A METABOLOMICS ANALYSIS AND THE DEVELOPMENT OF A FLUORESCENT ASSAY FOR KYNURENINE FOR THE DIAGNOSIS OF SEPSIS IN TRAUMA PATIENTS

by

KAMLESH PATEL

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Physical Sciences for Health Doctoral Training Centre
College of Engineering and Physical Sciences
The University of Birmingham
June 2020

Abstract

Traumatic injuries are a leading cause of death worldwide. Whilst the physical injuries can be treated patients are still at risk of developing an infection and becoming septic. It is hard to diagnose sepsis in patients of trauma as the inflammatory response to sepsis is masked by the inflammatory response to the traumatic injury. Current diagnostic techniques involve culturing blood samples or identifying a site of infection, neither of which are fast enough to allow for the best patient outcome. As such there is much need for a point of care diagnostic test. This work attempted to tackle this problem in two ways. First by analysing a metabolomics data set to identify biomarkers of sepsis in burns patients. Second, by developing a quantitative fluorescent assay for the detection of kynurenine, a biomarker of sepsis.

The data analysis was performed on a metabolomics data set acquired from clinical samples from burns patients. Three different classifiers were used to create models on the data set, k-nearest neighbours, naive Bayes and logistic regression. Classifier performance was above average on the data set, with the k-NN technique providing an AUC value of 0.82 and sensitivity and specificity of 85 % and 70 % respectively on the early-sample class-balanced subset of the data. The t-test and minimum redundancy maximum relevancy (MRMR) feature selection techniques were performed on the data set along with lasso with logistic regression. The t-test and logistic regression identified glucose and lactate as being the most important features whereas MRMR identified glucose and not lactate. These are both already well known biomarkers used in controlling the outcome of sepsis patients.

A literature search identified the metabolite kynurenine as being a positive biomarker for sepsis in patients of trauma. A fluorescent molecule was synthesised which upon binding to kynurenine causes a large bathochromic shift in the emission spectrum of the fluorescent sensor from 470 nm to 560 nm.

A fluorescent assay was created using the standard addition technique to quantify the amount of kynurenine in a urine sample. The standard addition technique was chosen as it removes matrix affects which would be present when using biological samples. Tests performed on pure samples of kynurenine, gave results within 5 % of the actual concentration. Synthetic urine was then used, giving results within 10 % of the actual concentration.

HPLC experiments were performed to quantify the amount of kynurenine in actual urine samples. The HPLC protocol was validated with kynurenine in solutions of water

and synthetic urine, however complications arose when using urine which meant the clinical samples could not be accurately quantified. As such, a qualitative experiment showed linear fluorescence plots when performing the standard addition technique with actual urine samples.

The results indicate the assay will work in urine. Being fluorescence based it has the potential to move into a point of care device. Future work would go towards quantifying the kynurenine concentration in clinical samples and then optimising the assay parameters once known concentrations can be determined.

ACKNOWLEDGEMENTS

I would first like to say a big thank you to my supervisors, Paula Mendes, Janet Lord and Peter Tino. You all have so much expertise in such a variety of fields, it's been humbling to engage in discussions with you all.

I would also like to say a huge thank you to the Mendes group for supporting me and being good friends. Special mentions to Eduardo Anaya-Plaza, Stefano Tommasone, Marcos Fernandez-Villamarin, Francia Allabush and Alice Di Pasquale for their extra help in the lab and to Joshua Norman for the insightful conversations. A thank you to Monika Köpf for being a special person who will go out of her way to help everyone. I also need to acknowledge the staff who support the research carried out, especially the analytical department of chemistry with Chi Tsang, Allen Bowden and Cecile Le Duff who personally helped out.

I would like to thank everyone from the PSIBS and Sci-Phy CDT cohorts which brought in many interesting characters. It was a pleasure meeting you all and sharing Chemistry West whilst we had it. A special mention to Muhammed Rassul for always having board games and sweets hidden somewhere on campus. Ann Smith deserves a huge amount of credit for looking after us for most of our time on the course.

Thanks to Chris Stepanek for introducing me to climbing and the many friends that brought with it, Hollie, Rebecca, Sam, George and everyone else who turned up.

A thank you to Daniel Lighter and his family who have treated me as their own family, and to Zoe Schofield for always encouraging me to never give up.

My housemates, James, Kalvin and Joe. You always made life interesting outside of work, from late night runs, card games, cartoons, silly challenges and epic meals.

I would like to thank my friends from undergraduate, Adam, Tola, Dale, Matt, David, Ian, Jawad, Rich, Charlie and all the rest of the group.

Finally a huge thank you to my family. This PhD has been a test of everything and you've stood by me through it all. This wouldn't have been possible without you.

This work was financially supported by a Engineering and Physical Sciences Research Council grant (EP/L016346/1) through a studentship from the Physical Sciences for Health Centre for Doctoral Training.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The work documented in this thesis was carried out with the aim of using techniques from the engineering and physical sciences to work towards diagnosing the condition of sepsis in patients of traumatic injury and in creating a diagnostic point-of-care test.

This chapter gives an introduction to sepsis and traumatic injury and the problems they introduce. Information about metabolomics and point-of-care testing with regards to the problems of sepsis and trauma are discussed.

The research aims of this thesis are then stated along with information on how they will be realised, followed by the thesis contributions.

The final section covers the outline of the rest of the thesis.

1.1 Background

1.1.1 Traumatic injury, burns and sepsis

Sepsis is defined as "life-threatening organ dysfunction caused by a dysregulated host response to infection" [1]. The infections are normally caused by bacteria, but can also be caused by fungi, viruses, and parasites [2, 3]. Whilst sepsis can occur in anyone, patients of traumatic injury, such as burns, are at a much greater risk of developing sepsis [4–6].

One of the reasons for this is because large/deep burns severely damage the skin. The skin is one of the bodies main defence systems against infection, and burns and other traumatic injuries brake this defensive layer, making the site of injury vulnerable to pathogens until the wound has healed [4, 5].

Sepsis is hard to detect in trauma patients as infection and trauma both cause an inflammatory response with the same common physical symptoms, redness, pain, heat and swelling [7].

Diagnosing and treating sepsis is a growing concern for modern medicine. A review looking at the global incidence in 2017 showed 48.9 million cases of sepsis were recorded with 11 million sepsis related deaths which equated to 19.7 % of the total deaths worldwide [8].

One way to protect against infection would be to prescribe antibiotics to all patients with traumatic injuries. This would not only be costly to do but could also have a negative impact on patient outcome. In our bodies there are bacteria which perform useful functions, such as aiding in digestion. Antibiotics can attack the useful bacteria in our bodies, making it harder for a patient to recover [9]. This is especially bad if the patient was not at risk of infection in the first place, as in this case the patient receives no positive effect from the antibiotics.

At this moment in time there are movements to try to control the distribution and administration of antibiotics more tightly [10]. The unnecessary or reckless use of antibiotics can lead to the evolution of antibiotic resistant bacteria which then puts the whole population at risk [11].

Due to these downsides in the blanket prescription of antibiotics there is a current clinical need to be able to differentiate between inflammation from infection and inflammation from a traumatic injury.

1.1.2 Diagnosis and treatment of sepsis

Symptoms of sepsis can be highly variable based on the patient age, infectious organism, comorbidities, and site of organ dysfunction [12]. Common symptoms can include fever, hypothermia, and hypotension, but even these common symptoms may be absent in elderly and neonatal patients [13–15].

A positive sepsis diagnosis is given when systemic inflammatory response syndrome (SIRS) and an infection are identified. There is a problem associated with this though, in that within the time taken for a site of infection to become identifiable, the number of pathogens circulating around the body will have ample time to multiply. It is estimated that for each hour the administration of antibiotics is delayed there is a 6 % increase in the risk of mortality from sepsis [16]. As such, the early diagnosis of sepsis is of utmost importance.

The Surviving Sepsis Campaign (SSC) is a global collaborative initiative with the purpose of improving treatment and reducing mortality from sepsis. Guidelines released by the SSC state that when sepsis is suspected, sets of blood cultures are to be taken to identify different types of bacteria [17]. The problem with this is that positive blood cultures can take on the order of hours to days to verify the presence of bacteria. Also, a

report looking at patient admissions over 10 years with over 6.8 million patients identified that 47.1 % of patients with severe sepsis were culture negative, that is the cultures performed did not positively identify bacteria [18]. As such, testing for bacteria is both slow and unreliable.

Within the guidelines of the SSC it is stated that no currently identified biomarkers should solely be used to direct medical intervention [17]. Given the magnitude of the problem that sepsis presents this highlights a huge problem and an area which needs to be investigated.

1.1.3 Metabolomics analysis for biomarker identification

The metabolome is the complete set of metabolites within a biological sample [19]. The metabolome is of importance as it comes after the genome, transcriptome and proteome, and so provides information on the functional status of an organism [20].

Metabolomics looks to identify and quantify metabolites, small sized molecules < 1500 Da, from patient samples e.g. serum, urine, spinal fluid, saliva [21, 22]. Using modern analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy and liquid chromatograph mass spectrometry (LC-MS) it is possible to identify hundreds and even thousands of metabolites, however one limitation of these techniques is difficulty in labelling the molecules found [22].

Metabolomics studies have been used to identify useful biomarkers for a variety of diseases/conditions such as Alzheimer's disease, hepatocarcinoma, chronic kidney diseases and forms of cancer [23–25].

There have been few metabolomics analyses of sepsis with trauma using limited data and sometimes even animal data [26, 27]. If more data were to become available from human subjects, metabolomics could identify novel biomarkers.

When it comes to sepsis with trauma, the physical signs such as changes in temperature, heart rate, and respiratory rate are not enough as they will be brought upon by the traumatic injury, as such utilising the metabolome which best characterises a patient's phenotype shows promise in uncovering biomarkers for sepsis with trauma.

1.1.4 Point-of-care (POC) testing

The identification of biomarkers is not enough there also need to be methods which allow for the rapid detection of such biomarkers.

POC devices typically produce results quicker than laboratory based techniques. There are multiple reasons for this such as they are designed to use patient samples directly without preprocessing and the locality of the device to the patient cuts down on time overheads of passing samples along and collecting results. A study comparing a lab based test with a POC test for procalcitonin showed that results were consistent between test types but the POC test was performed in 25 minutes vs 2 hours and 24 minutes for the lab based test [28]. An improvement of almost 2 hours would have a significant impact on the outcome of patients with sepsis.

Trauma patients have the worst outcomes when it comes to sepsis and so if biomarkers of sepsis with trauma can be identified there will be a need for a point-of-care test for the biomarkers.

1.2 Research Aims

Based on what has been described about sepsis and trauma this work had three aims.

Aim 1. To analyse a metabolomics data set from burns patients to identify novel

biomarkers of sepsis in trauma patients.

A novel study at the Queen Elizabeth hospital in Birmingham, United Kingdom, allows for this goal to be explored. The SIFTI (Scientific Investigation of biological pathways Following Thermal Injury) study which has already been undertaken has managed to collect samples from patients with burn injuries and as part of this there is labelled, quantitative metabolomics data available for analysis. Clinical studies such as this can take years to acquire funding, receive approval and recruit enough patients and therefore this work was very lucky to be able to use the data provided. This produces the first aim.

The data provided is completely relevant to addressing the research gap of there being no known biomarkers for the diagnosis of sepsis in burns/trauma patients. The analysis will look to determine whether the data collected provides good discrimination between septic and non-septic patients of traumatic injury and will further aim to determine which metabolites provide the best discriminatory power.

Aim 2. To design an analytical test to detect and quantify identified biomarkers of sepsis.

Taking biomarkers identified through aim 1 or from other research, this aim will work on the first stage in the production of a point-of-care test. Work will be conducted to synthesise selective recognition molecules for the biomarkers. For any diagnostic test, the analyte in question needs to be detectable. Fluorescent chemosensors can be synthesised which exhibit selective binding to analytes depending on structure and exposed functional groups and so a fluorescent based approach will allow for this aim to be met.

Aim 3. To develop and validate the analytical test to make progress towards making a point-of-care test.

Once a method of detecting and quantifying the selected biomarker/biomarkers has been established in the lab the next aim would be to have the platform work with patient

samples. POC tests are typically performed bedside and as such there can be no preprocessing of patient samples. The analytical test developed will need to be able to work with raw patient samples, or the test will need to be modified to incorporate methods of sample processing. It will also be of great importance to see if the test works with the complicated media of a biological sample.

Biological samples are available from the SIFTI study which can be used to validate any analytical test. In order to verify the validity of the test this work develops, a second quantitative technique will need to be used to confirm the values of analyte in biological samples.

1.3 Thesis Contributions

The contributions of this thesis are towards the development of a point-of-care test through the development of a fluorescence based assay.

This work synthesised and characterised a fluorophore which selectively binds to kynurenine. The linear range of the fluorophore was determined to be from 1 - 200 µM which includes the expected biological range of kynurenine. The optimum excitation and emission wavelengths were determined so as to produce the best signal and remove background noise. Analyte binding kinetics were uncovered with respect to temperature, and the stability of the chemosensor in neutral and acidic solutions was assessed to determine the chemosensor should be kept outside of solution until needed. This characterisation is of great importance for the future work in creating a POC test.

A standard addition based assay was developed which gave good accuracy of kynurenine concentrations in water (5 % error) and synthetic urine (10 % error). It was shown that when using a microplate reader there is no need to wait for the sensor:analyte complexation equilibrium to occur as long as the microplate is read in one pass, which saves

data acquisition time and will increase throughput.

A linear fluorescence response was shown in patient urine samples which shows great promise for the assay to be usable and to be developed into a point-of-care device.

1.4 Thesis Outline

This thesis consists of 9 chapters. This first chapter introduced the problem and identified the research gap. The rest of the thesis is structured as follows.

Chapter 2 explains in more detail the problem of sepsis and traumatic injury. Current research in diagnostic techniques and biomarker identification are reviewed. Machine learning techniques are discussed.

Chapter 3 explains the machine learning techniques and methods used for the analysis of the metabolomics data set.

Chapter 4 explains details of the data set. The results from the analysis are shown along with discussion of them.

Chapter 5 details the experiments performed in order to create the fluorescent assay. The choice in biomarker (kynurenine) and fluorophore are explained and the results from the characterisation studies are given and discussed.

Chapter 6 covers the experiments performed when using the fluorescent assay to quantify analyte in increasingly complex solutions, water, synthetic urine and patient urine. The results from attempts to quantify kynurenine using laboratory based high performance liquid chromatography in order to verify sample concentrations are given with discussion.

Chapter 7 covers the materials and methods used for the synthesis and characterisation

of the fluorophore and information on the validation experiments.

Chapter 8 provides discussion of this work with respect to the research aims, with limitations and future work also considered.

Chapter 9 gives the conclusions of this work and briefly touches upon what would be done should the project be taken forward.

CHAPTER 2

BACKGROUND

This chapter explains the themes of this project in more detail. First starting with traumatic injury, burns, sepsis and inflammation. In order to explain the problem fully, it is important to explain the different forms of inflammation, and also to talk about the real world impact this condition has. Data for this work has been collected from burns patients, and so some background information on burns is also provided. The biological processes are explained to provide context to the problem and to impart how substantial a problem sepsis is for patients of trauma. Current research with regards to the diagnosis of sepsis is then discussed, mentioning limitations which promote further research.

The next section covers systems biology and -omics research. The importance and usefulness of metabolomics is explained, with examples of how it has already identified biomarkers for other conditions.

One of the major components of this work was to analyse a data set collected from a clinical study. This introduction chapter will aim to give an overview of the types of machine learning techniques available to use and will also provide information about data itself. A discussion on algorithm selection is providing highlighting the pros and cons of each technique.

The discussion of data processing techniques is followed by information on diagnostic testing, where benefits and limitations of laboratory and point-of-care tests are discussed. Finally fluorescent chemosensors are discussed with respect to their ability to detect biomolecules and their importance for forming the basis of point-of-care tests is covered.

2.1 Traumatic Injury, burns and sepsis

Traumatic injuries are a major cause of disability and death in the world and they are predicted to increase in proportion as the cause of total deaths worldwide [29]. The Centre for Disease Control and Prevention lists types of traumatic injuries and disorders in its Occupational Injury and Illness Classification System and they are defined as "the result of a single incident, event, or exposure" [30]. The category therefore includes not just damage arising from physical impacts, but also the impact of things such as chemicals and burns. Illness caused by pathogens, neoplasms or conditions which arise from long term complications are not included [30]. This work will focus predominantly on burn injuries, however the body's response to other types of traumatic injury will be similar with regards to their inflammatory response [29].

Burns injuries range in severity and are classified by the number of skin layers affected. Therefore, first degree burns affect only the upper layer of skin (the epidermis), second degree burns affect the upper two layers of skin (the epidermis and the dermis,) and third degree burns (also called full thickness burns) affect all three layers of skin (the epidermis, the dermis and the subcutis) [31–33]. The layers can be seen in fig. 2.1. A burn can also be classified as being local or a severe burn injury (SBI) depending upon the total body surface area (TBSA) affected by the burn, with a local burn having a TBSA < 20 %, and a SBI having a TBSA > 20 % [32].

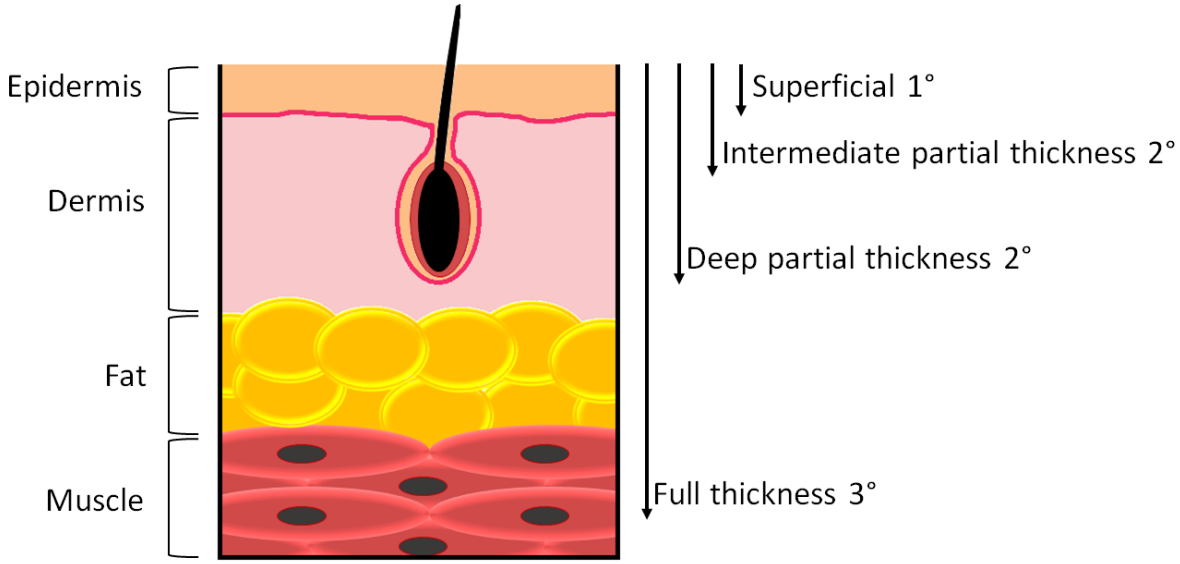


Figure 2.1: Types of burns and the layers of skin and tissue affected.

In the UK the NHS treated over 15,000 patients for burns and scalds in 2017, whilst in the US there were 486,000 burn injuries which received treatment in 2011 [34, 35]. The incidence of burns injury and the cost to treat them are very high; serious burns can cost the NHS £95,000 to treat [34]. High costs for treating severe burns are associated with the cost of staying in an intensive care unit, requiring multiple surgeries, constant monitoring and supplies such as blood and wound dressings [36].

2.2 Inflammation

Inflammation is a non-specific immune response which arises from infection or injury. Signs of inflammation include: redness, pain, heat and swelling [7]. These signs can be explained by responses such as increased blood flow, vasodilation, increased metabolism, leukocyte extravasation and release of soluble mediators, mainly pro-inflammatory cytokines, such as interleukin 1 beta [37].

Inflammation can be of two forms, acute or chronic. Acute inflammation is the first response to injury where the body aims to remove pathogens and control damage. Acute

inflammation should only last a few days. Conversely, chronic inflammation can persist for months to years. Chronic inflammation can be caused by autoimmune responses or pathogens which the body cannot remove. The work presented looks at data from patients immediately after injury. This will therefore be acute inflammation.

Inflammation can also be categorised as non-sterile or sterile depending upon the presence of pathogens. This is a more important distinction, and the focus of this work.

2.2.1 Non-Sterile Inflammation

Non-Sterile inflammation is caused by pathogens. Pathogens create an immune response through the activation of pattern recognition receptors present on immune cells [38]. Pathogens have molecules associated with them called pathogen-associated molecular patterns (PAMPs) which are detected as foreign to the host. Examples of PAMPs include peptidoglycan (a polymer which forms the cell wall of many bacteria), lipopolysaccharide and double-stranded RNA which forms the genetic material of some viruses [29].

2.2.2 Sterile Inflammation

Whilst non-sterile inflammation occurs due to the presence of pathogens, it is possible to trigger an inflammatory response without pathogens. Such a response is called sterile inflammation, which occurs due to the host's response to the death of its own cells as opposed to the invasion of pathogens.

When cells are damaged in the body there are two distinct mechanisms by with cell death occurs, under a controlled mechanism called apoptosis, or in an uncontrolled manner called necrosis.

In apoptosis the cell membrane maintains its integrity [38] and intracellular compo-

nents are broken down though enzyme activation via controlled signalling cascades. This results in very few molecules leaking out into the extra cellular matrix and thus an inflammatory response is not produced [38–40].

For traumatic injuries cell death occurs via necrosis which causes the release of cell contents called damage-associated molecular patterns (DAMPs) from the cells. DAMPs are molecules which can trigger an immune response. This immune response is similar to if the body were being attacked by pathogens, but as it is not, the inflammation caused is called sterile inflammation.

DAMPs are released from different locations in the cell including the nucleus, the cytoplasm and the mitochondria. DAMPs take the form of DNA/RNA and intracellular proteins [29,38,41]. Mitochondrial DNA acting as a DAMP can induce a significant sterile inflammation response. One theory for this is because mitochondria are believed to be endosymbionts and so were originally bacteria [42].

2.2.3 The acute-phase response

The acute-phase response is the name given to the initial set of processes that occur during inflammation. This response encompasses the actions of preventing further damage to tissues, the separation and destruction of pathogens and also the activation of repair processes which help return the body to normal function [43].

At the site of infection/injury two types of white blood cell (WBC), neutrophils and macrophages, will aggregate. The pair of WBCs along with endothelial cells and platelets will secrete the pro-inflammatory cytokines: Tumour Necrosis Factor-α (TNF-α), Interleukin-1β (IL-1β) and Interleukin-6 (IL-6) [43, 44]. Cytokines are small peptides used for cell signalling and the cytokines listed are chemotactic towards neutrophils and macrophages. It is this chemotactic gradient which leads to the aggregation of neutrophils

and macrophages at the site of infection/injury.

2.2.4 SIRS, Sepsis and Multiple Organ Failure

Inflammation is kept in balance through pro-inflammatory and anti-inflammatory mechanisms. When these inflammatory mechanisms are not regulated properly after trauma, systemic inflammatory response syndrome (SIRS) can develop. SIRS is an inflammatory response of the immune system which affects the whole body. SIRS is diagnosed when at least two of the following criteria are met [45]:

- Patient temperature above 38 °C or below 36 °C
- Heart Rate $>$ 90 beats per minute
- Respiratory Rate $>$ 20 breaths per minute, or, arterial carbon dioxide partial pressure ($PaCO_2$) $<$ 4.3 kPa
- White blood cell count of less than 4000 cells/mm^3, or, white blood cell count greater than 12,000 cells/mm^3, or, 10 % immature band forms (immature neutrophils)

The anti-inflammatory mechanism associated with SIRS is the compensatory anti-inflammatory response syndrome (CARS) and involves production of anti-inflammatory cytokines such as IL10. If both the pro and anti-inflammatory responses do not return to a homeostatic level the condition can progress to persistent inflammation, immunosuppression, and catabolism syndrome (PICS). As the name indicates, PICS consists of persistent signs of inflammation from a pro-inflammatory response, but also leaves patients in an immunosuppresed condition which makes them more susceptible to infection. SIRS and PICS increase non-specific immunity around the body as they trigger neutrophils, part of the innate immune system. This however leaves the body vulnerable to attack via pathogens which can lead to sepsis. Sepsis is diagnosed when SIRS can be diagnosed (at

least two of the above conditions are met) and there is also an infection. The increased neutrophil activity during SIRS (and PICS) causes them to move across damaged tissues and accumulate in the organs. When this happens the neutrophils will begin to destroy healthy tissue which leads to multiple organ failure (MOF) [29, 46, 47].

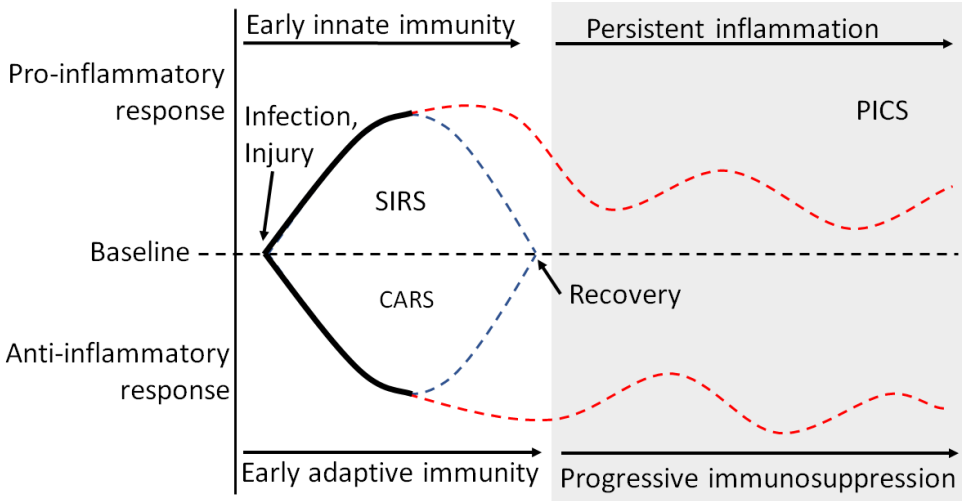


Figure 2.2: Diagram illustrating the relationship between pro-inflammatory and anti-inflammatory responses to injury/infection. At the point of injury/infection both pro and anti-inflammatory responses are triggered. The blue dotted line shows the intended behaviour where both responses return to the baseline and the patient recovers normally. The area in grey shows the onset of PICS, with the red dotted lines showing persistent inflammation along with a persistent anti-inflammatory response which suppresses the body's immune response.

A positive sepsis diagnosis is given when systemic inflammatory response syndrome (SIRS) and an infection are identified.

If sepsis is suspected the SSC recommend running at least two blood cultures (to test for aerobic and anaerobic pathogens) [17]. There are multiple problems with depending on blood cultures however. First, it has been reported to take a least 24 hours to get information of the potential organism [15]. Secondly it has been reported that 47.1 % of of patients with severe sepsis were culture negative and so it is reasonable that a patient could have sepsis whilst reporting a negative blood culture [18]. These two reasons combined with the fact that there is a 6 % increase in the risk of mortality from sepsis for each hour antibiotic therapy is delayed mean the use of blood cultures for diagnosis will not always

offer the best outcome for the patient [16].

Trauma patients, especially those of serious burns, are more likely to become subjected to sepsis, due to the fact that the skin, which normally provides to first barrier to pathogens, will be severely damaged [4–6]. As the sterile inflammation masks the non-sterile inflammation, better diagnostic technologies are needed.

Given these facts it can be seen that the early diagnosis of sepsis is essential, especially for trauma patients who already have serious morbidities, and so there needs to be new diagnostic tests made for this condition.

2.2.5 Current developments of diagnostic technologies for sepsis

Sepsis is a huge healthcare problem and so there is much research exploring different ways to diagnose sepsis. This section will look at diagnostic technologies under development.

Researchers from the University of Bath have developed a wound dressing which changes colour when a bacterial biofilm develops in the wound. The dressing works by encapsulating the dye molecule, carboxyfluorescein, inside vesicles. The vesicles are meant to mimic the membranes of host cells which the bacteria will release toxins to attack. The toxins cause the vesicles to burst, releasing the dye, causing a colour change [48, 49]. The colour change occurs because at high concentrations, carboxyfluorescein will self-quench, and it is only upon release from the vesicle and therefore upon dilution that the fluorescence can be detected [50]. A possible downside to this approach is that it requires the formation of bacterial biofilms to develop on wounds which is not a rapid process. In a paper published by the group, results were given for biofilms given 48 hours to develop on a tissue sample [51].

Researchers from Harvard University have been working on diagnosing sepsis by looking at neutrophil movement. Neutrophils are the most abundant type of white blood cell

in the human body and are a part of the innate immune system. As was discussed in the previous section, neutrophils play an important role in the progression of sepsis and can even be the cause of multiple organ failure. The research shows that neutrophils from patients with major burns have reduced migrational speeds across a microfluidic device they produced [52, 53]. The group later produced a microfluidic device which could process and capture neutrophils from a drop of blood [54]. The use of a microfluidic device relies upon the use of a microscope to view and record the neutrophil activity which was performed for 4 hours in the referenced study. As such, currently this kind of testing would be best performed in a hospital laboratory and would not be suitable for a bedside point of care test.

It is not just neutrophil mobility that is being studied. Neutrophil extracellular traps (NETs) are being studied too. NETs are an extracellular matrix of DNA, proteins and histones which are normally used as part of the innate immune system to kill bacteria and viruses. A study conducted in mice showed that by lowering NET activity, sepsis was attenuated [55]. Measurement of neutrophil NETs requires isolating neutrophils from blood, incubating with a stimulant such as lipopolysaccharide and then performing further tests to measure reactive oxygen species, nitric oxide and mRNA expression. These are separate tests and performing so many would make assessing neutrophil NET activity in a point of care device quite a challenge.

Procalcitonin (PCT) is a peptide which has been linked to sepsis and levels of it have been shown to increase in response to pro-inflammatory stimulus [56]. Its use as a biomarker in guiding clinical decisions with regards to sepsis has been reported to reduce antibiotic treatment duration [57, 58]. There has been much interest in procalcitonin over the years with many studies into its effectiveness with sensitivity values ranging from 76 % to 90 % and specificity values ranging from 70 % to 91 % [59–61]. Point of care devices have been made to detect procalcitonin, however the manufacturers of such devices state that "Increased PCT levels may not always be related to systemic bacterial infection", and

on the list of exceptions is included "The first days after a major trauma, major surgical intervention, severe burns" [62]. The SSC specifically mentions procalcitonin should not be used solely to make clinical decisions and it has separately been found to not be a good biomarker for critically ill patients [17, 63]. A review of sepsis biomarkers also concluded that PCT has limited ability to distinguish sepsis from other types of inflammation [64]. As such, procalcitonin is not a good biomarker for sepsis in patients of trauma, which as has been stated are one of the groups most likely to become afflicted with sepsis.

Lactate, a metabolite used in energy metabolism has also been identified as a biomarker for morbidity with respect to sepsis [65]. Lactate is measured as part of standard laboratory tests for patients in hospitals and it has been referenced in the Surviving Sepsis Campaign guidelines with respect to patient resuscitation as something to monitor, however this is not used for diagnosing sepsis. A systematic review of point of care lactate testing with respect to sepsis and infections concluded that there was not enough evidence to support the use of point of care lactate monitoring and the SSC highlights that the use of currently known biomarkers should only be used to support and supplement clinical assessment and that decisions with regards to antimicrobial therapy should never be made solely on the basis of biomarker data [17, 66].

The technologies and developments discussed here are where most of the research has been directed. Each technique currently has limitations as have been mentioned in their respective paragraphs. As such there is a still the need for the identification of biomarkers which can accurately diagnose sepsis. Due to the fact that early diagnosis of sepsis significantly improves patient outcome, a point of care device to measure any biomarkers would be extremely useful.

Systems biology could provide the key to identifying biomarkers for sepsis and this will be explained in the next section.

2.3 Systems Biology and Metabolomics

Systems biology is the combination of quantitative experimental techniques with computational and mathematical techniques for analysing and modelling biological systems [67]. A definition given by Kirschner for the field is as follows, "systems biology is the study of the behaviour of complex biological organization and processes in terms of the molecular constituents" [68]. What this means is that systems biology looks at how properties emerge by viewing systems holistically, as opposed to breaking down systems into individual constituents.

One type of data that is collected in systems biology research is -omics data. Omics can represent many fields within biology, common fields being genomics, proteomics and metabolomics which look at the genome, proteome and metabolome respectively (see fig. 2.3). This project involves the use of metabolomics data and so this will be explained in more detail.

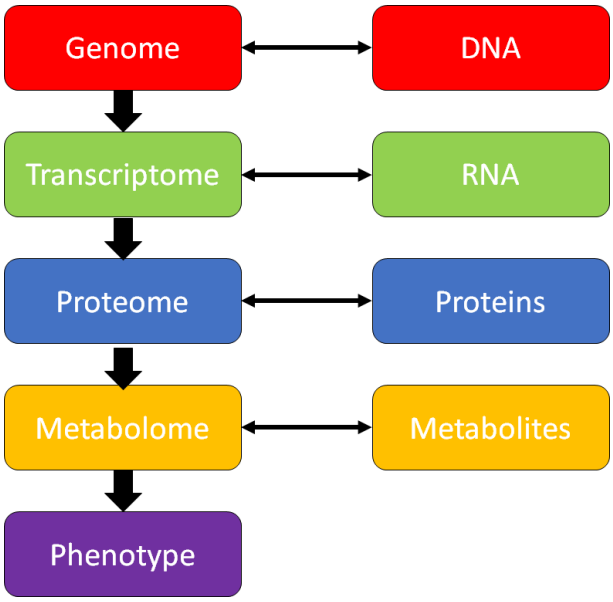


Figure 2.3: Relationship of genome, transcriptome, proteome and metabolome along with the constituents of each. The metabolome is downstream of all of the others and directly influences the phenotype.

2.3.1 Metabolomics

Metabolites are low molecular weight (typically < 1500 Da) organic and inorganic chemicals, which take part in enzyme-catalysed biochemical reactions, as reactants, products or intermediates [19, 69]. Metabolites can be endogenous (naturally produced in the organism) or exogenous (not naturally produced in the organism, e.g. drugs) [70]. Given the size restrictions, proteins are typically not metabolites, however small peptides are included, along with amino acids, carbohydrates, fatty acids and other small molecules.

The metabolome is the complete set of metabolites within a biological sample [19]. As it is downstream of the genome, transcriptome and proteome, the metabolome is of great importance is it characterises changes in the previous omics areas and provides information on the functional status of an organism (fig. 2.3) [20]. Metabolomics data is therefore a good candidate for research into the discovery of biomarkers for disease as changes in the metabolome reflect changes in DNA, RNA and proteins [67].

Metabolite information can be collected through analytical chemistry techniques such as mass spectroscopy (MS) and nuclear magnetic resonance (NMR) spectroscopy, where sampling times are on the order of minutes and high throughput options exist [19, 71, 72].

Metabolomics studies have been used to identify useful biomarkers for a variety of diseases/conditions such as endometriosis, ovarian and colorectal cancer, and asthma [23–25]. Therefore there is clear precedence for the use of metabolomics in biomarker discovery for sepsis in trauma patients.

Metabolomics has garnered such interest that a study at the Queen Elizabeth hospital in Birmingham, United Kingdom, was commissioned. The SIFTI (Scientific Investigation of biological pathways Following Thermal Injury) study which has already been undertaken was a longitudinal study which collected patient samples and records over a period of 6 months per patient. As the name suggests, patients with traumatic burn injuries

were targetted for the study. Samples have been processed with metabolites labelled which provides quantitative metabolomics data for analysis. This project has been given access to some of this data, namely data from days 1 - 4 since admission to the hospital.

2.3.2 Current developments in identifying biomarkers for sepsis in trauma

There is a clear need for biomarkers which can diagnose sepsis in trauma patients.

Metabolomics is still a relatively new field within medicine when compared to other data driven areas such as genomics. Despite this there are examples of metabolomics studies related to traumatic injury/burns and sepsis.

Orthogonal partial least-squares (OPLS) analysis has been used on metabolic data in a small study of twenty two patients. The result of this investigation was the identification of a metabolic phenotype which was able to predict sepsis with an area under curve (AUC) value of 0.778 for the receiver operating characteristic (ROC) curve [26]. This value still means there is some overlap between the classes but the model should predict correctly more often than not. The small sample size means the results should be validated through a larger study.

Another study used OPLS for a metabolomics study using rats. Potential metabolite biomarkers for sepsis were discovered. The biomarkers discovered are metabolites known to be involved in tissue damage [27]. This research needs to be validated with human serum data.

Metabolomics data is not just being used to predict sepsis. It can also be used to determine injury severity [73, 74]. Although metabolomics data could help with determining mortality, there are already several mortality scoring techniques which can be used and calculated quicker and easier than metabolomics data and so the application of these

findings may be limited.

As was previously mentioned, metabolomics is a new field when compared to the other fields such as genomics. The studies reported here highlight that metabolomics looks to offer promising information on disease states. The criticisms of the studies, namely small sample size or not using human patients are less relevant to the data from the SIFTI study which has been made available to this project. The SIFTI data made available contains 89 patient samples all taken from burns injured patients and so it contains more data than the previous studies mentioned and it is data from humans. Given the successes mentioned for other diseases, the SIFTI data could allow us to find biomarkers of sepsis.

2.4 Machine Learning and data

Specific techniques mentioned in this section which were used for this project are further discussed in more detail in chapter 3.

Machine learning is a computational technique which uses algorithms to process data in order to create models, find patterns or make predictions. The algorithms are able to process new data in order to improve their models and therefore predictive abilities. The output models may also be useful in identifying features of interest in the data sets provided.

In the language of machine learning observations are known as instances, explanatory/independent variables are known as features, and the different values of the dependent variable are called classes or outcomes.

Machine learning algorithms have been applied to fields as diverse as agriculture, bioinformatics, economics and speech recognition. Machine learning is therefore applicable for use in any area where it is possible to collect large amounts of data.

Machine learning techniques can be categorised through the way learning takes place. Two of the most common techniques are supervised and unsupervised learning with the work in this project utilising supervised learning techniques.

2.4.1 Supervised Learning

Supervised learning techniques consist of algorithms which learn a function to map inputs to outputs based on given input-output examples. The given examples labelled with output values are what provide the supervision and are often called training data. A supervised learning algorithm should be able to generalise the training data in order to provide output values for new unseen input data.

In order to test how well a supervised algorithm performs, it is important to test the function with data where the output values are known. Traditionally this is performed by a technique called cross validation which partitions the available data into a set for training and a set for testing. A variety of statistics can then be calculated by comparing the output generated from the function against the known output values for the test data.

Two important use cases for supervised learning techniques are that of classification and regression.

2.4.1.1 Classification

Classification is the process of determining which of a set of categories a data instance belongs to based on the values of its features.

To explain by example, a data instance may contain information on the words used within an email, and it may be desired to use these features to determine if the email belongs to the category of being a spam email or not.

In keeping with the theme of previous sections, a notable classification problem is to determine whether a patient has a specific disease when given some data collected on them. Classification is therefore used a lot with medical data, where it is not only useful in being able to determine the outcomes of patients but by looking at the models produced, it can be possible to determine which features contribute the most to the classification. It is therefore possible to identify biomarkers or clinical metrics useful for diagnostic purposes.

When there are only two classes to decide between, the problems are called binary classification problems and these represent the simplest case for classification. When there are more than two classes the problems are called multiclass or multinomial classification problems.

Some classification algorithms were designed only to handle binary classification problems. Examples of these types of algorithm are logistic regression and support vector machines.

Logistic regression creates a model using the logistic function, whose output value is a number bound between 0 and 1. When creating the model, if the classes used are represented by 0 and 1, the value output by the model when testing a new sample is the probability of the sample belonging to the class represented by 1. This technique is explained in more detail in chapter 3.

Support vector machines work by finding a hyperplane which separates two classes. The algorithm finds the hyperplane which has the maximal distance between samples from both classes and so the classification of a new sample is determined by its position in relation to this hyperplane. There are extensions which allow the use of support vector machines with multiclass problems, however traditionally it is a binary classification technique.

Binary classification can be used to accommodate multiple categories through use of the 1 vs all strategy. The 1 vs all strategy consists of creating a binary classifier

for each class where the outcome is whether an instance is a member of the class in question or not. With this technique, the number of classifiers needed is equal to the number of classes which means the training time will also increase proportionally with the number of classes. This strategy is not the only way to deal with multiple classes and there are classification techniques which can naturally handle multiple classes through the production of only one classifier. Examples of classification techniques which can perform multinomial classification are k-nearest neighbours and naive Bayes.

K-nearest neighbours classifies a new sample to the majority class of the k samples nearest to it in the training data. Naive Bayes creates a joint probability distribution across all features, for each possible class, and uses these to determine which class has the greatest probability given some feature values. Both of these techniques are explained in more detail in chapter 3.

2.4.1.2 Regression

Regression techniques are used when the task is to determine the value of an output, and so the dependent variable is continuous real valued, as opposed to being discrete valued as with classification. In order to do this, the relationship between input and output values must be estimated.

There are many different regression algorithms which can be used to create models for the data. Linear regression techniques use a linear combination of parameters create models. The simplest type, simple linear regression, is for the case of having one independent variable and one dependent variable. The algorithm aims to find a straight line which fits the data, thus linking the independent and dependent variables. With this simple case the independent and dependent variables can be plotted graphically as x and y axes respectively. A line of best fit can be found using the least squares fitting method. The gradient of the best fitting line provides information in how the two variables are

linked as it shows how a change in the value of one will cause a change in the value of the other.

Linear regression techniques can handle multiple independent variables, such as by finding a straight line of best fit in multi-dimensional space. The line of fit doesn't have to be a straight line however, the linear in the name linear regression comes from the parameter values for each feature, and so feature values can be modified by exponents or other operators.

Other regression techniques include polynomial regression, decision tree regression and random forest regression.

2.4.2 Machine Learning with metabolomics data

Machine learning algorithms are used for a variety of different problems. They are commonly used alongside artificial intelligence systems but they are also used with metabolomic data. K-Nearest Neighbours and rule based classifiers have been used to identify metabolites of interest which discriminate between two sets of mice [75]. The algorithms were used with many sets of data, each having different metabolites. By making comparisons with the metabolites and scores it is possible to isolate metabolites which have a large impact on the effectiveness of the model generated. Machine learning algorithms have also proved useful for finding biomarkers for inflammatory stimuli [76, 77].

2.4.3 The curse of dimensionality

The curse of dimensionality describes the problem of when a data set does not have enough samples to provide enough information for all of its dimensions (features) [78]. Data sets can posses an increasing number of dimensions as new and more advanced

data collection techniques become available, such as through increasing the resolution of spectrometers [79].

Having more information about a sample may seem beneficial at first thought, but increasing the dimensionality of data can lead to a data set becoming more sparse. As more dimensions are added, more samples are needed to fill in the data space. To provide an example with metabolomics data in mind, for a one dimensional data set looking at a single metabolite, it would be useful to have samples with concentration ranges from low to high. Imagining if 10 samples were enough to model the outcome based on this data, if a second metabolite were recorded, to create a similar model, 10 data points would be required per existing data point from the one dimensional space. That is for each concentration of metabolite 1, low to high, it would be useful to have concentrations from low to high in metabolite 2 for comparison. This would lead to a data set with 100 samples. If more metabolites are recorded, it can be seen more samples will be required and that this grows exponentially [79].

There is no fixed point for which the curse of dimensionality occurs. The SIFTI data contains 30 features and has 89 samples and so there are not a lot of samples per feature and so there is the possibility of this causing a problem.

One method to manage the curse of dimensionality is reduce the number of features used by a model. Feature selection methods exist which can be used with, or exist as part of machine learning algorithms which perform this, such as using logistic regression with the lasso penalty. Performance will be compared between algorithms using feature selection and those which do not to determine if this is actually an issue with the data set used.

2.4.4 Missing Data

When acquiring data for a study it is not always possible to collect a complete data set. In the omics fields due to the nature of how information is collected and processed it is difficult to identify the same genes, proteins or metabolites across all patients. Even when collecting data via questionnaire, participants may leave some questions unanswered. The information that has been collected is correct and valid and attempts should therefore me made to use it.

There are three types of missing data [80].

1. Missing completely at random (MCAR). Values in a data set are MCAR if the reason for the value being missing is independent of observable variables, unobservable parameters of interest and entirely at random. If data is MCAR, the missing values should not introduce any bias into the data set.

2. Missing at random (MAR). Values in a data set are MAR if the missingness is not at random but can be accounted for by observed variables which have complete information. MAR data is impossible to verify statistically and so substantive reasonableness behind the assumption is used.

3. Missing not at random (MNAR). Values in a data set are MNAR if the value of the missing variable is linked to the reason as to why the data is missing.

Missing data is a problem as some techniques require a complete data set to execute such as logistic regression and (unmodified) k-nearest neighbours which would fail to produce a result otherwise.

It is also possible to introduce bias in data depending on the missingness type such as MNAR data. An example of this bias going back to the questionnaire example would be respondents leaving a question about their earnings empty if they earned over a certain

amount of money. Analysing the data set, large values would be missing despite the fact they can exist and this would skew the recorded data set to lower means and variances.

2.4.4.1 Data Deletion

One of the simplest methods to deal with missing values in a data set is to delete any sample containing missing values until the data set remaining is complete. This technique is called listwise deletion, complete case analysis or case deletion [80–82]. The use of this technique can lead to the removal of useful data, can lead to substantial removal of data samples and if the missing data is not MCAR can introduce bias into models [80, 83].

Another deletion technique is called pairwise deletion. Pairwise deletion does not permanently remove data from the data set, instead the technique deletes samples with missing values only when the missing values are required [81]. This keeps the data set intact and allows recorded values to be used. However when calculating statistics on the features of a data set, the number of samples used for each feature/calculation can be different and this makes it more difficult to make comparisons between different statistics of the same data set.

The two techniques discussed involve deleting samples from a data set or from calculations. Sometimes it is necessary to delete features themselves if they contain substantial amounts of missing values [83].

Neither of these techniques will be used. The SIFTI data set does not contain any complete samples and so listwise deletion would delete the entire data set. Pairwise deletion would not be appropriate to use for this analysis as the analysis is performed on the full data set and not on subsets of features. It is common practice in metabolomics analysis to use imputation techniques [84]. These will be discussed in the next section.

2.4.4.2 Data Imputation

Instead of deleting missing values, imputation techniques aim to fill in missing values thus creating a complete data set. Strategies for determining values to impute are constantly being researched. The simplest strategies involve imputing a single value when a field is missing within a feature. A common value to use is the mean of the recorded values for a feature [80]. Imputing the mean causes the feature mean to remain the same, however it is not always an appropriate value such as when the feature is for a non-continuous value. When feature values are discrete/categorical, the mode may be imputed to have a similar affect. It is important to try to understand why values may be missing and so for metabolomics data, half of the minimum recorded value for a samples is sometimes imputed with the assumption being that the value was too low to be recorded.

The technique of multiple imputation has shown promise as an imputation strategy. For multiple imputation, copies of the original dataset are made and for each copy, the value to be imputed is drawn from a distribution of possible values. Each copy of the dataset will be made into a complete data set through imputation and each copy is used for analysis, with the final output taking into consideration the output when using each copy.

Imputation techniques are commonly used in metabolomics analyses [84]. Common values to impute include half the minimum value for each feature, which corresponds to the belief that the metabolite is present but below the limit of detection of the technique and also the mean, as this will have a less dramatic impact on the feature metrics.

For this work, deletion is not an option and so half-minimum, mean and multiple imputation techniques will be used. There is reasoning behind using the half-minimum value and so it should be explored and this can easily be compared to using mean imputation to compare effects. Multiple imputation has been shown to be resistant to different types of missing data and so it will also be used [85].

2.4.5 Algorithm choice

Three classification algorithms were chosen to be used for this project to perform the metabolmics analysis, k-Nearest Neighbours (k-NN), naive Bayes (NB) and logistic regression (LR).

The k-NN algorithm is a common starting point for classification problems as it is quick to perform. It doesn't require building a model as the model is just the training data. This can be a drawback when a data set is large as even though training is instant, testing would require calculating distances from each test sample to every training instance. Given the size of the SIFTI data this will not be a problem. Another positive aspect is the algorithm is non-parametric and so it makes no assumptions and forces no constraints on how the data should be shaped. However a downside is the algorithm cannot inherently tolerate missing values and so will require using an imputed data set unless modifications are made to the distance calculation. Data will need to be preprocessed before it can be used with the k-NN algorithm so as to standardise features, a z-score will do this, this will stop any features with larger magnitude values from dominating in the distance calculation. Another downside to this classifier is that class imbalance can greatly affect classifier performance. The full SIFTI data set does not exhibit strong class imbalance, however stratifying it into smaller data sets based on patient type/day of sample will cause the imbalance to worsen for some of the stratifications. For most of the stratifications the imbalance will be less than 57 % in favour of one class and so this should not cause a big problem. The k-NN classifier has been used with favourable results with genomics and proteomics data sets and so it should be viable to use for metabolomics data [86, 87].

The naive Bayes algorithm does not require a complete data set and so it will interesting to see how it compares against algorithms using imputation. With continuous variables the naive Bayes algorithm uses the probability distributions for each feature within each class to calculate a probability score. Downsides to the naive Bayes classifier

include performance penalties with continuous data if features aren't accurately modelled and the also the classifier assumes features are independent of one another. This could be a big problem as metabolites are often connected to one another, however if there are many intermediates between two metabolites a lot of biological processes can happen in between and so the relationship between metabolites may not be direct. A recent publication testing several machine learning algorithms for biomarker identification found naive Bayes to perform well and so it is worth testing [88].

An advantage of the logistic regression algorithm is that it can easily be used with regularization and so feature selection is embedded within it. This can reduce overfitting to data and also selects the most important subset of features. Another benefit is that the model produces weights (scores) for each feature used which show the impact each feature has on the output. These weights can be positive or negative which means LR will be able to not only inform which metabolites are useful, but also if it is an increases or decrease in their concentrations which is important. Like k-NN, LR can't tolerate missing values and so imputed data sets will need to be used, also data needs to be standardised so the feature weights can be comparable. Logistic regression has been used with metabolomics data to predict outcome and also identify metabolites of interest in patients with multiple sclerosis [89].

