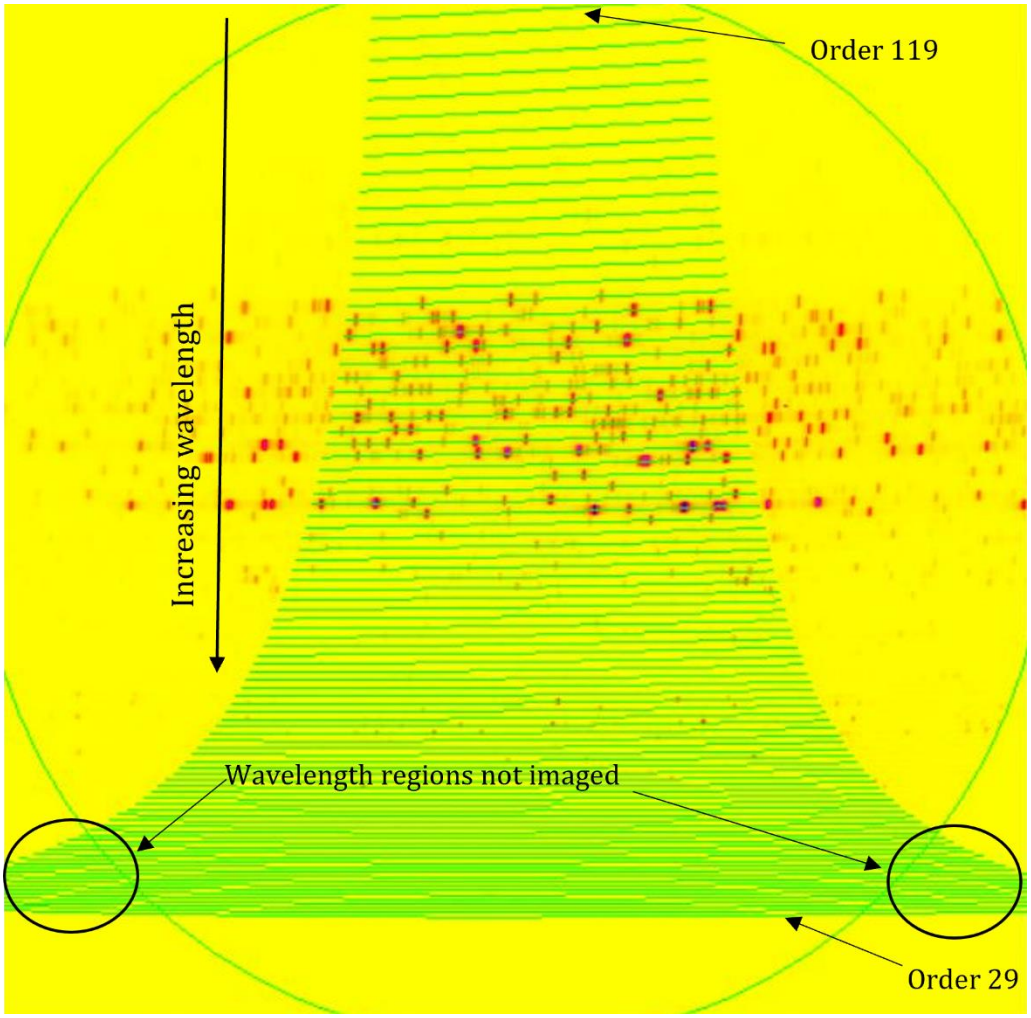

Figure 3.7: Echellogram of steel spectrum. Orders are given by green lines, each order contains a wavelength range.

Diffraction orders in Figure 3.7 are represented by the horizontal green lines. The uppermost green line corresponds to the highest order $m = 119$ and the lowest green line corresponds to the lowest order $m = 29$. The highest order $m = 119$ consists of the shortest wavelengths and the smallest wavelength range, in this case from 201.023 to 202.615 nm. The lowest order spans the longest wavelengths and possess the greatest wavelength range, from 816.875 to 838.393 nm. Therefore, the echelle spectrometer disperses UV wavelengths to the top of the CCD chip in this image and IR wavelengths to the bottom of the CCD chip. The green circle in Figure 3.7 shows the area of the CCD that is

illuminated and thus the range of wavelengths which can be measured. No light is collected outside of that circular area. Gaps in the measured spectra correspond to the regions where light is dispersed that are outside of the circle. The spectrometer was optimized to provide resolution and continuity for elements that have strong emission lines in and near the UV region, which are regions of interest to this work and for the majority of LIBS studies.

3.5 Bacteria Physiology and Sample Preparation

This section will give an overview of bacterial cell biology, focusing specifically on membrane chemistry and pathology. This is not meant to be an in-depth review of bacterial biology, but rather it is meant to provide the LIBS practitioner who may be unfamiliar with the fundamental knowledge of bacterial microbiology with an introduction to the topic. This chapter contextualizes why we have chosen to study specific species and gives a background on how these species' biology and how they impact human health. As well, the methods used for growing bacteria, sample preparation, and deposition of samples onto testing services is described here. More information on microbiological pathogenesis, diagnosis, and treatment can be found in the textbooks *Mechanisms of Microbial Disease* and *Bacterial Pathogenesis: A Molecular Approach*. Information on bacterial biology, culturing methods, and staining methods can be found in the undergraduate textbook *Microbiology: Canadian Edition*.

3.5.1 Bacteria Physiology

Bacteria are ubiquitous single-celled prokaryotic organisms that can cause disease. Bacteria lack membrane-bound organelles and a nucleus; however they still have DNA, ribosomes, and a plasma membrane. Most bacteria also contain an outer cell wall. Bacteria generally have three different shapes: cocci (round), bacilli (rod), or spirochete (spiral).

Bacteria can be grouped into two main categories based on membrane physiology, determined by the Gram stain. The two categories are Gram-positive, which stain purple, and Gram negative, which stain pink. Gram staining generally involves two basic steps. The first step is staining cells with a crystal violet dye, which initially all cells will take up. A solvent is then applied which dissolves the lipid layer of Gram-negative bacteria, causing the cells to release the crystal violet stain. The second step is a red counterstain, typically a

weakly water-soluble safranin. The red stain does not disrupt the Gram-positive bacteria but does stain the Gram-negative bacteria, resulting in the characteristic pink colour.[69]

A third lesser-known type of bacteria is the acid-fast bacteria; acid-fastness refers to the resistance of decolourization during an acid wash. The cell walls of acid-fast bacteria are different from the prior because they contain hydrocarbon chains that are woven throughout the cell wall.

Bacteria cell walls (and the rest of the bacterial cell) also contain varying concentrations of calcium, magnesium, sodium, and carbon. The ions are present due to ion exchange at the surface and the neutrals are present in the peptidoglycan.[71,70] These are the elements that we are primarily interested in for the detection and discrimination of bacteria. Table 3.1 shows the elements that we detect in the bacterial cell and the wavelengths of spectral lines for each element. Table 3.1 is located later in section 3.6.1 where it is necessary for a discussion of the variables used by the chemometric algorithms.

3.5.2 Bacterial Species Tested with LIBS

Five species of bacteria were chosen so that some representation was achieved across genus, shape, and Gram type. The species tested include *Escherichia coli, Pseudomonas aeruginosa, Staphylococcus epidermis, Mycobacterium smegmatis,* and *Enterobacter cloacae* which span acid-fast, Gram positive, and Gram negative, and include cocci and bacilli shapes.

E. coli is a motile Gram-negative rod. Non-pathogenic strains are commonly found in the intestines of warm-blooded animals and humans. These strains are beneficial because they produce vitamins, and because their large numbers help prevent other organisms from proliferating enough to cause infection. Harmful pathogenic strains can lead to health issues and are generally responsible for urinary tract infections (UTI), food borne infections, diarrhea, kidney failure, septicemia, pneumonia, and meningitis.[71,72] *E. coli* is easy and quick to grow, making non-pathogenic forms an ideal organism to grow and study using LIBS.

S. aureus is a non-motile Gram-positive coccus. Non-pathogenic strains of *S. aureus* are commonly found on epithelial surfaces, including the skin, nostrils, intestines, and upper respiratory tract. Pathogenic strains can enter the bloodstream through breaches in the skin and can adhere to plastic surfaces which can ultimately lead to bloodstream infections associated with catheters. Other pathogenic *S. aureus* strains can cause respiratory infections, pneumonia, and skin infections.[73,74] Pathogenic strains *S. epidermis* and *S. aureus* are also the leading cause of septicemia.[74] Many strains of *S. aureus* have become resistant to antibiotics, including methicillin-resistant *S. aureus* (MRSA).[73] There are no vaccines for MRSA and there are limited treatment options. A quick and targeted treatment is necessary, making a fast and accurate diagnosis imperative.

P. aeruginosa is a motile Gram-negative rod that is ubiquitous in the environment. It exists in water and other wet surfaces, soil, plants, and artificial environments such as hospitals.[73] Pathogenic *P. aeruginosa* is responsible for many nosocomial infections, as it can invade the body through breaches in the defense system, such as wounds and burns, making it an opportunistic bacterium. It is also particularly dangerous to those who have a weakened immune system, existing conditions, and diseases. It is responsible for eye infections arising from scratches on the cornea from contact lenses, burn infections on burn victims, and lung infections in patients with cystic fibrosis.[74] *P. aeruginosa* is also resistant to many antibiotics.[75]

M. smegmatis is an acid-fast, rod-shaped, bacteria and is non-pathogenic. It is found in water, soil, and plants and is an opportunistic organism, mainly infecting the immunocompromised. *M. smegmatis* can be cultured and grown quickly and requires only biosafety 1 level facilities, making it an ideal organism to work with in the laboratory. It is also ideal to use as a substitute for *M. tuberculosis* which is a much slower growing organism in culture; visible colonies often take a few days to form. The application of LIBS to *M. tuberculosis* and other *Mycobacterium* can thus be very useful in identifying and diagnosing this organism quickly. *Mycobacterium* can only be stained using acid-fast dyes because it has a waxy outer envelope which makes it resistant to the Gram stain.[74]

E. cloacae is a motile Gram-negative rod. Non-pathogenic *E. cloacae* exists in the human gut and on the skin, as well as on fruits and vegetables. Pathogenic strains can cause respiratory infections, urinary tract infections, and sepsis in immunocompromised patients.[73,74] *E. cloacae* is a common cause of nosocomial infections as it can enter the body through any skin lesions and is transported easily through surgical equipment. The nosocomial infections that result from this are typically bacteremia, endocarditis, septic arthritis, osteomyelitis, and ophthalmic infections.[74] *E. cloacae* also exhibit resistance to common antibiotics such as ampicillin and penicillin making them difficult to treat.[75] *E. cloacae* are difficult to distinguish from other Gram-negative species and thus difficult to diagnose. Testing with LIBS was done on *E. cloacae* to test if the technique could accurately distinguish between *E. cloacae* and other Gram-negative species.

3.5.3 Bacteria Sample Preparation

The bacteria samples prepared and studied in this work were initially provided by Ms. Ingrid Churchill of the department of integrative biology at the University of Windsor. From the initial stock solutions, more bacteria were cultured by our lab group on tryptic soy agar (TSA) nutrient media plates. The TSA plates are a general-purpose culture medium that consists of soybean meal, casein, NaCl, K_2HPO4, and dextrose. This media provides all the necessary nutrients needed to grow bacteria. To prepare the plates, 1 g of TSA powder is completely dissolved in an Erlenmeyer flask containing 25 mL of water. The flask opening is then covered by aluminium foil and autoclaved at a temperature of 121°C to ensure the mixture is sterilized. After removing the flask from the autoclave, it is left to cool for 30 minutes before pouring into a petri dish. The solution is poured slowly to avoid any bubbles from forming and sets for 2 hours. All TSA plates were prepared by our lab group.

Bacteria from stock solutions can be cultured on the surface of TSA nutrient media. Bacteria from the stock solutions were cultured by pipetting 100 µL of a specific bacterial species onto the surface and tilting the plate to allow the solution to spread across most of the plate. This step was repeated for each species of bacteria. Plates were incubated at 37°C for 24-72 hours. After incubation, a repeatable quantity of bacteria was harvested from each plate using a wooden toothpick and placed in labelled 10 mL centrifuge tubes

containing 5 mL of DI water. The initial concentration of each suspension made was determined by optical densitometry (absorbance). Absorbance measurements were also made of fractional dilutions of the stock solution. Fractional dilutions were also made from the stock solution by performing serial dilutions. The dilutions made were 1/5, 1/10, 1/50, 1/100, and 1/500. Absorbance measurements were performed on these to confirm concentrations.

Bacterial sample preparation was designed to mimic clinical collection of bacteria from surfaces in the body using swabs. First, 100 µL of bacterial solution is pipetted onto a small steel plate, shown in Figure 3.8a. The steel plate is heated on a hot plate at 200°C for two minutes and twenty seconds to remove excess moisture. Once the steel plate has cooled, 10 µL of deionized water is pipetted onto the tip of a sterile swab. The dampened swab is then used to swab the bacterial film off the steel plate, shown in Figure 3.8b. The swab is placed in a centrifuge tube containing 1 mL of water and vortexed for 15 seconds to shake all the bacteria off the swab and into the water, shown in Figure 3.8c. The swab is removed from the centrifuge tube and discarded. Alternately, samples of water containing bacteria were pipetted directly into the centrifuge insert rather than being pipetted onto the steel plate. This circumvented the swabbing step and was developed as a deposition method to simplify the testing of fluid clinical samples such as blood and urine, discussed in Chapter 6 of this thesis.

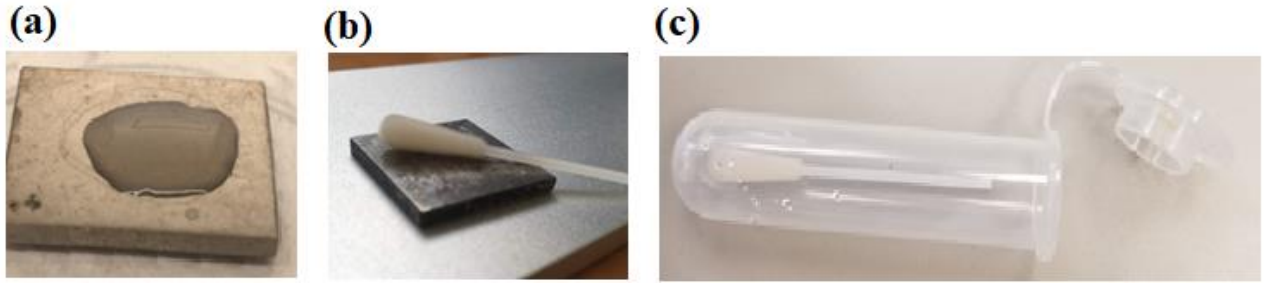

Figure 3.8: Process of bacterial deposition (a) sample is pipetted onto metal plate and excess moisture is evaporated off (b) sample is swabbed off plate using a damp swab, and (c) swab is vortexed in 1 mL of water for 15 seconds to shake off cells.

Solutions of bacteria that have been swabbed are deposited onto a 0.45 µm pore size nitrocellulose membrane filters that are 9.5 mm in diameter (HAWP04700, Millipore Inc.). The filter is held in place by a custom-fabricated centrifuge insert, and a custom-built

aluminium metal cone is placed on top of the filter to concentrate the bacteria to a spot of 1 mm in diameter. The cone and centrifuge insert achieve rapid concentration of the bacteria during centrifugation and are made with easily obtainable materials allowing them to be readily implemented in the clinic.

The drawing of the centrifuge pieces are shown in Figure 3.9. The insert was designed and 3D printed by a previous student in our lab group.

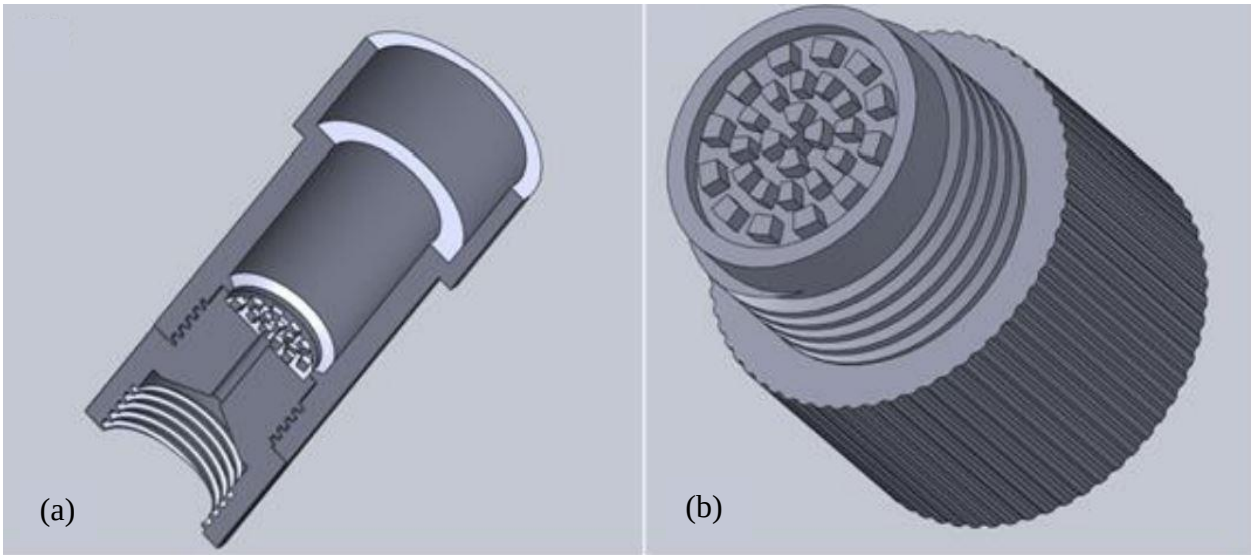

Figure 3.9: Centrifuge insert design in cross section. Filter paper is placed on base of insert in (b). Body of insert in (a) is screwed onto base of insert. Figure adapted from ref [75].

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. Both components of the centrifuge tube are made of plastic. The diameter of the outside of the centrifuge tube is 14 mm. The bottom piece is 9.5 mm in diameter and holds the filter. There is a small hole in the center of the bottom piece which allows water and other fluids to exit the centrifuge tube; this is shown in Figure 3.9b. The longer top piece screws on to the piece holding the filter and contains a seal to hold the filter in place. The wider portion of the top piece is 17 mm in diameter and allows the centrifuge insert to sit on the lip of the centrifuge tube. The length of the assembled insert is 40 mm, and it can hold a total of 1.5 mL of fluid.[75]

The metal cone is shown in Figure 3.10a. The cone was designed such that it would sit inside the centrifuge tube and press into the filter, shown in Figure 3.10b. The diameter of the hole at the base of the cone is 1 mm. During centrifugation, the liquid suspension containing bacteria passes through the hole at the base of the cone and is deposited onto the filter. The purpose of the metal cone is to rapidly concentrate bacterial cells on the filter to improve the limit of detection. With the addition of the metal cone, the limit of detection (LOD) was calculated to be 10,865 cells per laser shot which is an improvement by a factor of 10 compared to the LOD using only the centrifuge insert which was previously approximately 90,000 cells per laser ablation event.[76] The LOD was calculated by constructing a calibration curve from several concentrations of bacterial suspensions. A colour map of the concentration of cells onto a filter using the centrifuge insert and metal cone is shown in Figure 3.11.

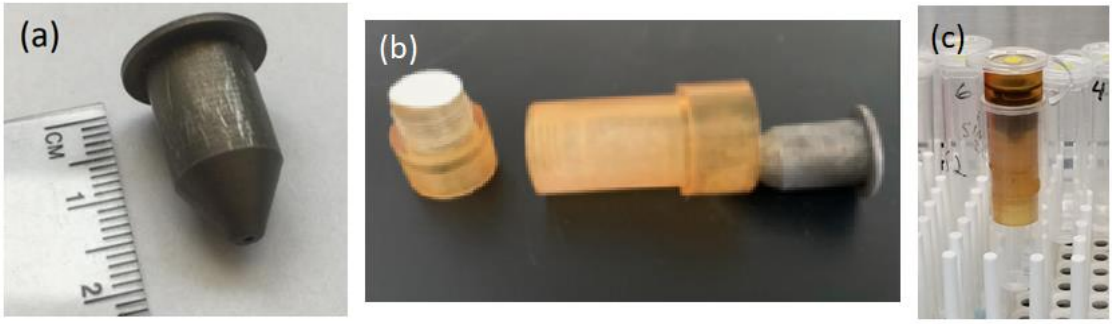

Figure 3.10: (a) Metal cone for rapid concentration of bacteria, shown next to ruler to give approximate size. (b) Centrifuge insert shown before assembly. A filter is placed on the bottom piece of the centrifuge insert before it is assembled. (c) The assembled centrifuge insert with cone is placed in a centrifuge tube.

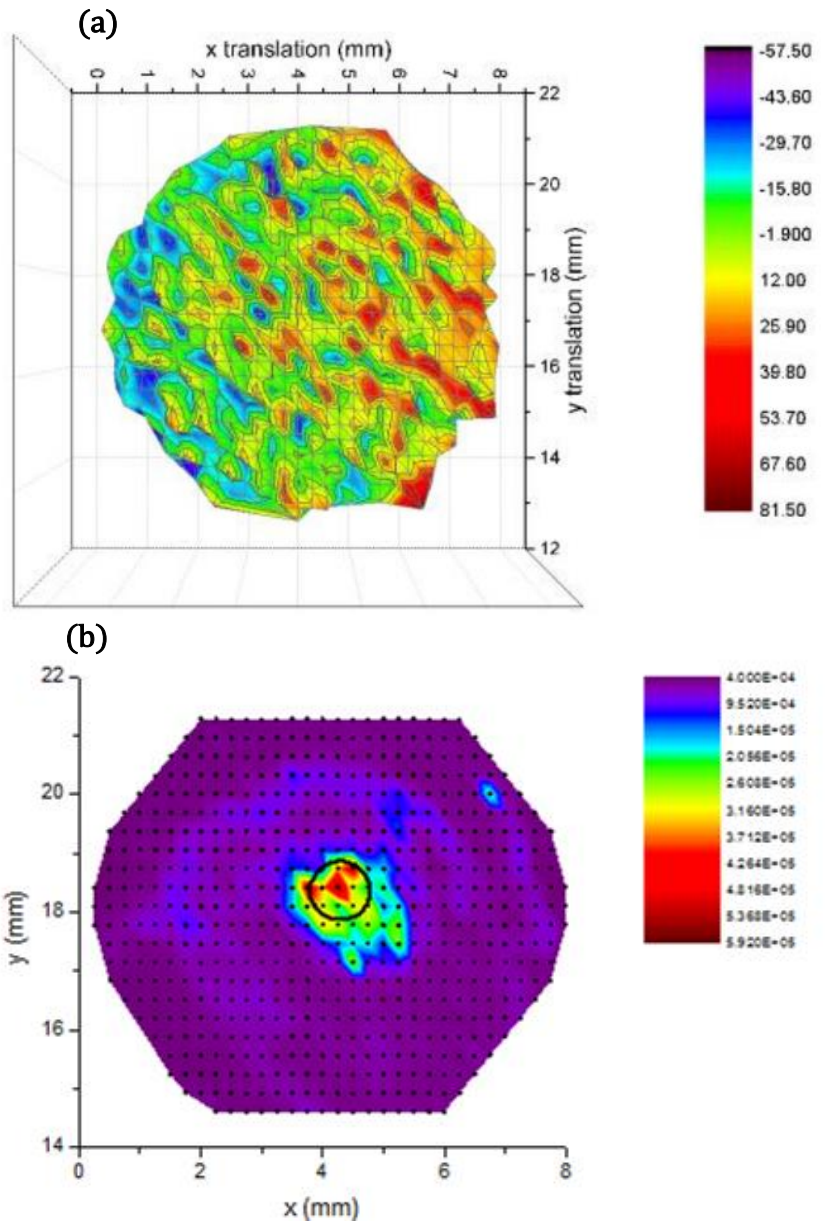

Figure 3.11: (a) Colour map of the concentration of cell deposition on filter after centrifugation through the centrifuge insert, and (b) colour map of the concentration of cell deposition on filter after centrifugation through the centrifuge insert and metal cone. Figures adapted from ref [75,76].

The assembled centrifuge insert is placed in a clean centrifuge tube and the metal cone is placed into the top of the assembled centrifuge tube, as shown in Figure 3.8c. 1 mL of the water from the centrifuge tube containing the swab is pipetted into the cone. Alternately, samples of water containing bacteria were pipetted directly into the centrifuge insert rather than being pipetted onto the steel plate. This avoided the swabbing step completely and was done to simulate fluid clinical specimens, which is discussed further in Chapter 6

of this thesis. The centrifuge tube is closed and placed in a centrifuge (PowerSpin BX, unico) with a counterweight. Samples are centrifuged at 5000 rpm, 2500g's of force for 5 minutes to pull the liquid suspension through the centrifuge insert. After centrifugation, the insert is taken out of the centrifuge tube and disassembled. The filter is removed from the bottom piece of the insert and placed onto a stainless-steel plate that has double sided sticky tape.

Samples were ablated in an argon gas environment; argon flow was set to 20 standard cubic feet per hour (SCFH). The laser used for ablation is the aforementioned Nd:YAG 1064 nm (Quanta Ray LAB-150-10, Spectra Physics) with 10 Hz repetition rate and 10 ns pulse. The gate width (τ_w) is set to 20 μs, and the gate delay (τ_d) is set to 2 μs for all bacterial experiments. The spacing between each laser shot is minimized to achieve the most amount of data per filter to increase the total amount of spectra in our library. The spacing between our laser shots is 0.15 mm. An example of a typical bacterial spectrum observed using such experimental parameters is shown in Figure 3.12.

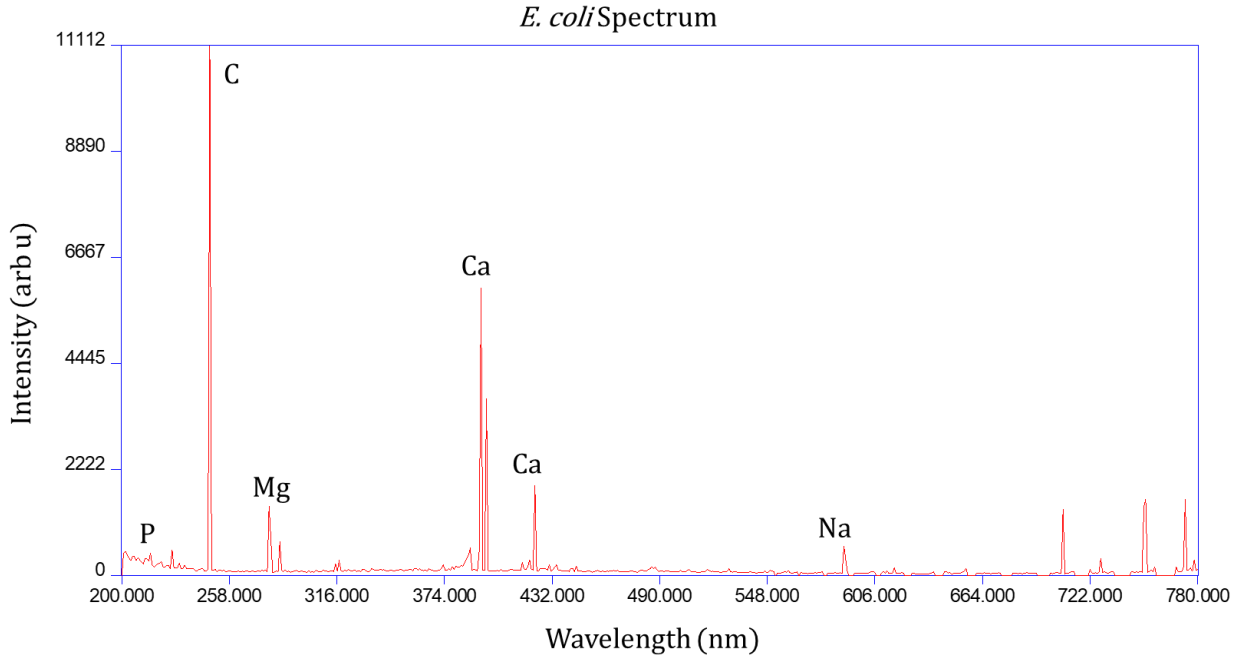

Figure 3.12: A typical spectrum of E. coli bacteria. All lines important for bacterial identification are labelled.

3.6 Overview of Chemometric Analysis

Chemometric algorithms reduce the amount of data needed for an accurate discrimination. Initial work on chemometric analysis in conjunction with LIBS began in 2003 and is discussed in greater detail in Chapter 2 section 2.2. Numerous chemometric methods have been explored over the years, beginning with linear correlation techniques to compare emission intensities. More advanced techniques have been developed and used, including linear regression models, chemometric algorithms, and neural networks.[77] This section will focus on the basic principles of chemometric techniques, specifically on DFA, PLSDA, and artificial neural networks (ANN). This section will also discuss the data models and methods used by our group when using chemometric algorithms. For more information on the chemometric algorithms discussed, the reader is directed to the textbook *Applied Chemometrics for Scientists*.

3.6.1 Data Models Used in Chemometric Algorithms

Initial chemometric analyses done by our group used the measured intensities of 13 of the most intense emission lines normalized to the sum of all the measured intensities as independent variables. These 13 lines encompassed 5 elements of interest for bacterial targets: calcium, magnesium, sodium, phosphorus, and carbon. This was the simplest model we have used to date and is called the "lines" model. The work of Gottfried et al. showed that the ratios of line intensities and summed line intensities, as opposed to individual lines, showed improved discrimination ability.[78] Based on this work, the original lines model was refined to form ratio model 1 (RM1). RM1 used the same 13 lines in multiple complex ratios to form 24 independent variables that included the 5 elements of interest: calcium, magnesium, sodium, phosphorus, and carbon. Later, RM1 was adapted to form RM2, which used the same 13 lines as previous models with an additional 67 complex ratios to form 80 independent variables.[79] More recently, a third ratio model was developed with 19 lines to include all the lines seen in bacterial spectra, as opposed to the largest 13 lines we see. This was called ratio model 3 (RM3) and it contains 19 lines and 145 simple ratios to make 164 independent variables.[78] A table with a complete list of the lines used in RM3 and their respective variable names are given in Table 3.1. The details of

the development of these models and the chemometric results can be found in references.[78,79,80]

Table 3.1: Emission lines used for discrimination and the respective variable associated with each. These emission lines and variable designations are used in RM3.

Emission Line	Variable Designation
C 247.856	c
P 213.618	p1
P 214.914	p2
P 253.398	p3
P 253.56	p4
P 255.326	p5
P 255.491	p6
Mg 279.079	mgii1
Mg 279.553	mgii2
Mg 279. 806	mgii3
Mg 280.271	mgii4
Mg 277.983	mgi1
Mg 285.213	mgi2
Ca 317.933	caii1
Ca 393.366	caii2
Ca 396.847	caii3
Ca 422.673	cai1
Na 588.995	na1
Na 589.593	na2

Some of the work in this thesis is done using RM3. More recent work has been done using a model I constructed based on the previous RM3, this model is called ratio model 2.5 (RM2.5). It consists of 15 emission lines and 92 simple ratios, for a total of 107 independent variables. Less lines were chosen to use in this model to avoid overfitting of the algorithms we use, and to eliminate emission lines that gave little to no information. Many of the lines that were removed from RM3 to create RM2.5 were phosphorus and magnesium lines that had consistently low or zero intensity. A table with a complete list of all lines used in RM2.5 is shown in Table 3.2. A complete list of the RM3 and RM2.5 ratios that are used in this research can be found in Table A.2 in Appendix A.

My work has been conducted using three algorithms: DFA, PLSDA, and ANN. These algorithms used precompiled data that has been collected over time for detection and classification of bacterial species. We call the precompiled data the spectral library. The library is organized in a ratio model Excel sheet, shown in Figure 3.13. The data is organized this way because these chemometric algorithms require the data to be passed in the form of two matrices. One of these matrices is the x-block, which is row vectors representing the data model. The other is the y-block, which is a single column vector holding information on the identity of the sample (the class). The file names for each spectrum are on the far left, they are the label for the data point. The assigned class of the data point is given in the class column. The emission lines used in the analysis are in the top row, in this case the 15 elemental emission lines for RM2.5 are shown. The first of 92 simple ratios are shown outlined in purple to the far right.

Table 3.2: Emission lines and the respective variable designation associated with each. These emission lines and variable designations are used in RM2.5.

Emission Line	Variable Designation
C 247.856	c
P 213.618	p1
P 214.914	p2
P 253.56	p4
Mg 279.079	mgii1
Mg 279.553	mgii2
Mg 279. 806	mgii3
Mg 280.271	mgii4
Mg 277.983	mgi1
Ca 317.933	caii1
Ca 393.366	caii2
Ca 396.847	caii3
Ca 422.673	cai1
Na 588.995	na1
Na 589.593	na2

File Name (Raw Intensity)	Class	C247 (c)	P213 (p1)	P214 (p2)	P253.5 (p4)	Mg279 (mgii1)	Mg279.5 (mgii2)	Mg279.8 (mgii3)	Mg280 (mgii4)	Mg285 (mgi)	Ca317 (caii2)	Ca393 (caii3)	Ca396 (caii4)	Ca422 (cai1)	Na588 (na1)	Na589 (na2)	p1/c
051519_25922ecoliFifth_001	1	0.292	0.0063	0.0032	0.0018	0.0059	0.165928	0.009931	0.087474	0.0108	0.018	0.251	0.1341	0.0138	0.0124	0.0056	0.021654
051519_25922ecoliFifth_002	1	0.336	0.0101	0.0026	0.002	0.0048	0.157413	0.008654	0.080682	0.0109	0.018	0.233	0.1281	0.0097	0.0097	0.0069	0.030127
051519_25922ecoliFifth_003	1	0.281	0.0086	0.0051	0.0015	0.0054	0.176191	0.008571	0.098573	0.0146	0.018	0.226	0.1191	0.0143	0.0135	0.01	0.030677
051519_25922ecoliFifth_004	1	0.145	0.0059	0.0026	0.0012	0.0077	0.213256	0.014587	0.126477	0.009	0.018	0.325	0.1356	0.0048	0.0053	0.0037	0.04072
051519_25922ecoliFifth_005	1	0.394	0.0087	0.0049	0.002	0.0032	0.135545	0.006507	0.074171	0.0062	0.019	0.212	0.1112	0.0068	0.0102	0.0062	0.022186
051519_25922ecoliFifth_006	1	0.246	0.0055	0.0042	0.0018	0.0058	0.189961	0.011535	0.107098	0.01	0.021	0.248	0.1277	0.0072	0.0083	0.0055	0.022365
051519_25922ecoliFifth_007	1	0.29	0.0102	0.0033	0.0013	0.0041	0.156109	0.008594	0.088069	0.0126	0.017	0.244	0.1339	0.0112	0.0125	0.0076	0.035242
051519_25922ecoliFifth_008	1	0.282	0.011	0.006	0.0025	0.0053	0.160099	0.01015	0.088055	0.0146	0.016	0.243	0.1336	0.0116	0.014	0.0074	0.039049
051519_25922ecoliFifth_009	1	0.272	0.005	0.0043	0.0018	0.0042	0.1636	0.00704	0.092464	0.01	0.02	0.26	0.1368	0.0083	0.0107	0.0065	0.018385
051519_25922ecoliFifth_010	1	0.264	0.0071	0.0042	0.0018	0.0055	0.1911	0.011412	0.107	0.0092	0.018	0.248	0.1285	0.0078	0.0095	0.0054	0.026761
051519_25922ecoliFifth_011	1	0.268	0.0066	0.0062	0.001	0.0043	0.169791	0.00956	0.09564	0.0161	0.018	0.25	0.1325	0.0139	0.0166	0.0093	0.02452
051519_25922ecoliFifth_012	1	0.319	0.0066	0.0057	0.0022	0.0059	0.148948	0.006971	0.085715	0.0161	0.013	0.224	0.1225	0.0143	0.0194	0.0099	0.020592

Figure 3.13: Example of RM2.5 Excel sheet used for chemometric analysis. The data labels are outlined in green, the designated class is outlined in blue (y-block vector), the 15 emission lines used are outlined in red, and the first row of the 92 simple ratios is outlined in purple. The 15 emission lines in red and the ratios in purple represent the x-block.

The data library in the Excel sheet is used to form a model in either PLSDA or DFA. Internal validation models are formed when the class for all data points entered is known. The model can be tested using external validation, which is done by removing subgroups of the library data and feeding them to the model with no known class. Subgroups of data can include whole filters, dilutions, species, and strains. The performance of the model with this inputted data can then be recorded and is an indication of how well the model can classify unknown spectra. After classification with internal and external validation, the sensitivity and specificity of the model is recorded, along with the sensitivity of the externally validated group. These results are often summarized in truth tables. Truth tables contain percentage values of true positive, true negative, false positive, and false negative results. To optimize a clinical test, the true positives and true negatives must be maximized while the false positives and false negatives are minimized. False positives can cause undue stress for a patient and increase the burden on the health care system, while false negatives cause a potentially dangerous condition to be overlooked. To optimize LIBS as a rapid point-of-care diagnostic tool, these parameters need to be met.

3.6.2 Discriminant Function Analysis (DFA)

Discriminant function analysis is a statistical procedure that determines the differences between groups and finds the fewest number of dimensions needed to determine the difference accurately and reliably between groups. In essence, DFA focuses on maximizing the separability between known classes while minimizing scatter within each class. DFA can be performed on N classes, but here the simplest case of a 2-class test will be considered. DFA uses information from both classes to project data onto a new axis that

maximizes the separation of the classes; in the case of a 2-class system the new axis would be a line. More mathematically, based on the initial inputted data, DFA gives a discriminant score based on the closeness of an unknown row vector to class A or B. The discriminant score is given by equation 9:

$$D_{AB} = (\bar{X}_A - \bar{X}_B) \cdot S^{-1} \cdot X^T - \frac{1}{2}(\bar{X}_A - \bar{X}_B) \cdot S^{-1} \cdot (\bar{X}_A + \bar{X}_B) \quad (9)$$

where $\bar{X}_A$ and $\bar{X}_B$ are the average vectors for groups A and B, X is the unknown group being classified, and S represents the pooled variance-covariance matrix for the groups.[80] The first term in the equation determines a score value, the second term of the equation determines if the score value is positive or negative. In the case of a 2-class test, a positive value would indicate that X is closer to class A, while a negative score for X puts it closer to class B. The variance-covariance matrix S handles outlier data by scaling the score based on the variance within each group. This helps to consider points in group A or B that are farther from the majority of the data to classify correctly.[80]

In expanding this model to N groups, only N-1 scores are needed for classification of an unknown group. This results in an N-1 dimensional space where the unknown group sits in between other known groups. The minimum distance found between the unknown group and another class indicates which group it most likely belongs to.[80] DFA will always classify the unknown group with a known group; for example, consider the case of deionized water as an unknown group inputted to a 3-class test between *P. aeruginosa, E. coli,* and *S. aureus.* DFA will not classify deionized water as its own group and instead match it with the group that has the least variation with the unknown data.

In DFA, discriminant functions are given scores in decreasing order to their relevance to the discrimination. Discriminant function 1 (DF1) will therefore include the largest part of the variance between groups, discriminant function 2 (DF2) will contain less variance, and so on to discriminant function N (DFN) which will contain the least amount of variance. The more discriminant functions used corresponds to finer detail, but often the most information can be elucidated from the first few discriminant functions. In Figure 3.14, an

example discriminant function plot is shown for a 5-class bacteria test. In all the analyses described in this thesis, IBM SPSS Statistics v.21 was used to perform DFA.

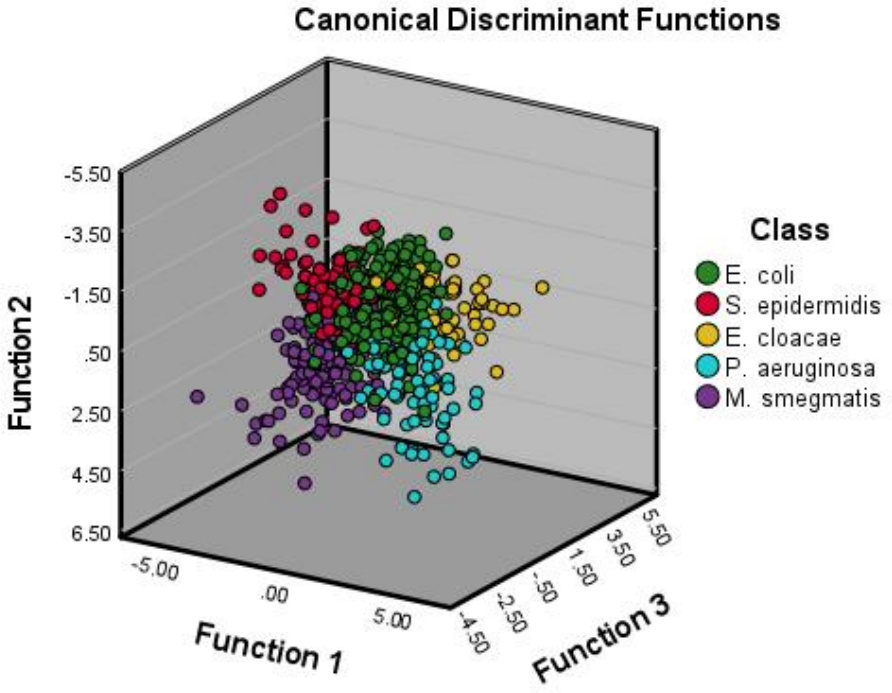

Figure 3.14: Example output of discriminant function analysis using IBM SPSS Statistics. This model consists of 5 bacterial species.

To conduct DFA, the library must contain at least the same number of independent variables as the model to avoid overfitting. For optimal classification, there should be more data points than variables, best practices call for 10 times more data than independent variables. For the RM3, there should be at least 164 data points, and 92 data points for the RM2.5. DFA also assumes a Gaussian distribution of data scattered about an "average" spectrum, so for best possible results the data inputted should be Gaussian. Our bacteria data is not Gaussian in nature. There is high variability in the data that exists more as an equal spread of intensities than a Gaussian distribution. Figures of histograms that show the total measured intensity (the sum of all the measured line intensities) of some filters of bacteria are included for reference in Figure 3.15. These histograms show that our data is not typically Gaussian and doesn't have a consistent statistical distribution.

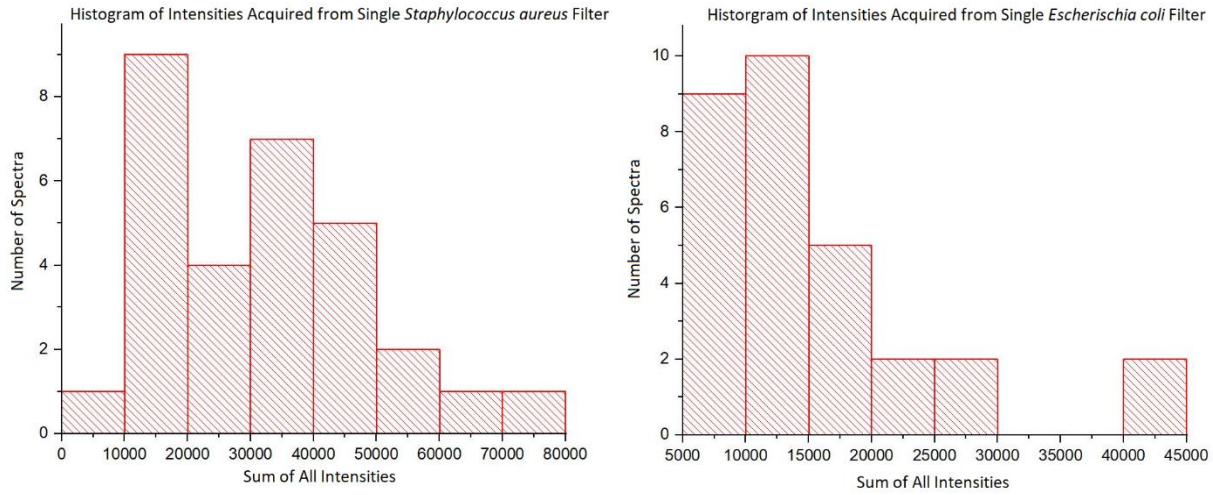

Figure 3.15: Histograms of data acquired from filters of S. aureus *(left) and* E. coli *(right). The* S. aureus *distribution resembles a bimodal distribution. The* E. coli *distribution resembles a left shifted Gaussian distribution with a discontinuity at higher intensities.*

3.6.3 Partial Least-Squares Discriminant Analysis (PLSDA)

PLSDA is a multivariate variable reduction technique that finds the maximum variance between groups and can be used for classification and analysis of spectral data. PLSDA cosntructs latent variables (LV's) out of the independent variables given by the ratio models. The LV's can then be used as predictor variables to construct a calibration curve based on the spectral data or to perform discrimination. Using LV's, PLSDA can calculate a predictor score for a member of a data set; the method of class prediction by PLSDA is different from DFA. DFA compares predictor scores to multiple classes simultaneously, while PLSDA compares predictor scores once per class. The predictor score is a +1, -1, or 0, and represents a 'yes' or 'no' answer for the data point in question.[80] A Bayesian line representing a threshold between the two classes is generated using Bayesian statistics. If an unknown sample is entered into the algorithm, it is tested once against each class present in the model, given a predictor value, and will either fall above or below the class threshold. Above the Bayesian line indicates a positive result, below the Bayesion line indicates a negative result. It follows then that a feature of PLSDA that is not present in DFA is the potential for a null result. PLSDA can therefore indicate that the unknown data is a species the algorithm is not familiar with, as opposed to DFA which forces the unknown data to commit to a class. This would be useful in the event that a new antibiotic resistant strain had formed, or if a sample was contaminated with another species of bacteria.

Shown in Figure 3.16 is an example of a two class discrimination between *E. coli* and ultrapure water. The PLSDA program used for this work is PLS_toolbox v.8.7.1 combined with Matlab 2016b v.9.1 (Eigenvector research, Inc.).

