# Design, Development, and Validation of an End-User Photo-Thermal Sensing Platform for Rapid Detection and Quantification of Analytes in Fluidic Samples

Derek Hayden

A thesis submitted to the facility of graduate studies in partial fulfillment of the requirements for the degree of Master of Applied Science

Graduate program in Mechanical Engineering

York University

Toronto, Ontario

December 2024

# Abstract

The ability to detect the presence of specific analytes and quantify their titers in fluidic samples is essential in many industries, spanning from food industries to law enforcement to healthcare and beyond. The existing technologies used for this purpose require the use of specialty equipment by trained professionals in a laboratory setting to function (e.g., mass spectrometry or ELISA) which greatly increases the cost and time taken to receive actionable results. Portable and inexpensive tests exist – Lateral Flow Immunoassays – however these tests are only qualitative and frequently have an inferior limit of detection. To date, several sensing devices have been designed to interrogate these LFIAs and decrease their limit of detection, however, these devices are often prohibitively expensive. This thesis outlines attempts to design and validate a sensing platform which could inexpensively enhance the limit of detection of LFIAs. The prototype is then validated through both lab-based and human experiments.

# Dedication

For my parents, who guided me to where I am today.

For Patricia, who has been by my side through it all.

For Carl, who is on his own journey now.

Thank you all.

# Acknowledgements

I would like to thank Professor Tabatabaei for his continued assistance and patience over the course of this thesis.

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms/Symbols

AuNP – Gold Nanoparticles

ELISA – Enzyme-linked immunosorbent assay

FD-PTR – Frequency-Domain Photothermal Radiometry

FTIR – Fourier Transform Infrared Spectrometry

IR – Infrared Radiation

LFIA – Lateral Flow Immunoassay

LOD – Limit-Of-Detection

NETD – Noise Equivalent Temperature Difference

PBS – Phosphate Buffered Saline

PCB – Printed Circuit Board

PCR - Polymerase Chain Reaction

PTR – Photothermal Radiometry

SERS – Surface Enhanced Ramen Spectroscopy

THC – Tetrahydrocannabinol

# Chapter 1: Introduction

## 1.1 Background

The need for accurate detection, and if possible – quantification - of analyte(s) in fluidic samples is abundant in a broad spectrum of industries and/or research fields. Food and water safety requires that water and agricultural samples be tested for contaminants such as heavy metals [1-3] or other contaminants like pathogens [4,5]. Modern agricultural practices also require these tests to properly identify products which have been genetically modified [6]. Medical point-of-care also make use of testing to screen or diagnose at the bedside [7,8], including screening for COVID-19 [9]. Endocrine health monitoring is also an industry which makes use of this sort of testing, including the most common and widely-used: pregnancy tests [10].

There are several tools and methods which are used to achieve these desired detections. The first method, largely considered to be the 'gold-standard' is mass spectroscopy. There are a number of mass spectrometers, each with their own intricacies, but the basic working principle is that a portion of the sample is ionized, those ionized particles are passed through a magnetic field, and the amount of deflection experienced by the particles is measured to extrapolate their mass [11]. These pieces of equipment are very precise in targeting the desired analyte, as well as incredibly accurate in determining the relative concentration of said analyte within the sample being tested. This performance comes at a cost, however – these pieces of equipment are expensive, must be located in a laboratory to allow for sample preparation, and must be operated by an experienced professional. The time it takes to process a single sample also means that processing large volumes of samples is not possible. For these reasons, testing a sample with mass spectroscopy does not address the need in applications that require rapid and low-cost administration of tests.

Another analytical tool for measuring specific analytes in a fluid sample is the enzyme-linked immunosorbent assay (ELISA). There are several varieties of ELISA tests, but they all involve preparing a multiplex of wells with antibodies, adding the test fluid, and assessing the resulting colour change as the fluid reacts with the antibodies in the well [12]. Because the antibodies in the well bind only with the target analyte this type of experiment is very specific for the analyte in question, and since many wells can be prepared at once, it can be repeated simultaneously. The drawback to this type of experiment is that it too must be carefully prepared in a laboratory

setting by professionals, meaning that this test is relatively time consuming and expensive, not addressing the need in applications that require rapid and low-cost administration of tests.

The Polymerase Chain Reaction (PCR) is another analytical method which facilitates the rapid reproduction of DNA, even when starting with a very small starting sample [13]. This method has many applications, for example it can be used to identify pathogens in a fluid sample, or for forensic identification. While relatively rapid and effective at identification tasks, it is another method which must be carried out by a professional under strict environmental conditions and with costly equipment. Therefore, testing a sample can be costly and take time to wait for the sample to process.

The above-mentioned laboratory-based analytical methods offer the best detection and quantification performance at high cost and turn-over time to a broad spectrum of industries and sectors, however in many industries the need is for rapid and low-cost detection and quantification. Many such applications (e.g., roadside drug of abuse testing) greatly benefits from tests which are specific and precise enough to generate actionable results in few minutes. Such tests allow for performance of rapid on-site testing which enables informed decision making to prevent losses and liabilities. A test which attempts to achieve these goals is the Lateral Flow Immunoassay (LFIA), also known as Rapid Tests. The LFIA is paper-based assay is functionally similar to the ELISA test in that it relies on the action of antibodies to capture the specific analyte being tested for. LFIAs are inexpensive and easily mass-manufactured, as well as being so simple to use that it can be reliably performed by a lay-person. Interpreting the results of such an assay is simple – either one or two lines are visible, (depending on the style of LFIA) indicating whether the sample has tested positive or negative. Most beneficially, the LFIA can be performed on-site, generating results in mere minutes. For these reasons, there are already many industries which make use of LFIAs [14], such as COVID-19 screening [9,15], testing for contamination of seafoods [4], or, perhaps most widely known, as pregnancy tests [10]. Owing to widespread uses of LFIAs, the global rapid tests market size was valued at $33,329.15 million in 2020, and is projected to reach $97,606.33 million by 2030 [16].

The LFIA does have its own drawbacks; the tests typically provide a binary positive or negative response with respect to some predetermined concentration (aka. limit of detection). A binary response such as this makes a standard LFIA unsuitable for quantification of analyte, instead

serving a confirmation of the presence of an analyte. Also, because the tests are typically interpreted visually, the concentration of analyte must be relatively high before it can be made visible and confirmed reliably which effectively leads to relatively high limit-of-detection (LOD) for a typical LFIA (in some cases, too high to be of use). Due to these limitations, LFIAs are frequently believed to have low accuracy and are only approved as initial screening tools, requiring follow up confirmation with laboratory-based analytical tools. If the limitations of LFIAs could be alleviated, many more applications would be able to reap the benefits of on-site, user-friendly testing. For example, an accurate and sensitive LFIA enables low-cost and rapid screening of diseases and biomarkers by a patient at home.

The drawbacks of the LFIA have to do with a loss of meaningful signal amongst the noise of the system. A device which could increase the output signal of the LFIA could alleviate the stated drawbacks. As will be discussed in the next chapter, the response of an LFIA increases as the concentration of analyte in the fluid increases. If this response could be reliably measured, instead of relying on a binary visual interpretation, then not only would the LOD decrease, but some level of quantification would also be made possible.

While there have previously been many attempts to amplify the signal produced by the LFIA using a variety of different methods, one particular method which has been proven to be effective is that of photothermal radiometry (PTR) [17]. This method thermally excites the LFIA using photonic methods and interprets the resulting thermal signature to estimate the underlying concentration of analyte in the sample being tested. Our group has been previously demonstrated that the photothermal method is quite effective in decreasing the LOD of LFIAs, as well as allowing for quantification [18,19]. The devices developed to perform these tasks, however, are prohibitively expensive (> $350) and immobile enough that they counteract the positive benefits of using LFIAs in the first place. What our team's previous studies accomplish is a rigorous proof-of-concept, demonstrating that such a method is a viable solution, if it could be achieved with a suitably inexpensive device then the vision of inexpensive and sensitive at-home detection could be realized.

The goal of this thesis is to produce a suitably portable and inexpensive device which can decrease the LOD and quantify the results of a variety of LFIAs. The vision for this device is that one day it could become a household device, used to interpret a wide range of LFIAs, digitizing

and uploading the quantified results via an IoT connection (ie. Bluetooth or Wi-Fi). With enough households making use of these devices, a massive quantity of high-quality data could be generated, processed, and used by relevant professionals to track elements such as public health, which is currently unfeasible when using purely "analog" LFIAs. Additionally, by designing a standalone device rather than a smartphone-based device, the end-user's privacy can be more protected by not having health results implicitly tied to a personal device.

## 1.2 Specific Aims

The specific aims which will be demonstrated to have been achieved are as follows:

1. Designing and developing a low-cost, end-user photothermal device with internet-of-things (IoT) connectivity capable of:
    a. Significantly decreasing the Limit of Detection of lateral flow immunoassays
    b. Enabling quantification of analytes with commercially available, nominally binary LFIAs
2. Optimizing and validating the performance of the device on a large collection of rapid tests spiked with various concentrations of control positive solutions
3. Investigating the performance of the device in the field and on human samples
4. Investigate possibility of utilizing Machine Learning models for data interpretation and quantification, and comparing their performance with those of conventional signal processing

The thesis is structured into 5 chapters. Chapter 1 is an introductory chaptering highlighting the motivation behind the thesis and the significance of the problem being tackled. It also lays out the specific aims of thesis and its structure.

Chapter 2: Literature Review outlines the function and uses of Lateral Flow Immunoassays, as well as different methods found in literature for interpreting their results. The physics of Frequency Domain Photothermal Radiometry is then explored, as well as previous studies which

have made use of the principle. Finally, various machine learning methods for post-processing results are outlined.

Chapter 3: Methods begins by describing the iterative design process used to arrive at the final prototype device, as well as a set of custom LFIAs. Then, the various validation studies are outlined, including details about how the LFIAs were prepared – either in a lab or through human experiments. Lastly, a detailed overview of some of the various post-processing algorithms are explained.

Chapter 4: Results and Discussions details the results of the three experiments used to validate the prototype. Discussion about the design of the experiments and how they attempt to validate different functions of the prototype is included.

Chapter 5 concludes this thesis, outlining what was accomplished and exploring what next steps could be taken to further the goals of this project.

# Chapter 2: Literature Review

## 2.1 Lateral Flow Immunoassays

The lateral flow immunoassay (LFIA) is an easy-to-use, inexpensive, nitrocellulose-paper based assay which rapidly provides the user with its binary positive vs. negative result [20,21]. The use of an LFIA is simple; a fluid sample is first collected and depending on manufacturer specifications, a buffer solution may be added. A specified amount of solution is deposited in the sample collection well, mixing with the elements in the sample pad, before being pulled through the length of the test through capillary action. As the fluid passes the test and control lines, marker elements (normally gold nanoparticles (AuNPs)) are immobilized on test and control lines. The amount of immobilization on the test line is directly correlated to the concentration of analyte in the fluid. These markers exhibit visible optical colors, serving as indicators of either a positive or negative test result.

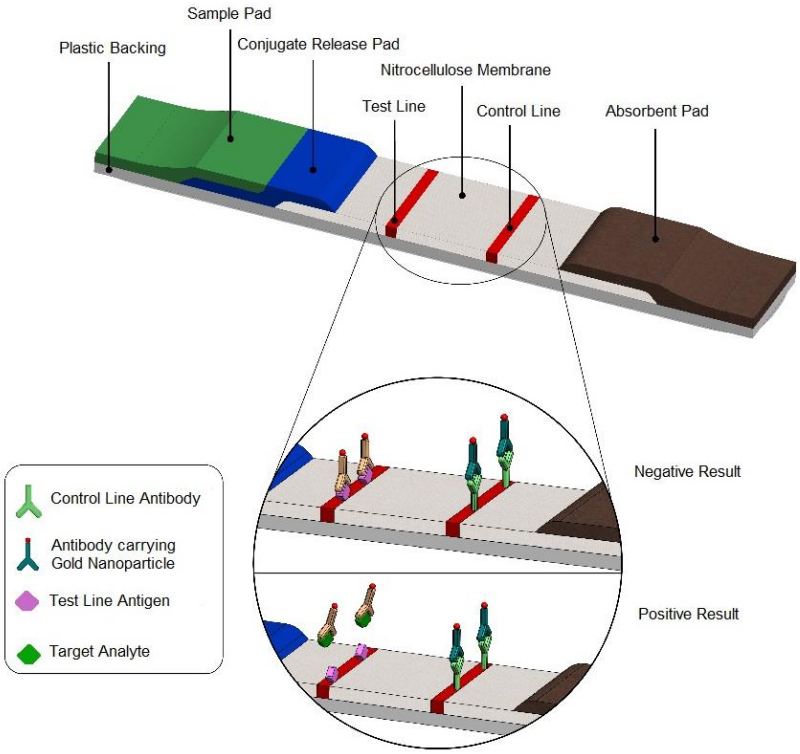



**Figure 1:** Rendering of a conventional, competitive-style Lateral Flow Immunoassay. Figure reprinted with permission from [22], © MDPI 2022

Figure 1 [22] illustrates a representative LFIA. There are six main components to the LFIA; from left to right these components are the backing, sample pad, conjugate pad, nitrocellulose body, absorbent pad, and the test/control lines. The backing is generally made of a polyvinyl chloride (PVC) or other opaque plastic and provides mechanical support for the test. The sample pad contains buffers and other additives which ensures that the fluid will flow smoothly and evenly up through the paper, as well as acting as a filter for larger, irrelevant particulate. One top of the sample pad is the conjugate release pad, which houses and stabilizes the detector particles; generally, these are nanoparticles which have been doped with some form of antibody (ie. a labelled antibody) related to the analyte being tested for. The vast majority of conventional LFIAs make use of colloidal gold nanoparticles (AuNPs; ~50nm diameter), which readily absorb specific parts of the visual spectrum of light to appear reddish to human eye. The body of the test is usually made of a nitrocellulose paper, which enables capillary action to pull the test fluid smoothly. An absorbent pad sits opposite to the sample pad and wicks away excess fluid, allowing for more fluid to pass through the lines, increasing the efficacy of the test.

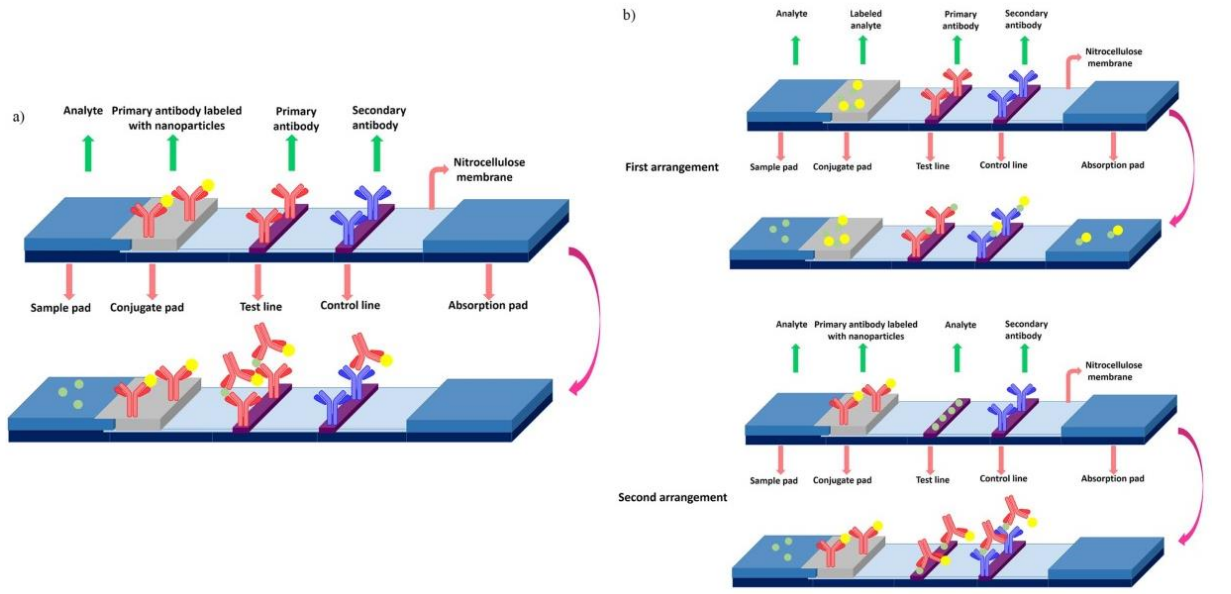



**Figure 2:** The two styles of LFIA **a)** Sandwich Style, where a dark test line indicates a positive test **b)** Competitive Style where a dark test line is a negative result. Figure reprinted with permission from [20], © 2016 Elsevier B.V.

The lines of an LFIA vary in composition depending on what the LFIA is testing for, and what style of test is being performed. Illustrated in Figure 2, there are two major categories of LFIA test: direct or ‘sandwich’ style, and competitive. Sandwich style tests are generally used for larger analyte targets; their conjugate pads contain antibodies which have been labelled with a

nanoparticle. When the test fluid is added to the sample pad, the analyte being tested for binds with the labelled antibody. As the fluids are pulled up the nitrocellulose paper, the bound targets encounter the test line. This test line contains antibodies which will bind with the conjugated pair, fixing them in place. Unbonded, labelled antibodies will pass the test line and be captured by the control line. When the test is complete, the more analyte the test fluid contained, the greater the concentration of conjugate on the test line. As such, for a sandwich style test, two lines is a positive result, one line is negative.

Competitive LFAs work differently, primarily being used to test physically smaller analytes, since these cannot bind to more than one site at a time. In this case, the conjugate pad contains the target analyte which has been labelled with some form of nanoparticle, and the lines have the antibodies affixed. When the sample is added, if there is analyte present, it will compete with the analyte in the conjugate pad for the spaces on the test line, leading to a less prominent test line. If there is no presence of analyte in the sample, then only labelled analytes will become affixed on the test line. In the case of competitive assays, two lines indicate a negative test result, while one line indicates positive.

LFIAs can also be more complex than this, for example, multiplexing [23] is a method wherein multiple test lines are arranged one after another, each testing for a separate analyte. This allows for higher throughput testing, as well as allows the collection of less fluid, which can be beneficial when the fluid collection is invasive (ie. blood collection). There is also ongoing work put into fluid preconcentration, where a fluid's analytes are concentrated before being submitted to the LFIA, thereby increasing the limit of detection of the LFIA.

Whether the LFIA is sandwich or competitive style, the next step after it has developed is interpretation. The vast majority of commercially available LFIAs are made with the intention that they will be interpreted visually [24]. This means that most LFIAs can produce either a positive or a negative result with respect to some cutoff concentration of analyte, as determined by the manufacturer. As has been mentioned previously however, the concentration of the labelled antibody on the test line – in both styles of LFIA – is proportional to the concentration of analyte within the fluid being tested. This means that visually interpreting the results of the test overgeneralizes the results, and using a more standardized method for interpretation could retrieve the information which would otherwise be lost.

## 2.2 Alternate Methods of LFA Interpretation

There have been many different methods for interpreting the results of LFIAs proposed in recent times [25,26]. Many of these methods have to do with changing the type of nanoparticle used when labelling the antibody in the conjugate pad, allowing for different styles of interpretation. In this section we will review some of these methods in greater detail.

### *2.2.1 Colourimetric*

The simplest route to interpret the LFIA test results is the use of colourimetric methods, wherein the colour of the test and control lines are measured by way of a camera, and their intensity is determined from those measurements. The method and rigor of these imaging methods vary greatly; some experiments have made use of specific lighting and imaging techniques [27], while others rely on ambient lighting and a smartphone camera [28]. Once the measurements have been made, the data is generally processed to extract the colourimetric intensity of three regions – the test line, control line, and the background paper. The background is treated as the baseline or background noise signal, and then the test line intensity is normalized to that of the control line. This process allows some level of control over differences in lighting conditions between experiments.

The benefits of this method are that it is a simple experiment to design and execute, requiring no specialty equipment beyond a camera and lighting, and requiring no changes to the typical LFIA which uses AuNPs as the label. It has been previously demonstrated that changing the nanoparticle to one which is better suited for colourimetric analysis can be effective [29], however this runs contrary to the main draw to this method, which is its simplicity and ease of execution.

The main problem with colourimetric methods which cannot be easily solved, even with a new nanoparticle, it inherent in its design. The nitrocellulose paper which makes up the body of the LFIA is highly reflective at visible wavelength, whereas the AuNPs tend to absorb these wavelengths. This means that when the test line is strong, there is an obvious difference between the response of the paper and that of the line, but when the test line is weak, only a small portion

of the light is absorbed. This means that colourimetric methods fundamentally have an inherently low signal-to-noise ratio when compared to other methods.

### *2.2.2 Fluorescent Imaging*

When selecting a label for the antibodies to be used as the detector in an LFIA, a fluorescent dye, [30] or nanoparticle such as a quantum dot [31] or semiconducting polymer dot (Pdot) [32] can be selected. Once the test is completed, a fluorescent light source is used to illuminate the test and control lines. The resulting light can be focused to a device through optical elements, and the fluorescent intensity can be measured.

This method avoids the issue of low signal-to-noise ratio inherit in the colourimetric method because the fluorescent labels are highly specific. That is, when an ultraviolet or fluorescent light is incident upon the test line, the surrounding paper does not fluoresce, whereas the fluorescent labels do. Any response measured must therefore originate from the labelled particles, and therefore the signal-to-noise ratio is high.

The drawback to this method is that many of the fluorescent particles are unstable, such that short shelf life would greatly hinder attempts to become widely adopted. The active sensing optical systems are also complex, requiring specialty filters and sensing elements to accurately capture the fluorescent signal, meaning that the setups are often bulky and always expensive.

### *2.2.3 Ramen Spectroscopy*

Ramen Spectroscopy is another method of interpreting LFIAs which relies on Ramen scattering. When a photon, incident on a molecule, either increases (Stokes-Ramen scattering) or decreases (Anti-Stokes-Ramen scattering) the vibrational energy of the molecule, the energy of the photon is changed. Molecules can be designed to specifically change the energy signature of a photon which has interacted with it, and this signature can be detected through spectroscopy. The signature is very specific to the molecule, and thus extremely low concentrations of the molecule can be detected using this method [33]. This overall effect can be greatly increased if the molecule has been previously adsorbed to a metal surface, and this is known as Surface-Enhanced Ramen Spectroscopy (SERS).

To use this method in conjunction with a LFIA, nanoparticles (typically metal to enable SERS) have Ramen Spectroscopic labels adsorbed to them. After the LFIA test is complete, a laser of a specific wavelength is shone on the sample and the reflected signal is captured and spectroscopically interrogated to determine the concentration of the label on the lines of the assay. It has been demonstrated on LFIA tests for HIV-1 [34] and avian influenza H7N9 [35] that this is an effective strategy, often increasing the limit of detection by order(s) of magnitude. Similar to that of fluorescent methods, the drawbacks to this method of interrogation are that the equipment required to perform this ramen spectroscopy are expensive and generally not portable, again running contrary to the main purpose of LFIAs being low-cost end-user solutions.

As can be seen from the comparison of these methods, there are several challenges when it comes to developing proper interrogation techniques for LFIA. While colourimetric methods are attractive because they use LFIAs which require no modification to the nanoparticle labels, the method is haunted by its fundamentally low signal-to-noise ratio. Fluorometric and Ramen spectroscopic methods overcome this hurdle with the use of markers to increase the signal-to-noise ratio, but this requires a fundamental change to the labels in the LFIA, accompanied with expensive equipment for interpretation. The method which attempts to reconcile these opposing issues, overcoming the signal-to-noise problem while also requiring no fundamental changes to the label of the LFIA, are photothermal methods.