The previous section has covered topics to do with machine learning and data processing. The next section looks at diagnostic testing.

2.5 Diagnostic Tests

When biomarkers have been identified to diagnose a condition, it is important to develop analytical tests which can be performed to detect/measure the chosen biomarkers. These tests can typically fall into two categories, laboratory-based and point-of-care based tests.

As the name implies, laboratory based testing is when tests are performed in a laboratory.

Point-of-care testing, sometimes called bedside testing, is defined as "diagnostic testing at or near the site of patient care, wherever that medical care is needed" [90]. The definition specifically addresses that the testing should be available wherever the care is needed, as such POC tests and testing devices must be relatively small, portable and unobtrusive as patients can present themselves not just in hospitals but also GP surgeries/outpatient centres and even in public emergencies where paramedics may be required to perform such tests [90].

It will typically take more time to get the results from laboratory based tests than POC tests. Lab tests will require a trained professional to perform the test and verify the results, whereas POC tests can normally be performed by personnel not trained in clinical laboratory sciences. This means POC devices can be used in more settings and even by patients themselves.

It has been reported that lab based tests can offer better sensitivity, specificity, precision and accuracy when compared to POC tests [91]. The difference between accuracy of laboratory and POC tests is not always significant and will depend upon the tests being performed. A comparison of laboratory and POC tests for procalcitonin showed no significant difference between the quality of the test, despite the laboratory test taking two hours longer [28]. A comparison of laboratory and POC glucose analysis from critically ill patients determined there was no significant difference between the POC and lab based results [92]. However a later study in glucose testing in critically ill patients found the POC tests differed by as much as up to 20 % from the laboratory test value, with a further study recommending the verification of hypo and hyper-glycemic results from POC devices with laboratory tests [93, 94]. A study analysing differences between lab and POC tests for troponin in an emergency department found no significant difference and recommended the use of the POC testing to provide the fasted results [95]. A study

looking at monitoring haemoglobin in patients concluded that the POC test results were the same as the lab tests, but the POC test offered the ability to provide continuous measurements and so would be seen as an improvement over the lab based tests [96]. As such it can be seen that there is conflicting evidence on the differences in quality of POC tests when compared to lab test results. The conclusion should be it is important to judge each test individually, as a sweeping statement suggesting POC tests are less accurate does not make much sense given the number of different POC tests available.

Despite conflicting results regarding the accuracy of POC tests, there is a huge move towards POC testing.

In order to develop a POC test for a biomarker there needs to be a way to selectively identify the biomarker. Antibody based systems have been used in the past due to their high specificity. However the use of antibodies has its own negatives such as high cost, difficult preparation and issues with physical and chemical stability [97]. As such, fluorescent chemosensor can be utilized. These will be discussed in the next section.

2.6 Fluorescent Chemosensors

A fluorescent chemosensor is a compound with a binding site, a fluorophore and a mechanism by which the two sites can communicate [98]. There are a huge variety of fluorescent chemosensors and this section will attempt to cover some of the application domains they cover, but first, fluorescence itself will be explained.

2.6.1 Fluorescence

Fluorescence is the phenomenon of a molecule absorbing and then subsequently emitting light. The absorbed photon promotes an electron into a higher energy level and a new

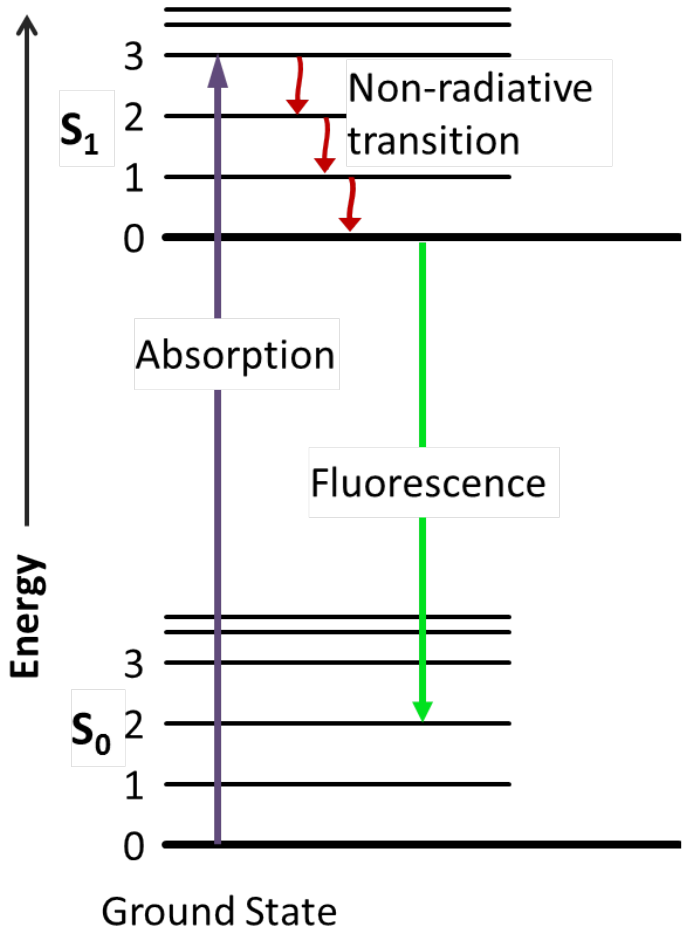


Figure 2.4: Example Jablonski Diagram. S_0 and S_1 label electronic energy levels. The labels 0 - 3 label vibrational energy levels within each electronic energy level. The diagram shows an electron being promoted to a higher energy level through the process of absorbing energy (purple arrow) such as from a photon of light. The electron then loses energy through non-radiative transitions (red arrows). The electron then returns to the ground state electronic energy level (green arrow) and as it does this it loses energy which for fluorescence is in the form of a photon.

photon is released upon relaxation back to the ground state. Typically the photon emitted will be of a longer wavelength and thus lower energy than the absorbed photon, however it is possible for the emitted photon to have more energy than the absorbed photon in the case of upconverting nanoparticles. Upconverting nanoparticles are a special case where multiple photons are absorbed per each emitted photon and so the emitted photon has higher energy [99]. The difference in wavelength between absorbed and emitted light is called the Stokes shift. The causes for this difference will typically be dependent upon the fluorophore and the environment it is used in, however common causes for the emission of lower energy photons include electron energy loss through vibrational relaxation and also electrons entering higher vibrational energy levels within their ground states. The energetics of fluorescence can be illustrated with the aid of a Jablonski diagram with an example presented in fig. 2.4.

2.6.2 Fluorescent chemosensors for the detection of ions and biomolecules

Fluorescent chemosensors have been around since 1867, where the first reported fluorescent chemosensor was a technique to detect the aluminium ion (Al^{3+}) [100].

Since then, many chemosensors have been developed for other ions, such as those essential to live, such as sodium (Na^{+}) [101], potassium (K^{+}) [102] and calcium (Ca^{+}) [103, 104]. These chemosensors typically have an large aromatic structure with a site the metal ion may complex to.

It has been shown that chemosensors can be used with biological samples where a range of ions can be detected such as chloride ions, oxygen for the determination of blood oxygen concentrations and hydrogen ions for the determination of pH [105].

Chemosensors have come a long way form just being able to detect ions in solutions.

They are now actively developed to detect biomolecules. One group reports the ability to detect the metabolite α-Ketoglutarate within serum using a fluorescent chemsosensor [106]. This is currently a lab based test but as it works in serum the possibility to move to a point-of-care system is well within reach.

A fluorescent chemosensor to detect glutathione, a biomarker for cancer, and Cu^{2+} ions has been developed [107]. This chemosensor is actually colorimetric and so it exhibits visible colour changes upon binding to gluththione or Cu^{2+} ions.

As has been shown, chemosensors can be made to detect a multitude of ions and molecules. The final part of this chapter will look at their development into point-of-care devices.

2.6.3 Chemosensors and POC testing

Detecting fluorescence typically involves the use of large instruments such as fluorometers and microplate readers. However small benchtop/handheld fluorescence detectors have been designed [108–110]. As specific fluorescence assays will only require one excitation wavelength and one emission wavelength, large and expensive monochromator/filter wheel based systems are not necessary, more simple optical and filter based systems can be used.

For some fluorescent assays there are visible colour changes are high analyst concentrations. There will still be a colour change at lower concentrations, it will just be more subtle, but that was using the human eye. Cameras can more accurately detect the difference between two colours. One such point-of-care system to exploit this was designed to detect haemoglobin within blood samples using acoustic waves to modulate the fluorescent particles and a smart phone to detect the colour changes [111].

Another potential avenue for a point of care device which has been developed is the use of microfluidic paper-based analytical devices in conjunction with a smart phone to take

photos and analyse the results [112]. Continuing the theme of combining microfluidics with fluorescence, there also exists a point-of-care device which uses and optical fibre spectrometer with a microfluidic chip to detect avian influenza virus [113].

As can be seen fluorescence sensing devices are leading the way in advancing point-of-care diagnostic technology. Chemosensing platforms can provide an advantage over traditional analytical techniques for biomarkers such as nuclear magnetic resonance spectroscopy, mass spectrometry and chromatographic techniques as these existing techniques cannot currently be miniaturised to make them bedside instruments.

2.7 Conclusion

This concludes the background and literature review chapter. The next chapter deals with the machine learning techniques and methods used for the metabolomics analysis part of this work.

CHAPTER 3

MACHINE LEARNING TECHNIQUES, METHODS

3.1 Introduction

One of the aims of this project was to try to determine whether metabolic data can be used to create an accurate predictive model for the development of sepsis in patients suffering from burns injuries.

This chapter will discuss the techniques used in this project to build a classifier for sepsis, and also the attempt to perform feature selection in order to pick out the most relevant metabolites for classification.

The techniques used are:

1. k-Nearest Neighbours
2. Naive Bayes
3. Logistic Regression

These techniques are all supervised learning techniques [114, 115]. This means that they require a labelled dataset for training a model which can then be used to classify

new data samples. Given that this property is true for all of the techniques used, first it will be discussed how the data is partitioned into training and test datasets, subsequently preprocessed, and then the discussion will move onto how the individual techniques work. Finally feature selection and methods of evaluating classifier performance are discussed.

3.2 Training and Test Datasets

As was mentioned before, the techniques used are supervised learning techniques, and so the algorithms must be provided with data to train a model on. In order to test the model produced

To test how well the algorithms work, we need to provide them with data to train on and then use different data for testing. The data used to test the algorithms needs to be labelled, and so the classes of each sample will be known, in order to determine whether the classifications provided by the algorithms are correct. Because of this, completely new data where the outcome is unknown can not be used to test the classifiers, but instead data from the original dataset will be used. To do this, the original dataset must be partitioned. There are many ways to partition a dataset to provide training and test sets, but this work uses the technique of k-fold cross-validation.

3.2.1 k-fold cross-validation

For k-fold cross-validation, the full dataset is partitioned into k folds. For this research stratified k-fold cross-validation was performed and so the distribution of classes in each fold will be similar to the distribution of classes across the whole dataset where possible. By keeping the class distributions in each fold similar to that of the entire dataset, the training set should relate more closely to the complete dataset as opposed to if random

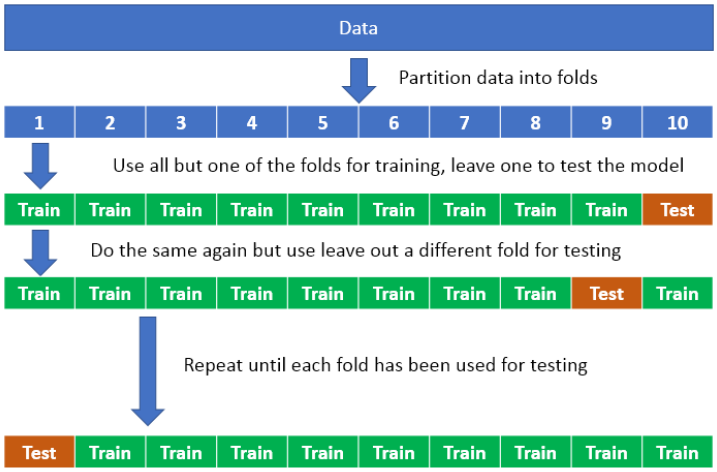


Figure 3.1: Example of k-fold cross-validation where k = 10.

sampling were used to determine each fold. After the folds are made, k rounds of testing are performed (one per fold), with each round using k-1 folds as training data and the remaining fold as test data. Each round uses a previously unused fold for the test fold. This allows for all of the data to be used as test and training data and results can be collected by analysing the performance from each fold.

3.3 Z-score

For the k-nearest neighbour and logistic regression techniques the data was z-scored before use. This means that for each metabolite, all values had the metabolite mean subtracted from them and were divided by the metabolite standard deviation, which gave each metabolite a mean of zero and a standard deviation of one.

$$\text{z-score(x)} = \frac{x - \bar{x}}{\sigma} \tag{3.1}$$

This is a method of standardisation, and is a useful way to preprocess the data. Depending on the technique used to analyse the data, metabolites which have naturally

greater concentrations could overshadow those which naturally have lower concentrations. The absolute values are not as important as the differences between samples, and so standardising the data in this way stops metabolites with larger values from dominating the classification.

3.4 k-Nearest Neighbours (k-NN)

K-Nearest Neighbours is a non-parametric machine learning algorithm which can be used for both regression and classification problems. The k-NN algorithm can be described as a lazy algorithm as it uses the complete set of training data provided to it without any modification, as such it does not produce any models of its own. Because of this, training is very fast and easy, the data just needs to be partitioned and the creation of the training set is the training stage. It is during the testing stage that the complexity increases, as the distances between each training sample and each test sample need to be calculated in-order to determine which k samples are closest (the k-nearest neighbours) to each test sample. When multiple neighbours are used, the new data point will be assigned to the class which has the majority neighbours.

Mathematically the algorithm for the classifier works as follows [116, 117].

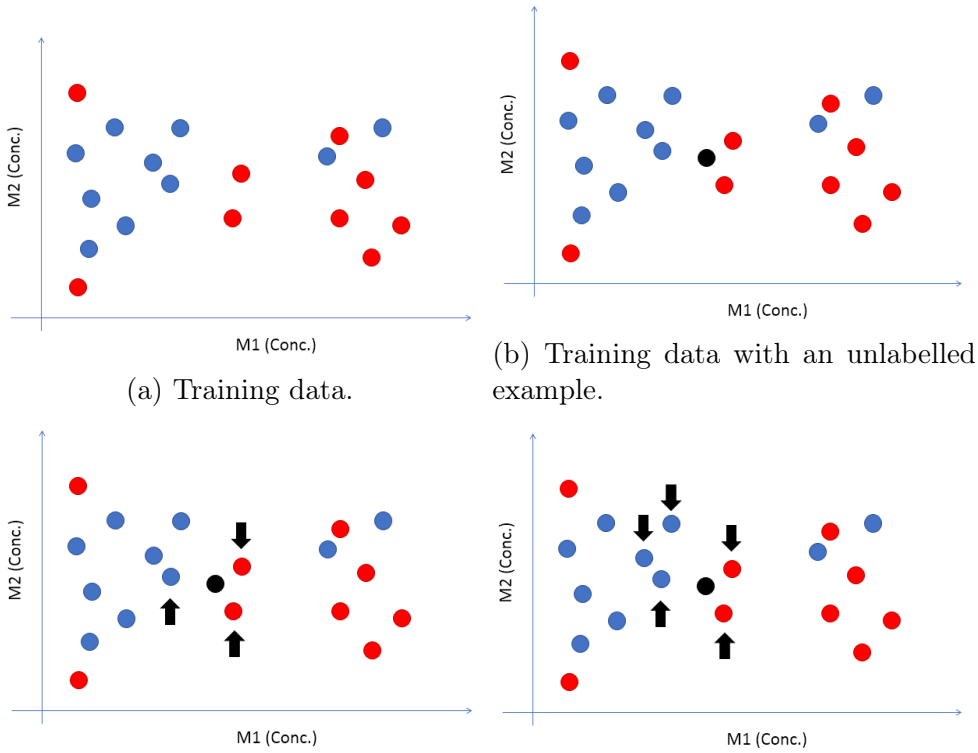
$$y_t = \underset{c \in \{c_1,\ldots,c_n\}}{\arg\max} \sum_{x_i \in N(x_t,k)} I(y_i = c) \tag{3.2}$$

Where $I(.)$ is the indicator function which returns 1 if the condition evaluates to true or 0 if the condition evaluates to false. And where $N(x_t, k)$ returns the set of k nearest neighbours of sample X_t from the training data.

Many uses for k-NN only require classification and so eq. (3.2) is applicable . For medical data, amongst other types of data, it is typical to produce Receiver Operating

Characteristic (ROC) curves to assess a classifier. To create a ROC curve, probability values are required for a classification and the classification threshold is changed to determine how the classifier reacts to such a change. In order to get probabilities from a k-NN classifier one simply needs to take the sum of each class occurrence within the k nearest neighbours and divide it by k, so the probability is equal to the proportion of the class occurrences as is given by eq. (3.3) [118].

$$P(Y = c \mid X = x_t) = 1/K \sum_{x_i \in N(x_t,k)} I(y_i = c) \tag{3.3}$$



(a) Training data.

(b) Training data with an unlabelled example.

(c) When $k = 3$ the new sample will be classified to the red class.

(d) When $k = 5$ the new sample will be classified to the blue class.

Figure 3.2: An illustration of how to classify using k-NN.

Figure 3.2 illustrates how the k-NN classifier works with a simple example. First it shows a training dataset with two classes indicated by different colours (red and blue), then, a new sample (the black circle) is introduced which needs to be classified. The

nearest neighbours for k values of 3 and 5 are shown, indicating the effect changing k can have on the quality of the classification.

3.4.1 Calculating Distances

The k-NN algorithm requires a distance function in order to determine which training points are nearest to the test point. For this, the square distance was used which is similar to the Euclidean distance function but without calculating the square root, as shown in the equations below:

Given two points p and q of n dimensions,

$$\text{Euclidian Distance} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{3.4}$$

$$\text{Squared Distance} = \sum_{i=1}^{n}(p_i - q_i)^2 \tag{3.5}$$

The square distance can be used for this algorithm as the distance value is only used to determine which neighbours are the nearest. As the differences between points in the same dimensions are squared, the distance values are all positive. If the square root of the distance values were taken, larger squared values will have larger square root values and so the points considered using the square distance are the same points which would be used if the Euclidean distance were used. Not calculating the square root speeds up computation time as fewer calculations are needed [119, 120].

Equation (3.5) requires that the two samples have the same dimensions. As mentioned earlier this is not the case for all samples. In order to be able to calculate a distance between all pairs of samples the data and/or the distance equation needs to be modified.

There are multiple methods which can be employed to handle missing values [121]. For this project it was decided to try three methods:

1. Leave the cells empty, no imputation (NI).

2. Half-min Imputation (HM). Replace the missing values in each column with half of the minimum value in the column. This technique makes the assumption that the metabolites are present but just in small concentrations and so may not have been detected.

3. Mean Imputation (M). Replace the missing values in each column with the column mean.

When half-min and mean imputation are used, eq. (3.5) can be used. For method 1, the distance function is modified such that eq. (3.6) below is used.

Given two points, p and q, which have comparable data for m metabolites:

$$\text{Distance} = \frac{1}{m}\sum_{i=1}^{m}(p_i - q_i)^2 \tag{3.6}$$

The samples given to eq. (3.6), p and q, have comparable data in m metabolites, but either may have concentration values for metabolites not present in the other sample. This data must be ignored as there is nothing to compare it to.

3.4.1.1 Comparing samples with missing values

When leaving in samples/metabolites with missing values, comparing samples to calculate a distance value requires another step.

In order to make the comparison it is necessary to find out which metabolites two

Table 3.1: AND logic table.

Input 1	Input 2	Output
0	0	0
0	1	0
1	0	0
1	1	1

samples have in common. To do this efficiently the AND operator is used. The logic table for AND is shown below.

It is possible to create logical matrices in Matlab which have a 1 where an actual concentration value is and a zero where there is none. This step is facilitated by filling in the empty cells in the data spreadsheet with an unrealistically high value (e.g. 100,000). This value will not appear as an actual concentration in the dataset and so it is used as a placeholder and can be searched for to find the empty cells. It is possible to do this without a placeholder value, by using not a number (NaN) values, however this limits the functions which can be used, as NaN values in Matlab do not behave as traditional numbers.

By creating logical vectors for each pair of samples to be compared, the logical vectors can have the AND operation performed on them. The result is another logical vector. For this to work, both vectors p and q need to be of the same dimensionality (n) and the resultant vector will also be of n-dimensionality. On vectors (and matrices), the AND operation compares values with the same index (the same position in both vectors/matrices). From table 3.1 it can be seen that when using logical values to indicate concentration values being present, the resultant value will only be a 1 if both input values are 1, and so the metabolite needs to be present in both samples as illustrated in fig. 3.3.

The sum of the values in the output vector provides the number of comparable metabolites (value m in eq. (3.6)). To calculate the distance according to eq. (3.6), first both data points (the actual concentration vectors, not the logical vectors) are multiplied by the logical output vector. This will zero the unmatched metabolites and only keep the

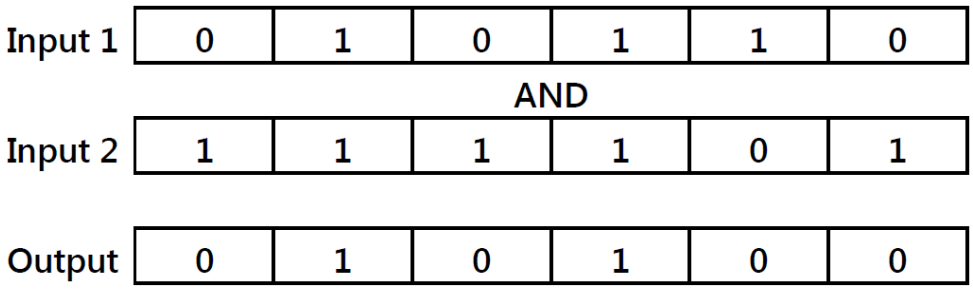


Figure 3.3: Example of two samples having the AND operation applied to them.

comparable metabolite concentrations. The squared distance function can now be used from eq. (3.5) as the zeroed values will not contribute to the overall distance. The distance can then be scaled by dividing by the number of comparable metabolites (m).

The half-min and mean imputed samples will have no missing values and so when using those techniques all samples can be compared directly to one another without needing the above check.

3.4.2 Classification

Once the distances have been calculated, the nearest k points are selected and the number of neighbours for each class is calculated.

The number of neighbours used is user defined and will affect the quality of the outcome. There are many heuristics for the optimal value of k, such as using $k = \sqrt{n}$ where n is the number of samples in the training data. Despite the heuristics available, the optimal value for k is very data dependant and so in order to determine the optimal value for k, the algorithm should be run for multiple k values and the value which is consistently the best should be chosen. In order for the classifier to be robust, therefore not affected by noise in the dataset, the value for k should be sufficiently large so noise will be ignored.

3.4.2.1 Scaled Voting

The data provided is subject to class imbalance, as such one of the classes is a more common outcome than the other. Having imbalanced classes can affect the quality of results produced by the k-NN algorithm. To try to lessen the effect, the totals for neighbours in each class are scaled according to the class distribution.

In the code, the number of neighbours from the non-sepsis class is scaled, so it is as if there were the same number of samples for each class.

3.5 Naive Bayes (NB)

The naive Bayes algorithm uses Bayesian probability statistics to generate a probability value for a data sample belonging to a specific class. As the name implies, it utilises Bayes' theorem which is given in eq. (3.7). Bayes' theorem provides the value for a conditional probability, that is the probability of an event happening knowing that another event has already happened. This is relevant to the classification problem as we would like to know the probability of a patient being in the disease state, given the information collected on them. Bayes' theorem is often expressed in a non-mathematical way, such as in eq. (3.8). When looking at eq. (3.7), for the classification problem the hypothesis would be the patient having sepsis, and the evidence would be their metabolomic data. In order to calculate this posterior probability, it is necessary to calculate the likelihood, which is the probability of a patient having that metabolomic data (the evidence) given that they already have sepsis (the hypothesis), and this is multiplied by what is called the prior. The prior is what we currently believe the probability of the hypothesis being true to be. Given the example being used, the prior for a patient having sepsis could be the calculated as the fraction of patients admitted who had sepsis against the total number of patients admitted. The probability of the evidence in this example will be the probability of seeing

the patient's metabolomic data irrespective of their outcome. Therefore the evidence is calculated as the probability of seeing the metabolomic data given a patient having sepsis multiplied by the (prior) probability of having sepsis, and adding to this the probability of seeing the metabolomic data given a patient not having sepsis, multiplied by the (prior) probability of not having sepsis.

$$P(Hypothesis \mid Evidence) = \frac{P(Evidence \mid Hypothesis)P(Hypothesis)}{P(Evidence)} \tag{3.7}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \tag{3.8}$$

$$P(C_1 \mid x) = \frac{P(x \mid C_1)P(C_1)}{(P(x \mid C_1)P(C_1)) + (P(x \mid C_0)P(C_0))} \tag{3.9}$$

Naive Bayes Classifier for a binary classification problem (classes being 0 or 1), calculating the probability for a sample belonging to the class labelled 1.

$$P(x \mid C_i) = \prod_{j=1}^{d} P(x_j \mid C_i) \tag{3.10}$$

$$P(Evidence) = \sum_{i} P(x \mid C_i)P(C_i) \tag{3.11}$$

Equation (3.9) shows the formula behind the naive Bayes classifier for a binary classifier where the result is the probability of a sample x belonging to class 1.

Given a dataset of d dimensions and a sample x which also has d dimensions, the naive Bayes classifier calculates the probability values of each individual feature appearing in a

chosen class (in the equation class 1 is chosen). The feature probability values per class are multiplied together to give the likelihood of having sample x given the chosen class. The likelihood value is multiplied by the prior probability of the class to give the posterior probability.

The naive Bayes classifier can easily be used with features which are discrete/categorical as the likelihood and prior values can easily be calculated. For the likelihood values $P(X = x \mid C_i)$ can be calculated by $\frac{\sum_i^N X_i=x}{N}$ where N is the number of samples. The prior probability $P(C = c)$ can be calculated by $\frac{\sum_i^N C_i=c}{N}$.

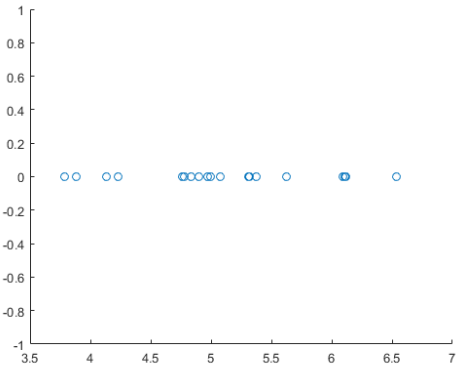
Continuous data is not as simple to work with. As has previously been mentioned, for discrete data it is only necessary to tally the total number of observations of each feature value across all samples and divide by the total number of samples. For continuous data, this is typically not possible, as it is possible there are values which are valid but which have not been seen in the training data. One method to deal with continuous data is to discretize it.

3.5.1 Probability Density Functions

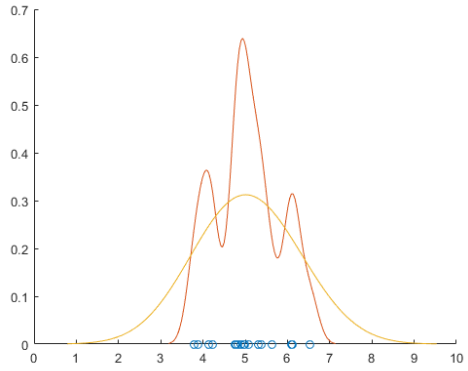
In order to use the naive Bayes classifier with the metabolomics dataset, probability density functions (PDFs) needed to be built for each metabolite. Probability density functions transform a set of discrete data values into a continuous probability function. When looking at the denominator of Bayes' Theorem the probability of the evidence needs to be calculated, which in the case of this project would be the probability of a patient's set of metabolite values appearing. It is therefore necessary to construct PDFs because the metabolite concentrations can be any from a range of continuous values, but the dataset itself will only contain a small subset of these values. If testing the probability of a specific value for a metabolite appearing in the dataset, it is highly likely there will be no repeat, and this is true for all samples. Part of this is due to the nature of the

data. The metabolite values are recorded to a precision which would require many more samples to have a good chance of seeing the same value more than once. As can be seen in the denominators of eqs. (3.8) and (3.9), the probability of the evidence for a binary classification problem is the summation of the probabilities of the evidence coming from either of the available classes. In order to calculate these probabilities, PDFs are created for each metabolite, per class.

Probability Density Functions were constructed using the Parzen–Rosenblatt Window technique. This works by fitting a kernel function around each point and summing up the values from all of them. There are a variety of kernels which can be used, with a popular one being the normal (Gaussian) distribution. Figure 3.4 provides an illustration of a PDF generated for a metabolite. An important parameter when using the Parzen-Rosenblatt Window technique is that the PDF generated can differ greatly based on the bandwidth used for the kernel function. When using a normal distribution, the bandwidth affects how wide the peak will be. If the bandwidth is two narrow, there will be regions of extremely low probability between clusters of datapoints. If the bandwidth is too wide, it will smooth out the clusters of datapoints. Figure 3.5 shows examples of how changing the bandwidth can change the distribution produced. The bandwidth here can be seen as an analog to bin widths when plotting a histogram. A histogram can also be used to visualise the distribution of datapoints, however histograms do this in a discrete way due to the use of bins. If the bins too wide, the data will be split between very few bins, if the bins are too narrow, they may fail to collect multiple points to show an underlying density distribution.

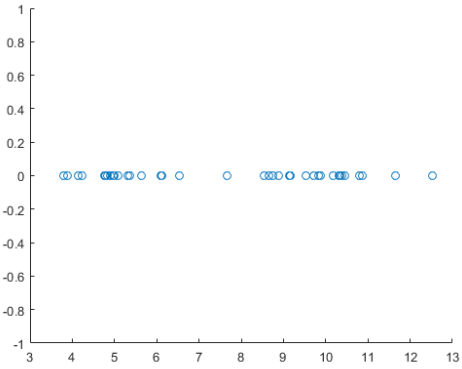


(a) 20 points taken from a normal distribution with Mean = 5, Var = 1.

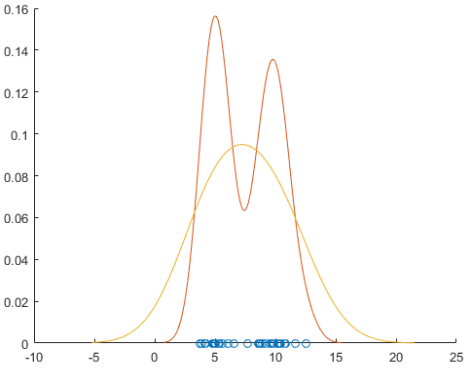


(b) Example PDFs. Bandwidth = 1 for yellow PDF, bandwidth = 0.2 for orange PDF.

Figure 3.4: Samples data points from a normal distribution and PDFs using large (1) and small (0.2) bandwidths. The large bandwidth produces a smooth normal distribution whereas the small bandwidth produces a distribution with multiple peaks.



(a) 40 points from two normal distributions. Half with Mean = 5, Var = 1, half with Mean = 10, Var = 1.



(b) Example PDFs. Bandwidth = 3 for yellow PDF, bandwidth = 1 for orange PDF.

Figure 3.5: Samples data points from two normal distributions and PDFs using large (3) and small (1) bandwidths. Using a large bandwidth creates a smooth normal distribution whereas the smaller bandwidth produces a distribution with peaks around the means of each original distribution.

3.6 Logistic Regression (LR)

Logistic Regression (LR) is a classification technique which falls under a class of techniques called generalized linear models (GLM). LR is typically used for binary classification

problems, and whilst it can be extended to multiple-class problems, for this project binary classification is the goal [118]. Unlike the k-NN and NB techniques seen previously, LR creates a model from the training data, and this model fits data using the logistic function eq. (3.12). The logistic function produces a sigmoid curve, as can be seen in fig. 3.6. This curve is bound between 0 and 1, and so the output can therefore be interpreted as a probability.

$$f(x) = \frac{e^x}{1 + e^x} \tag{3.12}$$

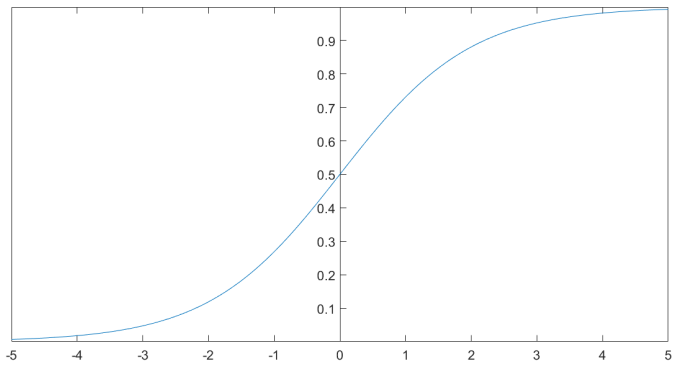


Figure 3.6: Example of a Logistic Curve

The model takes the form of $\beta_0 + \beta_1 * x_1 + \ldots + \beta_m * x_m$ where β_0 is the intercept for the model, $\beta_1 \ldots \beta_m$ are weights for each feature and $x_1 \ldots x_m$ are the feature values from a sample. LR is therefore a parametric classification technique, with the intercept and weights being the parameters which need to be computed. These parameters are estimated through a technique called maximum likelihood estimation (MLE). MLE maximises a likelihood function which provides an estimate for how good the predictions will be when using a given model given some training data. An example of the likelihood function is given in eq. (3.13) which calculates the residual sum of squares (RSS) error The mathematics behind MLE are beyond the scope of this project, however it is a procedure which has been optimised in many software packages and this project will take advantage

of that fact.

$$\text{Likelihood} = \min \arg \text{beta} \sum_{i=1}^{n} (y_i - f(X_i))^2 \tag{3.13}$$

In order to get a probability value from the model, the β values are used with the feature values of a sample and eq. (3.14) is used. The binary LR classifier returns a probability value for the outcome evaluated being true. For this project the outcome is a patient having sepsis, and so the probability value returned is the probability of the patient having sepsis.

$$P(Y = 1 \mid X) = \frac{e^{(\beta_0 + \sum_{i=1}^{m} \beta_i * x_i)}}{1 + e^{(\beta_0 + \sum_{i=1}^{m} \beta_i * x_i)}} \tag{3.14}$$

3.6.1 Multiple Imputation

The logistic regression technique requires a complete data set to work and so imputation will be required to use it. For this classifier the technique of multiple imputation was chosen.

The process of multiple imputation follows creating models for each feature in a data set and then using those models to generate values to impute into the incomplete data set. This process is repeated multiple (n) times, creating n different data sets upon which classifiers are produced for each imputed data set [80]. When testing is performed, each test case will be used with all n models, producing n different results which are then averaged to produce one overall result [122]. In the case that a test sample has missing values, multiple imputation is performed on it n times and all n new complete samples are passed to each of the n classifiers and the output is averaged across all responses [122].

The technique used to model each feature will be to create PDFs for each feature.

PDFs have been explained under the naive Bayes section, however unlike for the NB classifier which creates PDFs for each feature per class, for multiple imputation PDFs will be created only per feature, with all classes used [122].

3.7 Feature Selection

Feature selection will be important in determining a subset of metabolites. This work will use three different methods of feature selection. The first method is using the lasso penalty with logistic regression which is an embedded feature selection technique, which means it is incorporated in the model creation algorithm. The other two feature selection techniques are the use of t-test and minimum redundancy maximum relevance (MRMR) algorithms, which are examples of filter based feature selection techniques, which means they act on a data set before a model is created.

3.7.1 Lasso

The logistic regression technique creates a model by assigning weights to each feature. The value of the weights show how a change in the feature value will change the probability. The sign of the weight is important as it shows whether a feature has a positive or negative affect on the outcome. For this project, a positive weight means that an increase in a metabolite will cause an increased in the probability for sepsis and conversely a negative weight means an increase in metabolite concentration decreases the likelihood of sepsis.

By standardising the data through the use of z-scores, the weights show the contribution each metabolite has on the overall probability, and so metabolites with greater weights can be seen as more important to the outcome.

To take this a step further, a LR model can be created whilst using a penalty function.

A typical penalty function is the lasso penalty also known as the l_1 norm and is shown in eq. (3.15). This penalises a model based on the absolute value of all of the weights, and so models which assign non-zero weights to more features are penalised more heavily, thus this penalty promotes reducing the number of features used [123].

$$\text{Penalised likelihood} = \text{Likelihood} - \lambda \sum_{i=1}^{m} |\beta_i| \tag{3.15}$$

where λ is a parameter which scales the penalty value. A very large λ value will cause large penalties for non-zero weights and so the models produced will favour having most weights set to zero. A small value for λ will cause models with few zero weights, and so the value of λ needs to be tuned to produce models with the desired number of features.

3.7.2 Two-sample t-test

The two-sample t-test determines if there is a significant difference between two populations of data. For feature selection for a binary classification problem, the two-sample t-test is performed on each feature, splitting each feature into two groups based on the class label. The t-test is then performed to determine the significance of the difference between the two groups.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \tag{3.16}$$

where $\bar{x}$ and $\bar{y}$ are sample means, s_x and s_y are the sample standard deviations, and n and m are the sample sizes.

P-values are determined from t values from eq. (3.16) using degrees of freedom and values from Student's t-distribution [124]. Whilst it is common to threshold based on p-values when claiming significance, for feature selection, features can be ranked by p-

value for comparison. Whilst simple, the t-test has been shown to perform well with high dimensional biological data and so it should perform well with metabolomics data [125].

3.7.3 Minimum redundancy maximum relevance (MRMR)

The MRMR algorithm is more complicated than the t-test. It works by calculating the mutual information between pairs of features eq. (3.17). The mutual information measures how much uncertainty can be reduced from one variable by knowing another. For use with continuous data, the data can be discretized by binning. Equation (3.18) which determines the relevance of a feature set which needs to be maximised and eq. (3.19) which determines the redundancy in a set of features which needs to be minimised in order to rank features [126]. The MATLAB implementation of the MRMR algorithm discretizes continuous variables by binning into 256 bins, or the number of values in a feature if this is less than 256.

$$I(X,Z) = \sum_{i,j} P(X = x_i, Z = z_j) log \frac{P(X = x_i, Z = z_j)}{P(X = x_i)P(Z = z_j)} \tag{3.17}$$

where X and Z are features.

$$V_s = \frac{1}{|S|} \sum_{i \in S} I(c, f_i) \tag{3.18}$$

where c is the class labels, f_i are features, S is the set of features considered, |S| is the size of the feature set.

$$W_s = \frac{1}{|S|^2} \sum_{i,j \in S} I(f_i, f_j) \tag{3.19}$$

where f_i and f_j are features, S is the set of features considered, |S| is the size of the feature set.

3.8 Evaluating Classifier Performance

The accuracy of each technique can be calculated in terms of the percentage of the test data which is accurately classified. This can be calculated per fold of class validation and then averaged [118].

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^{k} \text{Accuracy}_i \tag{3.20}$$

Where k is the number of folds used for cross validation and $Accuracy_i$ is the percentage of correctly classified samples for a specific fold.

The accuracy measurement is not always the best statistic to evaluate classifier performance however, as the accuracy value alone does not provide information on the samples the classifier performed well or not very well on. An example of when the accuracy measurement would not be a good enough measurement would be when trying to diagnose a rare medical condition. If we had a dataset where only 1 in 1000 samples were from a patient with the condition, and we were to train a classifier on this dataset, a classifier which simply predicted not having the condition every time would have an accuracy of 99.9 %. Looking at the accuracy alone, this may seem like a great success, but it ignores the importance of being able to provide an accurate positive classification for the condition.

In order to better evaluate the performance of a classifier, a confusion matrix can be created. A confusion matrix is a table which shows the differences between the predicted outcomes from a classifier against the actual known outcome. This way, the number of true positives and true negatives can be seen, along with false positives and false negatives. A true positive is a correct positive prediction, and true negative is a correct negative prediction, a false positive is a positive prediction for a negative sample and a false negative is a negative prediction for a positive sample.

Table 3.2: An example of a confusion matrix which shows how true positives, false positives, true negatives and false negatives can be determined based on the predicted and actual classifications of samples.

		Actual Class	
		Disease	Not Disease
Predicted Class	Disease	5 (True Positive)	1 (False Positive)
	Not Disease	4 (False Negative)	2 (True Negative)

Some useful metrics to calculate from the TP, TN, FP, FN values are given below. The True positive rate (TPR) is also called sensitivity or recall is shown in eq. (3.21).

$$\text{TPR} = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{3.21}$$

True negative rate (TNR) is also called specificity or selectivity given in eq. (3.22).

$$\text{TNR} = \frac{TN}{N} = \frac{TN}{TN + FP} \tag{3.22}$$

One way of displaying the data from a confusion matrix, which is commonly used with medical data, is to create a receiver operating characteristic (ROC) curve [127]. A ROC curve plots the true positive rate against the false positive rate of a classifier output whilst changing the classification threshold. The classification threshold is the threshold value which is used to determine a classification, so for a threshold value of 0.5, any values returned from the classifier <0.5 would be classified as the negative class whereas values from ≥ 0.5 would be classified as the positive class. By changing the threshold value, the values in the confusion matrix will change and a ROC curve shows this change to give an idea of how good the classifier is and also for comparison against other classifiers. An example of a ROC curve is given in fig. 3.7.

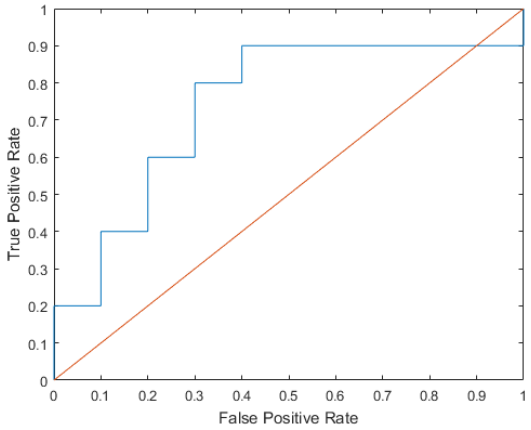


Figure 3.7: An example of a ROC curve.

When using a ROC curve, the area under the curve (AUC) value is also provided for each classifier. The AUC value gives an indication of how good a classifier performs, with a perfect classifier having an AUC value of 1 and all other classifiers having a value <1 [127].

ROC curves and AUC are not affected by class imbalance in a data set which can be a common issue for medical data sets, however they are still used extensively in the literature and so this work will do the same [59–61, 127].

$$\text{PPV} = \frac{TP}{TP + FP} \tag{3.23}$$

$$F_1 Score = 2 * \frac{PPV * TPR}{PPV * TPR} = \frac{2TP}{2TP + FP + FN} \tag{3.24}$$

Precision-recall curves plot recall against precision and do account for class imbalances [128]. These types of curves do not appear often in literature for clinical applications and so they will not be used. However as class imbalance is a problem, the F_1 score will be provided as the interpretation is simple, a value of 1 indicates the best performance with the lowest value possible being 0. The equation behind the F_1 score is given in eq. (3.24).

It uses the previous equations and also the positive predictive value which is also called precision which is given in eq. (3.23).

CHAPTER 4

METABOLOMICS ANALYSIS

4.1 Data

The following information concerns the data used for this project and is taken from the thesis of Dr Christopher Wearn who provided the data for this project [129].

The data used in this project was collected from patients as part of the SIFTI study (Scientific Investigation of biological pathways Following Thermal Injury) (UKCRN ID: 13654) and were admitted to the adult Burns Centre at Queen Elizabeth Hospital Birmingham (QEHB), UK (2012-2014).

The full information on patient recruitment and treatment is included in the thesis of Dr Wearn [129]. The information relevant to this project regarding the metabolomics data has been included below.

4.1.1 Data collection and processing

Blood samples were collected from patients, allowed to clot, centrifuged and then aliquots were stored at -80 °C. When needed, a sample would be thawed, centrifuged again with

a fraction taken from the middle of the sample so as to avoid the lipid rich top layer to process further. The fraction would then centrifuged in a 3000 Da molecular weight cut-off filter. Glass vials were prepared with NMR buffer containing 2,2,3,3-tetradeuteropropionic acid (TMSP), and an aliquot of serum filtrate was mixed with this solution in the glass vial. ^{1}H NMR spectra were collected from the processed patient serum samples. The spectra were autophased and calibrated based on the TMSP peak. The spectra were baseline corrected and the water region and TMSP peaks were removed. The spectra were divided into bins set by chemical shift (0.005 ppm, 2.5 Hz) were normalised and g-log transformed.

The spectral data were analysed by principal component analysis (PCA), partial least squared discriminant analysis (PLS-DA) and orthogonal PLS-DA (OPLS-DA). Variable importance in projection (VIP) scores were calculated to determine which peaks contributed the most to OPLS-DA models. These peaks were then identified using the Human Metabolome Database and Chenomx NMR suite software. Chenomx was also used to calculate the concentrations of metabolites in samples in relation to the TMSP standard. It is these metabolites and concentration values which are used in the dataset used by this project.

4.2 Data set Information

The dataset utilised by this project was processed on 17th March 2015. The NMR peaks from the patients were matched to those labelled from a healthy control patient, of which there were 30 metabolites labelled.

The dataset has samples from 57 patients where:

- 37 were Adult Resus (Age 16 - 64, $\geq$15 % TBSA) with a total of 69 records

- 5 were Elderly Resus (Age $\geq$65, $\geq$15 % TBSA) with a total of 9 records
- 15 were Elderly FT (Age $\geq$65, $\geq$1 % TBSA full-thickness burns but $<$15 % TBSA) with a total of 26 records

Samples were removed from the dataset when the patient was diagnosed as septic on the day of or just prior to the day of sampling. Samples from patients who died early ($<$ 3 days from admission) were also removed. Samples from control patients were not used in the analysis as the goal is to distinguish between sterile and non-sterile inflammation and healthy controls will not be inflicted with serious inflammation of either kind.

A table of the metabolites identified through the NMR analysis is provided in table 4.1. As this is an untargeted approach, metabolites already known to be related to sepsis were not targeted for this data acquisition. Therefore the data set contains the metabolites which were able to be identified based on the NMR data without focusing on any specific metabolites. An untargeted approach such as this allows for the discovery of new biomarkers.

Table 4.1: Metabolites identified through NMR analysis

2-Hydroxybutyrate	3-Hydroxyisobutyrate	3-Hydroxyisovalerate
Acetate	Acetoacetate	Acetone
Alanine	Betaine	Carnitine
Choline	Creatine	Creatine phosphate
Creatinine	Dimethyl sulfone	Ethanol
Formate	Glucose	Glutamine
Glycine	Histamine	Isoleucine
Lactate	Methanol	Phenylalanine
Proline	Pyruvate	Threonine
Tyrosine	Valine	Xanthine

4.2.1 Data set stratification

Patients were stratified into 3 categories, adult resus (AR), elderly resus (ER) and elderly full thickness (EFT). The reason for this is because the response to burns in the elderly has

been shown to be different than in non-elderly patients, where elderly patients exhibited delayed hyper-metabolism and increases in hyperglycemic and hyperlipidemic responses [130]. The differences in metabolomic response between adult and elderly patients should be apparent in the data set and so separating the two cohorts should produce a benefit in classifier performance. The adult resus and elderly resus groups can be combined into a separate category representing severe injury. The elderly resus set by itself has too few samples to be used on its own and so will only be used in conjunction with the adult resus samples.

The patient groupings can also be subdivided based on the day from admission that the samples were collected on. Early samples were collected on either days 1 or 2, and later samples were collected on days 3 or 4. This stratification was performed as earlier diagnosis would be more beneficial and so it will be interesting to see if there are important differences between sample collection dates.

Table 4.2 shows the data stratifications created from the full data set. EFT D1D2 and EFT D3D4 have not been created, this is because the number of samples would be very low for those groups. The table also shows that some of the stratifications have quite a large class imbalance, such as D3D4 and EFT. The scaled vote technique for the k-NN classifiers was designed to reduce the effect of class imbalance and will be tested on these data sets. The naive Bayes classifiers will likely be impacted, however as stratified cross validation is performed with the class distributions maintained across data sets, the priors will reflect this imbalance. For logistic regression, class imbalance should not affect the weights associated with each feature but only the intercept term [131] and so important features should still be identified. Changes to the intercept can affect performance but it was decided to determine initial performance before deciding on more advanced techniques.

Table 4.2: Information about the different stratifications of the data set. For each stratification, the number of patient samples, the number of samples from patients diagnosed with sepsis, the number of samples from patients who did not become septic and the percentage of septic samples against total samples is included. The septic percentage shows the amount of class imbalance in each stratification.

Data set	Patients	Septic	Not Septic	Septic %	Not Septic %	Completeness
Full	89	39	50	43.82%	56.18%	72.02%
D1D2	54	27	27	50.00%	50.00%	71.73%
D3D4	35	12	23	34.29%	65.71%	72.48%
AR	58	28	30	48.28%	51.72%	69.48%
AR D1D2	36	19	17	52.78%	47.22%	69.72%
AR D3D4	22	9	13	40.91%	59.09%	69.09%
ARER	65	33	32	50.77%	49.23%	70.62%
ARER D1D2	41	23	18	56.10%	43.90%	71.30%
ARER D3D4	24	10	14	41.67%	58.33%	69.44%
EFT	24	6	18	25.00%	75.00%	75.83%

4.2.2 Cross-Validation

For this project, stratified cross-validation was used, with the typical fold number being 10, however this was reduced to 5 in situations where required by the data. In table 4.2 it can be seen that for the AR D3D4 and EFT stratifications, there are only 9 and 6 septic patient samples respectively. Using 10 fold cross validation would lead to folds without any septic patient samples which can have an affect on the confidence bounds of the ROC curve. Folds of 5 or 10 were chosen as these have been reported in the literature to be commonly used values [132].

4.3 Classifier performance results

The results given in this section are from analysing the data set stratifications mentioned in the previous section with the techniques discussed in chapter 3. The k-NN techniques are no imputation (NI), half-min imputation (HM) and mean imputation (M) and each technique is also used with and without scaled voting (sv). The naive Bayes and logistic

regression with multiple imputation techniques are also used with metrics reported. The values given in each table of results are the k value for the number of neighbours which when used gave the best AUC value, the 95% confidence interval value (calculated from the 5 or 10 folds of cross validation), the sensitivity and specificity for the best point on the ROC curve and the F_1 score. Results not reported for the NB or LR_MI techniques are due to not being able to create PDFs of the features in a training set due to not enough data.