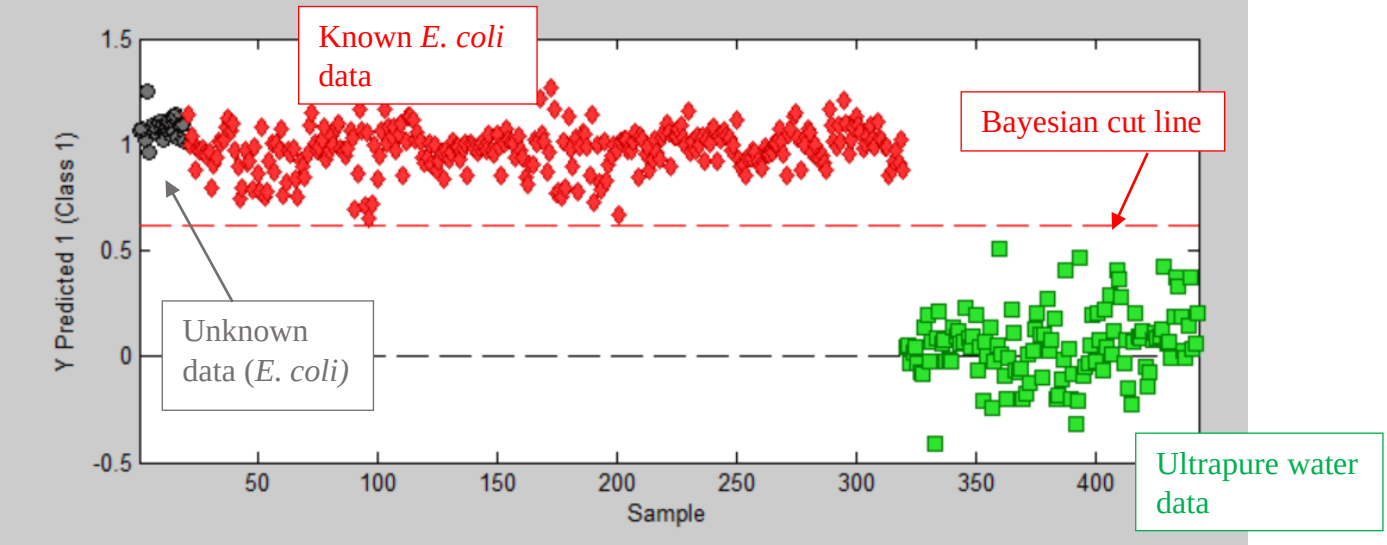

Figure 3.16: PLSDA 2-class discrimination plot between E. coli *and ultrapure water. A data set of* E. coli *has been removed from the model and entered as unclassified data to test the robustness of the model.*

PLSDA, as with DFA, requires attention to avoid overfitting. This is done in the same way as DFA, by ensuring that the number of independent variables is equal to or does not exceed the number of inputted data points. Overfitting can lead to higher rates of false positives. Misclassified spectra can potentially be avoided by adjusting the number of LV's for classification. In the PLSDA program used for this work, the LV's can be adjusted manually or can be assigned automatically by the program. This however does not always work, and the best course of action is acquiring more data for an optimal classification result. Throughout this thesis, a mixture of automatic and manual assignments of LV's were used.

3.6.4 Artificial Neural Networks

Artificial Neural Networks (ANN) are based on the structure of the brain. They create a non-linear model to establish a relationship or pattern between an input and an output. Relationships are created through supervised learning, which requires training data that has a known output value.[81]

The structure of ANN typically consists of an input layer with input nodes, an output layer with output nodes, and one or more hidden layers with hidden nodes. The input layer takes information given to the system and the number of input nodes corresponds to the number of inputs. The hidden layers are the layers between the input and output nodes and this is where most of the computation takes place in the algorithm. The hidden layer takes values from the input nodes and uses these values to produce another value that's passed on to the next layer. The output layer is the outputted values of the algorithm. The output nodes we use are the bacterial classes. Each layer is connected to the next or previous layers by weights and nodes from the same layer cannot be connected. The weights that connect each node represent the strength or importance of a specific input.[82]

ANN algorithms have two primary structures, feed forward and feed back. Only feed forward algorithms were used in this work, therefore only feed forward algorithms will be discussed. The feed forward method constrains data to only move in the forward direction through the algorithm, from the input nodes, to hidden layers, to output nodes. Feed forward algorithms are primarily used for classification and pattern recognition purposes.[81] A schematic diagram of what a typical feed forward algorithm looks like is shown in Figure 3.17. The way this algorithm works is information is received at the input nodes in the form of a vector. Then each input is multiplied by a corresponding weight. Each product of this operation is added up to give a weighted sum. The weighted sum passes through an activation function which ensures that the output is non-linear. This process is how information passes from one layer to the next. During the first iteration, random values are assigned to each weight and the values of these weights are modified for each iteration to produce a predicted output close to the known output. The algorithm will stop modifying the weights when the error between the input and output reaches a minimum.

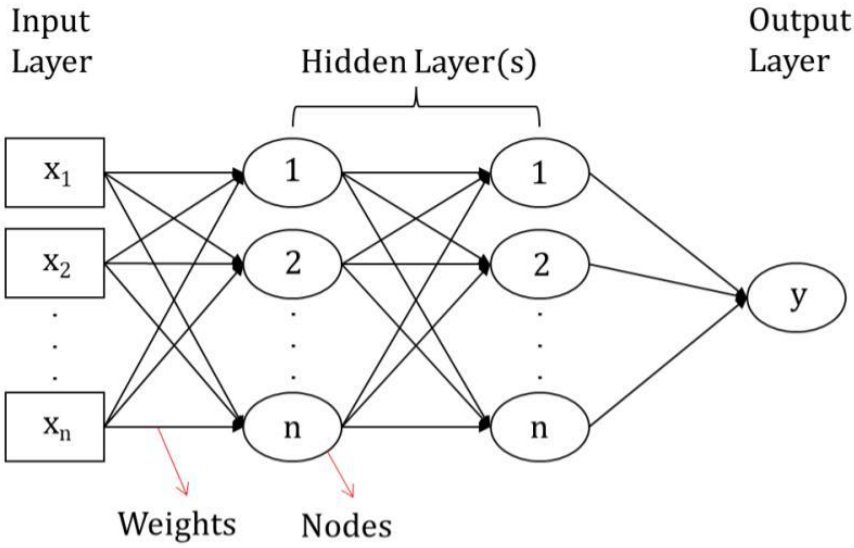

Figure 3.17: Typical artificial neural network feed-forward algorithm schematic.

3.6.5 Sensitivity, Specificity, and Classification Error

When using a clinical test to diagnose an infection or disease in a patient, the accuracy of the clinical test is highly important. It must be able to correctly identify the disease, and also be able to correctly determine if the disease is not present. This is the concept of a true positive and a true negative. The rate of true positives and true negatives can be measured by finding the sensitivity and specificity of a test. Chemometric algorithms calculate sensitivities and specificities when performing a discrimination to give the user an idea of the accuracy of the test.[83]

The sensitivity is the proportion of samples that result in a positive test result and are genuinly positive. The sensitivity also depends on the number of false negatives, which are also known as type II errors. False negatives occur when the test returns a negative result when it is actually positive. Sensitivities are often reported as percentages. The formula for the sensitivity is described by the following equation:

$$\boldsymbol{Sensitivity} = \frac{TP}{TP + FN} \tag{10}$$

Where TP is true positive, and FN is false negative. The optimal medical test would have a sensitivity, or true positive rate, of 100%. This means that the medical test would diagnose

100% of all cases of disease. Any sensitivity lower than 100% would mean that some cases of disease were missed. Though catching 100% of cases is desirable, it comes at a cost. In the case of 100% sensitivity the test would likely be so sensitive that it reports positive results in people not carrying the disease or infection. To balance this, a quantity called the specificity is required.[84]

The counterpart to the sensitivity is the specificity, which is the true negative rate of the test. This is the proportion of the samples that result in a negative test result that are genuinly negative. False positives occur when the test returns a positive result when there is no disease present, this is known as a type I error. Specificities are also often reported as percentages, and the specificity is described as the following:

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP} \tag{11}$$

Where TN is true negative, and FP is false positive. An optimal medical test would have a specificity of 100%, meaning that no one is misdiagnosed. Values of specificity lower than 100% would mean false positives occur. False positives are problematic for the healthcare system, as they could cause unnecessary procedures, medications, and undue stress on the patient. However, as with sensitivity, 100% specificity can be problematic as well. Having a perfectly specific test could result in all members of the population testing negative for a disease even if they do have it.

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 for the other. Ideally, the sensitivity and specificity for a medical test should be 100%, but there is no medical test that can currently achieve this level of accuracy. Rather, the focus is on achieving a balance between the two values. To summarize the overall perforance of a medical test, classification accuracy is the metric used, which is defined as the fraction of predictions a model or test got right to the total number of predictions. This can be summarized as:

$$\boldsymbol{Classification\ Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

[57] Cremers, D., Radziemski, L.J. (2013). Introduction, *Handbook of Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 1-18). Wiley.

[58] Thakur, S. N., Singh, J.P. (2020). Fundamentals of LIBS and Recent Developments. In Thakur, S. N., & Singh, J.P. (Eds.), *Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 3-19). Elsevier Science.

[59] Senesi, G., Senesi, N. (2016). Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. *Analytica Chimica B.*

[60] Hecht, E. (2002). Modern Optics: Lasers and Other Topics. In Black, A. (Ed.), *Optics* (4th ed., pp. 581-639). Pearson.

[61] Thakur, S. N., Singh, J.P. (2020). LIBS for Aerosol Analysis. In Thakur, S. N., & Singh, J.P. (Eds.), *Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 3-19). Elsevier Science.

[62] Cremers, D., Radziemski, L.J. (2013). Basics of the LIBS Plasma, *Handbook of Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 1-18). Wiley.

[63] Marvin, J. (2018). *Signal Optimization and Enhancement of Laser-Induced Breakdown Spectroscopy for Discrimination of Bacterial Organisms.* [Master's thesis, University of Windsor].

[64] Dawood, M. (2015). *Space and Time Characterization of Laser-Induced Plasmas for Applications in Chemical Analysis and Thin Film Deposition*. [Doctoral dissertation, University of Montreal]. http://hdl.handle.net/1866/12347.

[65] Thakur, S. N., Singh, J.P. (2020). Study of the Different Parts of a Tokamak Using Laser-Induced Breakdown Spectroscopy. In Thakur, S. N., & Singh, J.P. (Eds.), *Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 3-19). Elsevier Science.

[66] Miziolek, A. W., Palleschi, V., & Schechter, I. (Eds.). (2006). *Laser Induced Breakdown Spectroscopy.* Cambridge University Press.

[67] Rai, V. N., Thakur, S. N.(2020). Physics and Dynamics of Plasmas. In Thakur, S. N., & Singh, J.P. (Eds.), *Laser-Induced Breakdown Spectroscopy* (2nd ed., pp. 3-19). Elsevier Science.

[68] Installation Guidelines Echelle Spectra Analyzer ESA 3000, LLA Instruments GmbH, Berlin, Germany, 2005.

[69] Engleberg, N. C. (2007). Diagnostic Principles. In N. C. Engleberg, V. J. DiRita, T. Dermody, & M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 503–512). Chapter, Lippincott Williams & Wilkins.

[70] Marquis, R. E., Mayzel, K., Carstensen, E. L. (1976). Cation exchange in cell walls of Gram-positive bacteria. *Can J Microbiol, 22*(7), 975-982. https://doi.org/10/1139/m76-142.822931.

[71] Salyers, A. A., Whitt, D. D. (2002). *Bacterial Pathogenesis: A Molecular Approach* (2nd ed., pp. 84-100). ASM Press.

[72] Schaechter, M. (2007). Introduction to Pathogenic Bacteria. In N. C. Engleberg, V. J. DiRita, T. Dermody, & M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 503–512). Chapter, Lippincott Williams & Wilkins.

[73] Center for Disease Control and Prevention. (2019). *Antibiotic Resistant Threats in the United States, 2019*. U.S. Department of Health and Human Services. https://doi.org/10.15620/cdc:82532.

[74] John, J.F, Jr., Sharbaugh, R.J., & Bannister, E. R. (1982). Enterobacter cloacae: bacteremia, epidemiology, and antibiotic resistance. *Reviews of Infectious Diseases*, *4*(1), 13–28. http://www.jstor.org/stable/4452695.

[75] Malenfant, D. (2016). *Indfluences on the Emissions of Bacterial Plasmas Generated Through Nanosecond Laser-Induced Breakdown Spectroscopy.* [Master's thesis, University of Windsor].

[76] Paulick, A. (2018). *Development of Laser-Induced Breakdown Spectroscopy as a Rapid Diagnostic Tool for Bacterial Infection.* [Master's thesis, University of Windsor].

[77] Salyers, A. A., Whitt, D. D. (2002).The Second Line of Defense: Antibodies and Cytotoxic T Cells. In *Bacterial Pathogenesis: a Molecular Approach* (2nd ed., pp. 84-100). Chapter, ASM Press.

[78] Gottfried, J. L. (2011). Discrimination of biological and chemical threat simulants in residue mixtures on multiple substrates. *Analytical and Bioanalytical Chemistry*, *400*(10), 3289–3301. https://doi.org/10.1007/s00216-011-4746-4.

[79] Putnam, R. A., Mohaidat, Q. I., Daabous, A., & Rehse, S. J., (2013). A comparison of multivariate analysis techniques and variable selection strategies in a laser-induced breakdown spectroscopy bacterial classification. *Spectrochimica Acta Part B: Atomic Spectroscopy, 87,* 161-167. https://doi.org/10.1016/j.sab.2013.05.014.

[80] Brereton, R. G. (2007). *Applied Chemometrics for Scientists*. (1st ed). Wiley.

[81] Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, *26*, 195-197. https://doi.org/10.1038/nbt1386

[82] Malekian, A., & Chitsaz, N. (2021). Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In P. Sharma & D. Machiwal (Eds.), *Advances in Streamflow Forecasting* (pp. 115–147). Elsevier Gezondheidszorg. https://doi.org/10.1016/B978-0-12-820673-7.00003-2.

[83] Lalkhen, A. G., McCluskey A., (2008). Clinical tests: sensitivity and specificity*, Continuing Education in Anaesthesia Critical Care & Pain, 8(6), 221-223.* https://doi.org/10.1093/bjaceaccp/mkn041.

[84] Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(1), 45–50. https://doi.org/10.4103/0301-4738.37595.

Chapter 4: Discrimination of *E. coli* From Sterile Water

LIBS is often touted as being a fast diagnostic technique with minimal to no sample preparation. However, the reality of LIBS on bacteria is that detecting bacteria with minimal to no sample preparation requires a 'solid' or pelletized bacterial colony. These solid pellets of bacteria do not represent what a clinician would sample from the body, and over represents the number of bacteria that would be present in a clinical sample. To achieve the high numbers of bacteria in solid pellets, culturing for 24-78 hours would be required.

Other groups have attempted methods to easily detect trace numbers of cells. One such method which has proven to be very sensitive is serologically tagging cells with unique elemental nanoparticles. The immediate drawback of this is that a priori knowledge of the identity of the cells is required to correctly tag them, or they would have to be tested against all known immunoassays in parallel. Another approach was to deposit a small number of cells on inorganic substrates that do not exhibit the same lines as bacterial specimens, such as aluminium or steel disks. A drawback to this approach was the lack of porosity of the substrates; if any liquid is present with the cells an additional overnight evaporation step would be required to accurately identify them. As well this technique has only been shown to be effective for sampling small volumes of liquid, between 5 to 10 µL.[85]

Our group has addressed the problem of sample preparation time with disposable nitrocellulose filters. These filters are porous and allow significantly larger volumes of bacterial suspensions to pass when centrifuged. Addition of a concentration cone placed on the filter concentrates bacteria and has lowered the limit of detection of *Escherichia coli* to approximately 10,000 CFU's, from a previous 50,000 CFU's.[86] Being able to reliably achieve high signal from bacterial specimens when compared to the background signal is crucial for LIBS-based pathogen identification. Therefore, much of the work in this chapter focuses on efforts to decrease the level of background signal in our spectra from filter and water so that the signal from bacterial specimens is comparatively higher. This chapter also focuses on enhancing detection of bacteria by optimizing chemometric algorithms for discrimination between bacteria and water.

4.1 Calibration Curve and *E. coli* Spectral Library

For this study, LIBS data was acquired for various concentrations of *E. coli*, 'blank' nitrocellulose filters, and 'sterile' water nitrocellulose filters. The blank nitrocellulose filters are filters with nothing deposited on them. The sterile water filters are nitrocellulose filters with sterile water deposited on them. 20 to 30 spectra were obtained from each filter, and the data spans two years. To quantify the LIBS signal, normalized spectral intensities were graphed. The results of this study are shown in Figure 4.1.

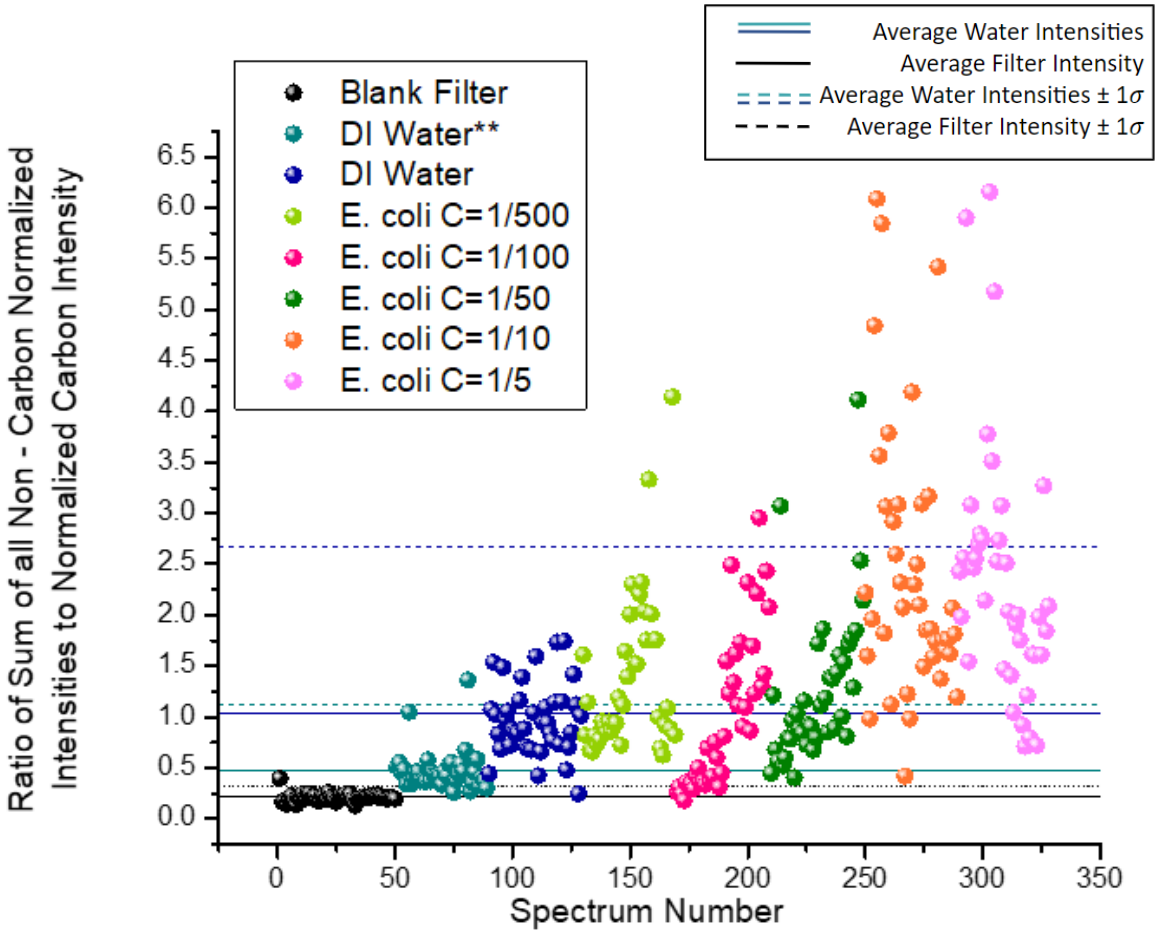

*Figure 4.1: LIBS bacterial curve of growth constructed from serial dilutions of E. coli (**represents a 24-hour drying period after deposition).*

The graph plots the ratio of the sum of all non-carbon normalized intensities to the normalized carbon intensity for each spectrum. Bacterial spectra contain strong emission lines from calcium, magnesium, sodium, and phosphorus, as opposed to blank filter spectra which contain only strong emission from carbon. Bacterial spectra therefore maximize the

numerator and minimize the denominator of the vertical scale in Figure 4.1. The blank filter spectra have the lowest ratio, at approximately 0 since they contain only carbon. Deionized water, represented by the light and dark blue, contained some contamination resulting in ratios between 1 and 2. Increased intensity above the deionized water show the presence of bacteria, with the general trend of increasing intensity for increasing concentration of bacteria.

While Figure 4.1 signal intensity trends upwards with increasing bacterial concentration, it can be observed that there is a large amount of variability between single-shot bacterial spectra. We have attributed the large variance in the data in part to the interaction of the LIBS apparatus with a non-uniform surface. LIBS ablation produces the most reproducible data on solid uniform surfaces; this is shown clearly by the blank filter data which have the smallest amount of variability. In contrast, bacterial depositions resemble a non-uniform film, with bubbles, cracks, and gaps. It is this reason that LIBS ablation repeatability on a bacterial target is poorer than the uniform filter surface.[85] At times the bacterial film ablates well, leading to a maximum value of approximately 6 in Figure 4.1. When the bacterial film does not ablate well, we see an intensity that is on par with deionized water. This variability in the data makes detecting bacteria on a background of water very difficult. The graph essentially leads to 2 important research questions: how can the background signal be as close to zero as possible, and how can the repeatability of the non-zero bacterial spectra be improved. Overcoming the similarity between water and low intensity single-shot spectra of bacteria is the primary focus of this chapter.

It should also be noted here that although the increasing concentrations of bacteria followed a general upwards trend in ratio, there was no consistent linear curve of growth. This can be explained by the bacterial suspensions being a true suspension rather than a solution. A solution contains evenly dispersed solvent and can be characterized by a linear curve of growth. In suspension, bacteria tend to clump together or form strings, making the solution more heterogenous and rendering a true solution difficult.[87]

4.2 Lowering the Background Signal

This section details several methods for lowering the background signal and improving repeatability. Different waters for bacterial storage and sample preparation were investigated, and vigorous cleaning steps were also tested to further reduce background signal of the water. Methods of outlier rejection were also studied to curb the issues of filter reproducibility.

4.2.1 Improving Cleaning Procedures for Background Reduction

I address the first question of reducing the background signal by investigating cleaning procedures on pieces used for deposition of bacteria onto nitrocellulose filter. This study was done by myself and colleague Sydney Sleiman. Deposition procedure was previously discussed in Chapter 3 section 3.5.3. For this study we tested several different cleaning procedures on the metal plate, the metal cone, and the swab. The results on the metal plate and swab and swab will then be discussed first, followed by the metal cone, thereby following the chronological order in which the tests were conducted.

The metal plate and the metal cone were cleaned using either bleach and water, or ultrasonicating in acetone and methanol for 5 minutes each. We chose to ultrasonicate because it is a more vigorous cleaning procedure than using bleach and water. We used 'uncleaned' pieces as our control, where we defined uncleaned as using pieces without cleaning after previous use for bacterial deposition. Deionized water was deposited onto nitrocellulose filter for each cleaning procedure. We wanted to achieve a low background to detect any contamination while also having fluid running through the cone to ensure that we were properly simulating deposition of bacteria on the nitrocellulose surface. The swab was cleaned by vortexing it in deionized water before using it to pick up the bacteria to shake off any potential contamination. Each cleaning procedure was performed once per deposition on only one metal piece at a time.

First the metal plate was tested. Spectra of the uncleaned metal plate, metal plate cleaned in bleach water, and metal plate ultrasonicated in acetone and methanol for 5 minutes are shown in Figure 4.2. Figure 4.3 shows a bar graph comparing the normalized intensities of the 3 cleaning methods with error bars. As suspected, the uncleaned metal

plate spectrum showed higher contamination; specifically higher contamination was seen for magnesium and calcium. There was however very little difference between cleaning in bleach water and cleaning with ultrasonication. As per Figure 4.3, the carbon line and all metal lines are within error of each other, and no marked improvement is seen. The same is true for the spectra; there is no marked improvement between our original cleaning procedure with bleach and water and the new procedure of ultrasonication. This led us to conclude that the metal plate was not a major source of the contamination we see in our deionized water samples.

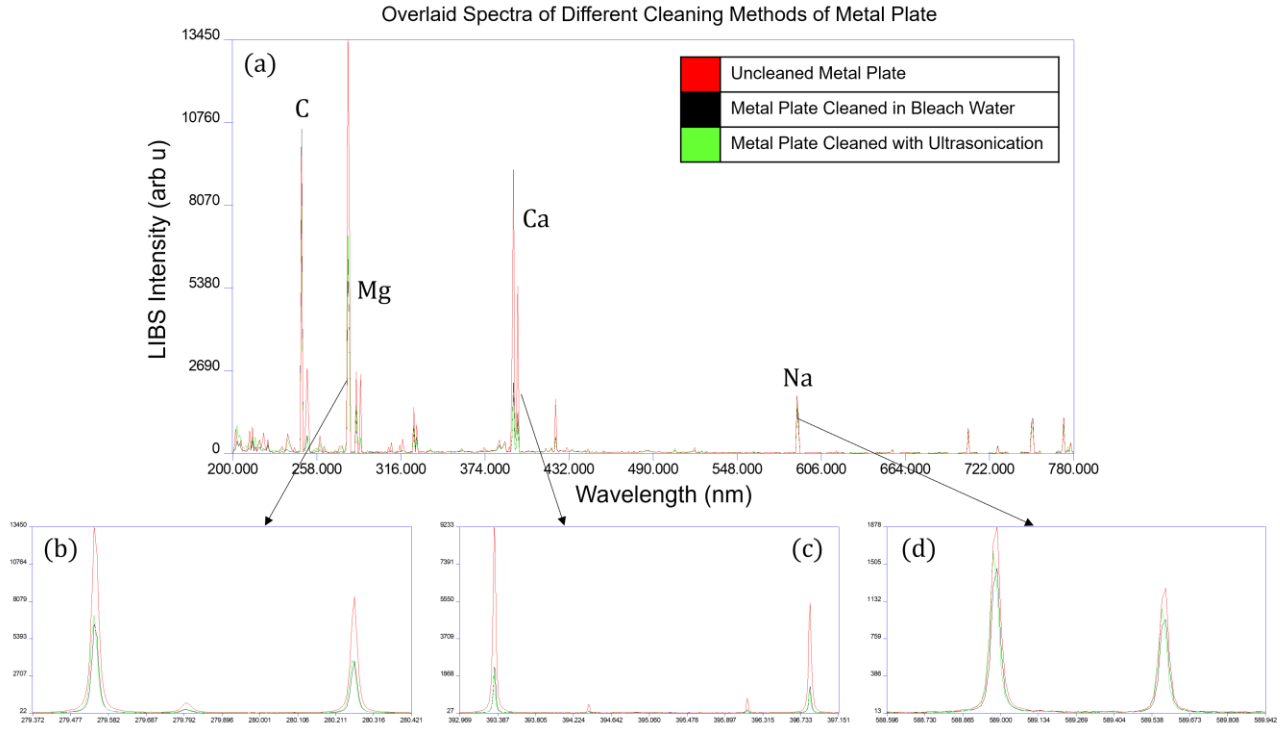

Figure 4.2: Spectra comparing different metal plate cleaning procedures; (a) shows an overlaid spectra of uncleaned metal plate (red), metal plate cleaned with bleach and water (black), and metal plate cleaned with ultrasonication (green). (b) Comparison of magnesium lines for 3 cleaning methods, cleaning reduces magnesium line intensity. (c) Comparison of calcium lines for 3 cleaning methods, cleaning reduces the calcium line intensity. (d) Comparison of sodium lines for 3 cleaning methods, cleaning reduces sodium line intensity.

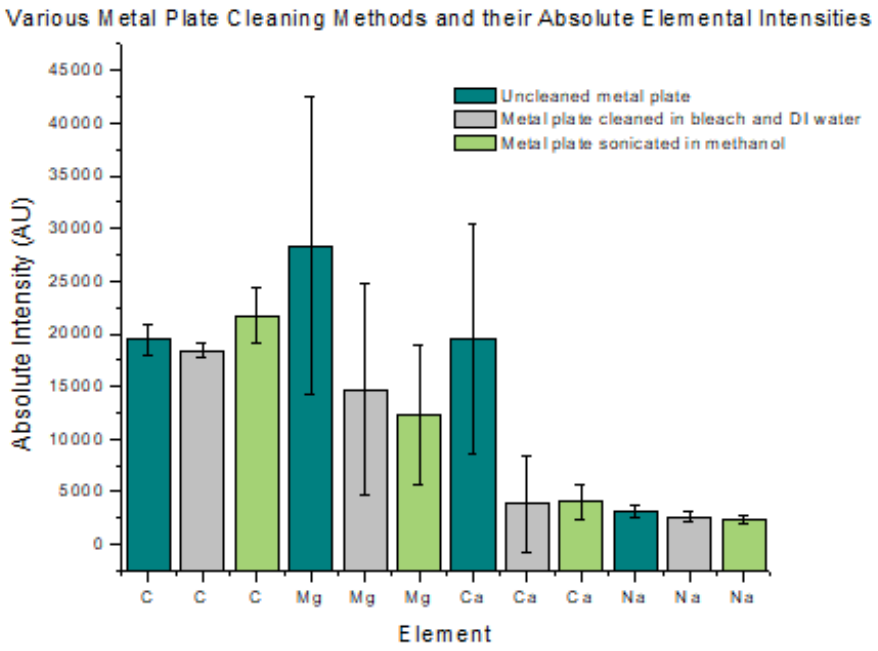

Figure 4.3: Bar graph comparing elemental intensities for 3 cleaning procedures. Cleaning appeared to reduce intensities of lines contributing to the background, although most intensity reductions are within error.

Pre-cleaning of the swab that is used to collect bacteria was tested. We also tested the effect of the metal cone and swab together on potential contamination. The results of this are given in Figures 4.4 and 4.5. Depositing deionized water onto filter without the metal cone showed very similar signal between pre-cleaned swab and uncleaned swab. This is evident by the similarity between the red and black spectra in Figure 4.4, and the similarity between the grey and purple bars in Figure 4.5, leading to the conclusion that pre-cleaning does not substantially effect background signal. However, using the metal cone to deposit the sample shows consistently higher signal than the case where no metal cone was used. Additionally, the spectra acquired with the metal cone were virtually identical in intensity. These observations led to 2 conclusions; one being that the swab is fairly sterile and pre-cleaning makes no difference. The second conclusion suggested that the major source of contamination likely came from the metal cone. Therefore, rigorous cleaning of the metal cone was investigated to reduce this background signal.

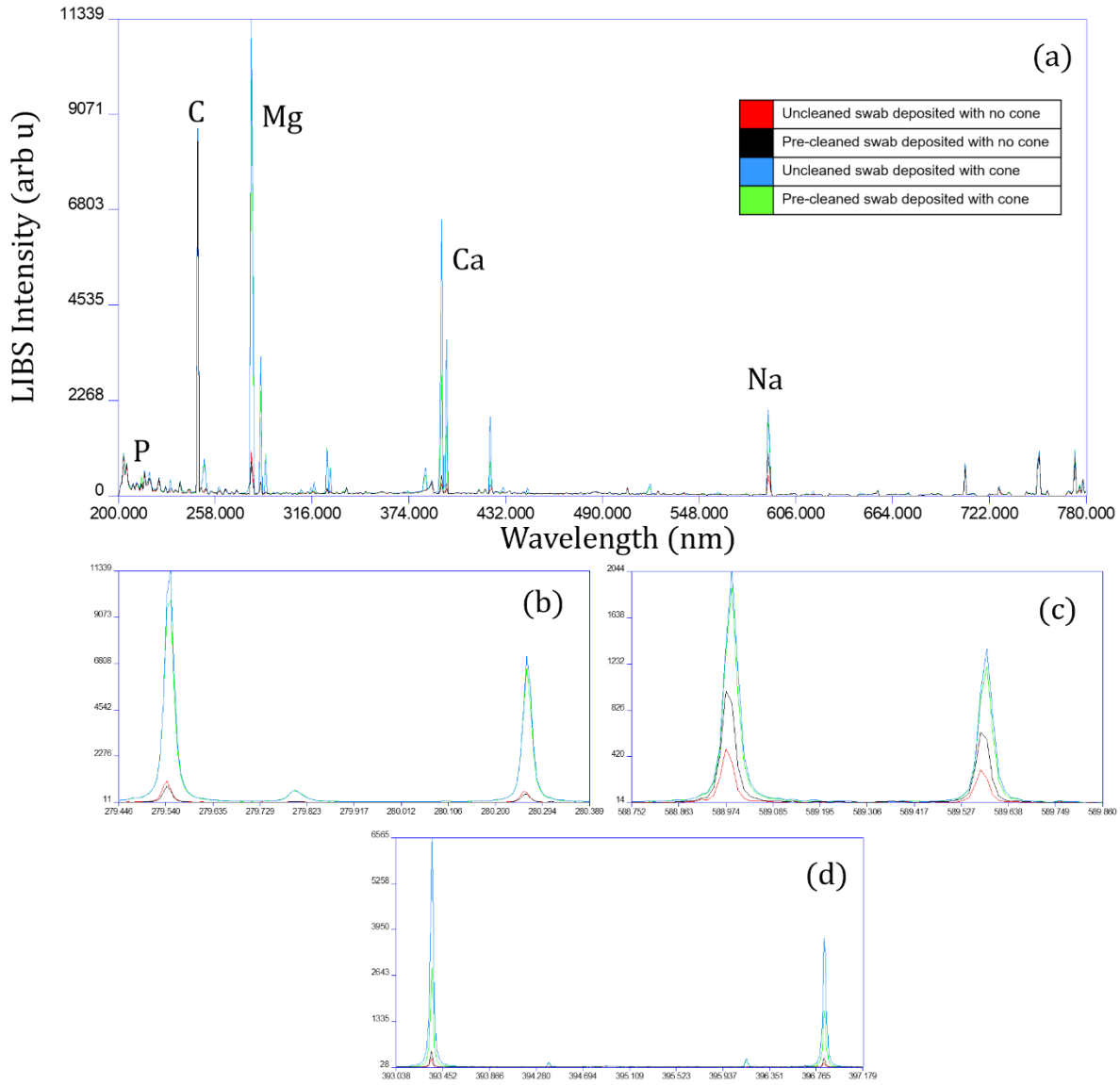

Figure 4.4: (a) Spectrum of swab cleaning procedures. Green spectra: pre-cleaned swab, water deposited through cone. Blue spectra: uncleaned swab, water deposited through cone. Black spectra: pre-cleaned swab, cone not used. Red spectra: uncleaned swab, cone not used. It is evident that the swabs used with the metal cone produced higher signal due to the concentration of the fluid on the filter and contamination of the water by the cone. (b) Close-up of magnesium line.

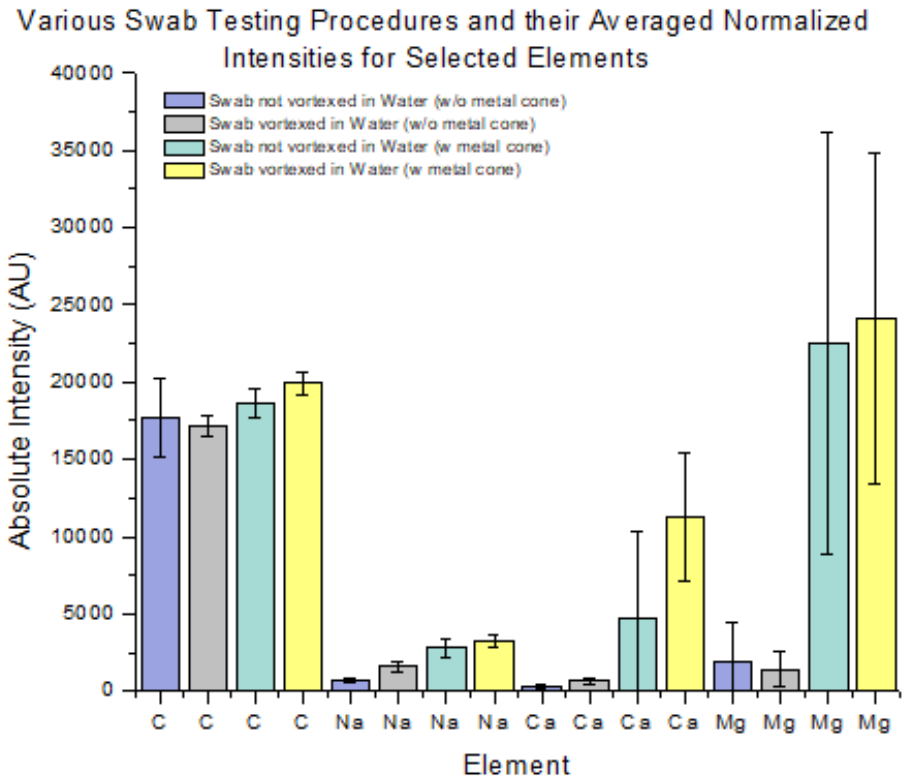

Figure 4.5: Bar graph comparing the elemental intensities for 2 cleaning procedures and 2 deposition methods. The samples deposited with the metal cone have significantly higher intensity than those deposited without cone. However, pre-cleaning of the swab makes no statistical difference for most lines.

For the metal cone, the same cleaning and deposition procedures were used as the metal plate. The resulting spectra of these 3 cleaning methods are shown in Figure 4.6 and a bar graph comparing the normalized intensity of our most important lines is shown in Figure 4.7. When analyzing the spectrum, the ultrasonication reduces the sodium line and reduces the calcium line, while the magnesium line remains relatively the same. This is supported by the graph in Figure 4.6 where it is clear that the sodium line has been significantly reduced, the calcium line has been reduced within error, and the magnesium line is relatively the same. This observation makes sense since the cone acts as a concentration step, so if it is concentrating bacteria, it may also be concentrating contamination. Based on the reduction of the sodium line and slight reduction of calcium, it was concluded that the cone was a source of contamination and it was recommended that the metal cone be ultrasonicated in acetone for 5 minutes followed by methanol for 5 minutes for all bacterial sample preparation. All other cleaning and deposition methods remained the same as before.

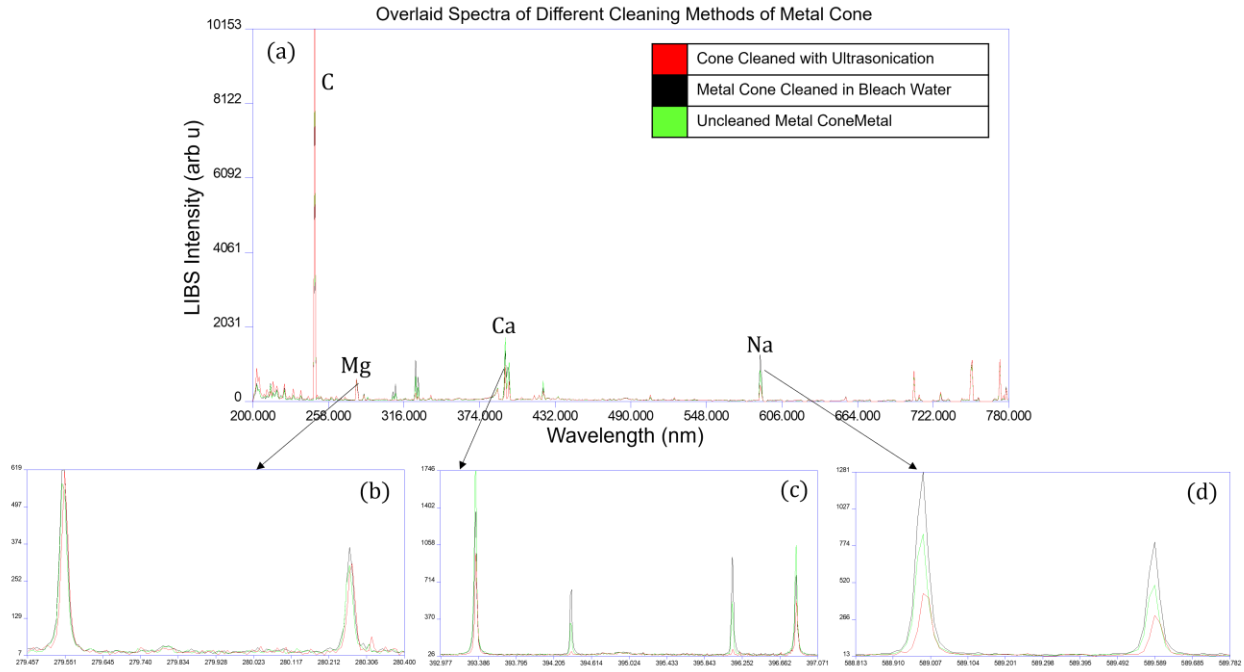

Figure 4.6: Spectra comparing different metal cone cleaning procedures; (a) shows an overlaid spectra of uncleaned metal cone (green), metal cone cleaned with bleach and water (black), and metal cone cleaned with ultrasonication (red). (b) Comparison of magnesium lines for 3 cleaning methods, cleaning does not appear to reduce magnesium line intensity. (c) Comparison of calcium lines for 3 cleaning methods, cleaning reduces the calcium line intensity. (d) Comparison of sodium lines for 3 cleaning methods, cleaning reduces sodium line intensity.

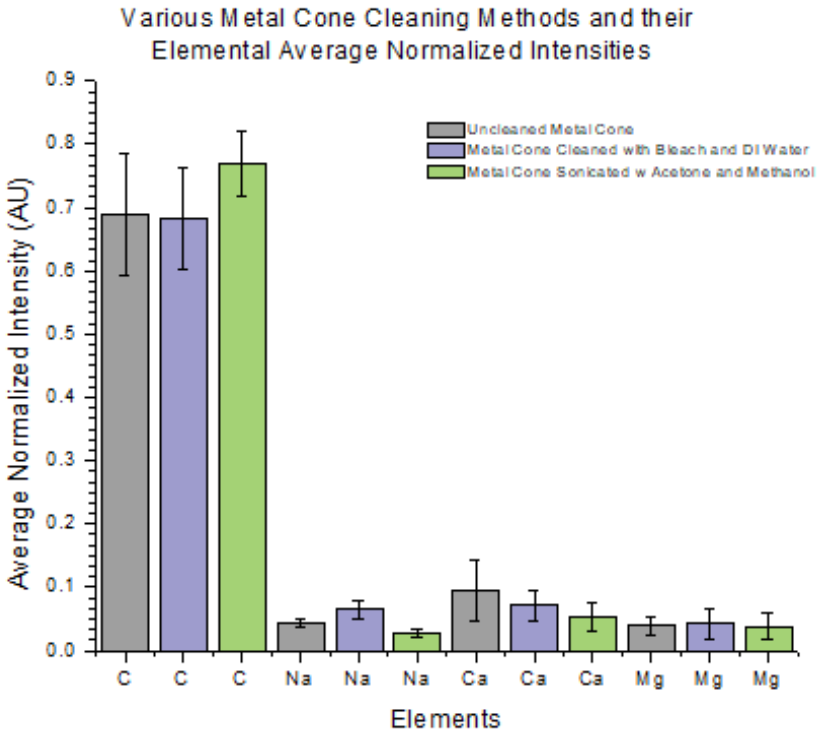

Figure 4.7: Bar graph comparing elemental intensities of 3 cleaning methods for the metal cone. Cleaning appeared to significantly reduce the intensity of the sodium line. The calcium line intensity appears to be gradually reduced for each cleaning method within error. The magnesium line intensity is relatively the same across all methods.

4.2.2 Investigation of Filter Media to Lower the Background Signal

Along with cleaning procedures, other filter media were investigated to see if any other porous filter would offer a lower carbon signal. The filter we use currently, a Millipore nitrocellulose filter, has a very high carbon signal, which can interfere with the measurement of carbon emission from the bacteria. The intensity of the carbon line also prevents the use of higher spectrometer amplification values since the light from the carbon could damage it. The filter contains traces of other elements important to bacterial detection as well including calcium, magnesium, and sodium, but these are in such small amounts that they do not interfere with bacterial detection. Two other types of filter media were investigated: Millipore Durapore filters (GVWP01300, Millipore Inc.) and Millipore glass microfiber filters (1825-090, Whatman). The spectral intensity of each of these filters was analyzed to determine which filter offered the lowest background signal and solved the problem with the high carbon intensity. A semi-log plot comparing these intensities is shown in Figure 4.8 and a figure comparing the filters after laser ablation is shown in Figure 4.9.

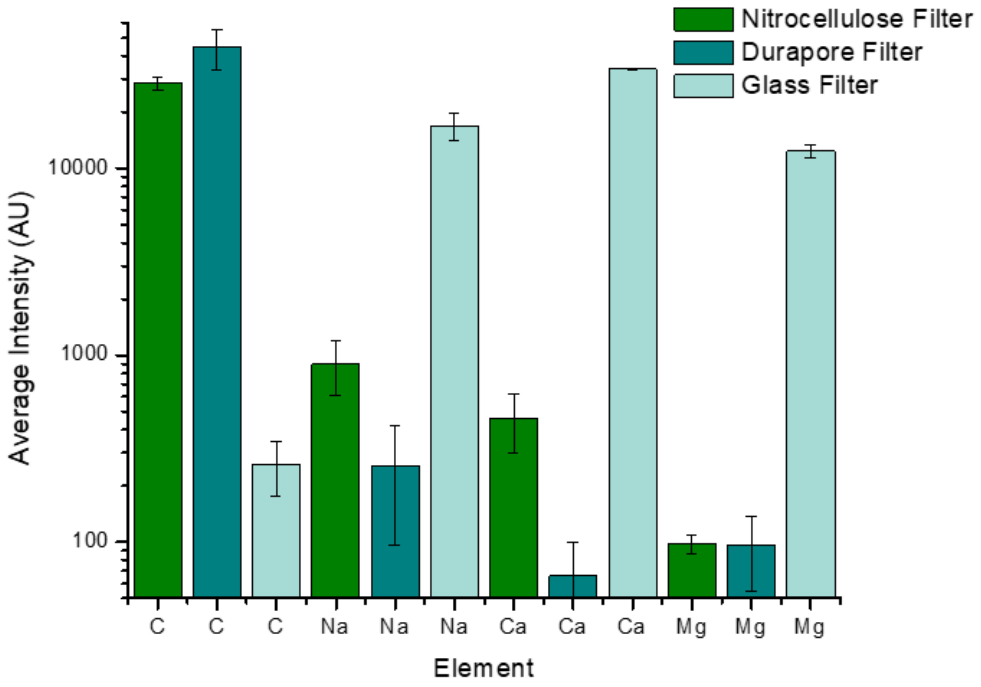

Figure 4.8: Semi-log plot of line intensities for nitrocellulose (green), Durapore (dark blue), and glass microfiber (light blue) filters.

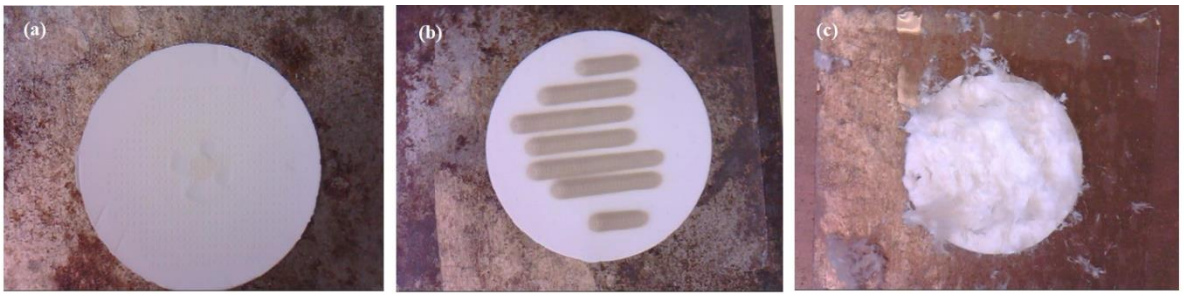

Figure 4.9: Images of (a) nitrocellulose, (b) Durapore, and (c) glass microfiber filters after ablation. Laser ablation on the Durapore filter in (b) causes scorching and destruction of testing surface on the glass microfiber filter in (c).