## 2.3 Photothermal Radiometry

### *2.3.1 Surface Plasmon Resonance*

Fundamental to all LFIA photothermal interpretation methods is a concept known as Localized Surface Plasmon Resonance (LSPR) [36]. When a collection of free electrons coherently oscillates in a metal, this is referred to as a plasmon. These plasmons can be induced on the surface of bulk metals by introducing photons of suitable wavelength. An incident photon on the surface of a metal excites the electric field within the skin of the metal, which generates a surface wave, or more specifically a Propagating Surface Plasmon Polariton (PSPP). These PSPPs do not penetrate deep into the skin of the metal and are quickly dispersed over the surface of the metal [37].

In an AuNP, where the diameter of the particle is smaller than that of the incident photon, the electrons in the entire bulk of the nanoparticle resonate coherently. This is because the nanoparticle is so small; while the plasmon usually cannot penetrate deep into the metal, this metal is so thin that the wave effectively reaches the entire bulk mass of the particle. As this resonance decays, the bulk of the particle increases in temperature and it emits thermal infrared radiation, also known as LSPR resonance absorption [38]. The photon wavelengths which are most effective at inducing this resonance are dependant on several factors, such as particle diameter and the surrounding environment, but the fact remains that for any AuNP there are a range of wavelengths which are very effectively absorbed.

The heat and primarily Infrared Radiation (IR) emitted by the decaying LSPR in the excited AuNPs can be viewed as a specific signature in the same sense that fluorescence or Ramen signatures were in their respective interrogation methods. That is, excitation of AuNPs and their infrared thermal response are at completely different spectral ranges, eliminating any chance of crosstalk between the two and inherently enhancing the detection sensitivity (aka. a hybrid detection method). A major benefit to the photothermal method for LFIA interpretation is that significant majority of existing LFIAs already use AuNPs, so no changes to manufacturing are required to perform this form of interrogation.

### *2.3.2 Thermal Contrast Imaging*

The simplest method to leverage this thermal signature is thermal contrast imaging. In this method, a light source of a wavelength which excites the AuNPs in the LFIA is shone across the length of the LFIA for a specified length of time. Simultaneously, the infrared radiation produced by these lines are measured, usually by way of a thermal camera. The difference in infrared radiation (indicating a temperature change on the surface) compared to the baseline ambient at each spatial point of the LFIA is recorded. The measurement at a blank spot on the LFIA is often considered to be an adequate measurement of the noise of the system, so its value is used as a background to compare the results of the test line with [39,40].

This method effectively leverages the phenomenon of SPR to overcome the problem of signal-to-noise ratio found in colourimetric methods without requiring changes being made to the nanoparticle label typically found in LFIAs. However, this method is still limited by the fact that

the change in temperature which is measured by the thermal camera is fully dependant on the ambient temperature at which the test is being performed at. This is an obvious hinderance when the goal is an assay which can be prepared on-site, regardless of what the temperature happens to be. There is a method of PTR which can overcome this obstacle, known as Frequency-Domain Photothermal Radiometry (FD-PTR).

### *2.3.3 Frequency Domain Photothermal Radiometry (FD-PTR)*

The method of FD-PTR takes the concept of thermal contrast imaging a step further by introducing the idea of thermal waves. These methods, developed primarily for use in non-destructive testing [41], have been used to probe beneath the surface of materials, elucidating delamination or defects that would otherwise be invisible. The driving idea behind this method is that thermal waves induced in the medium will be reflected or otherwise impaired as they encounter discontinuities or other changes in the material properties of the medium being probed. These interrupted thermal waves are then shifted in both phase and amplitude, and those shifts can be used to infer the depth of the incongruity within the medium.

The medium can be thought of as being semi-infinite (ie. thickness much larger than the depth of penetration of thermal waves) and having a surface which absorbs the incoming light. The mathematical phrase describing the propagation of surface heating through a semi-infinite medium is as follows [41]:

$$T(z,t) = \frac{A_0 Q_0}{2\sqrt{\rho c k \omega}} e^{\left(\frac{-z}{\mu}\right)} e^{j\left(\omega t - \left(\frac{z}{\mu}\right) - \left(\frac{\pi}{4}\right)\right)} \tag{2.1}$$

$$\mu = \sqrt{\frac{2\alpha}{\omega}} \tag{2.2}$$

Where $Q_0$ is the intensity of the light source (assumed to be of a single wavelength), $A_0$ is the fraction of light which is absorbed by the surface, $\rho$ is the density of the medium, $c$ is the material's specific heat capacity, $k$ is the sample's thermal conductivity, $\omega$ is the light source's modulation frequency, $\alpha$ is the thermal diffusivity, and $\mu$ is the thermal diffusion length. As

illustrated in equation 2.2, the thermal diffusion length is inversely proportional to the modulation frequency; a lower modulation frequency increases the thermal diffusion length.

The two elements of equation 2.1 which can be controlled in an experiment are $Q_0$, the intensity of the light source, and $\omega$, the modulation frequency. In general, the higher the intensity of light source chosen the higher the increase in temperature, resulting in a higher magnitude of signal. Modulation frequency can also be tuned such that the depth of interest is represented well by this expression. In general, low frequency modulation frequency will penetrate deeper into the medium, and higher frequencies will target shallower depths. The effect of modulation frequency on thermal-wave inspection depth is mathematically evident in equation 2.1 since thermal diffusion length is in the denominator of the decaying negative exponent signal envelope.

As the temperature through the depth of the medium changes, Infrared Radiation (IR) is produced. To determine the total amount of IR at the surface, the IR contribution at all depths through the medium must be considered. Then, the emissivity of the medium should also be factored in. The depth integral which describes this is as follows [42,43]:

$$s(l,t) \propto \overline{\mu_{IR}} \int_0^l T(z,t) e^{-\overline{\mu_{IR}}\, z}\, dz \tag{2.3}$$

Where *s(l,t)* is the full-depth signal received by the thermal cameral or sensor at time *t*, *l* is the full depth of the sample, and *T(z,t)* is the induced thermal wave field at depth *z* and time *t*. If we assume that the system has reached a steady state, the term $\overline{\mu_{IR}}$ is the spectrally averaged IR emissivity and absorptivity of the medium.

Based on equation 2.3, the measured infrared signal can be mathematically modeled as:

$$s(t) = A sin(\omega_0 t + \varphi) \tag{2.4}$$

This introduces a new term $\varphi$, which is a term to describe the difference in phase induced by the propagation of the thermal waves through the medium. $A$ is the term which broadly describes the amplitude of the measured signal. Investigating equation 2.1 reveals that the terms with the greatest impact on $A$ are $Q_0$, the intensity of the light source, and $A_0$, which refers to the surface

absorptivity of the source light. In an experiment, $Q_0$ will be kept constant, meaning that the absorptivity $A_0$ is directly correlated with $A$. Since we also know that this absorptivity is dependant on the concentration of AuNPs, we can conclude that $A$ will be highly correlated with the underlying concentration of AuNPs, and therefore the concentration of analyte in the fluid sample.

Finding the amplitude of $s(t)$ from experimental measurements is achieved through quadrature demodulation of acquired infrared signals. To do so, the measure signal is divided into quadrature and in-phase elements and then weighted and low pass filtered to eventually enable calculating signal's amplitude and phase. The process is mathematically demonstrated in the following formula [43]:

$$
\begin{aligned}
&Asin(\omega_0 t + \varphi) \xrightarrow{Reference} \begin{bmatrix} \sin(\omega_0 t) \times Asin(\omega_0 t + \varphi) \\ \sin(\omega_0 t + 90) \times Asin(\omega_0 t + \varphi) \end{bmatrix} \\
&\xrightarrow{Mixing+Weighting} \begin{bmatrix} \frac{A}{\sqrt{2}}[\cos(\varphi) - \cos(2\omega_0 t + \varphi)] \\ \frac{A}{\sqrt{2}}[\sin(\varphi) - \cos(2\omega_0 t + \varphi + 90)] \end{bmatrix} \\
&\xrightarrow{Low\ Pass\ Filter} \begin{bmatrix} \frac{A}{\sqrt{2}}\cos(\varphi) = S^0 \\ \frac{A}{\sqrt{2}}\sin(\varphi) = S^{90} \end{bmatrix} \\
&\xrightarrow{Transformation} \begin{bmatrix} A = \sqrt{(S^{90})^2 + (S^0)^2} \\ \varphi = arctan\left(\frac{S^{90}}{S^0}\right) \end{bmatrix}
\end{aligned}
\tag{2.5}
$$

The quadrature demodulation method is time-domain processing of signals which is normally favoured in older systems using analog electronics. Today's digital systems can effectively perform Fourier analysis of infrared signals via the fast Fourier transform (FFT) algorithm to rapidly determine the amplitude and phase of signal at the modulation frequency used in the measurements. While the FFT method is technically less memory efficient, its memory requirement is minimal for today's digital systems, especially for systems with single element sensors (like the ones in this thesis) that produce data on the order of kilobytes. Since the FFT method is simpler to implement, it will be used for this project.

## 2.4 Previous FD-PTR Designs

FD-PTR analysis of LFIAs is a patented innovation of our group which has gone through several transformation since inception in 2018. Experiments have previously been performed, by our group, which demonstrated the efficacy of the FD-PTR on LFIAs in the past. The following two devices directly precede the device which was designed for the purpose of this thesis.

### *2.4.1 Tabletop System*

The initial design [18] served as an initial proof of concept, utilizing a 30w 808nm laser, producing a relatively high intensity beam of 2.7 W/cm$^2$, illuminating the entire LFIA at once with a modulation frequency of 1Hz. A thermal camera (Xenics Gobi 640, Belgium, spectral range 8-14µm) was used with a frame rate of 100 frames per second to measure the resulting infrared radiation. The system diagram can be seen in Figure 3.

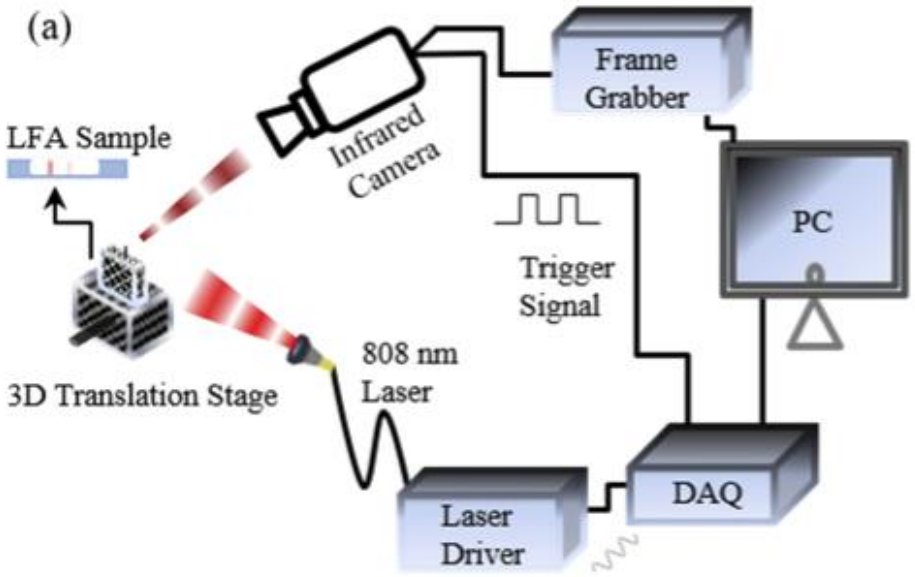



**Figure 3:** Schematic of the initial tabletop system. This system required the use of an external PC and DAQ to properly handle the data being produced by the thermal camera. Figure reprinted with permission from [18] © Optical Society of America

Data collected by the thermal camera is processed on-line through the formula outlined by equation 2.5. The trigger signal produced by the DAQ is used to modulate the laser and the sampling rate of the camera. This signal is also passed to the PC as the reference signal required for demodulation of the incoming data. Once demodulated, the outcome of the process is an intensity value for each pixel in the camera's frame, which can be seen in Figure 4(a). These values are then averaged over the width of the LFIA, and the average values can be plotted as seen in Figure 4(b). The intensity of the control line remains relatively consistent, with large changes in the intensity of the test line which correlate with the changes in concentration of THC in the sample used to dope the LFIA.

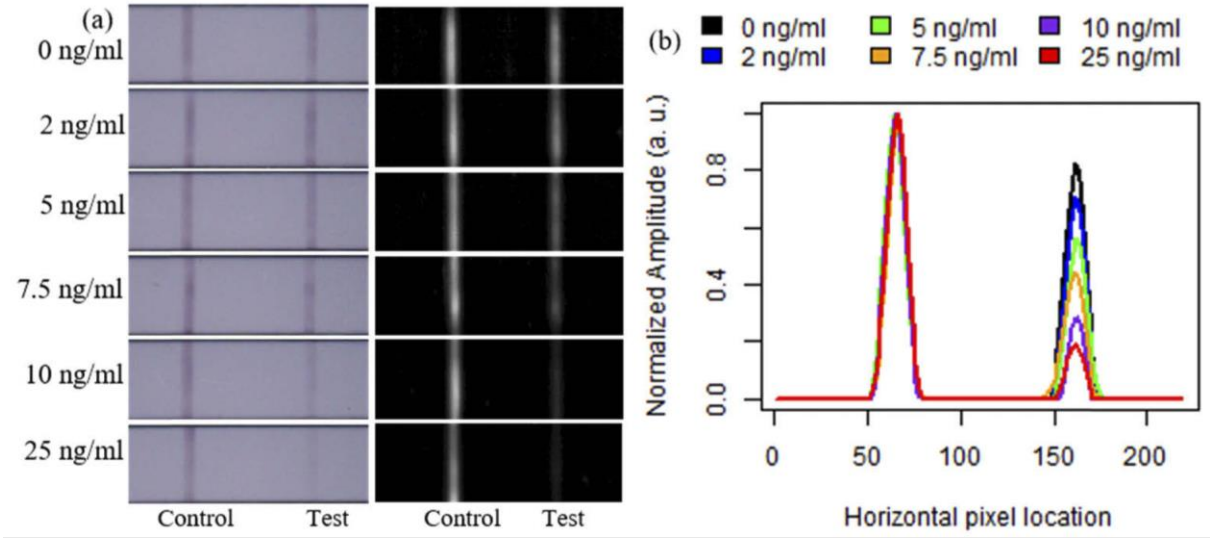



**Figure 4:** a) Conventional image and intensity profile of a set of representative LFIAs, and b) The intensity of response averaged over the width of the LFIA for each concentration. Figure reprinted with permission from [18] © Optical Society of America

In this experiment, a total of 6 concentrations were prepared, 10 LFIAs were doped with each concentration, and each LFIA was interrogated 5 times. The resulting data was then placed in concentration categories and a 95% confidence interval calculated for each concentration, as illustrated in Figure 5.

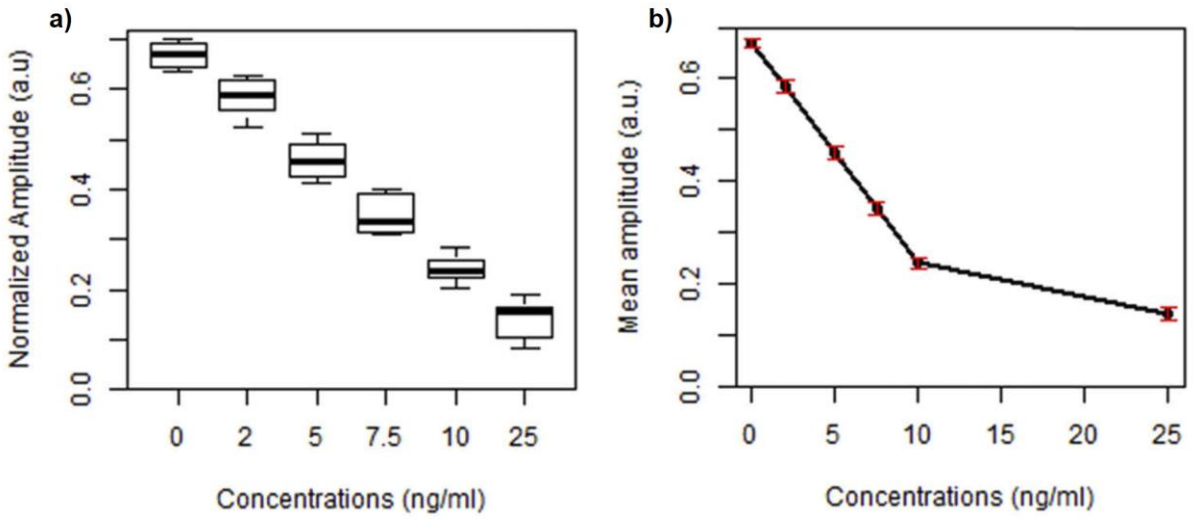



**Figure 5: a)** The responses of LFIAs prepared at different concentrations of THC and **b)** The same data, scaled by concentration and with 99% confidence intervals. Figure reprinted with permission from [18] © Optical Society of America

These charts demonstrate that LFIAs prepared with different concentrations of THC generate responses which are clearly discernable from one another. Even in the case of the lowest cutoff concentration tested (2 ng/mL), positive and negative responses can be readily resolved from one another. In other words, this device was able to greatly lower the limit of detection to 2 ng/mL (compared to the visual LOD of 10 ng/mL). Beyond this, the device was also able to semi-quantitatively classify LFIAs into their respective concentration bin.

The performance of this device is as good as one can hope for without making modifications to the LFIA itself. The major drawback, however, is that this device was a large tabletop setup, making use of a thermal camera and supporting equipment which costs upwards of $30,000. As a proof of concept, this experimental setup was a success, however if the system is to be turned into an end-user device, the cost and size needed to be drastically reduced without a major decrease in performance.

### *2.4.2 Preliminary Handheld Reader*

The next iteration [19] was an improvement on the previous design, exchanging the expensive thermal camera with a much more affordable thermal camera meant for use with a smartphone. This camera also required external optics in the form of a zinc-selenide lens. The laser selected for this purpose was also an 808nm wavelength, however it had a maximum luminous power of 600mW. These features are all labelled in Figure 6.

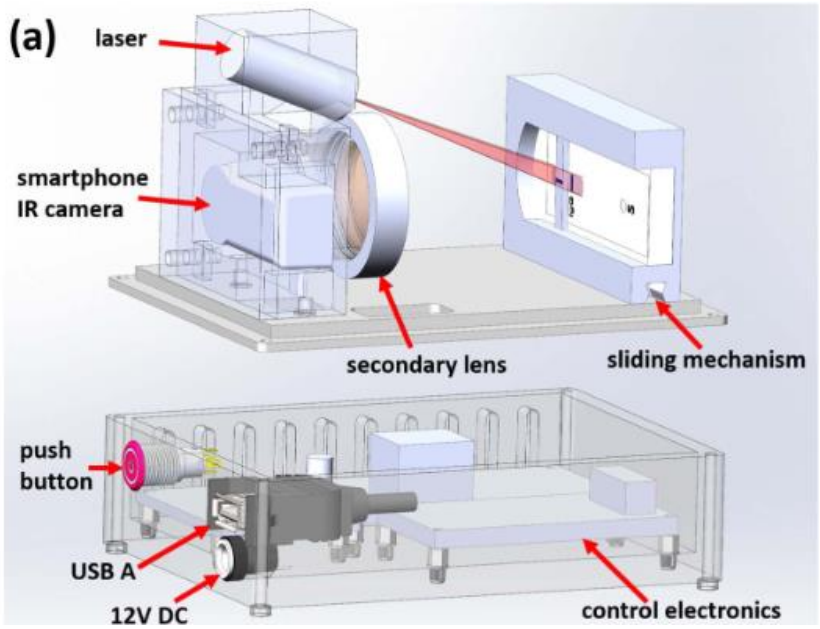



**Figure 6**: Schematic of the portable device which used a much less powerful laser and computer when compared to the large system depicted in Figure 3. Figure reprinted with permission from [19] © [2021] IEEE

This device was validated using COVID-19 LFIAs, prepared with known concentrations. While the LFIAs themselves test for both IgG and IgM, these two test lines used different nanoparticles, and only the IgG response could be resolved using the 808nm laser. As can be seen in Figure 7, a very clean response was extracted from the set of LFIAs using this device. Much the same as the previous iteration, the limit of detection can be greatly decreased, in this case reliably reaching 0.2 µg/mL. We again see the clear delineations between the different concentrations, indicating that classifying a response between the different concentrations bins is a trivial task.

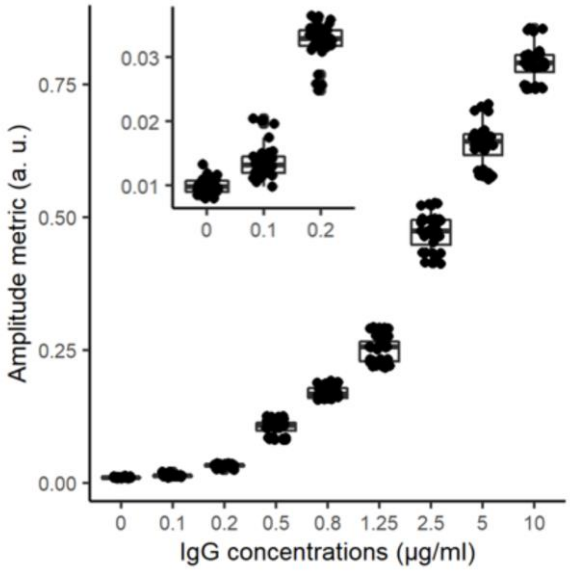



**Figure 7**: The dataset of interrogations of COVID-19 LFIAs, illustrating the difference in thermal wave response as a function of concentration. The LFIAs tested in this experiment are sandwich style, so it is expected that the amplitude of response rises with the concentration. Figure reprinted with permission from [19] © [2021] IEEE

This setup's overall cost was reported at approximately $350; a great improvement from the previous iteration, however not inexpensive enough to be a household device for analysis LFIAs costing about 10 cents each. While more compact than its predecessor, this system was still larger than what could be reasonably considered handheld, not to mention the fact that it still required to draw power from a wall plug rather than an on-board battery. There was also no on-board processing, or method for connecting to the IoT and directly uploading results. In all, there was room for improvement on the device before it could be considered an end-user device, but strides had been made by greatly reducing system cost while maintaining performance.

### 2.5 Data Processing in FD-PTR

Through the development of these previous device iterations, a general data processing algorithm has been developed to extract meaningful data from the raw signals. This conventional signal processing algorithm will be used for processing on this project as well, with some modifications which will be explained in the next section. In addition to the conventional processing methods, many different machine leaning algorithms were trained to further increase the performance of the device. The function of these machine learning algorithms will also be outlined.

### *2.5.1 Conventional Data Processing*

Conventional data processing is the process of transforming existing data into a generalized model which can be used to classify new data. This is a static (or rule-based) process, in that once a model or an algorithm for processing data has been established, this process does not change when the model is exposed to new data. The steps of conventional processing can be seen in Figure 8, and will explored in detail.