Table 4.3: Full data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	13	0.78	0.08	0.64	0.80	0.68
k-NN_NIsv	13	0.78	0.08	0.62	0.82	0.67
k-NN_HM	10	0.70	0.07	0.41	0.84	0.51
k-NN_HMsv	10	0.69	0.07	0.41	0.84	0.51
k-NN_M	31	0.75	0.08	0.49	0.86	0.58
k-NN_Msv	31	0.75	0.08	0.64	0.74	0.65
naive Bayes		0.78	0.10	0.69	0.78	0.70
LR_MI		0.75	0.07	0.74	0.70	0.70

Table 4.4: D1D2 data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	9	0.82	0.14	0.85	0.70	0.79
k-NN_HM	20	0.77	0.12	0.81	0.63	0.75
k-NN_M	17	0.77	0.16	0.48	0.93	0.62
naive Bayes		0.72	0.17	0.63	0.93	0.74
LR_MI		0.71	0.15	0.70	0.70	0.70

Table 4.5: D3D4 data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	7	0.84	0.13	0.75	0.83	0.72
k-NN_NIsv	7	0.83	0.13	0.83	0.83	0.77
k-NN_HM	15	0.72	0.18	0.83	0.74	0.71
k-NN_HMsv	15	0.75	0.18	0.83	0.74	0.71
k-NN_M	8	0.82	0.13	0.83	0.74	0.71
k-NN_Msv	8	0.81	0.13	0.83	0.74	0.71
naive Bayes						
LR_MI		0.82	0.15	0.67	0.91	0.73

Table 4.6: AR data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	15	0.84	0.08	0.79	0.70	0.75
k-NN_NIsv	15	0.84	0.08	0.79	0.70	0.75
k-NN_HM	12	0.75	0.07	0.50	0.90	0.62
k-NN_HMsv	12	0.76	0.07	0.46	0.93	0.60
k-NN_M	20	0.79	0.09	0.61	0.80	0.67
k-NN_Msv	20	0.80	0.09	0.82	0.63	0.74
naive Bayes		0.69	0.12	0.61	0.80	0.67
LR_MI		0.67	0.11	0.68	0.70	0.68

Table 4.7: ARD1D2 data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	14	0.86	0.09	0.74	0.88	0.80
k-NN_NIsv	14	0.88	0.09	0.84	0.82	0.84
k-NN_HM	14	0.78	0.13	0.74	0.76	0.76
k-NN_HMsv	14	0.80	0.13	0.84	0.71	0.80
k-NN_M	5	0.82	0.08	0.58	0.88	0.69
k-NN_Msv	5	0.83	0.08	0.63	0.88	0.73
naive Bayes						
LR_MI		0.76	0.17	0.68	0.94	0.79

Table 4.8: ARD3D4 5-fold CV data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	2	0.72	0.18	0.33	1.00	0.50
k-NN_NIsv	2	0.76	0.19	0.56	0.85	0.63
k-NN_HM	3	0.67	0.16	0.22	1.00	0.36
k-NN_HMsv	3	0.69	0.16	0.33	1.00	0.50
k-NN_M	3	0.75	0.22	0.67	0.77	0.67
k-NN_Msv	3	0.74	0.22	0.67	0.77	0.67
naive Bayes						
LR_MI						

Table 4.9: ARER data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	8	0.80	0.12	0.82	0.69	0.77
k-NN_NIsv	5	0.81	0.10	0.67	0.84	0.73
k-NN_HM	4	0.78	0.10	0.85	0.56	0.75
k-NN_HMsv	4	0.80	0.10	0.82	0.63	0.75
k-NN_M	22	0.78	0.09	0.91	0.50	0.76
k-NN_Msv	22	0.79	0.09	0.76	0.66	0.72
naive Bayes		0.71	0.09	0.64	0.78	0.69
LR_MI		0.72	0.09	0.79	0.72	0.76

Table 4.10: ARERD1D2 data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	11	0.86	0.09	0.78	0.89	0.84
k-NN_NIsv	8	0.85	0.12	0.91	0.67	0.84
k-NN_HM	19	0.75	0.09	0.96	0.39	0.79
k-NN_HMsv	3	0.77	0.12	0.87	0.50	0.77
k-NN_M	11	0.88	0.13	0.83	0.72	0.81
k-NN_Msv	11	0.89	0.13	0.61	1.00	0.76
naive Bayes						
LR_MI		0.77	0.12	0.70	0.89	0.78

Table 4.11: ARERD3D4 data set results.

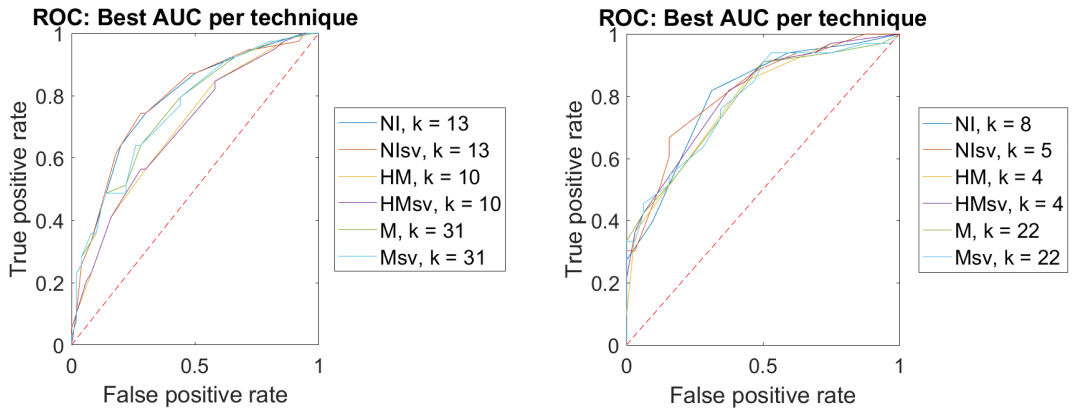
Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	2	0.79	0.15	0.40	1.00	0.57
k-NN_NIsv	2	0.82	0.15	0.80	0.79	0.76
k-NN_HM	9	0.66	0.23	0.80	0.71	0.73
k-NN_HMsv	3	0.67	0.23	0.90	0.50	0.69
k-NN_M	6	0.72	0.23	0.60	0.79	0.63
k-NN_Msv	6	0.75	0.23	0.90	0.57	0.72
naive Bayes						
LR_MI						

Table 4.12: EFT 5-fold CV data set results.

Technique	k	AUC	95 % CI (±)	Sensitivity (TPR)	Specificity (TNR)	F1 Score
k-NN_NI	10	0.75	0.19	0.83	0.83	0.71
k-NN_NIsv	10	0.75	0.19	0.83	0.83	0.71
k-NN_HM	10	0.75	0.14	0.83	0.78	0.67
k-NN_HMsv	10	0.75	0.14	0.33	1.00	0.50
k-NN_M	10	0.81	0.11	0.83	0.78	0.67
k-NN_Msv	10	0.82	0.10	0.83	0.78	0.67
naive Bayes						
LR_MI		0.75	0.31	0.33	1.00	0.50

4.3.1 ROC curves

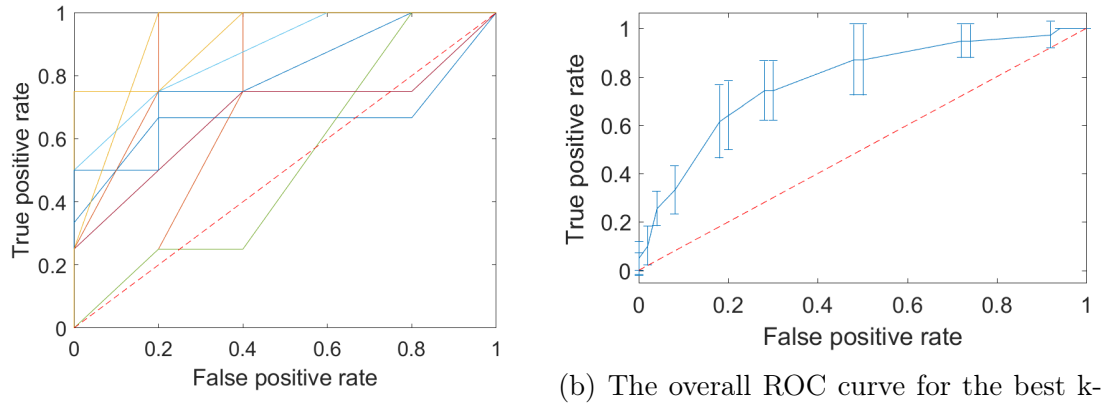
Below are samples of the ROC curves produced by the created classifiers. The sample was chosen in order to highlight things such as comparing the k-NN techniques, showing per fold variance from cross validation, the confidence intervals (95 %) associated with the curves, and the NB and LR techniques.



(a) The ROC curves with the best AUC values per k-NN technique when using the Full data set.

(b) The ROC curves with the best AUC values per k-NN technique when using the ARER data set.

Figure 4.1: Showing the best ROC curves for each k-NN technique for two data sets, a the Full data set, and b the ARER data set. With the Full data set it can be see that the NI method performs the best, followed by M and then HM imputation techniques. The ARER data set shows that this is not consistently the case.



(a) The ROC curves per fold for k-NN NIsv classifier on the Full data set.

(b) The overall ROC curve for the best k-NN NIsv classifier on the Full data set showing confidence bounds set at 95 %.

Figure 4.2: Showing the variance across folds when using NI technique with k = 13 on the Full data set.

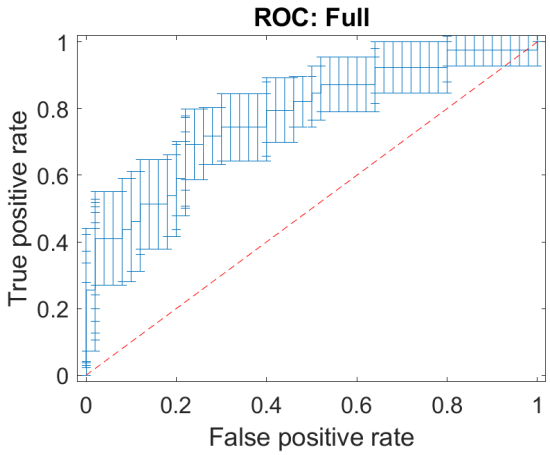


Figure 4.3: The ROC curve for the naive Bayes technique on the Full data set showing confidence bounds set at 95 %.

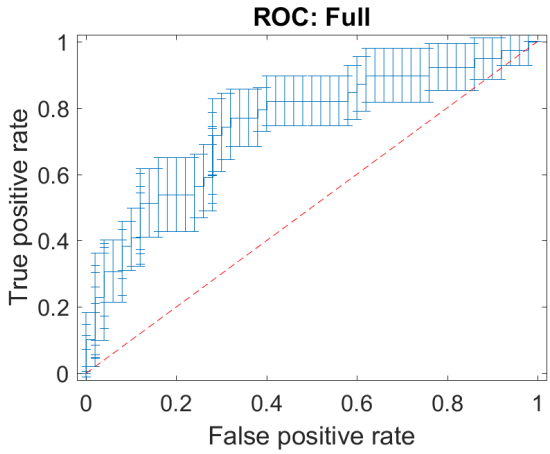


Figure 4.4: The ROC curve for the logistic regression technique on the Full data set showing confidence bounds set at 95 %.

4.4 Classifier performance discussion

The stratifications of this data set had different levels of class imbalance. The scaled voting technique was introduced to the k-NN algorithms in order to compensate for class imbalance. The elderly full thickness data set contains the largest class imbalance, 75 % non-septic patients, however it is also one of the smallest data sets and so a single misclassification in a fold can greatly impact the metrics recorded. Looking at the results

in table 4.12 it is seen that the scaled vote did not affect the classifier performance for the no imputation and mean imputation techniques. For half-min imputation there is a difference, however not the difference which would be expected, as the true negative rate increases and the true positive rate decreases. The scaled vote should have allowed for more positive predictions to have been made but the TNR being 1 shows there were no false positives and the lower TPR score shows there were more false negatives.

Looking at the other data sets it can be seen that the affect of the scaled vote varies. In the Full data set (table 4.3) there is a pronounced improvement in TPR when scaled voting is used with mean imputation, going from 0.49 to 0.64. Similar improvements when using the scaled vote technique can be seen in other data sets, such as in the D3D4 data set for the NI methods the TPR increases from 0.75 to 0.83. The ARERD3D4 data set (table 4.11) is imbalanced in favour of the negative non-septic class (58 %). For this data set the expected effect of vote scaling would be to increase the influence of the positive septic samples and thus lead to more samples being given a positive classification. This is a good data set to look at, as the optimal number of neighbours for the NI and M imputation methods was also the same when using vote scaling for each respective method, thus the impact of the scaled vote can be isolated. There are dramatic increases in the true positive rates when scaled voting is applied to each method, for NI the value doubles from 0.4 to 0.8, and for M the value increases by 1.5 times from 0.6 to 0.9. These are coupled with decreases in the true negative rate which for the NI technique went from 1 to 0.79 and for the M technique from 0.79 to 0.57. For the NI non-scaled vote method, a TNR of 1 meant that there were no false positives, however the TPR value of 0.4 shows there were a lot of false negatives. Here the classifier is favouring a negative classification which is likely due to the class imbalance. Vote scaling corrects this improving the AUC and F_1 score values.

The affect of vote scaling is not consistent across the data sets. For the ARD1D2 data set (table 4.7) the class imbalance is in favour of the positive class (53 %) and so

the vote scaling should correct for this and allow more negative class predictions to be made. The actual effect of the vote scaling led to increases in the TPR values for each technique and decreases in the TNR for NI and HM techniques. Each classifier pair (non-sv and sv) used the same k value and so the difference should be from the scaling factor. The differences between the AUC and F_1 scores between scaled vote and non-scaled vote methods vary from no change at all up to a 5 % difference, however the difference can be a reduction in the scores. Because of this, based on this data alone it can not be said that the vote scaling method performs the function is was intended to do. When the effect works for the better, the changes in TPR and TNR can be quite substantial as mentioned in the previous paragraph, however there is no pattern as to when it would provide this benefit. Further analysis could be performed on artificial data sets designed for testing class imbalance methods which may uncover the scenarios where vote scaling works out to be beneficial.

Looking at the different strategies to deal with missing data with the k-NN classifiers uncovers that not imputing values typically gives the best performance, giving the best AUC and F_1 scores in the majority of the data sets. Figure 4.1 illustrates that sometimes there is a clear advantage by not imputing values and other times the classifiers perform similarly. Typically the half-min imputation method performs worse than the mean imputation method which correlates with published data [84].

The naive Bayes classifier works by only looking at available data, which is similar to the no imputation technique created for the k-NN algorithm. The NI technique only compared matching features in samples which would in theory, reduce the amount of usable data when compared to the naive Bayes technique. This is because the naive Bayes technique constructs PDFs per feature per class, and therefore every feature in a test sample will have a probability value assigned to it for each class, whereas the NI technique required features to match between samples. Because of this, it would have been expected that the naive Bayes algorithm would at least perform similarly, whereas

it can been seen that it performs worse.

One reason for the naive Bayes algorithm performing worse than k-NN is that it depends strongly on the PDFs it uses. Without enough records per feature, the PDF may not accurately resemble the underlying distribution of the feature and that would cause the probability values to be poor. This limitation of small samples sizes was realised from the fact that it was impossible to produce PDFs for some data splits when using cross validation and so only the Full, D1D2, AR and ARER stratifications allowed the algorithm to complete. The number of folds was reduced from 10 to 5 for some of the data sets which failed to work with the NB implementation, however there were still some folds which did not contain enough values per feature per class to create the PDFs and so getting the algorithm to work on these data sets was abandoned.

Logistic regression with multiple imputation was used as the final classifier type. Multiple imputation requires modelling the feature set and this was done through the creation of PDFs as with the naive Bayes approach, as such similar difficulties arose when trying to do this for some data sets and so results are not available for the ARD3D4 and ARERD3D4 data sets. The LR_MI classifiers performed similarly to the naive Bayes classifiers, which is typically worse than the k-NN classifiers. As with the NB technique, the performance is likely hindered by the feature modelling performed where the PDFs may not be representative of the actual data.

Whilst some of the techniques presented managed to generate classifiers with AUC values going up to 0.89 for some data sets, it is important to take into account the variation per cross validation fold. Figure 4.2 shows the per fold ROC curves and how variable they are along with the ROC curve with confidence bounds highlighting the range the classifier actually performs in for the Full data set using the k-NN_NIsv technique. Figures 4.3 and 4.4 show the confidence bounded ROC curves for the NB and LR_MI techniques respectively for the Full data set. Ideally when performing cross validation there would be minimal variance between the results from each fold which would be indicative of the

folds being representative of the complete data set. The variation seen implies the folds differ a lot, which again reinforces the difficulties in generating PDFs for each feature for the NB and LR_MI techniques.

4.5 Stratification discussion

The complete data set was stratified by the conditions of patient type (adult and elderly) and also by sample time early, days 1 and 2, and late, days 3 and 4.

Looking at tables 4.3 to 4.5 it can be seen that in this instance, splitting the data set actually had a positive impact, with classifiers performing better on each split respectively when compared to the full data set, excluding the NB and LR classifiers for D1D2 which performed only slightly worse than on the full data. If there were no difference between the early and late samples it would be expected that the classifiers would perform better on the full data set as there is more data available. One thing to consider however is that the confidence intervals are larger for D1D2 and D3D4 compared with the full data set. This shows there was more variance across the cross validation folds. One explanation for this is simply because there are fewer samples in the split data sets, which means each fold has fewer samples and misclassifications will have a greater impact on the calculated metrics.

Table 4.6 gives the results of the stratification which removes all elderly patient data. When compared against the full data set in table 4.3 the k-NN techniques produce better classifiers, however the NB and LR do not. The confidence intervals remain similar between data sets with most being unchanged and the largest change of 0.04 for the LR method. This indicates that the consistency in performance across the folds is the same when using either of these data sets. Therefore the reduction in the number of samples has not had the same effect as noticed earlier.

Splitting the AR data set up by sample time shows improved classifier performance with ARD1D2 but decreased performance for with the later samples ARD3D4 as can be seen in tables 4.7 and 4.8. The ARD3D4 data set does not contain a lot of samples and so 5-fold cross validation was performed. Despite this the variation between the folds is very high with the confidence intervals ranging from $\pm$ 0.16 to $\pm$ 0.22. The ARD3D4 data set on its own should therefore not be used to build classifiers. The fact that classifier performance improved for the ARD1D2 data set over the AR data set does indicate that there are useful differences between the sample dates and so for further studies, if more data can be collected this stratification would still be recommended.

Results from excluding only the elderly full thickness patients from the data set are given in tables 4.9 to 4.11. The results repeat what has been seen previously, a minor increase in classifier performance when comparing ARER with the Full data set for all k-NN techniques, along with another increases when looking at the ARERD1D2 split, along with worse performance when looking at ARERD3D4. Again, having fewer samples on the D3D4 days appears to be affecting performance, but the increased classification scores when looking at ARERD1D2 against ARER does give further credence to there being differences between metabolite concentrations on the different days since admission.

The elderly full thickness EFT data set contains the strongest class imbalance and is among the data sets with the fewest samples. The classifier performance on this data set was worse than the Full data set, as would be expected. The class imbalance in this data set is heavily towards the non-septic patients (75%), however the classifiers do still manage to achieve good true positive rates, indicating the classifiers are not just choosing the dominant class. Despite this, the confidence intervals are large and so per fold performance varies greatly.

Classifiers using the early D1D2 splits for all patient type stratifications gave better AUC and F_1 scores than the respective data sets including the D3D4 samples. Also the data sets without the EFT patients obtained better AUC and F_1 values for all k-NN

techniques. Therefore, it can be seen that stratifying the data set by sample date from admission and removing EFT patients provides a benefit when performing classification on this data set, however attention needs to be paid to the number of samples remaining.

4.6 Feature selection results

The features selected via different techniques are displayed below. The techniques used are logistic regression with multiple imputation, half-min imputation and mean imputation, which provide results of an embedded feature selection process. There are also results from using the t-test and MRMR algorithms, which provide results of filter based feature selection.

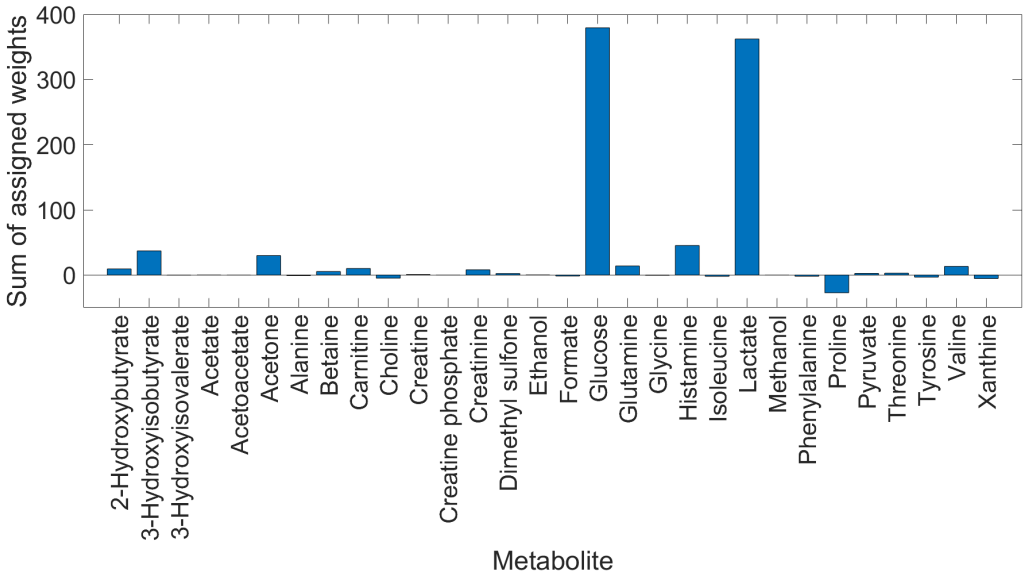


Figure 4.5: The sum of the weights associated to each metabolite across all of the models produced for the Full data set when using logistic regression with multiple imputation.

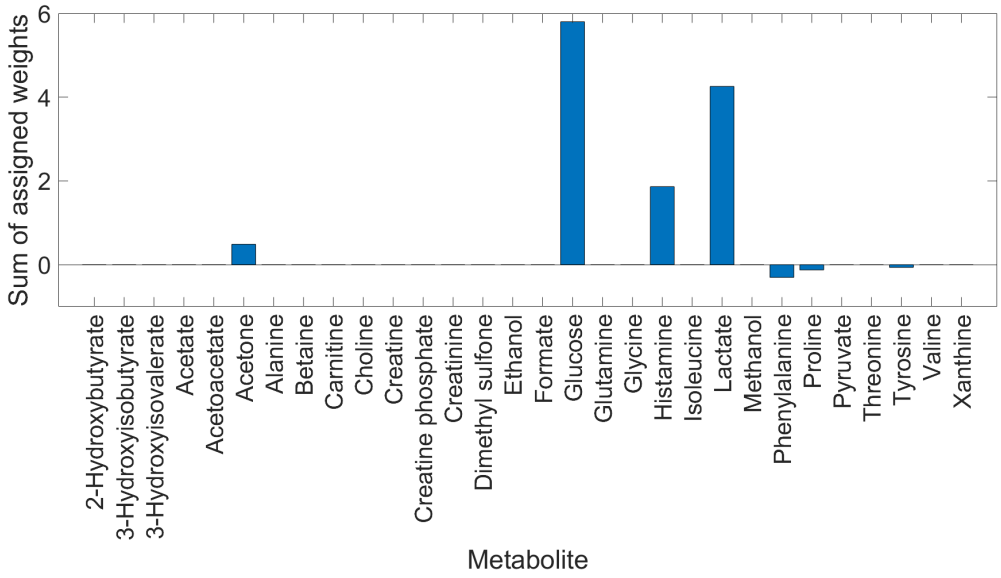


Figure 4.6: The sum of the weights associated to each metabolite across all of the models produced for the Full data set when using logistic regression with half-min imputation.

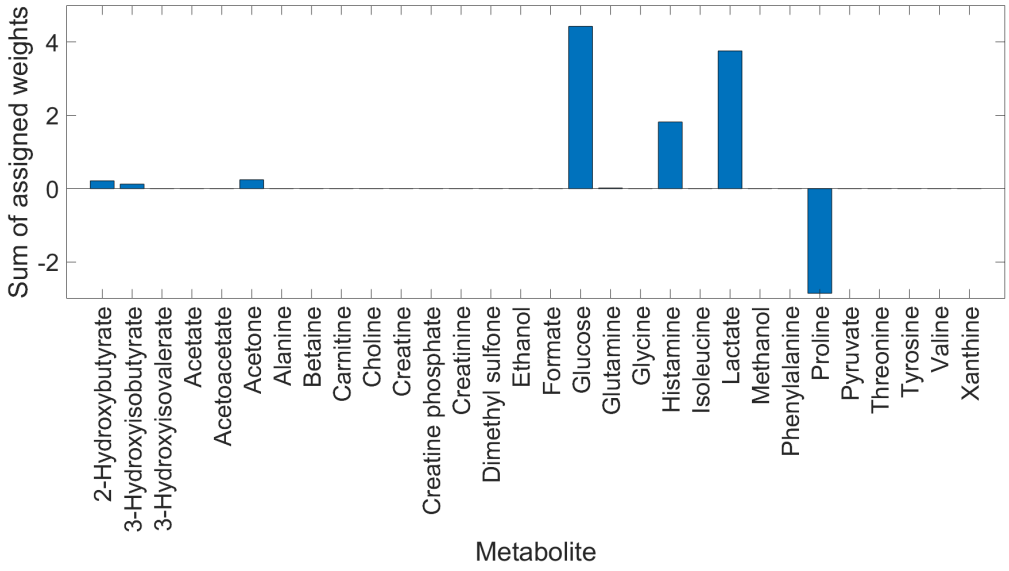


Figure 4.7: The sum of the weights associated to each metabolite across all of the models produced for the Full data set when using logistic regression with mean imputation.

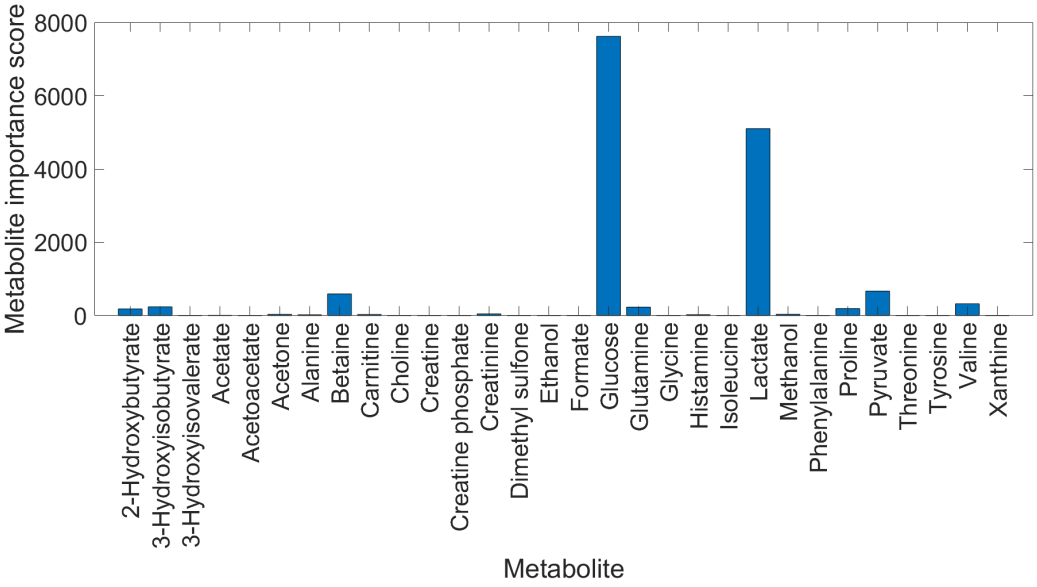


Figure 4.8: The metabolites identified as being significant when using t-tests for feature selection on the Full data set. The bar heights are the reciprocal of the p-value from performing the test. This is because a lower p-value signifies greater significance.

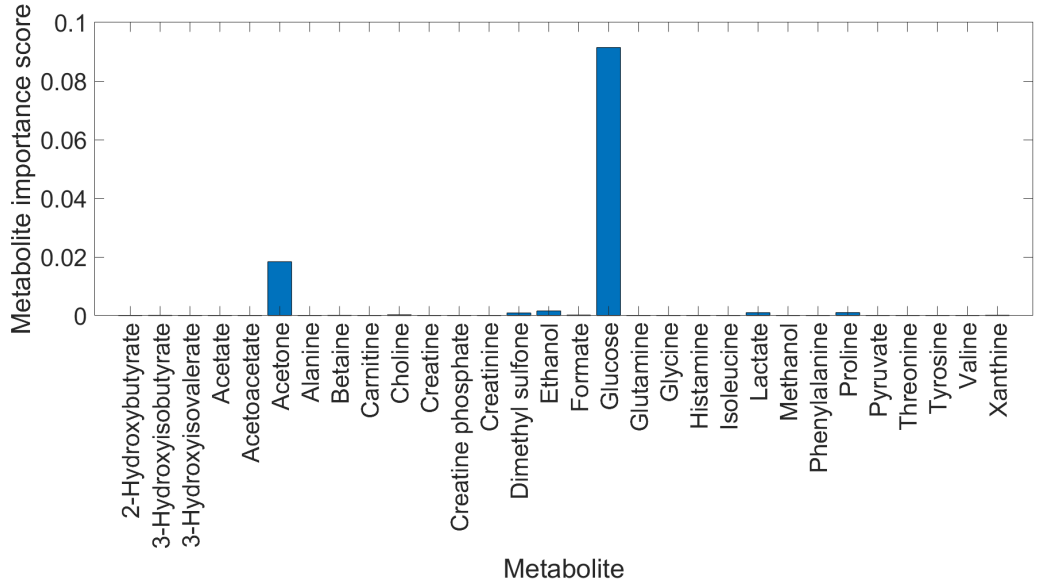


Figure 4.9: The metabolites identified as being significant when using the MRMR algorithm for feature selection on the Full data set.

4.7 Feature selection discussion

A benefit of using logistic regression is that it can perform feature selection automatically when paired with a penalty function. The lasso penalty function was used with this work. Because the overall aim of this project is to identify a subset of metabolites with high predictive power for sepsis in order to create a point of care device to detect them, it was decided to run the logistic regression algorithm and have it set to produce models with up to 5 metabolites, however for the planned work for the next part of this project, the fewer metabolites the better.

By looking at the weights assigned to each feature in the models created, it is possible to determine which metabolites are deemed the most important. This can be done by summing up the total weight assignments. The summed weights for the Full data set are shown in fig. 4.5 as an example. The two metabolites glucose and lactate stand out as being the most important features for the classifier.

To determine if multiple imputation had a profound effect on the identified features, LR was performed with a half-min and mean imputed Full data set with the results shown in figs. 4.6 and 4.7 respectively. The results from LR with multiple imputation show that more metabolites were used across the models, however the size of the columns indicates that they must metabolites outside of glucose and lactate were either used sparsely or just given very low weights. The half-min imputed feature graph mimics the multiple imputation feature graph, however the mean imputed graph identifies proline as a negative biomarker, that is a greater proline concentration is linked to a reduced likelihood of sepsis.

Using the t-test for each feature shows that glucose and lactate are again the most important features in the data set as shown in fig. 4.8.

The MRMR algorithm tries to keep the most relevant features but also remove re-

dundancy, and so if any metabolites are closely connected through metabolic pathways, this algorithm should remove the redundant metabolites. Interestingly this technique highlights glucose as the main feature of interest, the peak for acetone is almost 5 times smaller.

The result of the MRMR algorithm not reporting lactate but instead only glucose implies that the two are connected and so there would be redundancy in having both. The metabolic pathways of the two molecules are connected where glucose can be turned into pyruvate and then lactate [133]. This shows that the method works as expected.

Looking further into glucose and lactate, they are already currently recommended metabolites to monitor as declared by the Surviving Sepsis Campaign [17]. As such, there are already point of care devices for both metabolites [66, 92].

There is little literature available on proline in combination with sepsis however one report indicates that a higher proline concentration is indicative of sepsis, which would contradict the result from the logistic regression which gave it negative odds [134]. The paper actually shows that proline is also linked with glucose and lactate through energy metabolism. As such it would seem that it should be a positive biomarker and so there may be something wrong with the acquisition of the data set with respect to peak identification and fitting to standards from the original NMR data. Therefore it would be something to look for in a targeted metabolomics study, but not something to design a point of care device for given the current data.

4.8 Conclusions

The k-NN algorithm was used and adapted to handle missing values without imputation through the use of a custom distance function which only looks at matching pairs of features and scales the result accordingly. This was combined with mean and half-min im-

putation, and all three techniques were used with and without vote scaling based on class imbalances. The algorithm performed best without any imputation and the vote scaling did offer some inconsistent performance benefits. Classifier performance was increased by removing the EFT patient set from the Full data set. Improvements when comparing early sample collection date with the later sample collection dates may have been caused by changes to the class imbalance, or the reduction of samples comprising a set, as there were fewer D3D4 samples. The increased class imbalance when splitting the data set this way means accuracy could be increased by predicting the majority class more often. The large confidence bounds for each classifier show that the per fold accuracy varies greatly and the classifier is therefore not stable.

The naive Bayes algorithm was employed due to the fact that it can operate with an incomplete data set. Whilst this is true, it does require enough data to be able to create probability density functions to accurately model the data which does exist. This was not the case for most of the data stratifications for the given data set. The classifiers it did produce offered worse average performance than the k-NN classifiers and also suffered from large per fold variation.

Classifiers produced from logistic regression with multiple imputation did not perform as well as the k-NN techniques. The ROC curves still had large confidence bounds and so it appears the multiple imputation did not perform as hoped. This is likely due to the same reason the naive Bayes technique failed to execute on some data sets, namely a lack of feature information to accurately model each feature.

The logistic regression models had the advantage of having feature selection embedded within them and this was one of the reasons this technique was chosen. Despite the performance not being the best the metabolites of interest were investigated. These features turned out to be glucose and lactate, biomarkers which are already recommended to be monitored for sepsis and already have available devices do perform such monitoring. Given the importance given in the feature selection graphs towards lactate and glucose

it strongly suggests the other metabolites would not strongly classify the data without them.

During this analytical process, literature searches were being performed to better understand the metabolomics research in the field and to see if any novel biomarkers were reported, especially with regards to sepsis and traumatic injury. The next chapter will explain how a biomarker was chosen from literature to progress this project.

CHAPTER 5

ASSAY DEVELOPMENT

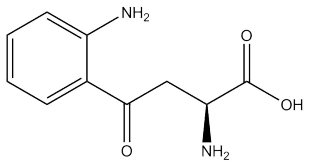
5.1 Introduction

The previous work utilising machine learning algorithms on a metabolomics data set managed to identify known clinical markers for sepsis diagnosis, however it did not provide any novel diagnostic biomarkers. Throughout the machine learning phase, literature searches were performed to improve understanding of the problem, research already conducted, and to try to identify novel biomarkers specifically for trauma patients. Having read about the tryptophan metabolic pathway and how its conversion to kynurenine showed significance in sepsis diagnosis, it was decided that the project should continue with the development of an assay for this small molecule. The aim to be to detect and quantify it within biological samples.

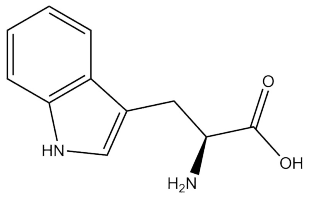
5.2 Kynurenine

Kynurenine was the metabolite chosen to continue this project. Kynurenine is a metabolite produced from the tryptophan metabolic pathway. Tryptophan is an essential amino

acid, and in humans, its metabolic pathway has two paths. One pathway leads to the formation of serotonin and melatonin, with the former being a neurotransmitter and the latter being a hormone, and the other pathway leads to the formation of kynurenine and follows through to niacin, a form of vitamin B3 [135, 136]. The structures of kynurenine and tryptophan are shown in fig. 5.1. Tryptophan is broken down along the kynurenine pathway by the indoleamine 2,3-dioxygenase (IDO) enzyme . The breakdown products of tryptophan along the kynurenine pathway have been linked to several biological functions and disease states and most importantly, this pathway has been linked with inflammation and the immune system [137].



(a) Kynurenine



(b) L-Tryptophan

Figure 5.1: Molecular structures of Kynurenine and L-Tryptophan

5.2.1 Connection with sepsis

A team from the University Hospital Düsseldorf and Johann Wolfgang Goethe University examined the possibility of kynurenine as a biomarker of sepsis in trauma patients. For their study, they analysed data from 60 trauma patients of various injuries, including peripheral fractures and abdominal, thoracic and cerebral injuries [138]. The concentrations of kynurenine and tryptophan were measured from blood serum samples and the results can be seen in fig. 5.2. The figure shows an increase in serum kynurenine concentration, a decrease in serum tryptophan concentration and so an overall increase in the

kynurenine:tryptophan ratio. As can be seen, the kynurenine concentrations measured in patients who did not develop sepsis lies within the range measured from the healthy control subjects. The kynurenine concentration for the patients who did develop sepsis is greater than this concentration and so kynurenine can act as a positive biomarker for sepsis. This is not the only study to find a link between tryptophan metabolism and sepsis and there are others which are linked to trauma and mortality. Multiple studies identify the role of the IDO enzyme in immunodeficiency and death in trauma patients [135, 137, 139]. These studies all had quite low sample sizes, between 15 and 30 patients, however the results consistently show increased kynurenine concentrations and kynurenine:tryptophan ratios are linked with sepsis. A larger study monitoring 100 patients found not just kynurenine but other metabolites downstream of it could discriminate between non-septic and pre-septic patients [140]. Other reports have identified the role of kynurenine in hypotension during systemic inflammatory response syndrome and septic shock [141–144]. It can clearly be seen that Kynurenine has a strong connection with sepsis. The referenced studies typically used MS or HPLC to determine the metabolite concentrations in patient samples. MS and HPLC are not bedside point of care devices, and so if a simple assay could be designed to measure kynurenine concentrations it would have great clinical significance.

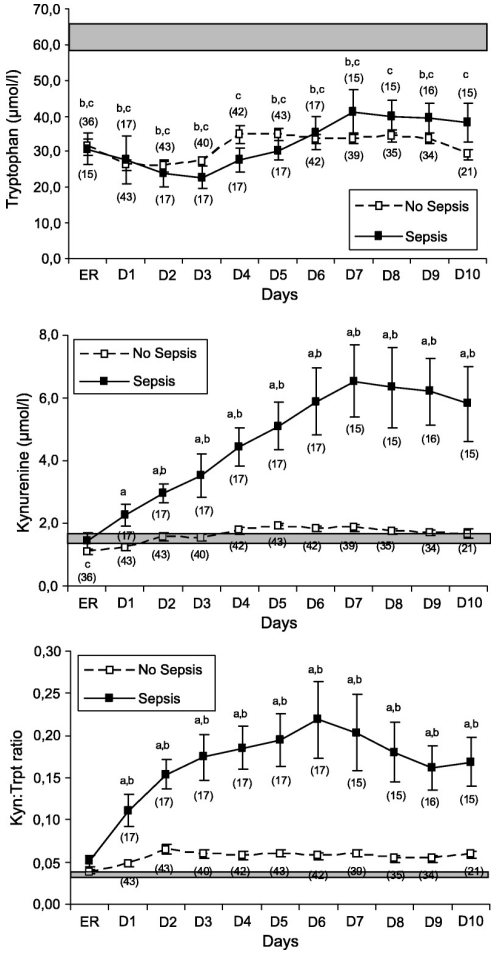


Figure 5.2: Graphs showing the measured concentrations of tryptophan, kynurenine and the kynurenine:tryptophan ratio across multiple days from patients of traumatic injury. The grey bar shows the concentration found in healthy control patients. This figure is reproduced with permission from [138].

5.2.2 Detecting Kynurenine

Absorption spectroscopy has the ability to detect analyte concentrations into the µM range, whilst fluorescence techniques can improve sensitivity by going down to the nM range in some cases. Fluorescence spectroscopy has also been shown to be more sensitive in urine and plasma than UV spectroscopy [145]. The concentrations of kynurenine which diagnose sepsis in serum are low (between 2 - 6 µM), and so photospectroscopic techniques are suitable for detection at these levels. Kynurenine is a weakly fluorescent molecule, with an excitation wavelength of 365 nm and an emission wavelength of 480 nm [146,147].

When in a biological sample, the signal from kynurenine is overlapped with that from other compounds [148]. These two issues make it very difficult to specifically detect and accurately quantify kynurenine in a biological sample.

In 2012 at the University of Missouri, Klockow and Glass designed a chemosensor for the detection of kynurenine [149]. They showed the chemosensor to bind selectively to kynurenine, and that when bound, the complex has an excitation wavelength of 555 nm and emission wavelength of 632 nm. This change in wavelength from kynurenine's natural emission wavelength can help in isolating the signal specifically for kynurenine.

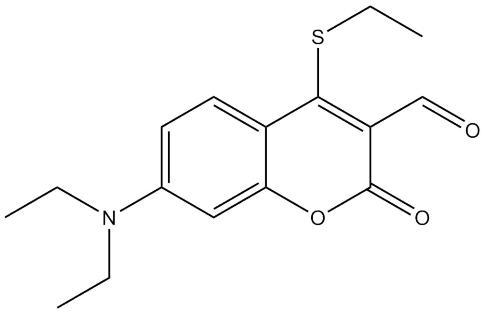


Figure 5.3: Chemical structure of the fluorescent sensor.

In this work it was decided that for the assay which was to be created, kynurenine would be the metabolite of interest based on the reasons given previously, and the chemosensor developed by Klockow and Glass would be the basis for it. There are multiple reasons as to why this chemosensor was chosen.

One reason for choosing this chemosensor is that it exhibits selectivity towards kynurenine compared to biomolecules with the same functional groups and similar size [149]. The tested biomolecules include glycine and the nucleosides adenosine and cytidine. Glycine is a simple amino acid which has amine and hydroxyl groups in an arrangement very similar to those in kynurenine. It does bind to the chemosensor but does not induce the same bathochromic shift which indicates that the rest of the structure of kynurenine has a role in the effect of this shift. Adenosine and cytidine are components of DNA and RNA and are therefore likely to be circulating in the blood especially around the site of

a traumatic injury. Both have amine and hydroxyl groups but neither produce the affect kynurenine does upon binding to the chemosensor. All three of these molecules exhibit lower association constants than kynurenine to the chemosensor which again highlights the selectivity towards kynurenine and how the chemosensor should be usable with the testing of biological samples.

Another reason for choosing this chemosensor is due to its molecular structure. The chemosensor contains a sulphide group, which when looking at the synthetic route in fig. 7.1, is added to the base coumarin structure through reaction with a thiol. In the future this thiol could be replaced with other molecules which could enhance the binding of the chemosensor to kynurenine and also change the wavelengths at which the complex fluoresces at [150]. The sulphide group could also be used to bind the chemosensor to a surface, where it could be used as part of a molecular imprint [151]. Usage as part of a surface could improve selectivity even more by forming a template around the chemosensor which is made for kynurenine thus inhibiting competitive interactions from other molecules [152].

The fluorescent chemosensor, which will be referred to as the sensor for this assay, was synthesised as described in chapter 7. The following sections in this chapter focus on the characterisation of the sensor and the initial tests conducted to create the assay.

5.3 UV-Vis spectroscopy

In order to test the synthesised sensor worked as expected, UV-Vis absorbance spectra were recorded with solutions of the sensor, kynurenine, and the sensor with kynurenine.

First there is the UV-Vis spectra of just the sensor at 10 µM concentration in fig. 5.4. In this graph there can be seen a peak absorbance at 469 nm, which is the expected position of the peak. It can also be seen that there is only a single peak in the wavelength

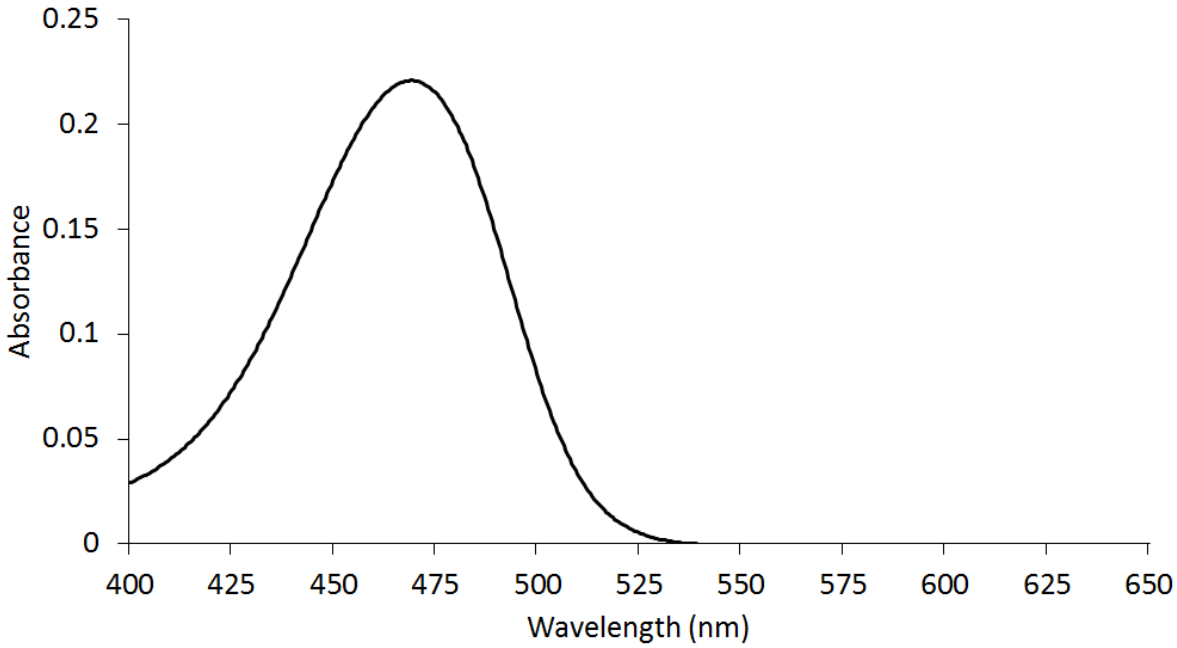


Figure 5.4: The absorbance spectra of a 10 µM solution of sensor.

range displayed of 400 - 650 nm and after the peak the absorbance value returns to zero. This indicates that there is nothing else in the solution which absorbs light in that wavelength range.

Next we have the UV-Vis spectra of two concentrations of kynurenine in fig. 5.5. The 6 µM solution is within the biological range we are aiming for and can barely be seen whereas the 500 µM is much higher in concentration and shows the absorbance spectra clearly with the expected peak for kynurenine at 360 nm.

Moving on to fig. 5.6, the sensor kynurenine mixture, the affect of increasing the kynurenine concentration can be seen in the absorbance spectra. Two changes are seen which indicate the sensor:kynurenine complex being formed. First there is a small decrease in the absorbance at 469 nm which indicates there is less free sensor in solution. Secondly there is the appearance of two new absorbance peaks at 526 nm and 555 nm, which are indicative of complexation with kynurenine.

The previous experiment provided confirmation the sensor was working as expected

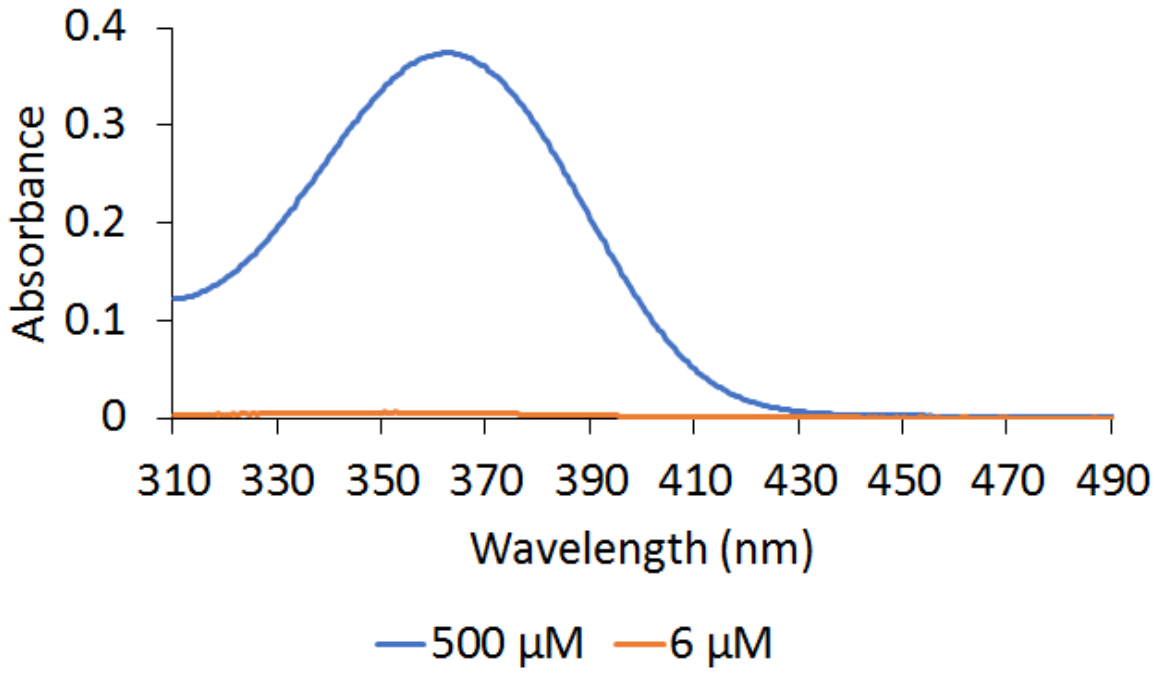


Figure 5.5: The absorbance spectra of two solutions of kynurenine. The orange line shows a 6 µM solution and the blue line shows a 500 µM solution.