Figure 4.8 shows that the Durapore filter has lower levels of sodium and calcium, however it has higher levels of carbon than the nitrocellulose filter. Since the carbon line is the main problem with the filter background signal, Durapore filters will not help the background signal. The glass microfiber filter has the lowest intensity of carbon of the three, however it has much higher intensities of sodium, magnesium, and calcium lines which will heavily interfere with bacterial lines. As well, ablation of the glass microfiber filter causes destruction of the testing surface which is shown in Figure 4.9c. The destruction of the testing surface creates a textured and more topographically complex surface which is unideal for LIBS. It was concluded that the nitrocellulose filter was the most ideal filter medium for testing bacterial specimens and all tests have continued to use nitrocellulose filters.

4.2.3 Investigating Water to Reduce Background Signal

After reducing the background signal with more aggressive cleaning procedures and investigating other filter media, the next step was to find a type of water that achieved the lowest background signal possible. Blank nitrocellulose filter gives the lowest background signal that we can obtain with the LIBS apparatus. A more realistic control is water deposited on nitrocellulose filter since bacteria are suspended in and prepared with deionized water. Therefore, our aim is to use the water that causes the lowest level of background signal. Several kinds of water were investigated with the goal of lowering the background signal present in LIBS spectra and to improve classification between bacteria and sterile water.

We tested distilled water, deionized water, and tap water. Tap water was collected from the sink in Essex Hall 390-2. A jug of distilled water was purchased from a store, and

deionized water was collected from the first-year chemistry labs on the first floor of Essex Hall. All three water samples were prepared the same way that bacterial samples were prepared, previously described in section 3.5.3. Tap water was used as a positive control, to ensure that unfiltered water gave the highest spectral signal, the most lines, or some combination of the two. Tap water contained all the same lines as our bacterial species with much higher intensities than other water types and of comparable intensity to the bacterial lines, confirming our suspicions. A spectrum of tap water is shown in Figure 4.10a; note the intense calcium line at 393 nm and the high magnesium lines around 279 nm. These lines would heavily interfere with the bacterial signal. The distilled water showed a marked improvement over the tap water, but still exhibited some intense lines, most notably the calcium line at 393 nm and the magnesium line at 279 nm, shown in Figure 4.10b. The deionized water, shown in Figure 4.10c, is similar to the distilled water but with one improvement; though the calcium line is still high, the magnesium line is much smaller. Deionized water however is still unideal because it is not very close to the more optimal blank filter spectrum, shown in Figure 4.10d, it is the water that provides the least amount of background signal.

A spectrum of the tap, distilled, and deionized water for easier comparison is shown in Figure 4.10e. All three spectra are overlaid on one another to show the drastic difference between tap and the other two waters. The overlay also shows the subtle differences between the distilled water in black and DI water in green. One of the major differences mentioned previously being the larger presence of the magnesium 279 line in distilled water. Otherwise there are approximately equal amounts of calcium and sodium in deionized and distilled.

Based on this conclusion, deionized water was used as the medium for bacterial suspensions and was used as the medium for all sample preparations from the period of 2019 to 2021.

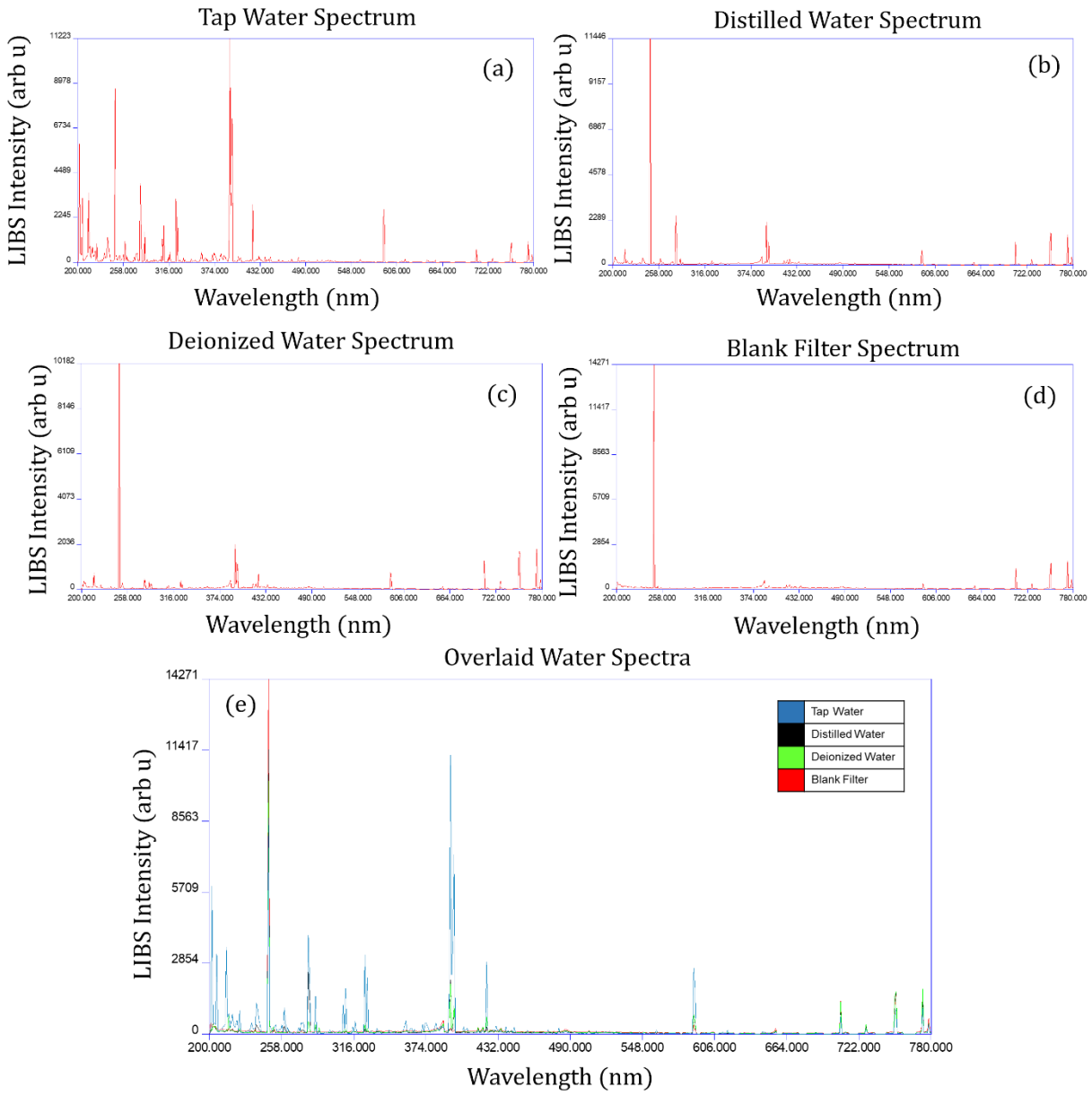

Figure 4.10: (a) A typical tap water spectrum. Several lines can be observed at high intensity, making this medium unideal when working with bacteria. (b) A typical distilled water spectrum. Calcium, magnesium, and sodium lines are clearly visible. (c) A typical deionized water spectrum. Magnesium lines are visible but small, calcium and sodium lines are larger and visible. (d) A typical blank filter spectrum, shown for comparison. Ideally, water spectra would be this low. (e) Water spectra overlaid with blank filter for comparison.

Though deionized water produced a low background signal, it was still not ideal as some of the lines present in water were the same as the bacterial lines and had the same intensity as low intensity bacterial spectra. Based on this and the observation that deionized water did not achieve reliable discrimination against other bacteria, I pursued using ultrapure water to address these concerns. Ultrapure water is water that reaches

very stringent levels of purity and has all possible contaminants removed, leading to high levels of resistance in the mega ohm region. To study the ultrapure water, I chose to use a new sample preparation procedure developed with the purpose of studying the separation of bodily fluids and bacterial cells. For this procedure the cone is cleaned by ultrasonication in acetone and methanol, the centrifuge inserts and tubes are cleaned in bleach-water solution. The filter and centrifuge insert are prepared normally (see section 3.5.3), and the cone is placed immediately into the centrifuge tube. 1 mL of ultrapure water was pipetted through the cone and the centrifuge tube was centrifuged at the same settings as those in section 3.5.3. Several filters of ultrapure water were analyzed this way. A typical ultrapure water spectrum that we see is shown in Figure 4.11a. From this spectrum it is apparent that ultrapure water offers a huge improvement over the other types of water. This is also seen in Figure 4.11b, where an ultrapure water spectrum is compared with a deionized water and blank filter spectra. It is clear from this that ultrapure water offers a consistently low signal that is almost identical to blank filter. It is also apparent from the scatter plot in Figure 4.12 that compared to the blank filter data, the ultrapure water has both a lower average intensity, given by the solid purple line, and a smaller amount of variance, given by the dashed purple line. Not only does ultrapure water achieve a background signal consistently as low as blank filter, it also achieves a reproducible background comparable to that of blank filter. Based on these findings, it was recommended that ultrapure water be used in sample prep and in bacterial suspensions.

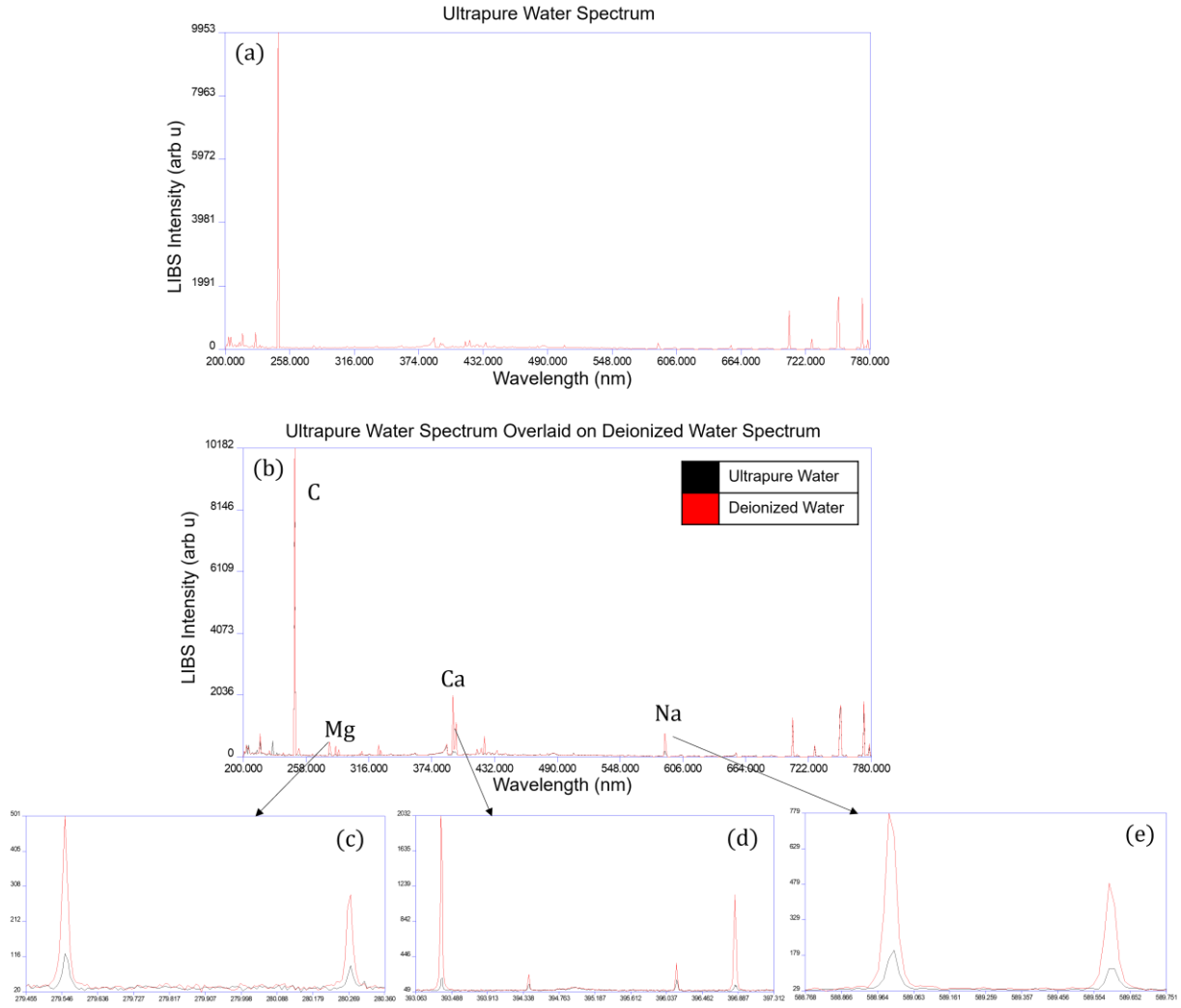

Figure 4.11: Comparison between ultrapure water spectrum and deionized water spectrum. (a) A typical ultrapure water spectrum, intensities of calcium, magnesium, and sodium lines are low. (b) Overlaid spectra of ultrapure and deionized water for comparison. Deionized water clearly has higher calcium, magnesium, and sodium emission. (c) Comparison of magnesium line intensity between the water types, magnesium emission is reduced with ultrapure water. (d) Comparison of calcium line intensity between water types, calcium emission is reduced with ultrapure water. (e) Comparison of sodium line emission between water types, sodium emission is reduced with ultrapure water.

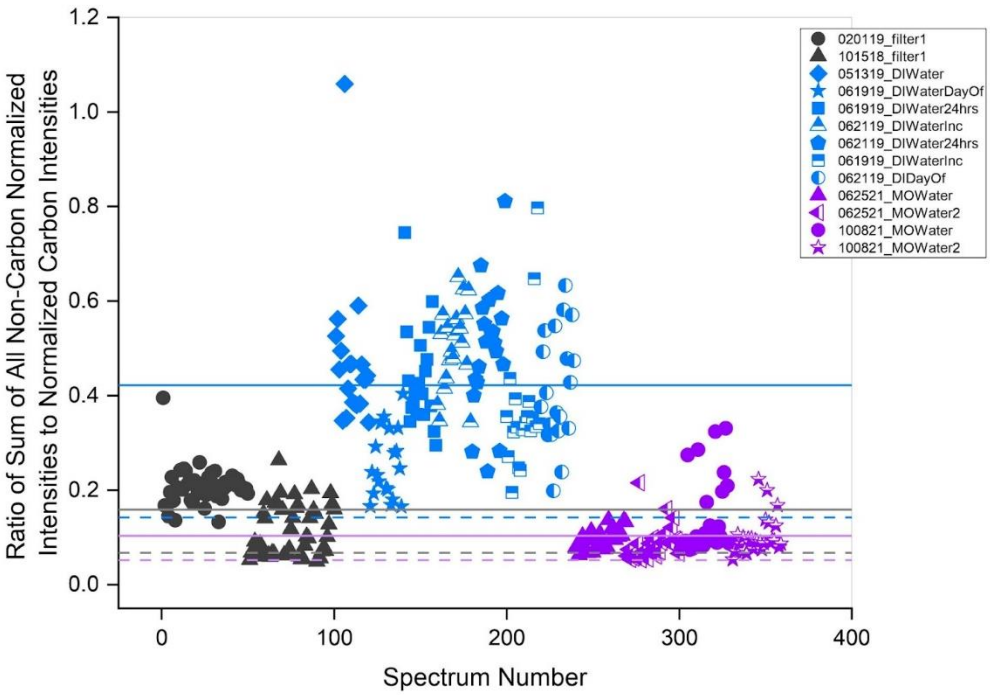

Figure 4.12: Graphical comparison of deionized water (blue), blank filter (black), and ultrapure water (purple). The average intensity of the ultrapure water (solid purple line) is lower than both the blank filter (solid black line) and deionized water (solid blue line). The dashed lines show the standard deviation of each species. Ultrapure water has a smaller standard of deviation than blank filter, deionzied water has the largest standard deviation.

4.3 Reducing the Scatter in Bacterial Data by Removing Outliers

Referring back to Figure 4.1, the scatter in the bacterial data ranges from as low as blank filter, to much higher. This variation posed some problems with the discrimination between species using PLSDA and DFA; poor separation along with low sensitivity and specificity were observed consistently as shown in Figure 4.13. In order to improve the discrimination, I investigated several outlier rejection schemes to remove the lowest intensity data. Outlier rejection is not a new topic in LIBS; it has been investigated extensively and has been implemented with great success.[88,89] LIBS is a process that depends on the laser-matter interactions and the homogeneity of the material, therefore it is sometimes necessary to remove outliers from the data, yet it is always necessary to retain data with sufficient variance to allow the construction of statistically robust chemometric models.[85,89] This fluctuation of the laser-matter interaction quality is especially applicable to the bacteria surface, as it contains many inhomogeneities, including cracks and bubbles that exist on the same scale as the laser beam. In the case of our bacteria, some spectra are not indicative of what is on the surface if they appear empty or

lower in intensity than previous spectra. This obviates the need for some outlier rejection scheme.

Several groups have investigated outlier rejection in conjunction with LIBS and chemometric analysis in the past, and some have implemented it successfully. El Haddad et al. have advocated for removal of outliers to improve classification results.[90] Sahoo et al. have studied outlier rejection prior to classification and found an increase in the accuracy of classification.[91] Cisewski et al. and Yueh et al. have also applied outlier rejection as a preprocessing step for classification of spore and tissues resulting in successful classification.[92,93] Implementing outlier rejection as a preprocessing step has the advantage of minimizing the effect of poor data on classification results. It has been noted by several of these authors that because chemometrics rely on high numbers of data for accurate results, removing any input data could potentially result in a decrease in classification accuracy.[6] However, this can be circumvented by collecting more data.

I investigated two main methods of outlier rejection. The first method I investigated was defining a predetermined threshold based on the deionized water intensity average. To find a threshold, emission lines for a single spectrum were first summed together, resulting in a total spectral intensity. The average of all individual spectral intensities was taken to determine the average total spectral intensity of water. The standard deviation of the average total spectral intensity was also found. Both the average and standard deviation of the water were plotted, the result is shown in Figure 4.1. Initially, it was planned that all bacterial spectra below the water average plus 3 standard deviations would be removed. Upon inspection of Figure 4.1 however, this would result in rejection of most of the data, thus it was decided to remove only those data points below water plus 1 standard deviation.

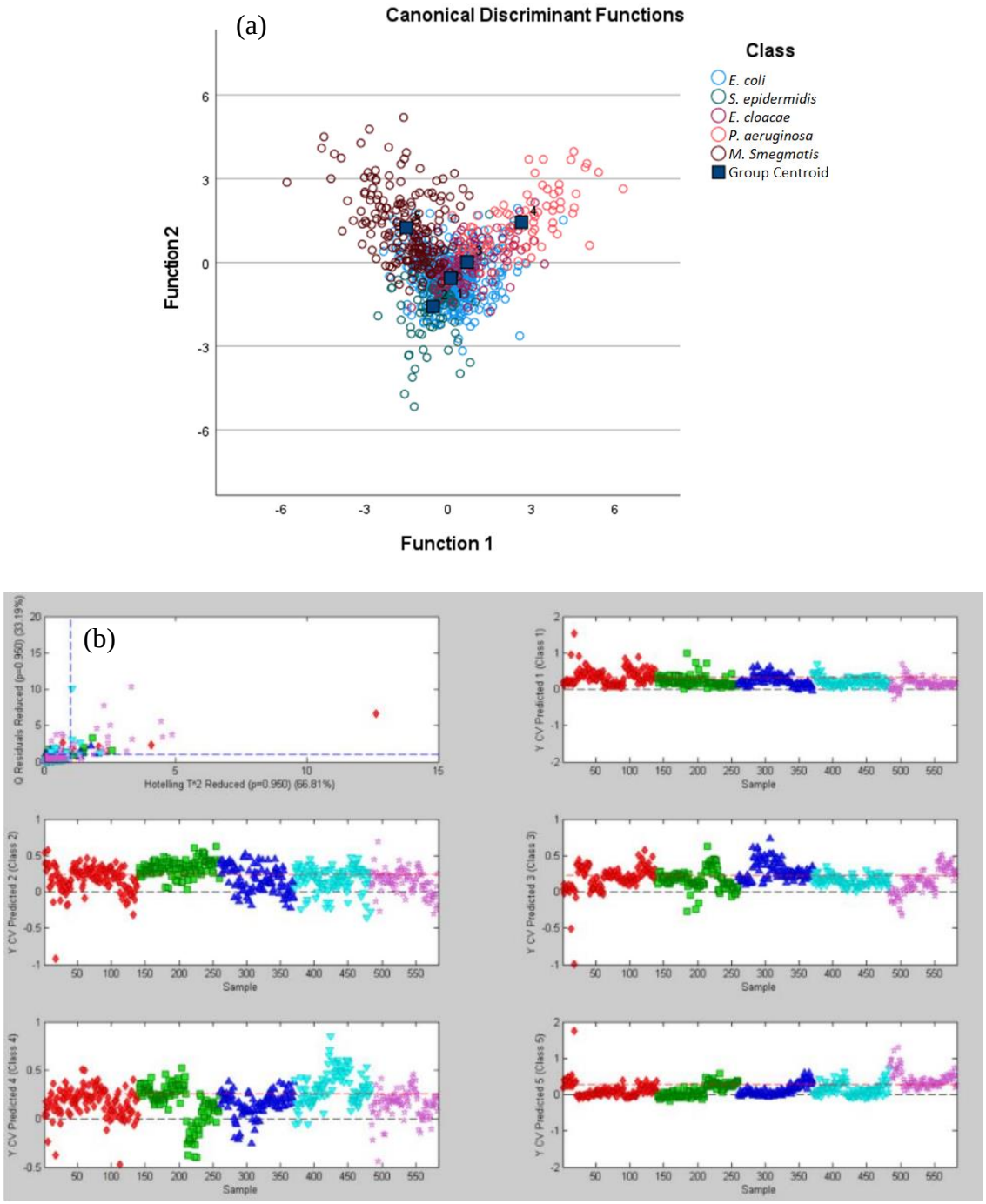

Figure 4.13: Discrimination between 5 species showing poor separation. (a) DFA discrimination between 5 species of bacteria, there is substantial overlap between groups. (b) PLSDA discrimination between 5 species of bacteria. E. coli *is shown in red,* S. aureus *in green,* P. aeruginosa *in dark blue,* E. cloacae *in light blue, and* M. smegmatis *in pink. In each test, each class is given a predictor score of 1 and the rest are given a score of 0. The Bayesian line defines the threshold between the specific bacteria and all other bacteria. There is little to no separation between groups.*

Outlier rejection was tested on both 'raw' intensity data and normalized intensity data for various concentrations of *E. coli* and *M. smegmatis*. Raw intensity data refers to data

that has not been normalized in any way and represents the brightness of the observed emission line. For both raw and normalized intensities, the total spectral intensity for each spectrum were compared to the water average plus one standard deviation. If a spectrum had lower raw or normalized intensity, it was removed from the data set. All data that passed the threshold test were included in a PLSDA analysis. All PLSDA analysis was performed using the MATLAB-based PLS_Toolbox (v.8.6.2 Eigenvector, Inc.). A discrimination between all concentrations of *E. coli* and *M. smegmatis* and between only 1/5 concentrations without exclusion of data was performed first to establish a baseline performance, this is shown in Figure 4.14a and 4.14b. In the PLSDA analyses shown in Figures 4.14 though 4.18, *E. coli* data is shown in red and *M. smegmatis* data is shown in green. Unclassified data to be tested is shown at the right as gray points. PLSDA visually differentiates between classes by the Bayesian cut line, shown by the red dashed line in Figure 4.14. Data was then excluded from these sets based on the predefined threshold and analyzed in PLSDA, this is shown in Figure 4.14c and 4.14d. A summary of the discrimination results are given in Table 4.1. For the discrimination between 1/5 dilutions, only 10 data points were excluded, 4 from *E. coli* and 6 from *M. smegmatis*. The discrimination with the exclusion of data performed well, but the exclusion caused the sensitivity to drop from 0.975 to 0.944. The specificities remained the same, at 1.000. Addition of the water threshold for data exclusion showed no marked improvement on the most concentrated bacterial dilutions.

For discrimination between all concentrations of bacteria, the initial sensitivity and specificity was 0.855 and 0.772, respectively. The data excluded based on the water threshold resulted in 58 spectra of *E. coli* being removed, and 161 spectra of *M. smegmatis* being removed. This resulted in an extremely unbalanced weighting between the two species, which is visually represented by Figure 4.14d. This unbalance reduced the sensitivity from 0.855 to 0.678. The specificity improved by a small amount, but the improvement was not large enough to warrant a change in the way we process our data. This imbalance of data could also be expected to occur with other species that often produced low intensity data, such as *P. aeruginosa*. Therefore, it was concluded that establishing an external threshold for outlier rejection created an unbalanced data set with

no increase in discrimination capability. It was decided based on this poor performance that no external validation of this model would be pursued. On the basis of this conclusion, an outlier rejection scheme that was internal to the species or the data set was investigated next.

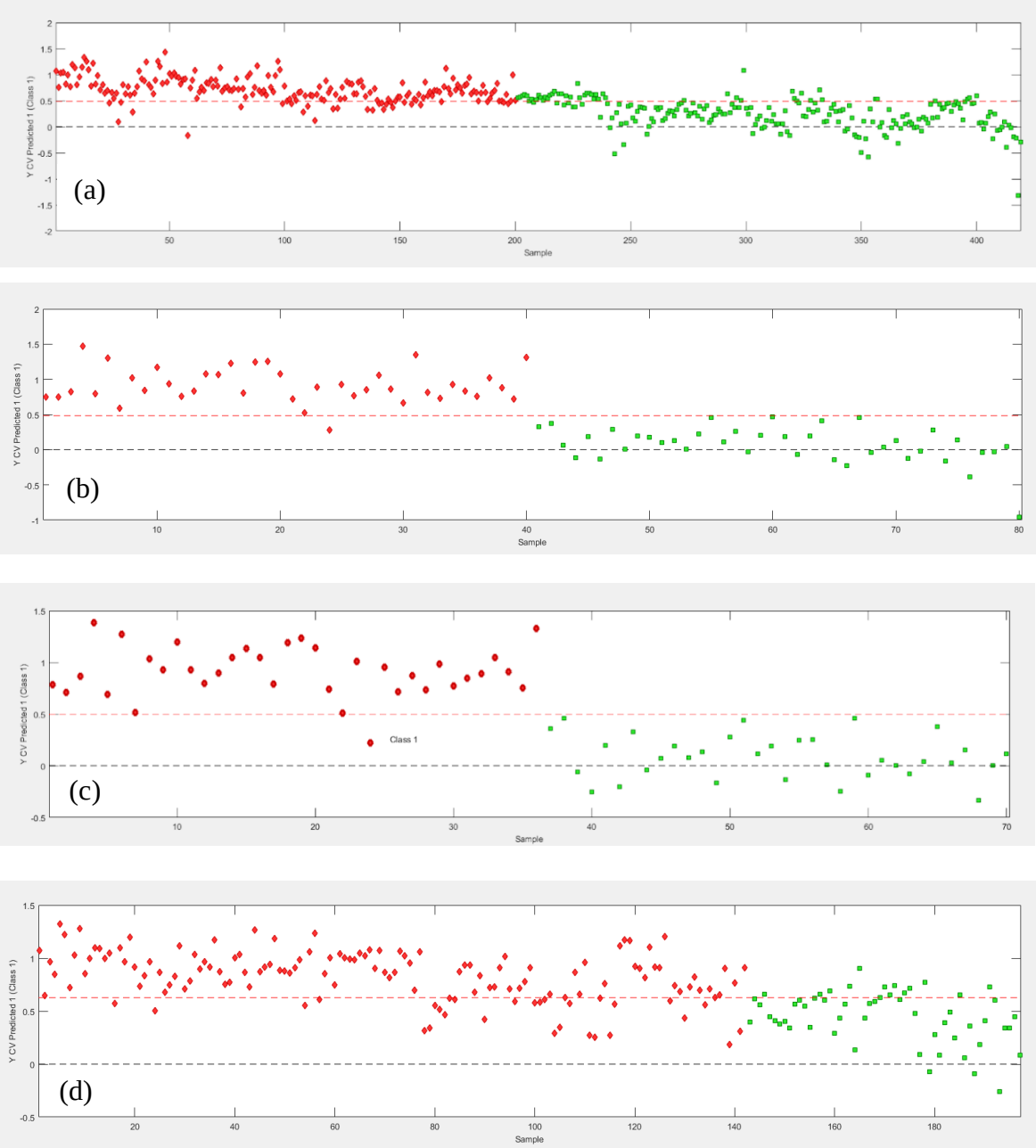

Figure 4.14: PLSDA discrimination for the water threshold analysis. (a) A discrimination without the exclusion of any data between E. coli *and* M. smegmatis*. (b) A discrimination between 1/5 concentrations of* E. coli *and* M. smegmatis *without the exclusion of data. (c) Discrimination after the exclusion of data between 1/5 concentrations of* E. coli *and* M. smegmatis*. (d) Discrimination after the exclusion of data between all concentrations of* E. coli *and* M. smegmatis.

Table 4.1: Summary of results from the outlier rejection study.

	Discrimination	Total Number of Data Files Included	Sensitivity (CV)	Specificity (CV)
Without Exclusion of Data	All E. coli vs All Myco	419	0.855	0.772
	E. coli 1/5 vs Myco 1/5	80	0.975	1.000
With Exclusion of Data	All E. coli vs All Myco	200	0.678	0.790
	E. coli 1/5 vs Myco 1/5	70	0.944	1.000

To establish an outlier rejection scheme internal to the individual filter data sets, histograms were made for each filter data set. Histograms were constructed using OriginPro 8 software. The number of bins for each histogram was automatically chosen by the software. Histograms of *E. coli* and *M. smegmatis* are shown in Figure 4.15.

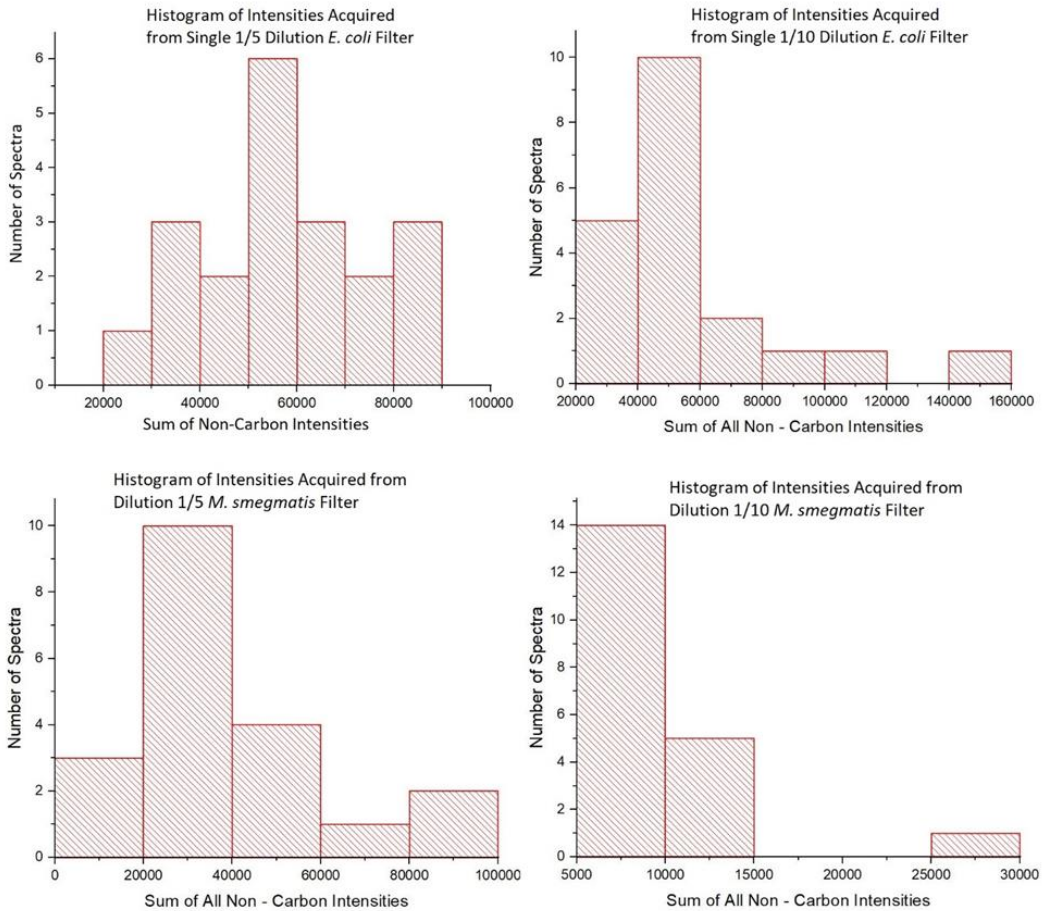

Figure 4.15: Histograms of (a) E. coli *1/5, (b)* E. coli *1/10, (c)* M. smegmatis *1/5, and (d)* M. smegmatis *1/10.*

The histograms were analyzed to establish any trends that would suggest the best way to remove data. It was found that higher concentrations of bacteria, namely 1/5 concentrations, exhibited shapes closer to a Gaussian curve. Lower concentrations showed more left-shifted distributions, often times with the largest number of data points being in the lowest bin. From these figures, it is clear that shots with lower intensity happen more frequently than those with larger intensity, however the larger intensity spectra are important to our analysis as they contain the most signal. All the spectra in the bin containing the weakest intensities were taken to represent the 'empty shots' as seen in Figure 4.1. These empty shots were removed from the library and a discrimination between 1/5 dilutions of *E. coli* and *M. smegmatis* was performed in PLSDA. These dilutions were chosen because 1/5 has the best quality data with the highest intensity spectra and therefore classifies more accurately than other dilutions. 1/5 dilution was also chosen because removing the bin containing the weakest intensities posed a problem for any dilutions of *M. smegmatis* lower than 1/5; for smaller dilutions, most of the spectra fell in the bin containing the weakest intensities. Using any dilution of *M. smegmatis* smaller than 1/5 would have resulted in an imbalance in the data. The results of the discrimination between 1/5 dilutions with and without data exclusion are shown in Figure 4.16 and the results of the sensitivity and specificity in Table 4.2. The exclusion of data improved the classification between 1/5 dilutions of *E. coli* and *M. smegmatis*.

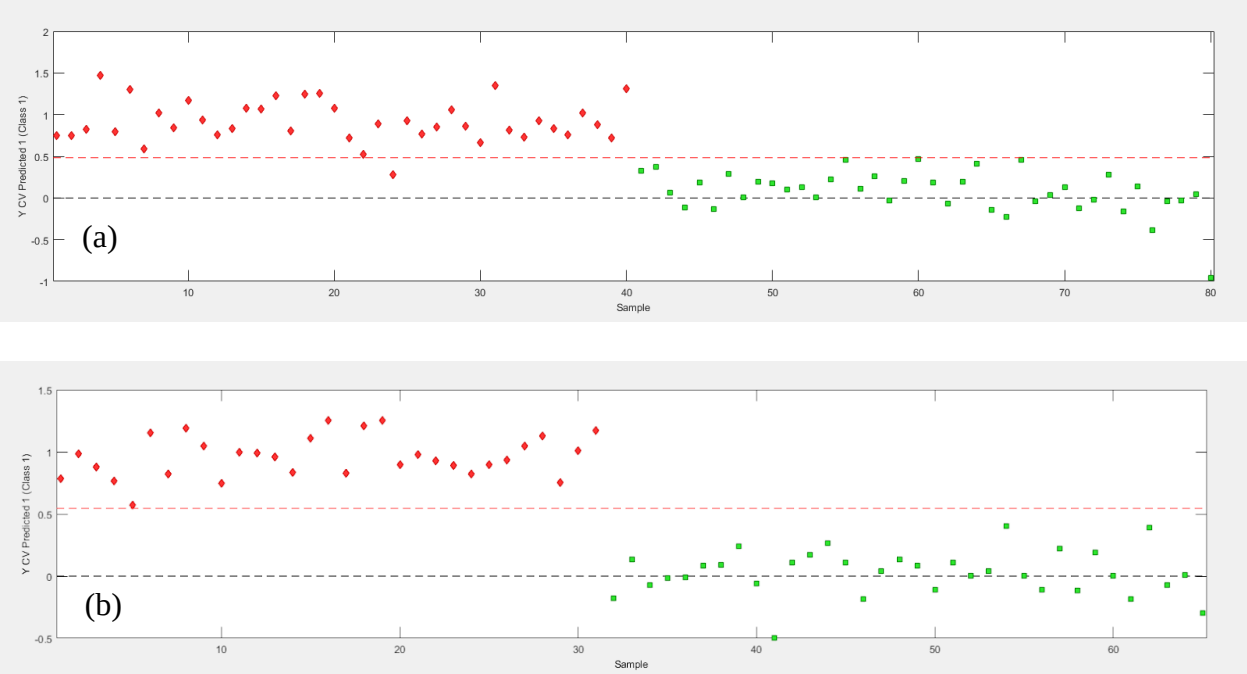

Figure 4.16: Discrimination between 1/5 dilutions of E. coli *and* M. smegmatis *(a) without and (b) with data exclusion.*

The PLSDA model constructed with 1/5 dilution and an exclusion of data was tested with several filters using external validation. First, a filter of 1/5 *M. smegmatis* was removed from the model and entered unclassified for an external validation test. As expected, 100% of the unclassified points classified as *M. smegmatis.* Next, a filter of 1/10 *E. coli* with no exclusion of data was tested in the model. The results of this test are shown in Figure 4.17a. The sensitivity of the 1/10 *E. coli* was 0.700, as only 14 of 20 points classified correctly. The specificity of this filter could not be calculated, as this is testing the rate of true positives that occur in the unclassified data set. True negatives in this set would simply be the number of *E. coli* spectra correctly misclassifying as *M.* smegmatis, which in this case is the same number of spectra that are true positives. Therefore, specificity for these external validation tests is redundant. The specificity was then compared to a PLSDA model built with 1/5 dilutions and no exclusion of data. The same *E. coli* 1/10 filter was entered unclassified into a PLSDA model with no exclusion of data and the result of this classification is shown in Figure 4.17b. 17 of 20 points classified correctly resulting in a sensitivity of 0.850. The library that had no data exclusion based on histograms performed better.

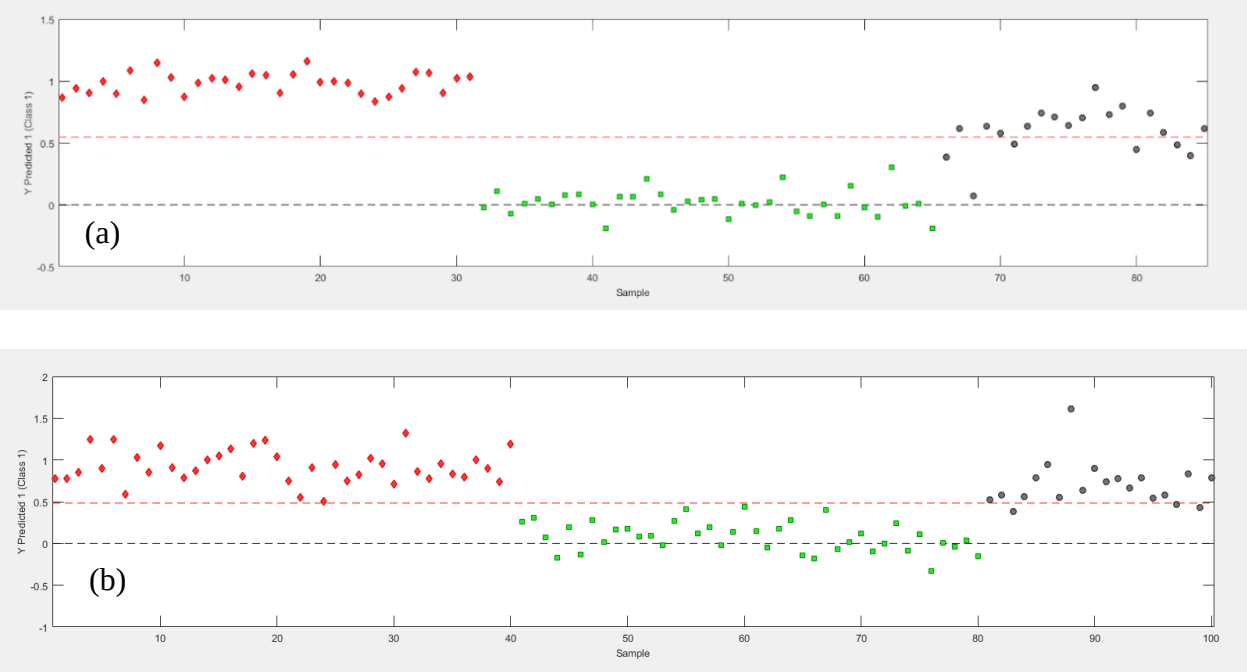

Figure 4.17: External validation of 1/10 E. coli *dilution in a model constructed from 1/5 dilutions of* E. coli *and* M. smegmatis. *(a) The external validation of* E. coli *1/10 dilution in a library with data exclusion based on histograms. (b) The external validation of* E. coli *1/10 dilution in a library where no data has been excluded.*

The same steps as above were repeated for the external validation using a filter of *M. smegmatis* 1/10 dilution. The results of this external validation are found in Figure 4.18. The performance of the discrimination in the library with data exclusion based on histograms showed 12 of 20 points correctly classified, resulting in a sensitivity of 0.600. However, the performance of the discrimination in the library that had no data exclusion was worse, with only 3 of 20 points classifying correctly resulting in a sensitivity of 0.15.

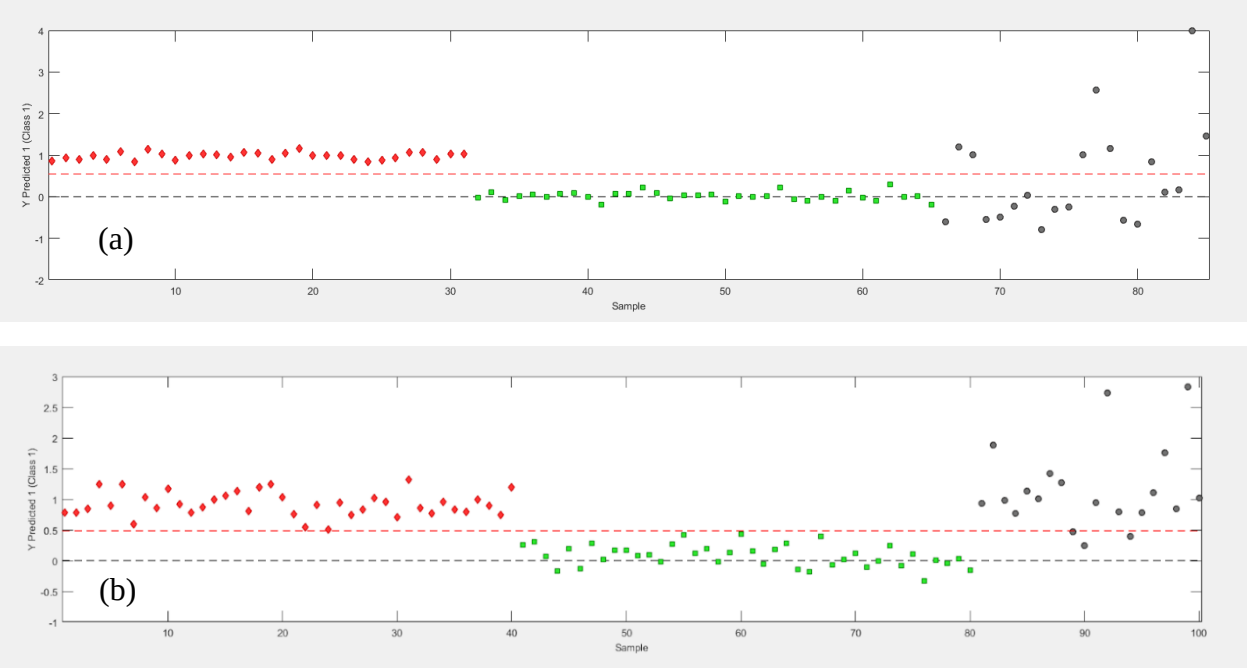

Figure 4.18: External validation of 1/10 M. smegmatis *dilution in a model constructed from 1/5 dilutions of* E. coli *and* M. smegmatis*. (a) The external validation of* M. smegmatis *1/10 dilution in a library with data exclusion based on histograms. (b) The external validation of* M. smegmatis *1/10 dilution in a library where no data has been excluded.*

Table 4.2: Comparison of sensitivity and specificity with no exclusion of data and with exclusion of data. Exclusion of data resulted in higher sensitivity and no change in specificity.

	Sensitivity (CV)	**Specificity (CV)**
Without Exclusion of Data	0.975	1.000
With Exclusion of Data	1.000	1.000

The results of these classifications show that outlier rejection based on histograms is not beneficial in all cases. It worsened the discrimination when testing with *E. coli* 1/10 but improved the discrimination when testing with *M. smegmatis* 1/10. As well, outlier rejection based on histograms also poses a data imbalance problem when imposed on species that have consistently low intensities. Finally, the time it takes to make each histogram and manually remove the data does not constitute the minute gain seen for some discriminations. For these reasons, outlier rejection of each filter based on histograms was rejected as a viable method to improve discrimination and decrease scatter.

Other methods to decrease this scatter have been addressed by our group in the past. One of these methods, pursued by Sydney Sleiman, was to use tween 20 on the bacterial suspensions. Tween 20 is a non-ionic detergent which is used to separate bacterial cells

from one another in clumps. Tween 20 is an amphipathic substance, containing a hydrophobic carbon tail and hydrophilic head group. Depending on the head group, the detergent can be anionic, cationic, non-ionic, or zwitterionic.[94] Tween 20 was investigated with the goal of making the bacteria more evenly distributed in solution, and thereby making the surface of the bacterial film on filter more even and the shot-to-shot reproducibility of bacteria higher. Tween 20 was added to *S. epidermidis* samples during sample preparation. Filters of *S. epidermidis* with no Tween 20 were used as a comparison. Each sample was shot using the LIBS setup and subsequently analyzed with a scanning electron microscope (SEM) operated by Sharon Lackie at the Great Lakes Institute for Environmental Research (GLIER). The SEM images of *S. epidermidis* with Tween 20 and without Tween 20 are shown in Figure 4.19. Included in this figure are a variation of back-scattered electron images and secondary electron images. These types of images were chosen because they show the topography of the sample best.

Analyzing these figures shows a more uniform deposition for the surfaces with Tween. At the 300.0 and 10.0 μm scale, the surfaces with tween are much flatter and do not posses cracks in the surface as the samples with no Tween do. At the 5.0 μm scale, it is clear that the Tween sample is flatter than the no Tween sample. However, examining this image shows that bubbles and holes in the film are still present and it is not completely uniform. These bubbles and holes are less apparent in the 300.0 and 10.0 μm images, nonetheless they are present when the images are examined closely. The data for the Tween experiments showed that no improved reproducibility was achieved by applying Tween 20, likely due to the persistent presence of bubbles in the surface of the film. As well, similar intensities were obtained for spectra regardless of the presence of tween. However, the application of Tween to reduce variability is still an area for possible future work.