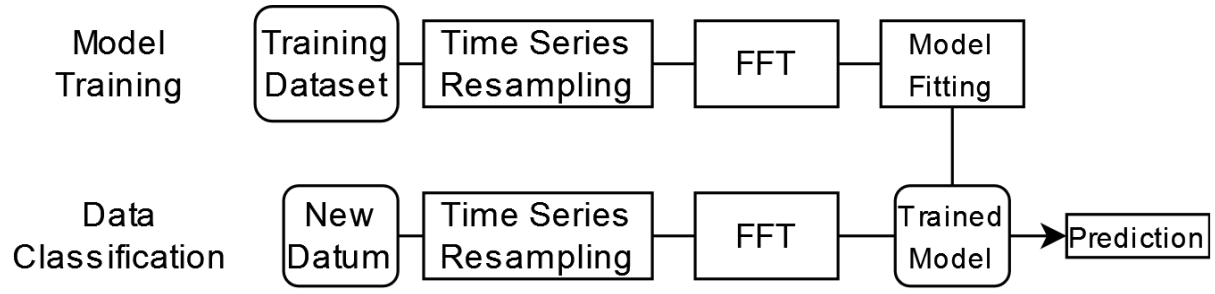



**Figure 8**: The conventional data processing scheme used in this thesis. Once a model has been built using the training data it will not change in response to new data.

As was explained in Section 2.3.3, FD-PTR relies on extracting the phase and amplitude of the thermal waves induced in the medium by the modulated light source. This can be accomplished in two ways. The first method is through the process outlined in Formula 2.4, where the on-line data frames are demodulated into quadrature and in-phase elements by mixing with the reference wave [43]. These signals are subsequently combined to determine the amplitude and phase. This method was used in both previous device iterations as this method is efficient with respect to computer memory.

An alternate method, developed partially in the most recent iteration, instead made direct use of an FFT to extract these amplitude and phase signals. Rather than performing on-line processing, the full data collection was stored in memory at a 3D dataset/movie. Then, the data corresponding to each pixel was treated as an individual time series, with the intensity fluctuating over time with respect to the modulation frequency. These individual time series could then be transformed to the frequency domain using an FFT, and the amplitude extracted from the known modulation frequency. Repeating this process for each pixel leads to calculation of amplitude (e.g., Figure 4(a)) and phases images from the 3D dataset. In the works reported in the literature, only the Amplitude image has been used for interrogation of LFIA. Interrogation process involves segmenting a space between the control and test lines of the LFIA to serve as a

measure of background/ambient noise. The average amplitude value of the background was then subtracted from the entire image to correct for any absorption of light by the white nitrocellulose paper. Then, test and control lines of the LFIA could be manually segmented and isolated from the rest of the background-subtracted Amplitude image. The image was then divided by average value of pixels in the control test segment to yield the normalized amplitude image. This image was then averaged row-wise to yield a profile from which the ratio of signals from test to control lines are calculated. These ratios (aka. normalized amplitude ratios) are normally between 0 and 1 and provide information on ratio of light absorption in test line compared to the control line.

The final step of post-processing is calibration. To perform this task, LFIAs which have known concentrations are used so that labelled data can be generated. For each concentration, the measured intensity is averaged; this value represents the center point of this concentration bin. The cutoff intensities between each bin are calculated by taking the midpoint between sequential intensity averages. After this process is complete, all the data points can be sorted by their respective intensities into a concentration bin. A cutoff concentration can then be selected, and a confusion matrix generated. The values from the matrix can then be used to generate performance metrics such as the true/false positive rate, accuracy, or specificity, to determine how effective the device is.

### *2.5.2 Machine Learning*

Professor Tom Mitchell states in reference to machine learning that "A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**." [44]. To paraphrase, machine learning is characterized by an increase in performance when presented with more experience (ie. data). The general machine learning process which makes use of this principle can be seen below in Figure 9, outlining how a general model is initially generated and then subsequently improved as new data is continuously fed through it.

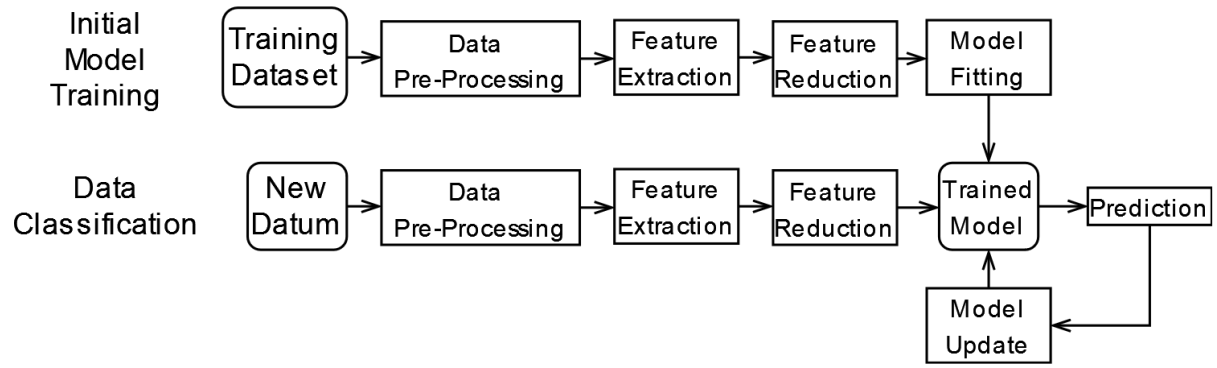



**Figure 9**: The machine learning process used in this thesis. Note that subsequent predictions can be used to update and improve the model.

The machine learning pipeline begins with data preparation [45]. Data often contains artifacts or irregularities which must be remedied before passing through the rest of the pipeline, otherwise they will lead to the algorithm picking up on patterns which do not actually exist. Missing values in the dataset must be removed and, in some cases, imputed and replaced. It is also good practice when working with sequential data series to resample raw signals so that they all have the same number of samples at an equal sampling rate.

Data is then typically transformed, combined, or simplified into features. This process coalesces the relevant data into a smaller set of features which can then be passed into the machine learning algorithm. Different features have different numerical values and distributions, meaning they will each influence the algorithm differently. Also, many machine learning algorithms make assumptions about the mean value and variance of data features, and as a result features must be scaled and normalized. There are several approaches to this [46], but the most common standardization practice is to remove the mean from all features and then scale them with respect to the standard deviation (aka. z-score normalization). This retains the information contained by the feature while ensuring that all features have the same numerical weight.

When training, it is also extremely important to keep separation between the data used to train versus that which is used to test. An algorithm whose performance is measured using samples which it has already been trained on is going to give biased results and seem to perform much more effectively than it can in a generalized case. For this reason, great care needs to be taken to avoid any of the data used for testing from “leaking” into the training dataset.

When presented with a labelled data set there are two main categories of machine learning algorithms: classification and regression. Regression type algorithms use the input data and estimate some numerical output value. Classification refers to algorithms which take in data and attempt to classify said data into one of several pre-existing categories, or classes. In the case of numerical values, these classes can be considered bins or ranges of potential values. A deciding function can also be performed on the output of a regression algorithm, effectively classifying the result.

There are many algorithms which can perform these tasks, text below offers a brief overview of the algorithms that have be utilized in this thesis.

#### *2.5.2.1 K-Nearest Neighbors (KNN)*

This algorithm [47] can be used for classification or regression, but the algorithm has only been used for classification in this thesis. K-Nearest Neighbors (KNN) is a relatively simple algorithm which creates a collection of each individual feature, including their associated label. When a new sample is to be classified, each feature is iterated through, and a user-defined number of nearest neighbors are collected. These neighbors each act as a vote as to which class the incoming sample belongs to. These votes can be uniform, in that each neighbor weighs their vote equally, or by distance, where a vote is proportional to the distance between the new feature and neighbor.

This type of algorithm is very effective at labelling unknown samples which are similar to training samples but can have trouble generalizing further. It is up to the user to set the hyperparameter of k, and the weighting schema. In many cases using distance as the weighting scheme outperforms the uniform weighting because a few features that are very similar can then outvote a number of relatively dissimilar votes. A higher value for k can help the algorithm against noise since there is a larger sample size for voting, however if k is too high it becomes computationally heavy to compute the results and it begins to consider votes from points which are very far from the new entry.

### *2.5.2.2 Support Vector Machine*

A support vector machine (SVM) [48] is an algorithm which places all training data points in N-dimensional space (where N is the number of features) and then attempts to fit a hyperplane separating the different classes. The subset of training data points which are closest to the boundary between classes become the support vectors, which make up each class boundary. The distance between two class boundaries is referred to as the margin, and the SVM algorithm attempts to maximize this margin so that the classes are kept as discrete as possible. The decision hyperplane is placed directly between these boundaries, and because the hyperplane is linear, it is relatively computationally simple to compute. However, most datasets are not linearly separable, and so the data must first be transformed.

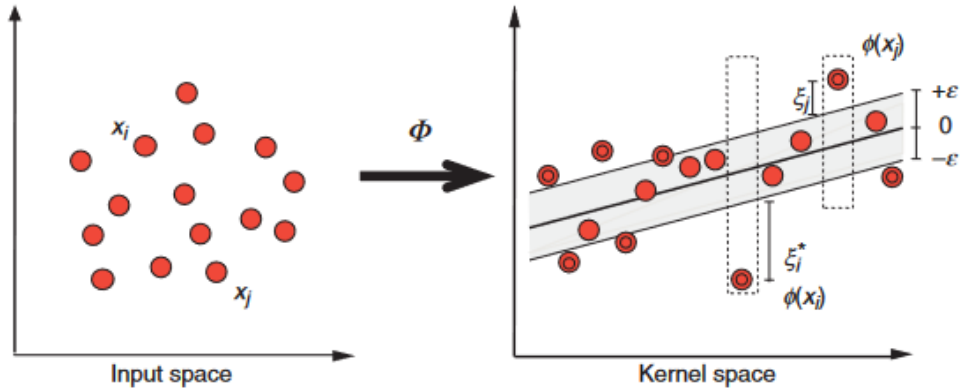



**Figure 10**: **a)** Raw feature data before transformation and **b)** after transformation by a kernel method. After the transformation, the data are now linearly separable. Figure reprinted with permission from [48] © 2014 John Wiley & Sons, Ltd.

To separate these non-linear datasets efficiently, a kernel method [49] is employed. The results of this kernel process are represented in Figure 10, panel a) is the raw data, panel b) is after a kernel transformation has been applied. The basis for this method is the fact that if a non-linear transformation is applied to a dataset, then the transformed data may be separated by a hyperplane in the higher dimension. This transformation process however can be computationally expensive to apply over the entire dataset. Crucially, applying a non-linear transformation to a pair of datapoints and then computing the dot product of the result can equivalently be performed by taking the dot product and then applying a transformation; this process is often less computationally expensive. This kernel method avoids the pitfall of expensive transformations by leveraging the fact that the transformation step can be avoided

completely by instead computing the distance (ie. dot product) of pairs of data points. After the kernel method has been applied with a suitable kernel, then hyperplane margin maximization can be computed. There are several different kernels which can be applied in the process of SVM training, the most common of these kernels are shown in **Table 1.**

**Table 1:** Support Vector Machine Kernels

| **Kernel Name** | **Kernel Function** | **Coefficients** |
|---|---|---|
| Linear | $\langle x, x' \rangle$ | - |
| Polynomial | $(\gamma \langle x, x' \rangle + r)^d$ | Constant Coefficient: r<br>Degree: d<br>Gamma: γ |
| Radial Basis Function | $e^{-(\gamma \|x, x'\|^2)}$ | Gamma: γ |
| Sigmoid | $tanh(\gamma \langle x, x' \rangle + r)$ | Constant Coefficient: r<br>Gamma: γ |

Each kernel, except for the simple linear kernel, has at least one hyperparameter which can be optimized by the user. The degree and gamma parameters to some extent control the complexity of the kernel; a highly complex kernel will be able to handle nonlinearities in data more easily, however it will also become more prone to overfitting.

#### *2.5.2.3 Decision Tree*

A decision tree [50] is a structure which separates data into nodes (or 'leaves') based on a series of binary decisions. It can be used for either classification or regression tasks, however only classification will be discussed here. When a new piece of data is to be classified, its features are evaluated at each branch of the tree until it eventually ends in a node which is the best guess for its class. Such a tree structure can be generated in many different ways, but a commonly used modern implementation of this algorithm is referred to as CART (Classification and Regression Tree) [51].

This algorithm begins with the assumption that the feature space can be divided up into a set of regions, and that each region is represented by some constant or classifier – ideally a class which well represents the response of the data contained in that region. The algorithm begins by

iterating through each feature at a set of split points, creating two separate regions at each step. The error of each split point must be measured; for classification tasks it is standard to use either the Gini Index or the Cross-entropy measure, as seen in Table 2.

$$p_{nk} = \frac{1}{n}\sum I(y_i = k) \tag{2.6}$$

Where $n$ is the number of observation ns in that region, $k$ is the class, and $y_i$ is the response.

**Table 2:** Error Functions for a Decision Tree

| **Error Function Name** | **Error Function** |
|---|---|
| Gini Index | $\sum_{k=1}^{K} p_{nk}(1 - p_{nk})$ |
| Cross-Entropy | $-\sum_{k=1}^{K} p_{nk}\log(p_{nk})$ |
| Misclassification Rate | $1 - p_{nk(max)}$ |

This evaluation is performed on each iteration and split point combination, and the combination which produces the minimum sum of errors is selected; this feature and split point are selected as the first node of the tree. This process then repeats iteratively on each branch or region of the tree, further increasing the specificity.

A user-defined end node size is specified, and a node of this size will no longer be split any further. Once all nodes have reached this point, the tree can be 'pruned', or simplified, through use of a cost-complexity function, typically the misclassification rate for classification problems. This cost-complexity function successively prunes the internal nodes which produce the least amount of difference in the final cost function, and then performs cross validation to measure the change. A simplified tree which produces the lowest error is the final algorithm.

There are several user defined parameters which can be modified when training a decision tree. The first is the splitting algorithm (ie. the equations in Table 2). They are all viable choices; however, both the Gini Index and Cross-Entropy are differentiable, and thus more suited for computational optimization. The other parameter defined by the user is the final node size. This can vary greatly depending on the number of samples in the dataset, as well as how resilient the user would like to make the tree to overfitting.

#### *2.5.2.4 Naïve Bayes Classification*

The Naïve Bayes Classification algorithm [52] utilizes Baye's theorem and naïve assumptions about the independence of variables in order to predict the classes of unknown datapoints. Bayes theorem states that the probability of y, given datapoints $X = [x_1, x_2 \dots x_n]$ is equal to the probability of y, times the probability of X given y, over the probability of X.

$$P(y|X) = \frac{P(y)P(X|y)}{P(x_i)} \tag{2.7}$$

This algorithm begins with the assumption that all variables are independent of one another.

$$P(X|y) = \prod_{i=1}^{n} P(x_i|y) \tag{2.8}$$

This assumption is not true in many real-world cases; however, violation of this assumption tends not to have a large impact on the performance of the algorithm. Since $P(x_i)$ is also constant, it can be removed from the final equation. The classifier now becomes:

$$y_{pred} = \max_{y} P(y) \prod_{i=1}^{n} P(x_i|y) \tag{2.9}$$

There are several methods for estimating $P(x_i|y)$ using the training dataset, but for the purpose of this report the gaussian function was utilized. This function assumes that the features tend to land in a gaussian distribution with respect to their respective class. The gaussian distribution function is given as follows:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{\left(x_i - \mu_y\right)^2}{2\sigma_y^2}\right) \tag{2.10}$$

Given the training data, for each feature and class pair, the mean and variance of the feature space can be fit using the maximum likelihood function. Simply, the mean is found by

calculating the mean values of the relevant training data, and the same is done with the variance. When an unknown data point is to be classified, the features are each input to this model and the class which produces the highest overall probability is predicted as the new data point's class.

This is a simple model which requires no hyperparameters.

#### *2.5.2.5 Neural Network*

An Artificial Neural Network [53,54] is a general term which refers to a multiple layered, interconnected network of nodes or 'neurons'. Feature data is passed into the network on the first layer and is propagated through the network before arriving at the final layer, which can then be further interpreted as either a class, or a continuous value in the case of a regression network.

There are many forms that a neural network can take, but for the purpose of this thesis only relatively simple networks were considered. A network has at least one hidden layer, that is, a layer containing a number of neurons, and in most cases each neuron is connected to all the neurons in the previous and following layer. It is possible to have neurons which 'skip' layers, or even create a feedback loop, however these architectures were not considered for the purpose of this thesis. There can be an arbitrary number of hidden layers, just as there can be an arbitrary number of nodes per layer, but the larger the network becomes, the longer it will take to train, the more memory it will take in the computer while running, and it can also begin to run the risk of overfitting to the training data. A basic example of a neural network with two hidden layers is illustrated in Figure 11.

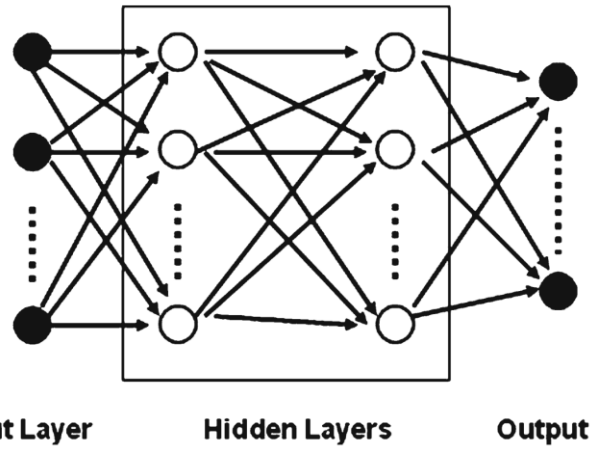



**Figure 11**: An example of a relatively simple ANN containing two fully interconnected hidden layers. Figure reprinted with permission from [54] © 2008 Humana Press, a part of Springer Science + Business Media, LLC

Each node within a neural network performs two steps. First, it generates a linear response by independently weighting the input from each neuron in the previous layer and summing all contributions. Then this linear combination is passed through a non-linear activation function. By passing these values through an activation function, the network can capture non-linear relationships between the features and the label, or response. There are several functions which can be used as activation functions, they are listed in Table 3.

**Table 3:** Activation Functions for use in Neural Networks

| **Activation Function Name** | **Activation Function** |
|---|---|
| Identity | $x$ |
| Logistic | $\frac{1}{1+e^{-x}}$ |
| Hyperbolic Tangent | $\tanh(x)$ |
| Rectified Linear Unit | $\frac{x-\lvert x\rvert}{2}$ |

The output of the activation function is then passed to the connected nodes in the next layer of the network. When training a network, the output is produced on the final layer of the network and can be compared with the known label. In the case of classification, the error of the prediction is calculated using the cross-entropy function (see Table 2). This error can then be used to modify the weights of the neurons through a process known as backpropagation [55,56]. The basic formula for this backpropagation is as follows:

$$\Delta w_j = \eta x_j \delta_j \tag{2.11}$$

Where $\Delta w_j$ is the adjustment to be made on the jth weight of the ith node, $\eta$ is the learning rate, $x_j$ is the value of the node, and $\delta_i$ is the error of the current node, which includes the difference between the actual and desired value of the current node, as well as the derivative of the activation function. Intuitively, this formula makes sense; if there is a large error between the output of the node and the desired output, then $\Delta w$ will be relatively large, and the larger changes will be made to the weights of the nodes which had the largest affect.

The last term of the backpropagation formula is the learning rate, which changes how large a 'step' the network takes every time it propagates. In theory, this could be a constant, however

this is generally not an efficient option; an ideal network would initially take large steps when it is first learning, and then slow down and take smaller steps towards an optimized solution as the model begins to converge. To control these learning rates, there are several other algorithms that can be used.

The first of these learning-rate algorithms is called Stochastic Gradient Descent (SGD) [57]. In this case, the gradient of the loss function is calculated to determine the vector of greatest descent, and the weights change to reflect this gradient descent. ADAM (Adaptive Moment Estimation) [58] is a similar algorithm to SGD in that it also first calculates the gradient of the loss function. The next step that the ADAM algorithm takes is to also calculate the first and second moment, which influence the amount that the weights change and helps to reach convergence with minimal increase in computational complexity. The third optimization algorithm is L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) [59], which functions somewhat differently from the other algorithms. This algorithm approximates the inverse Hessian which, similarly to the gradient descent approach, gives an indication of the direction and magnitude of change each weight should take to minimize the backpropagated loss function.

Of the various machine learning algorithms, the neural network has by far the largest hyperparameter space for user input. The number of layers, number of nodes for each layer, activation function, as well as solvers (and their related hyperparameters) all must be optimized for the task at hand.

# Chapter 3: Methods

This chapter will outline the methods used to design the device and related post-processing system, as well as the methods used to perform the validation experiments. Section 3.1 will outline the design process of the new device, from the initial concept, to experimentally determining the best components, to assembly. Section 3.2 will outline the experimental design of the validation experiments, as well as the procedure for carrying them out. Finally, Section 3.3 will include specifics about the various post-processing methods used.

## 3.1 Instrument Design Process

The overall goal of this device is to be able to lower the LOD of LFIAs (compared to manufacturer's approved method of visual interpretation) and to enable some level of quantification while remaining inexpensive enough to be an end-user device. The device has to be small enough to be handheld and battery powered, while also having enough power to run adequate active photonic elements simultaneously with the sensors and onboard computing. Another requirement was that the device would have the ability to connect wirelessly to a remote server or by Bluetooth to a smartphone to allow of Internet-of-Things interaction. This collection of requirements meant that the design would have to be highly optimized at every step of designing and testing, and that the entire system – from LFIA, to hardware, to post-processing – would need to be considered holistically. Table 4 depicts the design constrains and requirements/criteria that we formalized based on our experience and interactions with stakeholders/industry partners. The table also presents a list of our ambitions aimed at driving a paradigm shift in the LFIA industry.

**Table 4:** Summary of the guiding design constraints and requirements, as well as additional perks which would be ideal to achieve

| **Constraints** | **Requirements/Criteria** | **Ambitions** |
|---|---|---|
| Prototype must be low cost (Cost < $70 USD)<br><br>Prototype must be handheld (Volume < 1000cm$^3$) | Prototype must lower the nominal LOD by 50%<br><br>Prototype must be powered by an on-board battery<br><br>Prototype must demonstrate IoT connection | Prototype should lower LOD by an order of magnitude<br><br>Prototype should be a platform for any LFIA<br><br>Prototype should be easy to use for an inexperienced user<br><br>Prototype should integrate ML to improve performance |

It was determined very early in the project that if these requirements were to be met, there would need to be two major changes to the design when compared with the previous attempt at making a handheld device. First, the optical system would need to use much less power; switching from a laser to an LED would allow for this. This would lower the power requirements of the system and allow for it to be battery powered, as well as reduce the overall cost of the system. This benefit came at a cost; LEDs do not produce coherent light and cannot produce light at the same intensity of a laser, meaning the effective intensity reaching the LFIA would be greatly reduced. This reduction in turn means a significant reduction in the thermal signature produced, thereby decreasing the signal-to-noise ratio.

The second major change was that the thermal camera would have to exchanged for a single-element thermal sensor. This change would dramatically decrease the cost and complexity of the device however, this greatly reduced the data available for post-processing, meaning that the device would need to be immaculately optimized to maximize the value of the data being recorded. This change to the primary sensor would also later necessitate further optimizations to the post-processing system, including the introduction of machine learning algorithms in the place of conventional signal processing methods as had been used previously.