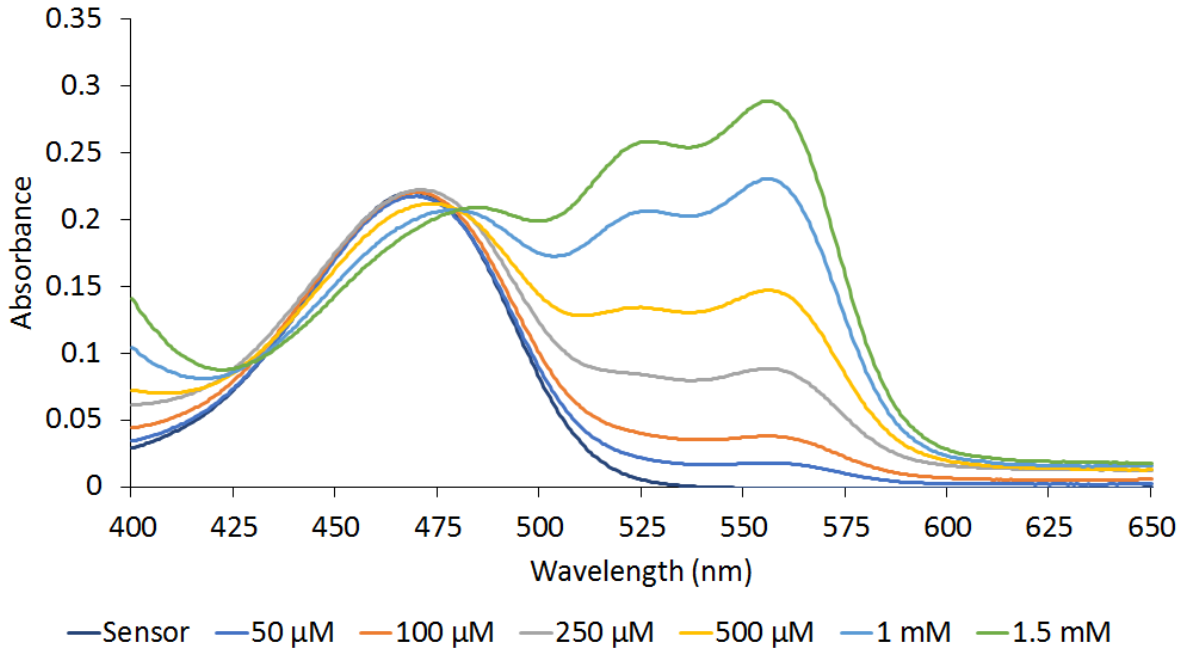


Figure 5.6: UV-Vis spectra with sensor concentration 10 µM and varying kynurenine concentrations.

from literature, however the concentrations of kynurenine used to produce the changes in the absorbance spectra were much higher than those expected for biological samples. For serum, the expected levels would be between 2 - 6 µM, whereas the lowest kynurenine concentration used in the experiments was 50 µM. The previous experiment used a fixed sensor concentration and varied the concentration of kynurenine. As there is an expected concentration range for kynurenine, another experiment was performed where the kynurenine concentration was fixed at 6 µM and the sensor concentration was changed. The goal was to determine if it was possible to produce a large change in the absorbance spectra with a lower, more biologically relevant concentration of kynurenine. The sensor concentrations used were (0.1, 0.5, 1, 2 and 6 µM) and so went from having kynurenine in excess to a 1:1 ratio of kynurenine:sensor. The results in Figure 5.7 show that this was not possible. As the sensor concentration increased, so did the peak of the unbound sensor, which was to be expected, but the two new absorbance peaks could not be seen. The concentrations used were possibly too low, as the maximum absorbance value was around 0.13, which is largest value given for unbound sensor. Without any sensor present the absorbance at 365 nm, the wavelength kynurenine absorbs at, is recorded as less than 0.01 which is a very low value, 10 times less than the unbound sensor of similar concentration and so absorbance spectroscopy for these concentrations of kynurenine will not make a viable technique.

Absorbance spectroscopy has been shown to be a less sensitive technique than fluorescence spectroscopy, with the latter able to go down to nanomolar detection, and so to counteract the low absorbance values of the new peaks with kynurenine present, it was decided to move to a fluorescence based system.

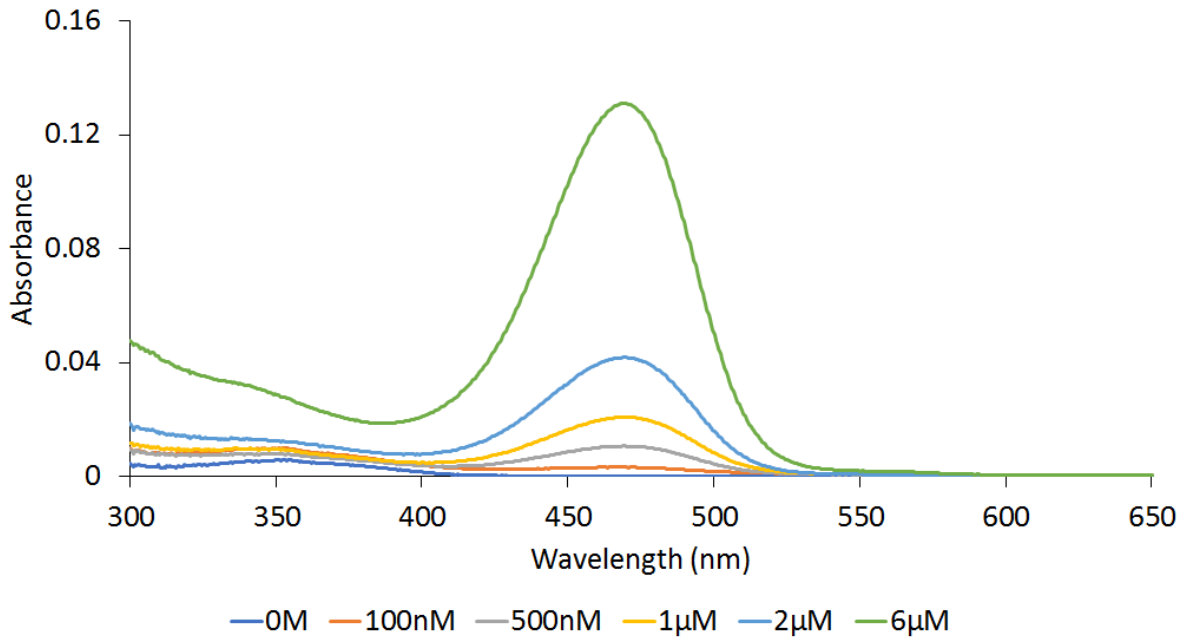
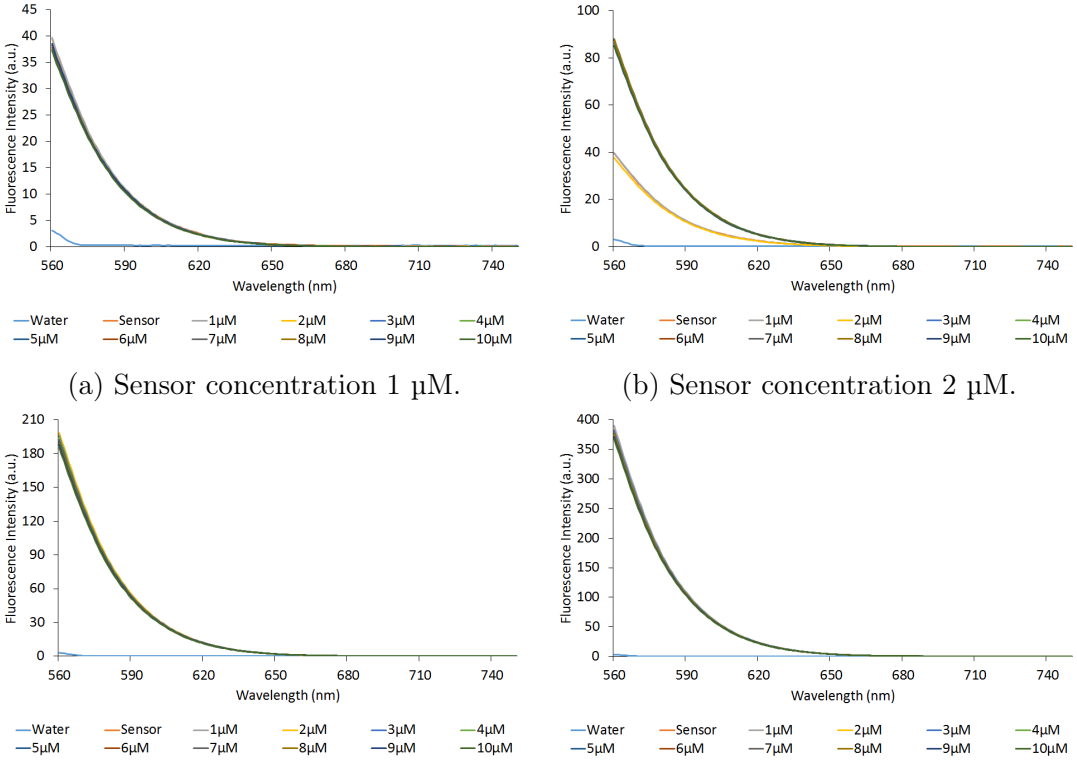


Figure 5.7: The absorbance spectra of varying sensor concentration from 100 nM to 6 µM, keeping the kynurenine concentration fixed at 6 µM.

5.4 Fluorescence Tests

Initial fluorescent tests consisted of titrations which kept the sensor concentration constant and increased the concentration of kynurenine. Excitation wavelengths were selected based on the absorbance spectra from the UV-Vis experiments and emission spectra were recorded by keeping a fixed excitation wavelength and measuring over a range of emission wavelengths. The excitation wavelengths used were 470 nm which is the peak for the unbound sensor and 555 nm which is the largest wavelength out of the two new peaks created by the complex. Figure 5.8 shows a subplot of graphs, where each new graph is for a different sensor concentration and within each graph there is a range of kynurenine concentrations being measured. The fluorescence intensity does not change a lot when exciting at 470 nm. This correlates with the UV-Vis spectra where there was no significant change in the absorbance at 470 nm. There are spectra within fig. 5.8b which show a decrease in fluorescence. This is believed to be caused by not thoroughly mixing the titrated solutions, as it happened only for the first two additions and then the spectra



(a) Sensor concentration 1 µM.

(b) Sensor concentration 2 µM.

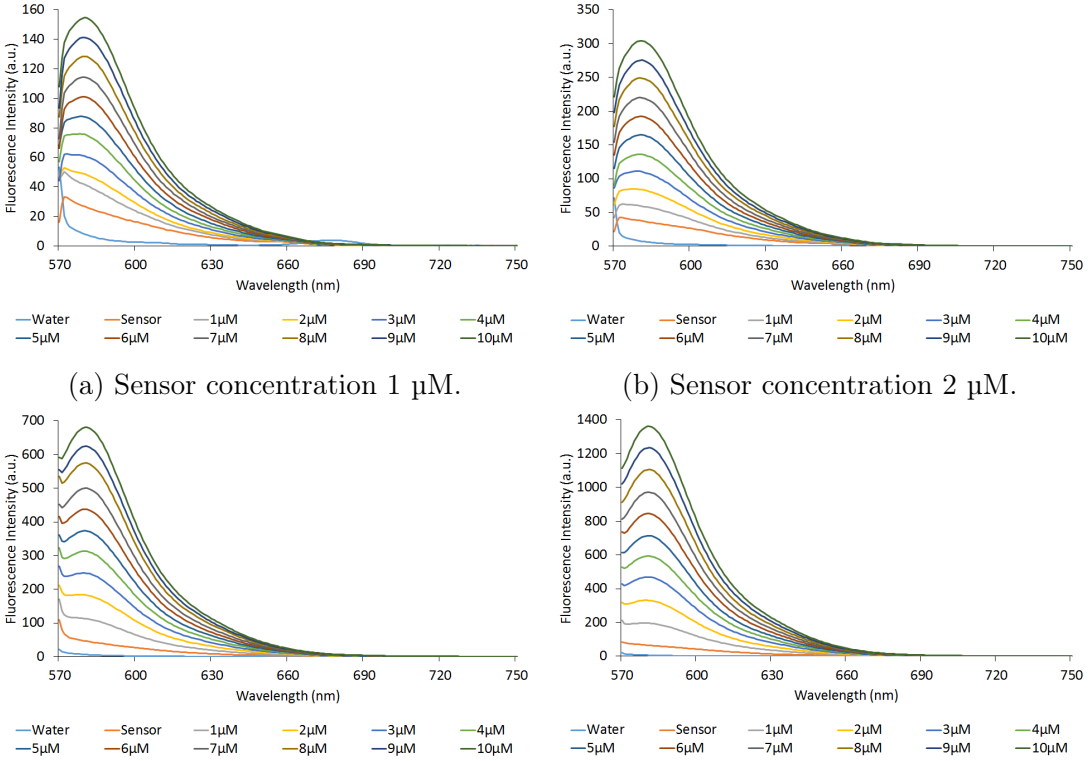
(c) Sensor concentration 5 µM

(d) Sensor concentration 10 µM

Figure 5.8: Fluorescence spectra of kynurenine in various sensor concentrations exciting at 470 nm

for all subsequent additions overlapped with each other. A change in fluorescence of the same kind cannot be seen at any of the other concentrations used either which supports the idea of incomplete mixing.

The fluorescence emission spectra when exciting at 555 nm display a defined increase in fluorescence after each addition of kynurenine. The bands are clearly visible, and even the spectra for 1 µM kynurenine concentration can be seen visibly separate from that of only the background fluorescence of the sensor. This validates the use of fluorescence over absorbance as a technique which can detect the complexation of kynurenine and the sensor even with low concentrations of kynurenine.



(a) Sensor concentration 1 µM.

(b) Sensor concentration 2 µM.

(c) Sensor concentration 5 µM

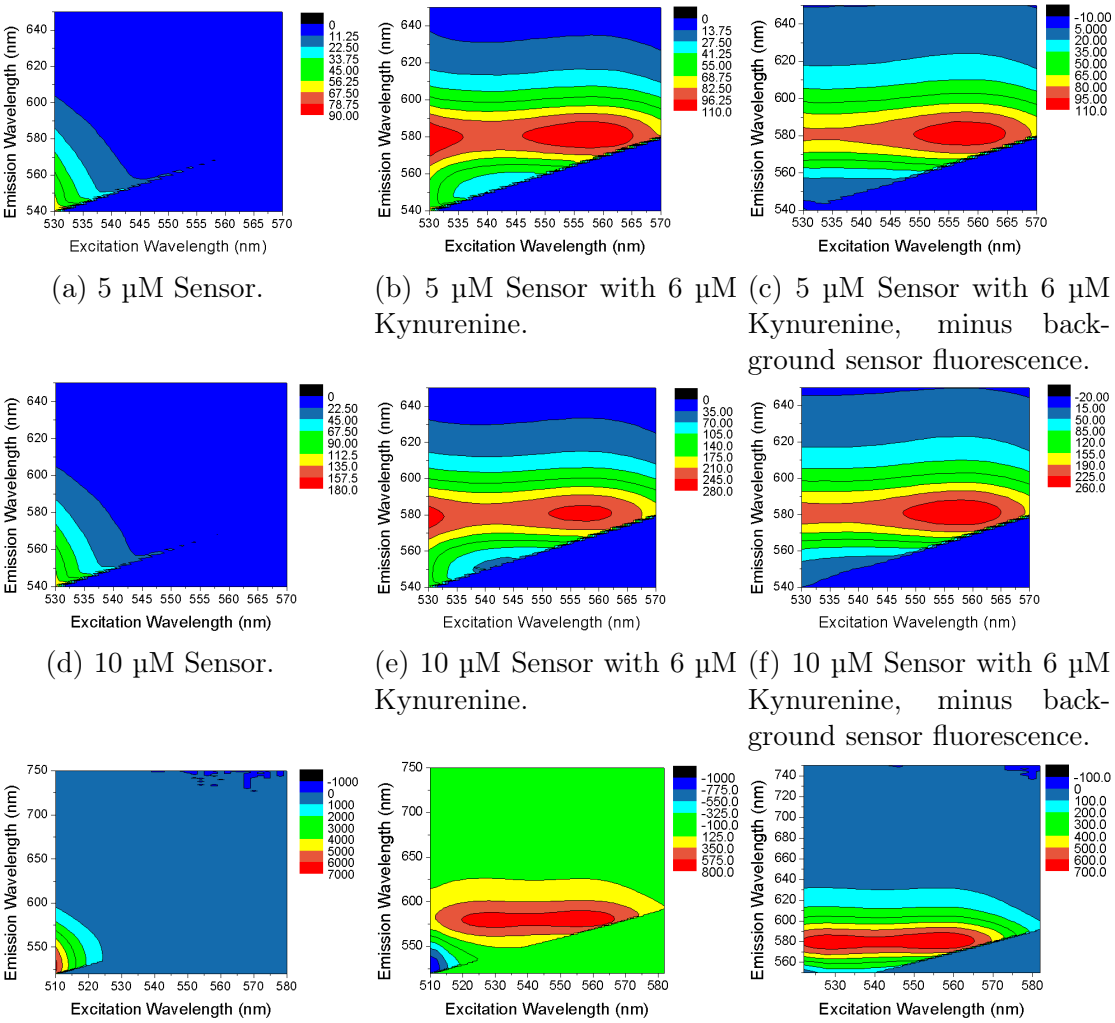
(d) Sensor concentration 10 µM

Figure 5.9: Fluorescence spectra of kynurenine in various sensor concentrations exciting at 555 nm

5.5 Absorption and Emission spectra

3D spectra were collected to determine the optimum excitation and emission wavelengths to use. A 3D spectra is a collection of spectra where emission and excitation wavelengths are changed. A contour plot is produced which shows which excitation wavelengths can produce maximal emission intensities and the wavelengths the measure them at.

When looking at just the sensor (figs. 5.10a, 5.10d and 5.10g) it can be seen that there is minimal background fluorescence from the unbound sensor in the wavelengths targeted for emission by the sensor:kynurenine complex. When kynurenine is added, the spectra appears to have two regions of maximum fluorescence intensity, both centered around a wavelength of 580 nm. Subtracting the background sensor fluorescence from that of



(a) 5 µM Sensor.

(b) 5 µM Sensor with 6 µM Kynurenine.

(c) 5 µM Sensor with 6 µM Kynurenine, minus background sensor fluorescence.

(d) 10 µM Sensor.

(e) 10 µM Sensor with 6 µM Kynurenine.

(f) 10 µM Sensor with 6 µM Kynurenine, minus background sensor fluorescence.

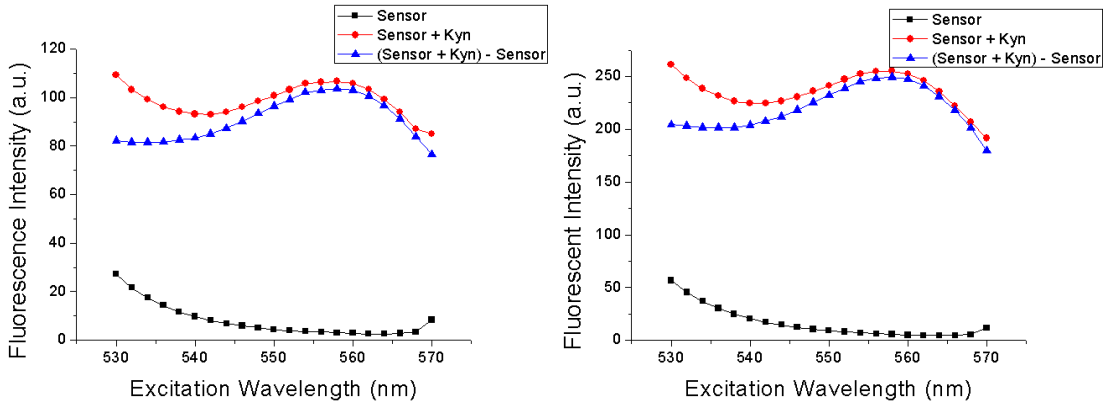
(g) 50 µM Sensor.

(h) 50 µM Sensor with 6 µM Kynurenine.

(i) 50 µM Sensor with 6 µM Kynurenine, minus background sensor fluorescence.

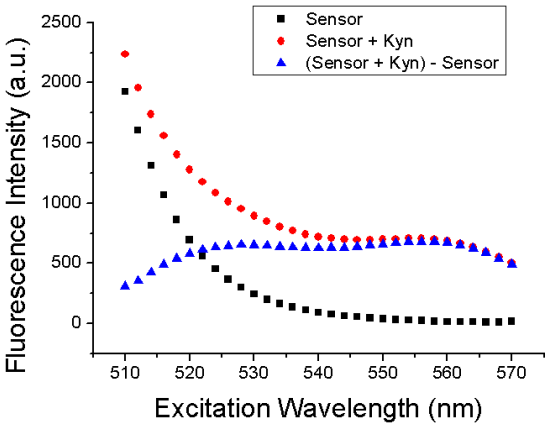
Figure 5.10: 3D Fluorescence spectra. Dark blue represents less fluorescence intensity and red represents the highest fluorescence intensity. The units are arbitrary units. Excitation wavelengths are given on the x-axis (nm) and emission wavelengths (nm) are given on the y-axis.

the sensor:kynurenine complex, figs. 5.10c, 5.10f and 5.10i, shows there is really only one maximum region and that by exciting with a wavelength between 550 - 565 nm, it is possible to reduce the background signal coming from the unbound sensor. To explore this further, the intensity values at an emission wavelength of 580 nm were taken for a range of wavelengths and plotted separately. The plotted data included the data for the sensor, the sensor with kynurenine and finally the sensor with kynurenine subtracting the initial sensor values and can be seen in in fig. 5.11. This data allowed for the excitation wavelength which gives the least background signal to be found, as can be seen in the graphs in fig. 5.11, where the (sensor + kynurenine) and the ((sensor + kynurenine) - sensor) values are closest to each other shows that the background sensor fluorescence is minimised. This value is at 560 nm. So the optimal excitation wavelength is at 560 nm and the optimal emission wavelength is 580 nm. Looking at figs. 5.10 and 5.11 it is clear that there are values around each wavelength that are usable, which could become useful for future experiments as the difference of 20 nm between excitation and emission wavelengths is quite a small gap and some machines may not allow measurements to be taken due without specialist equipment such as a dichroic mirror.



(a) Sensor Conc. = 5 µM, Kynurenine Conc. = 6 µM

(b) Sensor Conc. = 10 µM, Kynurenine Conc. = 6 µM



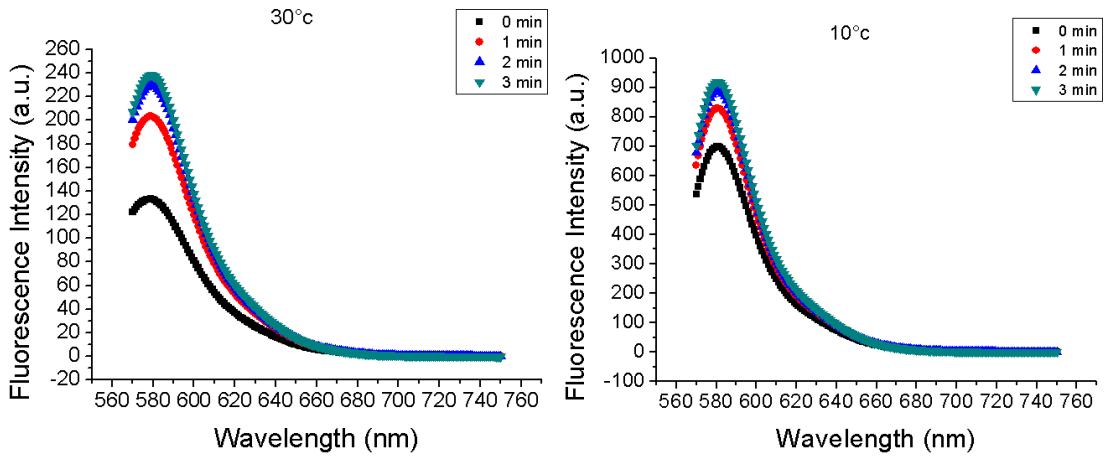
(c) Sensor Conc. = 50 µM, Kynurenine Conc. = 6 µM

Figure 5.11: Fluorescence emission intensity measured at 580 nm.

5.6 Temperature

Whilst conducting some initial tests with the sensor, it was found that the temperature of the sample has an affect on the fluorescence output, with the measured fluorescence decreasing as the temperature increases. When looking at the fluorescence intensity at 580 nm in fig. 5.12, at 30 °C in fig. 5.12a the value is 237.9, whereas at 10 °C in fig. 5.12b the value is 918.0, which means there is a 3.86 times increase in signal when reducing the temperature by 20 °C. To test this in the most extreme case, the solution was heated to 90 °C and compared to the signal from a solution with just sensor in it. The resulting

spectra, as seen in fig. 5.13, shows that the signals are almost identical. This demonstrates that as the temperature increases, the equilibrium for the complex shifts towards the uncomplexed side. A second point to notice from fig. 5.12 is that the fluorescence intensity took time to reach its maximal value. For the data collected in fig. 5.12, spectra were collected until there was no noticeable change in intensity. The solution was mixed thoroughly in the cuvette before being placed into the fluorometer, but this process was not regulated to be strict with timekeeping. This result is consistent with literature of different fluorophores, where the changes in fluorescence have been shown to be due to a combination of changes in binding equilibria and also changes in the photophysics of the molecules [153–155].The section below contains more detailed experiments which also look at the influence of time as well as temperature on the fluorescence intensity. The two are not independent of each other as it takes time for the sample to reach the desired temperature.



(a) Fluorescence at 30 °C

(b) Fluorescence at 10 °C

Figure 5.12: Fluorescence intensity of the same solution at different temperatures

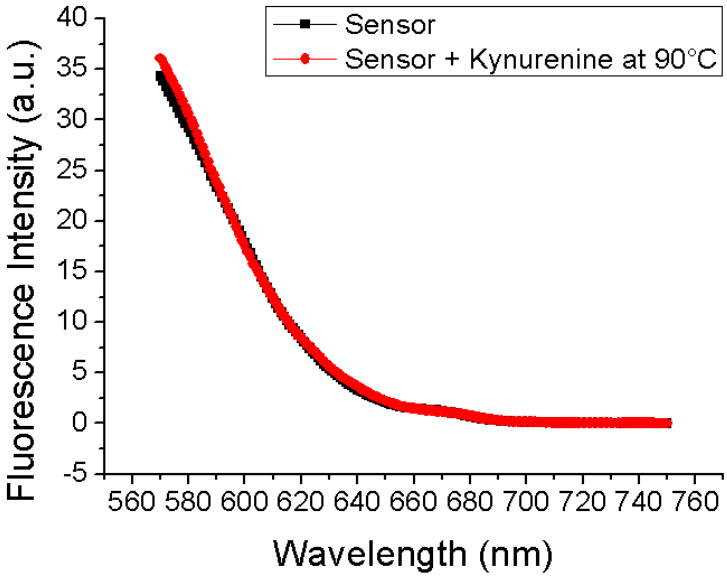
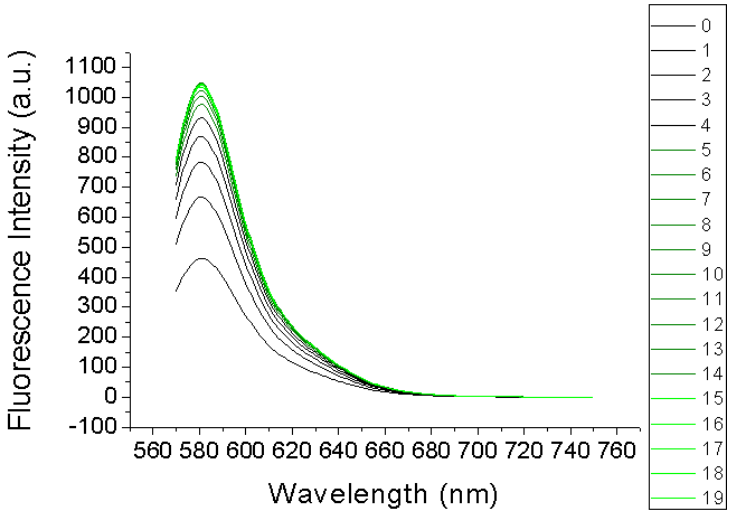


Figure 5.13: Fluorescence intensity at 90 °C

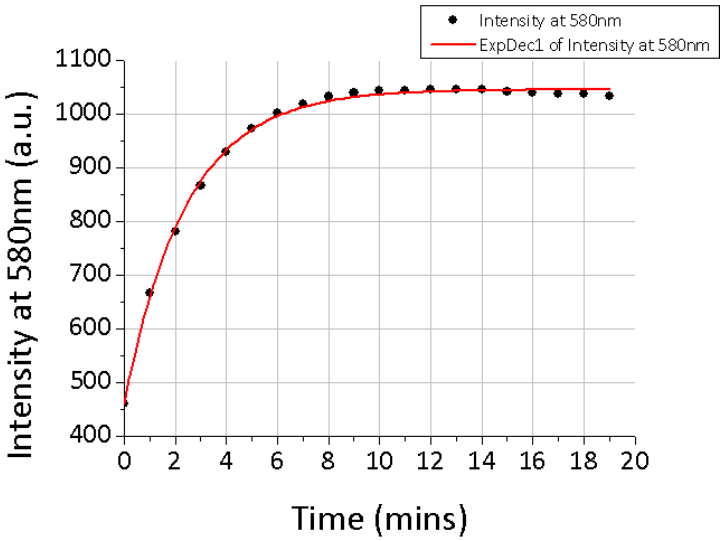
5.6.1 Equilibration time

Cuvettes will typically be filled using stock solutions at the ambient room temperature and so, when measuring at a fixed temperature using a fluorometer, it will take time for the temperature of the cuvette to equilibrate to that of the cell holder. To determine how much time would be needed for this, spectra were recorded every minute for 19 minutes (providing 20 measurements) to determine the optimal waiting time for the temperature to equilibrate. By performing this, it allows the parameters on the instrument to be set so that measurements are taken automatically after the desired time period. The data has been plotted (fig. 5.14) and tabulated (table 5.1) in order to see the true values at the peak emission. Time is taken from 0 minutes as a spectra is recorded immediately upon placement of the cuvette in the instrument until the 19th minute. Using the fitted data it is possible to determine when 99 % of the maximal value is reached, and this is from the measurement taken at the 9th minute.

The time determined by this experiment may only be relevant for cooling a sample to 10 deg from an ambient temperature of around 22 deg, however all future experiments



(a) Individual spectra



(b) Fitted data series taken from each spectra measured at 580 nm

Figure 5.14: Fluorescence intensity of the same solution at different temperatures

Table 5.1: Fluorescent intensity values at 580 nm from fig. 5.14.

Time (mins)	Intensity at 580 nm (a.u.)
0	461.257
1	666.459
2	782.458
3	867.509
4	929.999
5	973.955
6	1002.17
7	1020.56
8	1033.14
9	1040.19
10	1044.02
11	1045.5
12	1046.36
13	1046.11
14	1045.91
15	1041.97
16	1040.69
17	1038.3
18	1037.93
19	1033.96

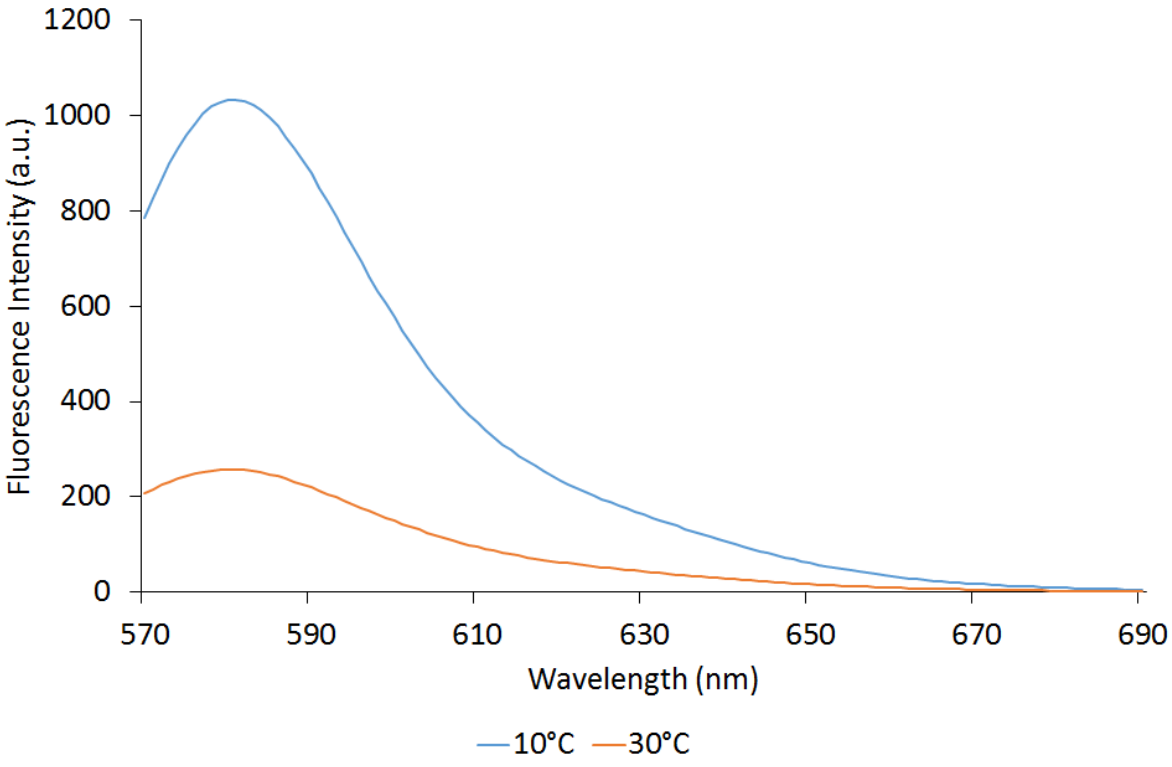


Figure 5.15: Fluorescence intensity after 20 mins

using the fluorometer will be conducted at 10 deg and so this is the only relevant measurement for the continuation of this work. The most important finding to learn from this is that sample measurements should be taken at the same temperatures when collecting a series of data. Previous experiments were all performed using active cooling/heating to maintain a constant temperature, although the temperature was not chosen to optimise the fluorescent signal.

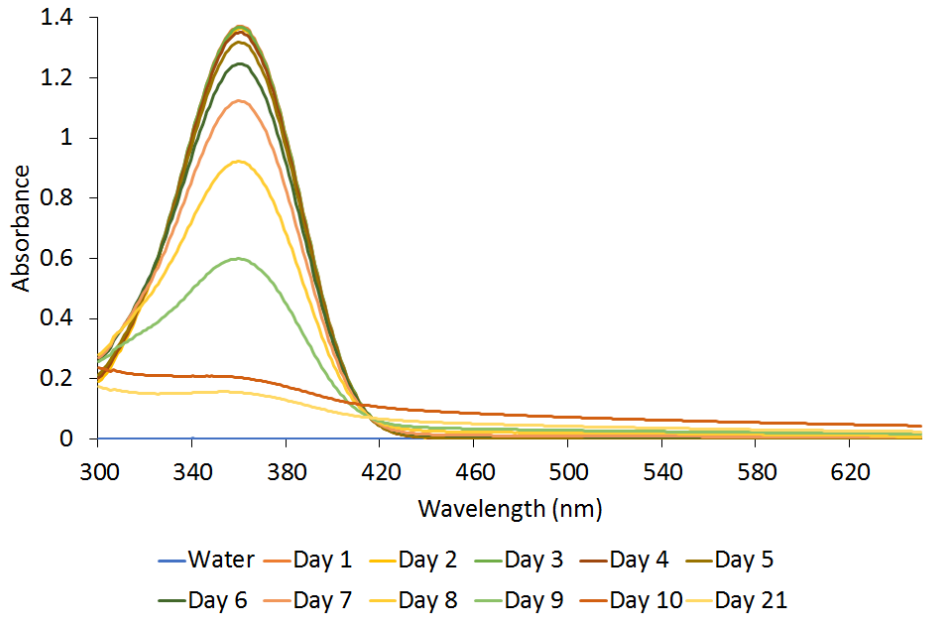
5.7 Stability

Previous experiments were set up through the use of fresh solutions of sensor and kynurenine. This involved weighing tiny amounts of the respective powders (fractions of a milligram) and then dissolving them into the necessary buffers. In order to save time in experimental preparation, it is normal to make up stock solutions which can then be used for many future experiments. As the goal of this project is to create an assay, it would be important to find out the stability of the sensor and of kynurenine in their respective buffer solutions. The importance of this comes from the fact that when testing the ability of an assay to accurately quantify an analyte, it is important to know the amount of analyte in the test solutions. UV-Vis spectroscopy measurements were used for this experiment as the absolute values will then be comparable across days. Two solutions were made for both kynurenine and the sensor, the first where the respective molecule was dissolved in water at neutral pH and the second where the respective molecule was dissolved in a solution of HCl at pH 1.

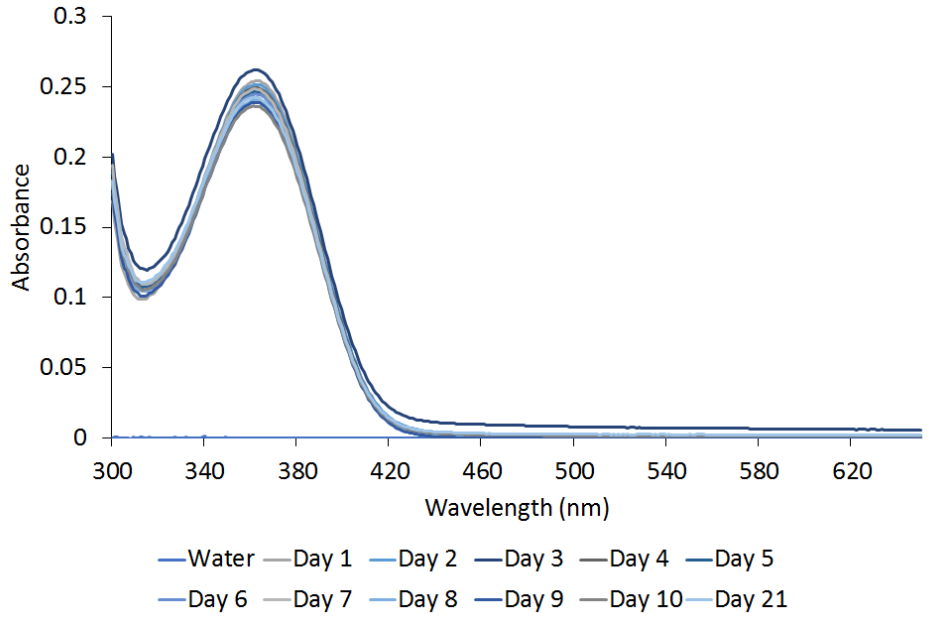
5.7.1 Stability of kynurenine

The results for 500 µM kynurenine can be seen in fig. 5.16. The results show that the absorbance values are always greater when in water at neutral pH as opposed to at pH 1,

however, in water there is a greater change between days. The absorbance values when dissolved just in water do not change drastically for the first 5 days.



(a) UV-Vis absorbance spectra for 500 µM Kynurenine at neutral pH across multiple days.

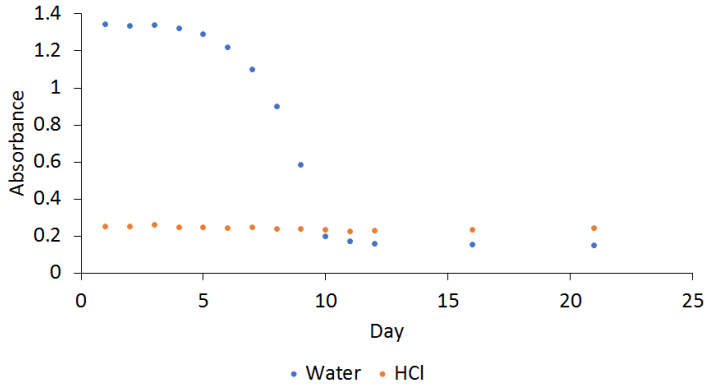


(b) UV-Vis absorbance spectra for 500 µM Kynurenine at pH 1 across multiple days.

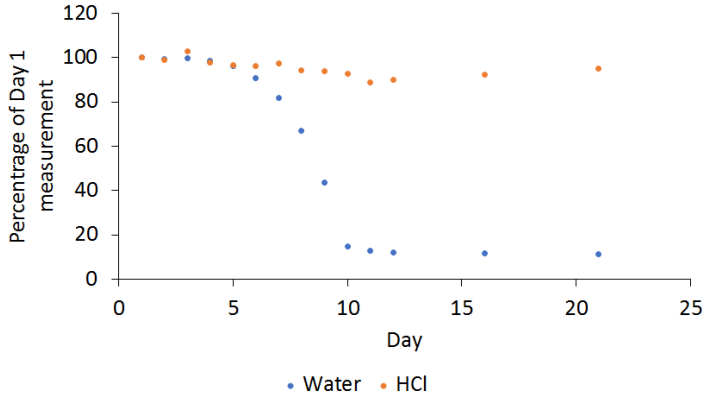
Figure 5.16: UV-Vis absorbance spectra for 500 µM Kynurenine across multiple days.

To examine the results from the experiment further, the absorbance values at 365 nm

have been considered in fig. 5.17. Figure 5.17a shows that after 5 days, the absorbance value begins to decrease sharply until the 10th day at a neutral pH, whereas at pH 1 although much lower, the absorbance value is much more consistent for the whole 21 days. Figure 5.17b is a plot showing each value as a percentage of the day 1 value. When at pH 1, the value goes as low as 88.8 % of the starting value at day 11. For comparison, at neutral pH by day 7 the value is 81.9 % that of the day 1 value, and by day 11 it is only 12.6 % of the day 1 value.



(a) UV-Vis absorbance value at 365 nm for 500 µM Kynurenine across multiple days.



(b) Absorbance values as a percentage of the Day 1 values.

Figure 5.17: UV-Vis absorbance value at 365 nm for 500 µM Kynurenine across multiple days.

The substantial difference in absorbance between the two samples even from day 1 was investigated further. One proposed reason is that of the chloride ions in solution causing the kynurenine to aggregate. To test this hypothesis, instead of adding HCl to the water

to change the pH, NaCl was added such that the chloride ion concentration should be the same as if changing the pH to 1 with HCl. The absorbance spectra were recorded along with one for kynurenine in pure water for comparison with the results shown in fig. 5.18. The graph shows no change in absorbance between the kynurenine solutions and so it is evident that the chloride ions are not the cause of the absorbance change from the previous experiment.

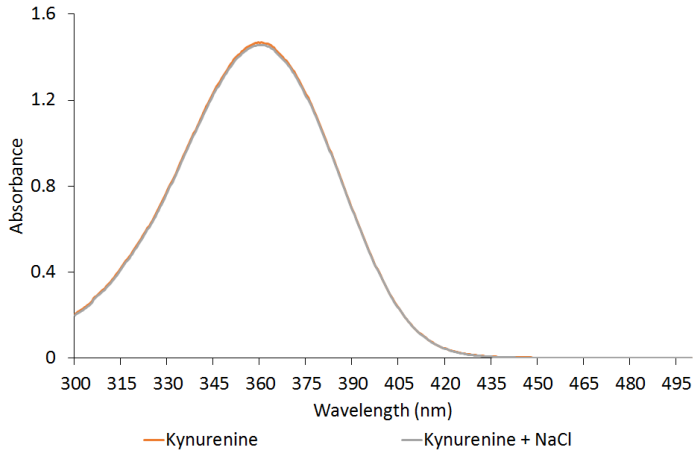


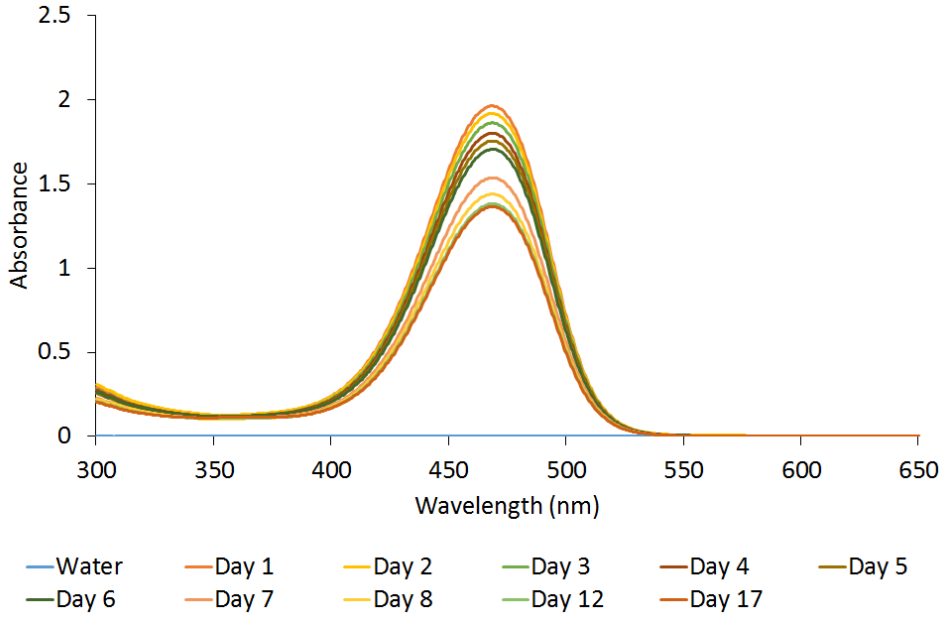
Figure 5.18: Absorbance spectra of 500 µM Kynurenine at neutral pH with and without NaCl added.

The overall result from these experiments is that kynurenine can be stored for 3 weeks in an acidic solution if a stock solution is to be used. Up to 7 days provided changes up to 3.5 %, after this after this the changes fluctuated between 5 - 11 %. Kynurenine should not be stored in water, but if necessary, for no longer than 5 days.

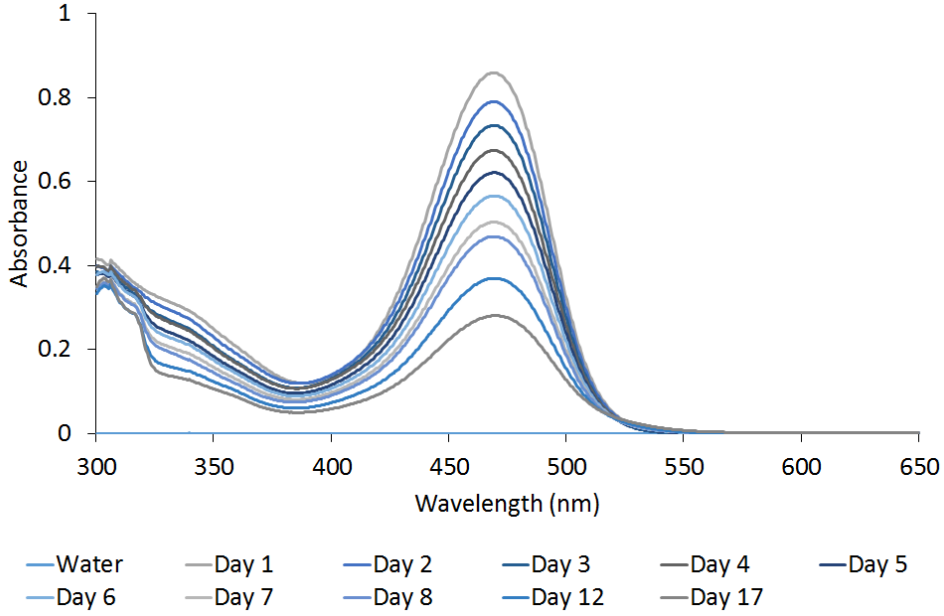
5.7.2 Stability of sensor

The results for the sensor at 50 µM can be seen in fig. 5.19. Again, as with kynurenine, the absorbance values for the pH 1 solutions are all below those of the neutral solutions. Different from kynurenine though is that the measured absorbance decreases for both in a more immediate manner. By day 3, the absorbance value for the solution at neutral pH

has decreased by 5.1 %, whereas by the second day the absorbance value for the solution at pH 1 has decreased by 7.9 % and by the third day it is down by 14.5 %. The rate of change is greater for the pH 1 solution, but neither prove to be stable. Figure 5.20 displays the results focussing on the absorbance at 470 nm.

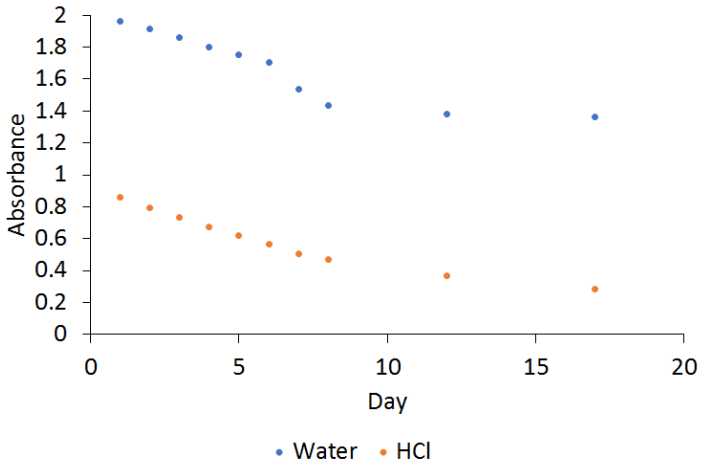


(a) UV-Vis absorbance spectra for 50 µM Sensor at neutral pH across multiple days.

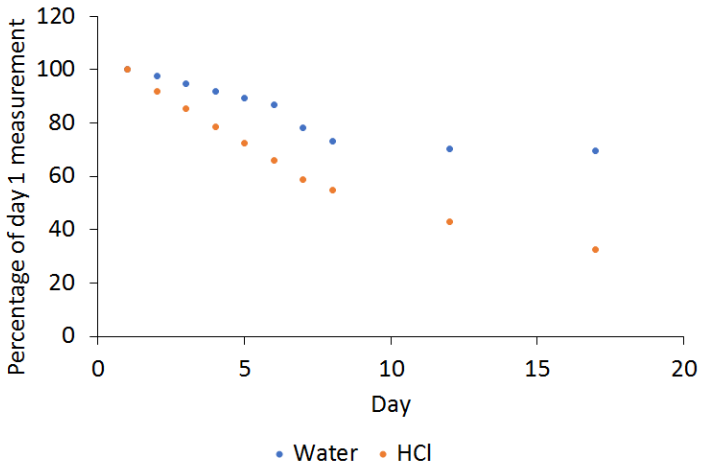


(b) UV-Vis absorbance spectra for 50 µM Sensor at pH 1 across multiple days.

Figure 5.19: UV-Vis absorbance spectra for 50 µM Sensor across multiple days.



(a) UV-Vis absorbance value at 470 nm for 50 µM Sensor across multiple days.