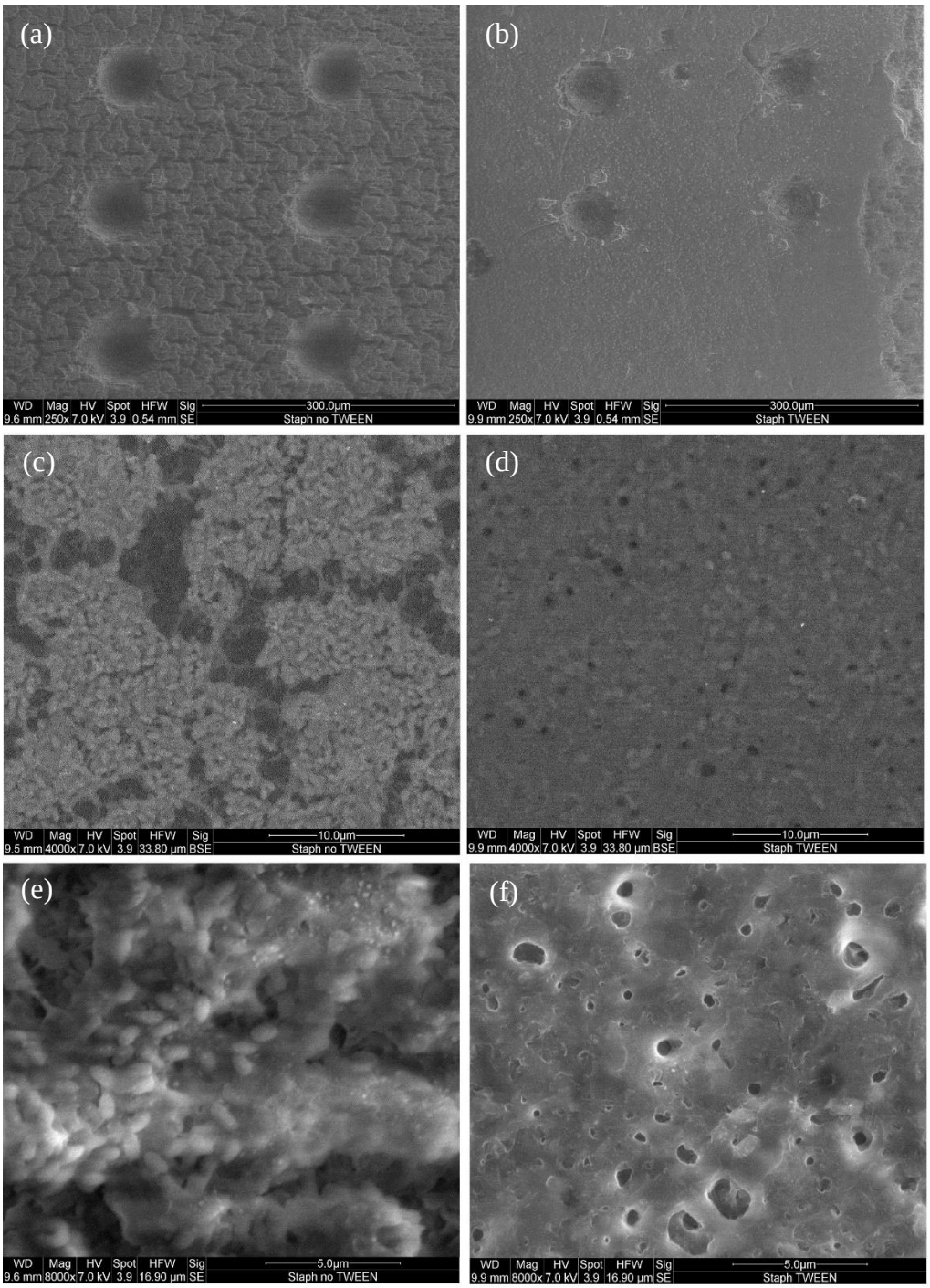

Figure 4.19: SEM images of S. epidermidis with *(a) no Tween at 300.0 µm, (b) Tween at 300.0 µm, (c) no Tween at 10.0 µm, (d) Tween at 10.0 µm, (e) no Tween at 5.0 µm, and (f) Tween at 5.0 µm. The deposition of* S. epidermidis *on the filters with Tween is much smoother and more uniform than filters with no Tween, however some bubbles still exist in the bacterial film, evidenced in (f).*

4.4 Detection of *E. coli* From Sterile Deionized Water

Using the previous results on deionized water, a study was conducted to see if *E. coli* could reliably and accurately be detected against deionized water. To do this, a MATLAB program called PLSDA was used to perform all discriminations between *E. coli* data at all dilutions and all deionized water data. 14 *E. coli* filters and 7 water filters, comprising 320 data shots and 139 data shots, respectively, were analyzed. Using 5 latent variables and 10-split venetian blinds cross-validation, a PLSDA model of single-shot *E. coli* and water was constructed with a sensitivity and specificity of 0.93 and 0.95, respectively. Next, external validation was conducted by removing one filter at a time and entering these filters back into the model unclassified. This resulted in 21 externally validated tests of the model. The sensitivity of each filter of *E. coli* was calculated by counting the number of spectra that correctly classified as *E. coli*. The specificity of each filter of water was calculated by counting the number of spectra that classified as water as opposed to *E. coli*, or the true negative rate. The average sensitivity of the single-shots on the 14 *E. coli* filters was 0.87, and the average specificity of the single-shots on the 7 water filters was 0.72. For *E. coli*, 7 filters had a true positive rate of 100%, and one filter had a true positive rate of only 33%. For water, 2 filters had 100% shots correctly classify as not *E. coli*, while the worst performing filter has only 35% correctly classify as not *E. coli*. An example of one of these external validations is shown in Figure 4.20. The red data points are the deionized water; the green points are the *E. coli* spectra. The grey points are the filter that was entered unclassified into the model. In this case, 100% of these points classified correctly.

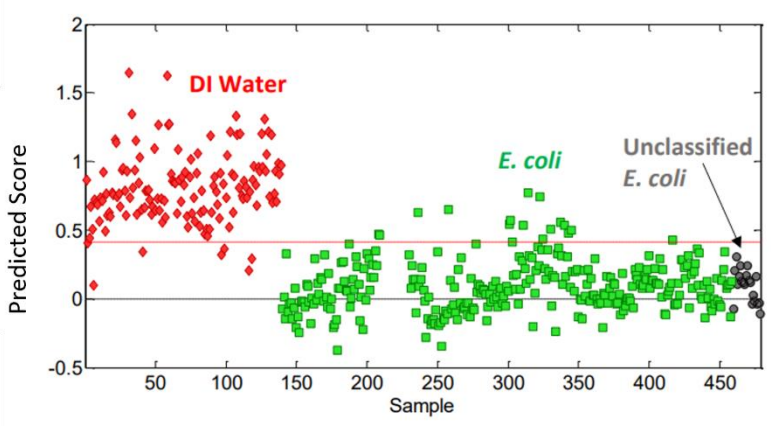

Figure 4.20: Example of an external validation for E. coli *and water. Adapted from ref [85].*

Though some filters resulted in high classification accuracy, others exhibited extremely low classification accuracy. This can be attributed to the high variability in the spectra as seen in Figure 4.1. To compensate for this, it was decided that the spectra from each filter would be summed to create one measurement per filter. This approach was motivated by the high variability and by the fact that clinically, only one diagnosis is required instead of 20 to 30. To accomplish this, two methods were investigated for summing the spectra. The first was using the "add-all" feature in Esawin, the second method was using Excel to sum individual line intensities.

The add-all function in Esawin functions as an average for spectra. In Esawin, the raw pixel intensities are added up for each individual spectrum to produce one spectrum. For each filter this is an average of 20-30 spectra. The effect of the add-all spectrum is to increase the signal to noise ratio so that smaller spectral features become more prominent, this is shown in Figure 4.21. In this figure it is apparent that the magnesium line in the add-all spectrum is bigger than the magnesium line in the single-shot spectrum. 21 add-all spectra were created for each of the 21 filters and an external validation was performed for each filter. The results for this test showed that 13 of the 14 *E. coli* filters correctly classified as *E. coli,* and 6 of the 7 water filters correctly classified as water. The sensitivity and specificity were therefore 0.93 and 0.86, respectively. Figure 4.22 shows a sample of one of these external validations.

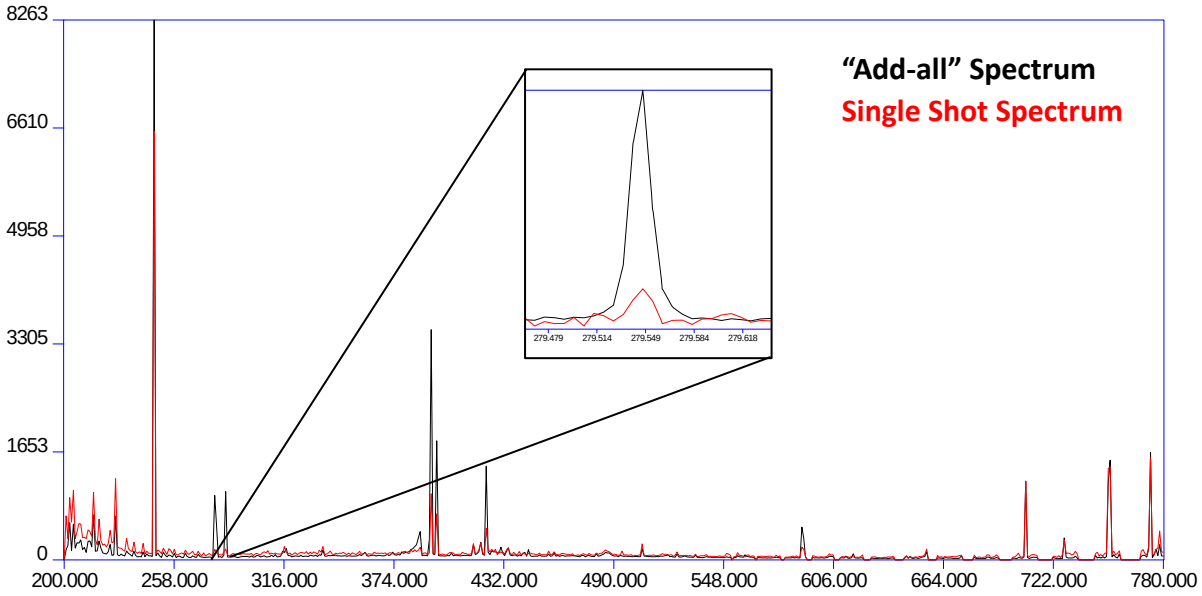

Figure 4.21: Esawin "add-all" spectrum (black) overlayed on a single-shot spectrum (red). The add-all spectrum reduces the noise in the spectrum and increases intensity of spectral features important to bacterial discrimination. This increased signal is shown by the close up of the magnesium 279 nm line where the add-all is higher in intensity than the single-shot spectrum.

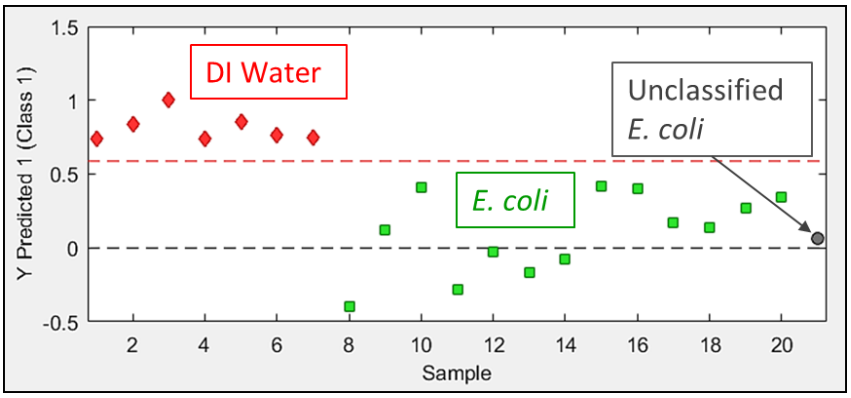

Figure 4.22: A PLSDA plot showing the classification of add-all spectra. The sterile DI Water is shown in red, the E. coli *is shown in green. The grey point is an add-all filter of* E. coli *entered into the PLSDA plot without any class information.*

With the Excel method, the 19 emission lines and their respective intensities for the 20-30 spectra per filter were first extracted from the Esawin software. The peak intensities were summed across the 20-30 spectra for each line, then the summed lines were normalized to the total spectral power of that summed spectrum, as usual. The complex ratios in RM3 were then calculated using the summed spectrum. The summed spectrum produced by Excel is similar to the Esawin "add-all" function, but not exactly the same. The difference observed between these two methods warranted both to be tested in PLSDA. The PLSDA results for the spectra summed in Excel showed 14 of 14 filters correctly

classified as *E. coli* and 7 of 7 water filters correctly classified as water. The sensitivity and specificity of this technique is therefore 1.000 and 1.000, respectively. The performance of the Excel summation is better than the Esawin add-all performance, leading us to conclude that any preprocessing step involving summation of shots across a filter should be done through Excel as opposed to Esawin.[85]

Though this method has shown improvement on classification, there is a problem with the small amounts of data being inputted into the algorithms. Ideally, the amount of spectra should be equal to the number of independent variables for the best results. We are attempting to address this currently and future work will focus on collecting enough data to meet this benchmark.

A more recent study inspired by the work on ultrapure water and the reduction of background signal focused on the same problem of discriminating between bacteria and water. For this study, a PLSDA model was constructed from 320 single-shot *E. coli* spectra and 123 single-shot ultrapure water spectra. The *E. coli* data included spanned all dilutions from 1/5 to 1/500 and resulted in a total of 14 filters of *E. coli.* The total number of water filters was 4; this is smaller than the previous test involving deionized water because we have been working with ultrapure water for less time and therefore have a smaller library. The model is shown in Figure 4.22 and has a sensitivity and specificity of 1.000 and 1.000, respectively. This is an improvement over the previous model, with perfect sensitivity and specificity.

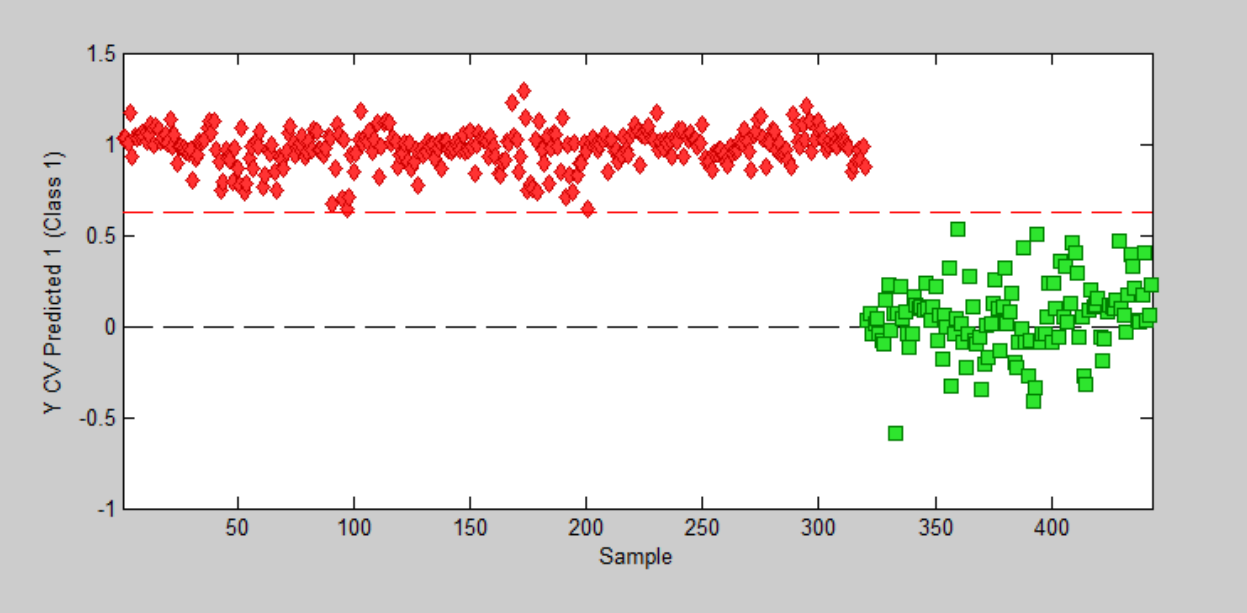

Figure 4.23: PLSDA model of all dilutions of E. coli *and ultrapure water. The sensitivity and specificity of the model is 1.000 and 1.000.*

External validation was also performed on this model using the same method as the previous test with deionized water. One filter of *E. coli* was removed from the model at a time and entered unclassified, the sensitivity of each fitler was calculated. The same was done for each ultrapure water filter, then the specificity of each filter was calculated. A sample of some of the results is shown in Figure 4.23 and Figure 4.24. 13 of the 14 *E. coli* filters had a sensitivity of 100%, with 1 of the filters having a sensitivity of 96.7%. This gives an average sensitivity of 99.7%.

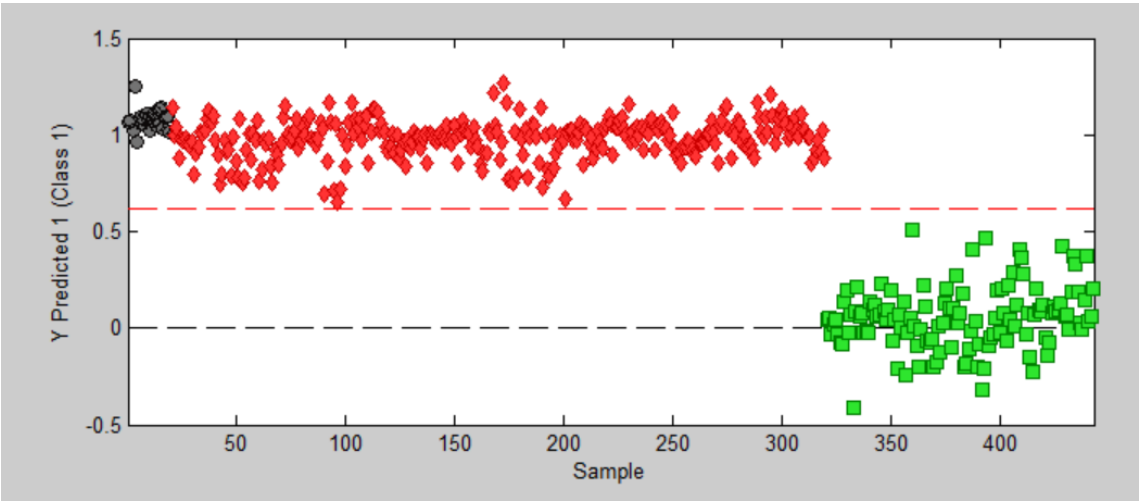

Figure 4.24: Discrimination between all dilutions of E. coli *(red) and ultrapure water (green). One* E. coli *1/5 dilution filter was entered unclassified into the model, which is shown in dark grey.*

4 of the 4 water filters classified with 100% specificity, an example of one of the external validations of water is given in Figure 4.25. In some cases, the sensitivity of the model was less than 100%, but was never lower than 99.4%. The average specificity of the ultrapure water is therefore 100%.

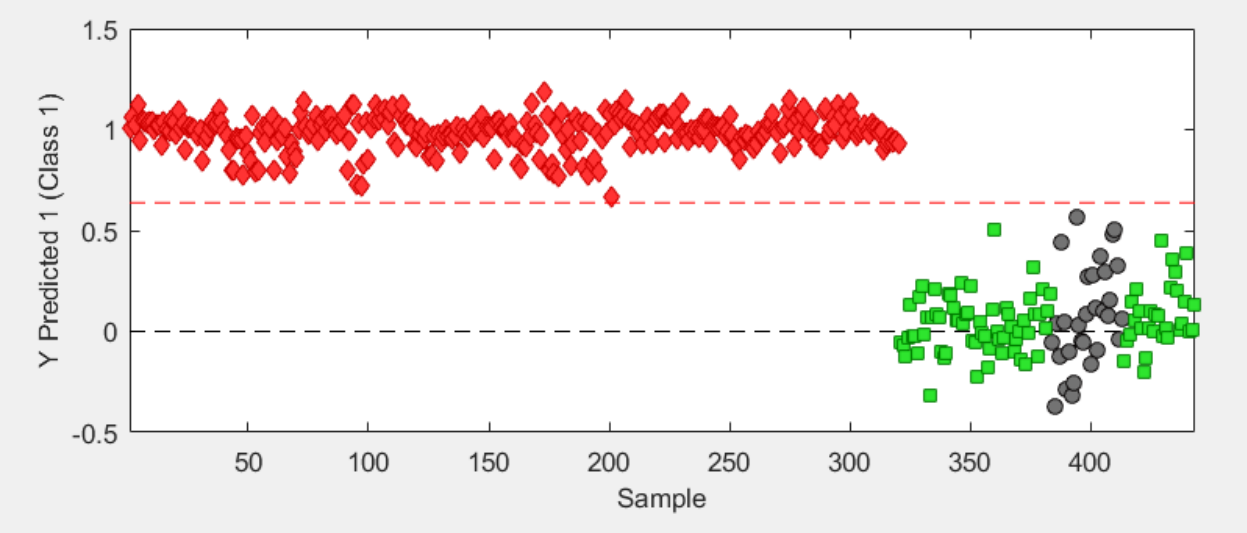

Figure 4.25: Discrimination between all dilutions of E. coli (red) and ultrapure water (green). One ultrapure water filter was entered unclassified into the model, which is shown in dark grey.

Initial studies were performed on classification of *E. coli* and ultrapure water add-all spectra. A total of 21 *E. coli* filters and 4 ultrapure water filters were used in the classification. More *E. coli* filters were used for this study because more data had been collected over time. A PLSDA model was built using these add-all filters, shown in Figure 4.26, and resulted in a sensitivity and specificity of 100% and 100%, respectively. External validation was performed on each filter in the same way as the previous studies. The results showed that 21 of 21 *E. coli* filters classified correctly, and 4 of 4 ultrapure water filters classified correctly. The average sensitivity and specificity of this technique is therefore 100%. These initial studies however may not be entirely representative of our ability to detect add-all spectra of bacteria because the data pool is small and imbalanced. The imbalance of the data is clearly seen in Figure 4.26, where the green *E. coli* spectra heavily outweigh the red ultrapure water spectra. More ultrapure water data is recommended to determine if this high accuracy of classification is true.

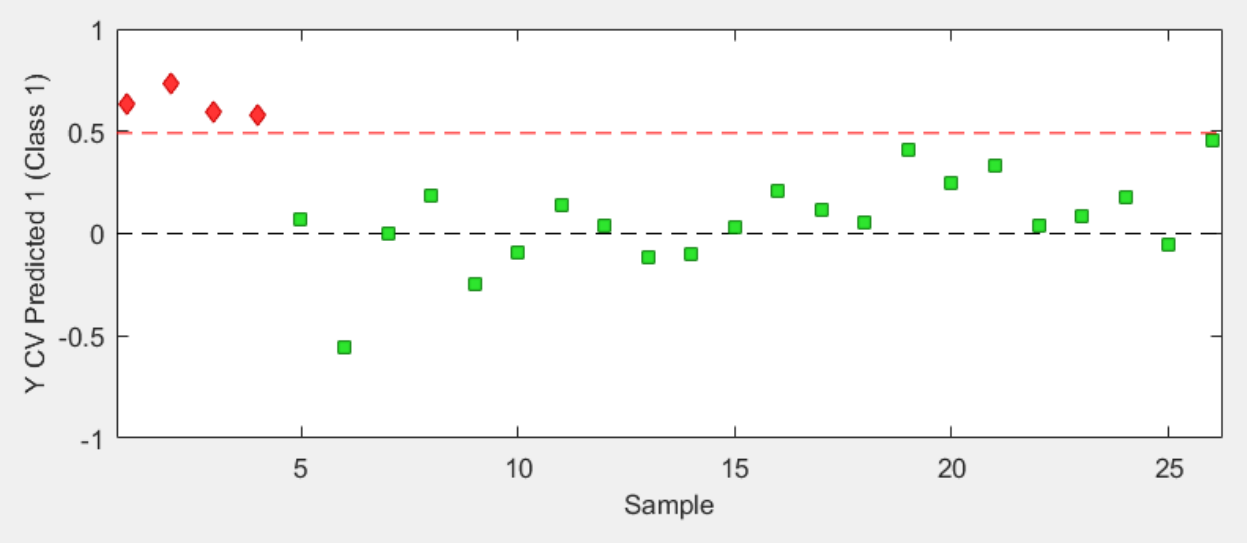

Figure 4.26: PLSDA classification between add-all ultrapure water spectra (red) and add-all E. coli *spectra of all dilutions (green). The sensitivity and specificity of the model is 100%, but the model is unbalanced due to the lack of ultrapure water data.*

These results for both the sensitivity and specificity using ultrapure water are an improvement over the results using deionized water. The single-shot classification performed as well as the classification with averaged spectra using Esawin "add-all", and slightly worse than the classification using Excel averaged spectra. However, both the method of averaging spectra and using ultrapure water allows us to detect the presence of bacteria with 100% sensitivity and specificity. As well initial results on the use of add-all ultrapure water spectra show the same trend of detection with 100% accuracy. The use of averaging spectra and the use of ultrapure water will therefore be used for all future studies involving the preparation and detection of bacteria.

This chapter detailed my attempts to reduce the scatter in the data and reduce the background intensity of spectra with no bacteria. It was concluded that washing of the cone with acetone followed by methanol reduced the background intensity of the spectra. The effect of the water was investigated, which showed that DI and distilled water resulted in about the same intensity of spectra, while ultrapure megohmic water showed reduced intensity of spectra. The reduction of line intensity in water improved detection of bacteria. Multiple substrates were investigated and it was found that nitrocellulose was the best substrate for our purposes. Other filters were investigated because of the high intensity carbon line of the nitrocellulose filter, which prohibited us from turning up the

amplification on the spectrometer to increase the intensity of other lines important to classification. However, these other filters showed both a high carbon line and increased intensities for magnesium, calcium, and sodium. Therefore, it was concluded that these filters were unideal for deposition with bacteria and nitrocellulose was continued to be used. Outlier rejection was investigated and showed no improvement. The addition of single-shot spectra to average out the low intensity spectra was investigated and was found to greatly improve the classification accuracy in PLSDA for detection of bacteria in sterile water. Originally, single-shot spectra classified with a sensitivity of 87 % and a specificity of 72 %. Addition of single-shot spectra improved the sensitivity and specificity to 100 % each.

[85] Blanchette, E. J., Sleiman, S. C., Arain, H., Tieu, A., Clement, C. L., Howson, G. C., Tracey, E. A., Malik, H., Marvin, J. C., Rehse, S. J. (2022). Detection and classification of bacterial cells after centrifugation and filtration of liquid specimens using laser-induced breakdown spectroscopy. *Applied Spectroscopy,* 76(8), 894- 904. https://doi.org/10.1177/00037028221092789.

[86] Paulick, A. (2018). *Development of Laser-Induced Breakdown Spectroscopy as a Rapid Diagnostic Tool for Bacterial Infection.* [Master's thesis, University of Windsor].

[87] Aragon, C., Penalba, F., Aguilera, J. A. (2005). Curves of growth of neutral atom and ion lines emitted by a laser-induced plasma, *Spectrochimica Acta Part B: Atomic Spectroscopy, 60*(7-8), 879-887. https://doi.org/10.1016/j.sab.2005.05.015.

[88] Pořízka, P., Klus, J., Prochazka, D., Képeš, E., Hrdlička, A., Novotný, J., Novotný, K., Kaiser, J. (2016). Laser-induced breakdown spectroscopy coupled with chemometrics for the analysis of steel: The issue of spectral outliers filtering, *Spectrochimica Acta Part B: Atomic Spectroscopy*, *123*, 114-120. https://doi.org/10.1016/j.sab.2016.08.008.

[89] Yaroshchyk, P., Death, D.L., Spencer, S.J. (2011). Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS, *Journal of Analytical Atomic Spectroscopy, 27,* 92-98. https://doi.org/10.1039/C1JA10164A.

[90] El Haddad, J., Canioni, L., Bousquet, B. (2014). Good practices in LIBS analysis: Review and advices, *Spectrochimica Acta Part B: Atomic Spectroscopy, 101,* 171-182. https://doi.org/10.1016/j.sab.2014.08.039.

[91] Sahoo, T. K., Negi, A., & Gundawar, M. K. (2015). Study of preprocessing sensitivity on laser induced breakdown spectroscopy (LIBS) spectral classification, *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 137-143. DOI: 10.1109/ICACCI.2015.7275598.

[92] Cisewski, J., Snyder, E., Hannig, J., Oudejans, L. (2012). Support vector machine classification of suspect powders using laser-induced breakdown spectroscopy (LIBS) spectral data, *Journal of Chemometrics*, *26*(5), 143-149. https://doi.org/10.1002/cem.2422.

[93] Yueh, F. Y., Zheng, H., Singh, J. P., Burgess, S. (2009). Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *64*(10), 1059-1067. https://doi.org/10.1016/j.sab.2009.07.025.

[94] Weiszhár, Z., Czúcz, J., Révész, C., Rosivall, L., Szebeni, J., Rozsnyay, Z. (2012). Complement activation by polyethoxylated pharmaceutical surfactants: Cremophor-EL, Tween-80 and

Tween-20, *European Journal of Pharmaceutical Sciences, 45*(4), 492-498. https://doi.org/10.1016/j.ejps.2011.09.016.

Chapter 5: Classification of Bacteria by Species – Diagnosing Bacterial Infections

The accuracy of a DFA test performed on 5 species of bacteria has previously been very high; sensitivities and specificities as high as 91% and 97% have been reported by Rehse et al.[95] However, these high numbers can be attributed to ablation of high numbers of cells on background-free nutrient-free agar. This is impractical for clinical use because many infections will not present with such high numbers of cells. As well, the nutrient-free agar background requires extra sample preparation that would add to diagnostic time. While this work demonstrated that bacterial species could be reliably and accurately discriminated, discrimination capability of spectra resulting from lower numbers of cells and with non-zero background must be investigated.

Dylan Malenfant showed that bacterial spectra could still be discriminated reliably and accurately when ablated on a filter medium. 30 µL of high concentrations of bacterial cells were deposited using a well-plate onto filter, shown in Figure 5.1, and ablated.[96] When the spectra were analyzed with DFA, there was a visible separation between groups and an average sensitivity and specificity of 98% and 99%, respectively. The DFA plot is shown in Figure 5.2.[96,97]

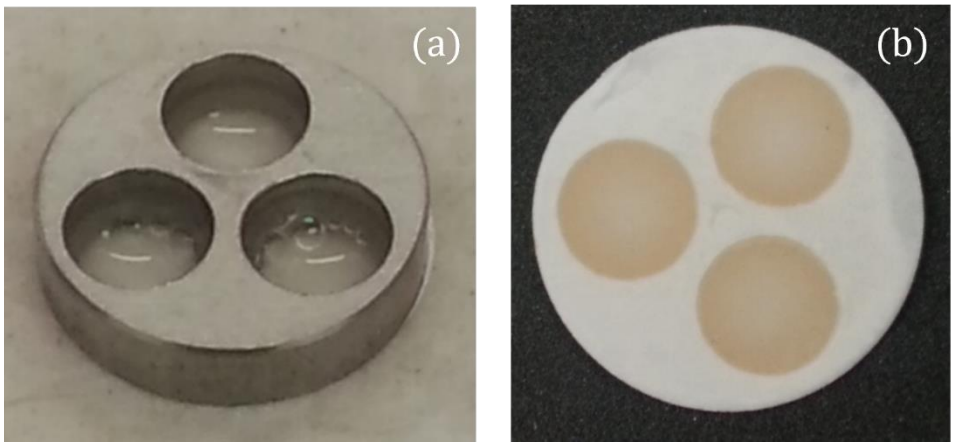

Figure 5.1: (a) Well-plate used for deposition. 30 µL of bacterial suspension is deposited into each well and left to dry. (b) After drying the well-plate is removed from the filter. Three pads of bacteria are left behind. Adapted from ref [96,97].

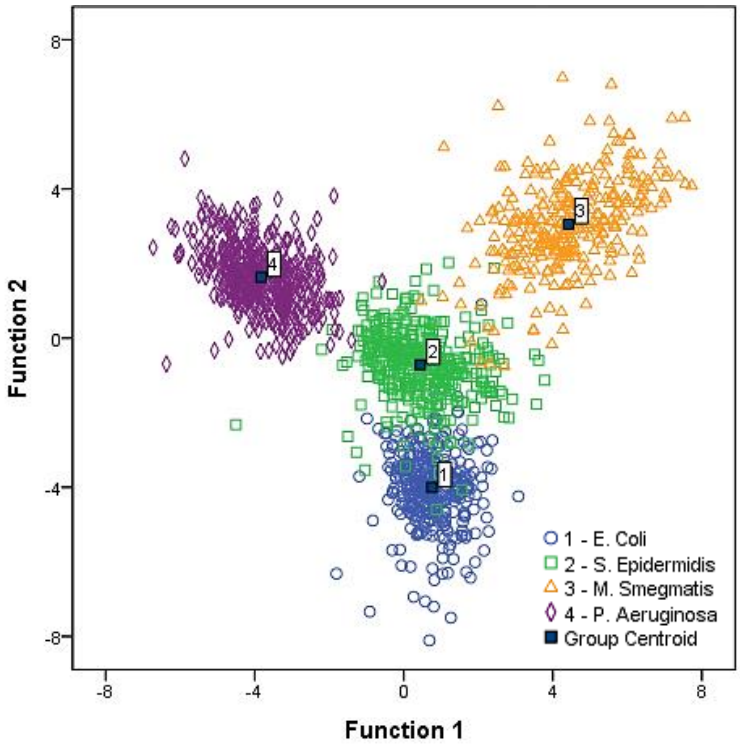

Figure 5.2: DFA plot resulting from bacterial spectra collected using the well-plate deposition method, adapted from ref [97].

The need for a more clinically relevant number of cells raises two questions: can the cells be detected, and can the cells be classified? Alexandra Paulick addressed the first question by designing a metal cone meant to concentrate small numbers of bacteria into a region with diameter 1 mm.[98] The results of this work, as addressed in Chapter 3, showed that they can be identified. We are now attempting to accurately classify fewer cells using DFA and ANN.

5.1 Classification of Bacteria Using Discriminant Function Analysis

In this work, several methods were used to optimize the classification of the DFA, including reducing the number of ratios in the ratio model, subtracting the contributions of the filter from the spectrum of bacterial, reducing the number of bacterial species used in classification, and using the 'add-all' method discussed in Chapter 4. The methodology and results of each method will be discussed below.

5.1.1 Results of the 5-Class Test

A 5-class test was attempted in DFA to test the performance of the low numbers of cells concentrated with the metal cone on the nitrocellulose filter, which created a non-zero background. 892 individual spectra were loaded into IBM SPSS Statistics using RM2.5 for a DFA analysis. The five species tested were *E. coli, S. aureus, E. cloacae, P. aeruginosa* and *M. smegmatis.* The number of spectra pertaining to each species was 400, 80, 113, 80, and 189, respectively. Few spectra were collected for some species because data acquisition of these species began later than others. The results of the DFA test show unideal sensitivity and specificity for clinical application. The sensitivities and specificities for each species are listed in Table 5.1 on page 111.

While some of the specificities are quite high, many of the sensitivities are low. As mentioned previously, an ideal medical test has a sensitivity and specificity of 100%. Several approaches were used to improve the DFA. Many of these approaches assumed that the carbon line from the nitrocellulose filter was not helpful in discrimination because it did not originate from the bacteria. First, we attempted to change the way the data was entered into DFA.

All the ratio model test used normalized data as the x-block vector. This means the data is normalized by dividing the sum of all 15 lines by the carbon line. We normalize by the carbon line because the carbon line is a consistent feature in the bacterial spectra. The carbon line in our spectra has a relative standard deviation (RSD) of approximately 5%, making it highly reproducible and reliable. Instead of using normalized data in the x-block vector, unnormalized 'raw' data was used instead. The results for the sensitivities and specificities for each species are in Table 5.2 on page 111. This model performed worse than the 5-class model using normalized data; sensitivities were far lower and classification errors were higher in most cases.

The next strategy adopted to remove the influence of the carbon line was subtracting the spectrum of the filter from bacterial spectra. To do this, an average value of each of the 15 lines was calculated for all filter spectra in our library. The average from each blank filter line was then subtracted from each single bacterial spectrum line. The average and

the new bacteria spectrum were both calculated using unnormalized data. The unnormalized result was entered in the RM2.5 sheet and used in DFA. The results were again much poorer than the original model using normalized data and are shown in Table 5.3 on page 111. The same subtraction process was repeated, and the result was normalized to determine if normalized data without the contribution from the filter would improve the discrimination. These results are shown in Table 5.4 on page 111 and were comparable to the original discrimination and the discrimination with the removal of filter contribution. Removal of the contributions from the filter therefore do not improve discrimination results between species.

Table 5.1: Results of DFA using RM2.5 with normalized data of 5 species.

RM2.5 With Normalized Data					
	E. coli	*S. aureus*	*E. cloacae*	*P. aeruginosa*	*M. smegmatis*
Sensitivity	58.8 %	68.3 %	58.8 %	48.7 %	75.0 %
Specificity	76.8 %	90.2 %	94.2 %	91.2 %	93.4 %
Classification Error	32.2 %	20.8 %	23.5 %	30.1 %	15.8 %

Table 5.2: Results of DFA using RM2.5 with unnormalized data of 5 species.

RM2.5 With Unnormalized Data					
	E. coli	*S. aureus*	*E. cloacae*	*P. aeruginosa*	*M. smegmatis*
Sensitivity	50.3 %	56.3 %	47.8 %	67.5 %	60.3 %
Specificity	81.2 %	91.2 %	91.7 %	90.8 %	84.5 %
Classification Error	34.3 %	26.3 %	30.2 %	20.9 %	27.6 %

Table 5.3: Results of DFA using RM2.5 with unnormalized data and subtracting contribution of filter from 5 species.

RM2.5 With Unnormalized Data and Subtracting Filter					
	E. coli	*S. aureus*	*E. cloacae*	*P. aeruginosa*	*M. smegmatis*
Sensitivity	60.0 %	44.4 %	38.8 %	38.1 %	61.3 %
Specificity	87.9 %	86.6 %	89.3 %	90.0 %	85.9 %
Classification Error	26.1 %	34.5 %	36.0 %	36.0 %	26.4 %

Table 5.4: Results of DFA using RM2.5 with normalized data and subtracting contribution of filter from 5 species.

RM2.5 With Normalized Data and Subtracting Filter					
	E. coli	*S. aureus*	*E. cloacae*	*P. aeruginosa*	*M. smegmatis*
Sensitivity	50.8 %	51.3 %	38.9 %	56.3 %	60.3 %
Specificity	76.2 %	87.5 %	92.8 %	91.0 %	87.7 %
Classification Error	36.5 %	30.6 %	34.1 %	26.4 %	26.0 %

Table 5.5: 10-fold CV in DFA results using RM2.5 with normalized data of 5 species.[99]

RM2.5 Using 10-Fold CV					
	E. coli	*S. aureus*	*E. cloacae*	*P. aeruginosa*	*M. smegmatis*
Sensitivity	48.7 %	57.5 %	41.5 %	50.0 %	48.1 %
Specificity	75.9 %	87.3 %	92.2 %	90.2 %	85.2 %
Classification Error	37.6 %	27.5 %	33.0 %	29.8 %	33.2 %

After the attempt to remove the carbon line did not improve results, a 10-fold cross validation (CV) was conducted in DFA. This process separates the dataset into 10 groups, 9 are used for training set and the tenth is used for validation. The test is performed 10 times until all groups are used for validation and the sensitivity and specificity are calculated. The results for this test are shown in Figure 5.3 and Table 5.4.[99] The 10-fold CV performed worse than all other tests, and the 3D plot of the discriminant analysis shows poor separation between groups.

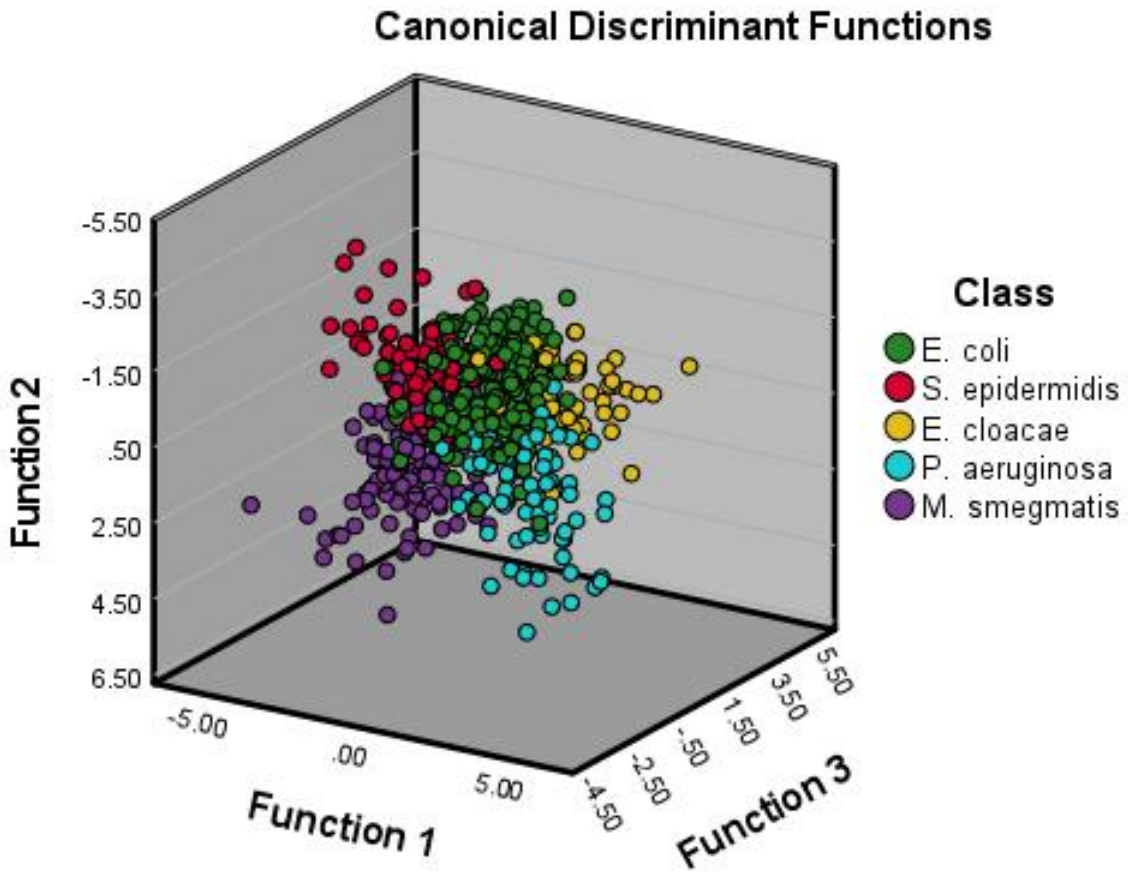

Figure 5.3: 10-fold CV in DFA on 5 species of bacteria.

Poor DFA sensitivity, specificity, and overall performance can likely be attributed to low numbers of spectra compared to the number of independent variables in the RM2.5. The RM2.5 contains 92 independent variables, and only 862 spectra were analyzed. It is recommended that for accurate statistics there should be at least 10 times the amount of data present in the model as there are independent variables. As well, classification accuracy can be degraded by imbalanced data sets. More spectra will need to be collected

to ensure that the dataset is balanced before attempting any future discrimination between 5 species.

5.1.2 Results of the 3-Class Test

A DFA discrimination was performed between only 3 classes of bacteria as an attempt to reduce the complexity of the model. It was theorized that reducing the complexity of the model may lead to improved performance. The 3 species of bacteria that were chosen for this model were *E. coli, S. aureus,* and *E. cloacae.* These species were chosen because they exhibited reliably high intensity compared to *M. smegmatis* and *P. aeruginosa*, as shown in Figure 5.4.[99] Figure 5.4 shows the same information as Figure 4.1 but includes all species. The figure demonstrates that *P. aeruginosa* (orange) had consistently low intensity spectra, most of which was consistent with sterile water. *M. smegmatis* did not have as many high intensity spectra as *E. coli*, *S. aureus,* and *E. cloacae.* Therefore, *P. aeruginosa* and *M. smegmatis* were removed from the DFA analysis.

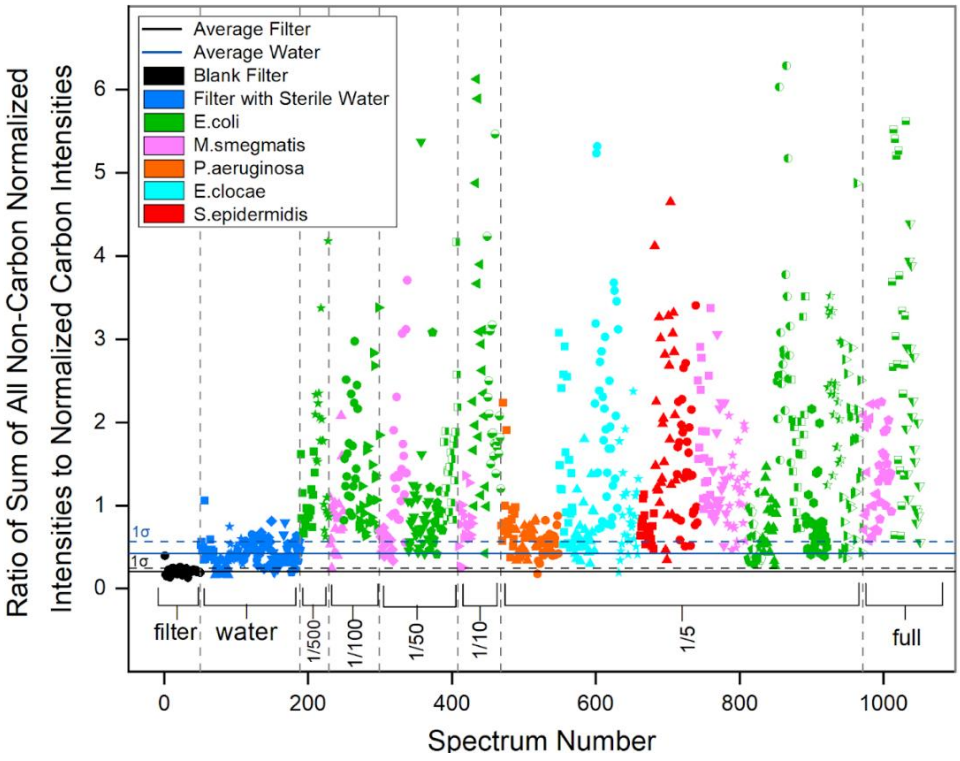

Figure 5.4: Scatter plot of individual bacterial spectral. The ratio of all non-carbon normalized intensities to the normalized carbon intensity is plotted against the spectrum number. Colours represent the different bacterial species; shapes represent individual filters of data. The average filter and water spectral intensity is given by the black and blue line, respectively. The ratios at the top represent concentrations of bacterial suspensions, with 'full' indicating stock solution. Adapted from ref [99].