### *3.1.1 Sensor Selection*

The single-element sensor is the crucial component of the entire device, and therefore it was necessary to ensure an optimal sensor was selected before the rest of the design could continue. Form factor, data output format (ie. digital or analog), and sensor cost were all initial considerations when selecting the sensor shortlist. This set of sensors was then subjected to an experiment to determine their respective Noise-Equivalent Temperature Difference (NETD). NETD is a measure of the total noise of a sensor, converting that noise into an equivalent temperature difference, meaning that precision below the NETD is effectively noisy and not informative; an optimal sensor will have a low NETD. This NETD is also a function of the integration time (equivalently, sampling rate) of the sensor, so it was necessary to evaluate them at a range of sampling frequencies.

The exploration began with the creation of an ice-water bath to act as the constant temperature target. Each sensor had its' sampling rate set, was placed facing the target at a constant distance, and collected data for 60 seconds. Once these collections were completed for each sensor at a set of different sampling rates, the data was processed. NETD calculations are relatively simple for single-element sensors, as only temporal temperature difference is measured, and no special considerations are required. The NETD was therefore measured by taking the standard deviation of the data over the course of each experiment; this number, measured in millikelvin (mK), was indicative of the amount of noise at a given integration time. Figure 12 depicts the results of this round of experiments, making clear the trends of each sensor as the NETD increases with sampling frequency.

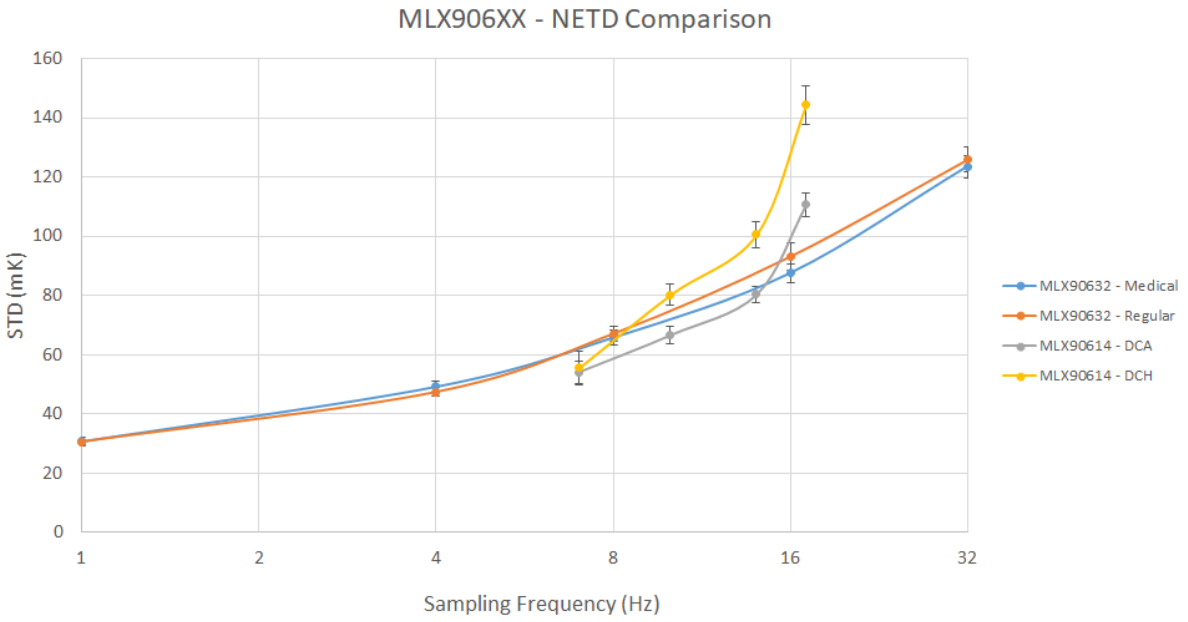



**Figure 12:** Graph of the NETD for the different sensors as a function of sampling rate. Error bars represent the standard deviation of repeated measurements on the same sensor.

Based on experience, it was initially assumed that a modulation frequency between 0.25 – 2 Hz would be used, and thus the bare minimum sampling rate required by the sensor would be 4 Hz as per the Nyquist–Shannon sampling theorem. This frequency range is justified by equation 2.1, where lower modulation frequencies are shown to increase the overall magnitude of the thermal field. Selecting a higher frequency than this would result in a lower amplitude, thereby increasing the noise of the system, and a modulation frequency that was significantly lower than this range would mean that fewer cycles could occur during a single experiment.

To ensure this criterion would always be met, as well as increase the number of samples acquired over the test time, it was decided that a sampling rate of 8Hz would be the target. As can be seen in Figure 12, the MLX90632 series of sensors performs better in this range. These sensors, while being marginally more expensive, also have a much smaller form factor when compared with the MLX90614 series, allowing for a more compact device. These sensors also communicate digitally, utilizing an Inter-Integrated Circuit ($I^2C$) protocol as opposed to a Pulse Width Modulation (PWM) schema, simplifying the process of creating their firmware and managing their interface with an Arduino.

### *3.1.2 Geometric Optimization*

With the sensor selected, it was decided that the next step would be to optimize the geometric layout of the device before moving on to selecting the active lighting. This was justified because an increase in the geometric optimization would in turn increase the effectiveness of the lighting system, allowing for a lower-powered solution to be selected. To begin, a relatively low powered (~500mW) 808nm laser was selected as the active lighting placeholder, with the dream of eventually moving onto LED lighting. The initial geometric orientation was the same as had been in previous iterations; the lighting and sensing modules are on the same side of the LFIA. A tabletop example of this setup can be seen in Figure 13, with the laser on the left shining incident on the test line directly in front of the sensor.

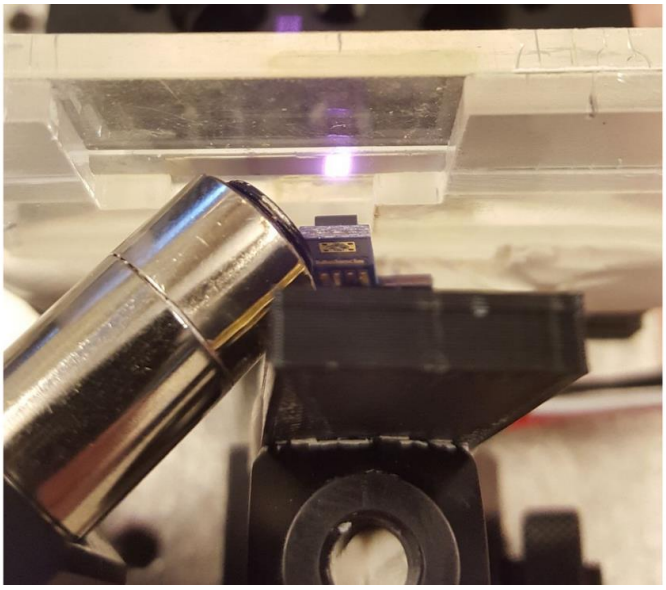


**Figure 13:** Initial tabletop setup utilizing the conventional testing orientation with an 808nm laser (appears white on camera). Note that the bulky exterior can of the laser makes it impossible to move any closer to the surface of the LFIA.

As can be seen, the laser and sensor are both oriented to face the line being tested, and when the other line must be tested, the entire optical slide is shifted to center the new line. The intensity of the infrared radiation captured by the sensor follows the inverse square law with respect to distance, as expected:

$$I \propto \left(\frac{1}{d^2}\right) \tag{3.1}$$

The power of infrared radiation received by the sensor is therefore highly influenced by the distance to the LFIA, and an increase in the power received can be granted by decreasing the distance between the sensor and the LFIA. Unfortunately, we were limited to how closely we could get to the LFIA before the size of the laser would impede on the senor's view. Figure 13 demonstrates the closest arrangement possible in this orientation.

The solution to this problem was in fundamentally changing the layout of the system from reflection-mode to transmission-mode. That is, instead of the lighting and sensor on the same side of the LFIA, the sensor was instead located behind the LFIA. This allowed for incredibly small distances between the light source, LFIA, and sensor, greatly increasing the received power and compensating for the dampening that the thermal waves experienced by passing through the PVC backing of the LFIA. Two different angles of the transmission-style setup is shown in Figure 14, take note of how much closer the sensor is to the LFIA compared to the configuration in Figure 13.

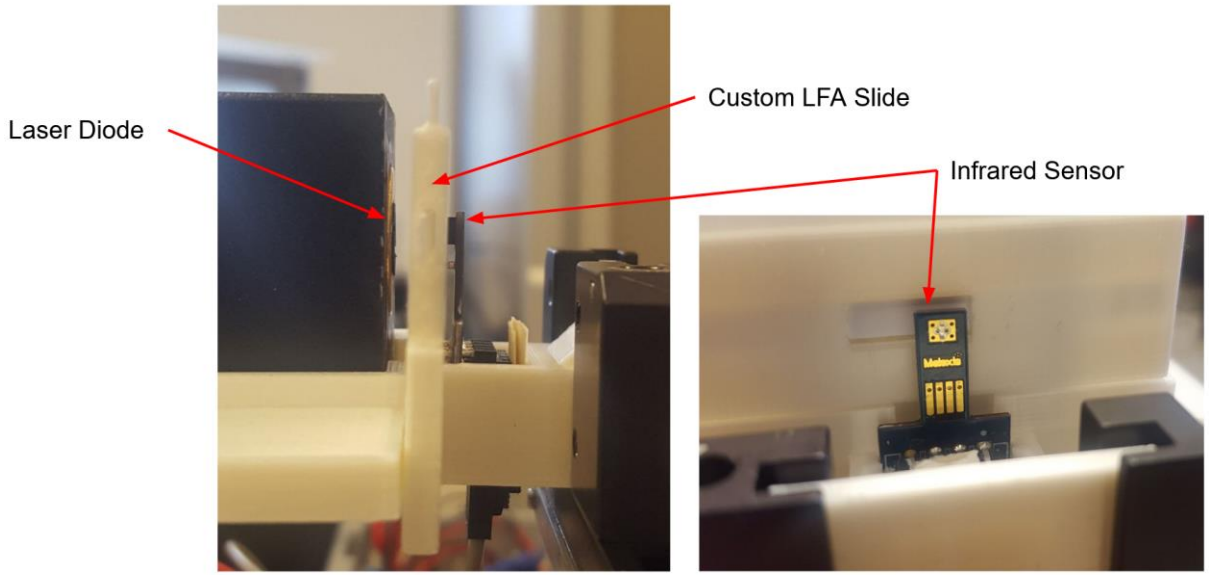



**Figure 14:** Initial tabletop setup for experimentation of the transmission style of LFIA interrogation. A custom LFA slide was required which would allow for a direct view of the backing from behind.

This transmission style of interrogation was a breakthrough in the device's development, as it not only increased the received power of infrared radiation (thereby allowing a low power LED to function in place of a laser), but it also made room for one sensor to fit over each line, allowing for simultaneous interrogation of both lines. This design had not been possible in previous

iterations, as thermal cameras have an optimal focal distance, and thus would not benefit from being held so close in proximity to the LFIA. In this way, the inherit properties of the single-element sensor was leveraged to become an overall advantage during the development of the device.

### *3.1.3 Experimental LFIA Design*

With this new transmission design, new possibilities opened in the manufacturing of the LFIAs. One of the desired outcomes for the design of this device was that the chemistry of existing LFIAs would not have to be changed for improvements in the performance to be made. There is some degree of design freedom when designing an LFIA without changing the chemistry; aspects such as the thickness of paper or specifications of the backing material could trivially be changed by a willing manufacturer with no changed to the underlying LFIA chemistry required.

To explore the possibilities and determine LFIA design changes were most likely to have a positive outcome, a model was created to simulate different material properties and their impact on the results of an experiment. This model was created in Matlab with a library called MCmatlab [60]. This library allowed for the modelling of a section of LFIA, including the backing, NC paper and embedded concentration of AuNPs, and incident lighting. A simulated collection could then be performed, and the results processed. This allowed for easily quantifiable differences in the performance of the LFIA with modification of material properties and dimensions. Two still captures from the simulation of a modulated signal incident on the LFIA is shown in Figure 16, where the red regions indicate local heating as a result of absorption.

With this model in place, a simulated annealing algorithm [61] was used to explore a range of material properties and dimensions to illuminate trends in LFIA design which could be leveraged to increase system performance. The results of this model indicated that thinner NC paper, as well as a clear backing, would theoretically improve the performance of the LFIA. Our team's partners at BTNX Inc. (an LFIA manufacturer and distributer) were key in creating a set of custom LFIAs which had clear plastic backings as opposed to the conventional, opaque backings.

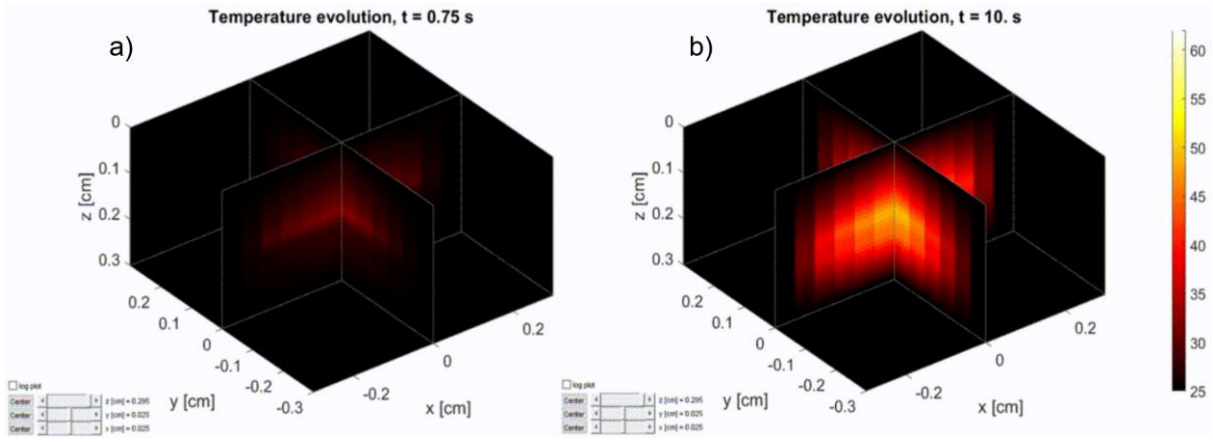



**Figure 16:** The simulated thermal response of the LFIA (a) initially, and (b) after 10 seconds. This simulation was performed with MCmatlab, a software used for performing Monte-Carlo simulations of light through turbid media.

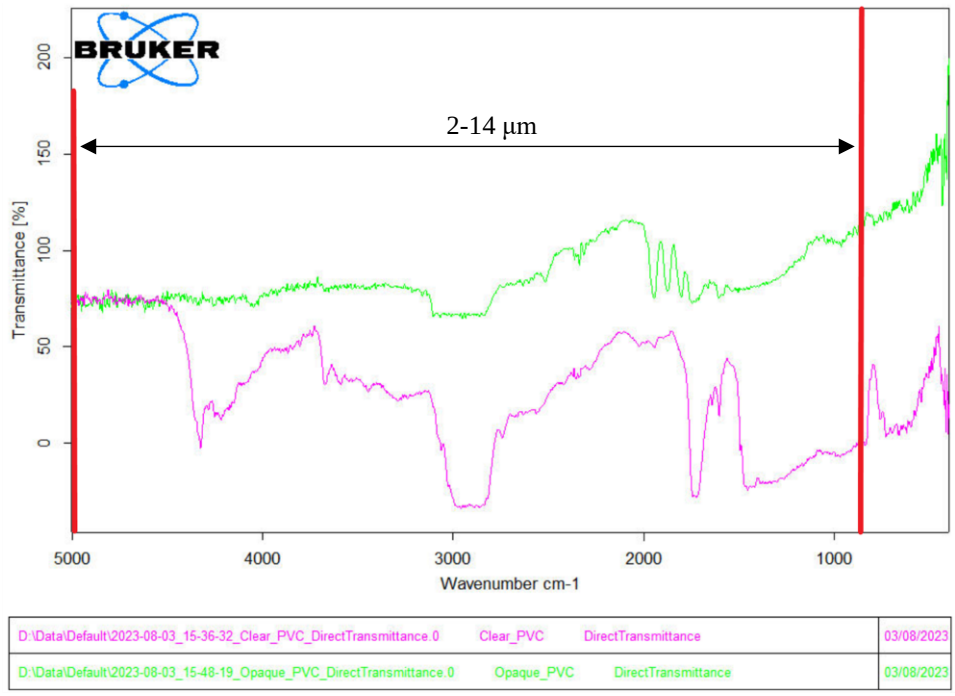



**Figure 15:** Results of the two potential PCV LFIA backings, clear (purple line) and opaque (green line). Vertical red lines indicate the sensitive range of the infrared sensor. The opaque backing clearly transmits a greater magnitude of IR.

These two types of plastic backing were subjected to the testing in a Fourier-Transform Infrared Spectrometer (FTIR), which quantified the transmissive properties of these materials at infrared wavelengths, and the results are seen in Figure 15. In the range of infrared radiation which the sensor is active, the opaque plastic backing is more transmissive than the clear backing.

We used the difference in infrared transmission to our advantage; the clear backing could face the lighting component, allowing for the incoming 532nm light to penetrate the backing and interact with the GNPs. The IR produced would not have to pass through a PVC layer to reach the sensor, and the IR released towards the PVC layer would be either absorbed (contributing to LFIA heating) or reflected towards the sensor. The orientation used in the past is illustrated in Figure 17(a) in contrast with Figure 17(d). Comparing these two illustrations, notice how the infrared radiation is reflected in one direction when a clear backing is used, rather than split between two directions as in the original configuration; this was a factor which would further increase our signal-to-noise ratio.

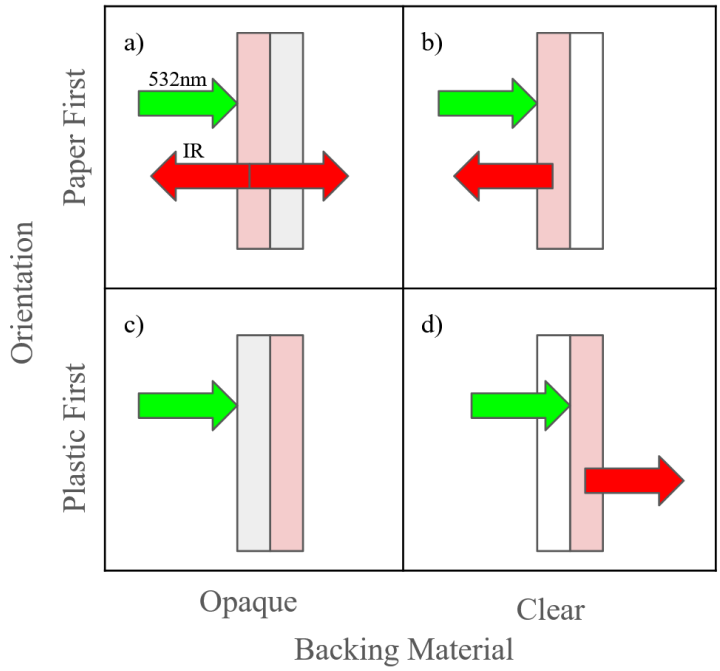



**Figure 17: a)** The 532nm light is absorbed and IR is released back towards the light, and can pass through the backing, **b)** 532nm light is absorbed and IR is released but blocked from passing through the clear backing, **c)** The opaque backing blocks 532nm light from being absorbed**, d)** 532nm light passes through clear backing, IR is released and blocked from returning back through the backing

The off the shelf LFIA was compared experimentally to the custom LFIAs in transmission mode, and it was found that the thermal waves detected were between 4-8 times the received power of the regular LFIAs. This increase in power directly lead to an increase in the precision at which the concentration of AuNPs on the lines of an LFIA can be detected. Therefore, with a minimal impact on the manufacturing process, the performance of the device and LFIA system could be greatly increased.

### *3.1.4 Lighting Optimization*

With the optimization of the geometry of the system and all the benefits it brings in place, the active lighting of the system could then be optimized. It is known that LSPR is the driving factor behind the heat and infrared radiation generation of the suspended GNPs, and the most effective wavelengths have been experimentally determined previously. These data are graphed in Figure 18, where the reflectance of the white paper is compared to the reflectance of the control line – considered to be "saturated" with AuNPs. The green line highlights the difference in reflectance as a function of wavelength. We notice immediately that the peak is near the 532nm wavelength.

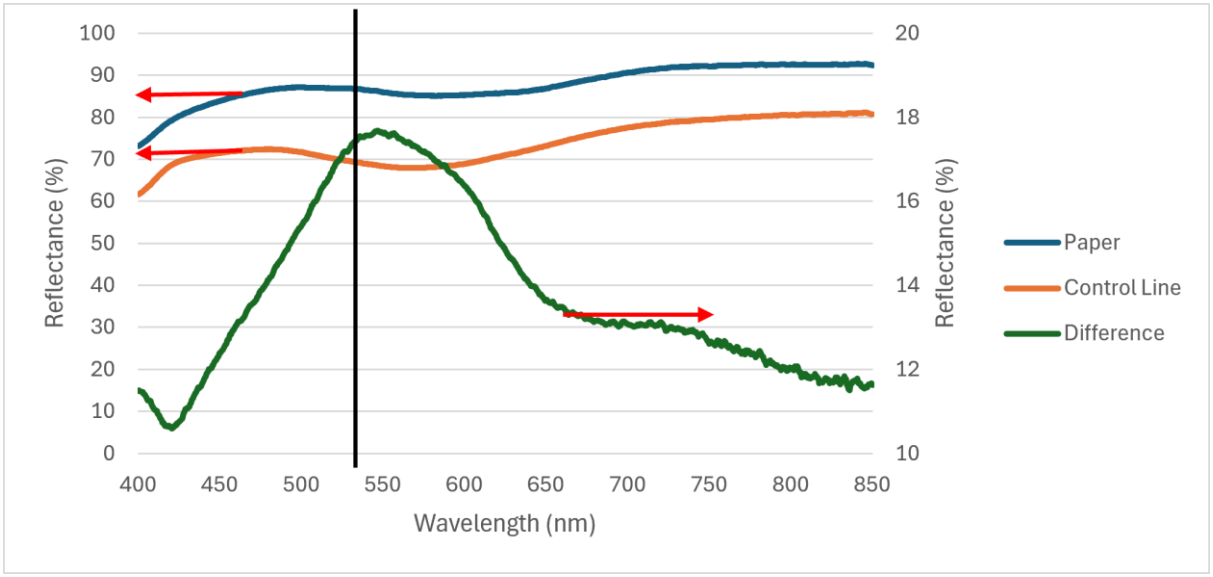



**Figure 18:** A graph illustrates the difference in reflectance between the white NC paper and the dark control line. It is assumed that if light is not reflected, it is absorbed, and therefore an area of high reflective difference between NC paper and the control line is desirable. The vertical black line approximates the 532nm wavelength, which roughly maximizes this difference

The peak absorption wavelength for this LFIA is around 530nm. Pragmatically, selecting an active lighting system is always a trade-off between cost, effective wavelength, and power of the lighting system. For example, in many previous cases an 808nm laser is used, even though it is not at the peak absorption wavelength. This is primarily because 808nm lasers can generally be much more powerful than 532nm lasers for the same cost, so the effective energy absorbed by the AuNPs was higher with 808nm lasers in our previous designs.