(b) Absorbance values as a percentage of the Day 1 values.

Figure 5.20: Absorbance values as a percentage of the Day 1 values.

As with kynurenine, to check if the chloride ions were responsible for the decreased absorbance between the two solutions, sensor solutions in pure water and also in pure water with added NaCl were prepared. The results turned out the same and the addition of chloride ions did not cause the absorbance to decrease as seen in fig. 5.21. Given these results, the sensor should not be stored in solution for more than a couple of days.

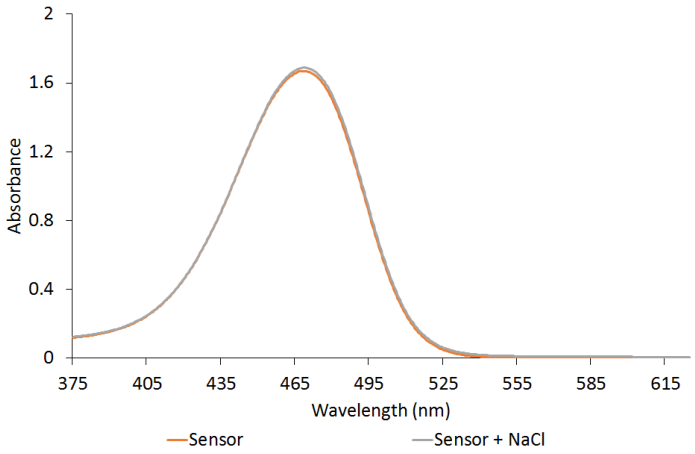
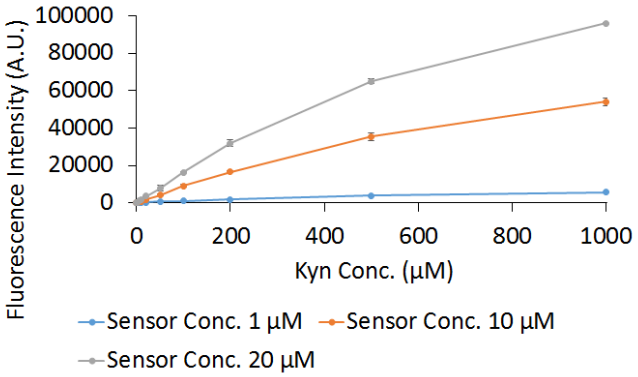


Figure 5.21: Absorbance spectra of 50 µM Sensor at neutral pH with and without NaCl added.

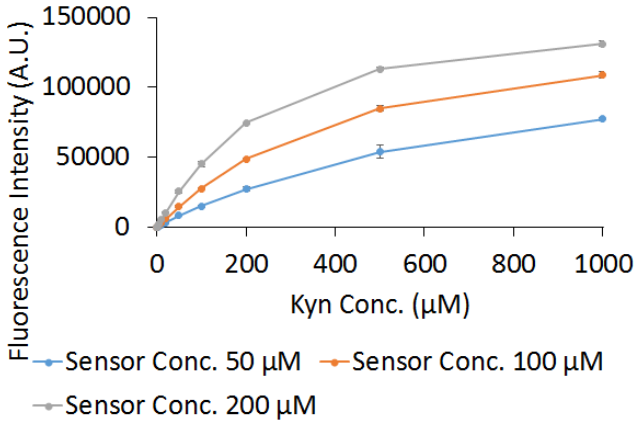
5.8 Linearity of sensor

For the standard addition technique to work, it is important to determine the linear range of the sensor. In order to do this, a range of fixed sensor concentrations were tested with a range of kynurenine concentrations and the fluorescence was measured. Figure 5.22 shows the results of these experiments. The results for 1 - 20 µM and 50 - 200 µM are plotted on separate graphs as they were taken on different days. As can be seen, for all of the concentrations chosen the responses appear to be logarithmic with increasing kynurenine concentrations. Each measurement was taken three times, with the average value plotted and the error bars representing two standard deviations each side giving a range where 95 % of measurements should be recorded assuming the values are normally distributed. Even with error bars of two standard deviations, the bars are not very visible on either graph with the exception of a few data points. The reason for this is that the scale of the graph is so large that it reduces the visibility of the error values, especially at the lower kynurenine concentrations where the response is much lower and this can be seen in fig. 5.23, which shows that the error values are higher at lower kynurenine concentrations compared to the average value used.

The next step for determining a linear range for the sensor was to use the values from the experiment in fig. 5.22, to determine which kynurenine concentrations fell on a linear line of fit. In order to do this, linear fittings were made using the least squares technique starting with the first three concentrations (0, 1 and 5 µM) and then creating new fittings by adding new concentrations to the set, one at a time. The starting point was set to the first three concentrations used as using two points would allow a straight line to connect them. As can be seen from the R^2 values in table 5.2, the higher sensor concentrations of 100 and 200 µM offer a lower linear range along with 1 µM. The sensor concentration of 10 µM doesn't show as good linearity at lower concentrations when compared to 20 and 50 µM concentrations. When using concentrations of 20 and 50 µM, the linear region is at least up to 100 µM kynurenine with the possibility of including 200 µM kynurenine.



(a)



(b)

Figure 5.22: Fluorescence emission of different concentrations of sensor with different concentrations of kynurenine. The fluorescence values plotted are the mean of three repeats and each error bar shows two standard deviations.

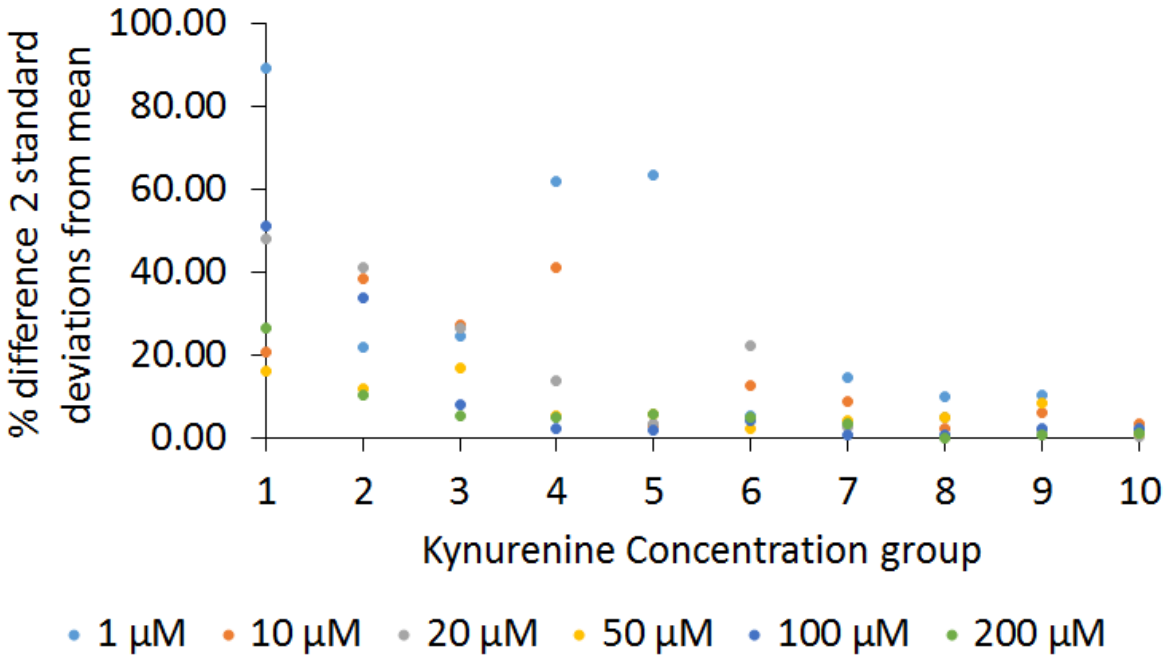


Figure 5.23: Graph showing two times the coefficient of variation for the response at each concentration. Concentrations are set as fixed continuous integers to remove bunching of the lower concentrations and thus provide better visibility. The values 1 to 10 represent the kynurenine concentrations 0, 1, 5, 10, 20, 50, 100, 200, 500, 1000 µM respectively.

Table 5.2: R^2 values when using different ranges of kynurenine concentrations.

	Sensor Concentrations					
Kyn Conc. (µM)	1 µM	10 µM	20 µM	50 µM	100 µM	200 µM
0 - 5	0.8212	0.9971	0.9962	1.0000	0.9981	0.9999
0 - 10	0.9197	0.9907	0.9991	0.9976	0.9996	0.9993
0 - 20	0.9616	0.9966	0.9992	0.9988	0.9999	0.9998
0 - 50	0.9878	0.9989	0.9970	0.9997	0.9998	0.9999
0 - 100	0.9862	0.9978	0.9990	0.9974	0.9990	0.9962
0 - 200	0.9913	0.9982	0.9997	0.9966	0.9954	0.9871
0 - 500	0.9914	0.9939	0.9917	0.9869	0.9656	0.9251
0 - 1000	0.9691	0.9758	0.9683	0.9588	0.9133	0.8437

As a means to directly compare the different sensor concentrations, the data from each

experiment per sensor concentration was normalised to its maximum value and so now each plot will be bound by a maximum value of 1 and the change in response of different concentrations of kynurenine can be seen with respect to the maximum value. As the values in table 5.2 show the linear range is likely to be around 200 µM kynurenine and so the normalised data was cut off to display this part of the response and is shown in fig. 5.24. The figure shows the trendlines for each sensor concentration and it can be seen that sensor concentrations less than 100 µM have the best R^2 values and it can even be seen that for 200 µM sensor the linear trendline does not fit the data well. The trendlines for 10 and 20 µM sensor concentrations both have similar gradients and x-intercepts, however the data from fig. 5.23 shows the 10 µM sensor concentration seems to exhibit greater variability at lower kynurenine concentrations and so the recommendation would be to use sensor concentrations of 20 or 50 µM when wanting to detect low kynurenine concentrations.

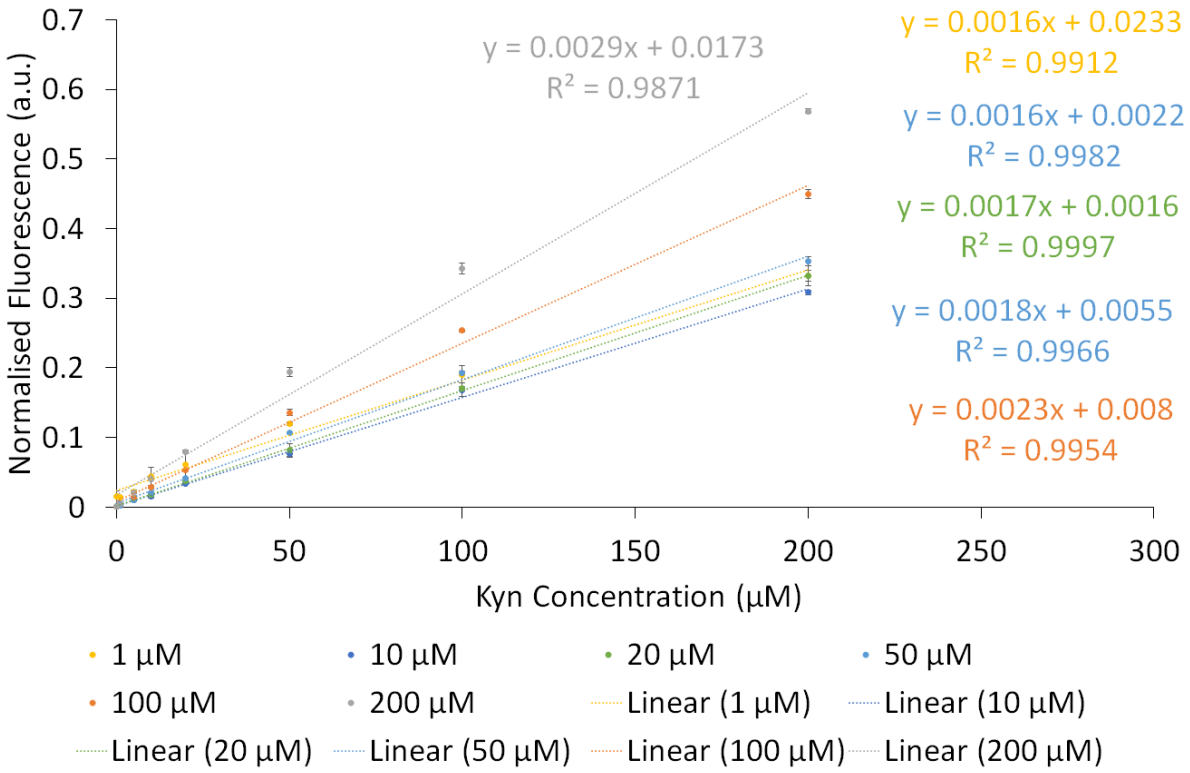


Figure 5.24: Graph showing the normalised response to different mixtures of kynurenine concentrations with different sensor concentrations. The complete data going up to 1000 µM kynurenine concentrations was normalised by dividing each response in an experiment by the maximum response. The data has then been clipped to show values for kynurenine concentrations up to 200 µM. The normalised fluorescence values plotted are the mean of three repeats with the error bars set at one standard deviation.

5.9 Conclusions

This chapter explains why kynurenine was chosen to be the target biomarker. It also explains the experiments undertaken to test the properties of a fluorescent molecule to determine if it is performing as the literature states, and also to determine the best parameters when creating an assay.

The fluorescent molecule was tested by UV-Vis spectroscopy and then fluorescence spectroscopy. The optimal excitation and emission wavelengths of 560 nm and 580 nm respectively was determined by a 3D fluorescence spectra.

The significant affect of temperature on the fluorescence output was discovered and can now be taken into account moving forward.

The stability of the sensor and analyte in water and pH 1 solutions was tested, with kynurenine being stable in acidic solutions, albeit the absorbance being drastically lower than in water. The sensor was not very stable in either condition over many days.

Finally the linear range of the sensor was determined across multiple sensor concentrations. It was determined that a sensor concentration of 20 or 50 µM would offer the best performance with the upper bound of the linear range in the 100 - 200 µM kynurenine concentration range.

CHAPTER 6

ASSAY VALIDATION

This chapter discusses the experiments performed to validate the assay as a quantitative technique. The previous chapter performed tests to check that the fluorescent sensor worked in uncomplicated media, found the optimal measurement parameters, and checked the viability of a fluorescence based assay.

6.1 Standard Addition of pure kynurenine solutions using cuvettes

In order to test the standard addition technique for the quantification of kynurenine, experiments were performed using known concentrations of kynurenine. It was decided that for the starting solution a concentration of 2 µM would be used as this was determined to be a good baseline in serum at least [138].

6.1.1 Experimental

An experiment was set up where a 2 µM solution of kynurenine was made to be the starting solution. 1.9 ml of the kynurenine solution was added to a cuvette along with 0.1 ml of a 1 mM sensor solution to give a cuvette sensor concentration of 50 µM.

A solution of 0.1 mM kynurenine and 50 µM sensor was made to be used for the additions to the starting solution.

Eight additions were made to the starting solution comprising of 10, 10, 20, 20, 50, 100, 200, and 500 µL, which translated to concentration additions from 0.50 to 31.3 µM.

Before each measurement the cuvette was allowed to sit in the fluorometer for 10 minutes to allow the contents of the cuvette to reach the temperature inside the cell holder when coming from the ambient temperature of the room. Fluorescence intensity was measured after this period had elapsed.

6.1.2 Results and discussion

Figure 6.1 shows the intensities for all measurements. The data shows that the greater addition concentrations appear to deviate more from the line of fit. Despite this the line of fit for all data points has an R^2 value of 0.9977 which indicates a good fitting and that it should be possible to use the standard addition technique across this concentration range. For the standard addition technique to work, the initial concentration and all subsequent additions should fall within the linear range of the sensor. As such, there is a greater risk of leaving the linear range when adding higher concentrations of analyte. This experiment starts with adding very low concentrations with respect to the initial concentration at first and then introduces greater additions at the end. To see the affect of adding greater concentrations, sets of consecutive additions have been analysed with

each larger set containing the previous set plus the next addition.

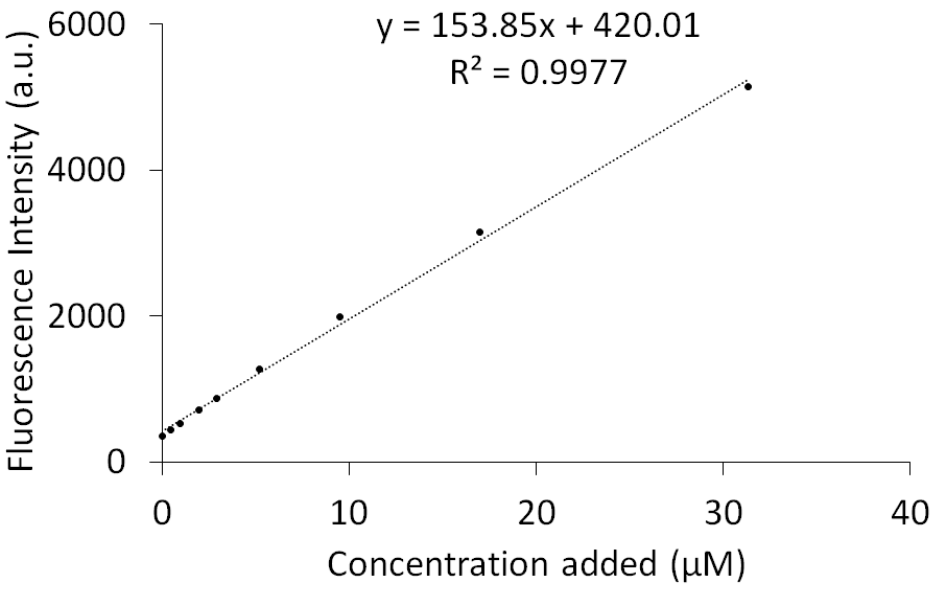


Figure 6.1: Fluorescence intensity of 2 µM starting solution and standard additions. The line of best fit is for all points and has an R^2 value of 0.9977. These results are of a single experiment and so there are no error bars available.

Figure 6.2 displays the predicted concentrations when using an increasing number of additions. The data shows that even though the previous work to reduce the influence of the background sensor signal when selecting the excitation and emission wavelengths, this signal still has an affect on the results which in this instance can make predictions worse by around 10 % of the actual value.

Looking at the values with the background sensor fluorescence subtracted, it can be seen that the best prediction comes from using the first 3 additions which give errors of less than 5 %, whereas the predictions made when using the first 6 additions onwards have error values of over 10 %.

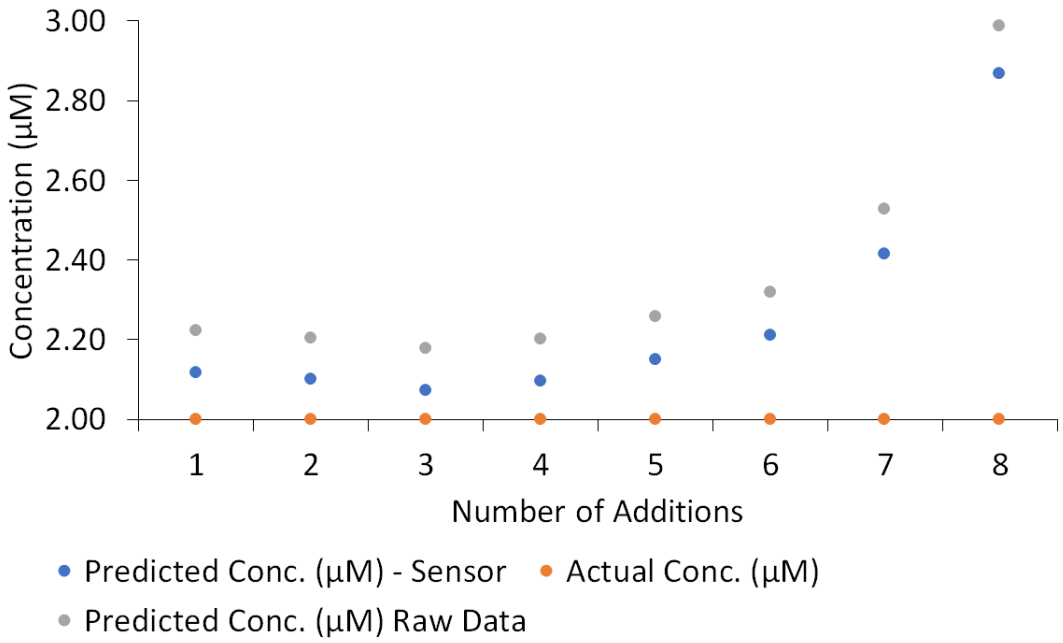


Figure 6.2: Concentration predictions when fitting against different numbers of additions. The actual concentration of 2 µM is shown by the orange dots. The grey dots show the predicted concentrations when using the raw data and the blue dots show the predicted concentrations when subtracting the background sensor fluorescence from the raw data.

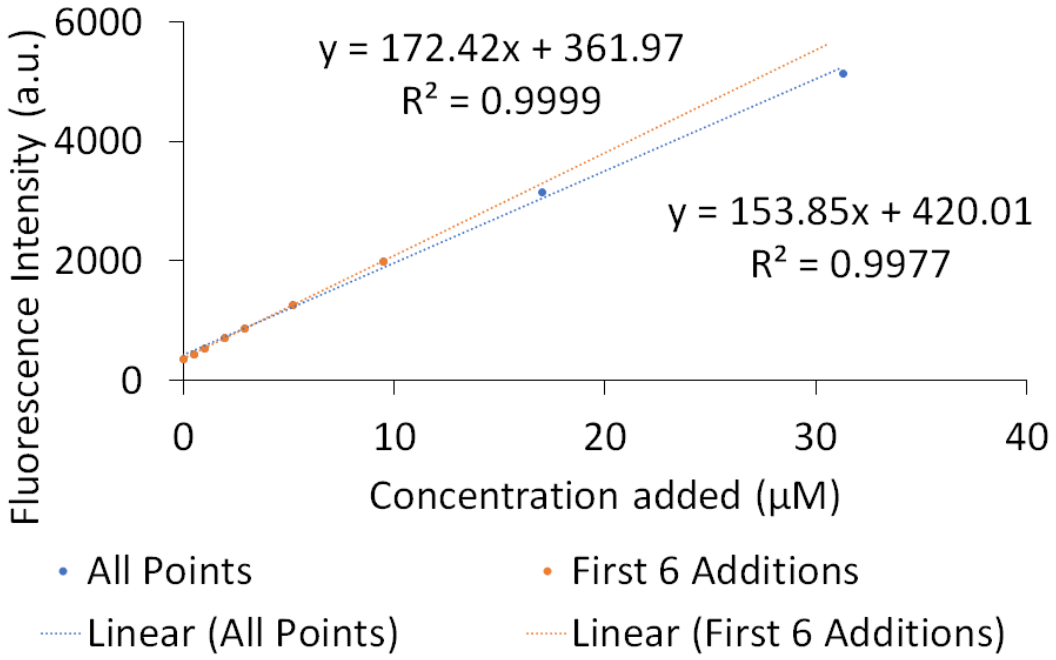


Figure 6.3: Fluorescence intensity of starting solution and additions showing the differences between the lines of fit created when using all points ($R^2 = 0.9977$) vs the first 6 additions ($R^2 = 0.9999$). This is a single experiment so no error bars are presented.

When using different sets of points to create a line of fit, fig. 6.3 shows going up to 6 additions gives a line of fit with an R^2 value of 0.9999. This is closer to 1 than the value of 0.9977 when using all data points and so indicates a better linear fit, however the improvement in linearity is not so substantial as to expect the predicted values to change from being around 5 % from the starting value to over 40 % away. On top of this all of the concentrations used should fall within the linear range of the sensor and an R^2 value of 0.9977 when fitting all of the data points indicates good linearity and so it would be expected that the whole set of points should provide good predictions. When looking at the line of fit using only the lower addition concentrations in fig. 6.3, the higher concentration points fall below this line which would cause the gradient of the line when using all points to be shallower and thus the predicted concentrations to be greater when using all points. One reason for why the recorded intensities would be lower than expected is the temperature. For this experiment the cuvette was allowed to sit in the cell holder for 10 minutes, which was determined from a previous experiment to be enough time for the fluorescence intensity to stabilise. However that experiment was for a fixed volume (2 ml). The larger additions changed the volume of the cuvette more dramatically and so could have required more time to equilibrate. Table 6.1 provides more details for this experiment. It can be seen that additions of larger volumes into the solution lead to greater differences from the actual value, and having more volume of solution in the cuvette would lead to a increase in the time needed for the solution to reach the temperature of the cell holder within the fluorometer. Because of this, by discarding the later additions, the assay has shown to be capable of accurately determining a concentration of 2 µM.

Table 6.1: Predicted concentrations and number of additions used to create the predictions. Information on the volume added, cumulative concentration within the cuvette and the percentage difference from the actual concentration are also provided.

No. of additions	1	2	3	4	5	6	7	8
Volume added (µL)	10	10	20	20	50	100	200	500
Cumulative conc. added (µM)	0.50	0.99	1.96	2.91	5.21	9.5	17	31.3
Starting conc. (µM)	2	2	2	2	2	2	2	2
Predicted conc. (µM)	2.12	2.10	2.07	2.10	2.15	2.21	2.41	2.87
% Difference from actual	5.88	4.99	3.66	4.80	7.52	10.48	20.70	43.31

This experiment has demonstrated that the standard addition technique can be used to quantify a concentration of analyte in a sample. The experiment also provides evidence that a large number of additions do not need to be performed. Accuracy is very good using 2 - 4 additions. It also brings to attention some of the limitations of the standard addition technique. For accurate results, it is essential to keep conditions consistent across additions, with small changes in R^2 values away from 1, although indicating good linearity, greatly affecting the accuracy. Due to this, experimental considerations need to be taken and the following section explains a change to the experimental set up.

6.2 Standard Addition and calibration curve of pure kynurenine solutions using a Microplate reader

To test the assay further it was decided to move from a fluorometer to a microplate reader. By utilising a microplate reader, data collection times can be drastically reduced as the fluorescence across the plate, and therefore all solutions, is measured at the same time. As the working volume used in each well will be the same, the temperature should be the same across the plate and so any temperature affects are minimised.

96-well plates provide enough sample wells for standard addition plots and a calibration curve to be recorded at the same time. This allows for a comparison to be made for the linearity of the sensor.

The fluorometer data showed the standard addition technique worked well for a solution of 2 µM kynurenine, a concentration which was based on the lowest concentration needed to be detected based on serum data. With the microplate reader a 6 µM kynurenine solution was tested. Accurate determination of this higher concentration could help the early diagnosis of sepsis as the concentration of kynurenine in serum has been shown to increase in patients who develop sepsis [138].

6.2.1 Standard Addition

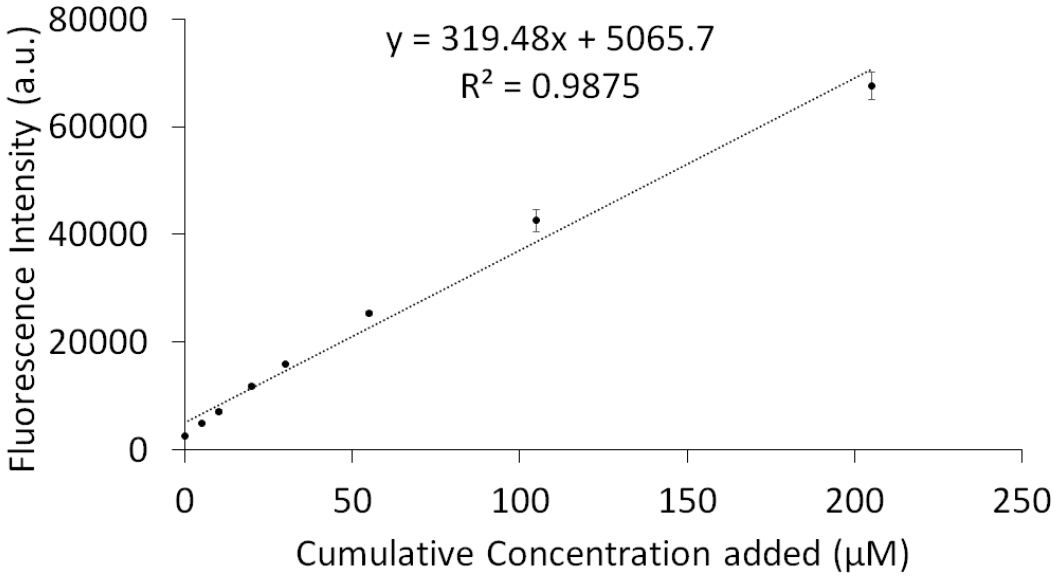


Figure 6.4: Fluorescence intensity of 6 µM starting solution and standard additions. The data plotted are the average fluorescence values across three repeats with the error bars representing one standard deviation.

For the standard addition plot, additions of 5, 10, 20, 30, 55, 105 and 205 µM were added to the starting solution. Figure 6.4 shows the fluorescence intensities recorded. From the figure it seems clear that the fluorescence signal from the higher concentration additions

do not belong to the linear response of the sensor. It can also be seen that if using the line of best fit from all of the additions, the value of the y-intercept is greater than that of the recorded value, being similar to that of the recorded value of the first addition, and also the slope of the line is shallower with respect to the first 6 additions. This will have the combined affect of over-estimating the concentration of the initial solution.

As with the fluorometer experiment, the background sensor signal has been subtracted and predicted concentrations of kynurenine have been calculated for all consecutive sets of additions with the results shown in fig. 6.5. As predicted based off of the intensity plot from the standard additions, the higher concentration additions reduce the accuracy of the predictions, with additions of up to 30 µM of kynurenine providing the greatest accuracy with the predictions being within 5.3 % of the actual value.

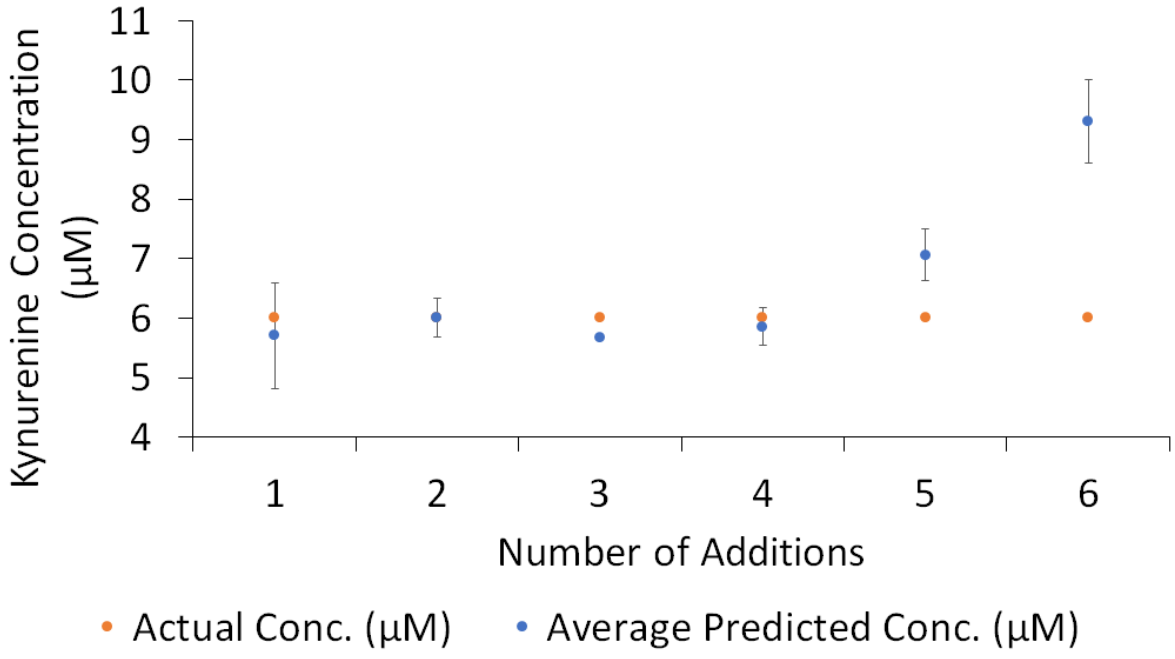


Figure 6.5: Concentration predictions when fitting against different numbers of additions. The actual concentration of 6 µM is shown by the orange dots. The blue dots show the average predicted concentrations from three experiments with the error bars showing the standard deviation of the predictions.

The fitting when using a single addition shows good accuracy, however the standard deviation for this point is quite large. Noise within the measurement could be the cause of

this. Figure 6.4 has error bars set to the standard deviation for each set of measurements, however as has been stated previously, due to the magnitude of the fluorescence values at higher concentrations, the error bars are not clearly visible on the graph. The standard deviation is provided in table 6.2 along with the coefficients of variation. There is no trend in the coefficients of variation which indicates no bias in the machine's response for specific concentrations. The data for this experiment show a greater variation in values recorded for the second addition when compared to the first, however when looking at the predictions made from the recorded values, incorporating the second addition provides a more accurate prediction with less variation. This appears to be due to how the different values for the starting measurement pair up with the values for the first addition. Each well of the same concentration can produce a range of fluorescence values when measured. This is due to a combination of errors which accumulate when creating the microplate such as from pipetting solutions and weighing small masses, and variability from the machine response itself. Because of this, if the recorded fluorescence from the first well is towards the low end of the response range, but the recorded fluorescence from the second well is towards the high end of the response range, the line of fit connecting the two points will be steeper than it should be and the y-intercept will by lower, thus the overall result will be expected to be lower than the actual value. Having multiple measurements in close proximity seems to have a stabilising affect on the predictions, as additions 2, 3 and 4 have smaller error bars than when using only the first addition. When taking the average prediction across the three experiments however, the accuracy is very good using 1 - 4 additions.

Table 6.2: Mean, standard deviation and coefficient of variation for the fluorescence intensity values. Figure 6.4 implies the additions of greater concentration have larger errors, however as can be seen when looking at the data in the table below, error values for the lower concentrations are similar in proportion to the average recorded values. The coefficient of variation values do not correlate with the error values displayed in fig. 6.5.

Addition	Cumulative Concentration Added (µM)	Mean	Standard Deviation	Coefficient of Variation (%)
0	0	2604.67	126.53	4.86
1	5	4912.33	203.35	4.14
2	10	7001.67	381.67	5.45
3	20	11729.00	352.18	3.00
4	30	15997.33	163.03	1.02
5	55	25435.67	391.68	1.54
6	105	42550.00	2110.47	4.96

Table 6.3: Mean, standard deviation and coefficient of variation for the predicted concentrations.

Addition	Cumulative Concentration Added (µM)	Mean	Standard Deviation	Coefficient of Variation (%)
1	5	5.71	0.94	16.47
2	10	6.02	0.28	4.58
3	20	5.68	0.08	1.36
4	30	5.85	0.36	6.08
5	55	7.07	0.48	6.83
6	105	9.31	0.71	7.65

6.2.2 Calibration Curve

For this experiment a calibration curve from 1 to 10 µM was created with the results shown in fig. 6.6. The R^2 value of 0.9947 indicates good linearity. There is no pattern to whether the points lie above or below the line and no other fitting represents the data

(e.g. exponential) and so a linear fitting is appropriate to use.

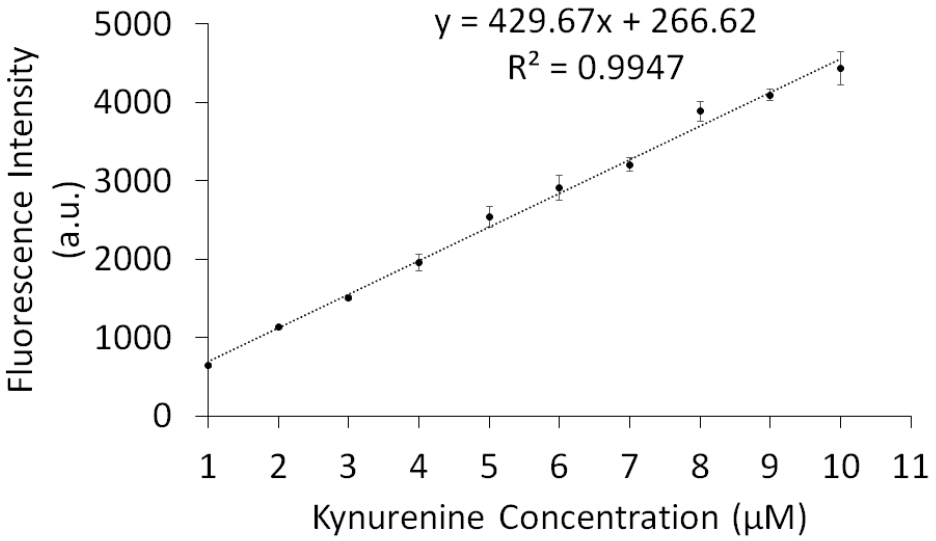


Figure 6.6: Calibration curve of 1 - 10 µM Kynurenine solutions. Plotted values are the mean fluorescence from three repeats with error bars set to one standard deviation.

Using the calibration curve to determine the concentrations of the samples the calculated concentration was 5.90 µM ± 0.29. This falls in line with the results from the standard addition experiments, where the most accurate results were 6.02 µM ± 0.33 and 5.85 µM ± 0.32.

For pure solutions of kynurenine it can therefore be seen that the accuracy of the standard addition technique is the same as using a calibration curve. The linear range of the sensor appeared to be lower than previous experiments indicated. This can be further investigated through the creation of calibration curves along with the standard addition experiments.

Given the quantitative accuracy in pure solutions, the next stage is to try the experiments in more complicated media.

6.3 Quantification of kynurenine in synthetic urine

To test the assay further in a more complicated media, experiments were carried out using synthetic urine. For this project the synthetic urine was purchased as the product Surine (purchased through Merck, produced by Cerilliant) which has been designed for use with assay development and is suitable for use with HPLC.

As with the previous microplate experiment, standard additions and calibration curves were produced. The amount of Surine used per well was 5 µL with all wells having a total volume of 200 µL. The reasoning behind this was due to the limited volumes available from actual urine samples. When moving to actual clinical samples, HPLC and fluorescence experiments would need to be performed along with repeats of these experiments so as such, 5 µL was determined to be a safe amount to use.

6.3.1 Standard Addition

Standard addition experiments were carried out using the Surine with average fluorescence intensities shown in fig. 6.7. The concentrations of kynurenine added were 5, 10, 20, 30, 55 and 105 µM.

The first analysis used the wells containing the 5 µM kynurenine in Surine solution as the starting point and fig. 6.8 shows the results of using the standard addition data to calculate this concentration. The data shows that when using a single addition, there is a lot of variance in the predictions. It also shows that going over 100 µM kynurenine may be leaving the linear range of the sensor. Therefore, ignoring the predictions when having added 5 and 100 µM kynurenine, the predictive accuracy of the standard addition assay is within 10 - 20 % of the actual concentration.

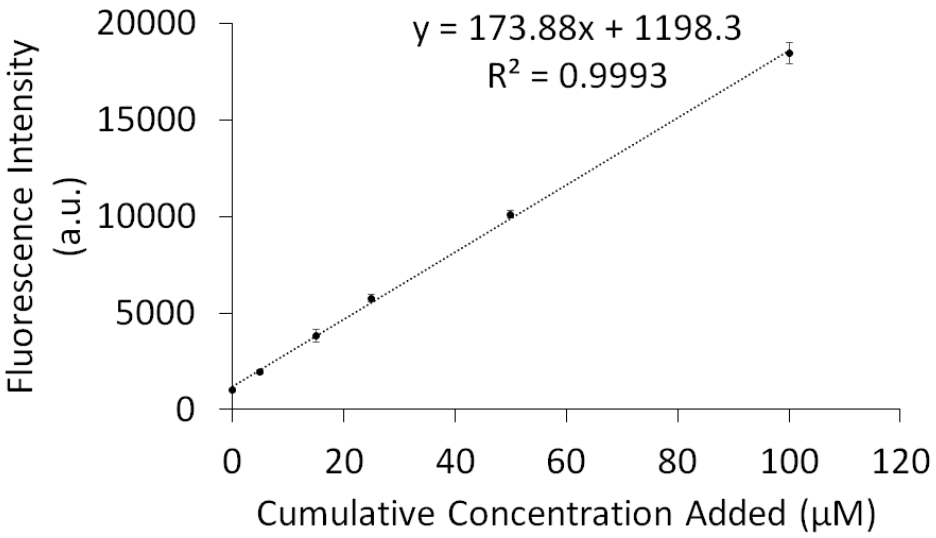


Figure 6.7: Fluorescence intensity of Surine solution and standard additions. The data plotted are the average fluorescence values across three repeats with the error bars set to one standard deviation. $R^2 = 0.9993$.

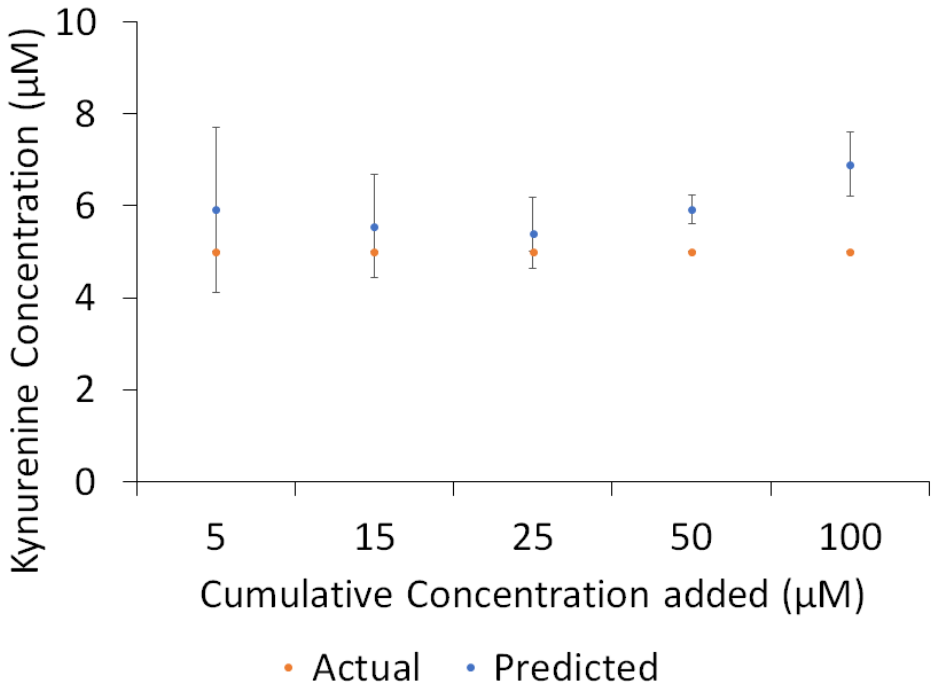


Figure 6.8: Predictions when using standard addition data with Surine solutions for a 5 µM solution. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

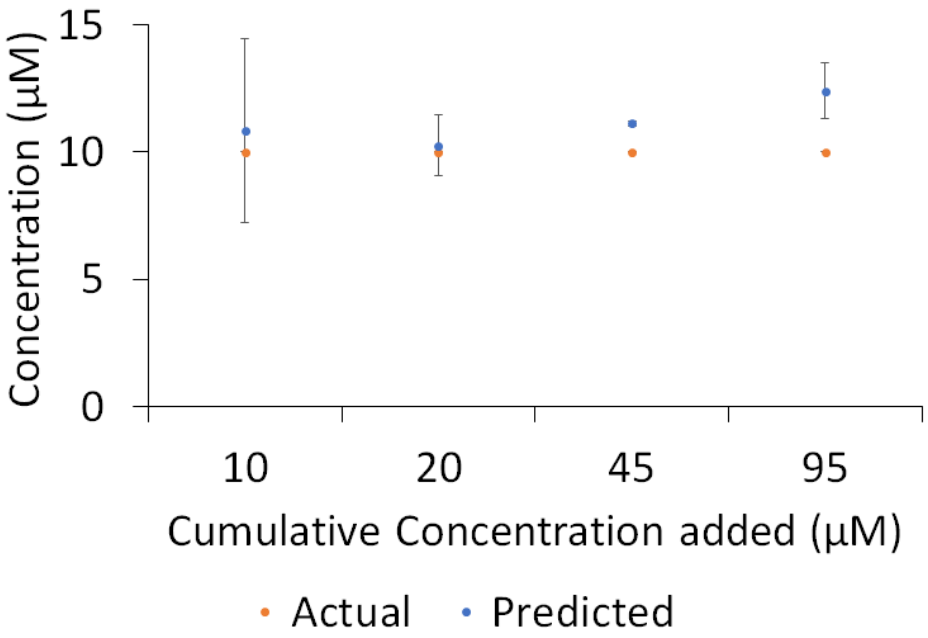


Figure 6.9: Predictions when using standard addition data with Surine solutions for a 10 µM solution. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

For this experiment the 10 µM kynurenine in Surine solution was also analysed with the results given in fig. 6.9. The results are similar to that of the 5 µM data, where the first predictions have a lot of variance and the final prediction appears to be just outside of the linear range of the sensor. The predictions when having added 20 and 45 µM kynurenine fall within the same under 20 % difference from actual as with the 5 µM solution.

6.3.2 Calibration Curves

A calibration curve was created with concentrations from 1 to 1000 µM and can be seen in fig. 6.10. The curve shows the linear range is the same as should be expected by previous experiments.

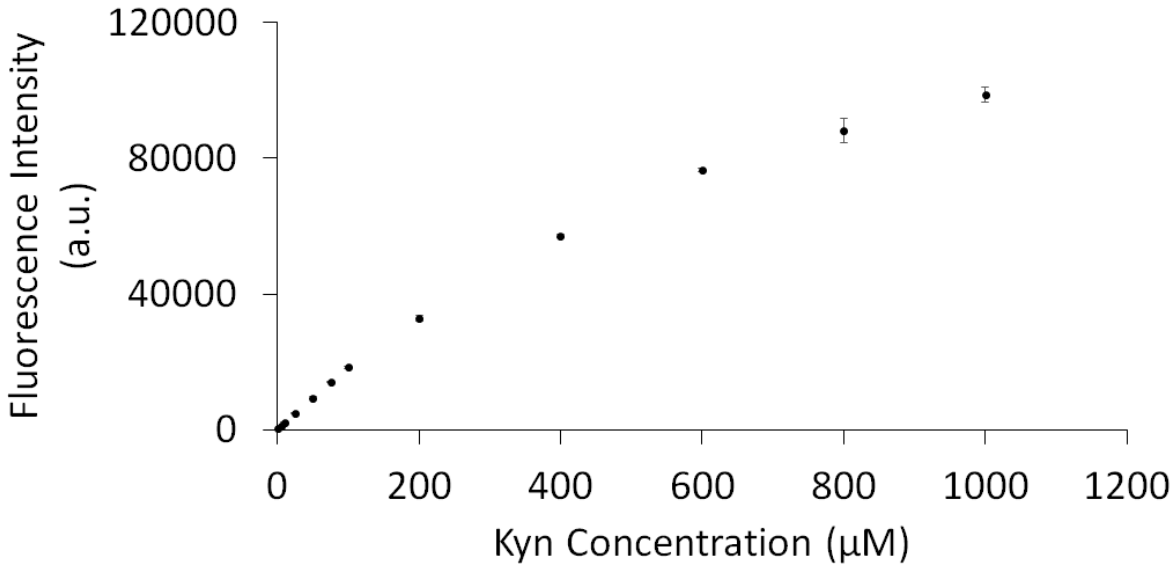


Figure 6.10: Fluorescence intensity of starting solution and additions. Plotted values are the mean of three experiments with the error bars set at one standard deviation.

As the calibration curve created for this experiment extends over a larger range than the previous experiment, it is possible to test multiple concentration ranges to see the affect on the predictive accuracy. Figure 6.11 shows the results of using different ranges for the equation of the calibration curve. The results for using the calibration curve points up to 100 µM are all similar, whereas including the next point, 200 µM, reduces the accuracy for the lowest tested concentrations. This shows the linear range of the sensor is at least up to 100 µM kynurenine. When using the calibration curve calculated using all of the points up to 200 µM kynurenine, the accuracy of the predictions for solutions >10 µM are all within 4 % of the actual value which is very good and shows that 200 µM must just be outside of the linear range. The predicted values when using this curve to calculate the 5 µM and 10 µM concentration solutions differed from the actual concentrations by 51.9 % and 21 % respectively, which is a huge difference from using any of the other set of points where the accuracy was within 7 % and 2 % respectively. Calculating the R^2 values for each regression line shows the values dropping below 0.999 when using 200 µM and above which again shows that when using the calibration curve, the curve should be made with all concentrations up to 100 µM kynurenine.

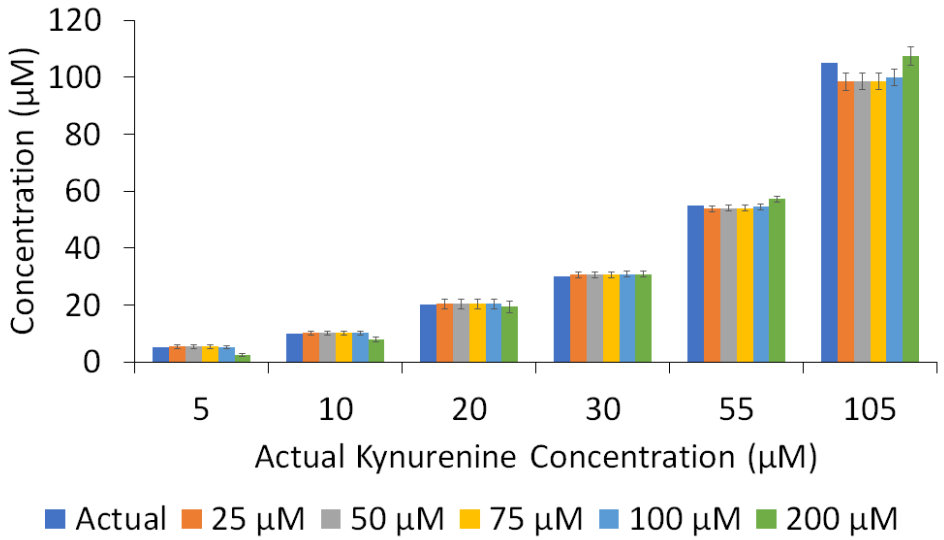


Figure 6.11: How predictions using different ranges from the calibration curve differ. Predicted values given are the mean of three experiments with the error bars set at one standard deviation. The bars are grouped by the actual concentration of the solutions, with the dark blue bar in each group providing a reference against the y-axis for the actual value. The orange bar is for predictions using the slope produced using values up 25 µM, the grey bar up to 50 µM, yellow bar for up to 75 µM, the lighter blue bar for up to 100 µM and he green bar for up to 200 µM. The predictions using up to 100 µM show great similarity, which indicates the slope created up to 100 µM has the same linearity across all points measured. The predictions when using the calibration curve points up to 200 µM are less accurate for lower concentrations and are different to the grouping visible for up to 100 µM, which indicates 200 µM is outside of the linear range.