An RM2.5 model was constructed using 1731 single-shot normalized spectra with 569 *E. coli* spectra, 598 *E. cloacae* spectra, and 564 *S. aureus* spectra. This represents a relatively well-balanced data set which should return the most accurate results and statistics. The RM2.5 was entered in DFA, the results of the sensitivity and specificity are given in Table 5.6 and the resulting discriminant function plot is given in Figure 5.5. The sensitivity and specificity results for the 3-class test are comparable to the previous results using 5 species, and thus are still low. The results from the 10-fold CV on 3 classes are shown in Table 5.7. The 10-fold CV only offers a very small improvement over these results, not a large enough improvement to be clinically useful.

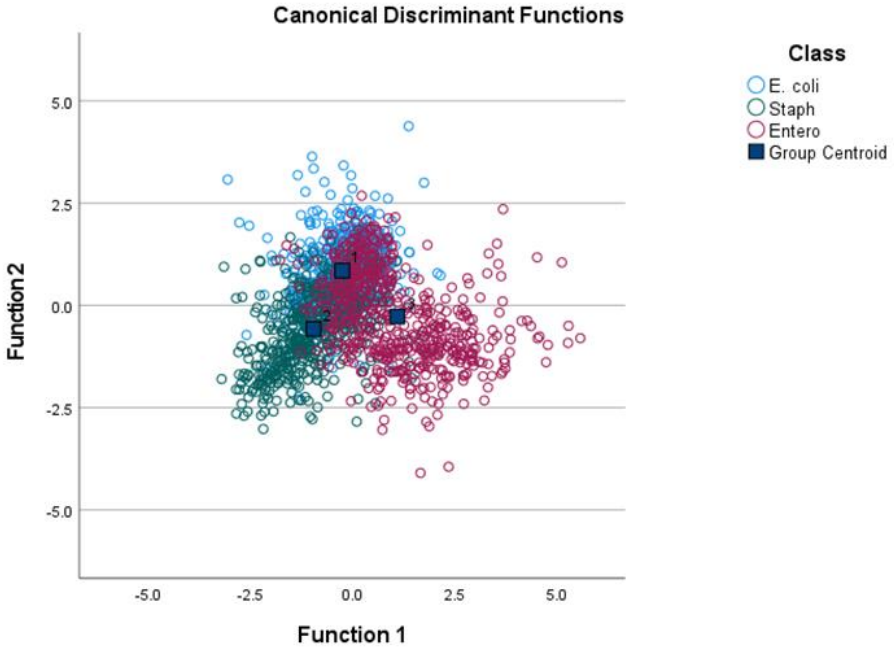

Figure 5.5: DFA of normalized data without 10-fold CV.

Table 5.6: Results of DFA using RM2.5 with normalized data of 3 species.

RM2.5 With Normalized Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	70.00%	60.00 %	60.00 %
Specificity	71.51 %	80.00 %	92.04 %
Classification Error	28.40 %	25.60 %	24.70 %

Table 5.7: Results of 10-fold CV DFA using RM2.5 with normalized data of 3 species.

RM2.5 With Normalized Data (10-Fold CV)			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	73.28 %	64.93 %	60.10 %
Specificity	71.78 %	84.93 %	92.53 %
Classification Error	27.47 %	25.07 %	23.68 %

Based on the previous recommendation from Chapter 4, the discrimination of the 'add-all' spectra were investigated. The method used to add spectra was the Excel summation method previously found to be the highest performing method in PLSDA. All 30 spectra from each filter were summed into one value for each of the 15 lines, and after summation each line was normalized to the carbon line. There were 58 filters total, thus 58 normalized data points were used in the RM2.5. The distribution of species was once again well-balanced, with 19 filters of *E. coli,* 19 filters of *S. aureus*, and 20 filters of *E. cloacae.* This discrimination performed worse than the previous 3-class test, with the results given in Table 5.8 and the discriminant function plot given in Figure 5.6.

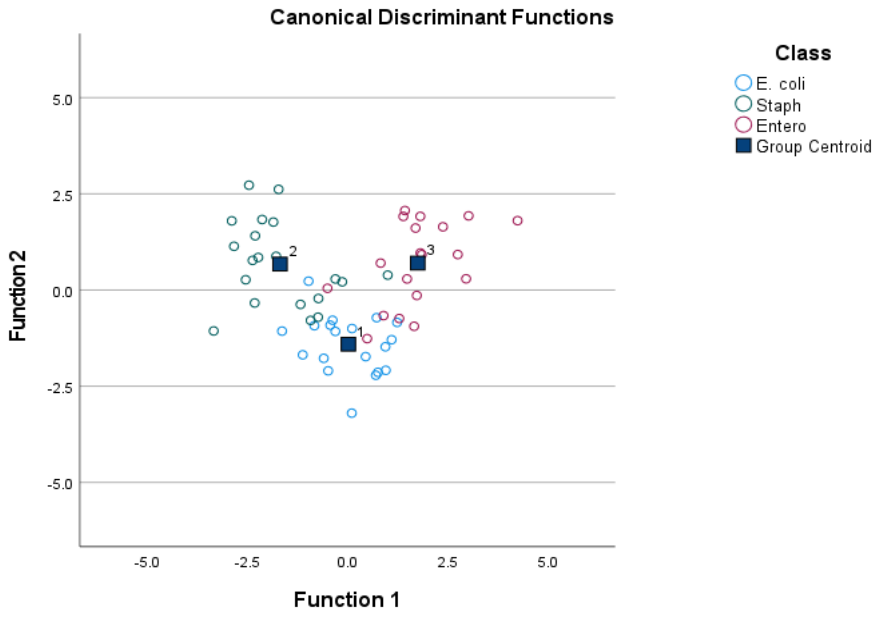

Figure 5.6: Discrimination between 'add-all' of 3 species.

Table 5.8: Results of discrimination between 3 species by summing single-shot spectra.

RM2.5 With Normalized Summed Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	50.00%	60.00 %	50.00 %
Specificity	74.35 %	70.00 %	87.18 %
Classification Error	39.10 %	35.80 %	30.10 %

This result was counterintuitive based on previous results in the PLSDA, but it may be explained by examining the structure matrix, the tolerance test, and the tests of equality of group means outputted by the DFA. The structure matrix shows how each predictor variable is correlated with the discriminant functions. The higher the number, the more highly correlated it is to a specific discriminant function; this is how the DFA decides which predictor variable to use in each function. The tolerance test determines which variables can remain in the regression equation. The test of equality of group means determines which independent variables will contribute to the model. After comparing these three parameters between the single-shot discrimination to the add-all discrimination, more independent variables were rejected from use in the add-all model than in the single-shot model. In the single-shot model, 49.53% of variables were rejected from the model, while in the add-all, 81.30% of variables were rejected from being used in the model. Another discrimination was attempted by removing the variables that were unused in the original add-all discrimination, however there was no improvement in the result. This suggested that the data was too poor to be discriminated in DFA, limiting the algorithm's performance.

The poor performance of DFA led us to develop and investigate a new classification algorithm, an artificial neural network (ANN). A general description of how ANN works and the parameters used in this experiment are covered in greater detail in Chapter 3, section 3.6.5.

5.2 Classification of Bacteria Using Artificial Neural Networks

An analysis using ANN was done to compare the performance of ANN to DFA. The same 3 species as in the DFA were used. Our ANN algorithm was initially developed in Python by Alayna Tieu and uses the libraries Pandas, Numpy, Tensorflow, keras, and Scikit-Learn. A brief description of how ANN algorithms work is given in Chapter 3.6.5. The ANN algorithm developed consisted of 92 input nodes, corresponding to the number of independent variables in the RM2.5 sheet. The algorithm consisted of 1 hidden layer where the number of nodes in this layer was optimized for each run using an optimization algorithm to optimize the number of epochs, written by Alayna Tieu, prior to using the ANN. The output nodes correspond to the number of classes being identified, which is set by the user. Other ANN parameters to optimize are the epochs, the batch size, number of hidden nodes, and the patience. An epoch is one iteration through the entire data set.[100] Our epoch number was optimized for each data set. The batch size is the number of items or nodes that are updated together during a given epoch. Our batch size was 32, which means that 32 nodes were updated per epoch. The patience determines when the ANN stops, it is meant to minimize the loss, the loss being the difference between an expected outcome and a current outcome.[100] The patience was optimized for each data set tested in the ANN.

Before the data is classified with ANN, the program randomly splits the data up into a training set and a testing set. The testing set was created by removing 20% of the data from the training set. Typically, the larger the training set the better the performance.

The resulting average sensitivity and specificity of the 3 bacterial species using the ANN optimize algorithm as well as the ANN were 64.62% and 82.29%, respectively. These values were calculated by running the ANN multiple times and using the sensitivity and specificity obtained from each run to compute the average sensitivity and specificity. The average was calculated because the ANN produces slightly different results each time it runs since it randomizes the files chosen for the test set. Overall, the performance of the ANN were highly similar to the results of DFA, and therefore the ANN still does not provide highly accurate results. The results are summarized below in Table 5.9.

To improve on these results, optimization of other parameters was investigated. A new ANN optimization algorithm was developed by Grace Johnson and August Baughan to optimize the patience and number of hidden nodes. This algorithm outputs the average sensitivity and specificity for each value of patience and number of hidden nodes into a spreadsheet. The user then chooses the patience and hidden nodes that correspond to the highest values of sensitivity and specificity, which are inputted into the ANN algorithm. For the 3 species, the ANN was run multiple times and an average sensitivity and specificity for each species was calculated. The average overall sensitivity and specificity was 66.78% and 83.36%, respectively. These results are shown in Table 5.10. The new optimization algorithm offered no improvement over the previous algorithm.

As with DFA, an add-all analysis was attempted in ANN as well based on the previous success in PLSDA. All 30 spectra from each filter were added to produce one spectrum per filter. The method used to create the add-all filters was the Excel method previously discussed. The results for the classification of 3 species is summarized below in Table 5.11. The average sensitivity and specificity of this method is 47.22% and 73.61%, respectively. The add-all approach performed worse than the single-spectra approach in ANN and performed worse than the add-all approach in DFA.

Table 5.9: Results of 3 species after optimization of epochs, followed by ANN analysis.

ANN With RM2.5 Normalized Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	63.33 %	66.09 %	64.44 %
Specificity	75.14 %	84.44 %	87.29 %
Classification Error	30.76 %	24.73 %	24.13 %

Table 5.10: Results of 3 species after optimization of patience and hidden nodes, followed by ANN analysis.

ANN With RM2.5 Normalized Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	60.14 %	69.78 %	70.42 %
Specificity	82.38 %	84.06 %	83.64 %
Classification Error	28.74 %	23.08 %	22.97 %

Table 5.11: Results of 3 species after add-all operation in Excel and optimization of patience and hidden nodes, followed by ANN analysis.

ANN With Add-All RM2.5 Normalized Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	33.33 %	69.78 %	70.42 %
Specificity	62.50 %	84.06 %	83.64 %
Classification Error	52.08 %	33.33 %	33.33 %

5.2.1 Data Reduction Using Principal Component Analysis

Based on other research in the field of LIBS, preprocessing of data before using ANN was attempted. Some researchers have reported on improved results when conducting a principal component analysis (PCA) analysis on data and using the principal component scores as independent variables in ANN.[101,102] The number of principle component scores retained from the PCA will therefore be the new number of input nodes in the ANN algorithm. PCA is a multivariate method that reduces the inputted data to a set of principal component (PC) scores and is often used for pattern recognition.[103] A PCA algorithm was

developed by Emma Blanchette and Grace Johnson in python using the libraries sklearn.decomposition, pandas, numpy, and mpl_toolkits. PCA scores were created for the RM2.5 data and 4 PC scores were kept. 4 PC scores were chosen because they encompassed 93% of the variance in the data. The ANN optimize algorithm developed by Grace Johnson and August Baughan was run using the 4 PC scores outputted by the PCA algorithm. The results of this test are summarized in Table 5.12 on page 122. The average sensitivity and specificity was 33.33% and 66.67%, respectively. PCA was attempted again with more PC scores kept to capture more of the variance in the data. 10 PC scores were kept, which captured 99.58% of the variance. The results of this were more favourable, but still offered no improvement over previous DFA and ANN results. The average sensitivity and specificity for this test was 66.46% and 83.34%, respectively. The results are summarized in Table 5.13 on page 122.

These results led us to conclude that we were likely at the limit with what 15 lines and their ratios could classify. Based on this conclusion, we decided to approach classification of species using a full spectrum as opposed to hand-picking 15 lines. This approach has been used in the past by other groups successfully.

Full spectrum analysis was done by extracting the raw intensity from each channel in the Esawin software into an Excel sheet. The data ranged from 200.915 nm to 779.915 nm. This results in 53,434 data points for each filter. The full spectrum data was fed to PCA and 10 PC scores were kept. The 10 PC scores were used in the ANN optimize algorithm to optimize patience and number of hidden nodes. Based on this algorithm a patience value and number of hidden nodes were chosen for the ANN. The results of this ANN analysis are shown in Table 5.14, page 122. The average sensitivity and specificity for this test was 94.18% and 97.02%, respectively. The average and individual sensitivities and specificities are higher than all previous tests between bacterial species.

These experiments were expanded upon by investigating mean-centering, a common data-preprocessing methods used in conjunction with PCA. Before performing PCA on the full-spectrum data, each variable was mean-centered. This was done by calculating the mean of the data across each wavelength and subtracting the mean from each data point of

the same wavelength. This was done for all 53,434 channels. 10 PC's were constructed from the mean-centered data. The 10 PC scores were used in the ANN optimization algorithm, and the ANN analysis resulted in similar sensitivity and specificity without preprocessing of data, likely within error of the previous scores. These results are summarized in Table 5.15 on page 122. Mean-centering provided good results, but no better than the results without the use of pre-processing. Due to the high computation requirements with no improvement observed, mean-centering is not recommended for this application. To verify that the algorithm was not fitting noise and instead looking at the important features of the spectra, classes from 1 to 3 were assigned randomly to each bacterial spectrum to purposely make the test fail. The test results showed random classification and poor performance, verifying that the algorithm is not fitting the noise. An example of this test is shown below in Table 5.16 on page 123, and this test was performed for all future PCA-ANN tests to confirm that the algorithm was not fitting the noise.

Table 5.12: Results of 3 species after reduction to 4 PC scores in PCA and optimization of patience and hidden nodes, followed by ANN analysis.

PCA-ANN With RM2.5 Data, 4 PC Scores			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	100 %	0 %	0 %
Specificity	0 %	100 %	100 %
Classification Error	50.00 %	50.00 %	50.00 %

Table 5.13: Results of 3 species after reduction to 10 PC scores in PCA and optimization of patience and hidden nodes, followed by ANN analysis.

PCA-ANN With RM2.5 Data, 10 PC Scores			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	58.82 %	62.50 %	78.07 %
Specificity	86.70 %	85.65 %	77.67 %
Classification Error	27.24 %	25.92 %	22.13 %

Table 5.14: Results of full-spectrum analysis of 3 species after reduction to 10 PC scores in PCA and optimization of patience and hidden nodes, followed by ANN analysis.

PCA-ANN With Full Spectrum Data			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	98.04 %	93.27 %	91.23 %
Specificity	97.71 %	97.22 %	96.12 %
Classification Error	2.13 %	4.28 %	6.33 %

Table 5.15: Results of mean-centering full-spectrum data before reduction to 10 PC scores in PCA and optimization of patience and hidden nodes, followed by ANN analysis.

PCA-ANN With Full Spectrum Data: Mean Centering			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	99.00 %	94.23 %	90.35 %
Specificity	96.78 %	98.14 %	96.60 %
Classification Error	2.11 %	3.18 %	6.53 %

Table 5.16: Results of class randomization to test if the PCA-ANN algorithm was fitting noise or fitting important features in the data.

PCA-ANN With Full Spectrum Data: Randomized Classes			
	E. coli	*S. aureus*	*E. cloacae*
Sensitivity	36.19 %	51.38 %	2.83 %
Specificity	55.81 %	44.55 %	94.86 %
Classification Error	54.00 %	52.04 %	51.16 %

Further investigation was done to confirm that 10 PC scores was optimal for both full-spectrum and the mean-centered full-spectrum. It was observed that less than 10 PC's resulted in lower sensitivity and specificity for both cases. It was also observed that greater than 10 PC's results in lower sensitivity and specificity for both cases. One exception to this case was the use of 20 PCs. In both cases, 20 PC's performed similarly to 10 PC's. It is therefore recommended that full spectrum analysis continue to be reduced to 10 PC scores with PCA, followed by analysis with ANN, with no pre-processing. 10 was chosen over 20 to avoid overfitting. These findings are summarized in Table 5.17 and Table 5.18.

Table 5.17: Results for PCA-ANN with full spectrum analysis. ANN optimization and classification algorithm is run with 5,8,10,12,15, and 20 PC's to demonstrate that 10 PC's is the optimal number.

		E. coli	*S. aureus*	*E. cloacae*
5 PC Scores	Sensitivity	75.00 %	82.69 %	92.98 %
	Specificity	98.17 %	96.79 %	80.29 %
	Classification Error	13.41 %	10.26 %	13.36 %
8 PC Scores	Sensitivity	91.18 %	87.50 %	98.25 %
	Specificity	98.62 %	98.15 %	91.75 %
	Classification Error	5.10 %	7.17 %	5.00 %
10 PC Scores	Sensitivity	98.04 %	93.27 %	91.23 %
	Specificity	97.71 %	97.22 %	96.12 %
	Classification Error	2.13 %	4.28 %	6.33 %
12 PC Scores	Sensitivity	91.18 %	92.31 %	91.23 %
	Specificity	98.62 %	94.44 %	94.17 %
	Classification Error	5.10 %	6.63 %	7.30 %
15 PC Scores	Sensitivity	93.14 %	89.42 %	92.98 %
	Specificity	97.71 %	98.15 %	91.75 %
	Classification Error	4.58 %	6.22 %	7.64 %
20 PC Scores	Sensitivity	98.04 %	97.12 %	91.23 %
	Specificity	95.87 %	98.15 %	99.03 %
	Classification Error	3.05 %	2.36 %	4.87 %

Table 5.18: Results for mean-centering of data before application of PCA-ANN on full spectrum data. ANN optimization and classification algorithm is run with 5,8,10,12,15, and 20 PC's to demonstrate that 10 PC's is the optimal number.

		E. coli	*S. aureus*	*E. cloacae*
5 PC Scores	Sensitivity	95.10 %	84.61 %	68.42 %
	Specificity	89.91 %	89.35 %	94.17 %
	Classification Error	7.46 %	13.02 %	18.71 %
8 PC Scores	Sensitivity	90.20 %	88.46 %	92.98 %
	Specificity	96.79 %	96.30 %	92.72 %
	Classification Error	6.51 %	7.62 %	7.15 %
10 PC Scores	Sensitivity	99.00 %	94.23 %	90.35 %
	Specificity	96.78 %	98.14 %	96.60 %
	Classification Error	2.09 %	3.81 %	6.53 %
12 PC Scores	Sensitivity	97.06 %	98.08 %	84.21 %
	Specificity	97.71 %	93.52 %	98.06 %
	Classification Error	2.62 %	4.2 %	8.87 %
15 PC Scores	Sensitivity	96.08 %	95.19 %	89.47 %
	Specificity	97.71 %	95.37 %	97.09 %
	Classification Error	3.105 %	4.72 %	6.72 %
20 PC Scores	Sensitivity	96.08 %	97.12 %	91.23 %
	Specificity	97.71 %	97.22 %	97.09 %
	Classification Error	3.105 %	2.83 %	5.84 %

The improvement of PCA-ANN over DFA can be attributed to ANN's nonlinearity. PLDSA and DFA are linear models and are typically used when the relationship between predictors in the x-block and responses in the y-block is linear. In nonlinear models, the predictors influence the response indirectly, or in a nonlinear way.[104] An easy way to test if

the data is linear or non-linear is to determine if the classes are linearly separable using a scatter plot.[105] To test if our data was better suited to a non-linear model, a scatter plot was created comparing 2 independent variables between 3 classes of bacteria, *E. coli, S. aureus,* and *E. cloacae.* If the data is linear, the classes should be separable by a line. This scatter plot is shown in Figure 5.7. It is clear when viewing the scatter plot that the classes cannot be separated by drawing a straight line. This data was analyzed after ANN was applied to the data to provide additional support for the recommendation that a nonlinear classification method such as ANN continue to be used for classification instead of DFA.

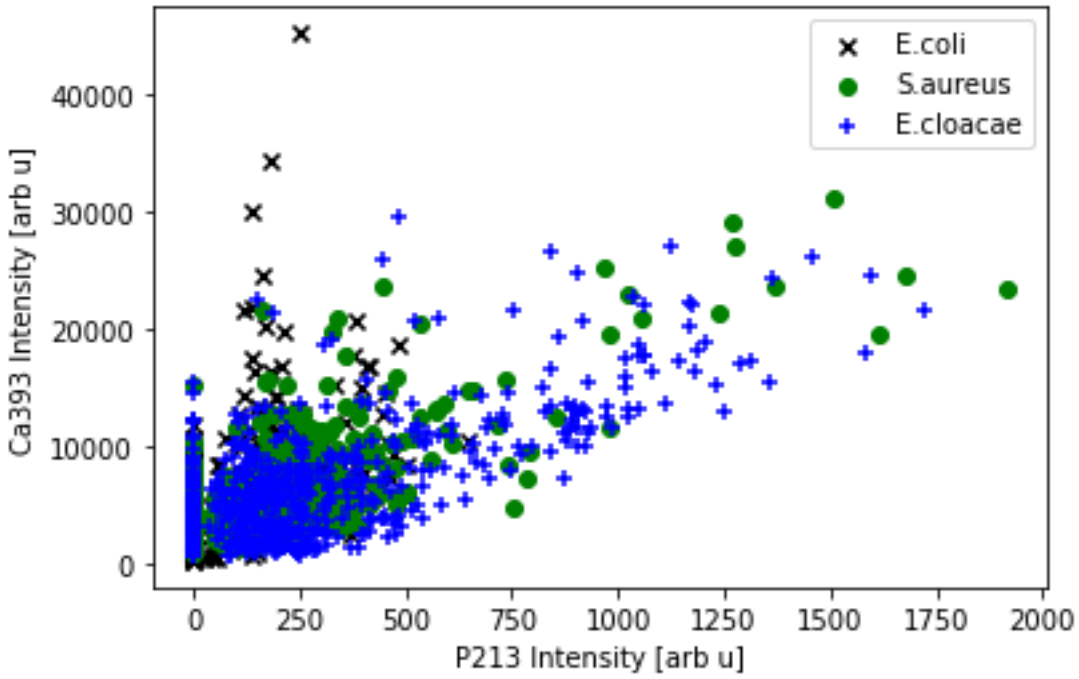

Figure 5.7: Scatter plot showing the separation of two different variables for 3 species of bacteria.

The improvement in classification due to reduction of data using PCA on full spectrum data can likely be attributed to PCA's ability to find differences between classes, and more importantly, what contributes most to these differences. As well, PCA reduces dimensionality which avoids any redundancy and highlights the data that is most useful for discriminating between species.[106]

All tests using PCA-ANN on full spectrum data were externally validated by splitting the data set into 2 parts; 80% of the data was used to build the library, 20% was tested against this library. When conducting an external validation in ANN using the method described

above, the data points are removed at random, therefore it is highly unlikely that an entire filter will be removed simultaneously, resulting in spectra that are always similar to those removed. While there is shot-to-shot variability within a filter, there is also substantial variability between filters. This motivated the next external validation test of removing entire filters and testing them against the model. This test is also clinically relevant, since a patient's bacterial sample will likely not exist in the model prior to them being tested.

A 2-class test was performed between *E. coli* and *S. aureus*. These species were chosen for the 2-class test because they are the simplest case. These species do not often get confused for each other in our classification algorithms, whereas *E. cloacae* tends to be confused for both *E. coli* and *S. aureus*, and *M. smegmatis* is never confused with any of these species. Whole filters were removed and tested against the model. The sensitivity was calculated for each filter based on how many spectra classified correctly. The average sensitivity of the *E. coli* filters is 85.43%, and the average sensitivity for *S. aureus* is 77.07%. Some filters of *E. coli* and *S. aureus* had 0% of shots classify correctly, so these filters were removed from the model. After removing these filters, the model was retested with the remaining filters to try and improve performance. Removing filters that performed poorly in the analysis greatly improved the average sensitivity for each species, which were calculated to be 93.76% for *E. coli* and 85.20% for *S. aureus* respectively. The improvement of this classification points to the removal of poorly performing filters to improve classification results. A summary of these results and the sensitivity for each individual filter classification is provided in Table 5.19 and 5.20.

Table 5.19: Sensitivity for each individual filter externally validated against the training model for the species E. coli.

E. coli	Predicted		
Sample #	*E. coli*	*S. aureus*	Sensitivity
1	0	29	0
2	27	3	0.9
3	0	30	0
4	30	0	1
5	30	0	1
6	30	0	1
7	30	0	1
8	22	8	0.733333333
9	30	0	1
10	30	0	1
11	30	0	1
12	30	0	1
13	30	0	1
14	30	0	1
15	30	0	1
16	30	0	1
17	25	4	0.862068966
Sum	434	74	0.8543307

Table 5.20: Sensitivity for each individual filter externally validated against the training model for the species S. aureus.

S. aureus	Predicted		
Sample #	*E. coli*	*S. aureus*	Sensitivity
1	18	12	0.4
2	25	0	0
3	0	30	1
4	9	6	0.4
5	30	0	0
6	0	30	1
7	0	30	1
8	0	30	1
9	0	30	1
10	0	30	1
11	4	26	0.866666667
12	25	5	0.166666667
13	0	29	1
14	0	30	1
15	0	30	1
16	0	30	1
17	8	22	0.733333333
18	0	30	1
Sum	119	400	0.7707129

To further test the PCA-ANN model's robustness, a 3-class test was performed between *E. coli*, *S. aureus*, and *E. cloacae.* The average sensitivity of *E. coli*, *S. aureus*, and *E. cloacae* are 77.95%, 64.73%, 32.16%. All 3 species did not perform as well as the previous external test. This was unsurprising as *E. cloacae* tends to classify as other species, most often *S. aureus*. The results for individual filters are summarized in Tables 5.21, 5.22, and 5.23. Upon inspections of the results, one can notice that samples 3 to 10, *E. cloacae* filters, performed much better than the rest of the *E. cloacae* filters. These filters were investigated further, which showed that these filters had abnormally high total spectral intensities when compared to all other data from the same species. Based on this finding, most of these samples were removed from the analysis and replaced with new data that was more representative of *E. cloacae* spectra. These results are summarized in Tables 5.24, 5.25, and 5.26. The average sensitivity of *E. coli*, *S. aureus*, and *E. cloacae* in this test are 71.06%, 61.27%, 65.18%. There is an improvement in *E. cloacae* sensitivity, but the sensitivity for *E. coli* and *S. aureus* decreases.

Table 5.21: Sensitivity for each filter that was externally validated against the training model for E. coli.

E. coli	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	0	27	2	0
2	20	10	0	66.66666667
3	0	24	6	0
4	28	2	0	93.33333333
5	30	0	0	100
6	21	8	1	70
7	29	1	0	96.66666667
8	30	0	0	100
9	12	1	17	40
10	19	11	0	63.33333333
11	25	5	0	83.33333333
12	29	1	0	96.66666667
13	26	4	0	86.66666667
14	8	22	0	26.66666667
15	26	1	3	86.66666667
16	29	1	0	96.66666667
17	29	0	0	100
Sum	382	67	29	0.8507795

Table 5.22: Sensitivity for each filter that was externally validated against the training model for S. aureus.

S. aureus	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	0	30	0	100
2	13	12	0	48
3	0	29	1	96.66667
4	3	2	10	13.33333
5	30	0	0	0
6	0	30	0	100
7	0	30	0	100
8	0	30	0	100
9	0	29	1	96.66667
10	0	29	1	96.66667
11	3	24	3	80
12	0	1	29	3.333333
13	1	22	6	75.86207
14	0	1	29	3.333333
15	0	30	0	100
16	0	29	1	96.66667
17	11	7	12	23.33333
18	2	1	27	3.333333
Sum	63	336	120	64.73988

Table 5.23: Sensitivity for each filter that was externally validated against the training model for E. cloacae.

E. cloacae	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	0	30	0	0
2	0	30	0	0
3	1	29	0	0
4	0	4	26	86.66667
5	0	7	23	76.66667
6	0	2	28	93.33333
7	0	30	0	0
8	0	1	28	96.55172
9	0	6	24	80
10	0	1	29	96.66667
11	0	29	1	3.333333
12	21	7	2	6.666667
13	12	18	0	0
14	8	12	10	33.33333
15	0	19	11	36.66667
16	30	0	0	0
17	0	29	1	3.333333
18	24	6	0	0
19	17	13	0	0
Sum	113	273	183	32.16169

Table 5.24: Sensitivity of each filter after external validation against the training model for each E. coli *filter.*

E. coli	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	0	27	2	0
2	20	10	0	66.66666667
3	0	24	6	0
4	28	2	0	93.33333333
5	30	0	0	100
6	21	8	1	70
7	29	1	0	96.66666667
8	30	0	0	100
9	12	1	17	40
10	19	11	0	63.33333333
11	25	5	0	83.33333333
12	29	1	0	96.66666667
13	26	4	0	86.66666667
14	8	22	0	26.66666667
15	26	1	3	86.66666667
16	29	1	0	96.66666667
17	29	0	0	100
Sum	361	118	29	71.06299213

Table 5.25: Sensitivity for each filter after external validation against the training model for each S. aureus *filter.*

S. aureus	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	1	4	25	13.333333
2	12	13	0	52
3	0	30	0	100
4	2	11	2	73.333333
5	30	0	0	0
6	0	30	0	100
7	0	30	0	100
8	0	30	0	100
9	0	25	5	83.333333
10	0	26	4	86.666667
11	7	18	5	60
12	0	1	29	3.3333333
13	28	1	0	3.4482759
14	0	6	24	20
15	0	30	0	100
16	0	30	0	100
17	0	26	4	86.666667
18	1	7	22	23.333333
Sum	81	318	120	61.271676

Table 5.26: Sensitivity of each filter after external validation against the training model for each E. cloacae *filter.*

E. cloacae	Predicted			
Sample #	*E. coli*	*S. aureus*	*E. cloacae*	Sensitivity
1	0	0	30	100
2	0	0	30	100
3	0	13	17	56.666667
4	0	16	14	46.666667
5	0	26	4	13.333333
6	15	0	15	50
7	1	0	29	96.666667
8	7	20	3	10
9	0	13	17	56.666667
10	26	4	0	0
11	1	11	18	60
12	22	0	8	26.666667
13	7	0	23	76.666667
14	3	0	27	90
15	0	0	30	100
16	1	0	29	96.666667
17	2	0	28	93.333333
18	0	0	30	100
Sum	85	103	352	65.185185

When comparing the results of the 3-class test to the previous 2-class test, it was noted that *E. cloacae* markedly decreased the sensitivity of the *E. coli* and *S. aureus.* To test if *E. cloacae* posed an issue for classification or was too difficult to classify with the other 2 species, we replaced *E. cloacae* with *M. smegmatis* for a 3-class test. Performance in the external validation test was markedly better for *E. coli* and *S. aureus* compared to the other 3-class test with *E. cloacae,* with average sensitivities of 84.44% and 77.26%. *M. smegmatis* achieved an average sensitivity of 99.76%. In this test, *M. smegmatis* was rarely confused with the other species, and the most common error was *E. coli* being confused for *S. aureus* and vice-versa. *M. smegmatis* represents the easiest case of classification, while *E. cloacae* appears to be the hardest to classify accurately. The reasons for this are likely due to the biochemical differences observed in each cell. *M. smegmatis* regularly has higher intensity magnesium peaks than calcium peaks, whereas *E. cloacae* mimics the biochemistry of *E. coli* and *S. aureus*, having higher intensity calcium peaks than magnesium peaks. We can therefore differentiate between more diverse species with higher accuracy than species more closely related. The results for individual filters are summarized in the below tables.

Table 5.27: Sensitivity of each filter that was externally validated against the training model for E. coli.

E. coli	Predicted			
Sample #	*E. coli*	*S. aureus*	*M. smegmatis*	Sensitivity
1	0	29	0	0
2	24	6	0	80
3	0	30	0	0
4	23	7	0	76.666667
5	30	0	0	100
6	28	2	0	93.333333
7	30	0	0	100
8	30	0	0	100
9	30	0	0	100
10	30	0	0	100
11	30	0	0	100
12	30	0	0	100
13	30	0	0	100
14	30	0	0	100
15	29	1	0	96.666667
16	30	0	0	100
17	25	4	0	86.206897
Sum	429	79	0	84.448819

Table 5.28: Sensitivity of each filter that was externally validated against the training model for S. aureus.

S. aureus	Predicted			
Sample #	*E. coli*	*S. aureus*	*M. smegmatis*	Sensitivity
1	0	30	0	100
2	13	0	12	0
3	0	30	0	100
4	3	12	0	80
5	30	0	0	0
6	0	30	0	100
7	0	24	6	80
8	0	30	0	100
9	0	30	0	100
10	0	30	0	100
11	3	27	0	90
12	29	1	0	3.3333333
13	0	29	0	100
14	0	30	0	100
15	0	29	1	96.666667
16	0	30	0	100
17	19	11	0	36.666667
18	2	28	0	93.333333
Sum	99	401	19	77.263969

Table 5.29: Sensitivity of each filter that was externally validated against the training model for E. cloacae.

M. smegmatis	Predicted			
Sample #	*E. coli*	*S. aureus*	*M. smegmatis*	Sensitivity
1	0	0	30	100
2	0	0	30	100
3	0	0	30	100
4	0	0	30	100
5	0	0	30	100
6	0	0	30	100
7	0	0	30	100
8	0	0	30	100
9	0	1	29	96.666667
10	0	0	30	100
11	0	0	30	100
12	0	0	30	100
13	0	0	29	100
14	0	0	30	100
Sum	0	1	418	99.761337

As with the previous 2 and 3-class tests, filters of *E. coli* and *S. aureus* that classified at 0% were removed from the 3-class analysis with *M. smegmatis.* Before removing these filters from the classification, we confirmed that they were the same filters that had previously failed in the 2-class test. Our results are shown in Tables 5.30 and 5.31 for *E. coli* and *S. aureus.* The *M. smegmatis* results remained the same and are shown in Table 5.29 above. The average sensitivity of the externally tested *E. coli* filters increased to 97.10%, and the average sensitivity of the externally tested *S. aureus* filters increased to 95.79%.

Table 5.30: Sensitivity for each filter that was externally tested against the model for the species E. coli *after filters that completely misclassified were removed.*

E. coli	Predicted			
Sample #	*E. coli*	*S. aureus*	*M. smegmatis*	Sensitivity
1	30	0	0	100
2	30	0	0	100
3	28	2	0	93.33333
4	24	6	0	80
5	30	0	0	100
6	30	0	0	100
7	30	0	0	100
8	30	0	0	100
9	30	0	0	100
10	30	0	0	100
11	30	0	0	100
12	30	0	0	100
13	30	0	0	100
14	28	2	0	93.33333
15	26	2	1	89.65517
Sum	436	12	1	97.10468

Table 5.31: Sensitivity for each filter that was externally tested against the model for the species S. aureus *after filters that completely misclassified were removed.*

S. aureus	Predicted			
Sample #	*E. coli*	*S. aureus*	*M. smegmatis*	Sensitivity
1	7	23	0	76.66667
2	0	30	0	100
3	0	15	0	100
4	0	30	0	100
5	0	25	5	83.33333
6	0	30	0	100
7	0	30	0	100
8	0	30	0	100
9	3	27	0	90
10	0	28	1	96.55172
11	0	30	0	100
12	0	29	1	96.66667
13	0	30	0	100
14	0	30	0	100
Sum	10	387	7	95.79208

While the result that more diverse species are easier to discriminate is unsurprising, advancements have been made discriminating between more closely related species. The overall sensitivity and specificity for closely related species was improved using full

spectrum analysis in PCA-ANN. Some improvements on the discrimination between *E. cloacae* , *E. coli*, and *S. aureus* need to be made. Most notably, the inter-filter reliability is low and should be addressed to improve reproducibility between filters. The above classifications where filters that classified poorly were removed indicated that our data tends to fail by filter. Increasing this reproducibility should improve the self-consistency within the database for each species and eliminate the need to remove poorly performing filters from training and testing sets. We have proposed to do this with a new design for a concentration piece, since the cone does not produce samples that are reliably the same.

5.3 Conclusions and Recommendations for Future Work

When using DFA to classify bacterial species, accuracy is low and not clinically relevant. No improvements were observed when preprocessing the data or eliminating data, therefore, DFA will no longer be investigated for the current sample preparation method. Using ANN with 15 lines and optimized numbers of epochs, patience, and hidden nodes also made some improvement. However, reduction of full-spectrum data by PCA followed by ANN increased values of sensitivity and specificity for all species. This method will be further investigated with lower concentrations of bacterial suspensions to determine if any improvement can be made on past results. It is recommended that PCA-ANN on full spectrum data continue to be used. This method will also be further investigated with external validation by removing whole filters. Methods of filter rejection need to be investigated to improve results of external validation in PCA-ANN.

For future studies, DFA could still be useful; it works faster than ANN and requires less computation power. A problem commonly experienced with our spectra is low signal of important lines such as phosphorus, and one common issue with LIBS is the shot-to-shot variability; both likely contribute to the wide spread of data around a group centroid and consequently poor discrimination as seen in the DFA plots. To increase overall signal intensity and reduce the shot-to-shot variability, we have been investigating deposition of silver nanoparticles onto bacterial targets to give a more consistent and enhanced signal. It is hypothesized that based on previous work on enhancement of LIBS spectra with silver and gold nanoparticles that we will be able to enhance the important lines present in our spectra and reduce the shot-to-shot variability.[107] A review of nanoparticle-enhanced LIBS

(NELIBS) is out of scope for this thesis. For more information on nanoparticle-enhanced LIBS, the reader is directed to the review listed in reference 106.

Finally, a review of our deposition method is needed. The current method of deposition using the cone placed inside the centrifuge insert occasionally leaks and causes bacterial deposition to be different between filters. As well, not all bacteria is caught on the filter as previously thought using the current deposition method. This finding will be covered in greater detail in Chapter 7. A new design for a concentration device has been proposed to increase inter-filter reliability and will be investigated in the future.

[95] Mohaidat, Q. I. , Sheikh, K., Palchaudhuri, S., Rehse, S. J. (2021). Pathogen identification with laser-induced breakdown spectroscopy: The effect of bacterial and biofluid specimen contamination, *Applied Optics*, *51*(7), B99-B107. https://doi.org/10.1364/AO.51.000B99.

[96] Malenfant, D. J., Gillies, D. J., Rehse, S. J. (2016) Bacterial suspensions deposited on microbiological filter material for rapid laser-induced breakdown spectroscopy identification, *Applied Spectroscopy*, *79*(3), 485–493. https://doi.org/10.1177/0003702815626673.

[97] Malenfant, D. J. (2016). *Influences on the Emissions of Bacterial Plasmas Generated through Nanosecond Laser-Induced Breakdown Spectroscopy* [Master's Thesis, University of Windsor].

[98] Paulick, A. E., Malenfant, D. J., Rehse, S. J. (2019). Concentration of bacterial specimens during centrifugation prior to laser-induced breakdown spectroscopy analysis, *Spectrochimica Acta Part B: Atomic Spectroscopy, 157*, 68–75. https://doi.org/10.1016/j.sab.2019.05.012.

[99] Blanchette, E. J., et al. (2022). Detection and classification of bacterial cells after centrifugation and filtration of liquid specimens using laser-induced breakdown spectroscopy, *Applied Spectroscopy, 76(8),* 894-904. https://doi.org/10.1177/00037028221092789.

[100] Afaq, S., & Rao, S. (2020). Significance of epochs on training a neural network, *International Journal of Scientific & Technology Research, 9*, 485-488.

[101] Zhao, Y., Wang, Q., Cui, X., Teng, G., Wei, K., Liu, H. (2020). Discrimination of hazardous bacteria with combination laser-induced breakdown spectroscopy and statistical methods, *Applied Optics*, *59*(1), 1329-1337. https://doi.org/10.1364/AO.379136.

[102] Yang, Y., Li, C., Liu, S., Min, H., Yan, C., Yang, M., Yu, J. (2020). Classification and identification of brands of iron ores using laser-induced breakdown spectroscopy combined with principal component analysis and artificial neural networks, *Analytical Methods, 12*(10), 1316-1323. https://doi.org/10.1039/C9AY02443C.

[103] Brereton, R. G. (2007). *Applied Chemometrics for Scientists*. (1st ed). Wiley.

[104] Kinhal, V. (2022, August 30). *All About ANN – Artificial Neural Networks and Chemometric Modeling for NIR Spectroscopy*. Felix Instruments. Retrieved September 15, 2022, from https://felixinstruments.com/blog/all-about-ann-artificial-neural-networks-and-chemometric-modeling-for-nir-spectroscopy/.

[105] Kumar, A. (2022, July 31). *Linear vs Non-linear Data: How to Know*. Data Analytics. Retrieved September 15, 2022, from https://vitalflux.com/how-know-data-linear-non-

linear/#:%7E:text=for%20Regression%20Problem-,Use%20Scatter%20Plots%20for%20Classification%20Problems,scatter%20plots%20representing%20different%20classes.https://vitalflux.com/how-know-data-linear-non-linear/#:%7E:text=for%20Regression%20Problem-,Use%20Scatter%20Plots%20for%20Classification%20Problems,scatter%20plots%20representing%20different%20classes.

[106] Ioele, G., De Luca, M., Dinç, E., Oliverio, F., & Ragno, G. (2011). Artificial neural network combined with principal component analysis for resolution of complex pharmaceutical formulations, *Chemical & pharmaceutical bulletin*, *59*(1), 35–40. https://doi.org/10.1248/cpb.59.35.

[107] Dell'Aglio, M., Alrifai, R., De Giacomo, A. (2018). Nanoparticle enhanced laser-induced breakdown spectroscopy (NELIBS), a first review, *Spectrochimica Acta Part B: Atomic Spectroscopy, 148,* 102-112. https://doi.org/10.1016/j.sab.2018.06.008.

Chapter 6: Detection of Bacterial Pathogens in Clinical Fluids

Hospital acquired infections, or nosocomial infections, affect millions of people a year and are often antibiotic resistant infections. Nosocomial infections often result from open surgical sites, catheters, or intubation with a ventilator. Because of this the most common infections are urinary tract infections (UTI), surgical infections, blood-stream infections (septicemia), and lower respiratory infections (pneumonia). These infections form because catheters and other equipment that are intravenous provide a site for bacteria to enter the body and a surface for bacteria to colonize and form a biofilm. Nosocomial infections are one of the leading causes of death. Hospitals must screen patients before releasing them for any nosocomial infection.[108]

Up to 80% of urinary tract infections are associated with a bladder catheter, and thus are a nosocomial infection.[108] Urinary tract infections are second in incidence only to respiratory infections, thus are extremely common and often antibiotics are sought to treat them. Diagnosis of a UTI relies on culturing urine, which must be performed shortly after acquiring the sample to avoid growth of other organisms. Culture is often time consuming, costly, and does not detect all bacteria as some cannot be cultured on nutrient-free media.[109] As well, the prevailing understanding until very recently was that urine is a completely sterile fluid. Recent studies have called this into question stating that female patients contain bacterial colonies in the bladder.[110] This increases the need for a highly sensitive and specific test for diagnosing the species present in the urine. This new information also increases the need for previous work done on the detection of bacteria in mixed cultures.[111] Many nosocomial UTI's develop from antibiotic resistant bacteria.[108]

A highly relevant and dangerous nosocomial infection is sepsis, or a bacterial blood infection. Sepsis is often caused by wounds (surgically induced or otherwise), and intravenous tubes or catheters. Sepsis is not always caused by bacteria, it can also be caused by fungi, viruses, or bowel leakage, but many cases are bacterial. The most common types of bacteria causing sepsis infections are gram negative enteric rods. Gram negative enteric rods cause approximately 300000 cases of sepsis per year where nearly one third of those infected die. Sepsis affects every organ in the body and if not caught quickly can

cause multi-organ failure. However, sepsis is highly difficult to diagnose and often presents differently for many patients. A blood culture is required for diagnosis, but not all blood cultures will produce a positive test result. Antibiotics are the preferred treatment, but even with the best treatment the mortality for patients who have reached shock is no better than 50%.[112]

In this chapter, I propose that LIBS could be used as a point-of-care technique directly on these sterile fluids to test for bacterial infections. The speed and specificity of LIBS will aid in delivering targeted treatment to patients experiencing potentially life-threatening nosocomial infections and should reduce the use of broad-spectrum drug in treating non-nosocomial UTI's. As well, the use of LIBS for screening patients before they leave the hospital will improve efficiency of patient discharge. Screening patients requires culturing, which uses up valuable time and resources, including hospital beds.

6.1 LIBS on Sterile Urine

Sterile urine was provided to the Rehse lab by the pathology lab at the Windsor Regional Hospital Ouellette campus. Samples were provided by Lucy DiPietro and Dr. Mohamed El-Fakharany. Only urine specimens that had already been tested negative for bacterial infection were provided in these preliminary experiments. Sterile urine was sampled and characterized with the LIBS system to understand how the spectrum of urine will affect the spectrum of bacteria. Several different urine samples from different patients were characterized to account for biochemical differences between samples.

Very little work has been done using LIBS on the analysis of urine. Some authors have investigated the use of LIBS to investigate the chemical composition of urinary and gall stones in order to understand their formation and pathogenesis.[113,114,115]Mohaidat et al. investigated the possibility of clinical application by discriminating between *Staphylococcus* species in urine, which was described in Chapter 2.[116]

To test urine samples, the deposition technique was changed from the previous swab technique to represent how this technology may be used in the clinic. In this deposition, the centrifuge piece was assembled with a 0.45 μm filter and plastic cone. Urine was vortexed before deposition in the centrifuge piece. 100 μL of urine was deposited directly into the

cone. The sample was centrifuged at 5000 RPM's for 5 minutes to deposit the urine on the filter. After centrifugation, the filter was removed from the centrifuge insert and mounted on a steel piece with double sided sticky tape. All the same LIBS parameters that were used previously to sample bacteria were used to sample urine.

Urine spectra from sterile urine cultures 1 and 3 are shown below in Figure 6.1. These spectra show that urine contains sodium, magnesium, carbon, and calcium. This is not surprising since the kidneys regulate salt content in the body and contain several ions and molecules, which includes sodium, calcium, and magnesium.[117] The urine cultures presented in Figure 6.1 come from two different patients. It is apparent that there are differences between these spectra; the sodium lines in Figure 6.1b are higher than the sodium lines in Figure 6.1a. As well, the magnesium lines are higher in Figure 6.1a than in Figure 6.1b. These differences in spectra highlight the large variety that exists between patients and their physiologic systems. Urine spectra from several urine samples showed relatively high sodium in comparison to other lines, and low amounts of calcium, magnesium, and phosphorus relative to other bacterial spectra. While the urine contributes to a non-zero background, the lines in the urine spectrum that interfere with the bacterial spectra are lower in intensity and are present in a different ratio with respect to the sodium lines. Therefore, it is expected that the urine will not interfere significantly with bacterial detection and diagnosis.