Several different active lighting options were investigated, with the goal being an LED to replace the laser which has been used historically. The reason for this change is because LEDs have a

much lower power consumption when compared with lasers, with the drawback that their beam is not collimated in the same way that a laser is. In general, this would be an issue when trying to focus the beam on the lines of the LFIA, however, because of the transmissive layout of the current design this was a non-issue, as the LED could be located less than a millimeter from the LFIA, making collimation effectively irrelevant.

### *3.1.5 Device Design*

At this time development was far enough along that it was time to migrate from the tabletop to a prototype device. This step required custom designed Printed Circuit Boards (PCBs) and housings to be designed and manufactured. The initial functional diagram can be seen below in Figure 19, with red indicating the flow of power and blue the flow of data.

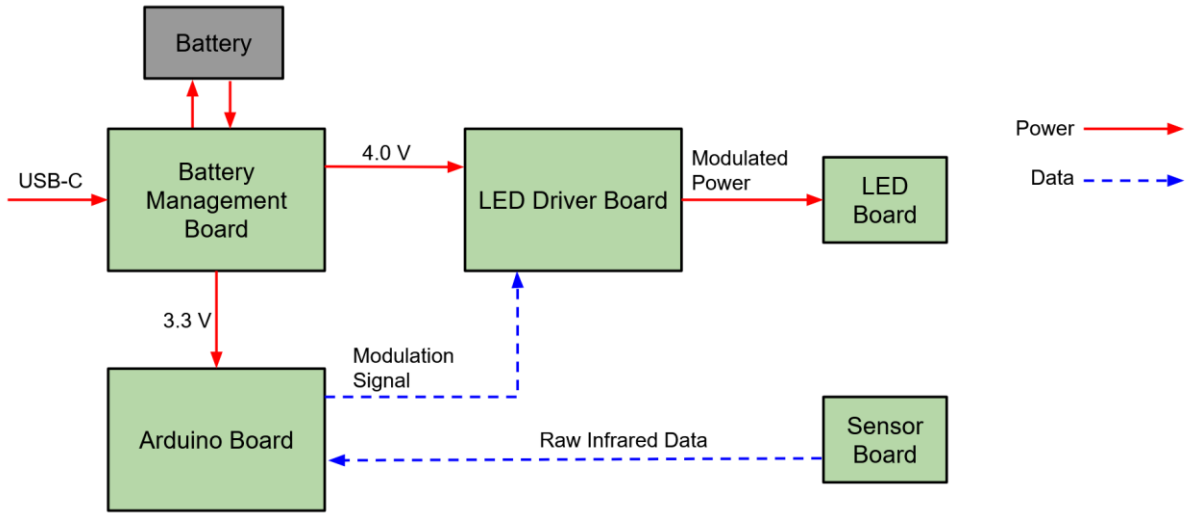



**Figure 19:** Functional diagram of the device, illustrating each PCB and their respective functions. The red arrows indicate the flow of electrical power, while blue indicates data transfer.

The simplest board to design was the one to house the Arduino, which can be seen in Figure 20. This allowed for power connections to be made between the Arduino and the battery management board, supplying its required 3.3V from either the battery or directly from the wall plug. The Arduino also read the incoming data from the two sensors via an $I^2C$ connection at the pre-determined sampling rate of 8 Hz. A modulation signal was also generated on demand with the press of a button. This modulation signal would be at a single frequency or a range of

frequencies to enable chirp or frequency keying techniques to be performed, as determined by the code. The last function of the Arduino was to upload the collected data to an external computer; this could be performed either over a USB connection or via Wi-Fi to a server.

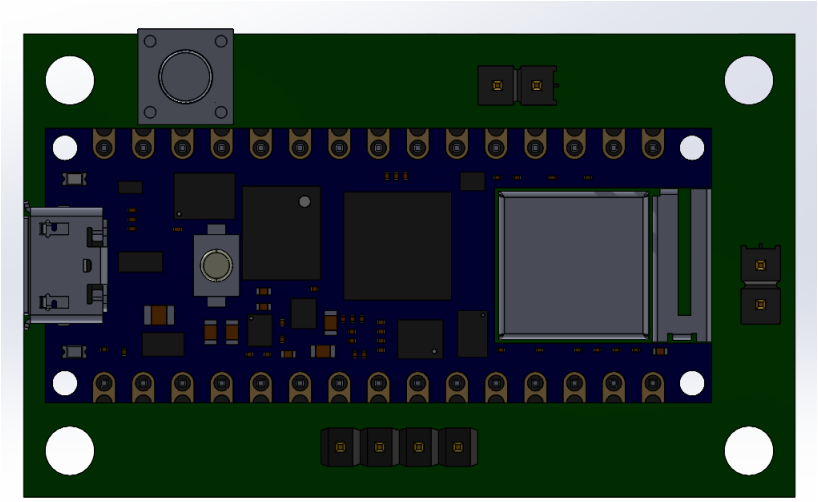


**Figure 20:** Arduino RP 2040 mounted on the custom PCB. Mounted on the board is an external button to initiate collection.

Figure 21 shows the battery management board regulated the charging of the on-board battery, indicating via an external LED when the battery needed charging. A switch on the side of the housing allows the user to change between external and battery power. The board would also receive external power via a USB-C connection and regulate the power to the Arduino and LED Driver board.

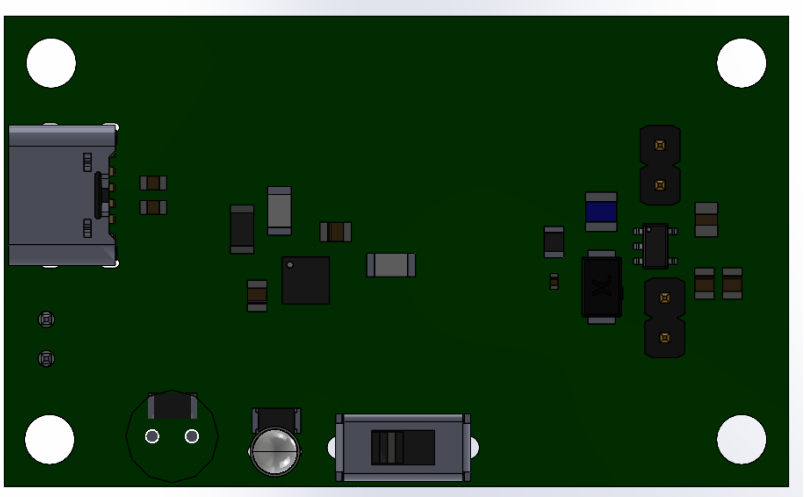


**Figure 21:** Custom battery charging/power control board. Allows for a USB-C connection, and a switch for controlling power source.

The LED Driver board illustrated in Figure 22 takes power from the Battery regulation board, and a data signal from the Arduino, to generate a powered signal sent to the two LEDs on the LED board. Also included is a manual potentiometer so that the power output of the LEDs could be adjusted for different experiments.

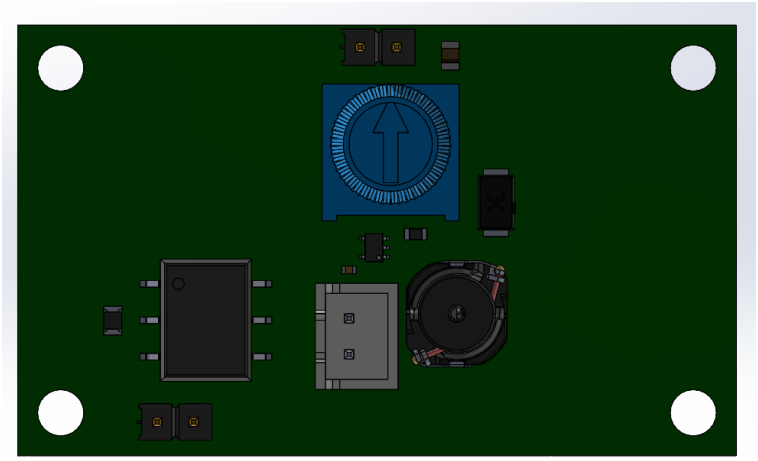


**Figure 22:** LED control and modulation board. The manual potentiometer allowed for some control over light intensity.

Figure 23 shows the LED and sensor boards. On the LED board the LEDs are arranged in series with one another, so that the illuminated output power is always identical (assuming no variations in the LEDs themselves). The final custom PCB is the Sensor board, which houses the two MLX90632 surface-mounted sensors. These sensors communicate via $I^2C$ connection to the Arduino. Just as the LEDs on their board, these sensors are spaced to specifically match the spacing of the LFIA to be tested. Because these sensors communicate over $I^2C$, it is trivial to address another, allowing for simultaneous testing of the multiple test lines of a multiplexed LFIA in subsequent iterations of this device.

**Figure 23:** LED board (left) and sensor board (right).

These PCBs were arranged in 3D printed housings, as seen in the rendering of Figure 24. There are two main sections of housing: the sensor and external housing. The sensor housing holds the senor and LED boards facing one another. The slot in the sensor housing allows for a specialty-printed LFIA holder to be slipped between the two boards with sub-millimeter clearance, maximizing illuminance. This sensor housing also has standouts on top to allow for mounting the LED Driver board. The external housing has standouts for mounting the Arduino and battery management boards, with the lid containing mounting brackets for the battery. The sensor housing is also printed in a white Polylactic Acid (PLA), while the external housing is in a black PLA. This allows for the 532nm light which is not absorbed by the AuNPs to transmit through the sensor enclosure and be absorbed by the external housing, effectively keeping any errant heat signal from influencing the sensors.

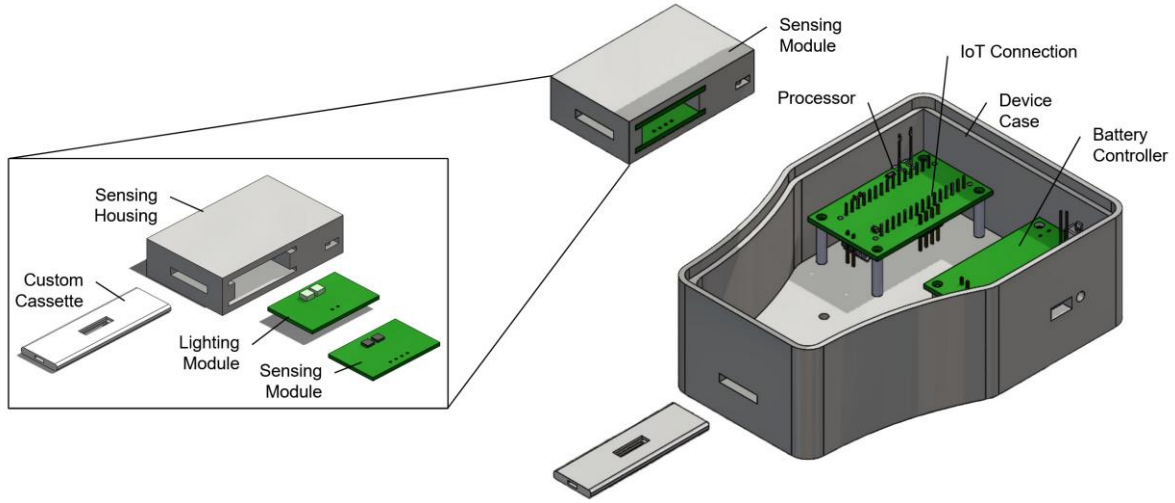



**Figure 24:** Rendering of the final device. The sensing module is housed inside the larger device casing, allowing for different sensing modules to be swapped out when interrogating an LFIA with different dimensions without requiring an entirely new device.

### 3.2 Materials and Samples

After the prototype instrument had been assembled, it was necessary to validate its function in a variety of different experiments. Successfully decreasing the LOD and enabling quantification in a variety of different tasks with different LFIAs would increase the confidence that the device has been well optimized and is on the right path towards an end-user product.

A series of three different validation studies were planned, each to test different aspects of the device as well as the LFIAs themselves. The first experiment would use the specially designed

LFIAs, which test for THC in lab-created artificial saliva samples. The second experiment would use off-the-shelf LFIAs for testing COVID-19 antibodies in artificial blood serum. Finally, a human study would be performed to test for THC in human saliva samples using the same LFIAs as in the first experiment.

### *3.2.1 COVID-19 IgG and IgM*

Immunoglobulins G and M (IgG and IgM respectively) are both produced by the body in response to an infection. For COVID-19, it is common that IgM is produced within a week of infection, while IgG takes longer to accumulate [62]. These compounds can be found in blood samples, and their presence can be detected and measured to determine the body's response to a COVID-19 infection; a need which has greatly increased in demand over the past couple of years.

### *3.2.2 Tetrahydrocannabinol (THC)*

Parent THC refers to a collection of different cannabinoids, but primarily tetrahydrocannabinol. When the body metabolizes this parent THC, several metabolites such as $\Delta^9$-THC and 11-hydroxy-$\Delta^9$-THC are generated which can be tested for in various ways depending on the purpose of the test. An aspect which makes testing cannabinoids particularly challenging is that many of these compounds are cross-reactive with one another, meaning that the presence of one compound can trigger a positive response in an experiment which is testing for another. This fact, combined with the fact that parent THC is a very difficult to use, means that it is industry standard to use a THC metabolite in place of parent THC when preparing tests. Because the rate of cross-reactivity is known, this process can be used even to calibrate the reactivity of LFIAs.

### *3.2.3 COVID-19 Antibodies - Lab Preparation*

To validate this device, a series of LFIAs had to be prepared with known concentrations. The LFIA selected for this validation study was for testing for COVID-19 antigens and antibodies in blood serum. These LFIAs were provided by BTNX Inc., and the control positive IgM/IgG solution was sourced from MyBioSource (*MBS355896*). A series of solutions were prepared by the process of serial dilution, with Phosphate Buffered Saline (PBS) used as the diluting solution. A total of 8 solutions, including one control negative containing no control positive solution, were prepared as per the manufacturer's instructions, with a range of concentrations between 0-10 µg/mL. Each of these solutions were then used to prepare 12 LFIAs, leaving a total of 96 prepared LFIAs. Examples of these LFIAs at each solution can be seen in Figure 25; these are sandwich style LFIAs, therefore a darker response indicates positive.

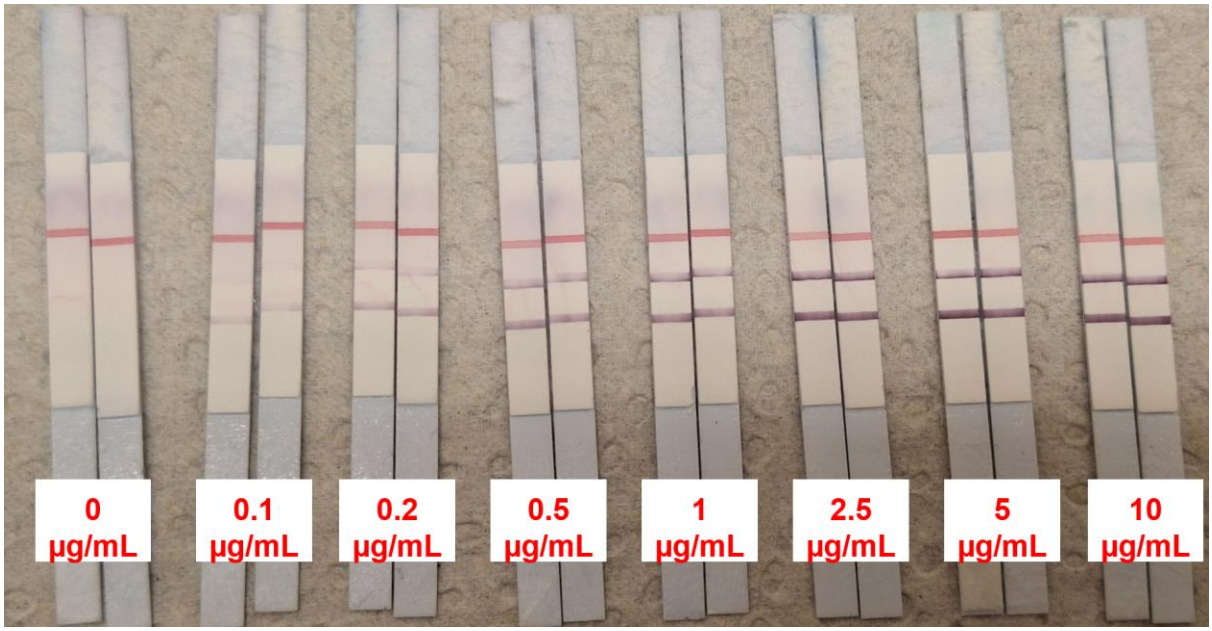



**Figure 25:** Representative samples of the prepared COVID-19 LFIAs.

Each LFIA was then repeatedly interrogated by the device three times. Interrogation lasted 60 seconds, while the sensor took measurements at a sampling rate of 8 Hz, giving a total of 480 total samples per line. The modulation frequency was set to a constant 0.5 Hz, a value which was known to be near-optimal from previous optimization experiments. This testing resulted in a total of 576 data sets ($96\ LFIAs * 2\ lines * 3\ repetitions = 576$). This process was repeated for each line, providing with a total of 576 data sets for each of the IgG and IgM lines.

### *3.2.4 THC – Lab Preparation*

As previously mentioned, our partners at BTNX created a custom, clear-PVC backed set of LFIAs to test for the Parent THC content in saliva. These new LFIAs were first to be tested in a lab setting before being used in a human validation study. LFIAs which are made to test for the presence of Parent THC often have a linear cross-reactivity with other THC metabolites such as $\Delta^9$-THC. Parent THC is a difficult substance to work with, and so it is industry standard to calibrate these LFIAs with such THC metabolites instead.

These clear-backed LFIAs were specified to have a cutoff concentration of 15 ng/mL of $\Delta^9$-THC. A solution containing 1.0 mg/mL of $\Delta^9$-THC was acquired from Millipore Sigma (*T-005*), and a serial dilution was performed using an artificial saliva solution to produce concentrations of 50, 25, 10, 7.5, 5, 2.5, 1, 0.5, and 0 ng/mL. An example of these LFIAs can be seen below in Figure 26(b). Note that these are competitive style LFIAs, which means that as the concentration increases, the test line will become lighter.

A total of five LFIAs were prepared at each concentration, leaving a total of 45 LFIAs. After these LFIAs were prepared, they were interrogated in a fashion very similar to that of the COVID-19 LFIAs; 60 second collections at a sampling rate of 8 Hz providing a total of 480 data points per line. These LFIAs were repeatedly tested a total of five times. This delivered a total of 450 data sets.

### *3.2.5 THC – Human Study*

A human study was also designed to use the LFIAs from the previous section to test for the presence of Parent THC in human saliva samples. This study, which was previously approved by York University Ethic's Protocol (*Approval Certificate #e2019-267*), invited participants that were familiar with cannabis use. These participants would provide a series of saliva samples, using a set of industry-standard saliva collectors, shown in Figure 26(a).

An initial collection was performed, and then the participant was asked to consume cannabis in the form of a joint. After 30 minutes, and then every 30 minutes after that, another saliva collection was performed until a total of 6 collections had been attempted. With each saliva sample, up to two LFIAs were prepared as per the manufacturer's specifications. The rest of each

saliva sample was then sealed and kept refrigerated. Once the test was complete, the collected samples were sent to a testing facility so that the actual concentration of THC and related metabolites could be measured through mass spectrometry. The LFIAs were allowed to dry for 24 hours before being interrogated, each being tested a total of five times.


**Figure 26: a)** Saliva collector apparatus given to participants. When the dye in the handle runs red the sample has been adequately collected and **b)** A representative collection of LFIAs prepared with human samples collected periodically (from left to right). Note that these are competitive style LFIAs, so a dark test line indicates a negative test

A total of 22 participants provided a set of samples for this experiment. Many of these sample sets contained less than the maximum 6 samples, as participants often had an issue producing enough saliva for a full sample. As well, the final four participants' samples were not tested as the samples were lost in transit and spoiled. This left a total of 18 participants, plus a set of control negative samples. Most of these samples were used to prepare two LFIAs each, however if the sample was small, occasionally only a single LFIA was prepared. In total, 187 LFIAs were prepared and tested, each LFIA being tested a repeated 5 times, granting 935 labelled data points.

## 3.3 Data Post Processing

All the data collected for this validation process have known associated concentrations, either because the LFIA was prepared using solutions with known concentrations, or in the case of human data, because the sample was subsequently tested with the ‘golden standard’ of mass spectrometry to measure the concentration of analyte. This means that each prediction made can be compared to the known result, and the performance of the device and model can be measured.

An overview of the data-processing steps for both classical signal processing and machine learning methods can be found in Chapter 2, here we will review only the specifics of each method used to process the raw data.

### *3.3.1 Classical Signal Processing*

This classical signal processing algorithm has been adapted from previous iterations of the device which used thermal cameras as their sensing apparatus. From the perspective of this algorithm, the data input from a single element sensor is a thermal camera with a single pixel. This meant that the algorithm could be greatly simplified, while maintaining the same general processing steps.

For this classical signal processing, only a single feature is considered: the intensity of the thermal wave at the modulation frequency. This is always the most informative feature, as the entire design of the device has been focused on this feature. Once this has been extracted a linear separation algorithm is performed on the training dataset.

The initial step was resampling the raw data. This resampling ensures that all datasets are uniform in their sampling timing, which can be somewhat variable in the raw dataset. This resampling also ensures that after the Fourier Transform is performed, the value at the modulation frequency will be exactly calculated. Generally, resampling is done to 100 millisecond time steps, leaving a resampled dataset of 600 points.

The next step is to remove the bulk heating factor of the signal. This step is performed by fitting and subtracting a 7$^{th}$ degree polynomial from the resampled data. A polynomial degree of 7 was selected because a lower degree does not fully capture the full bulk heating effect, but a higher degree tends to leave larger artifacts in the leftover data; artifacts are left by a 7$^{th}$ degree polynomial, but these effects are minimal. These two steps are illustrated in Figure 27(a) and (b).