6.4 Data acquisition time and variance

6.4.1 Use of initial measurements compared to last measurements

The temperature across a microplate should be consistent, however the microplates were still allowed to sit in the plate reader for 10 minutes, with measurements taken roughly each minute. The data analysis up to this point was performed on only the last measurement of each well per microplate, allowing time for sensor:analyte complexation to occur and also for the temperature to stabilise. One of the issues when using a cuvette based

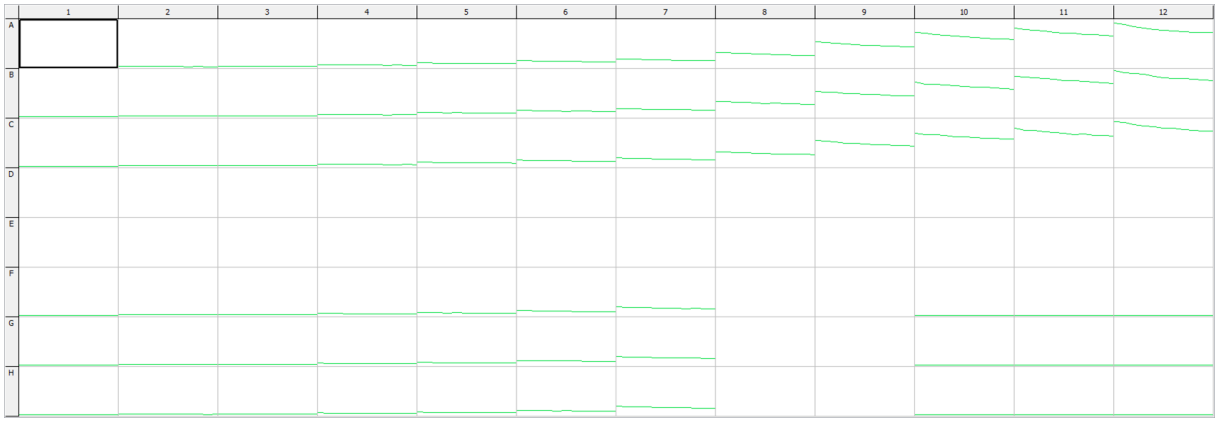


Figure 6.12: Overview of the previous synthetic urine microplate experiment. The intensity for each well is plotted over time and laid out in the same arrangement as the microplate used (example given in fig. 7.12). The top three rows are repeats of the calibration curve, with increasing kynurenine concentration towards the right. The bottom three rows are repeats for the standard addition (first 7 columns), blanks and background measurements (last three columns).

system was that if the temperature for each cuvette being measured were not the same, the maximum fluorescence intensity for each solution would not be comparable across each measurement to make an accurate standard addition plot. As the wells of a microplate are filled in advance and measured at roughly the same time, this should alleviate the affect of temperature, but also should mean that the waiting time for the microplate wells to reach ambient/machine temperature is not necessary. To test this, data from the first recording of the microplate was analysed and compared to the previous results using the last measurements.

The intensity values over time are shown in fig. 6.12 for the synthetic urine experiment. The figure shows that there is a decreasing trend in fluorescence intensity, especially for wells with greater fluorescence. If the results when using the first measurements are similar to those using the last measurement, this will demonstrate that there is no need to wait for the fluorescence values to stabilise, thus saving time when collecting data and increasing throughput for a potential clinical assay.

Starting with the calibration curve given in fig. 6.13, the predictions from both this and fig. 6.11 excluding the predictions for 5 and 10 µM solutions using the 200 µM

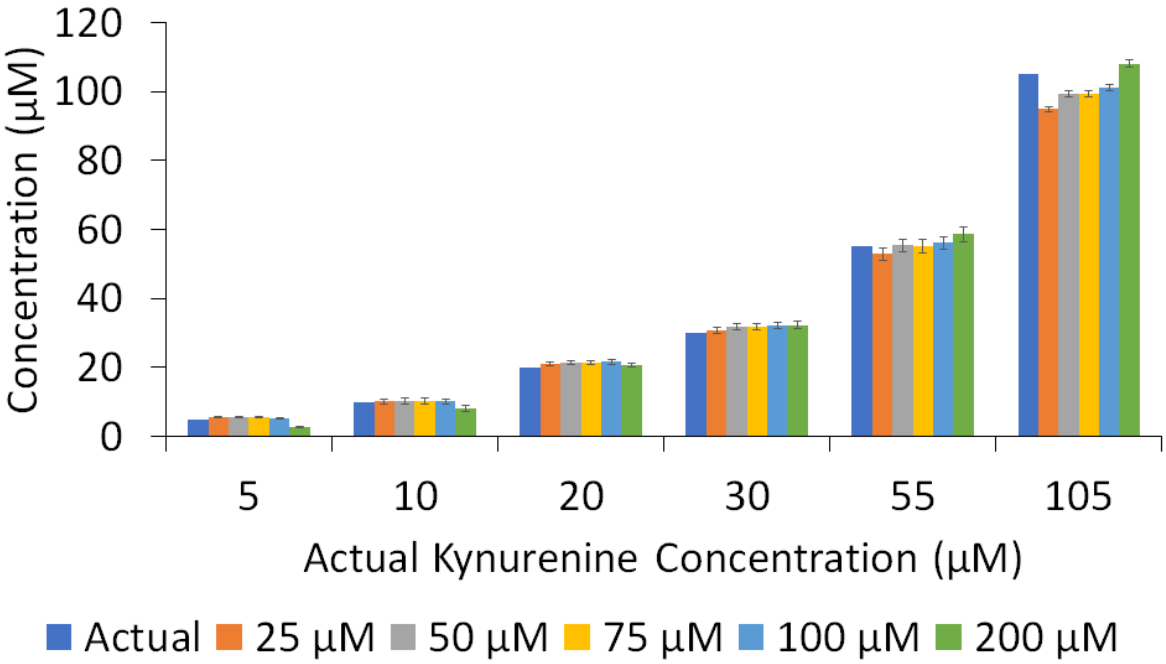


Figure 6.13: How predictions using different ranges from the calibration curve differ when using the first recorded values from the microplate reader. Predicted values given are the mean of three experiments with the error bars set at one standard deviation. The bars are grouped by the actual concentration of the solutions, with the dark blue bar in each group providing a reference against the y-axis for the actual value. The orange bar is for predictions using the slope produced using values up 25 µM, the grey bar up to 50 µM, yellow bar for up to 75 µM, the lighter blue bar for up to 100 µM and he green bar for up to 200 µM. The predictions using up to 100 µM show similarity, which indicates the slope created up to 100 µM has the same linearity across all points measured. These predictions also offer good accuracy where the calculated concentrations are within 10 % of the actual concentration. The predictions when using the calibration curve points up to 200 µM are less acccurate for lower concentrations and are different to the grouping visible for up to 100 µM, which indicates 200 µM is outside of the linear range.

calibration curve values are all within 12 % of the actual value. There is no trend to the magnitude of the standard deviation values (error bars) with no series of measurements having values consistently smaller than the other. Overall because of these facts, when using the calibration curve, once a microplate has been made there is no need to wait to get measurements compatible with one another.

The standard additions can also be analysed in the same manner as has been previously done. First looking at the standard addition predictions for 5 µM shown in figs. 6.8 and 6.14. The overall accuracy from the last measurements is better, with some values

being within 2 % of the actual value whereas the best from the first measurement series are within 6.5 %. The range of standard deviation values is comparable across data sets.

Figures 6.9 and 6.15 provide the results of the predictions for the 10 µM solutions. Here both data sets manage to provide at least one prediction within 2 % of the actual concentration, however the set of last measurements consistently provides better predictions than that of the first measurements. The standard deviation values do show more variance in the last measurement data set, which is mainly caused by the large standard deviation values for the first prediction of each set.

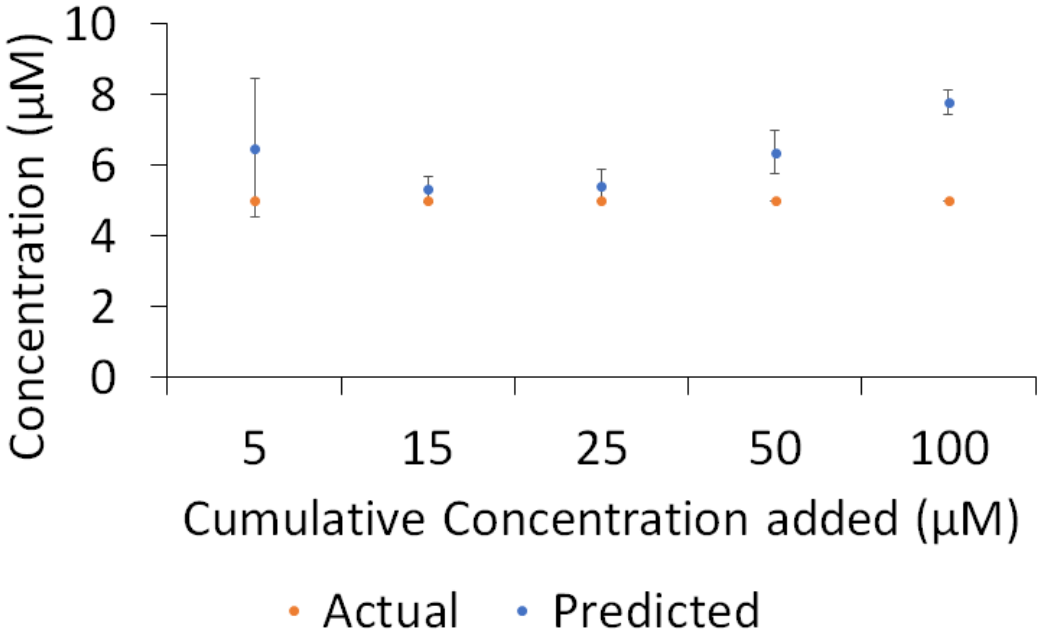


Figure 6.14: Predictions when using standard addition data with Surine solutions for a 5 µM solution using the first data recorded from the microplate reader. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

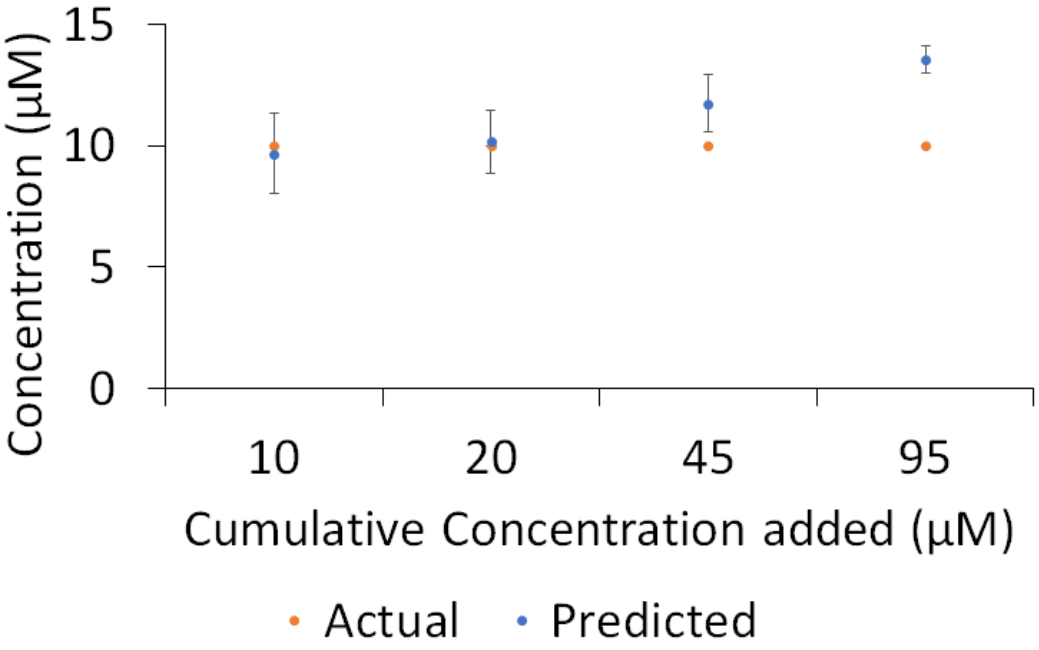


Figure 6.15: Predictions when using standard addition data with Surine solutions for a 10 µM solution using the first data recorded from the microplate reader. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

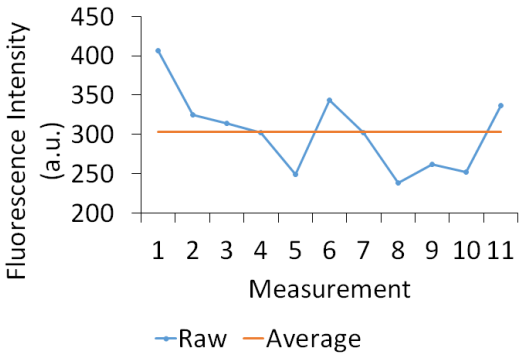
Given the results of the calibration curve which show no substantial difference between using the first or last measurements, the difference between the standard addition experiments is not great enough to justify using one over the other. Should the assay need to be performed with high throughput goals in mind, using the first measurements would really help speed up the time for data collection and maximise throughput.

6.4.2 Averaging all values

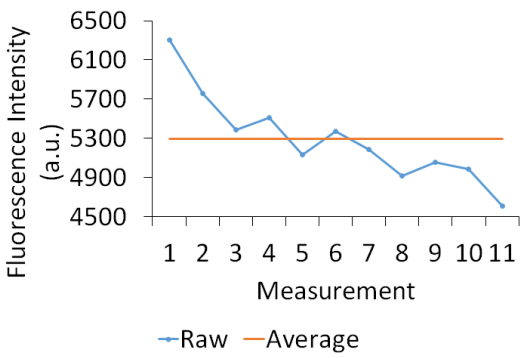
Results shown previously have had standard deviations which can range across multiple micromolar units, which when measuring the concentrations selected can mean discrepancies in the tens of percent. As has been mentioned previously, the cause of these large standard deviation values can be attributed to experimental error, such as errors in mass balances and pipetting, and also errors in measured values which is the noise of the in-

strument. A possible solution to minimise the affect of noise in the instrument would be to average the values from multiple measurements, which would eliminate the use of extreme values and eliminate the situation where a maximal measurement for a well can be paired with a minimal measurement for the next well in the standard addition, which will affect the intercept and slope values needed to determine the concentration.

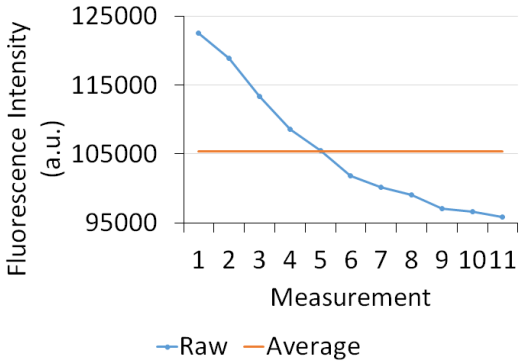
Figure 6.12 shows the intensity values recorded across the 10 minute time period for each well. For the wells with more fluorescent activity, it can clearly be seen that there is a trend towards decreasing fluorescence intensity over time. The microplate reader does not have a cooling system and so the lowest temperature setting is the ambient temperature inside the machine. It is possible for this temperature to be greater than that of the laboratory however as the machine is effectively insulated from the outside environment and also the microplates were made up in a different laboratory. Despite the apparent trend in the data, there are still spikes when looking more closely as shown in fig. 6.16.



(a) Well A1 from the microplate calibration curve containing a 1 µM kynurenine solution. The fluctuations in fluorescence can be clearly seen.



(b) Well A4 from the microplate calibration curve containing a 25 µM kynurenine solution. The fluorescence follows a downward trend but the fluctuations are still noticeable.



(c) Well A12 from the microplate calibration curve containing a 1 mM kynurenine solution. The fluorescence follows a downward trend and any fluctuations are less noticeable due to the strong downward trend.

Figure 6.16: Fluorescence intensity across time for different concentrations of kynurenine in sensor demonstrating how the intensity values fluctuate over time.

When using the averaged data to calculate concentrations using the standard addition technique the standard deviation of the results does decrease as shown in the smaller error bars in figs. 6.17 and 6.18 when compared to the previous results given in figs. 6.8, 6.9, 6.14 and 6.15. The average predicted values are not always improved however, and so averaging all cycles of measurements reduces the spread of data, as expected, but does not greatly impact the mean.

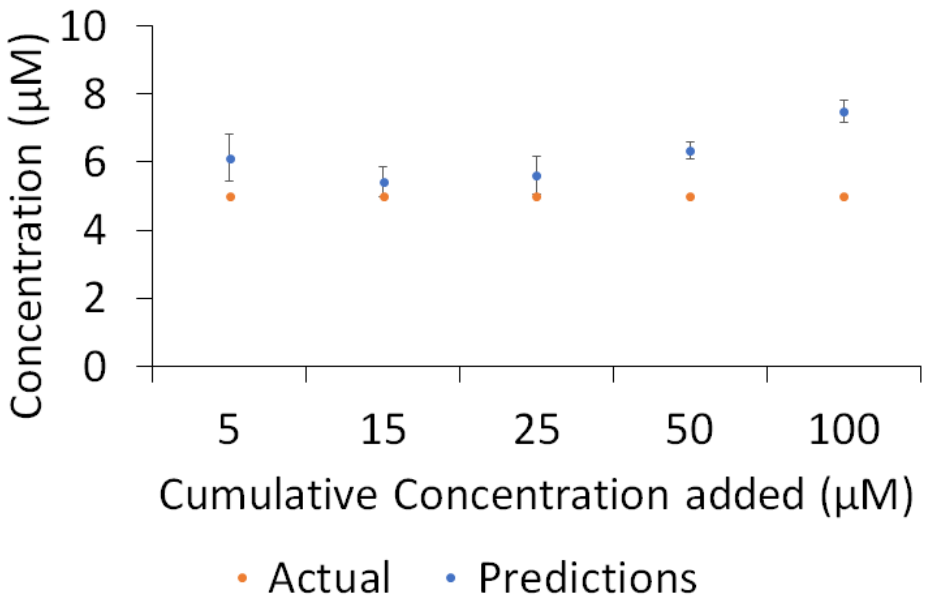


Figure 6.17: Predictions when using standard addition data with Surine solutions for a 5 µM solution using the averaged data recorded from the microplate reader. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

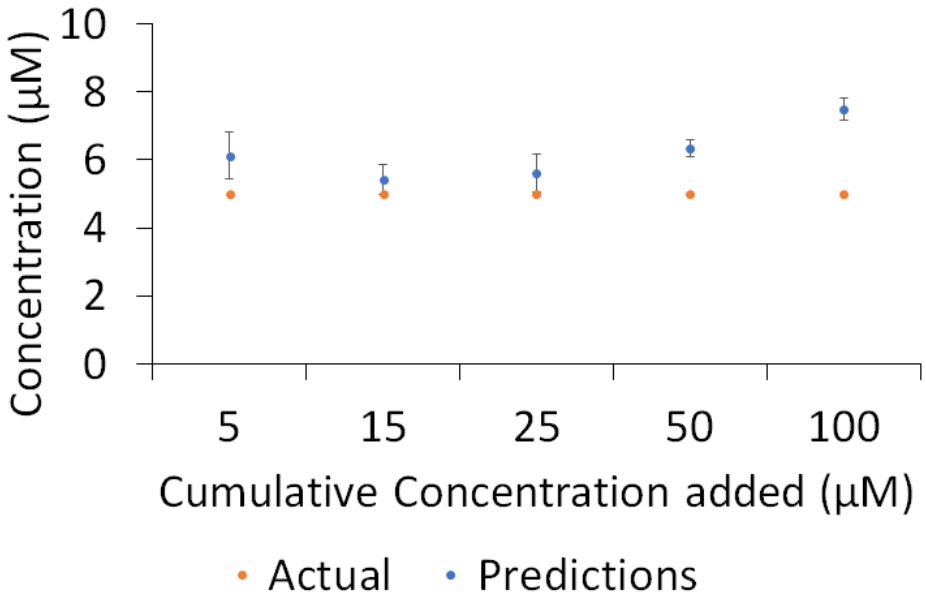


Figure 6.18: Predictions when using standard addition data with Surine solutions for a 10 µM solution using the averaged data recorded from the microplate reader. Predicted values given are the mean of three experiments with the error bars set at one standard deviation.

6.5 Quantification of kynurenine through HPLC

Kynurenine was identified within samples through mass spectrometry. This was a qualitative technique and just confirmed the presence of kynurenine. One of the steps needed to validate the fluorescence assay for biological samples is to be able to quantify the amount of analyte present in the biological sample so as to have a value to compare against.

There have been multiple procedures published on detecting kynurenine in biological samples using techniques such as HPLC, HPLC-MS and GC-MS [156–159]. For this project, HPLC was chosen for the quantification of kynurenine from biological samples with a protocol from literature found to do such a task [156].

In order to not waste clinical samples of which there are limited numbers and quantity of sample to use, HPLC protocols were tested first using pure kynurenine solutions, followed by synthetic urine solutions as with the fluorescent assay.

6.5.1 Pure kynurenine solutions

First, pure kynurenine solutions were used to check the protocol worked as expected. After some initial tests, it was determined the protocol worked, although with some differences from literature, mainly with the retention time of kynurenine. This is likely to be because the HPLC system available was slightly different to the one used from the publication, which will cause pressure differences in the system. The retention time for kynurenine remained consistent for samples run on the same day. A chromatogram for a 1 mM solution of kynurenine is provided as an example in fig. 6.19 where the peak for kynurenine is shown at 6.5 minutes.

The software for the HPLC recorded and calculated values for the time, width, height and area of peaks on the chromatogram. The next experiment determined whether a

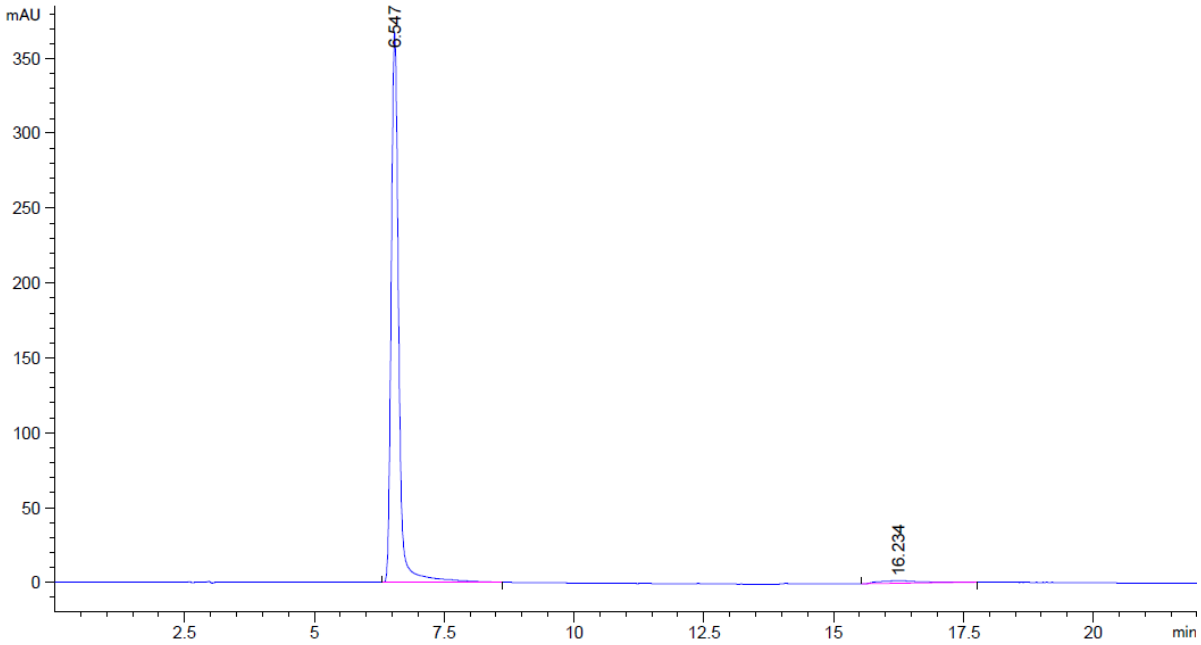


Figure 6.19: Chromatogram from HPLC of a 1 mM kynurenine solution.

calibration curve could be produced using the data the software provides. Chromatograms from different solutions of kynurenine, 5, 10, 25, 50, 100, 200, 500 and 1000 µM, were recorded. When using a 1 µM concentration of kynurenine it was not possible to detect the peak over the background fluctuations and so 5 µM was the lowest concentration to provide usable data. When making the calibration curve, the area values were first used. The area is the value of interest as this should be proportional to the concentration of the kynurenine solution as it takes into account both the height and width of the peak. The average values for the recorded areas plotted against the concentration of kynurenine solution used are shown in fig. 6.20 and the values have a linear fit with the R^2 value being 1.

Figure 6.20 shows error bars set to the standard deviation on each value, however the standard deviation values are so small for each set of measurements that the error bars cannot be seen, and so they have been included in table 6.4. Table 6.4 also includes the coefficient of variation which stays below 2 % for all values.

This result shows HPLC can be used to create a calibration curve and gives promise

in using HPLC to quantify kynurenine in more complicated media. The complication of needing to adjust the flow rate and having different retention times could be overcome by testing samples and creating calibration curves during the same time period.

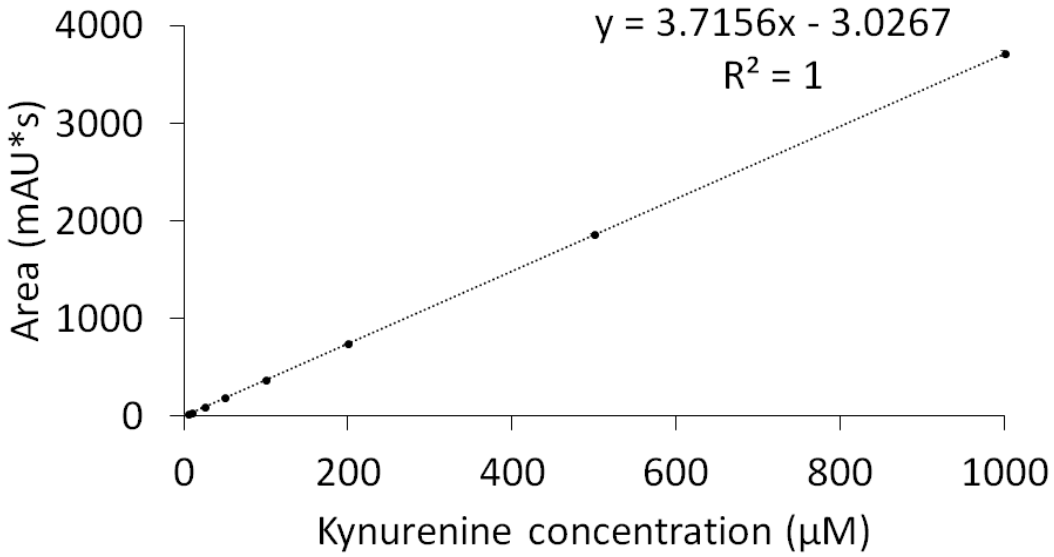


Figure 6.20: Plot of area values for absorbance from HPLC of pure kynurenine samples. Plotted values are the mean of three experiments with error bars set at one standard deviation.

Table 6.4: Mean, standard deviation and coefficient of variation values for the areas recorded.

Conc. (µM)	Mean Area	Standard Deviation	CV (%)
5	17.27	0.06	0.35
10	34.74	0.59	1.69
25	90.07	0.22	0.24
50	185.57	2.11	1.14
100	365.99	2.95	0.81
200	736.13	5.63	0.76
500	1855.28	8.03	0.43
1000	3713.20	23.16	0.62

6.5.2 Synthetic Urine

The next step involved testing the HPLC protocol with synthetic urine. As with the fluorescent assay, Surine was used for this.

Two calibration curves were produced, one with the pure kynurenine solution and one with the dosed Surine with both calibration curves being produced on the same day under the same conditions. An initial test run was performed so that the flow rate could be adjusted so as the pressure would not overload the system.

A calibration curve made from pure kynurenine solutions was created and on the same day, samples of the synthetic urine dosed with kynurenine were also processed. The concentrations of kynurenine for both sets of solutions were kept the same and were 25, 50, 100, 200, 400, 600, 800 and 1000 µM. The results from the pure kynurenine solutions were as expected from previous experiments and are shown in fig. 6.21. The results from the synthetic urine are very similar with all data points fitting a straight line giving an R^2 value of 1 as seen in fig. 6.22. It is not just the linearity which is common between both sets of solutions, as the absolute values lie close to each other as seen in fig. 6.23 where the greatest difference between any pair from the two sets of values is 5.79 %. The actual values and differences are given in table 6.5.

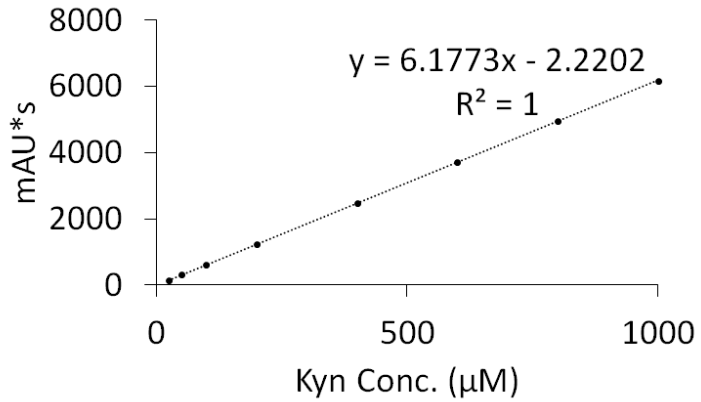


Figure 6.21: Plot of area values for absorbance from HPLC of pure kynurenine samples. Values displayed are from a single experiment, no error bars are displayed.

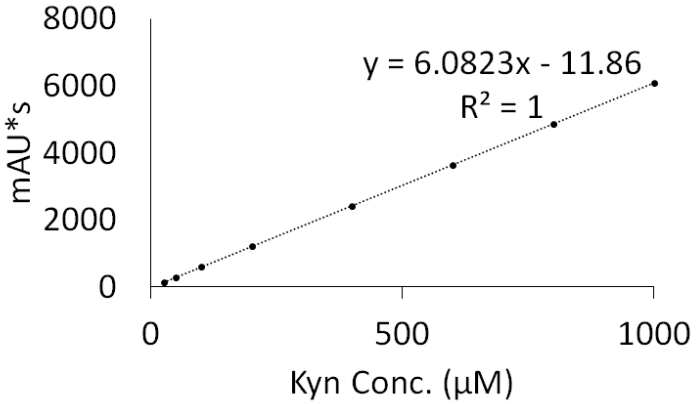


Figure 6.22: Plot of area values for absorbance from HPLC of dosed synthetic urine samples. Values displayed are from a single experiment, no error bars are displayed.

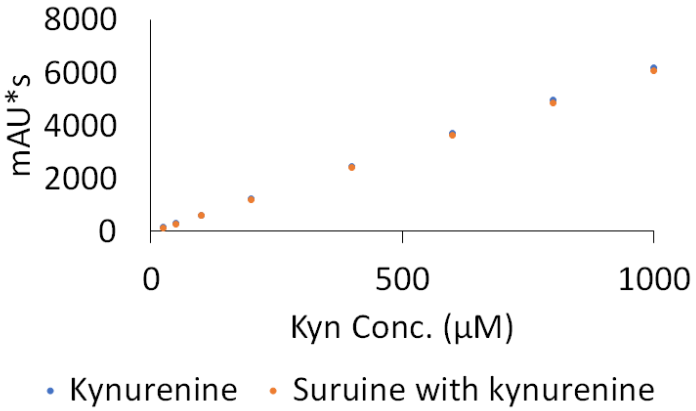


Figure 6.23: Plot of area values for absorbance from HPLC of dosed synthetic urine samples. Values displayed are from a single experiment, no error bars are displayed.

Table 6.5: Area values for kynurnine and synthetic urine solutions along with the percentage difference. All recorded values are set to 2 decimal places.

Conc. (µM)	Kyn	Surine	% Diff
25	151.23	142.46	5.79
50	306.42	294.25	3.97
100	615.02	596.73	2.97
200	1232.76	1206.69	2.11
400	2468.32	2423.48	1.82
600	3707.71	3621.31	2.33
800	4943.71	4848.70	1.92
1000	6170.04	6082.84	1.41

The calibration curves for pure solutions of kynurenine and kynurenine in synthetic urine have a strong similarity with the largest error being at the lowest concentration and this is still <6 %. As such, it should be possible to use either solution as the basis for a calibration curve. This experiment also shows that the HPLC protocol works in a more complicated media than just for pure solutions of kynurenine. The next stage for the project was to test actual urine samples using the sample HPLC protocol.

6.5.3 Urine

With HPLC able to detect and quantify the kynurenine concentration in pure solutions and synthetic urine solutions, the protocol was tested on actual urine samples. As done previously, a calibration curve was produced using pure kynurenine solutions for each set of experiments.

As was expected, the chromatograms for urine samples were more complex than any of the previous solutions.

It has been previously mentioned that the flow rate of the protocol needed to be modified even when using pure kynurenine solutions so as not to have too much pressure build up. This problem persisted with the urine samples to an even greater extent where the machine would often switch off after sample injection. In an effort to remedy this, some urine samples were filtered to try to remove the larger molecules and thus stop the excessive pressure when running the sample.

When running the filtered urines a calibration curve was made using the concentrations 20, 50, 100, 200, 400, 600, 800 and 1000 µM kynurenine. When analysing the results from urine samples, despite there being a peak for kynurenine present at the expected retention time based on the calibration solutions, the calculated concentrations ended up being negative. Figure 6.24 shows a chromatogram from one of the urine samples tested.

The peak for kynurenine is present at 10.42 minutes but it is not possible to calculate the concentration for this. This indicates that the calibration curve may be non-linear at low concentrations or that the integration of the kynurenine peaks in the urine samples is incorrect.

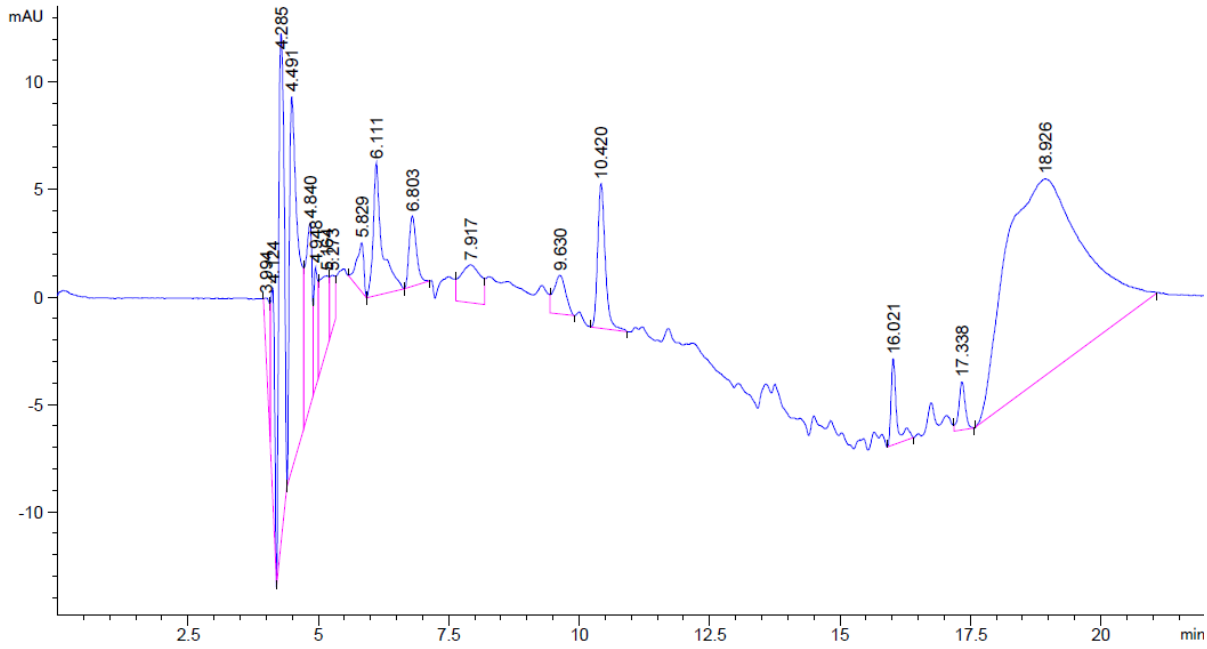


Figure 6.24: Chromatogram from a urine sample. The kynurenine peak is at 10.42 minutes, and is clearly defined. However when using the peak area value with the calibration curve a negative concentration is calculated.

To verify kynurenine was in the peaks, one urine sample was run on a semi-prep column, in order to collect the fraction, as this would allow more sample to be injected. The eluents used for these experiments caused problems when performing mass spectrometry on the collected fraction and so the fractions had to be preprocessed using the protocol given in chapter 7. When processed, mass spectrometry identified a mass which matches that of kynurenine and the spectra was similar to that of a fraction collected from a pure kynurenine solution submitted for comparison as shown in figs. 6.25 and 6.26. This provides verification that the HPLC protocol can successfully separate kynurenine from a urine sample.

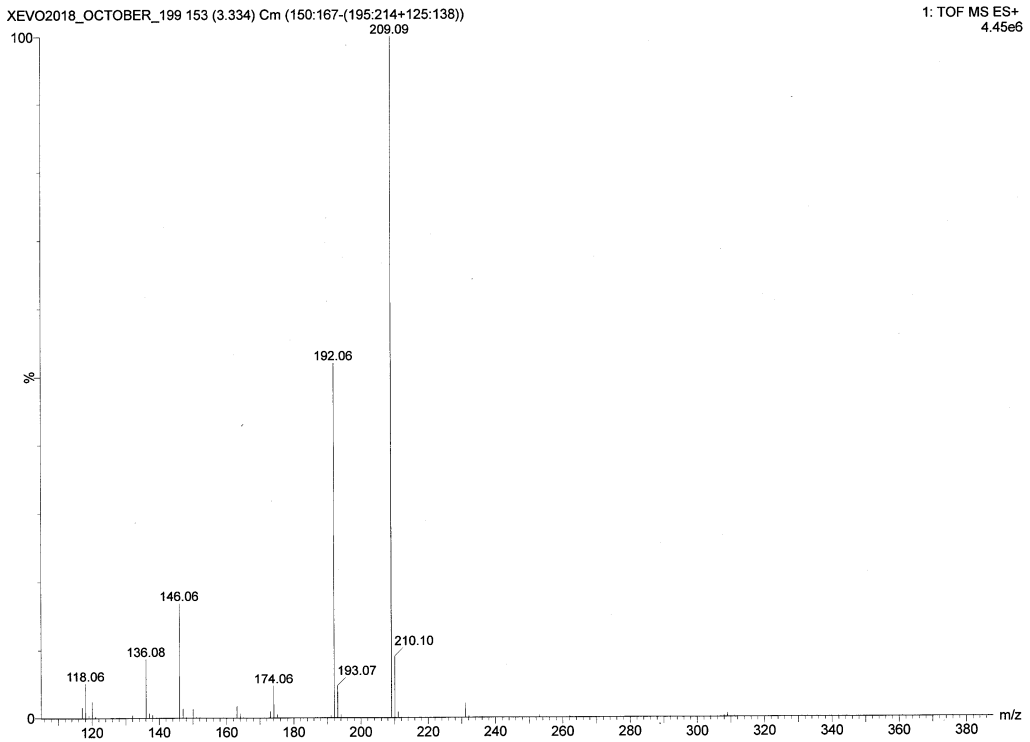


Figure 6.25: Mass spec of a fraction collected for a pure kynurenine sample using the HPLC protocol. The peak at 209.09 is for kynurenine.

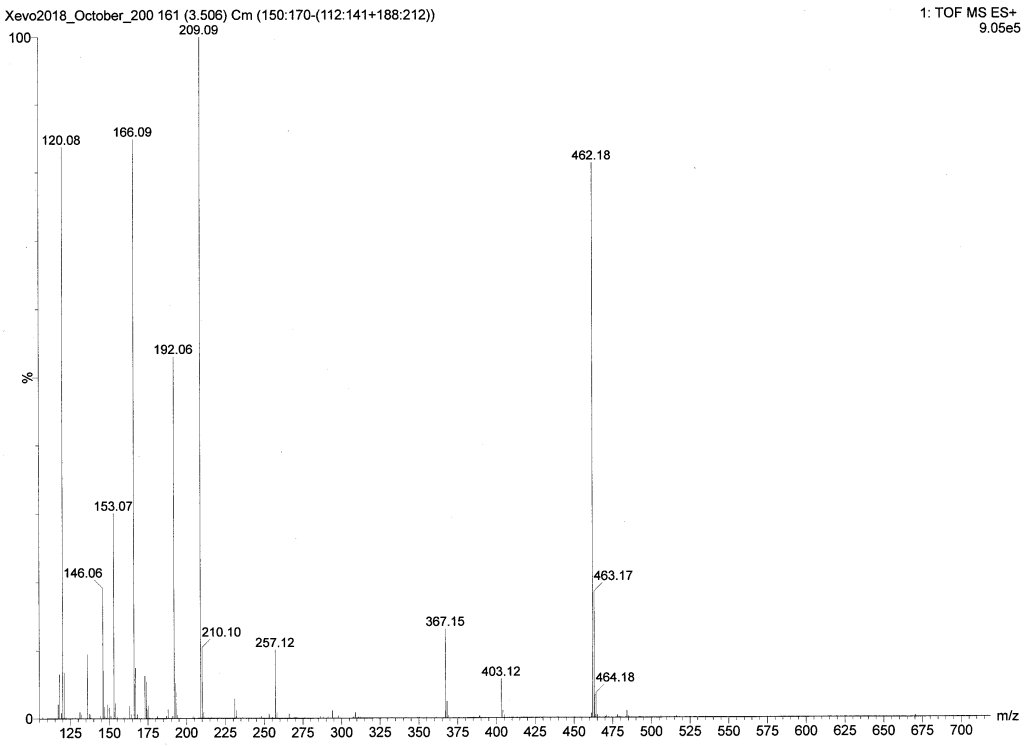


Figure 6.26: Mass spec of a fraction collected from a urine sample. The peak at 209.09 is for kynurenine.

Due to time constraints with the project it was not possible to optimise the use of the HPLC protocol to quantify the kynurenine concentrations within a urine sample and this will need to be done for future work.

6.6 Fluorescence tests with urine

Despite not being able to achieve sample quantification from HPLC, actual urine was used to see if an increasing fluorescence signal could be measured in the urine media. As mentioned previously, time constraints meant the HPLC protocol could not be optimised for urine samples, but the viability of using the fluorophore in actual urine needed to

be determined. If the fluorescence signal can be seen and the response to the standard additions is linear it would be possible to optimise the assay for urine in the future.

Two urine samples were chosen from the same patient, one from before a sepsis diagnosis and one from afterwards.

Looking at the sample from before a sepsis diagnosis, the standard addition plot is shown in fig. 6.27 and as can be seen, the response to the additions is linear. The predictions when using the standard additions are shown in fig. 6.28, where the average of the predictions when ignoring the first one is 126.26 µM ±9.2. When using the calibration curve (using solutions up to 40 µM as this gave the best R^2 value), the average predicted concentration is 127.06 µM ±14.43.

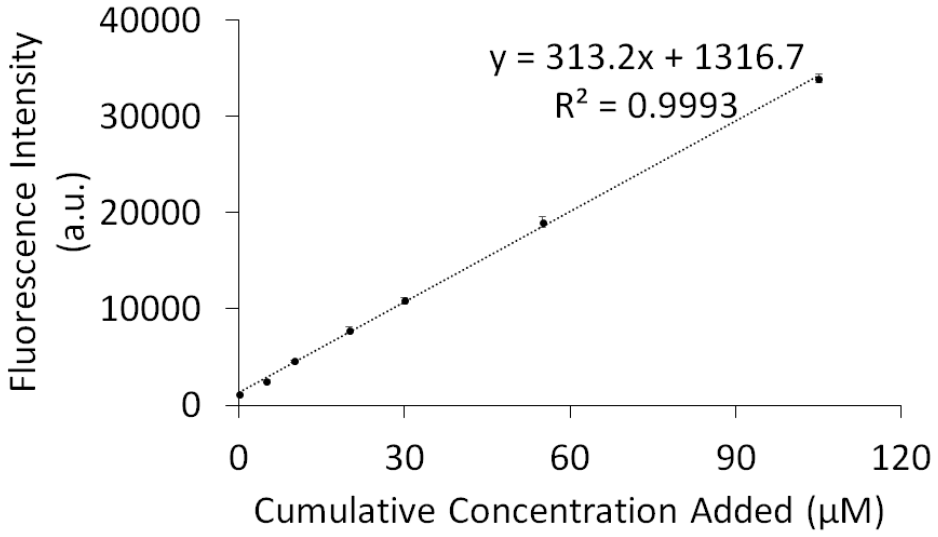


Figure 6.27: Standard addition plot of the before sepsis diagnosis urine sample. Plotted values given are the mean of three experiments with error bars set at one standard deviation.

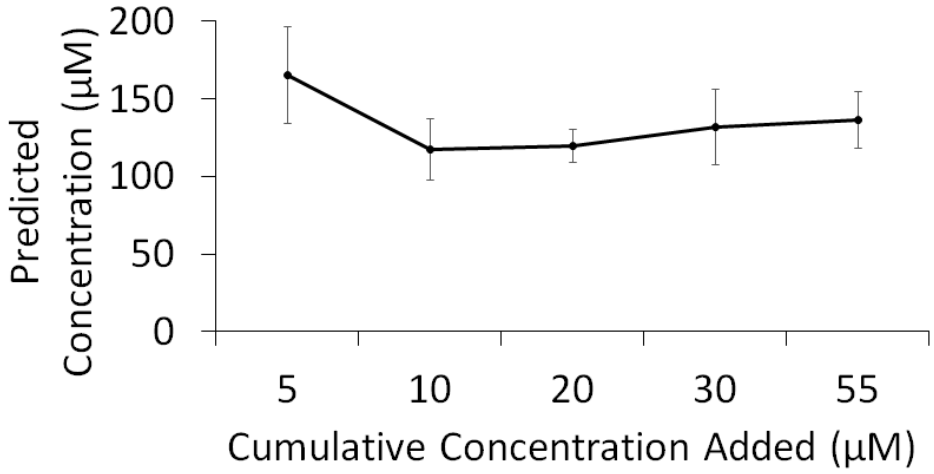


Figure 6.28: Predictions from the standard addition plot of the before sepsis diagnosis urine sample. Predicted values given are the mean of three experiments with error bars set at one standard deviation.

The standard addition plot and predictions for the after sepsis sample is given in figs. 6.29 and 6.30. The average of the predictions when using the standard additions is 597.53 µM ±43.15. The average prediction for the after sepsis samples using the calibration curve (using solutions up to 30 µM as this gave the best R^2 value) is 476.24 µM ±18.93.

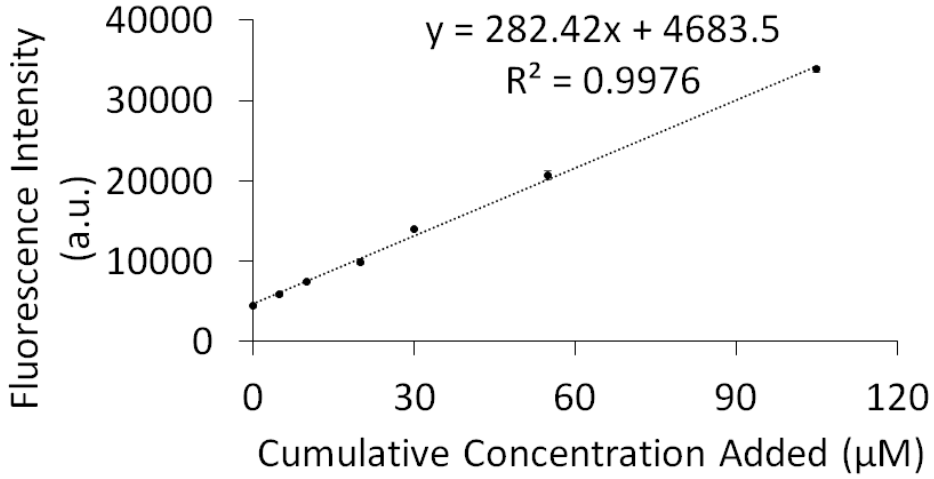


Figure 6.29: Standard addition plot of the after sepsis diagnosis urine sample. Plotted values given are the mean of three experiments with error bars set at one standard deviation.

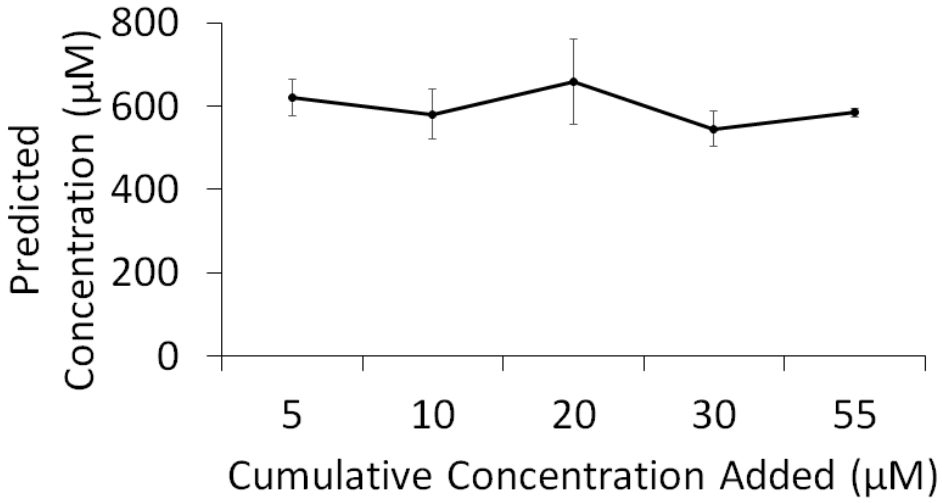


Figure 6.30: Predictions from the standard addition plot of the after sepsis diagnosis urine sample. Predicted values given are the mean of three experiments with error bars set at one standard deviation.