It should be noted that the spectra from the urine sample, while representative of the differences between patients, may not be entirely representative of a fresh urine sample. Our urine samples are stored in a fridge at 4°C and are kept and used for several months, solute can coagulate over this period of time or settle out of solution, perhaps making the line intensities smaller than what would be seen in fresh urine. However, it is difficult to continually obtain fresh samples and store them, so the urine is vortexed before sampling to ensure some solute is mixed in.

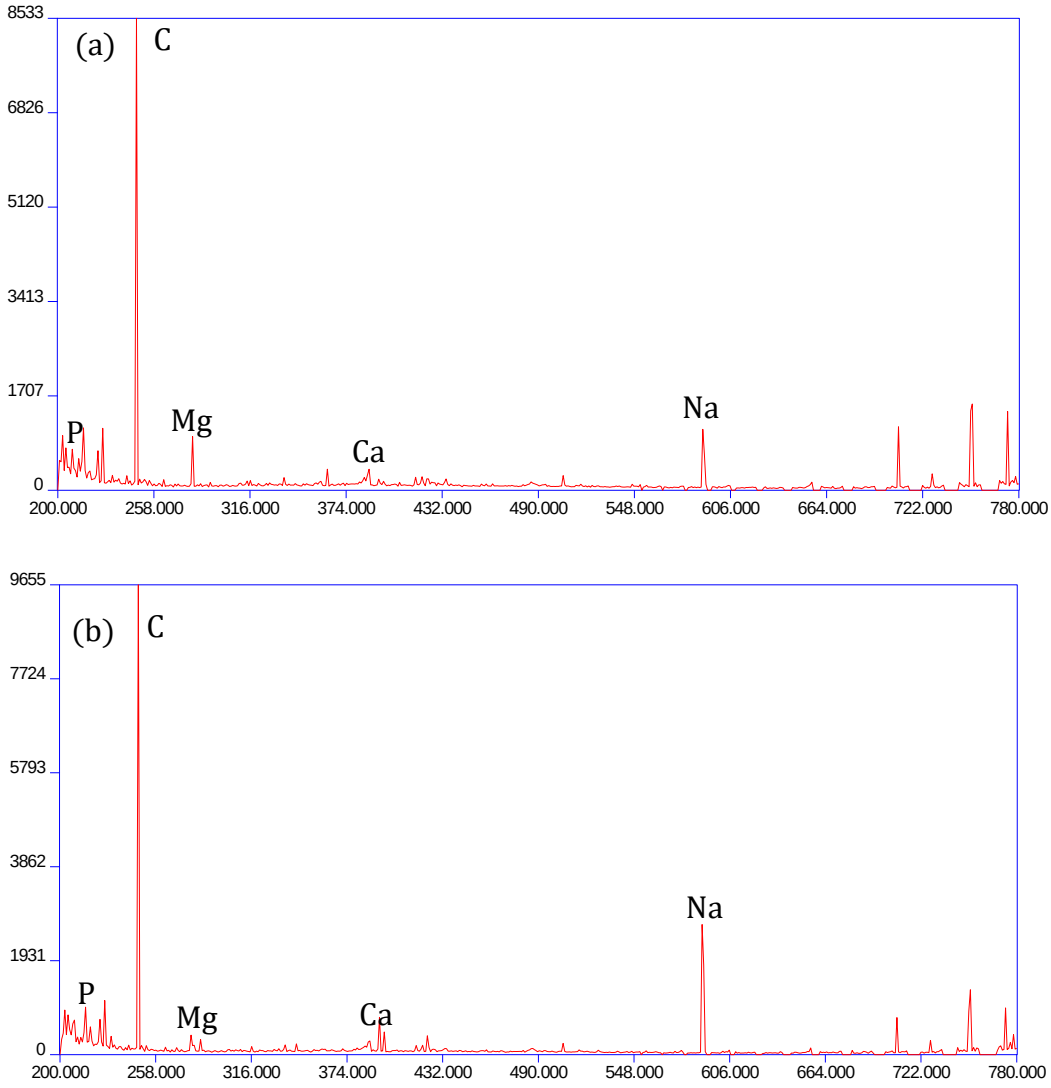

Figure 6.1: (a) Spectrum from sterile urine culture #3. The sodium and magnesium lines are visible. (b) Spectrum from sterile urine sample #1. The sodium and calcium lines are visible, and the sodium line is notably larger compared to spectrum (a). These urine samples highlight the variance between patients.

6.1.1 Detection of Bacteria in Sterile Urine

To determine if bacteria could be detected in sterile urine, the urine samples were 'spiked' with bacteria. This was accomplished by pipetting bacteria and urine into the same cone to create a solution of bacteria and urine. The centrifuge piece was assembled including filter and cone. 100 µL of a 1/5 bacterial suspension was pipetted into the cone. Then, 100 µL of a sterile urine sample was pipetted into the same cone. This effectively creates a mixture of bacteria in urine. The apparatus was centrifuged at 5000 RPM's for 5 minutes. Once centrifugation is complete the filter was removed and mounted on a steel

piece with double-sided sticky tape. All of the same settings were used when sampling with the LIBS apparatus.

The experimental procedure is simple and does not attempt to remove the bacteria from the urine. This is done for ease of use in the clinical setting, as less steps will require less time for diagnosis.

Figure 6.2 shows a spectrum of bacteria deposited in urine with a spectrum of urine overlaid to show the difference between the two. Calcium lines in bacteria are higher than that of urine, magnesium lines are relatively the same size, and sodium lines of bacteria are smaller than urine. Most notably, the phosphorus line in bacteria is higher than in urine, which indicates that bacteria are present. It is apparent from this spectrum that bacteria deposited in urine and urine are qualitatively different.

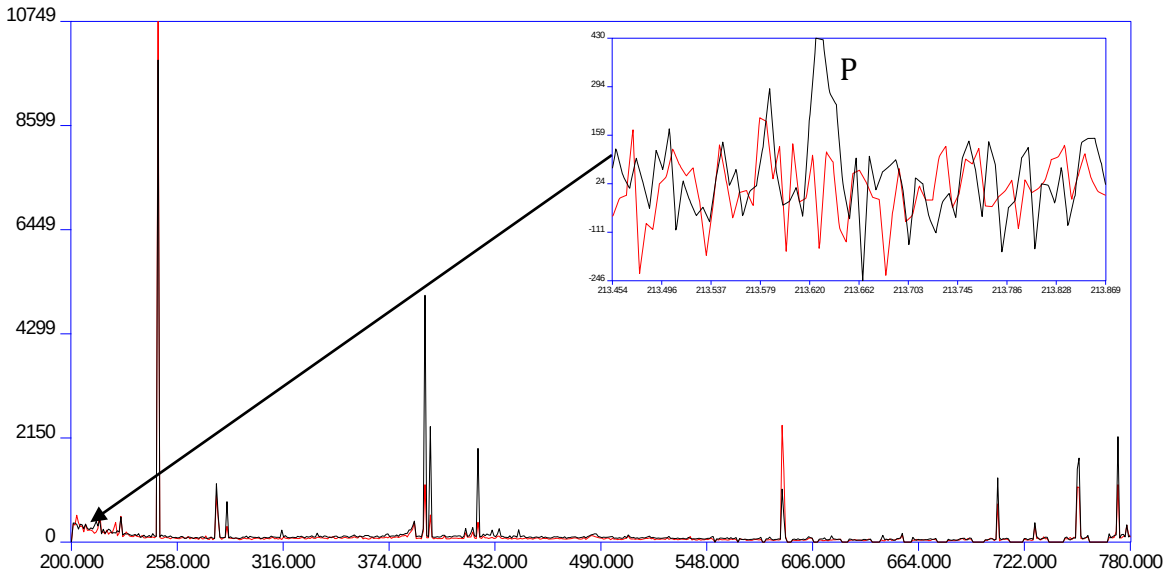

Figure 6.2: Overlaid spectra of urine (red) and bacteria in urine (black). The calcium lines are larger in the bacteria mixed with urine than in urine alone. The sodium lines are larger in urine. The magnesium lines of both the bacteria mixed with urine and urine alone are relatively the same. The phosphorus line of the bacteria mixed with urine is present, the phosphorus line in urine alone is absent; this is a major indicator of bacteria.

For the detection of bacteria in urine, PLSDA was used to build a model and test individual filters. RM2.5 was used to build the model in PLSDA. The model consisted of 240 total urine spectra which were listed as class 1, and 180 total bacteria in urine spectra that were listed as class 2. The bacteria spectra consisted of the species *S. aureus, E. coli,* and *E. cloacae.* Each species had two filters and 60 total single-shot spectra. The sensitivity and

specificity of the model in PLSDA without external validation is 100% and 100%, respectively. External validation was performed on the model by removing one filter at a time by entering it without any class information. This was performed for each filter of bacteria mixed with urine. The resulting sensitivity of each filter was calcualted, and the sensitivities of each filter was averaged to obtain an overall result. The same process was repeated for the urine filters to find the individual filters specificity and overall specificity. The average sensitvity for the bacteria in urine is 98.90%. The average specificity of the urine is 100%. This result is significant; it shows that PLSDA can detect the presence of any of the given bacterial species in urine when compared with urine that does not contain any bacteria. The results for individual fitlers is given in Table 6.1, and an example of an external validation in PLSDA is given in Figure 6.3.

Table 6.1: Sensitivity for each externally validated bacteria mixed with urine filter, specificity for each urine filter.

Externally Validated Filter	Sensitivity	Specificity
S. aureus 1/5 + Sterile Urine Filter #1	100%	--
S. aureus 1/5 + Sterile Urine Filter #2	100%	--
S. aureus 1/5 + Sterile Urine Filter #3	96.67%	
S. aureus 1/5 + Sterile Urine Filter #4	100%	--
E. coli 1/5 + Sterile Urine Filter #1	100%	
E. coli 1/5 + Sterile Urine Filter #2	100%	--
E. coli 1/5 + Sterile Urine Filter #3	100%	
E. coli 1/5 + Sterile Urine Filter #4	100%	
E. cloacae 1/5 + Sterile Urine Filter #1	93.33%	--
E. cloacae 1/5 + Sterile Urine Filter #2	90.00%	--
E. cloacae 1/5 + Sterile Urine Filter #3	100%	
E. cloacae 1/5 + Sterile Urine Filter #4	100%	
Sterile Urine Filter #1	--	100%
Sterile Urine Filter #2	--	96.67%
Sterile Urine Filter #3	--	93.33%
Sterile Urine Filter #4	--	100%
Sterile Urine Filter #5	--	93.10%
Sterile Urine Filter #6	--	100%
Sterile Urine Filter #7	--	100%
Sterile Urine Filter #8	--	100%

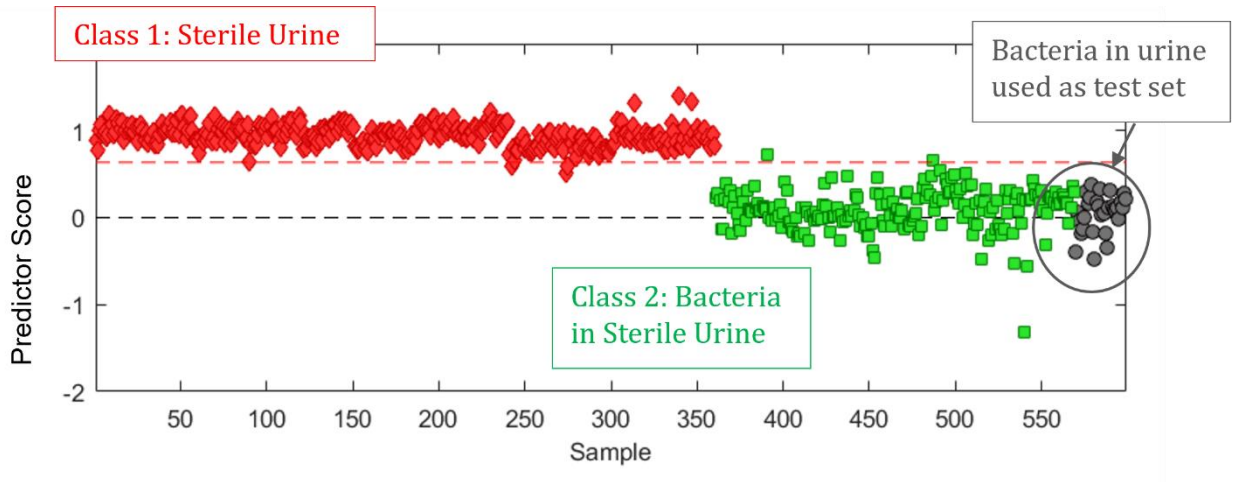

Figure 6.3: An example of an external validation in PLSDA. In this test, a filter of E. coli in urine was tested against the model.

6.1.2 Diagnosis of Bacteria in Sterile Urine

A DFA analysis was done using RM2.5 to determine if the 3 species of bacteria mixed with urine could be reliably classified separately. A total of 3 *S. aureus* filters with 90 single-shot spectra, 4 *E. coli* filters with 120 single-shot spectra, and 4 *E. cloacae* filters with 120 single-shot spectra were used. The results from the cross-validation in DFA are reported below in Table 6.2, and the DFA plot is shown in Figure 6.4. The specificity for all three species is high. The sensitivity for *S. aureus* is low compared to *E. coli* and *E. cloacae.*

Table 6.2: DFA results on 3 species in urine using RM2.5

RM2.5 With Normalized Data			
	S. aureus	*E. coli*	*E. cloacae*
Sensitivity	70.00 %	71.67 %	76.67 %
Specificity	86.67 %	86.67 %	85.83 %
Classification Error	9.58 %	5.65 %	3.27 %

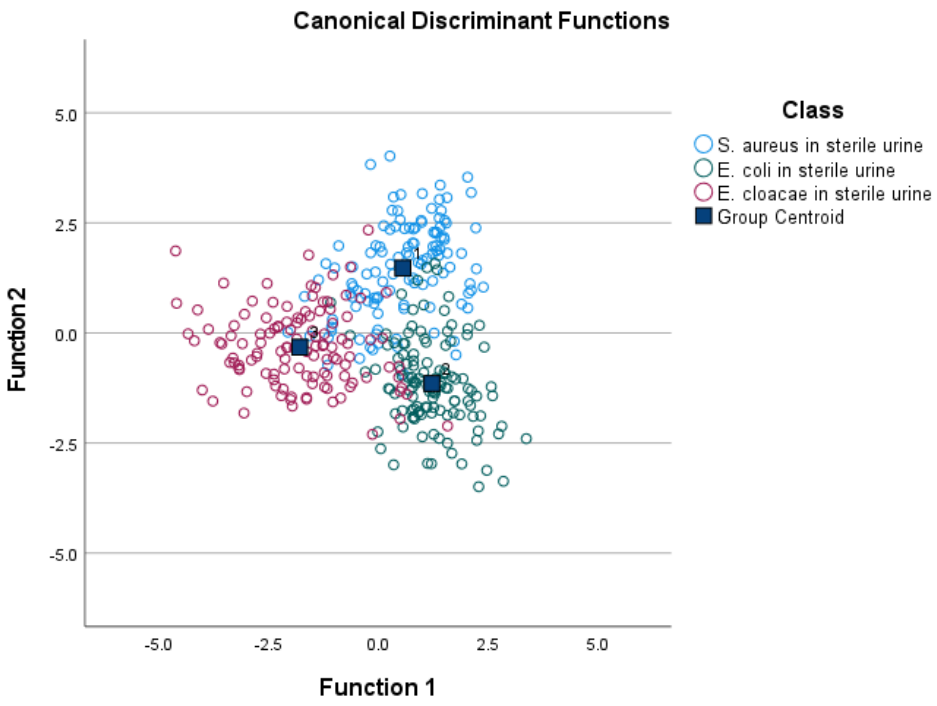

Figure 6.4: DFA plot between 3 species of bacteria mixed with urine.

To improve the sensitivity and specificity of the bacteria in urine, PCA-ANN on full spectrum data was investigated next. PCA was performed on 3 *S. aureus* filters with 120 single-shot spectra, 4 *E. coli* filters with 120 single-shot spectra, and 4 *E. cloacae* filters with 120 single-shot spectra. These were the same spectra as were used to obtain Figure 6.4. 10 PC scores were kept as it was previously determined in Chapter 5 that 10 was the optimal number of PC's. ANN on the 10 PCA scores yielded a higher sensitivity and specificity for all 3 species. The classification error also decreased markedly. The results are shown in Table 6.3.

Table 6.3: Results of PCA-ANN on full spectrum data of bacteria mixed with urine.

PCA-ANN With Full Spectrum Data			
	S. aureus	*E. coli*	*E. cloacae*
Sensitivity	100 %	100 %	91.67 %
Specificity	100 %	95.83 %	100 %
Classification Error	0.00 %	2.09 %	4.17 %

The model built to classify species with PCA-ANN can successfully and reliably discriminate between bacterial species. To further test the ANN, external tests were performed by splitting the data set into testing and training sets. As mentioned previously, our ANN algorithm program can randomly split data into training and testing sets to build our model. However, a clinician or operator will need to input the new data collected as testing data when it is collected from the patient. This was simulated by manually splitting our data into training and testing data after PCA scores are generated. Each filter was externally validated against the model and a sensitivity was calculated. The results are shown in Table 6.4, Table 6.5, and Table 6.6. There average sensitivity of *E. coli*, *S. aureus*, and *E. cloacae* are 75.83 %, 90.00 %, and 66.67 %, respectively. Since the model is already very small, removing these filters would be removing 25% of the data used for modelling; more data is needed to take this approach. However, removal of poorly performing filters should be investigated in the future after more data is collected based on the previous success demonstrated in Chapter 5.

Table 6.4: Sensitivity of each E. coli *filter removed from the model to be externally validated.*

E. coli	Predicted			
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	Sensitivity
1	0	28	2	93.333333
2	2	8	20	26.666667
3	0	25	5	83.333333
4	0	30	0	100
Sum	2	91	27	75.833333

Table 6.5: Sensitivity of each S. aureus *filter removed from the model to be externally validated.*

S. aureus	Predicted			
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	Sensitivity
1	30	0	0	100
2	18	12	0	60
3	30	0	0	100
4	30	0	0	100
Sum	108	12	0	90

Table 6.6: Sensitivity of each E. cloacae *filter removed from the model to be externally validated.*

E. cloacae	Predicted			
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	Sensitivity
1	1	4	25	83.333333
2	0	1	29	96.666667
3	0	7	23	76.666667
4	27	0	3	10
Sum	28	12	80	66.666667

6.2 LIBS on Sterile Blood

Sterile blood was provided to the Rehse lab by the pathology lab at the Windsor Regional Hospital Ouellette campus. Samples were provided by Lucy DiPietro and Dr. Mohamed El-Fakharany. Only blood specimens that had already been tested negative for bacterial infection were provided in these preliminary experiments. Blood specimens contained the anticoagulant sodium polyanetholesulfonate (SPS). Sterile blood was sampled and characterized with the LIBS system in the same manner as sterile urine to understand how the spectrum of blood will affect the spectrum of bacteria. Several different blood samples were characterized to account for the difference between several patients' physiology. To test sterile blood samples, the deposition procedure for urine was used.

Blood spectra from sterile blood cultures 4 and 6 are shown in Figure 6.5. These spectra show that blood, as with urine, contains sodium, magnesium, carbon, phosphate, and calcium. Most of these ions and minerals enter the bloodstream after digestion through the duodenum. Sodium is required in the blood to maintain osmotic pressure and is filtered through the kidneys into the urine. Sodium levels are regulated through the hormone aldosterone. Calcium is present in the blood due to bone desorption and resorption. The amount of calcium in the bone or blood is regulated through the hormones parathyroid hormone, calcitriol, and calcitonin. Phosphorus is present because the ion phosphate is used to build bones and teeth.[118] Most magnesium in the body is found in bones and cells, but some small amounts can be present in the blood.[119] The blood cultures presented in Figure 6.5 come from two different patients. It is apparent that there are differences between these spectra; the sodium lines in Figure 6.5a have higher intensity than the sodium lines in Figure 6.5b. The calcium lines in these two spectra are approximately the same intensity. The magnesium line in Figure 6.5a has a higher intensity than in 6.5b. As with the urine spectra, the differences between these spectra highlight the large variety that exists between patients and their physiologic systems. All other blood spectra showed relatively the same pattern of a more intense sodium line compared to all other lines, and also showed low amounts of calcium, magnesium, and phosphorus relative to other bacterial spectra. While the blood contributes to a non-zero background, the lines in the blood spectrum that interfere with the bacterial spectra are lower in intensity and are present in a different ratio than those in bacterial spectra. Therefore, it is expected that the blood will not interfere significantly with bacterial detection and diagnosis.

LIBS on blood was first performed by Melikechi et al. to characterize the blood spectrum. These studies were carried out on frozen whole blood samples from mice in a helium environment. They observed several iron and carbon peaks due to the hemoglobin, as well as calcium, magnesium, sodium, oxygen, potassium, nitrogen, and hydrogen, however no concentrations were calculated.[120] Most studies that use LIBS to study blood aim to detect and identify cancer markers. Markushin et al. used tag femtosecond LIBS in 2015 to detect cancer biomarkers in the blood to improve early detection of cancers. They determined that they could measure 0.01 U/mL of the cancer biomarker CA-125.[121] 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. They observed strong ion lines like calcium, sodium, and potassium, in the blood and weak lines of iron and magnesium. Using PCA, linear discriminant analysis (LDA), and k-nearest neighbours (kNN) classifications, they discriminated between samples from healthy patients and patients with lymphoma, finding that the algorithms performed with 99.7% accuracy and provided a diagnosis in 9 minutes. Discrimination between stages of lymphoma was attempted but unsuccessful.[122] Chen et al. also used LIBS to discriminate between lymphoma and multiple myeloma using whole blood samples. They observed that discrimination between 2 cancers was most accurate using kNN models, resulting in a sensitivity of 97.0% and a specificity of 95.6%.[123] Work on the diagnosis of melanoma using femtosecond LIBS on blood and tissue in mice was also done by Gaudiuso et al. in 2018. They attempted to diagnose cancer through direct analysis of spectra and by chemometric algorithms. They found that chemometric algorithms provided the most reliable results for detection of melanoma, with the Gradient Boosting performing best at 96.3% accuracy.[124] Blood serum was tested with LIBS for electrolyte concentration by Emara et al. to further characterize cancers. The presence or absence of electrolytes in the blood can provide a fuller picture of the disease, specifically the detection of hypokalemia, which can indicate later stage cancers. They found a significant difference in concentrations of potassium in the blood serum for healthy patients, patients with stage I colorectal cancer, and patients with stage II colorectal cancer.[125]

The body of work on detecting and diagnosing pathogens or illness present in the blood is smaller than the previously mentioned body of work on detection of cancer biomarkers in the blood. Omar Al-Jeffery et al. used laser-induced fluorescence (LIF) and LIBS for rapid detection of rubidium in blood to rapidly identify performance enhancing drugs. They were able to successfully detect trace levels of rubidium nitride at levels of 0.3%.[126] Gaudiuso et al. also investigated rapid diagnosis of gulf war illness using LIBS on blood samples. They were able to detect gulf war illness with a sensitivity of 100% and a specificity of 83.3%. However, these results came from a small scale study with few samples, so the authors have commented that more work needs to be done.[127] Finally, some work has been done on

detection of pathogens in blood by Wayua et al. and Multari et al. Wayua et al. focused on using LIBS on blood for detection of malaria. They used peak-free LIBS to detect the changes in concentrations of biometals that occur when infections are present in the body. They were able to show proof-of-concept of the predictive ability of ANN and PLS to determine the concentrations of Cu, Fe, and Zn. ANN provided accuracies of 73%, 68%, and 99%, respectively, while PLS provided accuracies of 70%. 68%, and 95%, respectively.[128] Multari et al. investigated the feasibility of detecting bacterial, viral, and parasite pathogens in the blood using LIBS. Using PLSDA and a decision tree algorithm, they were able to discriminate between the different types of biological pathogens with 96.6% accuracy and were able to detect blood containing pathogens 100% of the time.[129]

It should be noted that while these blood samples likely represent the diversity in patients, it is likely not entirely representative of what a fresh blood sample may look like. Blood samples are collected and stored in a fridge at 4°C for a period of a few months. Prolonged storage can cause *in vitro* hemolysis, meaning that there will be less blood cells present in a sample that has been stored for some period of time as opposed to a sample that's been collected more recently.[130] As well, *in vitro* hemolysis is also a problem in the clinic and can be caused by improperly collected samples, storage, and handling. This is particularly a problem in emergency departments where blood is collected quickly.[131] Before the blood can be given to our lab it must undergo a blood culture to ensure it is negative, which typically lasts 5 days. At the point that we receive the blood it is a week old and is likely already undergone some hemolysis. Blood can contain between 2 and 4 million cells per mL, and because of the extended time between collection of the blood sample and its use in the lab, it is unclear how many cells our samples contain. As well, the goal of clinical application is to sample the blood immediately after it has been drawn. The blood obtained from the pathology lab contains an anticoagulant to prevent the blood from clotting so it can be sampled and stored for extended periods of time. It is unclear how the absence of the anticoagulant will affect the spectra of the blood. The absence of the anticoagulant may decrease the level of sodium we observe in our spectra since the compound contains sodium. A possible solution to these issues is to separate the blood cells from the bacterial cells entirely, which will be covered in greater detail in Chapter 7.

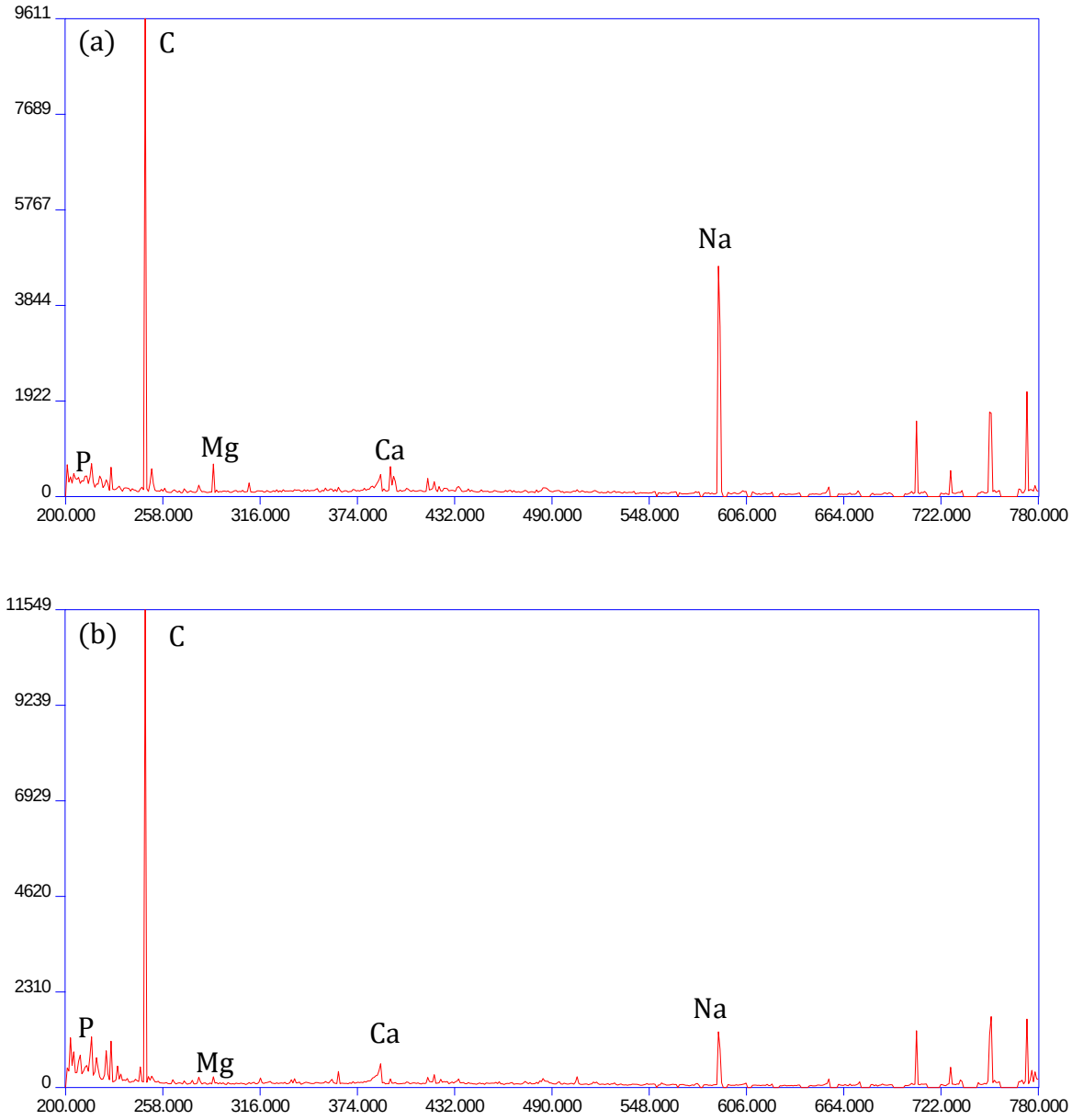

Figure 6.5: (a) Spectrum from negative blood culture 4. The sodium line has high intensity, and the calcium and magnesium lines are visible. (b) Spectrum from negative blood culture 6. The sodium line has a lower intensity compared to (a). The calcium line is visible.

6.2.1 Detection of Bacteria in Sterile Blood

For the detection of bacteria in blood, the blood samples were 'spiked' with bacterial suspensions using the same procedure as the urine. 100 µL of blood was pipetted into the plastic cone and centrifuge insert, followed by 100 µL of a 1/5 bacterial dilution, creating a mixture of bacteria and blood and therefore simulating a blood infection. The sample was

centrifuged at 5000 RPMs for 5 minutes. The filter was removed from the centrifuge insert and mounted on a steel piece with double sided sticky tape. To sample the bacteria and blood mixture, the same settings were used on the LIBS system as all previous bacterial samples.

This experimental procedure is simple and relies heavily on the difference between bacteria and blood. This methodology proved successful for us, but future work can be done to separate the blood cells from the bacterial cells for a more complex discrimination using a process called dual centrifugation.

A comparison between a blood spectrum and a bacteria mixed with blood spectrum is shown in Figure 6.6. Calcium, magnesium, and phosphorus lines in bacteria mixed with blood are higher than blood, and sodium lines of bacteria are smaller than blood. Most notably, the phosphorus line in bacteria is higher than in blood, which is an indicator that bacteria are present. It is apparent from this spectrum that bacteria mixed with blood and blood alone are qualitatively different.

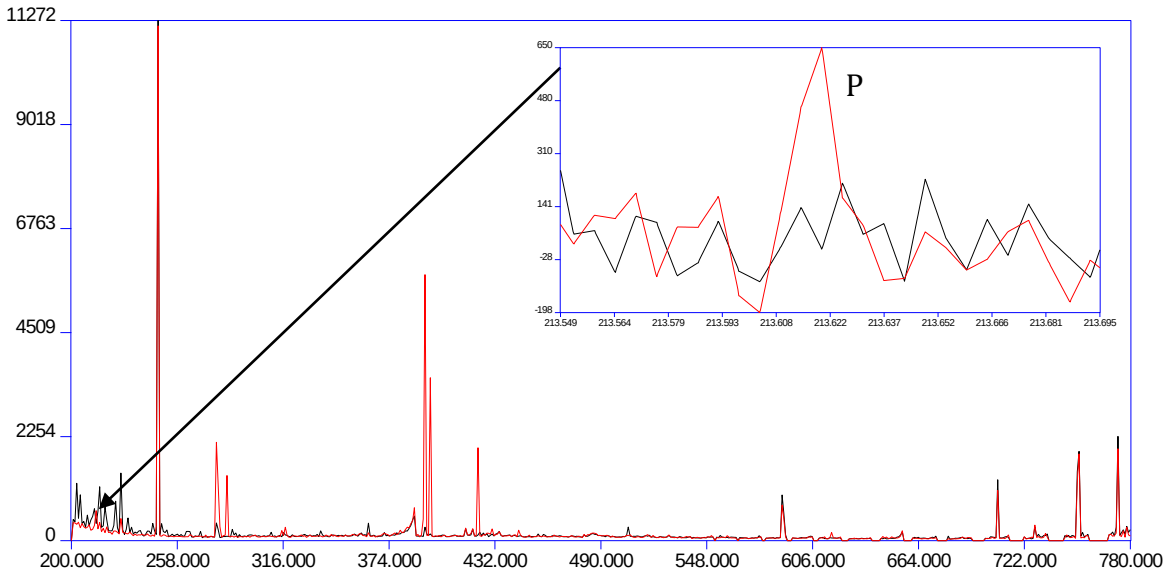

Figure 6.6: Overlaid spectrum of bacteria mixed with blood (red) and blood (black). The calcium, magnesium, and phosphorus line intensities are higher in the bacteria mixed with blood than in blood alone. The sodium line has a higher intensity in the blood alone than in the bacteria mixed with blood. The phosphorus line has a higher intensity in the bacteria mixed with blood than in blood alone.

For the detection of bacteria in blood, PLSDA was used along with RM2.5 to determine if spectra of bacteria mixed with blood were unique enough to classify separately from blood

alone. For this task 3 species mixed with blood, *E. coli*, *E. cloacae,* and *S. aureus,* were tested against blood alone. The blood class includes 206 individual spectra and 7 filters. For each species, there were 150 individual spectra of *E. cloacae*, 150 individual spectra of *E. coli*, 150 individual spectra of *S. aureus,* and 120 individual spectra of *P. aeruginosa.* The fourth species *P. aeruginosa* was added to the model because it is also a very relevant nosocomial pathogen. There are approximately 32000 cases of *P. aeruginosa* infections per year. These cases are typically caused by open wounds from surgery, ventilators, and catheters. There are approximately 2700 deaths per year from these infections.[132] A model was built using these single shot spectra and RM2.5, and the sensitivity and specificity of the model is 99.50% and 97.20%, respectively. The model was externally validated by removing one filter at a time from the model and re-entering it without any class information. This process was done for all filters containing blood mixed with bacteria. A sensitivity was calculated for each blood mixed with bacteria filter, and a specificity for each blood filter. The sensitivity and specificity is given for each filter in Table 6.7. An average sensitivity and specificity was calculated to be 96.31% and 98.57%, respectively. As with the PLSDA tests on urine, this result is significant because it suggests that any bacteria can be detected in this otherwise sterile fluid. Our technique is therefore sensitive to bacteria present in patient blood. An example of the external validation performed on a filter is shown in Figure 6.7 below.

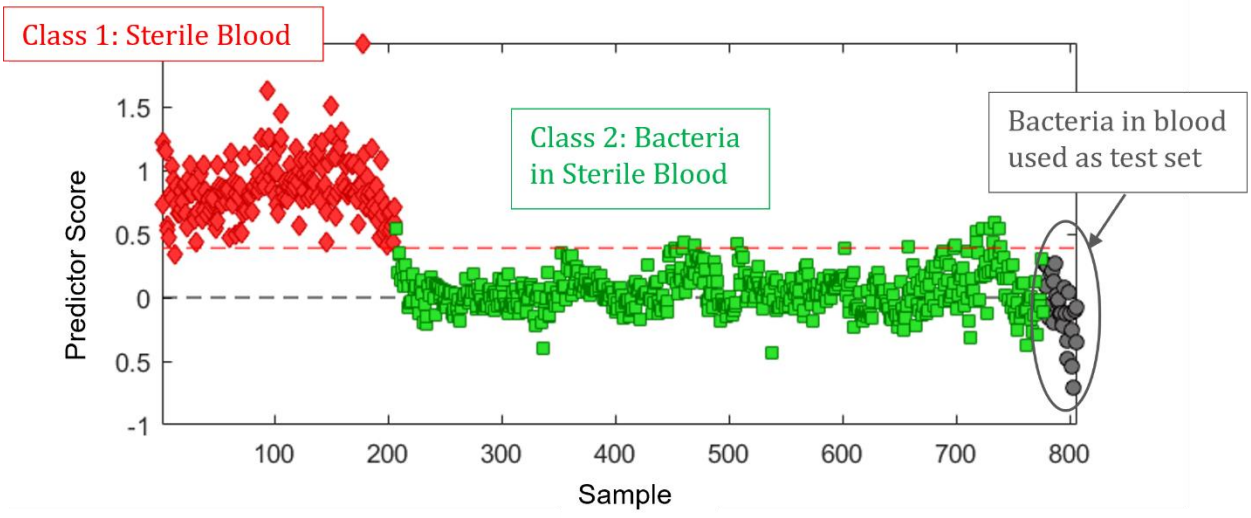

Figure 6.7: An example of external validation with PLSDA. In this test, a filter of P. aeruginosa *in blood is entered into the model without any class information. In this example, each spectrum classifies correctly as bacteria.*

Table 6.7: Sensitivity for each externally validated bacteria mixed with blood filters, specificity for each blood filter.

Externally Validated Filter	Sensitivity	Specificity
E. cloacae 1/5 + Sterile Blood Filter #1	96.67 %	--
E. cloacae 1/5 + Sterile Blood Filter #2	100 %	--
E. cloacae 1/5 + Sterile Blood Filter #3	100 %	--
E. cloacae 1/5 + Sterile Blood Filter #4	100 %	--
E. cloacae 1/5 + Sterile Blood Filter #5	100 %	--
E. coli 1/5 + Sterile Blood Filter #1	100 %	--
E. coli 1/5 + Sterile Blood Filter #2	100 %	--
E. coli 1/5 + Sterile Blood Filter #3	100 %	--
E. coli 1/5 + Sterile Blood Filter #4	80.00 %	--
E. coli 1/5 + Sterile Blood Filter #5	100 %	--
S. aureus 1/5 + Sterile Blood Filter #1	90.00 %	--
S. aureus 1/5 + Sterile Blood Filter #2	100 %	--
S. aureus 1/5 + Sterile Blood Filter #3	100 %	--
S. aureus 1/5 + Sterile Blood Filter #4	96.67 %	--
S. aureus 1/5 + Sterile Blood Filter #5	100 %	--
P. aeruginosa 1/5 + Sterile Blood Filter #1	90.00 %	--
P. aeruginosa 1/5 + Sterile Blood Filter #2	83.33 %	--
P. aeruginosa 1/5 + Sterile Blood Filter #3	93.33 %	--
P. aeruginosa 1/5 + Sterile Blood Filter #4	100 %	--
Sterile Blood Filter #1	--	93.33 %
Sterile Blood Filter #2	--	96.67 %
Sterile Blood Filter #3	--	100 %
Sterile Blood Filter #4	--	100 %
Sterile Blood Filter #5	--	100 %
Sterile Blood Filter #6	--	100 %
Sterile Blood Filter #7	--	100 %

6.2.2 Diagnosis of Bacteria in Sterile Blood

To determine if bacteria species present in blood could be distinguished from one another, a DFA analysis was done using RM2.5 on the 4 species of bacteria. 5 filters of *E. cloacae*, 5 filters of *E. coli*, 5 filters of *S. aureus*, and 4 filters of *P. aeruginosa* were tested in this model. Each class had 150 individual spectra. The results of the cross-validation between the 4 species are reported below in Table 6.8, and the DFA plot is shown in Figure 6.8.

Table 6.8: DFA results on 4 species mixed with blood using the RM2.5.

RM2.5 With Normalized Data				
	E. cloacae	*E. coli*	*S. aureus*	*P. aeruginosa*
Sensitivity	78.00 %	74.67 %	84.00 %	98.67 %
Specificity	92.00 %	93.56 %	92.89 %	100 %
Classification Error	15.00 %	15.89 %	11.56 %	0.66 %

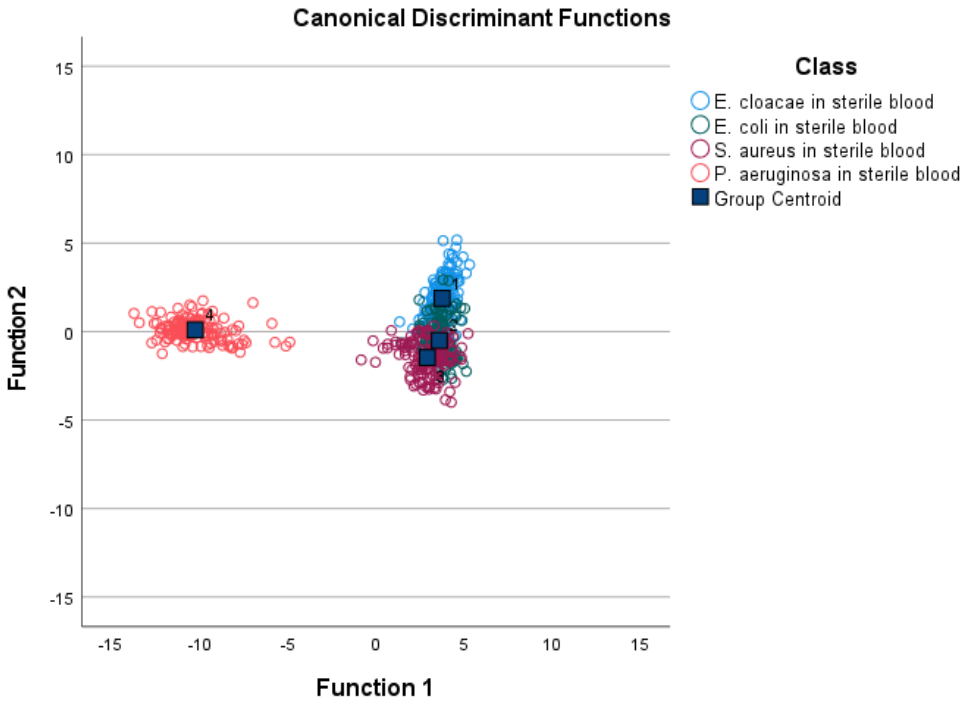

Figure 6.8: Classification of 4 species shown in DFA plot.

The sensitivity and specificity of *E. cloacae, E. coli,* and *S. aureus* are lower than *P. aeruginosa*, this is evident when examining the DFA plot. *P. aeruginosa* is clearly very

different from the other 3 species; it can be differentiated from them along discriminant function 1. The other 3 species are spread along discriminant function 2. The large spread along discriminant function 1 indicates most of the variance in the model is from the difference between *P. aeruginosa* and the other 3 species, and that there is lesser variance between the 3 species. While DFA can very clearly differentiate between *P. aeruginosa* and other bacteria, our method needs to be able to reliably detect and diagnose many other types of bacteria since the gram-negative enteric rods are also highly relevant to nosocomial blood infections.

To improve the differentiation between species, PCA-ANN was performed on full spectrum data of all the filters containing blood mixed with bacteria. 10 PC scores were generated using the PCA algorithm for each individual spectrum. These PC scores were used in the ANN algorithm. The number of hidden nodes used was 170 and the patience was 35. As with previous tests these numbers were chosen based on the ANN optimization algorithm. The results of this analysis are shown in Table 6.9. All filters classified perfectly in this test, so this model can successfully classify between different species present in blood with high accuracy. This model was therefore successful in reliably classifying species present in blood.

Table 6.9: Results of full-spectrum analysis in PCA-ANN.

PCA-ANN With Full Spectrum Data				
	S. aureus	*E. coli*	*E. cloacae*	*P. aeruginosa*
Sensitivity	100 %	100 %	100 %	100 %
Specificity	100 %	100 %	100 %	100 %
Classification Error	0.00 %	0.00 %	0.00 %	0.00 %

External validation of the model was tested again by removing whole filters from the model and inputting them without any class information. Filters from all 4 classes were tested using this external validation method, and the results are summarized in Tables 6.10, 6.11, 6.12, and 6.13. The average sensitivity for *E. coli*, *S. aureus*, *E. cloacae*, and *P. aeruginosa* are 80.67%, 65.33%, 92.67%, and 92.50%, respectively. Lower average sensitivity can be again largely attributed to filters that completely fail. Since the model is already very small, removing these filters would be removing 20% of the data used for

modelling; more data is needed to take this approach. Based on previous success, removal of poorly performing filters should be explored in the future.

Table 6.10: Sensitivity results for E. coli *filters removed from the model to be externally validated.*

E. coli	Predicted				
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	*P. aeruginosa*	Sensitivity
1	0	30	0	0	1
2	16	14	0	0	0.4666667
3	0	30	0	0	1
4	1	17	0	12	0.5666667
5	0	30	0	0	1
Sum	17	121	0	12	0.8066667

Table 6.11: Sensitivity results for S. aureus *filters removed from the model to be externally validated.*

S. aureus	Predicted				
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	*P. aeruginosa*	Sensitivity
1	30	0	0	0	1
2	1	0	29	0	0.0333333
3	30	0	0	0	1
4	30	0	0	0	1
5	7	23	0	29	0.2333333
Sum	98	23	29	0	0.6533333

Table 6.12: Sensitivity results for E. cloacae *filters removed from the model to be externally validated.*

E. cloacae	Predicted				
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	*P. aeruginosa*	Sensitivity
1	0	0	30	0	1
2	9	0	21	0	0.7
3	0	0	29	1	0.9666667
4	0	0	30	0	1
5	0	1	29	0	0.9666667
Sum	9	1	139	1	0.926667

Table 6.13: Sensitivity results for P. aeruginosa *filters removed from the model to be externally validated.*

P. aeruginosa	Predicted				
Sample #	*S. aureus*	*E. coli*	*E. cloacae*	*P. aeruginosa*	Sensitivity
1	0	1	0	29	0.9666667
2	8	0	0	22	0.7333333
3	0	0	0	30	1
4	0	0	0	30	1
Sum	8	1	0	111	0.925

The improved PCA-ANN performance over the DFA performance for both urine and blood can likely be explained by the non-linearity of the ANN model.[133] As stated in Chapter 5, DFA functions best on data that has a linear relationship between predictor variables and the resulting scores. Shown in Figure 6.9 is a similar plot that was shown in Chapter 5; a comparison between 2 independent variables is shown to demonstrate the lack of linear separation between the 3 classes. The non-linear nature of this data indicates that ANN will be better suited to classification in DFA. Figure 6.10 shows the same comparison between 2 independent variables. However, in this figure, there is a clear separation between *P. aeruginosa* and all other classes. This corroborates the finding that DFA can reliably discriminate between *P. aeruginosa* and all other species of bacteria. This clear separation is a good example of linearly separable data. Figure 6.11 shows a 3-class comparison with the blood and bacteria data to further demonstrate the need for ANN when discriminating between blood samples. While DFA works well on *P. aeruginosa*, ANN is needed for *E. coli*, *E. cloacae*, and *S. aureus.* As with the data in Chapter 5, PCA coupled with ANN also improves performance because PCA can find what contributes the most to differences between classes. PCA also reduces dimensionality which avoids redundancy in data.[134]

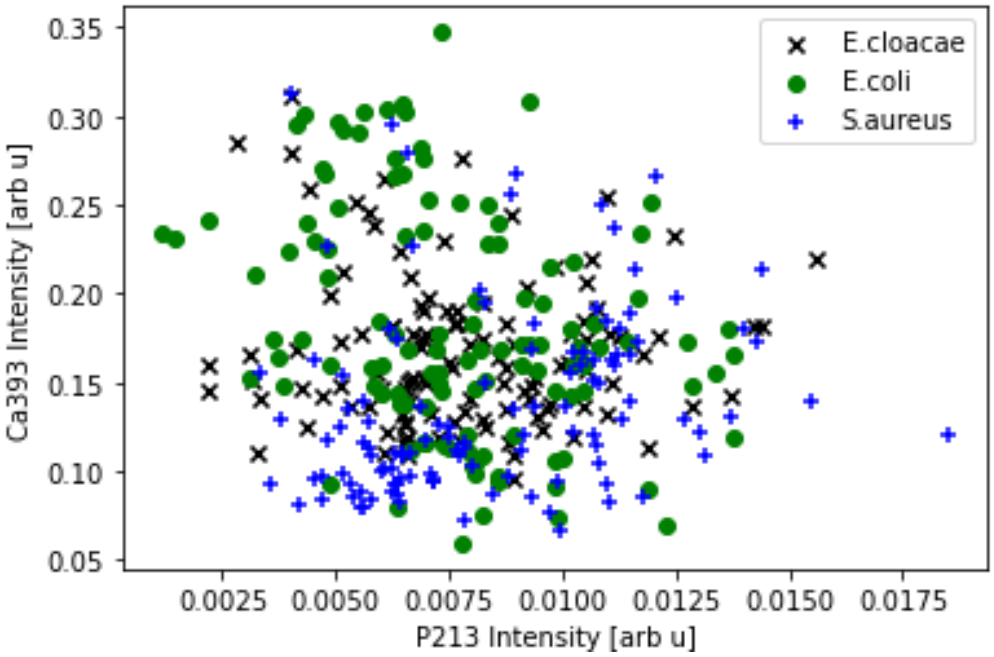

Figure 6.9: Scatter plot comparing independent variables phosphorus 213.618 nm and calcium 393.366 nm for all bacteria in urine data. There is no visible separation between classes.