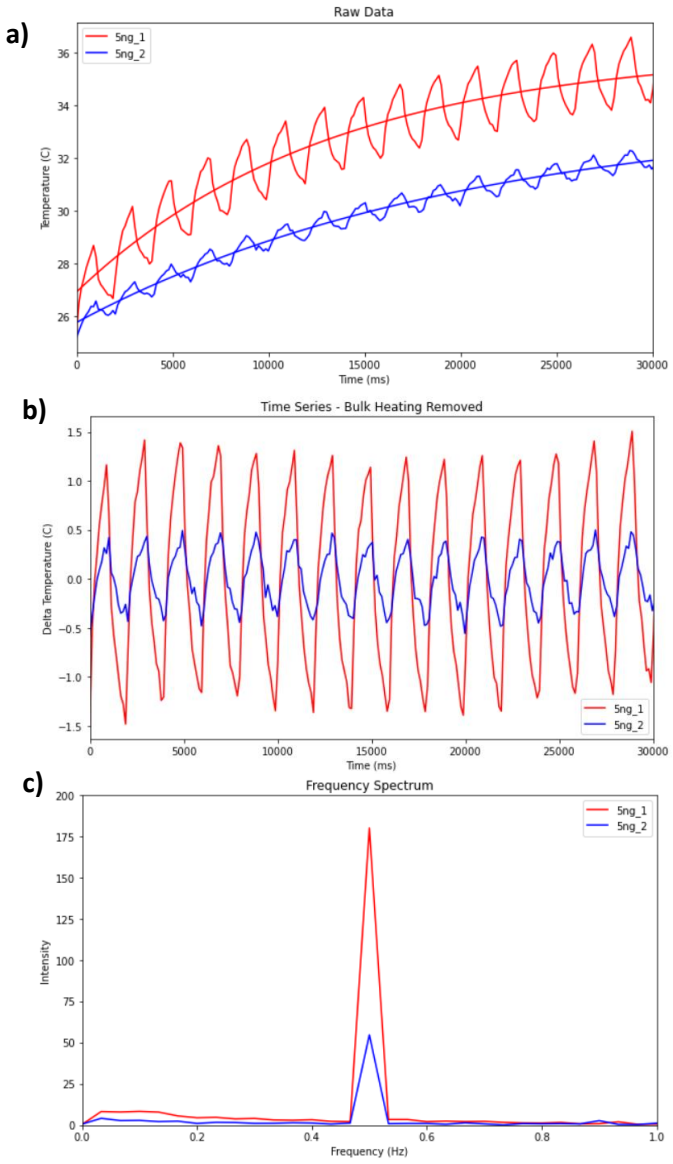



**Figure 27:** Classical signal processing being performed on an example data set. **a)** Fitting a polynomial to the raw data, **b)** the data after the bulk heating data has been subtracted, and **c)** the peak at 0.5Hz after the resulting periodic data is transformed to the frequency domain

The remaining data is a mostly periodic signal, which can trivially be transformed to the frequency domain through an FFT. Once transformed, the only value considered is the intensity at the modulation frequency; this value – referred to as the thermal wave intensity – becomes the singular feature for the line being tested. Importantly, for each LFIA there are at least two lines; the test and the control line, each with their own thermal wave intensity. The final step is to normalize the intensity of the test line to that of the control line. This step is done by simply dividing the test line value by the control line, nominally leaving a single normalized thermal wave intensity with a magnitude between 0 and 1. Ideally, the darker the test line, the closer to 1 this value will approach.

This process has reduced the raw data into a set of labelled, singular features. This dataset is then separated into training and test sets, and the training data is used to create a linear prediction model. This model is generated by separating the data into bins by concentration. Then, the mean normalized thermal wave intensity of each bin is calculated. Finally, the midpoint between the means of two subsequent bins is calculated and saved as the model's cutoff intensity for that bin. This is represented visually in Figure 28, with the vertical line representing the cutoff concentration and the horizontal line the model's separation between the groups.

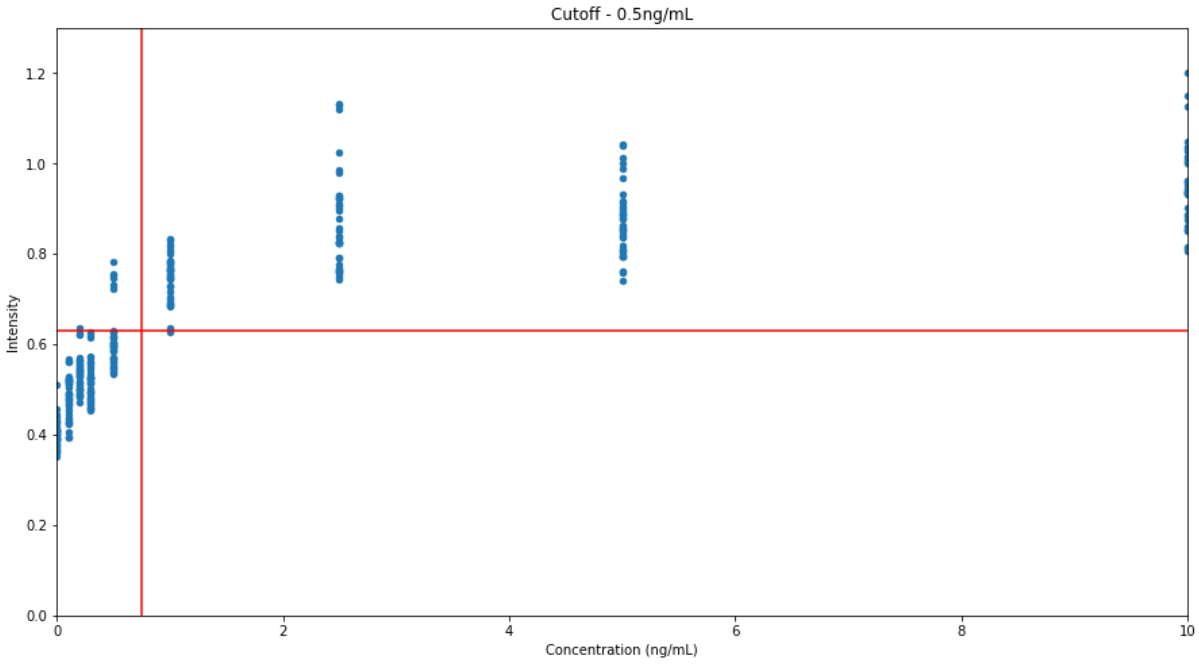



**Figure 28:** Example data being linearly separated by a conventional regression process.

Once this has been performed for each bin, the predictive model is complete, and the test data can be evaluated. The model can then quantify an input datum by fitting it in an intensity bin and reporting the corresponding concentration as its' prediction.

### *3.3.2 Machine Learning*

To contrast with classical signal processing to further increase the performance of the device, several different machine learning approaches were attempted. These approaches made use of many more features in the data than just the thermal wave intensity, however this feature remained the most informative feature by far.

These machine learning methods began with the same resampling step as in the classical signal processing approach. Once re-sampled, these datasets were separated by a process of Singular Spectrum Analysis [63], separating the signal into a trend, periodic, and residual signal. This meant that the bulk heating and periodic signal could be considered separately and the residual signal (ie. noise) could be discarded.

The periodic signal had its' thermal wave intensity extracted using the same method as in classical processing by way of an FFT. Features were also extracted from the bulk heating data, unlike in the classical algorithm. These features included the initial temperature, final temperature, and three indexes. These indices are calculated by normalizing the bulk heating signal to values between 0 and 1, and then selecting the indices where it passes through 0.25, 0.5, and 0.75. These indices relate to how quickly the temperature rises, which is a feature related generally to the concentration of AuNPs on the line. Another feature is generated by interpolating between this normalized bulk heating curve and a diagonal; higher values also indicate faster initial heating. When processed in this way, each individual line produces a total of 7 features. Standard thermal wave intensity normalization is then performed, and the two sets of features are combined, generating a set of 15 features for each LFIA interrogation.

A feature reduction step was then performed to reduce the number of features (and therefore the computational complexity of training the model) without too negatively influencing the outcome of the model. This process was usually carried out through a Mutual Information analysis [64], however a Recursive Feature Elimination [65] analysis was also performed to confirm.

The final datasets could then be split into their training and testing sets and used to train a model. Several machine learning models were selected for training, and their exact details will be discussed in Chapter 4, however each model went through a similar optimization process. For many machine learning models, there are a set of hyperparameters which influence the performance of the model given a particular dataset. These hyperparameters need to be tuned for a specific use, and this was done through an exhaustive grid search of potential hyperparameters. A model and ranges for each hyperparameter were first defined, and the model is trained iteratively with each combination of hyperparameter on the testing and training data until the best performance is determined. This search is exhaustive, and therefore computationally taxing, however for the relatively small number of samples (~1000 samples at maximum) and small number of features per sample, these models were quickly trained. For this reason, the more exhaustive search was selected, as in the grand scheme it did not add much more time to the overall process of training.

Once a given machine learning model had been tuned, it could be used to make predictions about the test data. These evaluations could then be used to judge the efficacy of the model, especially when compared to the conventional model.

# Chapter 4: Results and Discussion

The following chapter outlines the results of the optimization process, as well as the three most crucial validation experiments carried out over the duration of this project.

The strategy for design and validation had four steps:

1) An initial iterative design process to design and optimize the system; this was a series of dozens of small-scale experiments as adjustments were made
2) A sanity-check experiment with results that could be verified by comparing to results from the previous devices
3) A large-scale experiment, utilizing a larger sample size of lab-prepared LFIAs to further increase confidence in the system
4) An experiment with human participants, increasing confidence that the device works despite all the added complexities that come along with human factors

## 4.1 Device Optimization Process

The process of system optimization began approximately after sensor selection had taken place and lasted until the design freeze shortly before the experiment outlined in Section 4.2. This process included all aspects of the physical device (ie. component selection, geometry, orientations, etc.), as well as the post-processing software and algorithm. A part of this process is also represented by the design of custom LFIAs, but this is explained in Section 3.1.

One of the first aspects of the system to design and optimize was the software and algorithm. This was based on previous iterations [18,43] of the system, although greatly simplified to work with a single-element sensor rather than the data produced by a thermal camera. The software was written using Python, making it easy to connect over USB to the Arduino. Python also has many libraries for performing signal processing and machine learning, greatly expediting the creation and comparison of several different post-processing methods. The software suite also allowed for the simulation of an external server which the Arduino could remotely connect to

and upload data through Wi-Fi. This served as a proof-of-concept that the device could easily upload the collected data to a cloud service for data collection and aggregation.

The hardware of the device underwent many iterations as different parameters were optimized in tandem. Some physical parameters which had to be optimized and questions which had to be answered included:

- Selection of the lighting source: What wavelength, power draw, size, and type of lighting component will produce the highest response?
- Selection of additional components: Does adding additional focusing optics increase the received signal? If so, which lenses work best at the relevant wavelengths? How large, and at what distance should these lenses be placed? Can they be modified to further decrease the form factor of the device?
- Geometry of the components: What angle should the sensor and light source be at with respect to the LFIA? How close should these components be to the LFIA? How close can they be to one another before ambient heat influences the measurements? Is it possible to put the sensor and lighting source on opposite sides of the LFIA?
- Modulation Frequency: What frequency works best? Does this change between LFIAs? Is there any benefit to performing frequency sweeps or chirps compared to a single frequency?
- LFIA geometry: An LFIA has at least two lines, does each line need/can each line fit its own dedicated sensor and lighting component? Is it possible to have one sensor-lighting pair and move the LFIA to test the lines independently? Will this movement increase the variability of response? Can the additional monetary expense and power draw of two sensor-lighting pairs be justified?

The process of answering these questions and optimizing this design was iterative and time consuming. One parameter at a time would be varied until an optimal value was found, then that parameter would be held constant as another was improved. Initially, this was all performed on a tabletop optical table so that modifications could easily be made. Once a sufficient design was

settled on a handheld enclosure had to be designed and 3D printed. This process itself took several iterations before becoming fully functional.

In the end, optimal parameters were found: a 532nm LED proved to provide enough power to generate a clear thermal wave, but only if situated within a millimetre from the LFIA. The optimal orientation ended up being the sensor and LED on either side of the LFIA so that they could be as close as possible, far outperforming any orientation which included a focusing optical element for either the driving light source or thermal sensor. The optimal modulation frequency was approximately 0.5 Hz, higher frequencies resulted in lower thermal wave amplitude and lower frequencies meant the test length would have to be extended to accommodate long light cycles. It was also determined that even a sub-millimetre movement of the line of the LFIA in the light source would greatly impact the results, and therefore the additional cost of adding a sensor and light source for each line was a justified increase in complexity and cost.

Overall, this iterative process resulted in a relatively high-performing handheld device with IoT capabilities that could be powered by only an onboard battery. The design was frozen, and it was determined that the first validation experiment should be carried out.

### 4.2 THC Lab Experiment

The first step towards prototype validation was to compare the results from the prototype to a known benchmark. Previous iterations of the device were validated using LFIAs which test for THC in saliva samples; performing a similar experiment yielded results which can reasonably be compared to determine if the performance of the current prototype is adequate. Identical LFIAs were not used - this experiment made use of the custom-made clear-backed LFIAs described in Chapter 3 – so this experiment is not a direct comparison to previous results, but this experiment still acts as a reasonable initial benchmark for the status of the prototype.

The experiment begins with a set of 45 lab-prepared THC LFIAs contained a total of 9 different concentrations (including the control negative), with 5 LFIAs prepared at each concentration. Each of the 45 LFIAs were measured 5 times, and the resulting data were processed using the

conventional data processing method as machine learning methods had not yet been developed at the time of these initial experiments. These results can be seen in Figure 29.

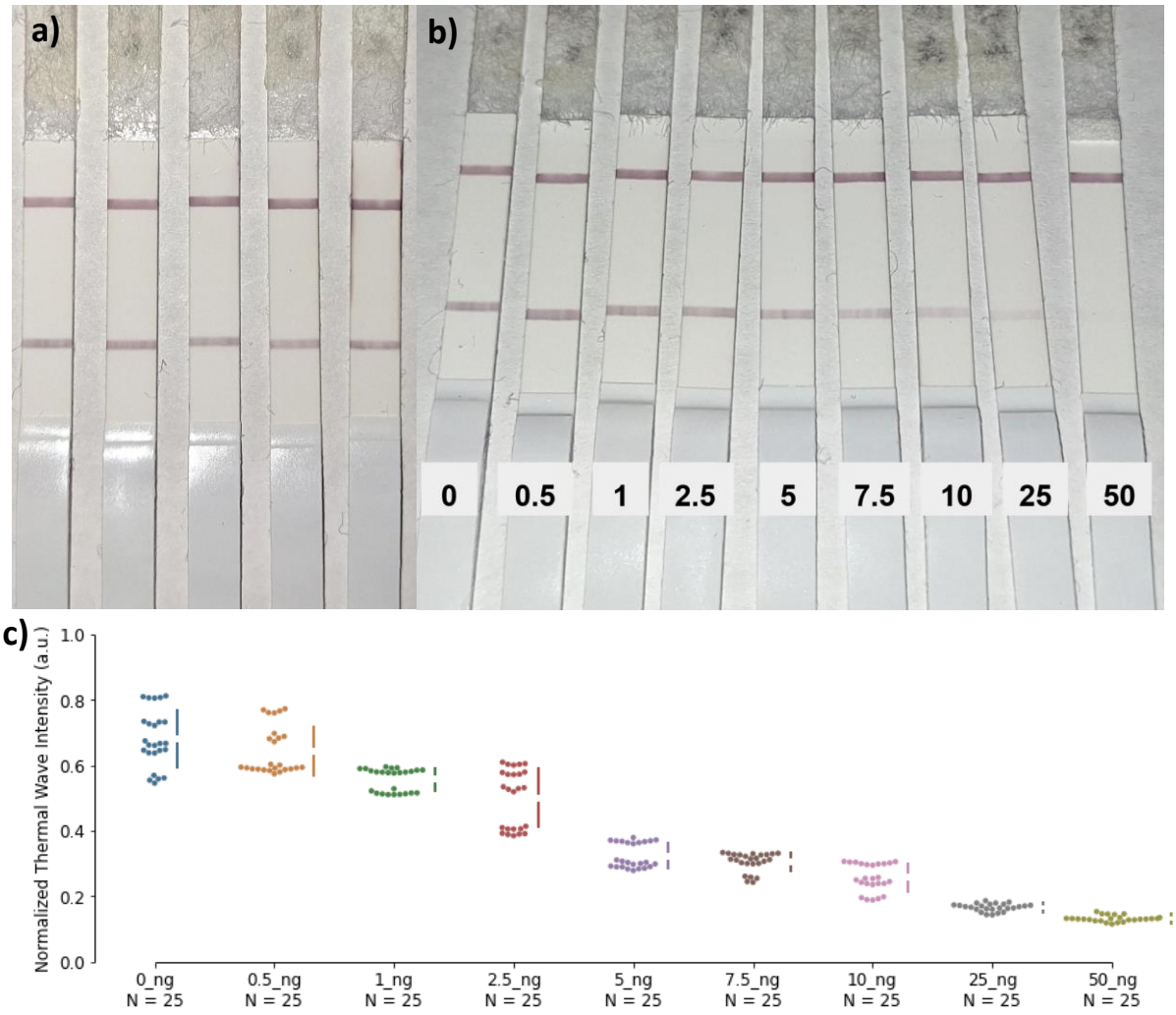



**Figure 29: a)** The five LFIAs prepared with control negative solution **b)** A representative LFIA from each of the concentrations, measured in ng/mL, and **c)** Individual responses from the lab-prepared THC LFIA dataset. Each colour represents its' corresponding concentration, and each point is a separate measurement. The clusters in each concentration result from testing the same LFIA multiple (5) times.

There are several trends that become apparent in Figure 29(c). The first is that, generally, as the concentration of THC increases, the overall signal decreases; this is expected behaviour for a competitive assay such as the these. As the concentration of THC increases, the competitive assays' test lines fade, indicating lower concentrations of AuNPs. This lower concentration means a decrease in the absorbed modulated signal, resulting in a lower measured thermal wave intensity.

One also notices that the spread of response from different LFIAs of the same concentration tends to decrease as the concentration increases. This becomes especially apparent at concentrations above 10 ng/mL, where the spread is very tight. This is because the LFIA's manufacturer recommended cutoff concentration is 15 ng/mL and doping the LFIA with concentrations which greatly exceed this cutoff result in a saturated response. Lower concentration produces a much less saturated response, which increases the variability.

The most important trend to note in Figure 29(c) is how the data is grouped in each concentration. The control negative (0 ng/mL) is the best example; there are clearly five different groupings of five data points each. Consider the top five points in the 0 ng/mL category; these are five repeated measured of the same LFIA by the prototype. The fact that there is so little variation in this group indicates that the prototype is highly optimized for repeatability. Now consider the four other point-clusters, each representing another LFIA doped with the same control negative solution as the first LFIA. The variability *between* LFIAs is far greater than the variability produced by repeated measures; in other words, the experiment is highly repeatable, but variable when reproducing the experiment. This indicates that the device has been very well optimized, to the point that the variability produced by the LFIA itself has a far greater impact than any variability produced by the device itself. Further optimization of the device will not significantly increase the quality of the results, LFIAs must be improved if better results are desired.

Figure 30 compares the results of previous devices (panel a) when testing THC LFIAs to this current experiment (panel b). Both Figure 30 (a) and (b) show measured variations of the normalized amplitude responses as a function of concentration of THC ±95% confidence intervals. The results of the previous generation of our innovation demonstrate statistically significant separations between 0, 2, and 5 ng/mL concentrations because the confidence intervals do not overlap. Similarly in our experiment, we see no overlap between 0 and 1 ng/mL, and an extremely significant distance between 1 ng/mL and 5 ng/mL. The previous device was also able to separate 5, 7.5, and 10 ng/mL categories, whereas this prototype can only differentiate between 5 and 10 ng/mL. This may be due to differences in the performance of the device, or it could be due to differences in the LFIAs themselves.

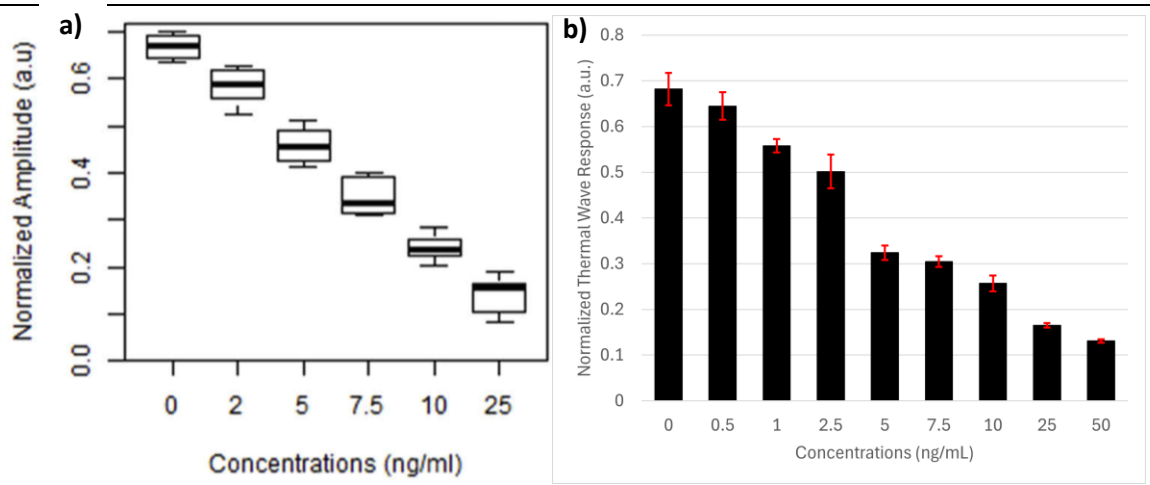



**Figure 30: a)** Line plot including the 95% confidence intervals of a previous system [18] when testing lab-prepared THC LFIAs and **b)** a line plot representation of our data with 95% confidence intervals. In both plots, concentrations with intervals which do not overlap one another are significantly different categories, and therefore sorting LFIAs into these discrete categories should be a trivial task. Figure (a) reprinted with permission from [18] © Optical Society of America

Overall, this experiment demonstrated that the device had been very well optimized because, under very similar circumstances, this prototype was able to resolve 1 ng/mL from 0 ng/mL, just as the previous, more expensive device was able to, compared to the manufacturer cutoff concentration of 40ng/mL. The variance produced by the device itself was shown to be orders of magnitude lower than that produced by the LFIAs. It was also demonstrated that, the 0 ng/mL category is separable from the 1 ng/mL. This sort of performance from an LFIA which was designed with a cutoff concentration of 40 ng/mL is a significant improvement. The results of this experiment inspired enough confidence to continue with larger scale experiments to more precisely validate this prototype.

### 4.3 COVID-19 Experiment

Another experiment was designed to more rigorously validate the performance of the prototype. For this experiment, LFIAs used to test for COVID-19 antibodies were used instead of THC. These assays are sandwich style, rather than competitive like the THC LFIAs, allowing for the performance of the device to be measured with each style. These LFIAs were also off-the-shelf LFIAs, giving us an idea of how well the prototype works with standard LFIAs with the opaque plastic backing. For this experiment, a much greater number of LFIAs were prepared, increasing the sample size and allowing for more conclusive results. This experiment would also be the first

to introduce and test different types of classification models to more rigorously describe the performance of the prototype.

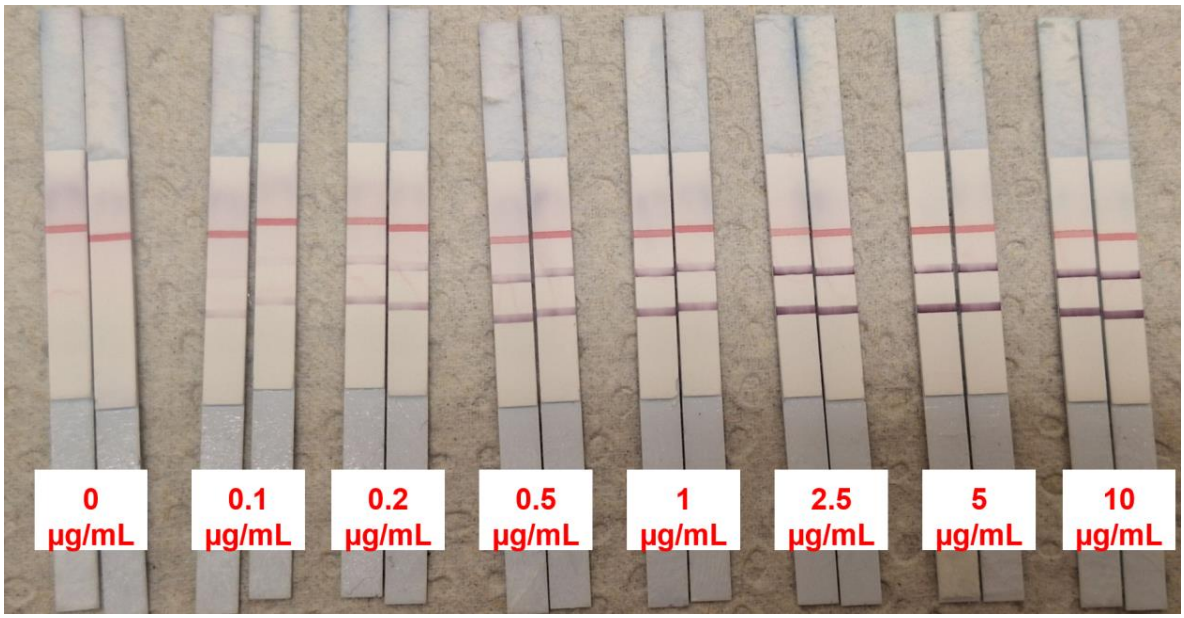



**Figure 31:** A representative sample of the LFIAs prepared for this experiment. Note the three lines, the top line is control, middle line is for testing IgG, and the bottom line is IgM. IgG was the focus of this experiment as it was a more subtle response at low concentrations.