Whilst the before and after predictions can't be taken quantitatively, the qualitative data is interesting. The difference in values between the two urine samples is caused by greater fluorescence in the after sepsis sample, which could imply a greater kynurenine concentration. The most important result from this experiment is that the response to additions when using actual urine is still linear which provides great potential for this assay going further.

6.7 Conclusions

This chapter covered the process of designing a quantitative assay using a previously synthesised fluorescent molecule.

The work starts with showing that the fluorescent response with pure solutions of kynurenine is linear and also highlighted the importance of keeping sample conditions the same. This spurred moving to a microplate reader set-up where the standard addition assay calculated concentrations within 5 % of the actual.

Following the success with the pure solutions, synthetic urine was utilised and again a linear response was produced. The results were not as good as in pure solutions but some of the accuracy values were within 10 % of the actual, and it was shown that the linear range is best used below 100 µM kynurenine.

A HPLC protocol was tested for use in determining the kynurenine concentration in clinical samples. The protocol was tested with pure solutions of kynurenine and kynurenine in synthetic urine and worked well. Problems were encountered when moving to urine samples where it was not possible to quantify kynurenine in the samples. The chromatograms did display a peak where kynurenine was expected to be and for a urine sample, this peak was collected and determined to be kynurenine. This shows there is potential to quantify kynurenine in urine samples using HPLC, only more time is needed.

Finally, despite not having a known concentration of kynurenine, the sensor was used in actual urine samples to see the affect of the urine on the fluorescent response. The urine samples provided a linear response with additions of kynurenine which shows there is promise in being able to fully develop a fluorescence based assay for kynurenine in urine samples.

CHAPTER 7

METHODS ON ASSAY DEVELOPMENT AND VALIDATION

This chapter details the materials, methods and equipment used for the synthesis of the fluorescent chemosensor, its validation and its development into a fluorescent assay.

7.1 Chemicals and Materials

All chemicals and reagents used were purchased from Merck, Thermo Fisher, VWR International, Alfa Aesar or TCI. All water used was ultrapure water from a Milli-Q system. Disposable cuvettes used were made of either polystyrene (PS) or Poly(methyl methacrylate) (PMMA) and were purchased from Thermo Fisher. Greiner Bio-One black, flat bottom, 96 well microplates were used for fluorescent microplate experiments.

7.2 Nuclear Magnetic Resonance Spectroscopy (NMR)

^{1}H NMR spectra were recorded on a Bruker AV300 (at 300MHz) or a Bruker AVIII400 (at 400 MHz) at room temperature. ^{13}C NMR spectra were recorded on a Bruker AVIII400 (at

100 MHz) at room temperature. Data was processed using MestReNova v12 (Mestrelab). The multiplicity of signals are expressed as: s = singlet, d = doublet, t = triplet q = quartet, m = multiplet. Coupling constants (J) are reported in Hz.

7.3 Mass spectrometry (MS)

Samples were analysed by either a Synapt G25 using an atmospheric solids analysis probe or a Xevo G2XS using electrospray ionization (both from Waters Ltd) with measurements taken using time of flight analysis on both systems.

7.4 Thin Layer Chromotography (TLC)

TLC was carried out on aluminium plates coated with silica gel 60 F254 (Merck). The TLC plates were visualised using either potassium manganate or ninhydrin dip and dried with a heat gun.

7.5 Synthesis of sensor

The fluorescent sensor was produced via a 4 step synthesis. The overall synthetic procedure is shown in fig. 7.1. The procedures were taken from papers by Liu et al [150], and Klockow and Glass [149].

The following sections go into more detail on the synthesis and the specific reaction steps carried out.

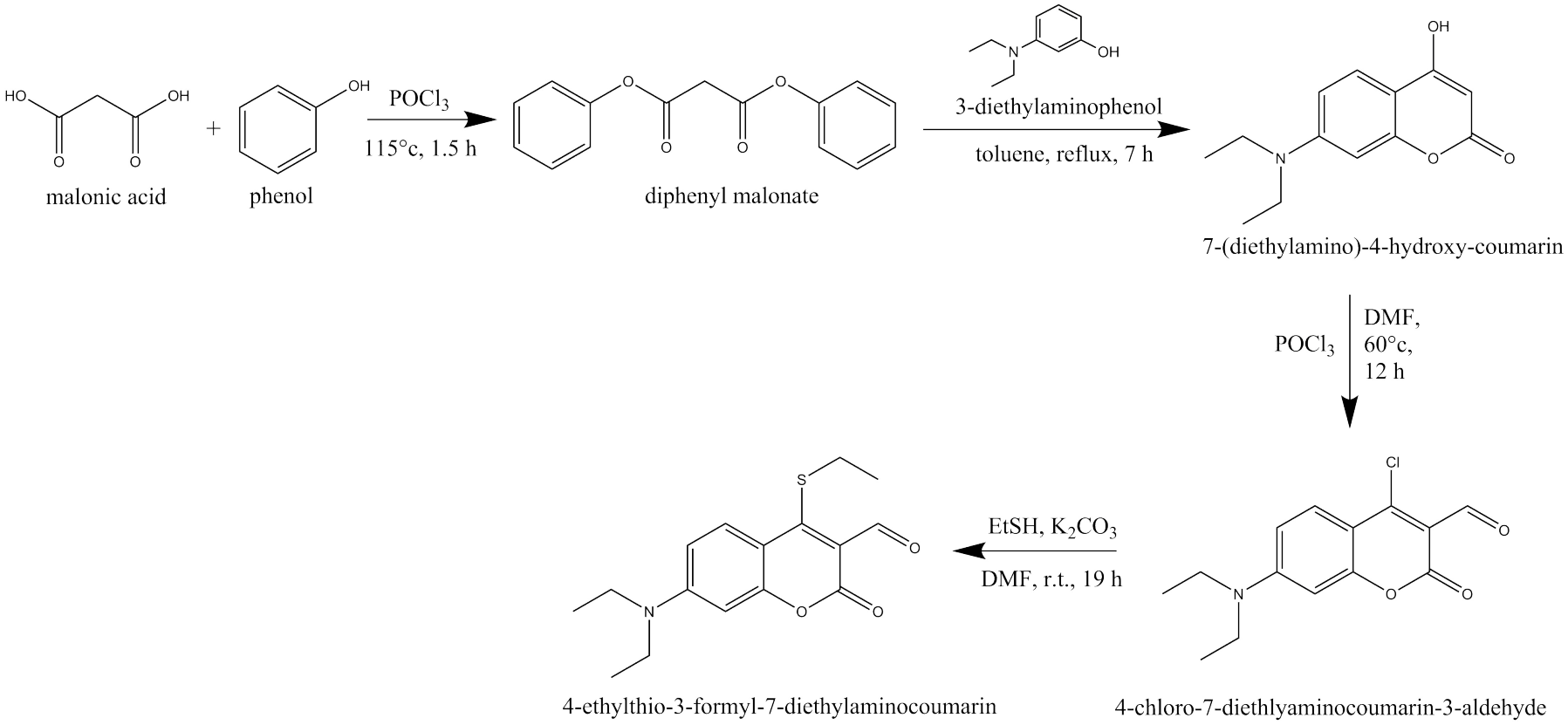


Figure 7.1: Synthetic Route for sensor

7.5.1 Synthesis of diphenyl malonate

Malonic acid (11.03 g, 106 mmol) was weighed and added to a round bottom flask (RBF). Phenol (20.08 g, 213 mmol) was weighed and added to the RBF. The RBF containing the malonic acid and phenol mixture was placed in ice to keep at 0°C and $POCl_3$ (11.5 mL, 123 mmol) was added slowly. The RBF was connected to a trap containing 1M NaOH solution. The mixture was heated at 115°C and constantly stirred until the release of HCl ceased (about 1.5 h). The upper layer was poured into 150 mL of water and three extractions with EtOAc were performed. Anhydrous magnesium sulphate was then added to the extracted solution to further remove any water, and the solution was filtered to remove the magnesium sulphate. This gave diphenyl malonate (21.3 g, 78.4 % yield) as a pale brown oil.

^{1}H-NMR (400 MHz, $CDCl_3$, 298 K) δ 7.41 (m, 4H, H_{ar}), 7.27 (m, 2H, H_{ar}), 7.16 (m, 4H, H_{ar}), 3.86 (s, 2H, CH_2).

MS (TOF AP^+): Calculated 256.07, Observed 257.16 $[M+H]^+$.

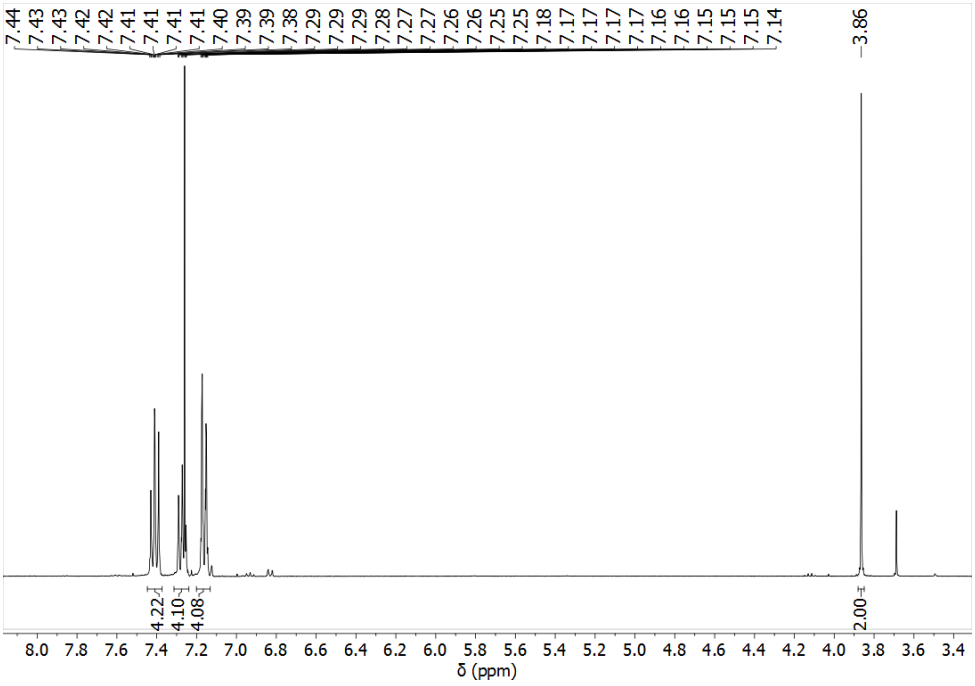


Figure 7.2: 400 MHz ^{1}H-NMR of diphenyl malonate

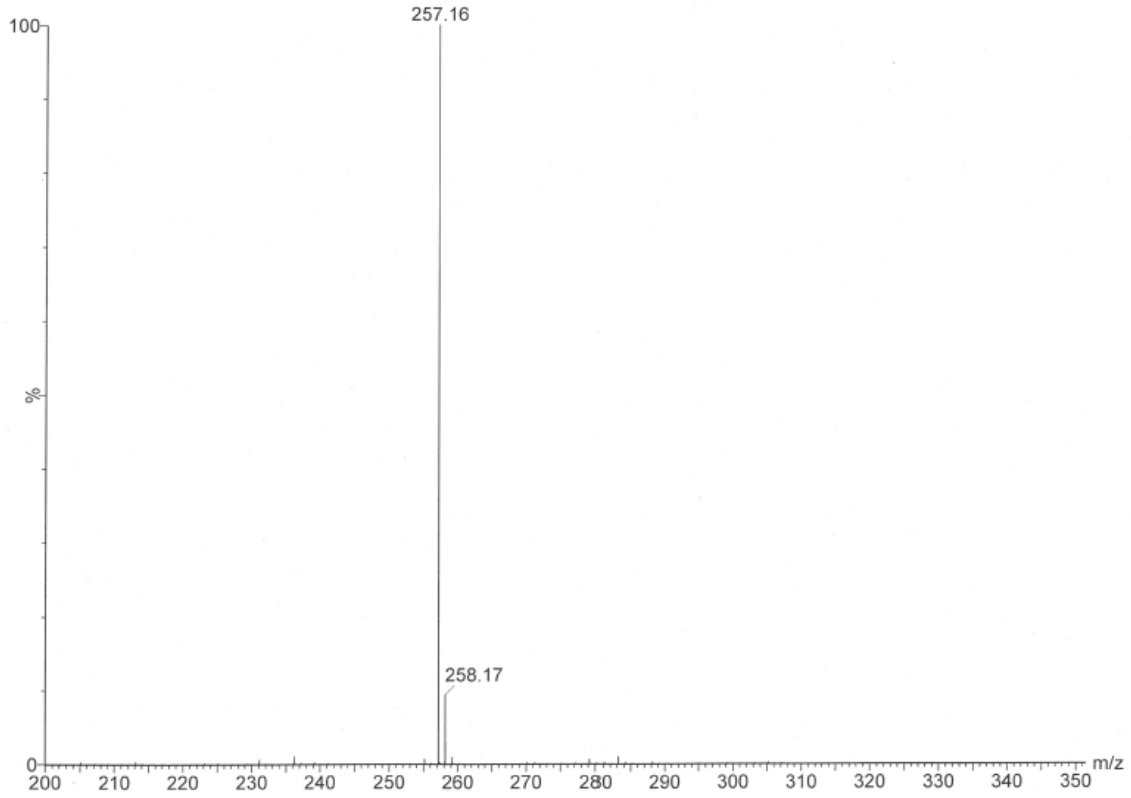


Figure 7.3: Mass spectra of diphenyl malonate. Calculated mass: 256.07, observed: 257.16 $[M+H]^+$.

7.5.2 Synthesis of 7-(diethylamino)-4-hydroxy-coumarin

Diphenyl malonate (20.5 g, 80 mmol) was dissolved into toluene (80 mL). 3-N,N-diethylaminophenol (13.2 g, 80 mmol) was added to this mixture. The mixture was refluxed for 7 hours. After the reaction had completed, the cake was filtered and washed with hexanes and then dried under vacuum. This gave 7-(diethylamino)-4-hydroxy-coumarin (6.93 g, 37.2 % yield), a light yellow solid.

^{1}H-NMR (400 MHz, DMSO-d_6, 298 K) δ 7.54 (d, J = 8.9 Hz, 1H, H_{ar}), 6.65 (dd, J = 9.0, 2.5 Hz, 1H, H_{ar}), 6.45 (d, J = 2.4 Hz, 1H, H_{ar}), 5.25 (s, 1H, CH), 3.40 (q, J = 7.2 Hz, 4H, NCH_2), 1.11 (t, J = 7.0 Hz, 6H, NCH_2CH_3).

MS (TOF ES^+): Calculated 233.17, Observed 234.1 $[M+H]^+$.

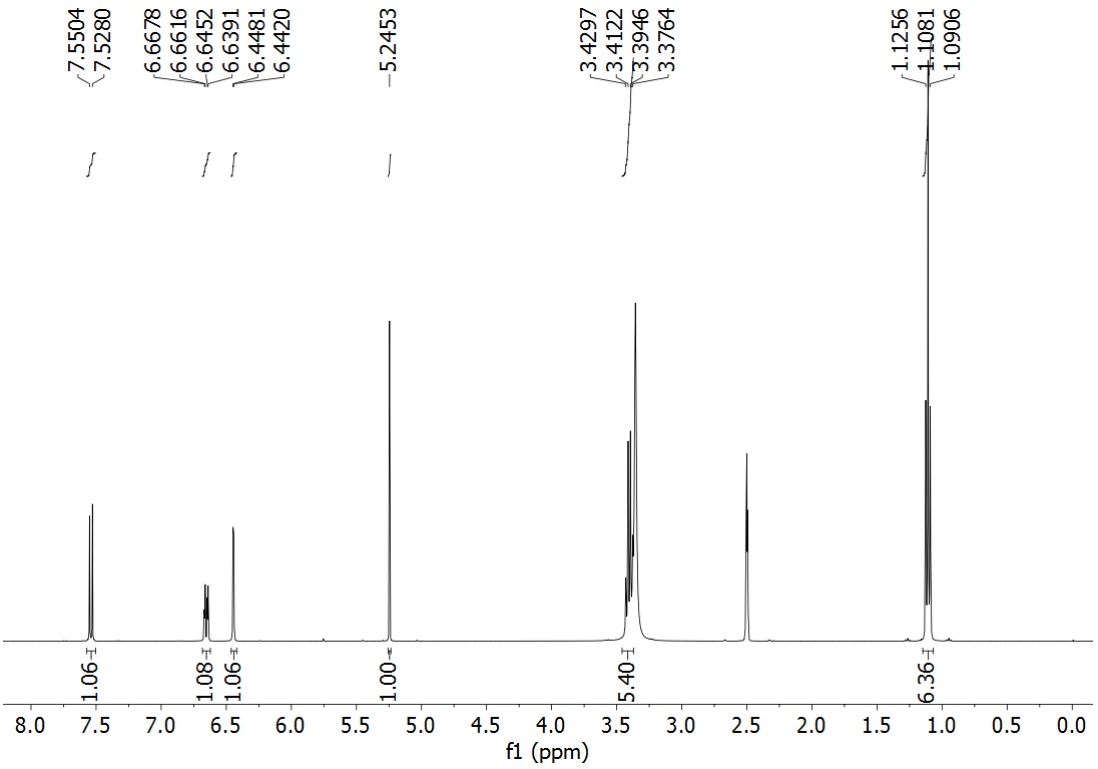


Figure 7.4: 400 MHz ^{1}H-NMR of 7-(diethylamino)-4-hydroxy-coumarin

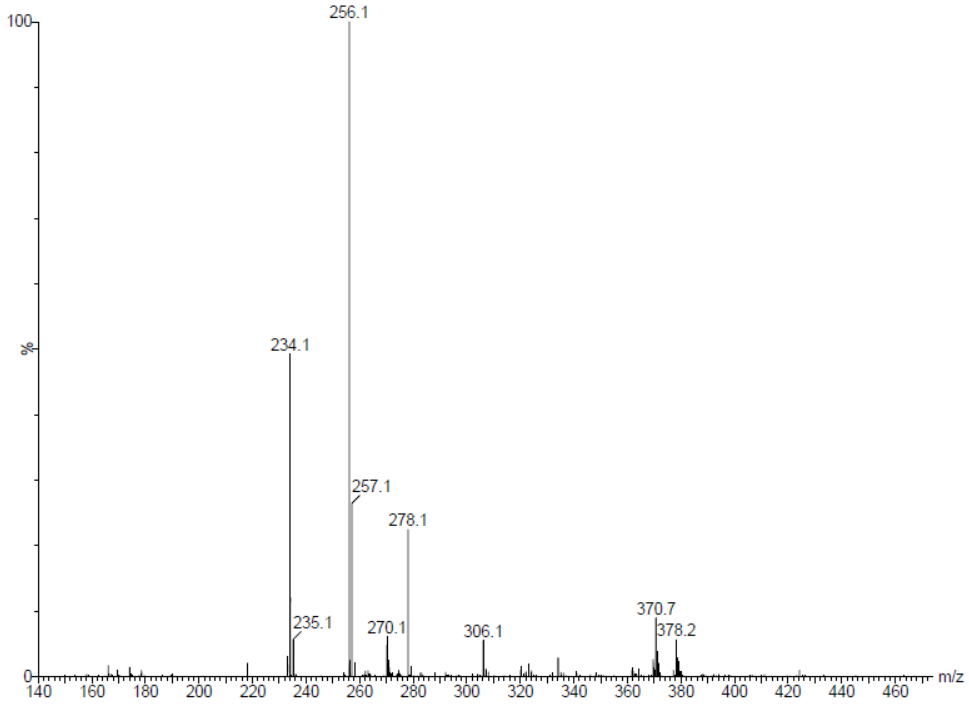


Figure 7.5: Mass spectra of 7-(diethylamino)-4-hydroxy-coumarin. Calculated 233.17, Observed 234.1 $[M+H]^{+}$.

7.5.3 Synthesis of 4-chloro-7-diethylaminocoumarin-3-aldehyde

Under argon, anhydrous DMF (2.8 mL) was added dropwise to $POCl_3$ (2.8 mL) at 20-50°C and stirred for 3 hours to yield an orange solution. 1 mL of $POCl_3$ was added to the solution after this time in case any had hydrolysed. 7-(diethylamino)-4-hydroxy-coumarin (2.33 g, 10 mmol, dissolved in 13.2 mL DMF) was added dropwise to the solution. The mixture was stirred at 60°C for 12 hours. The mixture was then poured in to 100 mL of ice water. NaOH solution (20 %) was added until the pH of the mixture was pH 14, to obtain a large amount of precipitate. The crude product was filtered, washed with water, dried and recrystallized with ethanol. This gave 4-chloro-7-diethylaminocoumarin-3-aldehyde (1.7 g, 60.9 % yield).

^{1}H-NMR (400 MHz, $CDCl_3$, 298 K) δ 10.29 (s, 1H, C*H*O), 7.83 (d, J = 9.3 Hz, 1H, H_{ar}), 6.69 (dd, J = 9.4, 2.5 Hz, 1H, H_{ar}), 6.43 (d, J = 2.5 Hz, 1H, H_{ar}), 3.48 (q, J = 7.2 Hz, 4H, NCH_2), 1.26 (t, J = 7.2 Hz, 6H, NCH_2CH_3).

MS (TOF ES$^+$): Calculated 279.07, Observed 280.1 $[M+H]^+$.

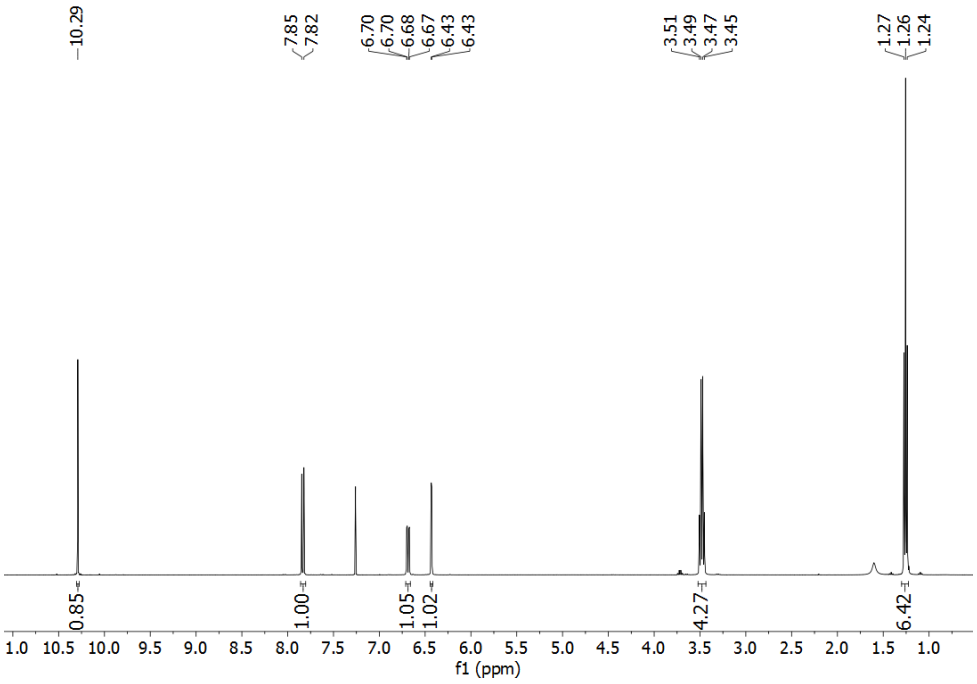


Figure 7.6: 400 MHz ^{1}H-NMR of 4-chloro-7-diethylaminocoumarin-3-aldehyde

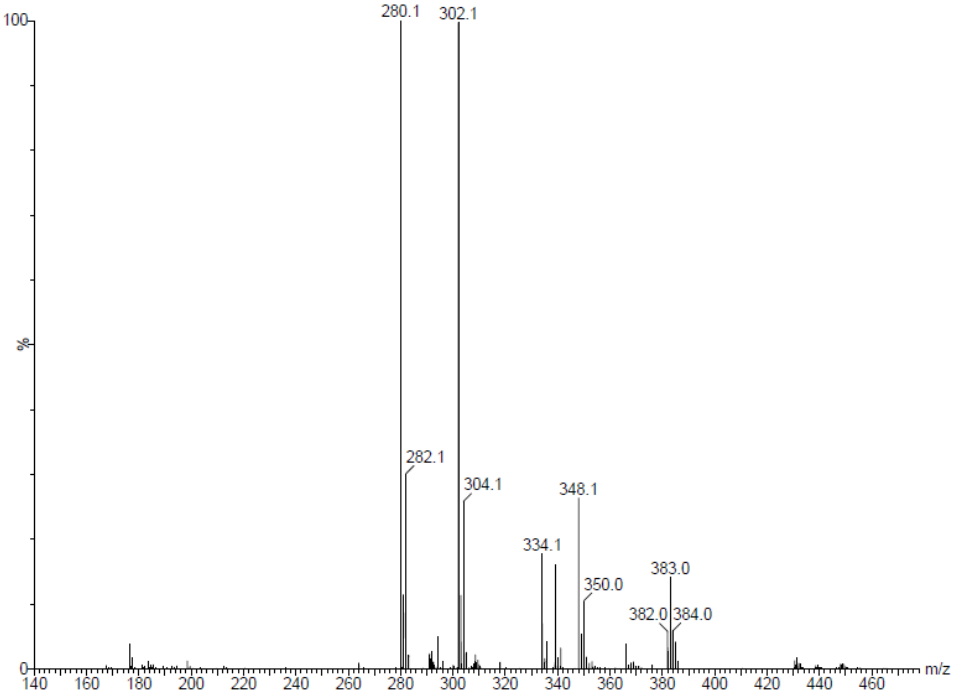


Figure 7.7: Mass spectra of 4-chloro-7-diethylaminocoumarin-3-aldehyde. Calculated 279.07, Observed 280.1 $[M+H]^+$.

7.5.4 Synthesis of 4-ethylthio-3-formyl-7-diethylaminocoumarin (Sensor)

4-chloro-7-diethylaminocoumarin-3-aldehyde (201 mg, 0.720 mmol) and K_2CO_3 (396 mg, 2.860 mmol) were measured into a flame-dried round bottom flask. A 0.1 M solution of ethanethiol in DMF (6.9 mL) was added to the starting material. The solution was stirred at room temperature for 19 h. The DMF was removed in vacuo and the remaining crude material was purified by chromatography (starting at 99:1 DCM/EtOAc and changing to 97:3 DCM/EtOAc). This gave the sensor (75.7 mg, 34.4 %) as a dark yellow powder.

^{1}H-NMR (400 MHz, $CDCl_3$, 298 K) δ 10.35 (s, 1H, C*H*O), 8.07 (d, J = 9.3 Hz, 1H, H_{ar}), 6.63 (dd, J = 9.4, 2.6 Hz, 1H, H_{ar}), 6.42 (d, J = 2.6 Hz, 1H, H_{ar}), 3.46 (q, J = 7.1 Hz, 4H, NCH_2), 3.08 (q, J = 7.4 Hz, 2H, SCH_2) 1.29 (m, 3H, SCH_2CH_3), 1.23 (m, 6H, NCH_2CH_3).

MS (TOF ES$^+$): Calculated 305.11, Observed 306.12 $[M+H]^+$

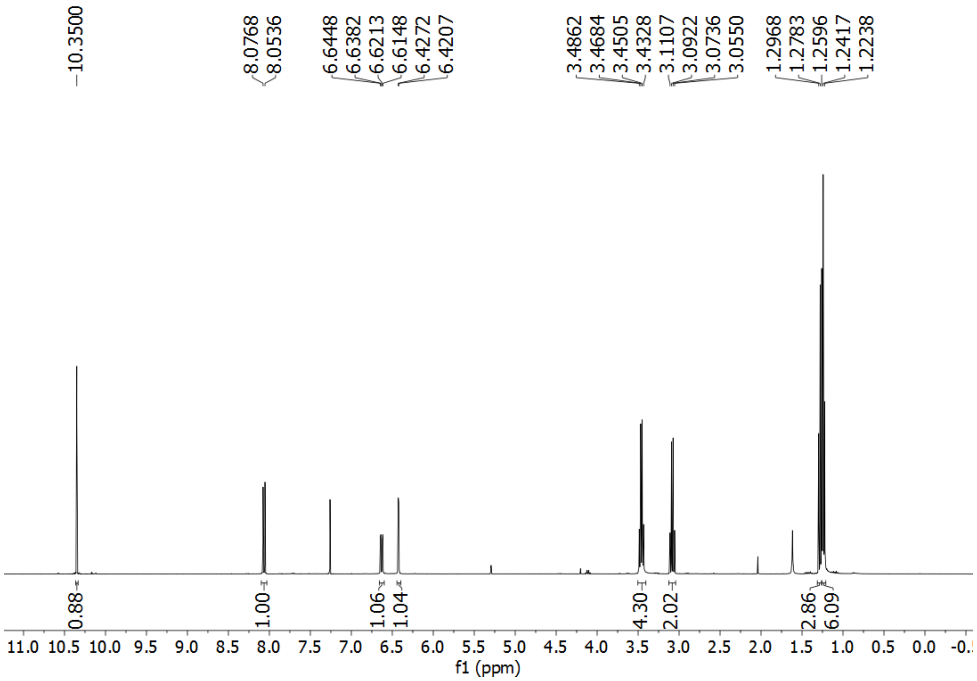


Figure 7.8: 400 MHz ^{1}H-NMR of 4-ethylthio-3-formyl-7-diethylaminocoumarin (Sensor)

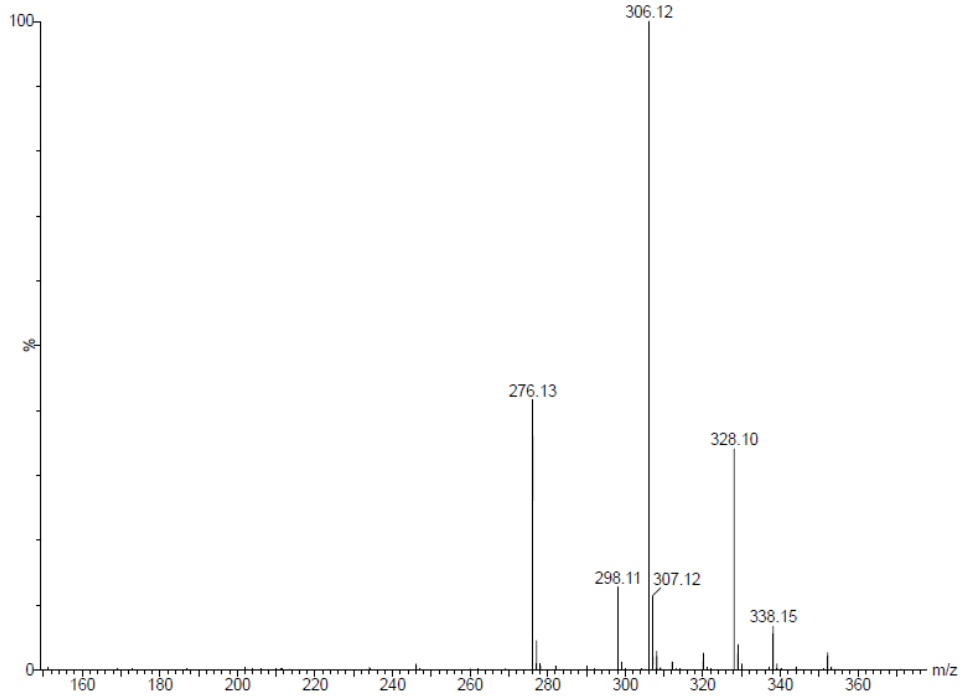


Figure 7.9: Mass spectra of 4-ethylthio-3-formyl-7-diethylaminocoumarin (Sensor). Calculated 305.11, Observed 306.12 $[M+H]^+$.

7.6 Buffer

For this work a pH close to 1 was required. In order to achieve this a buffer was made using hydrochloric acid (HCl) and potassium chloride (KCl). First 0.2 M solutions were prepared of HCl and KCl individually. 50 ml 0.2 M KCl was mixed with 134 ml 0.2 M HCl, and then the mixture was made up to 200 ml using ultrapure water. The pH of this solution was measured using a calibrated pH meter, and concentrated HCl was added dropwise to make any further adjustments to the pH.

7.7 Kynurenine and Sensor solutions

Throughout this project solutions of kynurenine and sensor were used.

Kynurenine solutions were made by weighing a desired amount of kynurenine and then dissolving into the pH 1 buffer so that the desired concentration was reached.

Sensor solutions were made by weighing a desired amount of sensor. The sensor was first dissolved in dimethyl sulfoxide (DMSO), using a volume such that the final amount of DMSO in the cuvette or microplate well does not exceed 1 %. The solution was then made to the desired concentration by adding pH 1 buffer.

7.8 Ultraviolet–visible spectroscopy (UV-Vis)

7.8.1 Technique

Ultraviolet–visible spectroscopy is a spectroscopic technique which measures the absorbance of light by a sample in the ultraviolet and visible wavelengths. There are many stages

for how this is achieved by a UV-Vis spectrophotometer. The first stage is producing the light to shine through the sample. This is normally achieved through deuterium, xenon or other lamps but recently can also be achieved by LEDs. The light produced by these light sources is typically not monochromatic, that is it will consist of a mixture of wavelengths of light. For UV-Vis spectroscopy, it is important to have light of only one wavelength entering the sample. This is so the absorbance of specific molecules can be isolated from the other molecules which may be present in the sample solution. To achieve this, monochromators or filters are used to isolate specific wavelengths of light. Monochromators have gratings which create different diffraction patterns which allow only specific wavelengths of light to pass through. The monochromatic light then passes through the sample where it can be absorbed by molecules in the sample. The light which is not absorbed passes through and is measured by a detector on the other side of the sample, in line with the incident beam.

Absorption as a process works through the interaction of light with the electrons in a molecule. Light can be viewed as quantised packets of energy, with the energy being related to the wavelength of the light. Electrons in a molecule occupy molecular orbitals which have discreet energy levels. If an electron is given a specific amount of energy, it can transition to a different molecular orbital at a higher energy level. The incident light provides the energy for this transition, and when the transition occurs, the light is absorbed by the electron and so removed from the transmitted light exiting the sample. For UV-Vis spectroscopy the transitions between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are what causes the change in the intensity of the transmitted light through the sample.

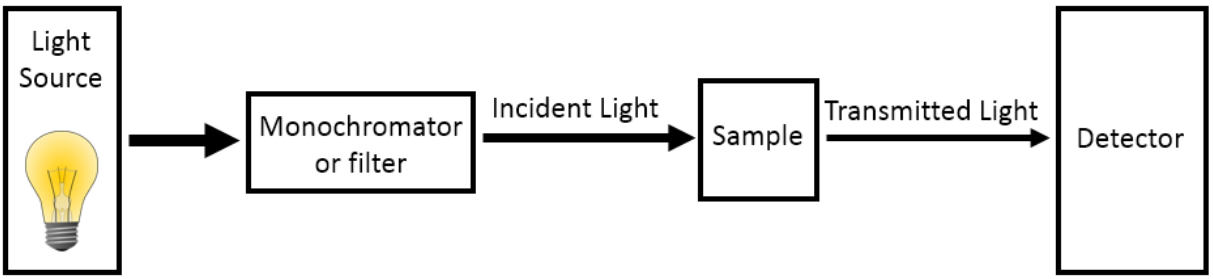


Figure 7.10: Illustration of how a UV-Vis spectrometer works. Light is first produced from the light source. This multi-wavelength light is reduced to a single wavelength via a monochromator/filter where it is then passed through the sample of interest. The light transmitted through the sample is then measured by the detector on the other side of the sample.

7.8.2 Experimental

UV-Vis experiments were carried out on either a UV-1800 (Shimadzu) or a Cary 5000 (Varian) spectrophotometer. A combination of quartz and disposable cuvettes were used.

7.9 Fluorescence Tests

Fluorescence work was carried out on a combination of a Jasco FP-8500 spectrofluorometer and a BMG Labtech CLARIOstar Plus microplate reader. Both machines use monochromators for selection of the excitation and emission wavelengths. The FP-8500 is cuvette based and was used with disposable cuvettes (PS and PMMA).

7.9.1 Fluorometer

Figure 7.11 shows the basic setup for how a fluorometer works. Light comes from a light source, passes through a filter/monochoromator and passes through the sample. Fluorescence light will be emitted in all directions, but only light perpendicular to the incident light will be collected by the detector.

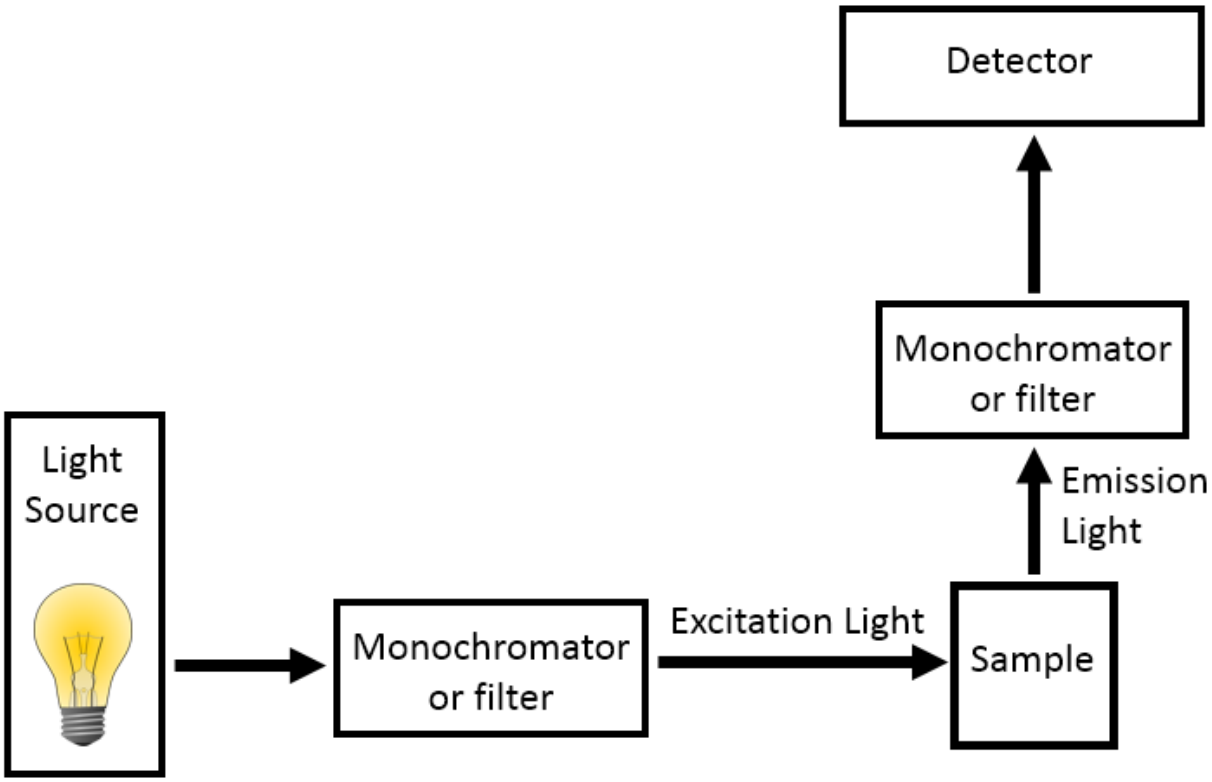


Figure 7.11: Illustration of how a fluorometer works. Multi-wavelength light is produced by the light source. This passes through a monochromator which allows only light of the set excitation wavelength to pass through to reach the sample. The fluorescence emission is measured perpendicular to the plane of excitation in order to reduce the effect of scattered light through the sample. This light then passes through another monochromator which allows only light of the set emission wavelength to pass through and reach the detector.

7.9.2 Microplate Reader

To reduce the time taken for data acquisition and also to reduce the volumes of solutions needed, a microplate reader could be used instead of a fluorometer. 96 well microplates were used, which allows for multiple experiments to be run on the same plate. An example layout for a microplate is given in fig. 7.12, which illustrates the possibility to test calibration curves with the standard additions. As the microplate is run at one time, the temperature should not vary between measurements, and so any waiting time for the temperature is only needed once, and that variance caused by slight temperature fluctuations should be minimised.

	1	2	3	4	5	6	7	8	9	10	11	12
A	1µM	5µM	10µM	15µM	20µM	25µM	30µM	40µM	50µM	100µM	200µM	500µM
B	1µM	5µM	10µM	15µM	20µM	25µM	30µM	40µM	50µM	100µM	200µM	500µM
C	1µM	5µM	10µM	15µM	20µM	25µM	30µM	40µM	50µM	100µM	200µM	500µM
D												
E												
F	Sample	x + 5 µM	x + 10 µM	x + 20 µM	x + 30 µM	x + 55 µM	x + 105 µM			Water	Buffer	Sensor
G	Sample	x + 5 µM	x + 10 µM	x + 20 µM	x + 30 µM	x + 55 µM	x + 105 µM			Water	Buffer	Sensor
H	Sample	x + 5 µM	x + 10 µM	x + 20 µM	x + 30 µM	x + 55 µM	x + 105 µM			Water	Buffer	Sensor

Calibration Curve

Blanks and Background

Standard Addition wells.
Sensor has been added to each well.

Figure 7.12: Microplate Layout

The microplate reader provides automatic focal height and gain optimisation. The focal height was allowed to be adjusted automatically for each run of each plate. The gain was set so that the well with the most expected fluorescence signal gave a fluorescence intensity of 50 % of the machines maximum range. This was done because when setting to a greater percent, fluctuations from the measurements would sometimes cause the photodiode to reach its upper limit and the measurement could no longer be used. Setting things up in this way meant that although the absolute fluorescence intensity values could not be compared between runs, the relative fluorescence was still comparable and by having a non-constant gain value, this allowed for the best dynamic range across microplates in order to provide the greatest accuracy per measurement.

7.10 Standard Addition

Standard addition is a quantitative technique used for determining the unknown concentration of an analyte when it is presented in an impure form. The technique is performed by taking a series of measurements using a method for which an increasing concentration of analyte produces a linear response. The first measurement is of the unknown sample and then several measurements are taken after additions of a known amount of analyte are added to the unknown sample. The results can then be plotted on a graph, where

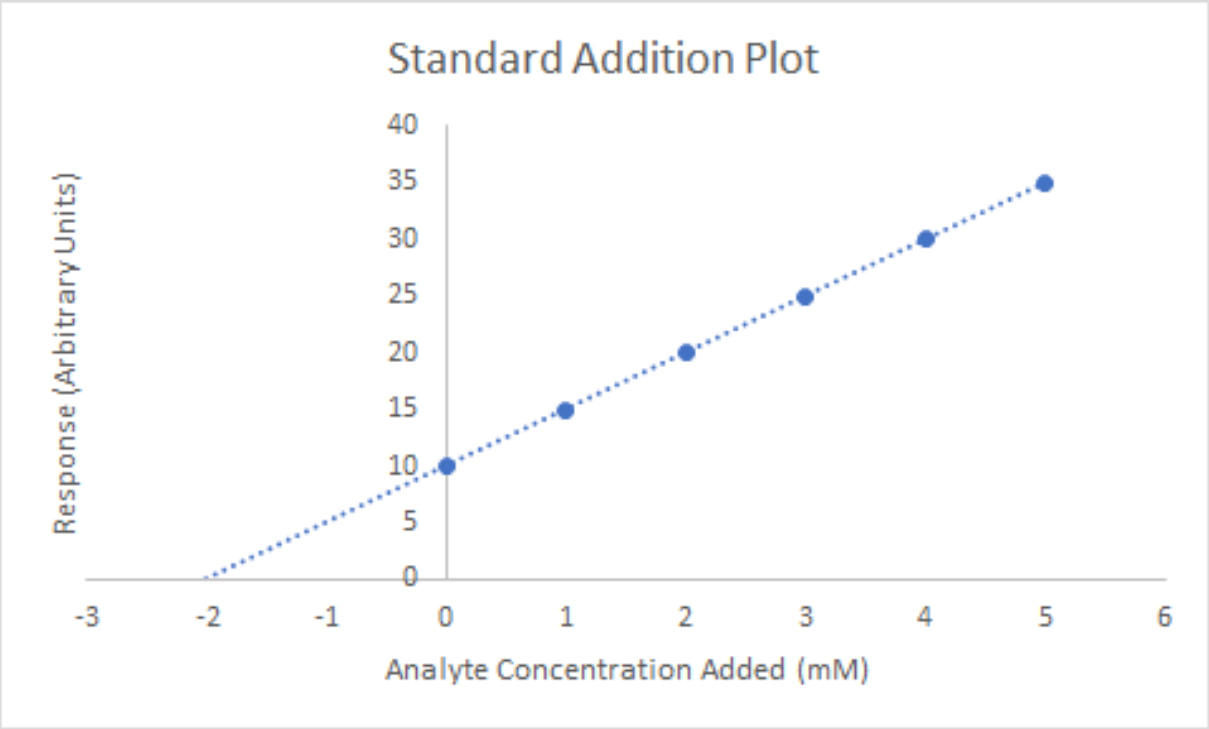


Figure 7.13: Example standard addition plot

the x-axis shows the concentration of analyte added to the sample, therefore the original sample starts at position 0 on the x-axis, and the y-axis is the response. A straight line should be fitted through all of the points, and then extrapolated backwards to find the x-intercept. The absolute value of the x-intercept can then be interpreted as the unknown concentration of analyte, and can be thought of logically as the concentration of analyte which would need to be subtracted from the sample in order to reduce the response to zero, and so therefore must be the starting concentration of analyte. The linearity of the response is important as a non-linear response is harder to accurately extrapolate to a zero response.

Figure 7.13 is an example of a standard addition plot, demonstrating extrapolation of a straight line fit to the x-axis intercept. The intercept of x = -2 here is indicative of a predicted concentration of 2 mM.

7.11 HPLC

7.11.1 Equipment

HPLC experiments were performed on an Agilent 1200 infinity series analytical/semi-prep. The system comprised of a G1311C 1260 Quat Pump VL, G1329B 1260 ALS autosampler, G1330B 1260 Thermostat, G1316C 1290 TCC, G1365D 1260 MWD VL UV-Vis detector and a G1364F 1260 FC-AS fraction collector.

A Phenomenex Kinetix C18/C18 Evo (5 µm particle size, 150 x 4.6 mm ID) column was used.

7.11.2 Eluents and gradient

For the determination of kynurenine the following eluents were used.

Eluent A: 20 mM Sodium Acetate, 30 mM Acetic Acid, 3 % Methanol, H_2O

Eluent B: 20 mM Sodium Acetate, 20 mM Acetic Acid, 10 % Methanol, 10 % Acetonitrile, H_2O

Table 7.1: HPLC elution protocol.

Time (Mins)	Eluent A (%)	Eluent B (%)
0	100	0
2	100	0
12	20	80
17	20	80
18	0	100
24	0	100
25	100	0
30	100	0

7.11.3 Fraction preprocessing

Fractions needed to be preprocessed before MS could be performed due to the use of sodium acetate.

A Phenomenex C18/ Kinetix Evo C18 column (2.6 µM particle size, 100 x 2.1 mm ID) was used for this.

Eluent A: H_2O with 0.1 % Formic Acid Eluent B: Acetonitrile with 0.1 % Formic Acid

Table 7.2: HPLC elution protocol for fraction preprocessing.

Time (Mins)	Eluent A (%)	Eluent B (%)
0	100	0
40	0	100
50	0	100
50.1	100	0
60	100	0

7.12 Urine filtering

To filter the urine samples, the samples were first thawed and then they were spun at 18200 rcf in an Eppendorf 5427R centrifuge for 10 minutes at room temperature. After this the liquid was then filtered through a 0.22 µM Pall Acrodisc syringe filter.

CHAPTER 8

DISCUSSION ON RESEARCH AIMS

8.1 Aim 1. To analyse a metabolomics data set from burns patients to identify novel biomarkers of sepsis in trauma patients.

The metabolomics data set was analysed using three different base algorithms, k-NN, NB and LR. Different imputation strategies were employed to handle missing values from the simple techniques of mean and half-min imputation to the more complicated multiple imputation.

The analysis did not identify novel biomarkers of sepsis. Instead it reaffirmed the currently known biomarkers lactate and glucose. This in itself is not a terrible result. The SSC guidelines state that lactate and glucose should be monitored and kept under control and this result reinforces that. Despite the research aim not being achieved this is still a positive result.

The stratifications of the data set showed better AUC and F_1 scores when the early samples (D1D2) were separated from the late samples (D3D4). This information could help direct future studies when it comes to grouping patient data. The stratifications

were also informative in showing that the EFT patients should not be joined with the AR or ER patient types.

It was shown that the scaled vote technique to try to manage class imbalance did not work consistently across the different stratifications and so more advanced techniques would need to be used.

The non-imputed k-NN algorithm performed the best against the two other techniques. Half-min imputation has been shown to be a poor imputation technique for -omics data and this work confirms that [84]. Half-min imputation had been proposed as the metabolite may not be recorded due to being present in concentrations lower than the detection threshold. Its poor performance compared to mean imputation implies this is not the case.

Despite being simple algorithms, on the unbalanced D1D2 data set the k-NN_NI achieved and AUC of 0.82 and specificity and sensitivity values of 85 % and 70 % respectively. Whilst not good enough to be used clinically it does show better than average discrimination between the classes. The per fold accuracy was not very good though and so more must be done to get stable results.

8.1.1 Limitations of work

Due to time constraints given to the separate parts of this project the techniques utilised for this work were very limited in their scope. The k-NN and NB algorithms were run without any feature selection. It would be useful to test these algorithms performance against using the full data sets and also reduced ones, especially the k-NN algorithms which performed the best overall.

The scaled value approach to dealing with class imbalance was very naive, and this probably explains why it worked poorly with inconsistent results.

The approach to multiple imputation was very basic with better techniques available [84, 160]. The PDFs were created using a built in algorithm which aimed to find the optimal bandwidth to fit normal distributions to. A manual approach may have been better, although this still would not have allowed the NB classifier to execute on the remaining data sets as this problem was due to lack of feature information.

The feature set contained within the data set may have limited the ability to identify novel biomarkers. Some untargeted metabolomics approaches manage to identify hundreds of biomarkers and techniques such as DNA micro arrays can generate hundreds to thousands of features whereas this data set was limited to 30 labelled metabolites.

Another limitation was in the use of only supervised techniques on this data set. Unsupervised techniques such as clustering could help uncover information hidden in the data.