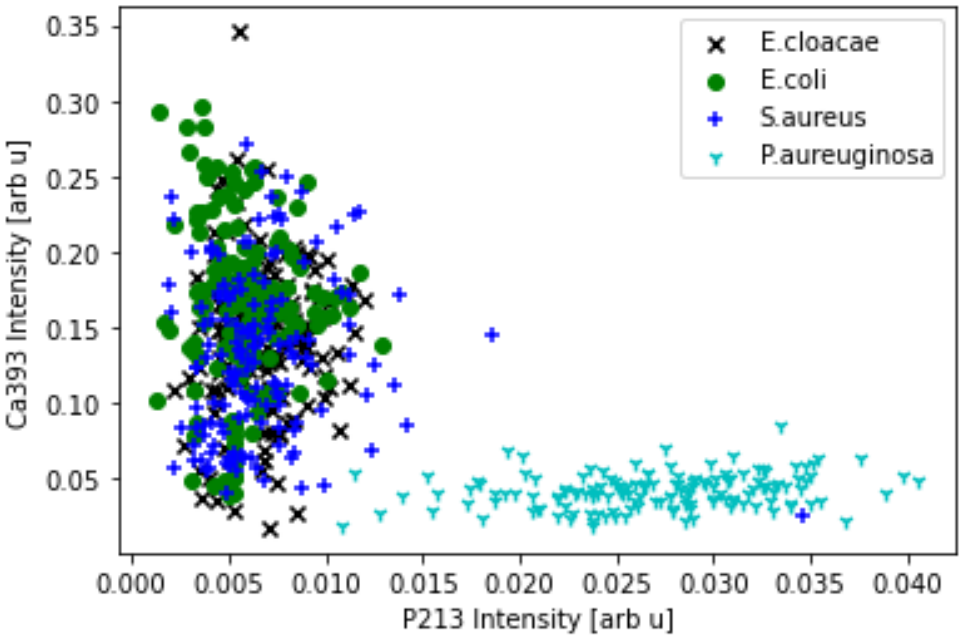

Figure 6.10: Scatter plot comparing independent variables phosphorus 213.618 nm and calcium 393.366 nm for all bacteria in blood data. There is no visible separation between E. coli, E. cloacae, *and* S. aureus, *but there is separation between* P. aeruginosa *and all other classes. This likely explains the high performance of* P. aeruginosa *in DFA and poor performance of the other 3 species.*

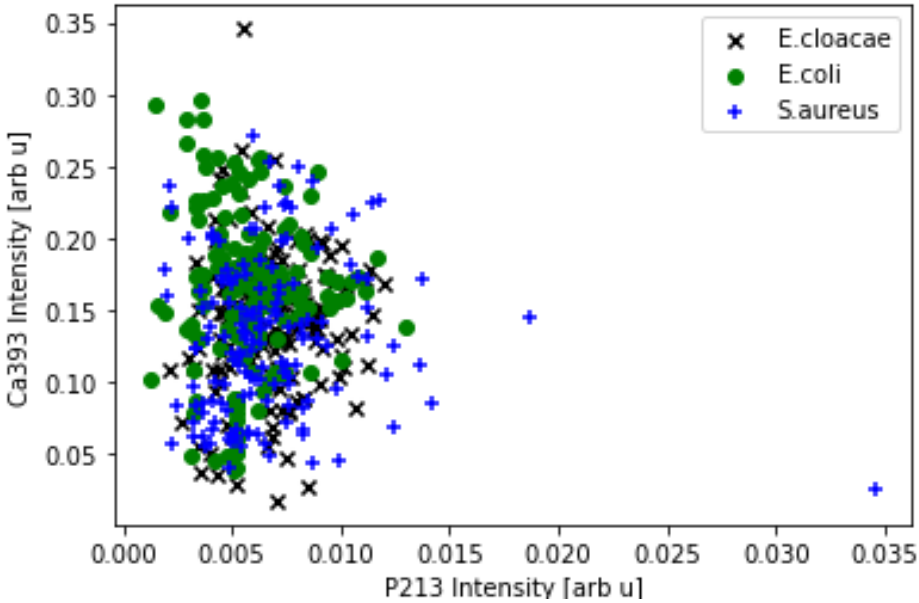

Figure 6.11: Scatter plot comparing independent variables phosphorus 213.618 nm and calcium 393.366 nm for bacteria in blood data. This plot demonstrates the need for ANN on the 3 species modelled.

6.3 Conclusions and Future Work

Sterile blood and urine deposited on nitrocellulose filters were characterized using LIBS. It was found that they both contain carbon, sodium, and trace amounts of magnesium and calcium. A deposition method was developed for sterile blood and sterile urine spiked with bacteria on these nitrocellulose filters. Using PLSDA, infected urine can be reliably distinguished from sterile urine, and infected blood can be reliably distinguished from sterile blood. To classify species present in bacteria and blood, PCA-ANN on full spectrum data produced the most reliable and accurate results.

Future work should be done to clarify the limit of detection of bacteria in blood and urine by investigating detection and classification ability of more dilute suspensions of bacteria. Future work should also focus on more closely mimicking clinical conditions that the blood and urine will be tested in. Fresh samples should be obtained and tested immediately if possible. As well, a greater understanding of how infections disperse and behave in the blood and urine samples that are initially drawn from patients will help us to understand the true clinical conditions that this technique will be implemented in. Understanding these clinical conditions will help us more accurately simulate them. Future work should also focus on improving the external validation results in blood and urine. Improvements to these results can likely be made by improving the repeatability of the deposition procedure, which is discussed in greater detail in Chapter 7.

[108] World Health Organization. (2002). Prevention of hospital-acquired infections: a practical guide, 2nd ed. G. Ducel, J. Fabry and L. Nicolle (Eds.). World Health Organization.

[109] Barza, M. (2007). Urinary Tract. In N. C. Engleberg, V. J. DiRita, T. Dermody, M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 564–572). Chapter, Lippincott Williams & Wilkins.

[110] Wolfe, A. J., Brubaker, L. (2015). "Sterile Urine" and the Presence of Bacteria. *European urology*, *68*(2), 173–174. https://doi.org/10.1016/j.eururo.2015.02.041.

[111] Rehse, S. J., Mohaidat, Q. I., Palchaudhuri, S. (2010). Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: the effect of mixed cultures and sample dilution on bacterial identification. *Applied Optics*, *49*(13), C27-C35. https://doi.org/10.1364/AO.49.000C27.

[112] Whitnack, E. (2007). Sepsis. In N. C. Engleberg, V. J. DiRita, T. Dermody, M. Schaechter (Eds.), *Mechanisms of microbial disease* (3rd ed., pp. 564–572). Chapter, Lippincott Williams & Wilkins.

[113] Jaswal, B.B. S., Singh, V. K. (2015). Analytical assessments of gallstones and urinary stones: A comprehensive review of the development from laser to LIBS. *Applied Spectroscopy Reviews,* *50*(6), 473-498. https://doi.org/10.1080/05704928.2015.1010206.

[114] Štěpánková, K., Novotný, K., Vašinová Galiová, M.V., Kanický, V., Kaiser, J., Hahn, D.W. (2013). Laser ablation methods for analysis of urinary calculi: Comparison study based on calibration pellets. *Spectrochimica Acta Part B,* *81*, 43-49. https://doi.org/10.1016/j.sab.2012.12.009.

[115] Anzano, J., Lasheras, R. J. (2009). Strategies for the identification of urinary calculus by laser induced breakdown spectroscopy. *Talanta* *79*(2), 352-360. https://doi.org/10.1016/j.talanta.2009.03.065.

[116] Mohaidat, Q. I., Sheikh, K., Palchaudhuri, S., Rehse, S. J. (2012). Pathogen identification with laser-induced breakdown spectroscopy: the effect of bacterial and biofluid specimen contamination. *Applied optics, 51*(7), B99–B107. https://doi.org/10.1364/AO.51.000B99.

[117] Sarigul, N., Korkmaz, F., Kurultak, İ. (2019). A new artificial urine protocol to better imitate human urine. *Scientific reports*, *9*(1), 20159. https://doi.org/10.1038/s41598-019-56693-4.

[118] Biga, L. M. (2019, September 26). *26.3 Electrolyte Balance – Anatomy & Physiology*. Pressbooks. https://open.oregonstate.education/aandp/chapter/26-3-electrolyte-balance/.

[119] Workinger, J. L., Doyle, R. P., Bortz, J. (2018). Challenges in the diagnosis of magnesium status. *Nutrients*, *10*(9), 1202. https://doi.org/10.3390/nu10091202.

[120] Melikechi, N., Ding, H., Rock, S., Marcano O, A., Connolly, D. (2008). Laser-induced breakdown spectroscopy of whole blood and other liquid organic compounds. *Optical Diagnostics and Sensing VIII, 6863*, 152-158. https://doi.org/10.1117/12.761901.

[121] Markushin, Y., Sivakumar, P., Connolly, D., Melikechi, N. (2015). Tag-femtosecond laser-induced breakdown spectroscopy for the sensitive detection of cancer antigen 125 in blood plasma. *Analytical and Bioanalytical Chemistry, 407,* 1849-1855. https://doi.org/10.1007/s00216-014-8433-0.

[122] Chen, X., Li, X., Yang, S., Yu, X., Liu, A. (2018). Discrimination of lymphoma using laser-induced breakdown spectroscopy conducted on whole blood samples. *Biomedical Optics Express, 9(3),* 1057-1068. https://doi.org/10.1364/BOE.9.001057.

[123] Chen, X., Li, X., Yang, S., Yu, Chen, D., Liu, A. (2018). Diagnosis of human malignancies using laser-induced breakdown spectroscopy in combination with chemometric methods. *Spectrochimica Acta Part B, 139,* 63-69. https://doi.org/10.1016/j.sab.2017.11.016.

[124] Gaudiuso, R., Ewusi-Annan, E., Melikechi, N., Sun, X., Liu, B., Felipe Campesato, L., Merghoub, T. (2018). Using LIBS to diagnose melanoma in biomedical fluids deposited on solid substrates: Limits of direct spectral analysis and capability of machine learning. *Spectrochimica Acta Part B, 146,* 106-114. https://doi.org/10.1016/j.sab.2018.05.010.

[125] Emara, E.M., Song, H., Imam, H., Elwekeel, W. M., Gao, X., Mohammed, M. M., Liu, S. (2022). 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*, 1081–1093. https://doi.org/10.1007/s10103-021-03355-5.

[126] Omar Al-Jeffery, M., Telle, H. H. (2002). LIBS and LIF for rapid detection of Rb traces in blood. *Optical Biopsy, 4613.* https://doi.org/10.1117/12.465241.

[127] Gaudiuso, R., Chen, S., Kokkotou, E., Conboy, L., Jacobson, E., Hanlon, E. B., Melikechi, N. (2021). Diagnosis of gulf war illness using laser-induced breakdown spectra acquired from blood samples. *Applied Spectroscopy, 76*(8), 887-893. https://doi.org/10.1177/00037028211042049.

[128] Wayua, D. M., Angeyo, H. K., Dehayem-Kamadjeu, A., Kaduki, K. A. (2022). 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). *Analytical Letters, 55*(17), 2669-2682. https://doi.org/10.1080/00032719.2022.2067862.

[129] Multari, R., Cremers, D. A., Nelson, A., Karimi, Z., Young, S., Fisher, C., Duncan, R. (2019). The use of laser-based diagnostics for the rapid identification of infectious agents in human blood. *Applied Microbiology, 126,* 1606-1617. https://doi.org/10.1111/jam.14222.

[130] Wan Azman, W. N., Omar, J., Koon, T. S., Tuan Ismail, T. S. (2019). Hemolyzed specimens: Major challenge for identifying and rejecting specimens in clinical laboratories. *Oman Medical Journal*, *34*(2), 94–98. https://doi.org/10.5001/omj.2019.19.

[131] Authority, P. P. S. (n.d.). *In Vitro Hemolysis: Delays May Pose Safety Issues | Advisory*. Pennsylvania Patient Safety Authority. Retrieved September 19, 2022, from http://patientsafety.pa.gov/ADVISORIES/Pages/200706_64.aspx.

[132] Center for Disease Control and Prevention. (2019). *Antibiotic Resistant Threats in the United States, 2019*. U.S. Department of Health and Human Services. https://doi.org/10.15620/cdc:82532.

[133] Kumar, A. (2022, July 31). *Linear vs Non-linear Data: How to Know*. Data Analytics. Retrieved September 15, 2022, from https://vitalflux.com/how-know-data-linear-non-linear/#:%7E:text=for%20Regression%20Problem-,Use%20Scatter%20Plots%20for%20Classification%20Problems,scatter%20plots%20representing%20different%20classes.https://vitalflux.com/how-know-data-linear-non-linear/#:%7E:text=for%20Regression%20Problem-,Use%20Scatter%20Plots%20for%20Classification%20Problems,scatter%20plots%20representing%20different%20classes.

[134] Kinhal, V. (2022, August 30). *All About ANN – Artificial Neural Networks and Chemometric Modeling for NIR Spectroscopy*. Felix Instruments. Retrieved September 15, 2022, from https://felixinstruments.com/blog/all-about-ann-artificial-neural-networks-and-chemometric-modeling-for-nir-spectroscopy/.

Chapter 7: Dual Stage Centrifugation of Bacteria in Clinical Specimens

Each millilitre of blood contains approximately 4 to 6 million cells and this large amount of cells may cause an increased carbon or phosphorus signal that would not be attributable to the bacterial cells in the blood, or may cause a false positive result.[135] In Chapter 6 it was found that blood has a non-zero contribution to the spectrum, with a visible sodium line, and smaller calcium and magnesium lines. However, as noted in Chapter 6, this may not be entirely representative of a fresh clinical sample, due to the high volume of blood cells that may be present upon sampling. For clinical application then, it may be necessary to filter out any blood cells that are in the blood sample taken. This motivated the study of dual stage centrifugation, with the aim of separating blood cells from bacterial cells using our current deposition procedure. Dual stage centrifugation could be a viable option for separating out blood cells from bacterial cells because blood cells are approximately 6 to 8 μm in diameter, whereas bacteria range from 0.5 to 2 μm in length.[136] This difference could allow us to separate out the larger eukaryotic human cells from the smaller prokaryotic bacterial cells by size.

During the study of dual centrifugation, it was found through several experiments described here that the deposition of bacterial cells upon our filters was not occurring nearly as consistently as anticipated or as required for reproducible data. The experiments below will demonstrate two substantial experimental problems that occurred during centrifugation deposition: the seal between the cone and the filter not working and the bacteria somehow going around the filter instead of resting on the filter deposition site. This experimental inconsistency is likely responsible for the scatter in Figure 5.4, but was not discovered until well after those experiments were completed.

7.1 Dual Stage Centrifugation Methods

To investigate dual stage centrifugation, a 2-step process was developed. Ideally, dual stage centrifugation will be performed in 1 step in the clinic. This can be accomplished using the centrifuge insert, which is shown in Figure 7.1.[137] The centrifuge insert was designed for single and dual stage applications; however, it was designed before the use of the cone. Therefore, concentration can only be achieved on the first filter with the current

design. However, larger cells or material would need to be filtered out first, followed by smaller cells. The current design is not conducive to the concentration of cells onto the bottom filter. For future application of the insert in a 1 step process, a new piece will have to be designed for concentrating bacteria onto the bottom filter.

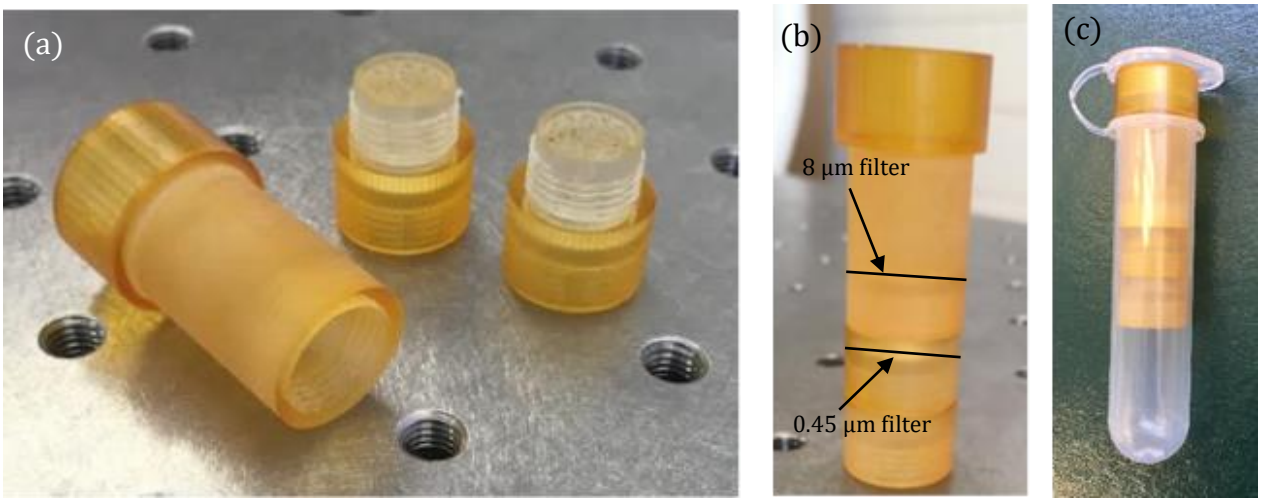

Figure 7.1: (a) Centrifuge insert tube (left) and bottom pieces (right). All pieces are screwed together for dual-stage centrifugation, as shown in (b) and (c). Black lines in (b) show where the filters would be placed in dual-stage centrifugation. Figure adapted from [137].

The 2-step process used to deposit bacteria is illustrated in Figure 7.2. First, 100 μL of a 1/5 bacterial suspension is pipetted into a centrifuge insert with no cone. The filter used for this has a larger pore size than the filter normally used to deposit bacteria. For this experiment, 8 μm filters were used. These filters will catch the larger blood cells, and this experiment will validate the thought model that bacteria will go through the filters with the larger pores. The sample is then centrifuged. After centrifugation, the tube is vortexed and the sample was removed from the bottom. The sample is then deposited into another centrifuge insert containing a cone and a 0.45 μm filter, which is centrifuged. The 0.45 μm filter is then sampled using LIBS to determine if bacteria went through the filter with the larger pore size and were deposited on the 0.45 μm filter. 3 species were tested with dual stage centrifugation: *E. coli, E. cloacae,* and *S. aureus.* To quantify how much bacteria got through the 8 μm filter, total spectral intensities were compared to filters of the same species that underwent single stage centrifugation.

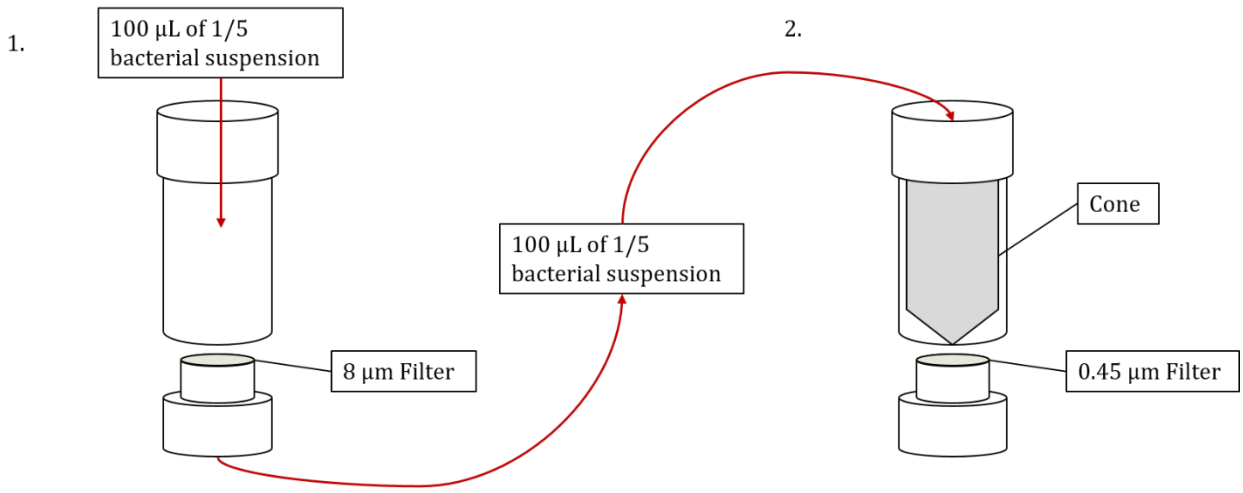

Figure 7.2: Procedure for dual stage centrifugation. 1/5 suspension of bacteria is pipetted into a centrifuge insert and deposited onto a filter of large pore size, in this case an 8 µm filter. The sample is centrifuged, removed from the bottom of the centrifuge tube, and pipetted into an insert containing the cone and a 0.45 µm filter.

To compare the spectral intensities, an average total spectral intensity was calculated for each species as well as the standard deviation on the mean. Because our data is highly noisy, the 10% highest and 10% lowest total spectral intensities were removed from each species, and an average was taken of the middle 80% of data. The averages of these species and the standard deviation on the mean for each are given in Table 7.1. A total of 3 trials were taken of the dual stage centrifugation filters and an average total spectral intensity was computed for each. These results are shown in Table 7.2. The average for the 3 trials was calculated and compared to the average for each species. The total spectral power was within error for *S. aureus*, slightly outside of the error for *E. coli*, and *E. cloacae* showed the largest difference between dual stage centrifugation and single stage centrifugation, falling far outside the error. Figure 7.3 shows a comparison of the total spectral power between single centrifugation and dual centrifugation for each species. Figure 7.4 shows the comparison between the average for dual and single stage centrifugation for all 3 species.

To further compare the single and dual stage centrifugation, a ratio of total spectral power was taken between dual centrifugation and single centrifugation. These ratios represent approximately the percentage of bacteria that went through the 8 µm filter and were deposited on the 0.45 µm filter. These ratios are shown in Table 7.3. These percentages are an estimate of how much bacteria comes through based on an average taken from regular samples. Based on the ratios and analysis of the figures, it can be

concluded that there are bacteria coming through the 8 μm filter, though it cannot be concluded with certainty if all cells come through the 8 μm filter. *S. aureus* and *E. coli* appear to have a higher percentage of cells coming through the 8 μm than *E. cloacae.* Since *E. coli* and *E. cloacae* are generally the same size and shape, the reason for this discrepancy is unclear at this time. There is some experimental error in these numbers, since we cannot control how much bacteria we pick up each time we make a sample, and we cannot control how each filter ablates due to the irregularity of the sample.

Table 7.1: Average total spectral power after single stage centrifugation for 3 species of bacteria, standard deviation, and the standard deviation on the mean for each.

	Average Total Spectral Intensity (arb u)	σ	σ_{Mean}
E. coli	17790.37	7050.558	371.597
E. cloacae	20901.48	12651.55	646.464
S. aureus	17034.95	6823.637	345.086

Table 7.2: Average total spectral power after dual stage centrifugation for 3 species of bacteria.

	Average Total Spectral Intensity (arb u)			
	Trial #1	Trial #2	Trial #3	Average
E. coli	12654.97	14663.27	22513.43	16610.56
E. cloacae	15520.93	20481.43	17695.00	18019.12
S. aureus	22280.30	15187.70	15497.87	17655.29

Table 7.3: Ratio of total spectral power of dual stage centrifuged samples to single stage centrifuged samples for 3 species.

	Percentage of Bacteria Deposited on Second Filter			
	Trial #1	Trial #2	Trial #3	Average
E. coli	71.13 %	82.42 %	126.55 %	100.02 %
E. cloacae	74.25 %	99.71 %	84.66 %	86.21 %
S. aureus	130.79 %	89.16 %	90.98 %	103.64 %

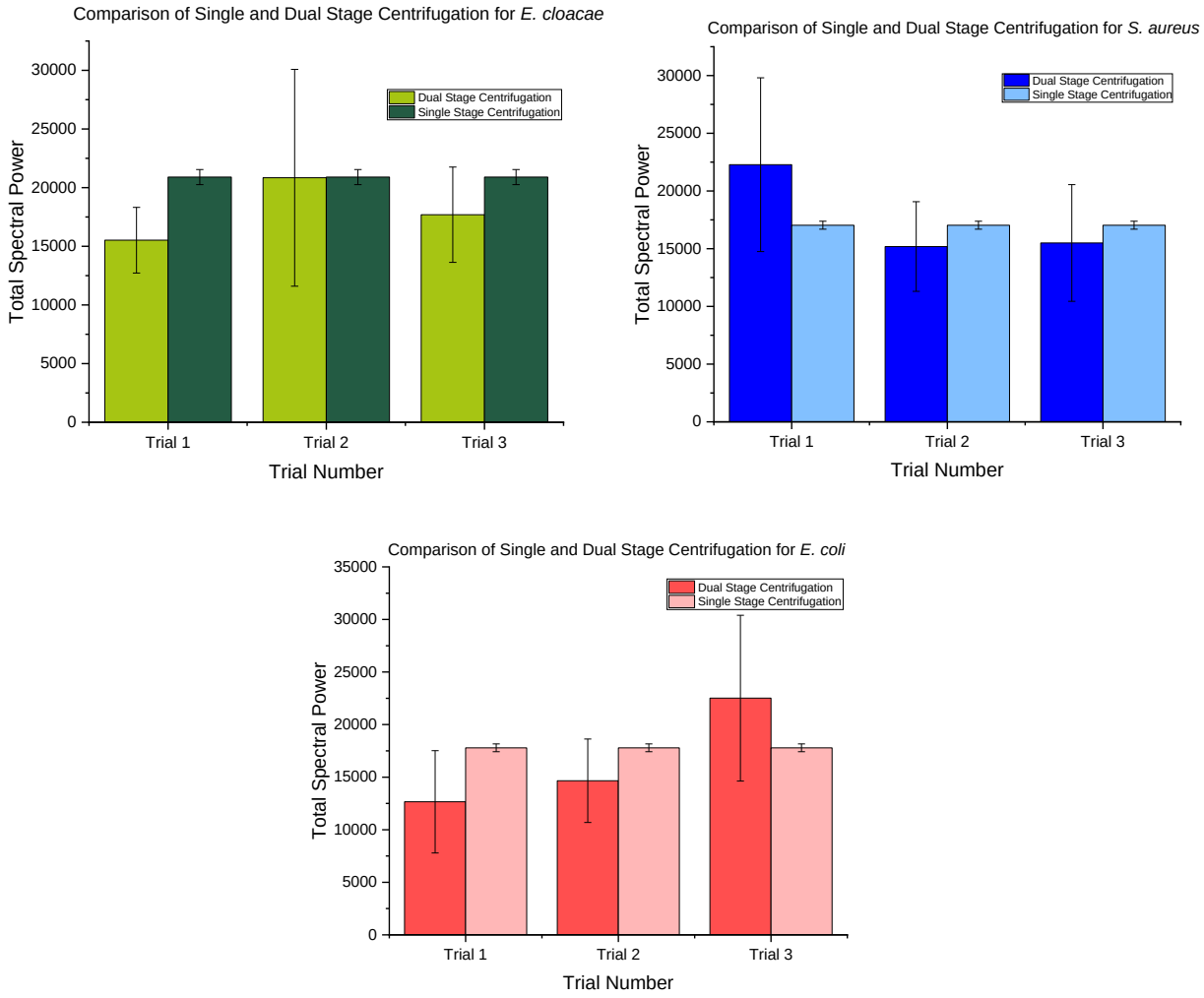

Figure 7.3: Bar graphs comparing summed intensities of 15 lines for single and dual stage centrifugation. 3 species are compared: (a) E. cloacae, *(b)* S. aureus, *(c)* E. coli. *The standard deviation of the dual stage filters is compared to the standard deviation on the mean of the single stage centrifugation.*

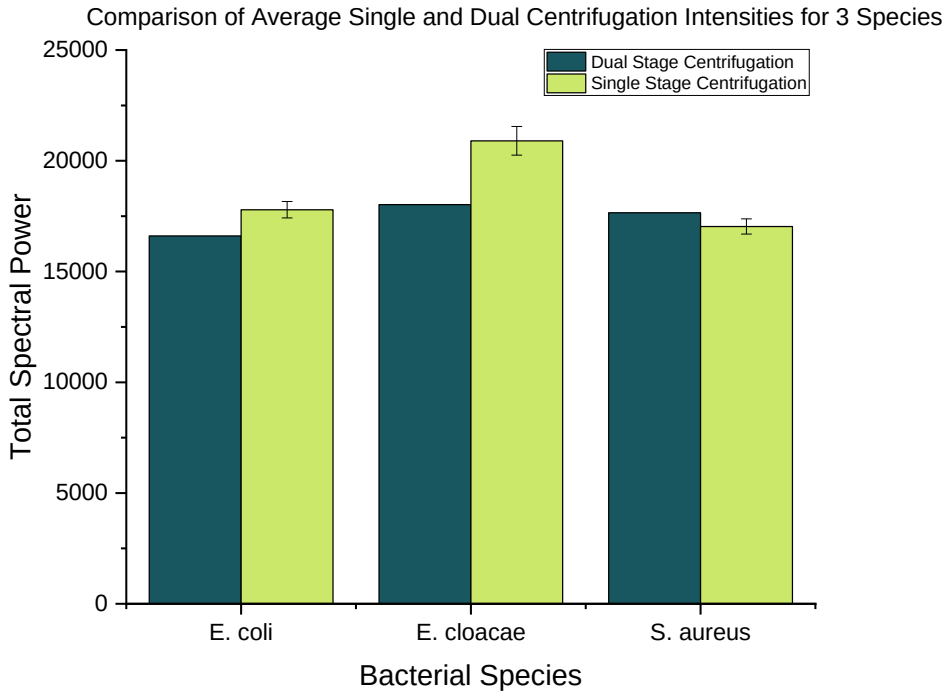

Figure 7.4: Comparison between average total spectral intensity for dual and single stage centrifugation for E. coli, E. cloacae, *and* S. aureus. *Dual stage centrifugation of* S. aureus *is within error of single stage centrifugation, dual centrifugation of* E. coli *is slightly outside of the single stage centrifugation error, and dual stage centrifugation of* E. cloacae *is lower than single stage centrifugation.*

There is currently some uncertainty with how the bacteria are deposited onto the filter during centrifugation. It is unclear if the elemental signature detected originates from ions washed off of the cell surface or released osmotically or via lysis, or if there are intact bacterial cells present. Often, the phosphorus line is a hallmark of the presence of bacteria. To verify that bacteria are being deposited as opposed to elements and ions from solution, the intensity of the most reliable phosphorus line was tracked next. As well, our understanding of how the filters worked also needed to be verified. Our original belief was that the small cells will completely pass through the 8 μm filter, and all will be caught on the 0.45 μm filter. To verify that we understand how the filters worked, the same procedure outlined in Figure 7.2 was repeated using a 0.45 μm filter in both the first and second stage. This experiment should produce a null result when the second filter is tested with LIBS because all of the bacterial cells should be caught on the top filter as the pore size is smaller than the cells. *E. coli, E. cloacae,* and *S. aureus* were tested with this method. For this experiment, the intensity of the phosphorus line was tracked to ensure that we were observing bacteria, as opposed to ions being washed off the bacterial cell surface.

First the 8 μm filter data was re-analyzed to determine the phosphorus line intensity after passage through the filter and compared to the single centrifugation data. These results are shown in Tables 7.5, and average phosphorus line intensities for each of the 3 species after single stage centrifugation are shown in Table 7.4. The ratios were calculated between the dual and single stage centrifugation, and these numbers are shown in Table 7.6. Comparing the average intensities of the single and dual centrifugation samples shows a decrease in the phosphorus signal of roughly half after passing through the 8 μm filter. This is corroborated by the percentages in Table 7.6. The same sources of error are present in this experiment as the previous one, however one more thing to note is that the error of smaller lines is much higher than larger lines. The phosphorus lines are among the smallest of the important lines we observe in our spectra, and have intensities that are not highly reproducible. Poor reproducibility of a small line is a known complication in LIBS. Despite the larger error in these measurements, the phosphorus line was however consistently lower in dual centrifugation samples. Because of the poor reproducibility, we cannot reliably state how high the percentage of passage through the 8 μm filter is, but due to the presence of phosphorus we can conclude that cells are passing through the 8 μm filter.

Table 7.4: Phosphorus 213 nm average intensity after single stage centrifugation for 3 species of bacteria, standard deviation, and the standard deviation on the mean for each.

	Average Phosphorus 213 nm Intensity (arb u)	σ	σ_{Mean}
E. coli	306.13	165.91	8.74
E. cloacae	487.70	352.54	18.01
S. aureus	355.13	169.70	8.58

Table 7.5: Average phosphorus 213 nm line intensity after dual stage centrifugation for 3 species.

Average Phosphorus 213 nm Intensity (arb u)				
	Trial #1	Trial #2	Trial #3	Average
E. coli	156.08	168.23	214.58	179.63
E. cloacae	197.44	191.86	272.45	220.58
S. aureus	197.86	211.19	251.44	220.17

Table 7.6: Ratio of phosphorus 213 nm line intensity of dual stage centrifuged samples to single stage centrifuged samples for 3 species.

Percentage of Phosphorus Line Measured				
	Trial #1	Trial #2	Trial #3	Average
E. coli	50.98 %	54.95 %	70.09 %	68.13 %
E. cloacae	40.48 %	39.34 %	55.86 %	45.23 %
S. aureus	55.72 %	59.47 %	70.80 %	61.99 %

After establishing that bacteria were present on the 0.45 μm filter after passage through the 8 μm filter, the blockage of cells after passage through a 0.45 μm filter was investigated using the same procedure outlined in Figure 7.2. The average phosphorus intensities were computed and compared to the average intensity after single centrifugation, these results are shown in Table 7.7. The results of this experiment were surprising; bacteria were observed on the second 0.45 μm filter after filtration through another 0.45 μm filter. The intensities observed once again were consistently lower than the average phosphorus line deposited through single centrifugation, but the intensities compared closely to the intensities observed after filtration through 8 μm filter. This finding suggested that the 0.45 μm filter does not catch all cells as originally thought. The percentages again were calculated by finding the ratio of dual centrifugation intensity to single centrifugation

intensity and are shown in Table 7.8. These percentages are comparable to the percentages observed when filtering with 8 μm filter first.

Table 7.7: Average phosphorus 213 nm line intensity after dual stage centrifugation for 3 species through the 0.45 μm filter.

Average Phosphorus 213 nm Intensity (arb u)			
	Trial #1	Trial #2	Average
E. coli	226.16	184.93	205.55
E. cloacae	213.5	243.3	228.4
S. aureus	221.00	197.63	209.32

Table 7.8: Ratio of phosphorus 213 nm line intensity of dual stage centrifuged samples to single stage centrifuged samples for 3 species.

Percentage of Phosphorus Line			
	Trial #1	Trial #2	Average
E. coli	50.98 %	54.95 %	68.13 %
E. cloacae	40.48 %	39.34 %	45.23 %
S. aureus	55.72 %	59.47 %	61.99 %

Due to the observation that bacteria was somehow coming through the 0.45 μm filter, further experiments were performed to determine how reliable filters are for catching bacteria. Experiments to quantify how much bacteria are caught by the smaller pore size filters were performed.

7.2 Investigation of Filter Efficacy

Originally, it was assumed that filters with pore sizes smaller than bacteria caught all of the bacteria. However, the previous results disagree with this initial assumption. It was observed via LIBS measurements of the phosphorus line that bacteria do appear to come through the 0.45 μm filter. This prompted several studies to determine which filters the bacteria could get through and approximately what fraction of bacteria the filters were catching. A study of pelletization was conducted after passage through several filters of different pore size. The goals of this study were to determine if a pellet of bacteria could be observed after a suspension passed through a filter and to measure how big the pellets

were relative to a control suspension of bacteria that had not passed through a filter. Next, measurements of the absorbance of the fluid that had passed through the filter were taken to attempt to quantify a cell concentration. LIBS measurements of these samples were also taken to compare to the optical densitometer measurements. The impact of the cone on filter efficacy was tested as well. Determining the efficacy of our filters will provide a more accurate estimate of our current limit of detection and may provide insight on how to deposit greater numbers of cells.

7.2.1 Pelletization After Filtration

A simple centrifugation test was devised to determine if there were bacteria passing through 8, 0.45, and 0.22 μm filters. 0.5 mL of *E. coli* was pipetted into 1.5 mL centrifuge tubes, which were then centrifuged to create a pellet. A picture was taken of the pellets before they were filtered and is shown in Figure 7.5. Before filtration, each pellet was approximately the same size. 3 centrifuge inserts were prepared with 8, 0.45, and 0.22 μm filters. The 8, 0.45, and 0.22 μm filter samples were then vortexed and removed from these 1.5 mL centrifuge tubes and deposited into centrifuge inserts with no cone. These samples were then centrifuged through the filter. After centrifugation, the filtrate at the bottom of the centrifuge tube was collected and placed back into the 1.5 mL tubes, centrifuged, and resulting pellet sizes were compared. A positive and negative control were also included for comparison; the positive control is an unfiltered bacterial sample, the negative control is ultrapure water. This is shown in Figure 7.6.

It is clear when comparing the before and after condition that all filters allow some bacteria to pass through. The pellets for all filters are smaller after filtration, but show that a significant amount of bacteria is not being caught on the filter. These pellets are also comparable in size to the pellet filtered through the 8 μm filter, which is counterintuitive.

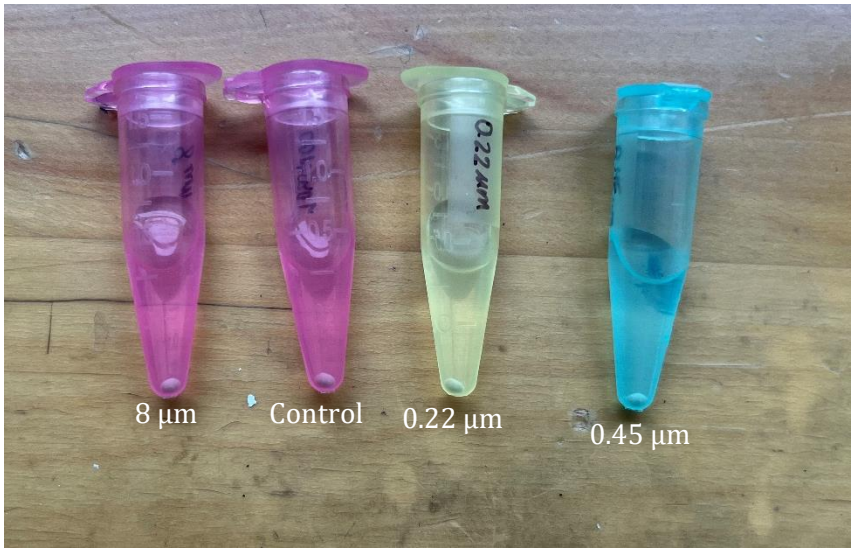

Figure 7.5: Bacteria pellets before centrifugation through a filter.

Figure 7.6: Comparison of pellet size after centrifugation through 8, 0.45, and 0.22 µm filters. A positive and negative control are also included; the negative control is purified water; the positive control is bacteria that have not been filtered.

The same experiment was performed using the cone, and the pellet size before and after were compared. This was to test if the cone had any positive or negative effect on the number of bacteria captured. In this case, there was variation in the initial amount of bacteria for each test tube, as shown in Figure 7.7a, so a comparison between the initial conditions could not be made. A comparison could be made between initial and final pellet for the same filter size after filtration, and it was found that bacteria could still pass through the filter. The initial condition is shown in Figure 7.7a, and the final condition is

shown in Figure 7.7b, c, and d. Therefore, the presence of the cone does not appear to mitigate the loss of bacteria through the filter, as had been hoped.

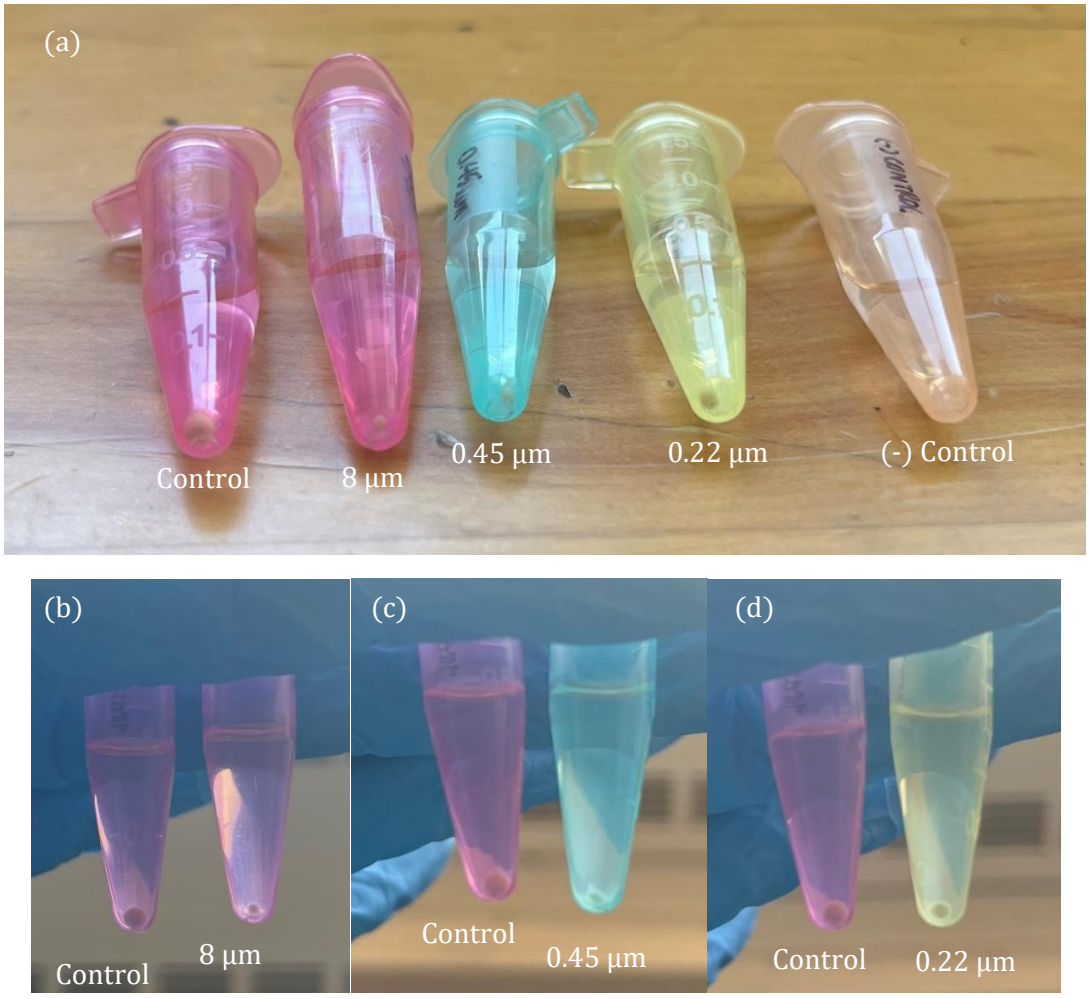

Figure 7.7: Comparison between bacteria pellet size (a) before filtration and (b),(c),(d) post filtration. Filtration was through the cone.

A more thorough comparison of the amount of bacteria passing through the filter was made by measuring the absorbance of the filtrate. 5 tubes containing 0.8 mL of *E. coli* 1/50 suspension were centrifuged in 1.5 mL tubes to ensure pellet sizes were comparable before testing. A lower concentration of bacteria was used to ensure that optical densitometer measurements were in the linear regime. Bacterial suspensions were deposited into centrifuge inserts containing either 8 or 0.45 µm filters. A 0.22 µm filter was not investigated here because it is not used in our sample preparation methods, and therefore is not entirely relevant. As well, since it also allows bacteria through, it does not provide a viable solution to the currently used 0.45 µm filter. Samples were made with and without

the cone. Samples were then centrifuged through the filter and filtrate was collected after. Filtrate was pipetted into 1 mL cuvettes for optical densitometer measurement. Absorbance measurements were also taken of ultrapure water and of an unfiltered sample. These absorbance values are compared in Table 7.9. Non-zero absorbance measurements indicate there are cells or cell fragments present in the water. As well, the 0.45 μm filter has lower absorbance values in both "with" and "without cone" cases, indicating that it is catching more bacteria.

Table 7.9: Absorbance measurements of water (negative control), unfiltered sample (positive control), 8 μm filter, and 0.45 μm filter. Absorbance are compared between filter size and between use of cone.

Sample	Absorbance		
(-) Control (Ultrapure H_20)	0.000		
(+) Control (Unfiltered bacteria)	0.086		
Without Cone	Abs.	With Cone	Abs.
8 μm	0.014	8 μm	0.041
0.45 μm	0.011	0.45 μm	0.026

LIBS measurements were also performed on the filters made with the cone to verify that cells were present on the filter and determine if the spectra collected from the 0.45 μm filter had a higher intensity than those from the 8 μm filter, which they should if the 0.45 μm filter was catching more cells. Spectra were not taken of the filters with no cone because the bacteria are too dilute to be measureable. A comparison of the spectra is shown in Figure 7.8. The comparison of spectra shows a higher intensity from the 0.45 μm filter, which corroborates the absorbance measurements. As well, the intensity of the phosphorus line was compared. This is shown in Table 7.10. The phosphorus 213 line is reliably higher for the 0.45 μm filter. The 214 nm and the 253 nm line are within error.

Table 7.10: Comparison of the phosphorus line intensity between filters.