As pictured in Figure 31, a total of 7 concentrations were prepared, plus a control negative, and used to prepare 12 LFIAs each, leaving a total of 96 LFIAs. Each LFIA was then interrogated three times on each of the IgG and IgM test lines, leading to 288 signal traces for IgG test line and 288 signal traces for IgM test line. These signal traces were then used to train different classification models. A conventional model as well as a wide range of different machine learning methods were trained using these data. These various classification methods could then be compared to determine which approach best suited the dataset.

### *4.3.1 Conventional Model*

The conventional model is a very simple regression-type model based on methods used in previous systems. This system begins with the single feature of normalized thermal wave intensity, which has been extracted using methods described in Chapter 3.

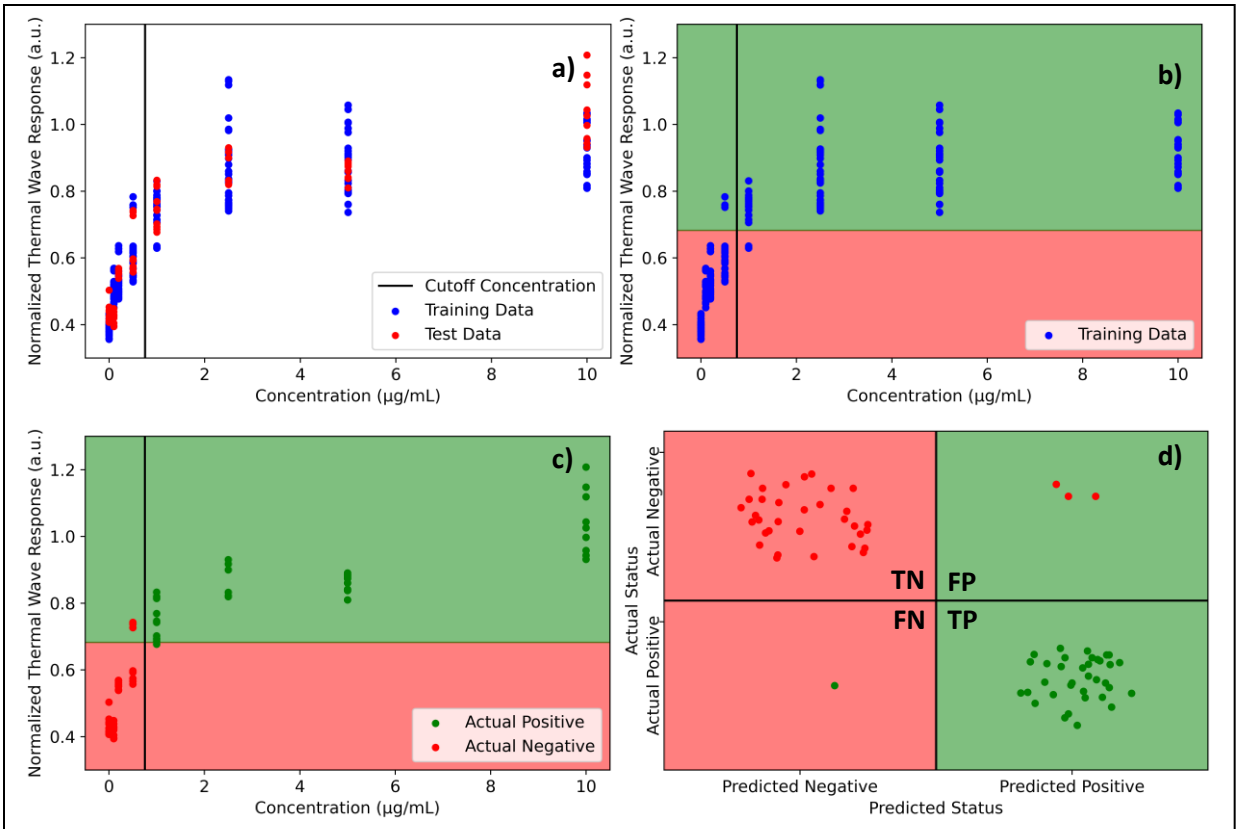



**Figure 32: a)** The full dataset plotting thermal wave response as a function of concentration. The dataset has been split into training data and testing data through a stratified split. Note the cutoff concentration, which has been arbitrarily set to 1ng/mL **b)** After the model has been trained with the training data, the coloured regions are the model's prediction of positive or negative with respect to the chosen cutoff concentration **c)** The testing dataset overlayed on the model's regions, **d)** The testing datapoints sorted into the categories of true negative, false positive, false negative, and true positive; this is known as a confusion matrix.

The labelled features are then randomly divided into training and test data using at 75%-25% split as can be seen in Figure 32(a). Specifically, a stratified split is used, which ensures that an equal number from each concentration are selected, and that data resulting from multiple interrogations of the same LFIA was never in both the test and training data set at the same time. This strategy serves to prevent data leakage from the test dataset into the model.

Once divided, the training data was used to create the conventional model. For each concentration, the average thermal wave response would be calculated. A midpoint between the averages which straddled the cutoff concentration we have selected – 1 ng/mL in this example - was calculated. This value becomes the model's prediction and is illustrated as the green (positive prediction) and red (negative prediction) areas in Figure 32(b).

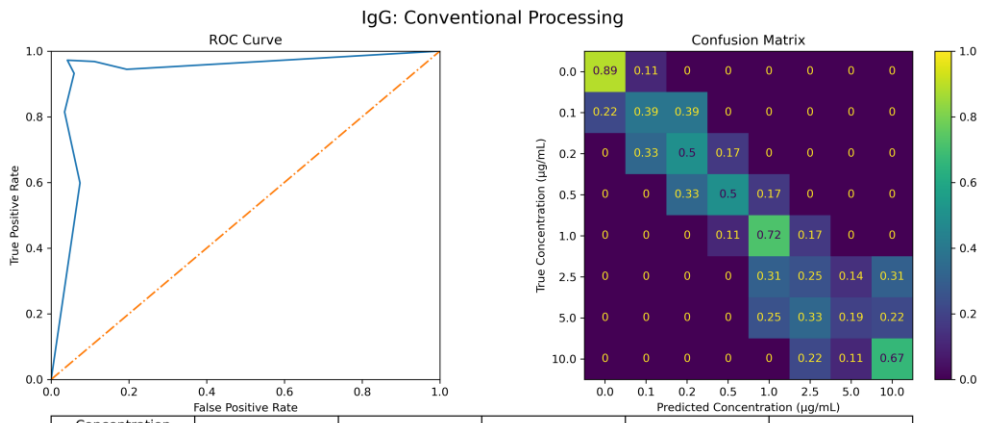
With the model trained, Figure 32(c) reintroduces the testing data and has the model make predictions about whether each point is positive or negative with respect to this cutoff concentration. These points are then sorted into one of four categories: true positive, true negative, false positive, and false negative (Figure 32(d)). This creates a confusion matrix, the values contained within can be used to calculate other important performance metrics used to measure the performance of the model. These values and how they are calculated can be viewed in Table 5.

**Table 5**: Different informative values and their formulas for calculating from the output of a confusion matrix

| **Feature** | **Sensitivity / True Positive Rate** | **Specificity** | **Accuracy** | **Positive Predictive Value** | **Negative Predictive Value** | **False Positive Rate** |
|---|---|---|---|---|---|---|
| **Formula** | $\frac{TP}{TP+FN}$ | $\frac{TN}{TN+FP}$ | $\frac{TP+TN}{Total}$ | $\frac{TP}{TP+FP}$ | $\frac{TN}{TN+FN}$ | $\frac{FP}{FP+TN}$ |
| **Result** | $\frac{35}{35+1} = 0.97$ | $\frac{33}{33+3} = 0.92$ | $\frac{35+33}{35+33+3+1} = 0.94$ | $\frac{35}{35+3} = 0.92$ | $\frac{33}{33+1} = 0.97$ | $\frac{3}{3+33} = 0.08$ |

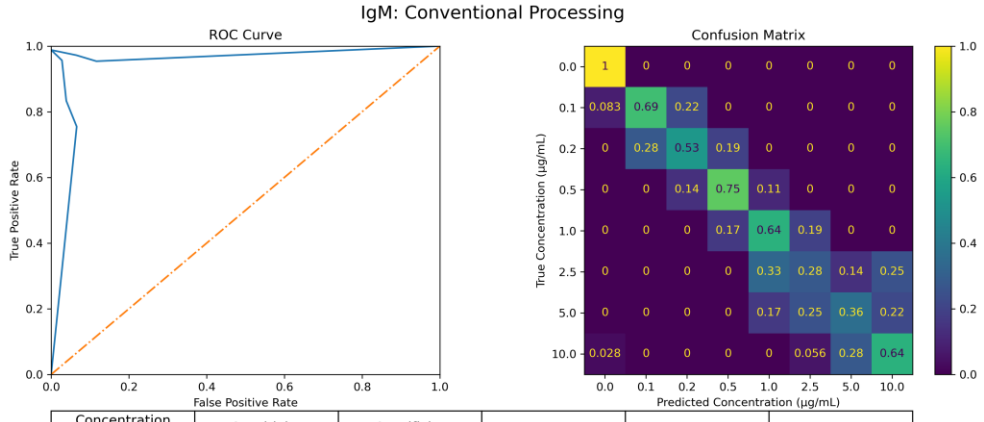
This illustrates how a confusion matrix is generated for a selected cutoff concentration. In practice, this process is performed iteratively over several different possible cutoff concentrations, each returning a varying degree of performance. This can be visualized utilizing a Receiver-Operating Curve (ROC), where for each cutoff the false positive and true positive rates are calculated as laid out in the table above and plotted. This process allows for different models to be directly compared in their performance on the same dataset.

All of the model's results, in the form of ROC, confusion matrix, and table of performance metrics, are found below in Figure 33 (IgG), and Figure 34 (IgM). These results considered together give an overview on how well the model performs at one-versus-one (ie. Binary) and one-versus-many classification tasks. This will be the standard array of results for each of the subsequent machine learning models to be tested.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.97 | 0.89 | 0.96 | 0.99 | 0.85 |
| 0.20 | 0.94 | 0.81 | 0.91 | 0.94 | 0.85 |
| 0.50 | 0.93 | 0.94 | 0.94 | 0.97 | 0.90 |
| 1.00 | 0.97 | 0.96 | 0.97 | 0.96 | 0.97 |
| 2.50 | 0.81 | 0.97 | 0.91 | 0.94 | 0.90 |
| 5.00 | 0.60 | 0.93 | 0.84 | 0.73 | 0.87 |

**Figure 33:** The results of the Conventional processing method on the IgG line.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.99 | 1.00 | 0.99 | 1.00 | 0.93 |
| 0.20 | 0.95 | 0.88 | 0.94 | 0.96 | 0.86 |
| 0.50 | 0.97 | 0.94 | 0.96 | 0.96 | 0.96 |
| 1.00 | 0.96 | 0.97 | 0.97 | 0.97 | 0.96 |
| 2.50 | 0.83 | 0.96 | 0.91 | 0.94 | 0.91 |
| 5.00 | 0.75 | 0.93 | 0.89 | 0.82 | 0.93 |

**Figure 34:** The results of the Conventional processing method on the IgM line.

This conventional model was mainly used as a point of comparison for other, more advanced machine learning models. This comparison will be facilitated with ROCs.

### *4.3.2 K-Nearest Neighbors*

One of the benefits of the machine learning models is that they are made to handle more than a single feature, as the conventional model is. As outlined in Chapter 3, a total of 15 features can be extracted from testing a single LFIA, and after performing the described feature reduction steps, a total of 9 significant features can be attributed to a test.

The first step of model training is identical to that expressed in Section 4.3.1, stratified splitting of the dataset into training and test data. The training data is then used to train the model, a process outlined in Chapter 2. An example of the results of training a model such as this can be seen in Figure 35, with the red and green regions indicating the model's prediction for data within said regions. The red and green data points indicate the actual position of each datum.

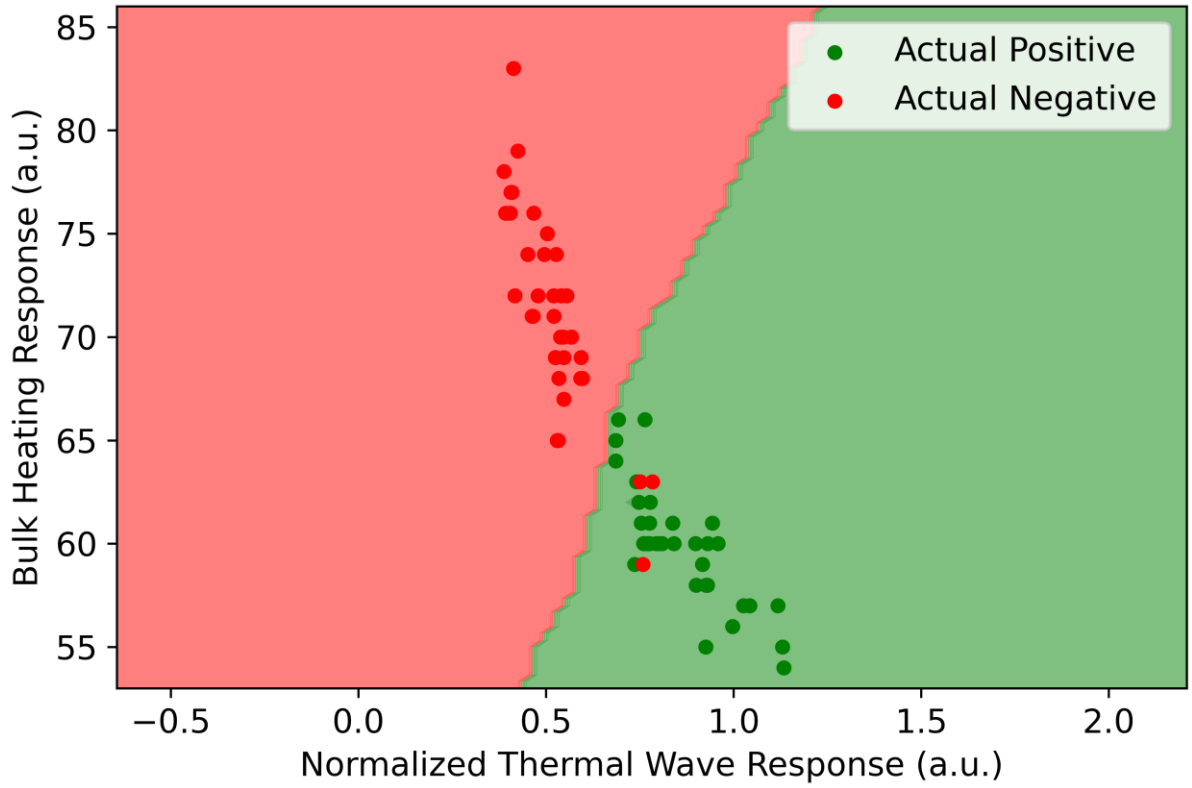
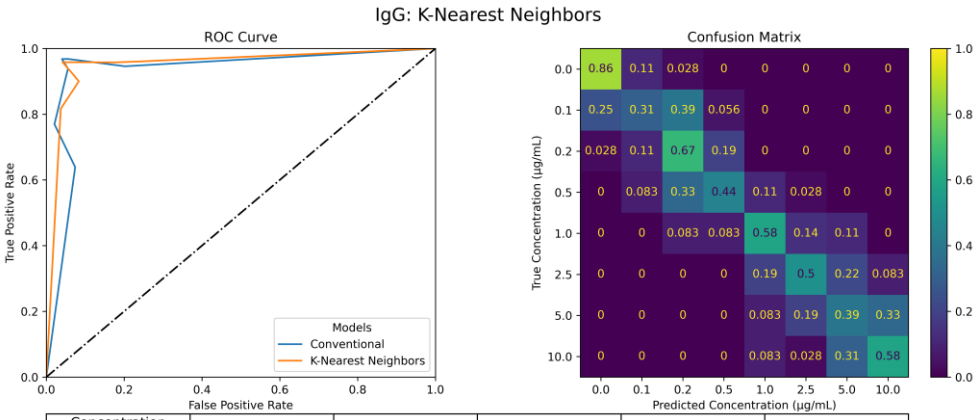



**Figure 35:** The same test dataset as in the conventional processing section after being sorted by a KNN model. The coloured regions of the graph are the model's predictions. This example model only used two features - thermal wave response and bulk heating – so that the results could be visually displayed. The actual model uses 9 features, and therefore would require a 9-dimensional graph to visualize properly

Once the model has been trained - just as with the conventional model – the data in the testing dataset can be used to make predictions, and those predictions are put into a confusion matrix, then used to calculate other significant figures of merit, such as the ROC.

So far, the only classification talked about has been a binary classification; that is, classifying a data point as either positive or negative with respect to some cutoff concentration. Another type of classification performed by these machine learning is multi-class classification. This type of classification attempts to sort a given data point into one of many different classes or concentration bins. Mathematically, training this model is identical to that of binary classification, but the result is a semi-quantitative model which can place a given data point into a specific bin, rather than just labelling it positive or negative. The results of this type of classification yields an extended confusion matrix, with the predicted bins on one axis and the true bin on the second. A perfect model would have data points only along the diagonal of the matrix: every point would be correctly classified.

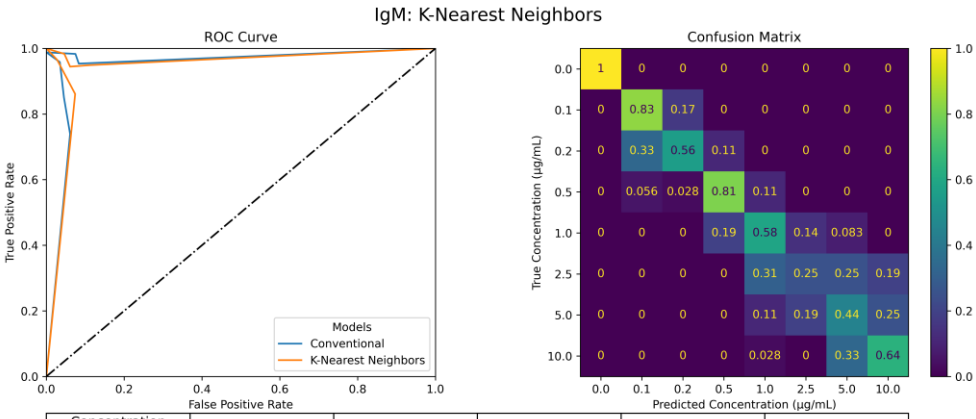



| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.96 | 0.82 | 0.93 | 0.97 | 0.71 |
| 0.20 | 0.96 | 0.77 | 0.91 | 0.92 | 0.88 |
| 0.50 | 0.90 | 0.92 | 0.91 | 0.95 | 0.85 |
| 1.00 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| 2.50 | 0.92 | 0.93 | 0.92 | 0.89 | 0.95 |
| 5.00 | 0.82 | 0.96 | 0.92 | 0.88 | 0.94 |

**Figure 36:** The results of the K-Nearest Neighbors processing method on the IgG line.

This algorithm has two hyperparameters, namely weighting schema and the number of neighbors to consider. The weighting could either be uniform (consider all neighbors equally) or distance based (consider the closest neighbors more important), and the number of neighbors could

between 2 and 20. Through a grid search optimizer, it was established a distance-based weighting schema with 18 neighbours was optimal. Very similar results are produced by marginally increasing or decreasing the number neighbours, but a distance-based measure had a more significant impact. The results of the KNN model can be seen in Figure 36 and Figure 37.



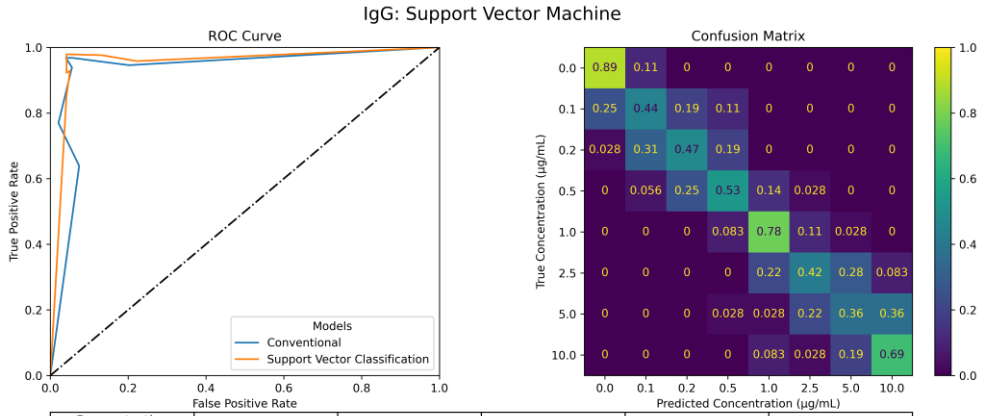

| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.20 | 0.95 | 0.89 | 0.93 | 0.96 | 0.86 |
| 0.50 | 0.98 | 0.95 | 0.97 | 0.97 | 0.97 |
| 1.00 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| 2.50 | 0.94 | 0.94 | 0.94 | 0.91 | 0.97 |
| 5.00 | 0.86 | 0.93 | 0.91 | 0.79 | 0.95 |

**Figure 37:** The results of the K-Nearest Neighbors processing method on the IgM line.

Given how simple and computationally inexpensive the method of K-nearest neighbors is, this is a good choice of machine learning method for these data. One of the drawbacks to this method however is the algorithm's relative inability to extrapolate new data points. For example, a set of samples at a concentration between two existing concentrations would often be sorted into the wrong category because the algorithm only labels new data by way of determining which set of existing data looks most like it. Therefore, at scale, this algorithm may not fully benefit from the large amounts of data being fed into it unless the number of neighbors considered was greatly increased. However, this increase would exponentially increase the computational complexity of the algorithm, rendering its initial benefits useless.