8.1.2 Further work

A more robust technique for multiple imputation could have been performed instead of relying on creating PDFs for features. Techniques such as multiple imputation by chained equations have shown to be promising for data imputation [160]. It has also been shown in literature that k-NN imputation is a good strategy, however the built in MATLAB function to perform this failed to execute due to the distribution of missing values affecting every sample [84].

Different approaches to dealing with class imbalance could have been employed such as over or under sampling of data.

A more traditional approach to metabolomics analysis could have been performed, such as using OPLS-DA. This would have incorporated dimensionality reduction and could have provided visualisations of the data. This would be beneficial to provide a

comparison with the more simple machine learning techniques currently used.

The k-NN algorithm with no imputation should also be further explored. The traditional approaches to -omics analysis such as OPLS-DA require complete data sets and so rely upon imputation which depending upon the technique used, can reduce classifier output by a lot. The k-NN_NI technique could be paired with a wrapper based feature selection method and the results of this would be interesting to see how it compares with other established techniques.

Unsupervised techniques can be performed such as clustering. With the whole data set clustering analysis could verify the difference in patient types and maybe uncover even more groups.

Applications to start a second SIFTI study have been submitted and if this is the case, there is the possibility to get access to new data. This could help fulfil the research aim of identifying new biomarkers.

8.2 Aim 2. To design an analytical test to detect and quantify identified biomarkers of sepsis.

This work successfully achieves what was aimed for at the beginning of this project. All of the background work for a standard addition assay was performed which significantly aided the next part of the project.

When kynurenine was chosen to be the biomarker to continue this project, a fluorescent chemosensor was identified which selectively binds to it. There was not a lot of information with regards to the chemosensor and work was carried out in characterising it. This characterisation included determining the optimal excitation and emission wavelengths, the affect of temperature on the sensor, the stability of the sensor in acidic and neutral

solutions and also determining the linear fluorescent range of it.

UV-Vis spectroscopy was rejects as a technique for quantifying kynurenine on its own or in conjunction with the sensor due to low absorbance values. As fluorescence exhibits greater sensitivity the move to fluorescence proved to be a good choice.

The linear fluorescent range was determined to be from 1 - 200 µM and different sensor concentrations were tested to show that concentrations of 20 - 50 µM gave the best response.

By looking at the excitation and emission data for the sensor and sensor:kynurenine complex it was determined that at the chosen emission wavelength the background signal from residual sensor was minimised, but a subtraction could be performed to minimise this affect even more.

The sensor was found to not be stable in neutral or acidic solutions. Keeping it as a powder could pose a problem and so a solution to this would need to be found when developing a point-of-care test.

8.2.1 Limitations of work

The sensor works best at pH 1. This could prove to be dangerous to use unless a way to achieve similar fluorescence at a more neutral pH is achieved. With regards to moving toward a point-of-care device, this is not optimal.

When determining the linear region of the sensor, the lowest concentration of kynurenine used was 1 µM which leads itself to be the fixed point. There is a possibility that the linear region goes below 1 µM and this should be determined.

8.2.2 Further work

This work did not validate everything already provided from literature itself. The fluorescence response curve with respect to pH should be determined to determine if a less dangerous pH is still suitable.

Determine the lower bound to the linearity of the sensor by entering the nM range. Along with this limits of detection and quantification also need to be calculated to establish limits for this assay.

Binding studies by isothermal titration calorimetry could be performed to determine the binding constant of the sensor with kynurenine [161].

8.3 Aim 3. To develop and validate the analytical test to make progress towards making a point-of-care test.

The standard addition technique was performed in different media from water to synthetic urine and was able to accurately quantify kynurenine within 10 % of the actual concentration. Standard additions were performed with patient urine samples and a linear response was recovered. As such this aim for this work was mostly achieved.

The assay speed was greatly increased by moving to a microplate reader from a fluorometer which allowed more experiments to be performed. Issues of temperature and binding equilibration time were determined to not manifest when using a microplate reader given the microplate is read in one pass.

The HPLC protocol showed the ability to isolate kynurenine in solutions of water and synthetic urine, however it failed to work with actual urine which meant there was no

way to validate the assay results with urine.

8.3.1 Limitations of work

The biological range targeted by this work was calculated from serum. Only estimates of the biological levels in urine are known and these are given with respect to creatinine to account for the dilution level of the urine. As such, levels of kynurenine in urine need to be determined or the assay can move onto using serum.

The amount of urine and synthetic urine used was only 5 µL. This represents 2.5 % of the total well volume of a microplate. This could be why the fluorescent signal could be detected when using the more complicated media. The low volumes were chosen as only 1 mL per patient sample was provided and this needed to be distributed between HPLC experiments and replicates for standard addition tests. A low volume also had the benefit of allowing the buffer to maintain a pH of 1 as adding too much unbuffered volume would change the pH.

8.3.2 Further work

The quantification of kynurenine in biological sample by HPLC or quantitative MS needs to be performed so as to be able to validate the results from using the assay on biological samples.

The assay needs to be further optimised for actual urine and should be tested with greater volumes than the 5 µL currently used. The pH will need to be recorded when using different volumes of urine and tests need to be performed to see if it has an effect on the linearity and signal from the sensor.

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

9.1 Conclusions

The work presented in the thesis spans multiple disciplines, from computational science to chemistry and chemical engineering with applications towards the biological and health sciences.

Chapter 4 goes into the analysis of metabolomics data acquired at the QE hospital. The results from chapter 4 show that the k-NN technique can produce classifiers with the greatest AUC values, however despite good average values, the actual variation per fold was quite dramatic, which showed the different partitions were not representative of the whole data set. The results from feature selection identified lactate and glucose to be the most relevant metabolites in the data set. These are already recognised metabolites and so sadly new biomarkers were not identified.

Chapter 5 presents the selection of kynurenine as a biomarker of interest. It also details the experiments carried out to characterise a fluorescent chemosensor which binds with kynurenine. Optimal excitation and emission were determined, the affect of temperature, the stability of the sensor and analyte in neutral and pH 1 solutions were measured, and

the linearity of the sensor was found for different sensor concentrations. All of these were necessary for devising the standard addition assay.

Chapter 6 presents the work done to create the assay. The standard addition technique produced results with an error of up to 5 % when using pure solutions of analyte, and when using a synthetic urine solution the errors stayed near 10 %.

Work to quantify kynurenine within urine samples by HPLC was also performed. Kynurenine could be quantified in pure and synthetic urine solutions however running urine samples in the HPLC caused some trouble. Because of this, the only experiments using urine were to determine if the fluorescence signal was linear when applying standard additions to a urine sample and this was the case.

9.2 Future Work

This project has many avenues for future work. The first is from the computer science and bioinformatics perspective. Whilst the classifiers produced were not stable across different folds of the data, they did provide promising average AUC values. These were quite simple algorithms to use and so more complicated techniques should be used to further analyse the data and address the challenges of class imbalance and missing data.

The main work of this project that is ripe for continuation is in developing the assay and eventually turning it into a point of care device. The accuracy of quantification in urine, or even serum needs to be determined. To do this, work must be continued using the HPLC on urine samples, or another means of quantifying kynurenine for comparison needs to be found.

As for developing the assay into a point of care device, there have already been numerous attempts from others to produce small benchtop/handheld fluorescence detec-

tors [108–110]. As the assay only requires one excitation wavelength and one emission wavelength, large and expensive monochromator/filter wheel based systems are not necessary, more simple optical and filter systems can be used.

When performing the assay using cuvettes, it was possible to see the solution change colour, normally with large concentrations of kynurenine. There must still be a colour change at lower concentrations, it will just be more subtle, but that was using the human eye. Cameras can more accurately detect the difference between two colours. Another potential avenue for a point of care device which has been developed is the use of microfluidic paper-based analytical devices in conjunction with a smart phone to take photos and analyse the results [112].

Overall there seems to be great promise in the use of fluorescence/optical sensors as point of care devices. If this technology can be coupled with a test that can use urine, it could provide a bedside instrument which uses non-invasive patient samples and provide results almost immediately.

LIST OF REFERENCES

[1] Mervyn Singer, Clifford S. Deutschman, Christopherwarren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean Daniel Chiche, Craig M. Coopersmith, Richard S. Hotchkiss, Mitchell M. Levy, John C. Marshall, Greg S. Martin, Steven M. Opal, Gordon D. Rubenfeld, Tomvan Der Poll, Jean Louis Vincent, and Derek C. Angus. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - Journal of the American Medical Association*, 315(8):801–810, 2016.

[2] Jean-Louis Vincent, Steven M Opal, John C Marshall, and Kevin J Tracey. Sepsis definitions: time for change. *The Lancet*, 381(9868):774–775, mar 2013.

[3] James A. Russell, Joel Singer, Gordon R. Bernard, Arthur Wheeler, William Fulkerson, Leonard Hudson, Roland Schein, Warren Summer, Patrick Wright, and Keith R. Walley. Changing pattern of organ dysfunction in early human sepsis is related to mortality. *Critical Care Medicine*, 28(10):3405–3411, 2000.

[4] David G. Greenhalgh. Sepsis in the burn patient: a different problem than sepsis in the general population. *Burns & Trauma*, 5:1–10, 2017.

[5] Deirdre Church, Sameer Elsayed, Owen Reid, Brent Winston, and Robert Lindsay. Burn Wound Infections. *Clinical Microbiology Reviews*, 19(2):403–434, apr 2006.

[6] Jim O'Brien. The Cost of Sepsis, 2015.

[7] Hajime Kono, Akiko Onda, and Tamiko Yanagida. Molecular determinants of sterile inflammation. *Current Opinion in Immunology*, 26(1):147–156, 2014.

[8] Kristina E. Rudd, Sarah Charlotte Johnson, Kareha M. Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V. Colombara, Kevin S. Ikuta, Niranjan Kissoon, Simon Finfer, Carolin Fleischmann-Struzek, Flavia R. Machado, Konrad K. Reinhart, Kathryn Rowan, Christopher W. Seymour, R. Scott Watson, T. Eoin West, Fatima Marinho, Simon I. Hay, Rafael Lozano, Alan D. Lopez,

Derek C. Angus, Christopher J.L. Murray, and Mohsen Naghavi. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*, 395(10219):200–211, 2020.

[9] M. P. Francino. Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances. *Frontiers in Microbiology*, 6(JAN):1–11, 2016.

[10] Karen Bush, Patrice Courvalin, Gautam Dantas, Julian Davies, Barry Eisenstein, Pentti Huovinen, George A. Jacoby, Roy Kishony, Barry N. Kreiswirth, Elizabeth Kutter, Stephen A. Lerner, Stuart Levy, Kim Lewis, Olga Lomovskaya, Jeffrey H. Miller, Shahriar Mobashery, Laura J.V. Piddock, Steven Projan, Christopher M. Thomas, Alexander Tomasz, Paul M. Tulkens, Timothy R. Walsh, James D. Watson, Jan Witkowski, Wolfgang Witte, Gerry Wright, Pamela Yeh, and Helen I. Zgurskaya. Tackling antibiotic resistance. *Nature Reviews Microbiology*, 9(12):894–896, 2011.

[11] C Lee Ventola. The antibiotic resistance crisis: part 1: causes and threats. *P & T : a peer-reviewed journal for formulary management*, 40(4):277–83, apr 2015.

[12] Derek C. Angus and Tom van der Poll. Severe Sepsis and Septic Shock. *New England Journal of Medicine*, 369(21):2062–2063, nov 2013.

[13] Jesús Blanco, Arturo Muriel-Bombín, Víctor Sagredo, Francisco Taboada, Francisco Gandía, Luís Tamayo, Javier Collado, Ángel García-Labattut, Demetrio Carriedo, Manuel Valledor, Martín De Frutos, María-Jesús López, Ana Caballero, José Guerra, Braulio Álvarez, Agustín Mayo, and Jesús Villar. Incidence, organ dysfunction and mortality in severe sepsis: a Spanish multicentre study. *Critical Care*, 12(6):R158, 2008.

[14] Mitchell M. Levy, Mitchell P. Fink, John C. Marshall, Edward Abraham, Derek Angus, Deborah Cook, Jonathan Cohen, Steven M. Opal, Jean-Louis Vincent, and Graham Ramsay. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Medicine*, 31(4):1250–1256, apr 2003.

[15] N Schaub, R Frei, and C Müller. Addressing unmet clinical needs in the early diagnosis of sepsis. *Swiss Medical Weekly*, jul 2011.

[16] J. Soong and N. Soni. Sepsis: Recognition and treatment. *Clinical Medicine, Journal of the Royal College of Physicians of London*, 12(3):276–280, 2012.

[17] Andrew Rhodes, Laura E. Evans, Waleed Alhazzani, Mitchell M. Levy, Massimo Antonelli, Ricard Ferrer, Anand Kumar, Jonathan E. Sevransky, Charles L. Sprung, Mark E. Nunnally, Bram Rochwerg, Gordon D. Rubenfeld, Derek C. Angus, Djillali Annane, Richard J. Beale, Geoffrey J. Bellinghan, Gordon R. Bernard, Jean Daniel Chiche, Craig Coopersmith, Daniel P. De Backer, Craig J. French, Seitaro Fujishima, Herwig Gerlach, Jorge Luis Hidalgo, Steven M. Hollenberg, Alan E. Jones, Dilip R. Karnad, Ruth M. Kleinpell, Younsuk Koh, Thiago Costa Lisboa, Flavia R. Machado, John J. Marini, John C. Marshall, John E. Mazuski, Lauralyn A. McIntyre, Anthony S. McLean, Sangeeta Mehta, Rui P. Moreno, John Myburgh, Paolo Navalesi, Osamu Nishida, Tiffany M. Osborn, Anders Perner, Colleen M. Plunkett, Marco Ranieri, Christa A. Schorr, Maureen A. Seckel, Christopher W. Seymour, Lisa Shieh, Khalid A. Shukri, Steven Q. Simpson, Mervyn Singer, B. Taylor Thompson, Sean R. Townsend, Thomas Van der Poll, Jean Louis Vincent, W. Joost Wiersinga, Janice L. Zimmerman, and R. Phillip Dellinger. *Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016*, volume 43. Springer Berlin Heidelberg, 2017.

[18] Shipra Gupta, Ankit Sakhuja, Gagan Kumar, Eric McGrath, Rahul S. Nanchal, and Kianoush B. Kashani. Culture-Negative Severe Sepsis: Nationwide Trends and Outcomes. *Chest*, 150(6):1251–1259, 2016.

[19] Warwick B. Dunn, David I. Broadhurst, Helen J. Atherton, Royston Goodacre, and Julian L. Griffin. Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews*, 40(1):387–426, 2011.

[20] Michael Lämmerhofer and Wolfram Weckwerth. *Metabolomics in practice: successful strategies to generate and analyze metabolic data.* John Wiley & Sons, 2013.

[21] Aihua Zhang, Hui Sun, and Xijun Wang. Saliva Metabolomics Opens Door to Biomarker Discovery, Disease Diagnosis, and Treatment. *Applied Biochemistry and Biotechnology*, 168(6):1718–1727, nov 2012.

[22] Simon Skibsted, Manoj K. Bhasin, William C. Aird, and Nathan I. Shapiro. Bench-to-bedside review: Future novel diagnostics for sepsis - a systems biology approach. *Critical Care*, 17(5), 2013.

[23] Aihua Zhang, Hui Sun, Guangli Yan, Ping Wang, and Xijun Wang. Metabolomics for Biomarker Discovery: Moving to the Clinic. *BioMed Research International*, 2015:1–6, 2015.

[24] Onur Turkoglu, Amna Zeb, Stewart Graham, Thomas Szyperski, J. Brian Szender, Kunle Odunsi, and Ray Bahado-Singh. Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature. *Metabolomics*, 12(4):60, apr 2016.

[25] Fan Zhang, Yuanyuan Zhang, Weiwei Zhao, Kui Deng, Zhuozhong Wang, Chunyan Yang, Libing Ma, Margarita S. Openkova, Yan Hou, and Kang Li. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. *Oncotarget*, 8(21):35460–35472, may 2017.

[26] Benjamin J. Blaise, Aurélie Gouel-Chéron, Bernard Floccard, Guillaume Monneret, and Bernard Allaouchiche. Metabolic Phenotyping of Traumatized Patients Reveals a Susceptibility to Sepsis. *Analytical Chemistry*, 85(22):10850–10855, nov 2013.

[27] Xin-Ru Liu, Xing-Feng Zheng, Shi-Zhao Ji, Yong-hai Lv, De-Yi Zheng, Zhao-Fan Xia, and Wei-Dong Zhang. Metabolomic analysis of thermally injured and/or septic rats. *Burns*, 36(7):992–998, nov 2010.

[28] Alexander Kutz, Pierre Hausfater, Michael Oppert, Murat Alan, Eva Grolimund, Claire Gast, Christine Alonso, Christoph Wissmann, Christian Kuehn, Maguy Bernard, Andreas Huber, Beat Mueller, and Philipp Schuetz. Comparison between B·R·A·H·M·S PCT direct, a new sensitive point-of-care testing device for rapid quantification of procalcitonin in emergency department patients and established reference methods – a prospective multinational trial. *Clinical Chemistry and Laboratory Medicine (CCLM)*, 54(4), jan 2016.

[29] Janet M. Lord, Mark J. Midwinter, Yen Fu Chen, Antonio Belli, Karim Brohi, Elizabeth J. Kovacs, Leo Koenderman, Paul Kubes, and Richard J. Lilford. The systemic immune response to trauma: An overview of pathophysiology and treatment. *The Lancet*, 384(9952):1455–1465, 2014.

[30] Centers for Disease Control and Prevention. Occupational Injury and Illness Classification System, 2013.

[31] NHS. Burns and scalds, 2018.

[32] Lindsay Vaughn and Nicole Beckel. Severe burn injury, burn shock, and smoke inhalation injury in small animals. Part 1: Burn classification and pathophysiology. *Journal of Veterinary Emergency and Critical Care*, 22(2):179–186, 2012.

[33] Michael G. Sowa, Lorenzo Leonardi, Jeri R. Payette, Karen M. Cross, Manuel Gomez, and Joel S. Fish. Classification of burn injuries using near-infrared spectroscopy. *Journal of Biomedical Optics*, 11(5):054002, 2006.

[34] Kate Rivera. Burn accidents costing NHS £20 million per annum, show latest statistics on National Burn Awareness Day, 2018.

[35] American Burn Association. Burn Incidence Fact Sheet, 2016.

[36] I Sahin, S Ozturk, D Alhan, C Açikel, and S Isik. Cost analysis of acute burn patients treated in a burn centre: the Gulhane experience. *Annals of burns and fire disasters*, 24(1):9–13, 2011.

[37] L. Ferrero-Miliani, O. H. Nielsen, P. S. Andersen, and S. E. Girardin. Chronic inflammation: Importance of NOD2 and NALP3 in interleukin-1β generation. *Clinical and Experimental Immunology*, 147(2):227–235, 2007.

[38] Hans-joachim Anders and Liliana Schaefer. Beyond Tissue Injury — Damage-Associated Molecular Patterns , Toll-Like Receptors , and In fl ammasomes Also Drive Regeneration and Fibrosis. *Journal of the American Society of Nephrology : JASN*, pages 1–14, 2014.

[39] Jérôme Pugin. How tissue injury alarms the immune system and causes a systemic inflammatory response syndrome. *Annals of Intensive Care*, 2(1):27, 2012.

[40] Hua Shen, Daniel Kreisel, and Daniel Robert Goldstein. Processes of sterile inflammation. *Journal of Immunology*, 191(6):2857–63, 2013.

[41] Grace Y Chen and Gabriel Nuñez. Sterile inflammation: sensing and reacting to damage. *Nature reviews. Immunology*, 10(12):826–37, 2010.

[42] Qin Zhang, Mustafa Raoof, Yu Chen, Yuka Sumi, Tolga Sursal, Wolfgang Junger, Karim Brohi, Kiyoshi Itagaki, and Carl J Hauser. Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature*, 464(7285):104–7, 2010.

[43] Heinz Baumann and Jack Gauldie. The acute phase response. *Immunology Today*, 1994.

[44] Fabrizio Ceciliani, Alessia Giordano, and Valentina Spagnolo. The Systemic Reaction During Inflammation: The Acute-Phase Proteins. *Protein & Peptide Letters*, 9(3):211–223, jun 2002.

[45] M. G. Davies and P.-O. Hagen. Systemic inflammatory response syndrome. *British Journal of Surgery*, 84(7):920–935, jul 1997.

[46] Lori F. Gentile, Alex G. Cuenca, Philip A. Efron, Darwin Ang, Azra Bihorac, Bruce A. McKinley, Lyle L. Moldawer, and Frederick A. Moore. Persistent inflammation and immunosuppression. *Journal of Trauma and Acute Care Surgery*, 72(6):1491–1501, jun 2012.

[47] E E Moore, F A Moore, R J Franciose, F J Kim, W L Biffl, and A Banerjee. The postischemic gut serves as a priming bed for circulating neutrophils that provoke multiple organ failure. *The Journal of trauma*, 37(6):881–7, dec 1994.

[48] N. T. Thet, S. H. Hong, S. Marshall, M. Laabei, A. Toby, and A. Jenkins. Visible, colorimetric dissemination between pathogenic strains of Staphylococcus aureus and Pseudomonas aeruginosa using fluorescent dye containing lipid vesicles. *Biosensors and Bioelectronics*, 41(1):538–543, 2013.

[49] Jin Zhou, Thet Naing Tun, Sung ha Hong, June D. Mercer-Chalmers, Maisem Laabei, Amber E.R. Young, and A. Tobias A. Jenkins. Development of a prototype wound dressing technology which can detect and report colonization by pathogenic bacteria. *Biosensors and Bioelectronics*, 30(1):67–72, 2011.

[50] Raymond F. Chen and Jay R. Knutson. Mechanism of fluorescence concentration quenching of carboxyfluorescein in liposomes: Energy transfer to nonfluorescent dimers. *Analytical Biochemistry*, 172(1):61–77, jul 1988.

[51] N. T. Thet, D. R. Alves, J. E. Bean, S. Booth, J. Nzakizwanayo, A. E.R. Young, B. V. Jones, and A. Toby A. Jenkins. Prototype Development of the Intelligent Hydrogel Wound Dressing and Its Efficacy in the Detection of Model Pathogenic Wound Biofilms. *ACS Applied Materials and Interfaces*, 8(24):14909–14919, 2016.

[52] Caroline N Jones, Molly Moore, Laurie Dimisko, Andrew Alexander, Bryan A Hassell, H Shaw Warren, and Ronald G Tompkins. Spontaneous Neutrophil Migration Patterns during Sepsis after Major Burns. pages 1–17, 2014.

[53] Kathryn L Butler, Vijayakrishnan Ambravaneswaran, Nitin Agrawal, Maryelizabeth Bilodeau, Mehmet Toner, Ronald G Tompkins, Shawn Fagan, and Daniel Irimia.

Burn Injury Reduces Neutrophil Directional Migration Speed in Microfluidic Devices. 5(7), 2010.

[54] Felix Ellett, Julianne Jorgensen, Anika L Marand, Yuk Ming Liu, Myriam M Martinez, Vicki Sein, Kathryn L Butler, Jarone Lee, and Daniel Irimia. Diagnosis of sepsis from a drop of blood by in a microfluidic assay. *Nature Biomedical Engineering*, 2018.

[55] David F Colón, Carlos W Wanderley, Marcelo Franchin, Camila M Silva, Carlos H Hiroki, Fernanda V S Castanheira, Paula B Donate, Alexandre H Lopes, Leila C Volpon, Silvia K Kavaguti, Vanessa F Borges, Cesar A Speck-hernandez, Fernando Ramalho, Ana P Carlotti, Fabio Carmona, Jose C Alves-filho, Foo Y Liew, and Fernando Q Cunha. Neutrophil extracellular traps (NETs) exacerbate severity of infant sepsis. pages 1–13, 2019.

[56] Morven S Edwards and Carol J Baker. Bacterial infections in the neonate. In *Principles and practice of pediatric infectious diseases*, pages 549–555. Elsevier, 2018.

[57] Vandack Nobre, Stephan Harbarth, Jean-Daniel Graf, Peter Rohner, and Jérôme Pugin. Use of Procalcitonin to Shorten Antibiotic Treatment Duration in Septic Patients. *American Journal of Respiratory and Critical Care Medicine*, 177(5):498–505, mar 2008.

[58] Lila Bouadma, Charles-Edouard Luyt, Florence Tubach, Christophe Cracco, Antonio Alvarez, Carole Schwebel, Frédérique Schortgen, Sigismond Lasocki, Benoît Veber, Monique Dehoux, Maguy Bernard, Blandine Pasquet, Bernard Régnier, Christian Brun-Buisson, Jean Chastre, and Michel Wolff. Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *The Lancet*, 375(9713):463–474, feb 2010.

[59] Meichun Tan, Yunxia Lu, Hao Jiang, and Liandong Zhang. The diagnostic accuracy of procalcitonin and C-reactive protein for sepsis: A systematic review and meta-analysis. *Journal of Cellular Biochemistry*, 120(4):5852–5859, apr 2019.

[60] Alan E. Jones, James F. Fiechtl, Michael D. Brown, Jason J. Ballew, and Jeffrey A. Kline. Procalcitonin Test in the Diagnosis of Bacteremia: A Meta-analysis. *Annals of Emergency Medicine*, 50(1):34–41, jul 2007.

[61] Canan Bacli, Hülya Sungurtekin, Ercan Gürses, Uğur Sungurtekin, and Bünyamin Kaptanoğlu. Usefulness of procalcitonin for diagnosis of sepsis in the intensive care unit. *Critical Care*, 7(1):85–90, 2003.

[62] Thermo Fisher Scientific. Reference Values for PCT in Sepsis Patients.

[63] David C. Chu, Anuj B. Mehta, and Allan J. Walkey. Practice Patterns and Outcomes Associated With Procalcitonin Use in Critically Ill Patients With Sepsis. *Clinical Infectious Diseases*, 64(11):1509–1515, jun 2017.

[64] Charalampos Pierrakos and Jean Louis Vincent. Sepsis biomarkers: A review. *Critical Care*, 14(1), 2010.

[65] Uma Krishna, Suresh P. Joshi, and Mukesh Modh. An evaluation of serial blood lactate measurement as an early predictor of shock and its outcome in patients of trauma or sepsis. *Indian Journal of Critical Care Medicine*, 13(2):66–73, jun 2009.

[66] Elizabeth Morris, David McCartney, Daniel Lasserson, Ann Van Den Bruel, Rebecca Fisher, and Gail Hayward. Point-of-care lactate testing for sepsis at presentation to health care: A systematic review of patient outcomes. *British Journal of General Practice*, 67(665):e859–e870, 2017.

[67] Iman Tavassoly, Joseph Goldfarb, and Ravi Iyengar. Systems biology primer: The basic methods and approaches. *Essays in Biochemistry*, 62(4):487–500, 2018.

[68] Marc W. Kirschner. The meaning of systems biology. *Cell*, 121(4):503–504, 2005.

[69] David S Wishart, Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie-aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeroncic, Paul Stothard, Godwin Amegbey, David Block, David D Hau, James Wagner, Jessica Miniaci, Melisa Clements, Mulu Gebremedhin, Natalie Guo, Ying Zhang, Gavin E Duggan, Glen D Macinnis, Alim M Weljie, Reza Dowlatabadi, Fiona Bamforth, Derrick Clive, Russ Greiner, Liang Li, Tom Marrie, Brian D Sykes, Hans J Vogel, and Lori Querengesser. HMDB : the Human Metabolome Database. *Nucleic Acids Research*, 35:521–526, 2007.

[70] Caroline H. Johnson, Andrew D. Patterson, Jeffrey R. Idle, and Frank J. Gonzalez. Xenobiotic Metabolomics: Major Impact on the Metabolome. *Annual Review of Pharmacology and Toxicology*, 52(1):37–56, 2012.

[71] J. William Allwood and Royston Goodacre. An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis*, 21(1):33–47, 2010.

[72] Helen G. Gika, Georgios A. Theodoridis, Robert S. Plumb, and Ian D. Wilson. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis*, 87:12–25, 2014.

[73] Maria Louisa Izamis, Korkut Uygun, Nripen S. Sharma, Basak Uygun, Martin L. Yarmush, and Francois Berthiaume. Development of metabolic indicators of burn injury: Very low density lipoprotein (VLDL) and acetoacetate are highly correlated to severity of burn injury in rats. *Metabolites*, 2(3):458–478, 2012.

[74] Yong Zhang, Bin Cai, Hua Jiang, Hong Yan, Hao Yang, Jin Peng, Wenyuan Wang, Siyuan Ma, Xiuwen Wu, and Xi Peng. Use of 1H-nuclear magnetic resonance to screen a set of biomarkers for monitoring metabolic disturbances in severe burn patients. *Critical Care*, 18(4):1–10, 2014.

[75] Aarti Gautam, Peter D'Arpa, Duncan E. Donohue, Seid Muhie, Nabarun Chakraborty, Brian T. Luke, Dmitry Grapov, Erica E. Carroll, James L. Meyerhoff, Rasha Hammamieh, and Marti Jett. Acute and Chronic Plasma Metabolomic and Liver Transcriptomic Stress Effects in a Mouse Model with Features of Post-Traumatic Stress Disorder. *PLOS ONE*, 10(1):e0117092, jan 2015.

[76] Anna L Swan, Kirsty L Hillier, Julia R Smith, David Allaway, Susan Liddell, Jaume Bacardit, and Ali Mobasheri. Analysis of mass spectrometry data from the secretome of an explant model of articular cartilage exposed to pro-inflammatory and anti-inflammatory stimuli using machine learning. *BMC Musculoskeletal Disorders*, 14(1):349, dec 2013.

[77] Anna L Swan, Dov J Stekel, Charlie Hodgman, David Allaway, Mohammed H Alqahtani, Ali Mobasheri, and Jaume Bacardit. A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. *BMC Genomics*, 16(Suppl 1):S2, 2015.

[78] Frances Y Kuo and Ian H Sloan. Lifting the curse of dimensionality. *Notices of the AMS*, 52(11):1320–1328, 2005.

[79] Michel Verleysen and Damien François. The Curse of Dimensionality in Data Mining and Time Series Prediction. In Joan Cabestany, Alberto Prieto, and Francisco Sandoval, editors, *Computational Intelligence and Bioinspired Systems*, pages 758–770, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[80] Donald B Rubin and A John Wiley. *Statistical Analysis with Missing Data Second Edition.*

[81] Philip L. Roth, Fred S. Switzer, and Deborah M. Switzer. Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, 2(3):211–232, 1999.

[82] Edgar Acuña and Caroline Rodriguez. The Treatment of Missing Values and its Effect on Classifier Accuracy. *Classification, Clustering, and Data Mining Applications*, (1995):639–647, 2004.

[83] Nicholas J Horton and Ken P Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American statistician*, 61(1):79–90, feb 2007.

[84] Riccardo Di Guida, Jasper Engel, J William Allwood, Ralf J M Weber, Martin R Jones, Ulf Sommer, Mark R Viant, and Warwick B Dunn. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 12(5):93, may 2016.

[85] Sutthipong Meeyai. Logistic regression with missing data: a comparisson of handling methods and effects of percent missing values. *Journal of Traffic and Logistics Engineering*, 4(2), 2016.

[86] R. M. Parry, W. Jones, T. H. Stokes, J. H. Phan, R. A. Moffitt, H. Fang, L. Shi, A. Oberthuer, M. Fischer, W. Tong, and M. D. Wang. K-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *Pharmacogenomics Journal*, 10(4):292–309, 2010.

[87] Bao Weisheng, Ceana H. Nezhat, Gordon F. Huang, Ying Qing Mao, Neil Sidell, and Ruo Pan Huang. Discovering endometriosis biomarkers with multiplex cytokine arrays. *Clinical Proteomics*, 16(1):1–9, 2019.

[88] Ying Xie, Wei Yu Meng, Run Ze Li, Yu Wei Wang, Xin Qian, Chang Chan, Zhi Fang Yu, Xing Xing Fan, Hu Dan Pan, Chun Xie, Qi Biao Wu, Pei Yu Yan, Liang Liu, Yi Jun Tang, Xiao Jun Yao, Mei Fang Wang, and Elaine Lai Han Leung. Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational Oncology*, 14(1), 2021.

[89] Kirsty E. Waddington, Artemis Papadaki, Leda Coelewij, Marsilio Adriani, Petra Nytrova, Eva Kubala Havrdova, Anna Fogdell-Hahn, Rachel Farrell, Pierre Dönnes, Inés Pineda-Torra, and Elizabeth C. Jury. Using Serum Metabolomics to Predict Development of Anti-drug Antibodies in Multiple Sclerosis Patients Treated With IFNβ. *Frontiers in Immunology*, 11(July):1–15, 2020.

[90] Gerald J. Kost, Nam K. Tran, and Richard F. Louie. *Point-of-Care Testing: Principles, Practice, and Critical-Emergency-Disaster Medicine*. 2008.

[91] Kathleen David. Point-of-Care Versus Lab-Based Testing: Striking a Balance, 2016.

[92] Teresita Lacara, Caroline Domagtoy, Donna Lickliter, Kathy Quattrocchi, Lydia Snipes, Joánne Kuszaj, and MaryClare Prasnikar. Comparison of Point-of-Care and Laboratory Glucose Analysis in Critically Ill Patients. *American Journal of Critical Care*, 16(4):336–346, jul 2007.

[93] Anjannette Cook, Delyla Laughlin, Margery Moore, Doreen North, Kathleen Wilkins, Gay Wong, Allyson Wallace-Scroggs, and Lisa Halvorsen. Differences in Glucose Values Obtained From Point-of-Care Glucose Meters and Laboratory Analysis in Critically Ill Patients. *American Journal of Critical Care*, 18(1):65–72, jan 2009.

[94] Adil I. Khan, Yolanda Vasquez, Jacquelyn Gray, Frank H. Wians, and Martin H. Kroll. The variability of results between point-of-care testing glucose meters and the central laboratory analyzer. *Archives of Pathology and Laboratory Medicine*, 130(10):1527–1532, 2006.

[95] Michael Juliano and Courtney Wason. Comparison of Point-of-Care Versus Laboratory Troponin Testing in an Emergency Department Setting. *Military Medicine*, 182(7):e1938–e1940, jul 2017.

[96] Denis Frasca, Claire Dahyot-Fizelier, Karen Catherine, Quentin Levrat, Bertrand Debaene, and Olivier Mimoz. Accuracy of a continuous noninvasive hemoglobin monitor in intensive care unit patients*. *Critical Care Medicine*, 39(10):2277–2282, oct 2011.

[97] Guijian Guan, Bianhua Liu, Zhenyang Wang, and Zhongping Zhang. Imprinting of Molecular Recognition Sites on Nanostructures and Its Applications in Chemosensors. *Sensors*, 8(12):8291–8320, dec 2008.

[98] Di Wu, Adam C. Sedgwick, Thorfinnur Gunnlaugsson, Engin U. Akkaya, Juyoung Yoon, and Tony D. James. Fluorescent chemosensors: The past, present and future. *Chemical Society Reviews*, 46(23):7105–7123, 2017.

[99] Markus Haase and Helmut Schäfer. Upconverting Nanoparticles Angewandte. pages 5808–5829, 2011.

[100] Anthony W Czarnik. *Fluorescent chemosensors for ion and molecule recognition*. ACS Publications, 1993.

[101] Huarui He, Mark A Mortellaro, Marc J P Leiner, Susanne T Young, Robert J Fraatz, James K Tusa, Roche Diagnostic Corporation, Hembree Park Drive, Roche Diagnostic Gmbh, and A Graz. A Fluorescent Chemosensor for Sodium Based on Photoinduced Electron Transfer. 75(3):549–555, 2003.

[102] Wen-sheng Xia, Russell H Schmehl, and Chao-jun Li. Synthesis and Study of A Molecular Fluorescent Chemosensor For Potassium. pages 387–389, 2000.

[103] Huarui He, Kenneth Jenkins, and Chao Lin. A fluorescent chemosensor for calcium with excellent storage stability in water. 1:197–204, 2008.

[104] Engin U Akkaya and Serhan Turkyilmaz. A squaraine-based near IR fluorescent chemosensor for Calcium. 38(25):4513–4516, 1997.

[105] James K Tusa and Huarui He. Critical care analyzer with fluorescent optical chemosensors for blood analytes companies involved in medical. 2005.

[106] Ye He, Zhiqian Guo, Pengwei Jin, Changhong Jiao, He Tian, and Weihong Zhu. Optimizing the Chemical Recognition Process of a Fluorescent Chemosensor for α-Ketoglutarate. *Industrial & Engineering Chemistry Research*, 54(11):2886–2893, mar 2015.

[107] Zhi-Gang Wang, Xiao-Jing Ding, Yu-Ying Huang, Xiao-Jing Yan, Bin Ding, Qing-Zhong Li, Cheng-Zhi Xie, and Jing-Yuan Xu. The development of coumarin Schiff base system applied as highly selective fluorescent/colorimetric probes for Cu2+ and tumor biomarker glutathione detection. *Dyes and Pigments*, 175:108156, apr 2020.

[108] Young Ho Shin, Jonathan Z. Barnett, M. Teresa Gutierrez-Wing, Kelly A. Rusch, and Jin Woo Choi. A hand-held fluorescent sensor platform for selectively estimating green algae and cyanobacteria biomass. *Sensors and Actuators, B: Chemical*, 262:938–946, 2018.

[109] P. Kozma, A. Lehmann, K. Wunderlich, D. Michel, S. Schumacher, E. Ehrentreich-Förster, and F. F. Bier. A novel handheld fluorescent microarray reader for point-of-care diagnostic. *Biosensors and Bioelectronics*, 47:415–420, 2013.

[110] Uwadiae Obahiagbon, Joseph T. Smith, Meilin Zhu, Benjamin A. Katchman, Hany Arafa, Karen S. Anderson, and Jennifer M. Blain Christen. A compact, low-cost, quantitative and multiplexed fluorescence detection platform for point-of-care applications. *Biosensors and Bioelectronics*, 117(March):153–160, 2018.

[111] Liying Zhang, Zhenhua Tian, Hunter Bachman, Peiran Zhang, and Tony Jun Huang. A Cell-Phone-Based Acoustofluidic Platform for Quantitative Point-of-Care Testing. *ACS Nano*, 14(3):3159–3169, mar 2020.

[112] Zijie Luo, Taoyuze Lv, Kangning Zhu, Yi Li, Lei Wang, J. Justin Gooding, Guozhen Liu, and Bin Liu. Paper-Based Ratiometric Fluorescence Analytical Devices towards Point-of-Care Testing of Human Serum Albumin. *Angewandte Chemie - International Edition*, 59(8):3131–3136, 2020.

[113] Rui-Qiao Zhang, Shu-Lin Liu, Wei Zhao, Wan-Po Zhang, Xu Yu, Yong Li, An-Jun Li, Dai-Wen Pang, and Zhi-Ling Zhang. A Simple Point-of-Care Microfluidic Immunomagnetic Fluorescence Assay for Pathogens. *Analytical Chemistry*, 85(5):2645–2651, mar 2013.

[114] M. J. Baxter. A review of supervised and unsupervised pattern recogniton in archaeometry. *Archaeometry*, 48(4):671–694, 2006.

[115] S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatica (Ljubljana)*, 31(3):249–268, 2007.

[116] Wei Liu and Sanjay Chawla. Class Confidence Weighted k NN Algorithms for. *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science*, pages 345–356, 2011.

[117] Harshit Dubey and Vikram Pudi. Class based weighted K-Nearest neighbor over imbalance dataset. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7819 LNAI(PART 2):305–316, 2013.

[118] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[119] Nick Roussopoulos, Stephen Kelley, and Frédéric Vincent. Nearest neighbor queries. In *Proceedings of the 1995 ACM SIGMOD international conference on Management of data - SIGMOD '95*, pages 71–79, New York, New York, USA, 1995. ACM Press.

[120] Q Peter He and Jin Wang. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *Semiconductor Manufacturing, IEEE Transactions on*, 20(4):345–354, 2007.

[121] José M. Jerez, Ignacio Molina, Pedro J. García-Laencina, Emilio Alba, Nuria Ribelles, Miguel Martín, and Leonardo Franco. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, 50(2):105–115, 2010.

[122] Benjamin Marlin. *Missing Data Problems in Machine Learning*. PhD thesis, University of Toronto, 2008.

[123] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, jan 1996.

[124] Paul A. Murtaugh. In defense of P values. *Ecology*, 95(3):611–617, mar 2014.

[125] Jianping Hua, Waibhav D. Tembe, and Edward R. Dougherty. Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition*, 42(3):409–424, 2009.

[126] C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. *Proceedings of the 2003 IEEE Bioinformatics Conference, CSB 2003*, 3(2):523–528, 2003.

[127] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[128] Suzanne Ekelund. Precision-recall curves – what are they and how are they used?, 2017.

[129] Christopher Michael Francis Wearn. *THE METABOLOMIC RESPONSE TO SEVERE THERMAL INJURY AND THE IMPACT OF AGE*. PhD thesis, University of Birmingham, 2017.

[130] Marc G. Jeschke, David Patsouris, Mile Stanojcic, Abdikarim Abdullahi, Sarah Rehou, Ruxandra Pinto, Peter Chen, Marjorie Burnett, and Saeid Amini-Nik. Pathophysiologic Response to Burns in the Elderly. *EBioMedicine*, 2(10):1536–1548, 2015.

[131] R. L. PRENTICE and R. PYKE. Logistic disease incidence models and case-control studies. *Biometrika*, 66(3):403–411, 1979.

[132] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer, 2 edition, 2017.

[133] M. Adeva-Andany, M. López-Ojén, R. Funcasta-Calderón, E. Ameneiros-Rodríguez, C. Donapetry-García, M. Vila-Altesor, and J. Rodríguez-Seijas. Comprehensive review on lactate metabolism in human health. *Mitochondrion*, 17:76–100, 2014.

[134] F. B. Cerra, J. Caprioli, J. H. Siegel, R. R. McMenamy, and J. R. Border. Proline metabolism in sepsis, cirrhosis and general surgery. The peripheral energy deficit. *Annals of Surgery*, 190(5):577–586, 1979.

[135] Gilles Troché, Matthieu Henry-Lagarrigue, Frédérique Soppelsa, Stephane Legriel, Aihem Yehia, Fabrice Bruneel, Jean Pierre Bédos, and Odile Spreux-Varoquaux. Tryptophan pathway catabolites (serotonin, 5-hydroxyindolacetic acid, kynurenine) and enzymes (monoamine oxidase and indole amine 2,3 dioxygenase) in patients with septic shock: A prospective observational study versus healthy controls. *Medicine*, 99(19):e19906, 2020.

[136] Nathalie Le Floc'h, Winfried Otten, and Elodie Merlot. Tryptophan metabolism, from nutrition to potential therapeutic applications. *Amino Acids*, 41(5):1195–1205, 2011.

[137] Martin Ploder, Andreas Spittler, Katharina Schroecksnadel, Gabriele Neurauter, Linda E. Pelinka, Erich Roth, and Dietmar Fuchs. Tryptophan degradation in multiple trauma patients: Survivors compared with non-survivors. *Clinical Science*, 116(7):593–598, 2009.

[138] Tim T. Lögters, Maurice D. Laryea, Jens Altrichter, Janina Sokolowski, Jindrich Cinatl, Jenny Reipen, Wolfgang Linhart, Joachim Windolf, Martin Scholz, and Michael Wild. Increased plasma kynurenine values and kynurenine-tryptophan ratios after major trauma are early indicators for the development of sepsis. *Shock*, 32(1):29–34, 2009.

[139] Katharina Pellegrin, Gabriele Neurauter, Barbara Wirleitner, Arthur W. Fleming, Verlyn M. Peterson, and Dietmar Fuchs. Enhanced enzymatic degradation of tryptophan by indoleamine 2,3-dioxygenase contributes to the tryptophan-deficient state seen after major trauma. *Shock*, 23(3):209–215, 2005.

[140] J. P. Zeden, G. Fusch, B. Holtfreter, J. C. Schefold, P. Reinke, G. Domanska, J. P. Haas, M. Gruendling, A. Westerholt, and C. Schuett. Excessive tryptophan catabolism along the kynurenine pathway precedes ongoing sepsis in critically ill patients. *Anaesthesia and Intensive Care*, 38(2):307–316, 2010.

[141] Yutang Wang, Hanzhong Liu, Gavin McKenzie, Paul K. Witting, Johannes Peter Stasch, Michael Hahn, Dechaboon Changsirivathanathamrong, Ben J. Wu, Helen J. Ball, Shane R. Thomas, Vimal Kapoor, David S. Celermajer, Andrew L. Mellor, John F. Keaney, Nicholas H. Hunt, and Roland Stocker. Kynurenine is an endothelium-derived relaxing factor produced during inflammation. *Nature Medicine*, 16(3):279–285, 2010.

[142] Dechaboon Changsirivathanathamrong, Yutang Wang, Dorrilyn Rajbhandari, Ghassan J. Maghzal, Wendy M. Mak, Clive Woolfe, Johan Duflou, Val Gebski, Cris G. Dos Remedios, David S. Celermajer, and Roland Stocker. Tryptophan metabolism to kynurenine is a potential novel contributor to hypotension in human sepsis. *Critical Care Medicine*, 39(12):2678–2683, 2011.

[143] Christabelle J. Darcy, Joshua S. Davis, Tonia Woodberry, Yvette R. McNeil, Dianne P. Stephens, Tsin W. Yeo, and Nicholas M. Anstey. An observational cohort study of the kynurenine to tryptophan ratio in sepsis: Association with impaired immune and microvascular function. *PLoS ONE*, 6(6), 2011.

[144] Jason P. Eiserich, Keri A. Hayakawa, and Carroll E. Cross. Sepsis and hypotension: Enter kynurenine, move over nitric oxide*. *Critical Care Medicine*, 39(12):2767–2769, dec 2011.

[145] Arturo Villegas-Navarro, Antonio Morales-Aguilera, and Artemisa Posada-Retana. Comparison of UV and fluorescence spectrophotometry for the quantification of a potent myotonia inducer: Anthracene-9-carboxylic acid, in plasma, urine, and saline perfusion fluids. *Journal of Pharmaceutical Sciences*, 72(10):1221–1223, 1983.

[146] Yasutsugu FUKUNAGA, Yasuhiro KATSURAGI, Takashi IZUMI, and Fumio SAKIYAMA. Fluorescence Characteristics of Kynurenine and N'-Formylkynurenine, Their Use as Reporters of the Environment of Tryptophan 62 in Hen Egg-White Lysozyme1. *The Journal of Biochemistry*, 92(1):129–141, 1982.

[147] Jorge E. Churchich. l-kynurenine: A fluorescent probe of serum albumins. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 285(1):91–98, nov 1972.

[148] Motomasa Atsumi, Ken-ichi Mawatari, Akari Morooka, Makoto Yasuda, Tomoko Fukuuchi, Noriko Yamaoka, Kiyoko Kaneko, Kazuya Nakagomi, and Naoto Oku. Simultaneous Determination of Kynurenine and Kynurenic Acid by High-Performance Liquid Chromatography Photoirradiation System Using a Mobile Phase Containing 18-crown-6. *International Journal of Tryptophan Research*, 12:117864691983455, jan 2019.

[149] Jessica L. Klockow and Timothy E. Glass. Development of a fluorescent chemosensor for the detection of kynurenine. *Organic Letters*, 15(2):235–237, 2013.

[150] Jing Liu, Yuan-Qiang Sun, Yingying Huo, Hongxing Zhang, Linfang Wang, Pei Zhang, Dan Song, Yawei Shi, and Wei Guo. Simultaneous Fluorescence Sensing of Cys and GSH from Different Emission Channels. *Journal of the American Chemical Society*, 136(2):574–577, jan 2014.

[151] Colin D. Bain, E. Barry Troughton, Yu Tai Tao, Joseph Evall, George M. Whitesides, and Ralph G. Nuzzo. Formation of Monolayer Films by the Spontaneous Assembly of Organic Thiols from Solution onto Gold. *Journal of the American Chemical Society*, 111(1):321–335, 1989.

[152] Philippa Mitchell, Stefano Tommasone, Stefano Angioletti-Uberti, James Bowen, and Paula M. Mendes. Precise Generation of Selective Surface-Confined Glycoprotein Recognition Sites. *ACS Applied Bio Materials*, 2(6):2617–2623, 2019.

[153] E. J. Bowen and J. Sahu. The effect of temperature on fluorescence of solutions. *Journal of Physical Chemistry*, 63(1):4–7, 1959.

[154] Ann E. Oliver, Gary A. Baker, Robert D. Fugate, Fern Tablin, and John H. Crowe. Effects of temperature on calcium-sensitive fluorescent probes. *Biophysical Journal*, 78(4):2116–2126, 2000.

[155] Basavaraja Jana, S. R. Inamdar, and Suresh Kumar Suresh. Effect of quencher and temperature on fluorescence intensity of laser dyes: DETC and C504T. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 170:124–130, 2017.

[156] Jianxing Zhao, Hong Chen, Peihua Ni, Bingxin Xu, Xuemei Luo, Yiming Zhan, Pingjin Gao, and Dingliang Zhu. Simultaneous determination of urinary tryptophan, tryptophan-related metabolites and creatinine by high performance liquid chromatography with ultraviolet and fluorimetric detection. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 879(26):2720–2725, 2011.

[157] Jianxing Zhao. Simultaneous determination of plasma creatinine, uric acid, kynurenine and tryptophan by high-performance liquid chromatography: Method validation and in application to the assessment of renal function. *Biomedical Chromatography*, 29(3):410–415, 2015.

[158] Haihua Lu, Jing Yu, Jun Wang, Linlin Wu, Hang Xiao, and Rong Gao. Simultaneous quantification of neuroactive dopamine serotonin and kynurenine pathway metabolites in gender-specific youth urine by ultra performance liquid chromatography tandem high resolution mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, 122:42–51, 2016.

[159] James A. Eckstein, Gina M. Ammerman, Jessica M. Reveles, and Bradley L. Ackermann. Simultaneous profiling of multiple neurochemical pathways from a single cerebrospinal fluid sample using GC/MS/MS with electron capture detection. *Journal of Mass Spectrometry*, 43(6):782–790, jun 2008.

[160] Melissa J. Azur, Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49, mar 2011.

[161] Alice Di Pasquale, Stefano Tommasone, Lili Xu, Jing Ma, and Paula M. Mendes. Cooperative Multipoint Recognition of Sialic Acid by Benzoboroxole-Based Receptors Bearing Cationic Hydrogen-Bond Donors. *Journal of Organic Chemistry*, 85(13):8330–8338, 2020.