Phosphorus Line	Intensity on 0.45 μm Filter (arb u)	Intensity on 8 μm Filter (arb u)
P213	284 ± 102	178 ± 60
P214	139 ± 73	80 ± 30
P253	59 ± 27	34 ± 14

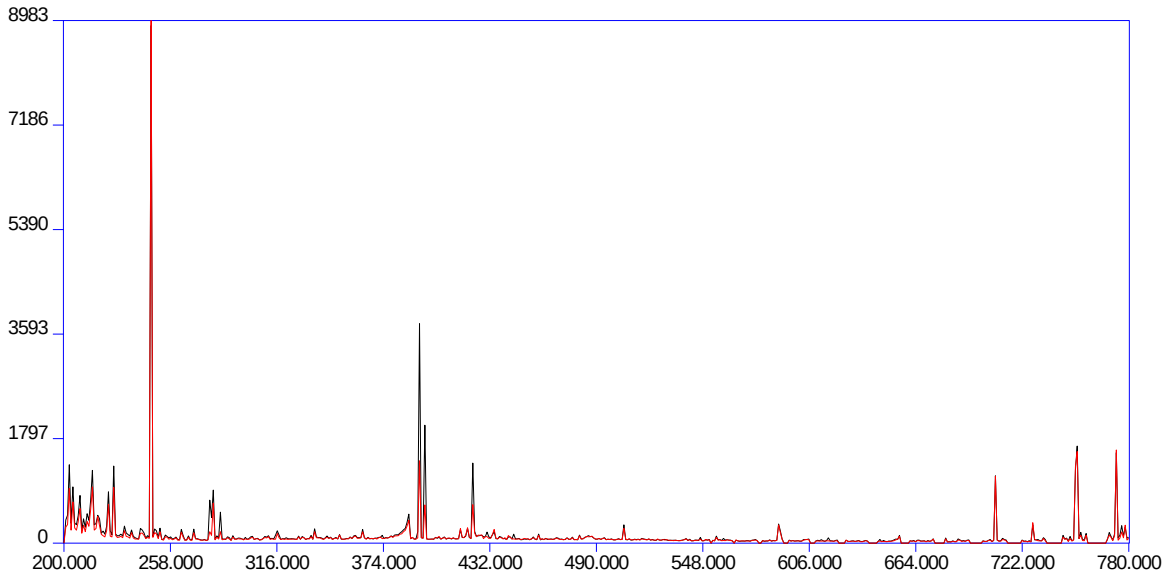

Figure 7.8: Comparison of LIBS intensity between 8 (red) and 0.45 (black) µm filter.

The reasons for bacteria getting through a filter that has a smaller pore size than the cell size is unclear. It was hypothesized that the force of the centrifugation caused bacteria to be forced through the filter, either by lysing the cells allowing them to fall through or forcing them to take a different path. To test this, different RPM values were investigated to determine if there was a value for optimized deposition. Higher amounts of cell deposition should result in a lower absorbance and higher LIBS intensity. A suspension of 1/20 *E. coli* was prepared and 0.8 mL were pipetted into centrifuge inserts containing 8 µm and 0.45 µm filters. For each filter, there were 4 samples made corresponding to 4 RPM values: 2000, 3000, 4000, 5000. Lower values were impractical to use because it took approximately 30 minutes or longer for complete deposition whereas a typical deposition can be completed in 5 minutes. Each sample was centrifuged through a filter. The filtrate at the bottom of each sample was collected after vortexing and placed in a cuvette for measurement. The experiment was then repeated to determine if the relationship between absorbance and LIBS was repeatable. The results are summarized in Figure 7.9 below.

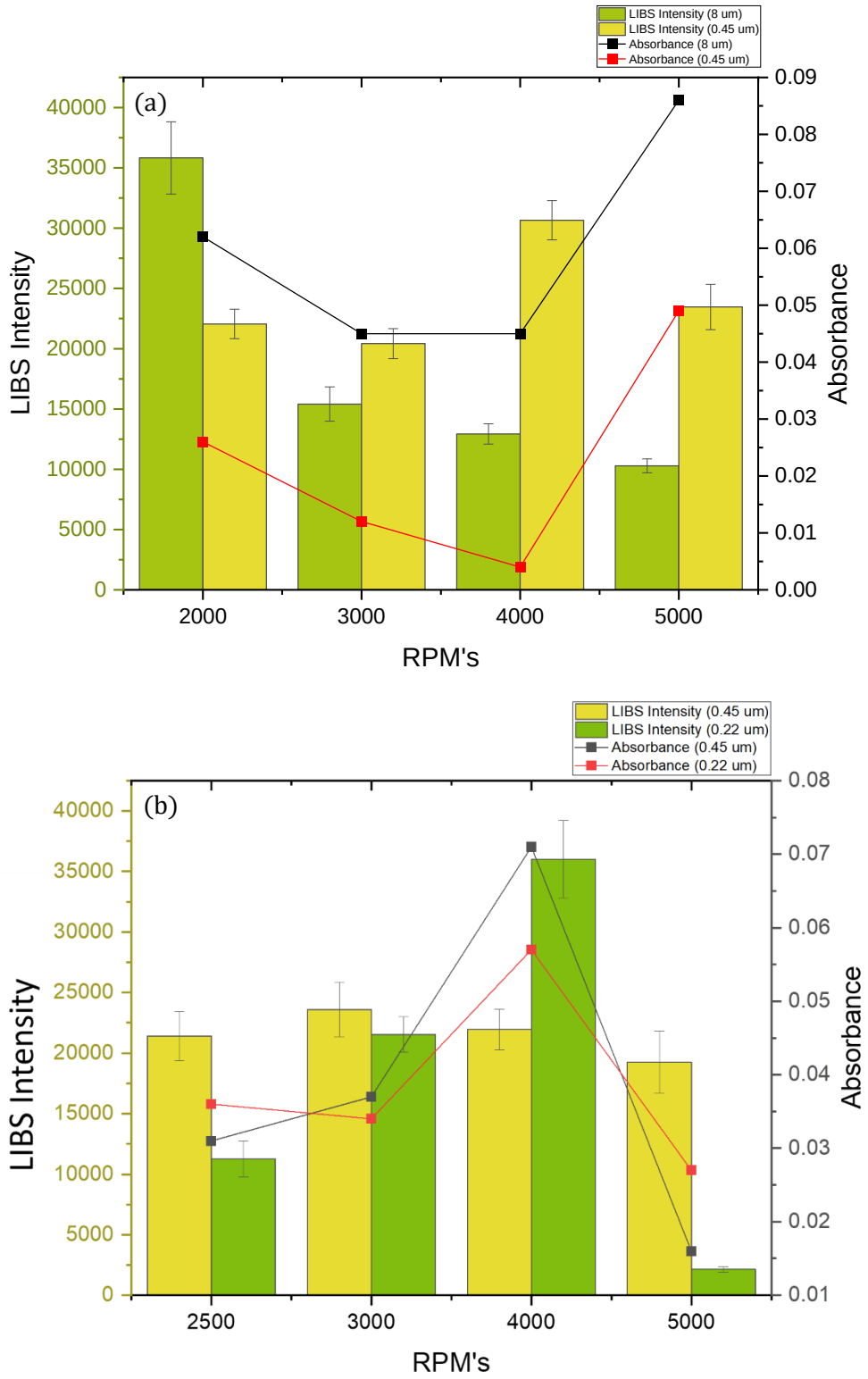

Figure 7.9: (a) Absorbance measurements for bacterial suspensions passing through filter (line graph) overlaid on LIBS intensity measurements. The absorbance values do not follow a clear or consistent trend with increasing RPM values. They are also uncorrelated with the LIBS measurements. (b) Repeat of absorbance experiments using different filter sizes; no pattern is clear between RPM's and LIBS intensities.

The results of this study did not confirm this hypothesis. Absorbance measurements were not inversely correlated with LIBS measurements as hypothesized, and followed no discernable or repeatable pattern with respect to RPM. It was expected that lower RPMs should have lower absorbance. This was not observed, except for the case of 5000 RPMs, therefore the centrifuge is most likely not lysing and pushing cells through the filter at higher forces. As well, for the 0.45 μm filters, there is no value of RPM that reliably improves LIBS intensity. Bacteria still pass through all filters including those with smaller pores and changing to a smaller RPM value will not catch them all. However, this study confirmed that at higher RPM's an 8 μm filter allows more bacteria through the filter due to its larger pore size, which confirms that using a filter with a larger pore size as the first step in dual centrifugation will let some bacterial cells pass.

To further test the permeability of our filters, 0.05 μm filters were purchased and tested using LIBS and the current deposition method. 0.8 mL of *E. coli* was deposited onto 0.05 μm filters using the cone and centrifuge insert. The samples were centrifuged, and the filters were removed and sampled with LIBS. The filtrate was centrifuged to observe if a pellet was present after filtration. After centrifugation a pellet was observed at the bottom of the centrifuge tube, indicating that bacteria did come through the filter, shown in Figure 7.10. It was also observed after centrifugation that the bottom of the cone does not create a good seal with the filter. The filters had a spot of discolouration due to water leakage from the cone, shown in Figure 7.11.

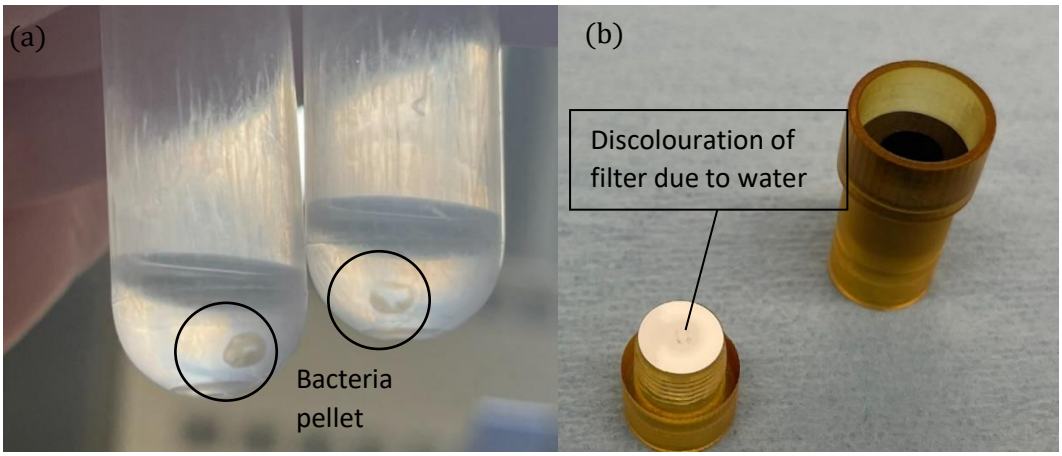

Figure 7.10: (a) Pellet of bacteria after filtration through a 0.05 μm filter. (b) Discolouration of filter due to leakage from the bottom of the cone.

Several 0.05 µm filters was sampled using LIBS. The intensities of the spectra were recorded and compared to intensities typically obtained on 0.45 µm filters. A comparison of the intensities are shown in Figure 7.11. It is clear from this figure that 0.05 µm filters produce intensities that are reliably smaller than 0.45 µm filters. Many of the 0.05 µm filters resulted in low intensity spectra and were below the average intensity obtained from bacteria deposited on 0.45 µm filters. This is likely due to the phenomenon observed in Figure 7.10 which showed an imperfect seal and leakage of water out from under the cone and around the filter. It was concluded that the 0.05 µm filters do not catch all bacteria and do not perform as well as 0.45 µm. Therefore, it is recommended that 0.05 µm filters are not used in sample preparation. As well, the problem does not appear to be pore size since bacteria that are larger than the pore size are going through.

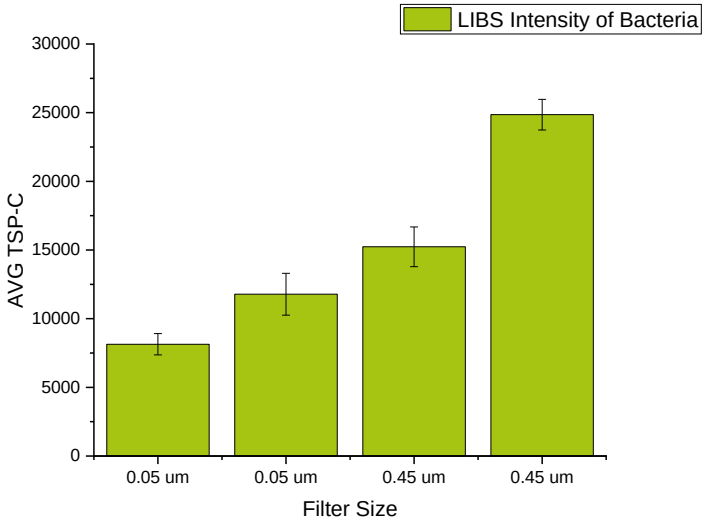

Figure 7.11: Comparison of LIBS total spectral intensities between 0.05 µm filters to 0.45 µm. It is clear that 0.05 µm filters do not produce a reliably high intensity.

Another potential explanation for the cells being able to circumvent filter deposition and form a pellet at the bottom is cell lysis. Typically, bacterial cells are stored in a phosphate buffer to relieve any osmotic pressure. However, we store our bacterial cells in ultrapure water to reduce the probability of any potential interferents in the spectra. Since the cells are not stored in a buffer solution, some lysis may occur prior to centrifugation, resulting in cell fragments reaching the bottom of the centrifuge tube instead of whole cells. However, we know that not all cells lyse when stored in ultrapure water because we

culture new plates of cells from old stocks, implying that not all cells are lysed and therefore non-viable. To test lysis of cells due to either osmotic pressure or the aforementioned force caused by centrifugation, 0.3 mL of 1/5 *E. coli* was centrifuged through 8 μm, 0.45 μm, 0.22 μm, and 0.05 μm filters. 0.3 mL of volume was used to ensure enough bacteria could be collected to form a pellet at the bottom of the centrifuge tube. After pelletization of the bacteria through centrifugation, the solution was vortexed, pipetted onto an agar plate, and then grown in the incubator for 48 hours to determine if the material passing through the filter consists of whole viable cells or just fragments of cells.

Pictures were taken at 24 hours and 48 hours after plating to compare the amount of colonies that grew. The photos taken after 24 hours of incubation are shown in Figure 7.12. From this figure it is clear that not as many bacteria are present after filtration as the control plate shown in 7.12a, but the smaller pore sizes do not filter out significantly more than the 8 μm pore size filter. Growth appears consistent and comparable across all plates regardless of pore size. The photos taken after 48 hours of incubation are shown in Figure 7.13. Consistent and comparable growth is seen here again across all plates that have been filtered prior to deposition for growth on agar. The presence of growth on all plates indicates that some cells are viable and are not being lysed due to centrifugation. Because some cells are not being lysed during centrifugation, which would allow them to pass through the smaller pore size filter, this study provides evidence towards the theory that some cells are going around the filter as opposed to through it. A potential solution for this would be to redesign the centrifuge piece used to concentrate the cells. More work needs to be done however to determine the dynamics of the fluid inside the centrifuge insert during the centrifugation process to further optimize cell deposition.

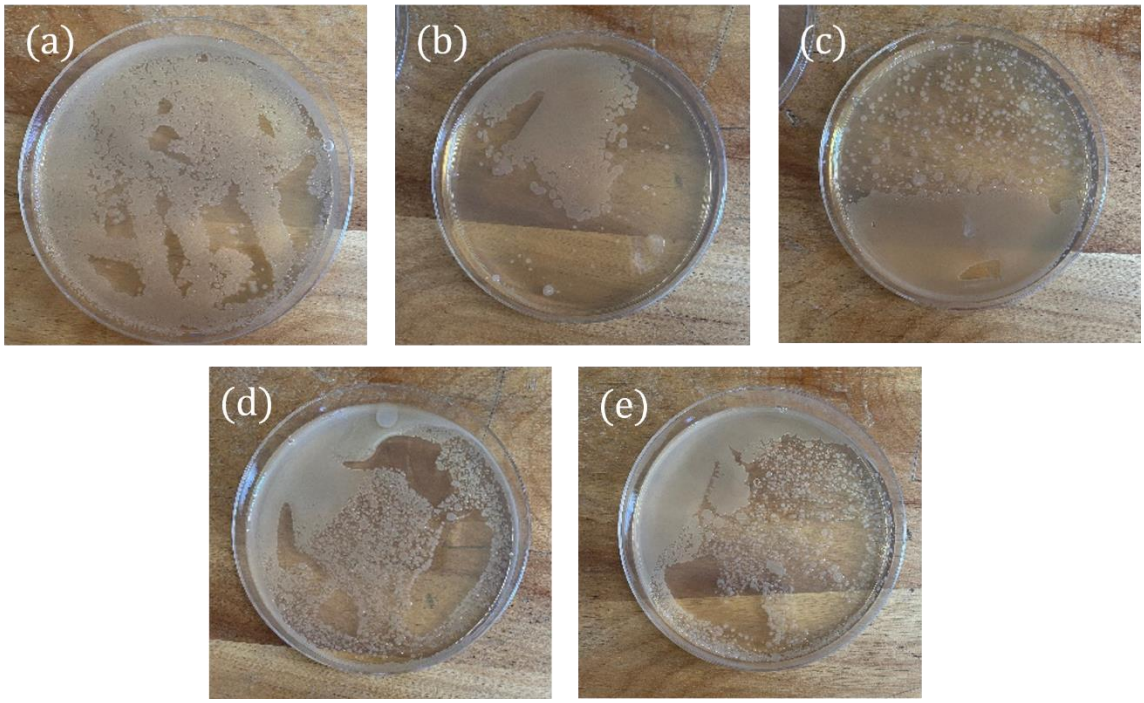

Figure 7.12: Pictures of growth on agar plates after 24 hours of incubation. (a) Bacteria that was not passed through a filter before plating. Bacterial suspension was plated following filtration with a (b) 8 µm pore size filter, (c) 0.45 µm pore size filter, (d) 0.22 µm pore size filter, and (e) 0.05 µm pore size filter.

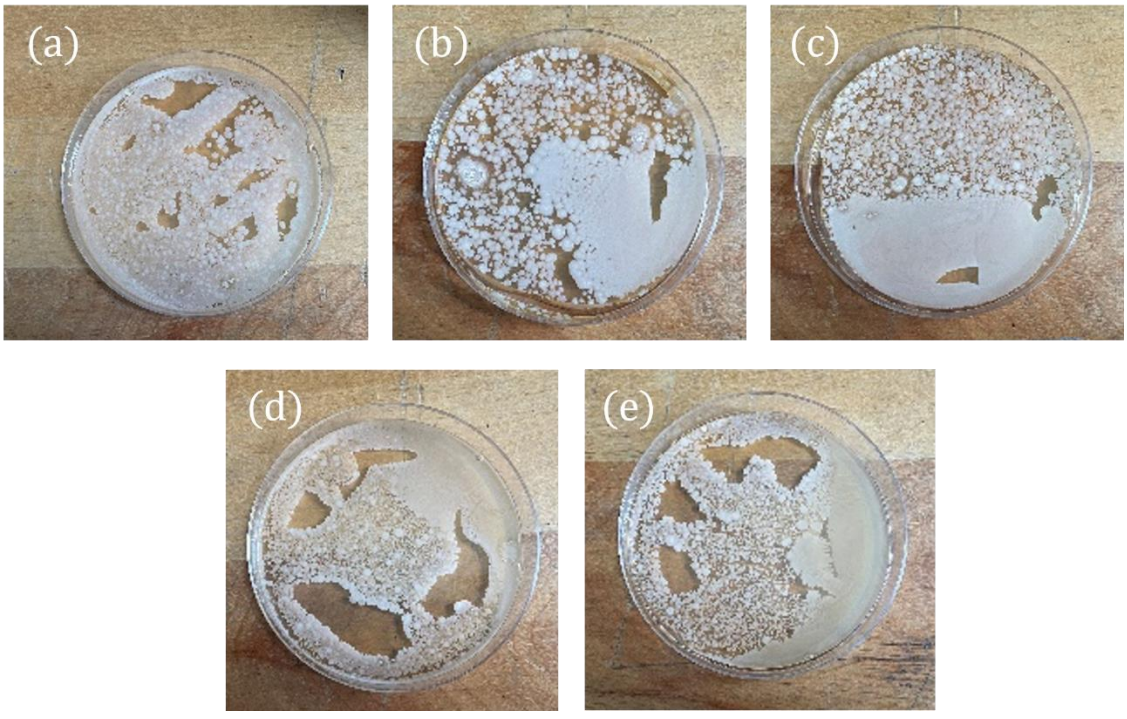

Figure 7.13: Pictures of growth on agar plates after 48 hours of incubation. (a) Bacteria that was not passed through a filter before plating. Bacterial suspension was plated following filtration with a (b) 8 µm pore size filter, (c) 0.45 µm pore size filter, (d) 0.22 µm pore size filter, and (e) 0.05 µm pore size filter.

It is also clear that the samples made with the cone have a higher absorbance than those made without the cone. A flaw in this experiment that may have caused this to happen is using a 0.8 mL volume of fluid. The cone and centrifuge tube can only hold 1 mL of fluid, and it has been observed in the past that higher volumes of fluid will leak around the side of the cone. Volumes between 0.75 mL and 1 mL are necessary for an accurate absorbance measurement. The difference between the values obtained with and without cone was attributed to this flaw in the centrifuge piece and cone. To resolve this, another trial of this experiment was performed. 100 µL of *E. coli* 1/5 suspension was pipetted into the centrifuge insert and cone, and centrifuged. Once centrifugation was complete, 0.8 mL of ultrapure water was added to the filtrate at the bottom in order to achieve enough volume to be measured by the optical densitometer. This new suspension was then vortexed and placed in cuvettes to be measured. Absorbance measurements were also taken for a negative control, which was ultrapure water, and a positive control, which was unfiltered 100 µL of bacteria. The results show that bacteria still come through the filter despite the presence or absence of the cone, but the absorbance measured is approximately the same between cone and no cone, shown in Table 7.11.

Table 7.11: Comparison of absorbance values after filtration through 0.45 µm and 8 µm filters with and without cone. Absorbance value were also compared between filters.

Sample	Absorbance		
(-) Control (H_2O)	0.000		
(+) Control	0.523		
With Cone	Abs.	Without Cone	Abs.
8 µm	0.011	8 µm	0.035
0.45 µm	0.022	0.45 µm	0.007

7.3 Conclusions and Future Work

Through analysis of the absorbance of several samples, it is clear that the filters used do not catch all of the bacteria, regardless of pore size or RPM speed. Evidence presented in this chapter suggests that bacteria goes around the filter instead of being deposited on it, most likely due to a poor seal between the filter and the centrifuge insert. This is especially evident in the deposition experiments done with no cone; however, it is unknown why this occurs. Attempting to use the cone to keep the bacteria concentrated also does not prevent bacteria from going around the filter, likely because the seal between the cone and filter is improper. The improper seal is shown schematically in Figure 7.14, which shows where the seal is supposed to be between the cone and filter. The red arrows in the figure represent the path that bacteria take around the edges of the filter. The leakage from the cone was also demonstrated in Figure 7.10b. The experimental inconsistency of how many cells are deposited is likely responsible for the high amount of scatter presented in Figure 5.4.

It is clear then that our deposition method is not as effective as once thought at reliably depositing cells on a filter. One solution for this is to redesign the centrifuge insert or the cone to create a better seal. Future work will focus on characterizing a flat metal disk, as opposed to the cone, to be used for concentration. The disk will be pressed between the 2 centrifuge insert pieces, hopefully reducing the amount of leakage and increasing the integrity of the seal, thereby increasing the number of cells deposited on the filter. A schematic of what this setup may look like using a metal disk is shown in Figure 7.15. With the disk used for concentration instead of the cone, the centrifuge inserts will likely not have to be redesigned for dual centrifugation. The current design if used in conjunction with the cone will only allow for concentration on the first filter, since the second filter would be on the second bottom piece screwed into the first bottom piece. There is no room for a cone to be placed on top of the second filter. If the disk concentrates as well as or better than the cone, the disk would eliminate the problem of redesign. It should also be noted that because less cells are being deposited on the filter than originally thought, our limit of detection is likely lower than originally anticipated.

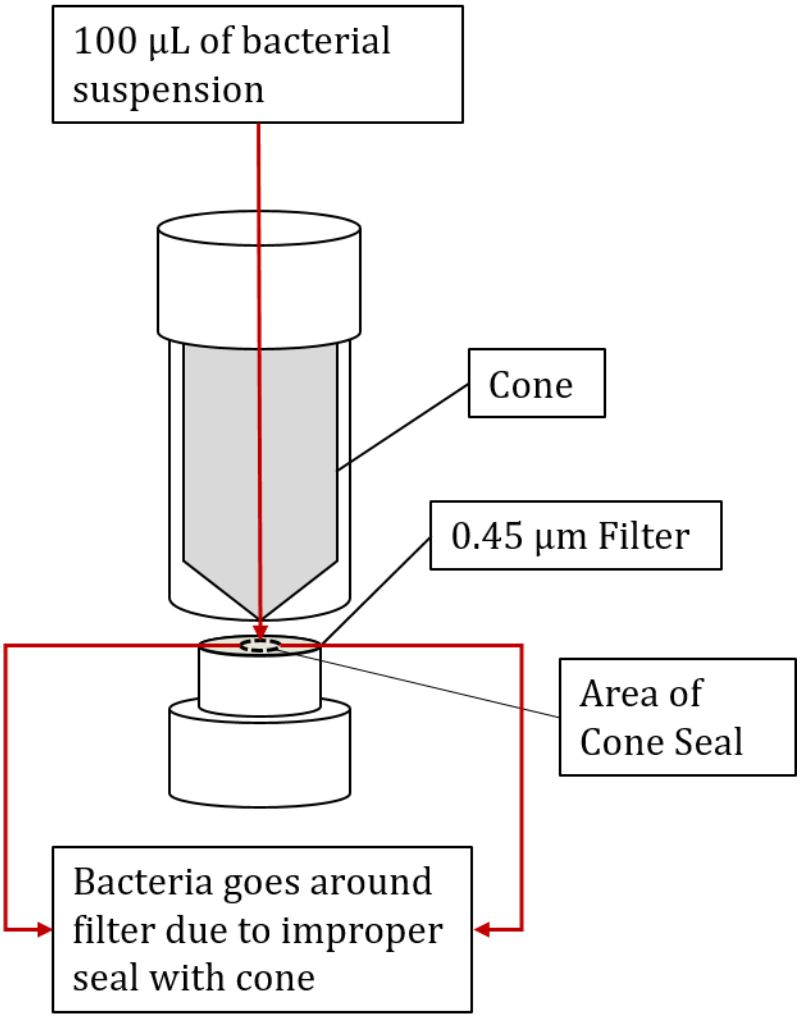

Figure 7.14: Schematic of the path of bacteria through the centrifuge insert and cone during the centrifugation process. The dashed line on the filter shows where the seal between the cone and filter should occur. Due to the improper seal between the 2 pieces, bacteria goes around the filter, due to a second improper seal between the filter and the centrifuge insert.

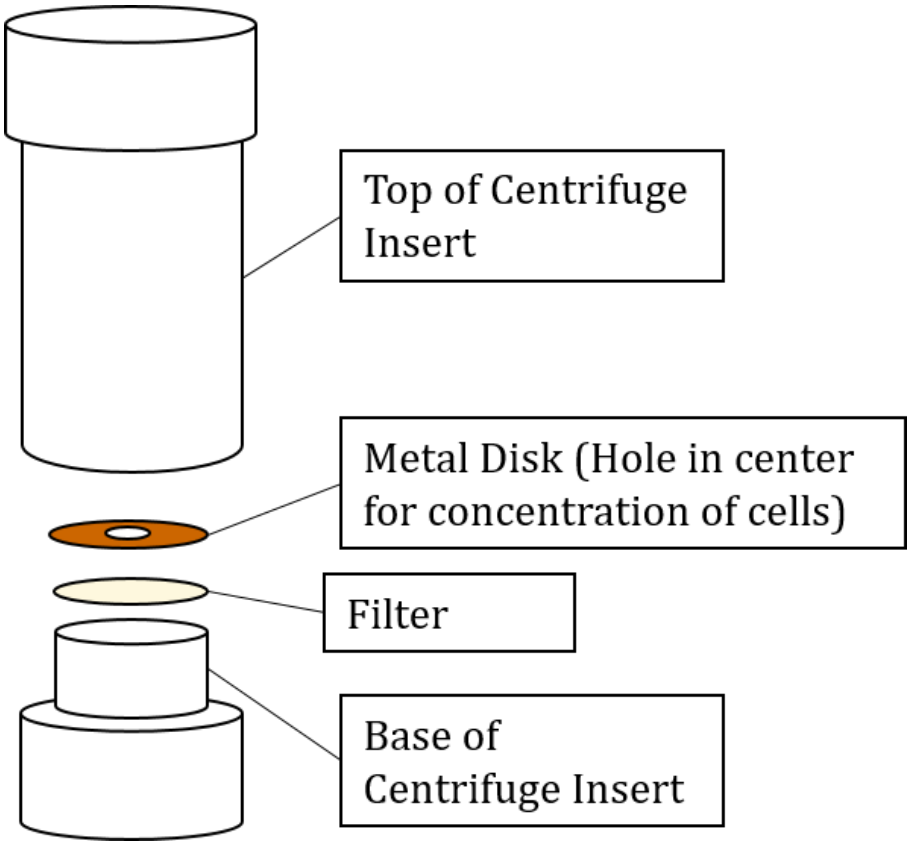

Figure 7.15: New design for concentration component of centrifuge insert. the disk will be between the top and bottom piece, potentially reducing the amount of leakage from the bottom and from around the sides.

The effect of phosphate buffer on our spectra should also be investigated. Cells are typically stored in a medium that reduces the osmotic pressure so no lysing occurs. While we know our cells are still viable since we reculture from stock solutions, it is unclear if any lysis occurs in the ultrapure water due to the higher osmotic pressure. Further investigation needs to be done on determining how the storage of cells in the appropriate buffer effects our spectra and ability to discriminate. Phosphate buffer solution should be characterized without the presence of cells and with cells to elucidate this. As well, investigating how the 'freshness' of our bacterial suspensions effects spectra and discrimination should also be investigated.

Finally, future work will also focus on optimizing dual stage centrifugation for the separation of blood and bacteria. Dual stage centrifugation through an 8 μm filter followed by a 0.45 μm filter was achieved, with the average total spectral intensity of the dual stage centrifugation samples being comparable to single stage centrifugation samples, and

phosphorus lines being measured at between 45%-68% for the 3 species tested. Dual stage centrifugation of *S. aureus* appears to have achieved the highest amount of bacterial deposition, but the reason for this is not clear at this time. Proof-of-concept has been achieved for dual stage centrifugation, and optimizing the amount of bacteria through the 8 μm filter while optimizing the amount deposited on the 0.45 μm filter will be the focus of future work. As well, a duplication of these experiments using these techniques on fresher clinical samples to ensure that we are modelling bacteria in blood correctly should be performed.

References

[135] *Complete Blood Count - Health Encyclopedia - University of Rochester Medical Center*. (n.d.). Retrieved September 19, 2022, from https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=167&contentid=complete_blood_count.

[136] Kinnunen, M., Kauppila, A., Karmenyan, A., & Myllylä, R. (2011). Effect of the size and shape of a red blood cell on elastic light scattering properties at the single-cell level. *Biomedical optics express*, *2*(7), 1803–1814. https://doi.org/10.1364/BOE.2.001803.

[137] Paulick, A. (2018). *Development of Laser-Induced Breakdown Spectroscopy as a Rapid Diagnostic Tool for Bacterial Infection.* [Master's thesis, University of Windsor].

Chapter 8: Conclusions and Future Work

8.1 Conclusions

The overarching goal of our research for the past several years has been to create a point-of-care diagnostic tool for rapid identification of bacterial infections. A rapid diagnostic tool will address the current antibiotic crisis occurring worldwide and accelerate the treatment process. Though we have made significant progress, several complications still needed to be addressed. When working with lower concentrations of cells, classification accuracy is poor between bacterial species. There is also high variability between filters which is causing poor classification. As well, no work had been done on testing bacteria in blood and urine. In working towards this goal, my thesis focuses on improvement of classification of species through outlier rejection, implementation of new algorithms to improve classification, as well as the development of testing procedures for bacteria in blood and in urine.

The problem of poor classification in chemometric algorithms was attributed to high scatter in the data as well as contamination from deposition pieces. Solutions to these problems were addressed in chapter 4. First, cleaning of the metal cone was investigated where it was found that ultrasonicating the cone in acetone and methanol reduced the background intensity of the spectrum. The effect of the water used in sample preparation was also studied, with the conclusion that ultrapure megohmic water has the lowest background intensity which is ideal for our experiment. Therefore, ultrapure megohmic water was used in bacterial suspensions and used in sample preparation. Other filters were also investigated to determine if any had lower carbon emission than the nitrocellulose filters. This study showed that nitrocellulose filters had the lowest emission of calcium, magnesium, and sodium lines and the highest emission of carbon. Since the sodium, magnesium, and calcium lines are all important to bacterial detection and classification, we have continued to use nitrocellulose filters. To address the scatter in the data, outlier rejection was investigated but did not improve classification results. A method that worked for improvement of classification was adding spectra together to reduce noise and remove the shot-to-shot variability. The average sensitivity and specificity of single-shot *E. coli*

classified against single-shot sterile water was 87 % and 72 %, respectively. After applying this method, the sensitivity and specificity improved to 100 % and 100 %, respectively.

For discrimination between species, application of DFA using 10-fold CV and adding spectra was unsuccessful for 5-class tests. Reducing the complexity of the model to a 3-class test and applying the same methods as a 5-class test did not improve the results. Due to the non-linearity of our data, we switched to developing and using an ANN algorithm. This only improved results marginally, which led us to add the preprocessing step of PCA before ANN to reduce the dimensionality of our data. As well, we chose to analyze the full spectrum as opposed to individual lines. This greatly improved our results for classification of species.

A new method of deposition was developed to simulate the testing of a blood or urine sample drawn from a patient. Blood and urine were characterized using the LIBS setup. Blood and urine were found to be fairly empty spectra, with the most prominent features being the carbon line from the filter and the sodium line. Infections were successfully simulated in blood and urine by 'spiking' each sample with our bacterial suspensions. The success of the simulated infection was made clear by the presence of bacteria lines on the fairly empty blood and urine spectra. A discrimination between sterile clinical fluids and simulated infections was performed with PLSDA, with identification of infected blood and urine occurring 100 % of the time. As well, classification of species present in blood and urine was carried out using DFA and PCA-ANN, with the best results being achieved with analysis of full spectrum data in PCA-ANN. This indicates that we can clearly detect and diagnose bacteria in a simulated sepsis or UTI infection.

Another method of deposition was investigated with the aim of separating out blood and bacterial cells. The studies were initially done to show proof-of-concept that the technique would work to separate out the two cell sizes, however it ended up showing that the deposition efficacy is not as high as originally anticipated. It was found that a large number of cells were not being deposited on the filter and instead ended up at the bottom of the centrifuge tube. Several studies were conducted varying the RPMs, filter size, and presence of cone to determine if the loss of cells could be reduced. It was found that none of

the above techniques allowed for capture of 100% of the cells, which led to the design of a new concentration piece that may be able to mitigate the number of cells going around the filter.

8.2 Future Work

The goal of our work is to develop an easy-to-use, low-cost instrument to be used in the clinic for rapid diagnosis. We have accomplished proof of concept by demonstrating accurate detection of bacteria in water and discrimination between species and strains. However, these results were accomplished with large amounts of cells, and most of the work over the past few years has been focused on retaining the same level of classification accuracy with fewer cells. This thesis and all future work aims to achieve the goal of high diagnostic accuracy with simple procedures, inexpensive equipment, and small numbers of cells.

Though classification of fewer cells was improved using PCA-ANN, work still needs to be done on classification of individual filters with this method. In a clinical setting, the diagnostic data will not be analyzed in an 80:20 split of training and testing. Rather, the testing data will be one sample or filter inputted by the physician against a library of training data. This poses a problem for our current procedures because external validation of individual filters was found to have a lower classification accuracy than an 80:20 external validation. To combat this, filter to filter variation needs to be reduced, which is being approached by designing a new concentration piece for use with our centrifuge insert. A schematic of what this might look like is shown in Figure 7.15 as a thin disk with a hole in the middle. The disk will span the entire width of the centrifuge insert and will sit between the bottom piece and top piece, potentially reducing or eliminating the leakage around the deposition area that was observed with the cone. Another approach to reducing the filter-to-filter variation is deposition of nanoparticles with the bacteria to enhance the spectra. Enhancement of spectra may eliminate the occurrence of low intensity filters. This is an ongoing area of research in our lab.

New algorithms may also be investigated if changing our deposition method to reduce filter-to-filter variability fails to improve classification with PCA-ANN. Work has been done

in the past on decision tree algorithms by Multari et al. as well as other authors. Decision trees make successive, 2-class decisions on samples entered into the model. For example, if this type of model was implemented in our work, the first test the algorithm would perform would be the presence or absence of bacteria. If bacteria were present, it would classify it against the species in the library that is easiest to identify, which would be *P. aeruginosa*. If the unknown species did not classify as *P. aeruginosa*, the decision tree would continue to perform 2-class discrimination tests until the unknown species found a match.

Filter-to-filter variation may also be affected by the coupling of the laser to the surface. We have observed numerous times that spectra coming from shots close together spatially on the filter are very different in intensity. Currently we use a 1064 nm laser for ablation, but it is known that lasers in the UV regime offer higher photon energies for bond breakage and ionization of the sample. UV lasers also have a lower penetration depth, particularly in water-containing targets, which allows for more energy per unit volume to be deposited in the sample, and therefore a greater ablation efficiency.[138,139] A 355 nm 10 ns Q-switched YAG laser is available in our laboratory for experiments to compare the ablation efficiency of the ultraviolet wavelength to the infrared wavelength with the bacteria targets.

Detection and diagnosis of bacterial pathogens could also benefit from the effects of dual-pulse LIBS. Dual-pulse LIBS increases the emissivity of the plasma, which increases emission intensities and therefore improves sensitivity of the technique. The most popular configuration studied that has shown the highest enhancement has been the orthogonal pre-ablation spark method.[140] This method however results in a wider or deeper crater, which may not be ideal for our testing substrate since it may result in more filter ablation instead of more bacterial ablation. Studies on the effect on the volumes of filter ablated using orthogonal pre-ablation LIBS are recommended. Another method that showed some enhancement was the collinear dual-pulse arrangement, where enhancement was mostly attributed to reheating of the plasma with the second laser pulse.[141] This method of dual-pulse may be more attractive than the previous as it could result in less filter ablation and enhancement of the bacteria lines.

To further understand the issues presented in Chapter 7 specifically and to understand deposition, concentration curves can be created using the optical densitometer for cell suspensions to understand what percentage of cells we are catching. A curve representing an unfiltered series of suspensions to serve as a control compared to a curve of filtered suspensions will allow us to calculate how many cells we are losing in deposition. As well, this curve will allow us to compare the efficacy of other deposition methods we hope to investigate using newly designed centrifuge insert pieces.

A more in-depth approach we could take could be to understand how the algorithms are classifying our data, and further how the filters are failing classification; this can be accomplished by understanding more in-depth how the algorithms work. Understanding of how the algorithms classify our data could be done using a technique called permutation feature importance. Permutation feature importance determines which features in a data set hold predictive power by shuffling the values for each feature. Breaking the relationship between the feature and the predicted result may result in a lower accuracy for the model, with larger decreases in performance meaning the feature holds more weight in the classification.[142] An in-depth analysis of each filter that misclassifies may give us an idea of how to improve our deposition and sampling.

In terms of clinical sample accuracy, future studies should focus on obtaining fresher samples of blood and urine, if possible, to ensure accurate representation of the data and performance in algorithms. As mentioned in Chapter 7, blood cells in samples can disappear over time, causing a misrepresentation of clinical conditions. A potential solution to this may be working closely with the staff in the hospital to test blood upon sample collection.

More studies need to be done to capture the diversity of bacterial concentrations that may exist within patients. Lower and higher concentrations of bacterial suspensions present in blood and urine need to be investigated to determine if they still classify correctly. LOD's also need to be found and quantified in blood and urine samples. Future studies should also focus on replicating the dispersion of the bacteria in blood and urine, since some species may not disperse evenly throughout the blood. For example, *S. aureus*

aggregate together surrounded by a fibre-based film, instead of evenly dispersing throughout the blood stream. However, behaviours and aggregation of cells in the blood stream is largely unexplored.[143,144]

A clinical fluid that could not be studied in this thesis due to infrequent testing and small lack of volume available in each test was cerebral spinal fluid. Future studies should focus on obtaining and testing this fluid to determine if diagnosis of bacteria in cerebral spinal fluid is feasible. Such experiments require the cooperation of our partners at the Windsor Regional Hospital and the availability of such specimens is out of our control.

[138] Russo, R.E., Mao, X.L., Borisov, O.V., Liu, H. (2000). Influence of wavelength on fractionation in laser ablation ICP-MS. *Journal of Analytical Atomic Spectrometry, 15*, 1115-1120. https://doi.org/10.1039/B004243I.

[139] Fornarini, L., Spizzichino, V., Colao, F., Fantoni, R., Lazic, V. (2006). Influence of laser wavelength on LIBS diagnostics applied to the analysis of ancient bronzes. *Analytical and bioanalytical chemistry, 385*(2), 272–280. https://doi.org/10.1007/s00216-006-0300-1.

[140] Pender, J., Pearman, B., Scaffidi, J., Goode, S. R., Angel, S. M. (2006). Laser-induced breakdown spectroscopy using sequential laser pulses. In Miziolek, A., Palleschi, V., Schechter, I. (Eds.) *Laser-induced breakdown spectroscopy: Fundamentals and Applications.* (pp. 516–538). Cambridge University Press.

[141] Cremers, D. A., Radziemski, L. J., Loree, T. R. (1984). Spectrochemical analysis of liquids using the laser spark. *Applied Spectroscopy*, *38*(5), 721–729. https://doi.org/10.1366/0003702844555034.

[142] Molnar, C. (2022, November 12). *8.5 Permutation Feature Importance | Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/feature-importance.html.

[143] Crosby, H. A., Kwiecinski, J., Horswill, A. R. (2016). Staphylococcus aureus aggregation and coagulation mechanisms, and their function in host-pathogen interactions. *Advances in applied microbiology*, *96*, 1–41. https://doi.org/10.1016/bs.aambs.2016.07.018.

[144] Pont, S., Fraikin, N., Caspar, Y., Van Melderen, L., Attrée, I., Cretin, F. (2020). Bacterial behavior in human blood reveals complement evaders with some persister-like features. *PLoS pathogens*, *16*(12), e1008893. https://doi.org/10.1371/journal.ppat.1008893.

Appendix A

Table A.1: Complete list of RM3 ratios used for discrimination.

Complete List of RM3 Ratios						
p1/c	p2/caii2	p4/mgii1	p6/mgii1	mgii2/caii2	mgi1/c	caii4/c
p1/mgii1	p2/caii3	p4/mgii2	p6/mgii2	mgii2/caii3	mgi1/caii2	caii4/na1
p1/mgii2	p2/caii4	p4/mgii3	p6/mgii3	mgii2/caii4	mgi1/caii3	caii4/na2
p1/mgii3	p2/cai1	p4/mgii4	p6/mgii4	mgii2/cai1	mgi1/caii4	cai1/na1
p1/mgii4	p2/na1	p4/mgi1	p6/mgi1	mgii2/na1	mgi1/cai1	cai1/na2
p1/mgi1	p2/na2	p4/mgi2	p6/mgi2	mgii2/na2	mgi1/na1	c/na1
p1/mgi2	p3/c	p5/c	p6/caii2	mgii3/c	mgi1/na2	c/na2
p1/caii2	p3/mgii1	p5/mgii1	p6/caii3	mgii3/caii2	mgi2/c	mgi1/mgii1
p1/caii3	p3/mgii2	p5/mgii2	p6/caii4	mgii3/caii3	mgi2/caii2	mgi1/mgii2
p1/caii4	p3/mgii3	p5/mgii3	p6/cai1	mgii3/caii4	mgi2/caii3	mgi1/mgii3
p1/cai1	p3/mgii4	p5/mgii4	p6/na1	mgii3/cai1	mgi2/caii4	mgi1/mgii4
p1/na1	p3/mgi1	p5/mgi1	p6/na2	mgii3/na1	mgi2/cai1	mgi2/mgii1
p1/na2	p3/mgi2	p5/mgi2	mgii1/c	mgii3/na2	mgi2/na1	mgi2/mgii2
p2/c	p3/caii2	p5/caii2	mgii1/caii2	mgii4/c	mgi2/na2	mgi2/mgii3
p2/mgii1	p3/caii3	p5/caii3	mgii1/caii3	mgii4/caii2	caii2/c	mgi2/mgii4
p2/mgii2	p3/caii4	p5/caii4	mgii1/caii4	mgii4/caii3	caii2/na1	cai1/caii2
p2/mgii3	p3/cai1	p5/cai1	mgii1/cai1	mgii4/caii4	caii2/na2	cai1/caii3
p2/mgii4	p3/na1	p5/na1	mgii1/na1	mgii4/cai1	caii3/c	cai1/caii4
p2/mgi1	p3/na2	p5/na2	mgii1/na2	mgii4/na1	caii3/na1	
p2/mgi2	p4/c	p6/c	mgii2/c	mgii4/na2	caii3/na2	

Table A.2: Complete list of ratios used in RM2.5 discrimination.

Complete List of RM2.5 Ratios				
p1/c	p2/caii3	mgii1/caii4	mgii4/caii3	caii4/na2
p1/mgii1	p2/caii4	mgii1/cai1	mgii4/caii4	cai1/c
p1/mgii2	p2/cai1	mgii1/na1	mgii4/cai1	cai1/na1
p1/mgii3	p2/na1	mgii1/na2	mgii4/na1	cai1/na2
p1/mgii4	p2/na2	mgii2/c	mgii4/na2	c/na1
p1/mgi1	p4/c	mgii2/caii2	mgi1/c	c/na2
p1/caii2	p4/mgii1	mgii2/caii3	mgi1/caii2	mgi1/mgii1
p1/caii3	p4/mgii2	mgii2/caii4	mgi1/caii3	mgi1/mgii2
p1/caii4	p4/mgii3	mgii2/cai1	mgi1/caii4	mgi1/mgii3
p1/cai1	p4/mgii4	mgii2/na1	mgi1/cai1	mgi1/mgii4
p1/na1	p4/mgi1	mgii2/na2	mgi1/na1	cai1/caii2
p1/na2	p4/caii2	mgii3/c	mgi1/na2	cai1/caii3
p2/c	p4/caii3	mgii3/caii2	caii2/c	cai1/caii4
p2/mgii1	p4/caii4	mgii3/caii3	caii2/na1	
p2/mgii2	p4/cai1	mgii3/caii4	caii2/na2	
p2/mgii3	p4/na1	mgii3/cai1	caii3/c	
p2/mgii4	p4/na2	mgii3/na1	caii3/na1	
p2/mgi1	mgii1/c	mgii3/na2	caii3/na2	
p2/caii2	mgii1/caii2	mgii4/c	caii4/c	
p2/caii3	mgii1/caii3	mgii4/caii2	caii4/na1	

Vita Auctoris

NAME:	Emma J. M. Blanchette
PLACE OF BIRTH:	Windsor, Ontario
YEAR OF BIRTH:	1999
EDUCATION:	B. Sc. [Honours] Physics, 2021 University of Windsor, Windsor, Ontario
	M. Sc. Physics, 2022 University of Windsor, Windsor, Ontario