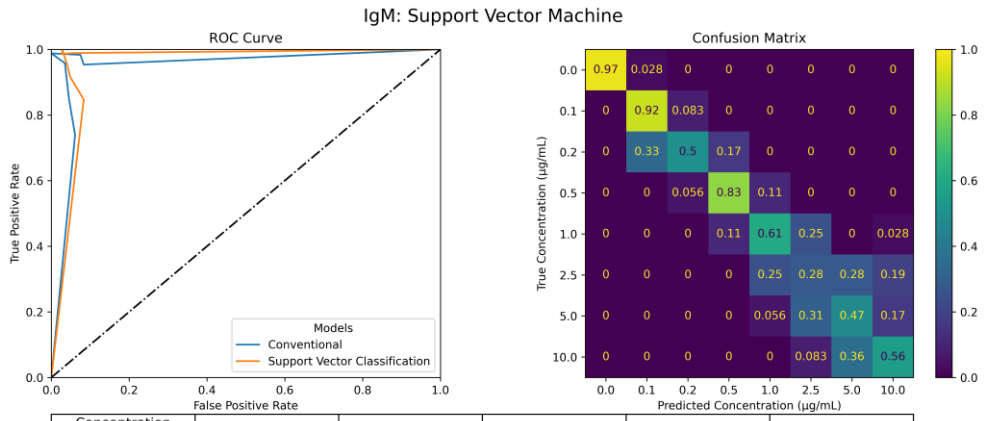
### *4.3.3 Support Vector Classification*

The Support Vector Classification algorithm utilizes one of several different kernels, and each kernel can have its own set of hyperparameters. The simplest linear kernel contains no hyperparameters. The rest of the kernels contain a gamma parameter which modifies the weight that a single sample will have on the model. This gamma parameter can either be scaled by the variation of the input features ("scale" mode) or just by the number of features ("auto" mode). The polynomial kernel has a degree hyperparameter of 2, 4, 6, or 8 explored by the grid search optimizer. The optimal hyperparameters were demonstrated to be the Radial Basis Function kernel with a gamma selected in "auto" mode, and these results are demonstrated below in Figure 38 and Figure 39.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.98 | 0.87 | 0.96 | 0.98 | 0.85 |
| 0.20 | 0.96 | 0.78 | 0.91 | 0.93 | 0.90 |
| 0.50 | 0.92 | 0.96 | 0.94 | 0.98 | 0.88 |
| 1.00 | 0.98 | 0.96 | 0.97 | 0.96 | 0.98 |
| 2.50 | 0.93 | 0.95 | 0.94 | 0.92 | 0.95 |
| 5.00 | 0.85 | 0.96 | 0.93 | 0.88 | 0.95 |

**Figure 38:** The results of the Support Vector Classification processing method on the IgG line.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 |
| 0.20 | 0.95 | 0.96 | 0.95 | 0.99 | 0.87 |
| 0.50 | 0.99 | 0.94 | 0.97 | 0.97 | 0.98 |
| 1.00 | 0.99 | 0.97 | 0.98 | 0.97 | 0.99 |
| 2.50 | 0.92 | 0.95 | 0.94 | 0.92 | 0.95 |
| 5.00 | 0.85 | 0.92 | 0.90 | 0.77 | 0.95 |

**Figure 39:** The results of the Support Vector Classification processing method on the IgM line.

It makes sense that the optimal kernel selected for this dataset was the Radial Basis Function, as this function is capable of most handily dealing with non-linearities. There is still some difficulty in differentiating between the higher concentrations, however this is more since at these concentrations the LFA itself is almost at its upper limit of detection, and thus would not be expected to produce significantly different results. This learning algorithm does well where it matters most: at the lower concentrations.

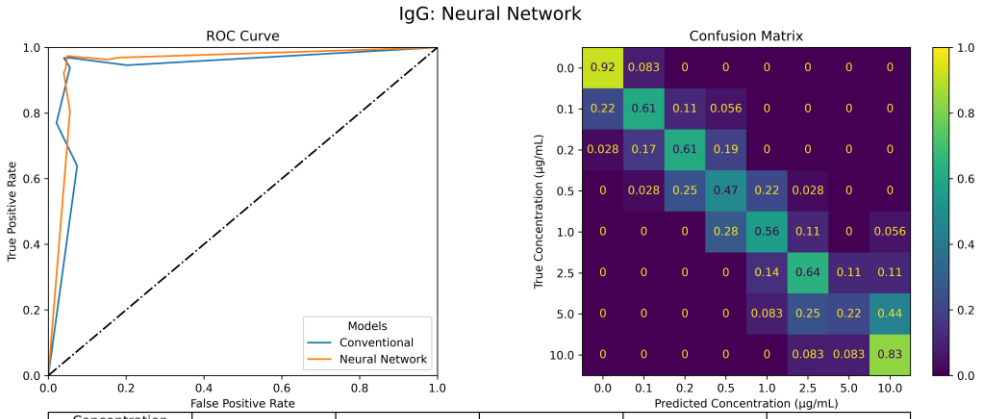
### *4.3.4 Neural Network*

The neural network was the most complex of the algorithms to train because it is made of a network of an arbitrary number of layers, each with their own number of nodes, meaning there is a theoretically infinite number of ways to construct the network. Several one-, two-, and three-layer networks were explored. The number of nodes per layer varied from about the same number as the number of features, to several times larger. A total of 25 different configurations were tested.

To add further complexity, there are several different “activation functions” to choose between for the nodes, including the identity, Logistic, Hyperbolic Tangent, and Rectified Linear Unit;

the equation for these functions is overviewed in Chapter 2. The final hyperparameter is the solver used to change the training rate adaptively and keep the network learning efficiently. There are four possible solvers, each explored in more detail in Chapter 2.

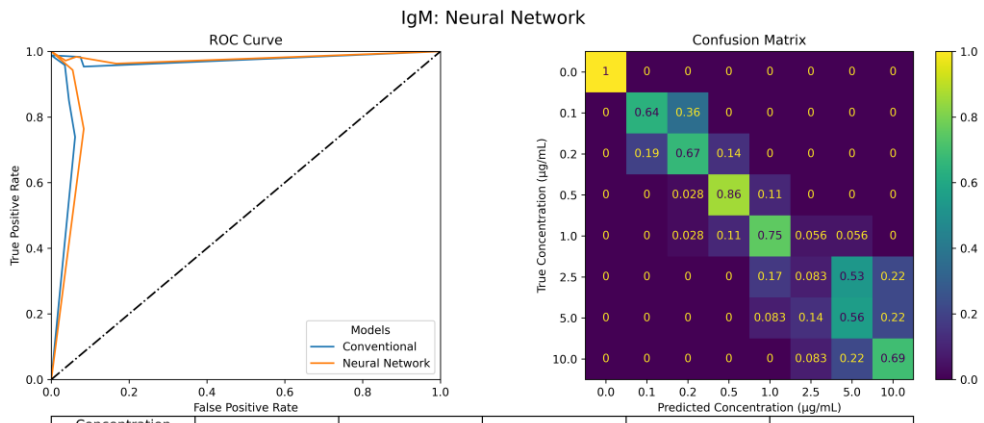
The optimal hyperparameters was found to be a three-layer network with 10, 40, and 10 nodes respectively, using the identity activation function and the SGM solver. The results of this model are tabulated in Figure 40 and Figure 41.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.97 | 0.82 | 0.95 | 0.97 | 0.82 |
| 0.20 | 0.96 | 0.85 | 0.93 | 0.95 | 0.90 |
| 0.50 | 0.92 | 0.96 | 0.94 | 0.98 | 0.88 |
| 1.00 | 0.97 | 0.95 | 0.96 | 0.96 | 0.97 |
| 2.50 | 0.94 | 0.96 | 0.95 | 0.93 | 0.96 |
| 5.00 | 0.81 | 0.94 | 0.91 | 0.84 | 0.94 |

**Figure 40:** The results of the Neural Network processing method on the IgG line.

This algorithm performed very well, which could have been expected from the most complex algorithm. The optimum solver selected was a stochastic gradient descent, however the other solvers tested both performed about as well with convergence times that were extremely comparable for a dataset of this size; in the future, when datasets become very large, an alternate solver may be more fitting. Interestingly, the optimum activation function was the identity function, which means applying no function at all. This was likely made possible because of the normalization process applied to all the features before fitting began.



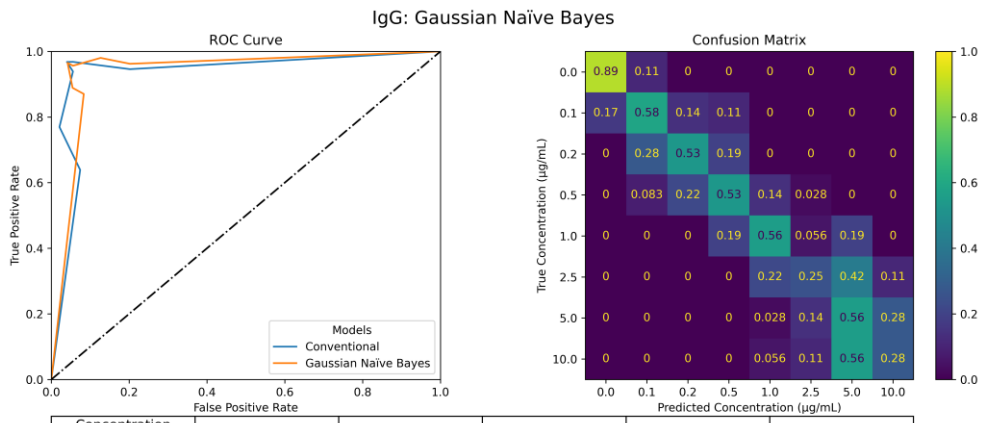

| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.20 | 0.96 | 0.83 | 0.93 | 0.95 | 0.89 |
| 0.50 | 0.98 | 0.94 | 0.97 | 0.96 | 0.97 |
| 1.00 | 0.97 | 0.96 | 0.97 | 0.97 | 0.97 |
| 2.50 | 0.94 | 0.95 | 0.94 | 0.91 | 0.97 |
| 5.00 | 0.76 | 0.92 | 0.88 | 0.80 | 0.93 |

**Figure 41:** The results of the Neural Network processing method on the IgM line.

### *4.3.5 Naïve Bayes Classification*

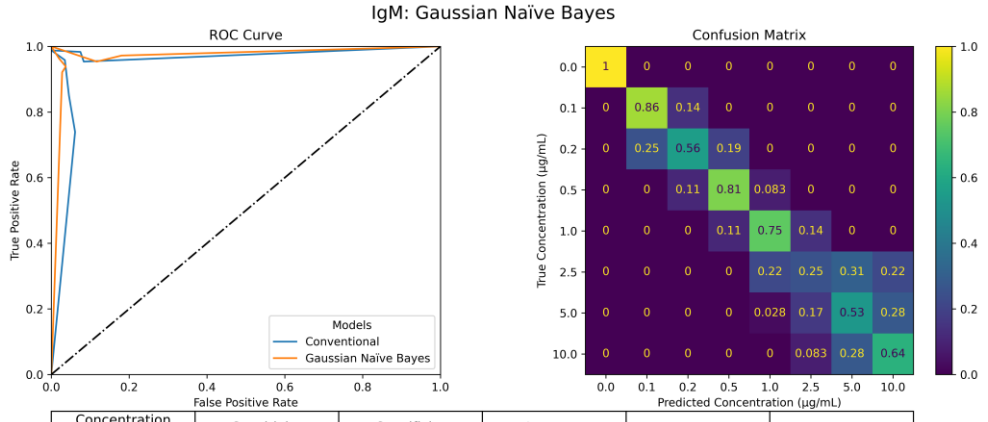
One of the biggest benefits of the Naïve Bayes Classifier is that it requires no hyperparameter tuning to function. The results of this model are shown in Figure 42 and Figure 43.

Despite being one of the easiest algorithms to train because of the lack of hyperparameters, this algorithm was one of the highest performers. Like all other algorithms, there was confusion at the higher concentrations, but the performance at lower concentrations was very good.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.96 | 0.94 | 0.95 | 0.99 | 0.76 |
| 0.20 | 0.87 | 0.92 | 0.88 | 0.97 | 0.71 |
| 0.50 | 0.89 | 0.95 | 0.91 | 0.96 | 0.84 |
| 1.00 | 0.97 | 0.96 | 0.96 | 0.96 | 0.97 |
| 2.50 | 0.98 | 0.87 | 0.91 | 0.82 | 0.99 |
| 5.00 | 0.96 | 0.80 | 0.84 | 0.62 | 0.98 |

**Figure 42:** The results of the Naïve Bayes Classification processing method on the IgG line.



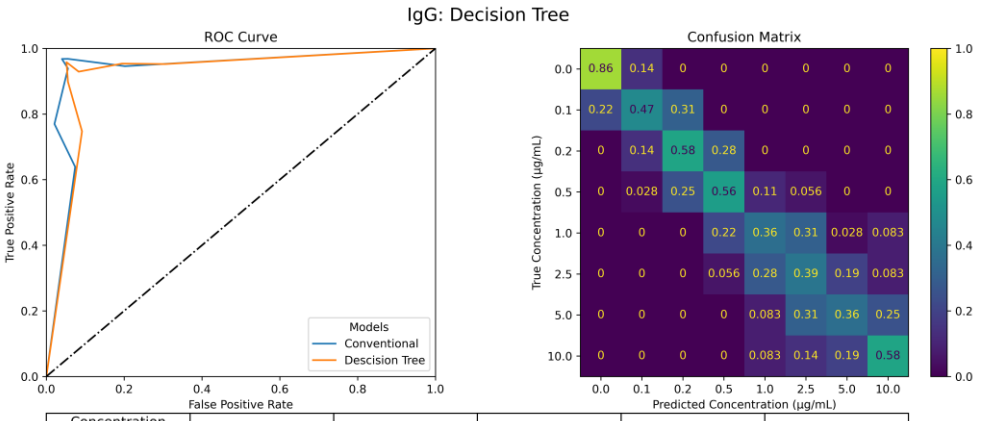

| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.20 | 0.92 | 0.97 | 0.93 | 0.99 | 0.81 |
| 0.50 | 0.94 | 0.96 | 0.95 | 0.98 | 0.91 |
| 1.00 | 0.96 | 0.98 | 0.97 | 0.98 | 0.97 |
| 2.50 | 0.95 | 0.88 | 0.91 | 0.83 | 0.97 |
| 5.00 | 0.97 | 0.82 | 0.86 | 0.65 | 0.99 |

**Figure 43:** The results of the Naïve Bayes Classification processing method on the IgM line.

### *4.3.6 Decision Tree*

There are several hyperparameters to consider in the decision tree. The first is the splitting criterion, which is either entropy, Gini, or log loss, all of which are explained in Chapter 2. There is also a hyperparameter for the minimum number of samples which are required before a node can be split, which was anywhere between 2 and 10.

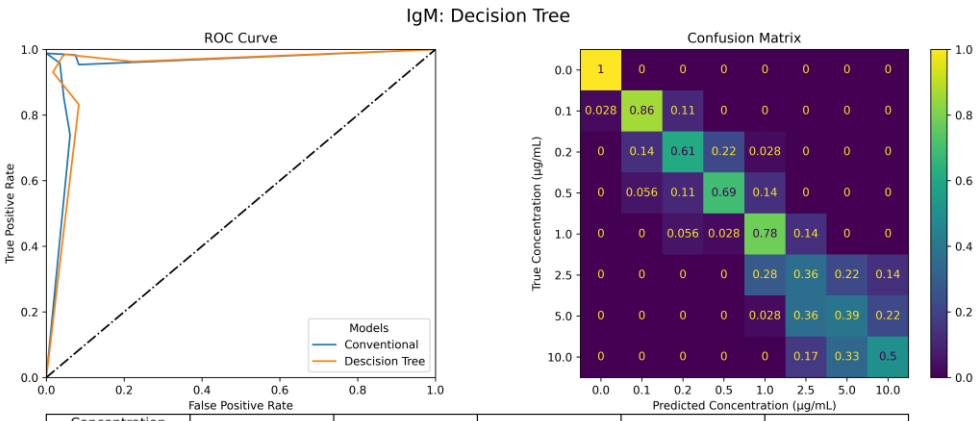
The optimum setup was determined to be an entropy-based splitting criterion with 8 minimum samples. The results are below in Figure 44 and Figure 45.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.95 | 0.69 | 0.92 | 0.96 | 0.66 |
| 0.20 | 0.95 | 0.81 | 0.92 | 0.94 | 0.87 |
| 0.50 | 0.93 | 0.92 | 0.92 | 0.95 | 0.89 |
| 1.00 | 0.96 | 0.95 | 0.95 | 0.95 | 0.96 |
| 2.50 | 0.90 | 0.94 | 0.93 | 0.91 | 0.94 |
| 5.00 | 0.75 | 0.91 | 0.86 | 0.73 | 0.91 |

**Figure 44:** The results of the Decision Tree processing method on the IgG line.

The decision tree algorithm was not the right choice for this dataset; its performance was below that of the conventional algorithm. This is likely because the underlying dataset is not truly made up of different categories, but rather well-spaced samples from a continuous feature space. Other algorithms have an easier time making sense of this somewhat continuous dataset, even if it is non-linear, whereas this method struggles to determine truly discrete labels.




| Concentration Cutoff (µg/mL) | Sensitivity | Specificity | Accuracy | PPV | NPV |
|---|---|---|---|---|---|
| 0.10 | 0.98 | 0.95 | 0.98 | 0.99 | 0.90 |
| 0.20 | 0.96 | 0.78 | 0.92 | 0.93 | 0.89 |
| 0.50 | 0.94 | 0.94 | 0.94 | 0.96 | 0.92 |
| 1.00 | 0.98 | 0.96 | 0.97 | 0.96 | 0.98 |
| 2.50 | 0.93 | 0.98 | 0.97 | 0.97 | 0.96 |
| 5.00 | 0.83 | 0.92 | 0.90 | 0.77 | 0.94 |

**Figure 45:** The results of the Decision Tree processing method on the IgM line.

### *4.3.7 Model Comparison*

Overall, both conventional and machine learning methods performed extremely well on the dataset. As can be seen in Figure 46(a) and (b), the area under all the curves approach 1, which would be a perfect model. We can notice slightly higher AUCs in the IgM graph, this is because the IgM lines were more distinct than the IgG lines on the LFIAs, making the thermal response bolder. We also notice that the least performant model was the Descision Tree, perfroming more poorly than even the conventional model.

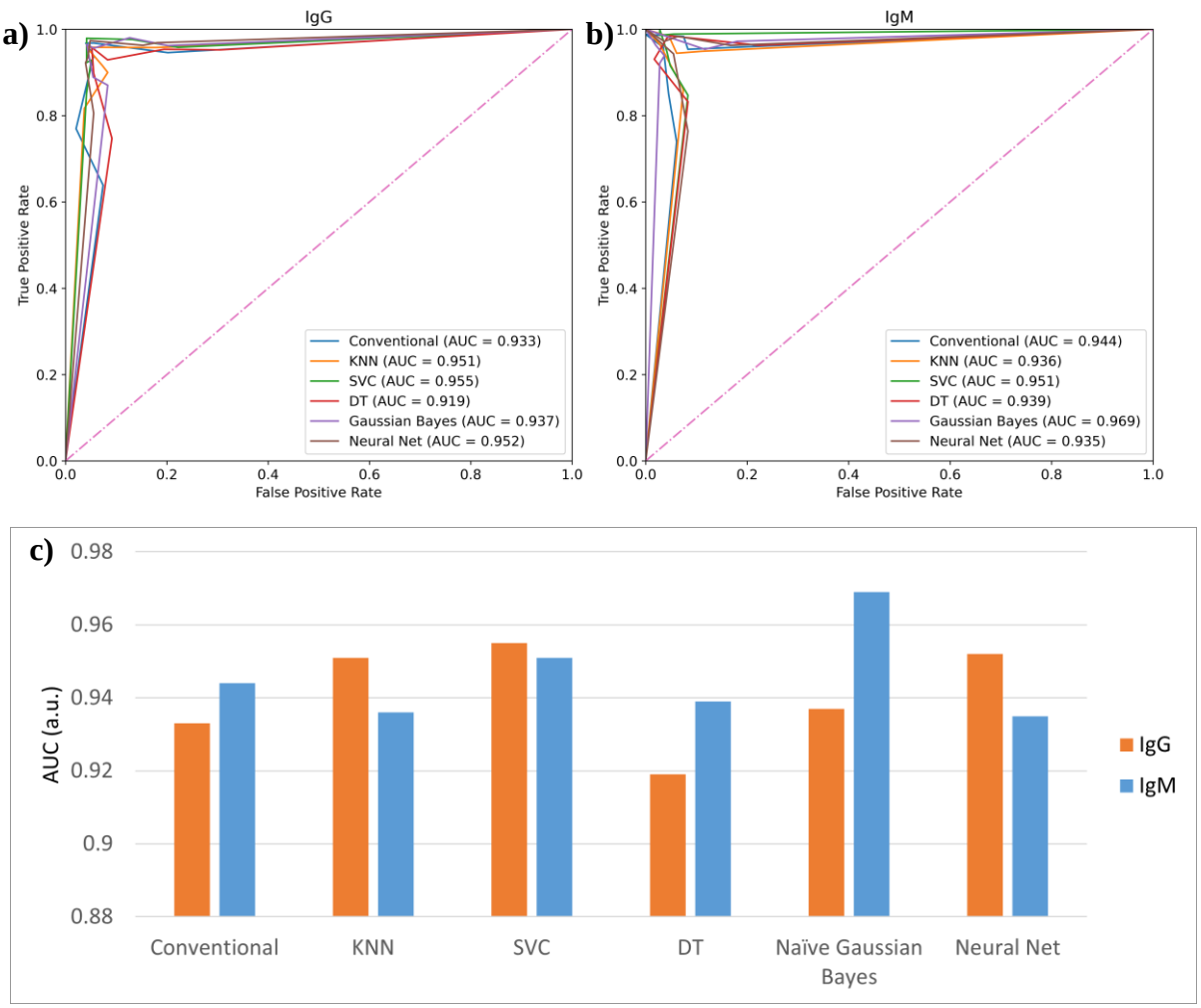



**Figure 46:** A comparison of all the models trained used on **a)** IgG and **b)** IgM. **C)** Compares all of the AUC values for each model.

The AUC is a decent measure of the overall performance of a model, but we should also pay attention to the specifics. Table 6 lists the accuracy and specificity values for IgG and IgM with a cutoff concentration of 0.1 ug/mL. This concentration is the closest to control negative and should be the most difficult task for these models. We see again, most of the models performed very well, with accuracy and specificity values averaging above 90%. Again, the model which stands out as the least performant is the Decision Tree. The conventional model does have a significantly lower specificity in the IgM category, just as the KNN and Neural Network have in the IgG category. The highest performing models are the Support Vector Classification, and Naïve Gaussian Bayes', both with near perfect performance in IgM and extremely good performance in the IgG.

**Table 6:** Comparing the accuracy and specificity values of each model for a cutoff concentration of 0.1 μg/mL.

| | IgG | | IgM | |
|---|---|---|---|---|
| | **Accuracy** | **Specificity** | **Accuracy** | **Specificity** |
| **Conventional** | 0.96 | 0.92 | 0.93 | 0.86 |
| **K-Nearest Neighbors** | 0.93 | 0.82 | 1.00 | 1.00 |
| **Support Vector Classification** | 0.96 | 0.87 | 1.00 | 0.97 |
| **Decision Tree** | 0.92 | 0.69 | 0.98 | 0.95 |
| **Naïve Gaussian Bayes** | 0.95 | 0.94 | 1.00 | 1.00 |
| **Neural Network** | 0.95 | 0.82 | 1.00 | 1.00 |

If we consider factors other than strictly performance, the Naïve Gaussian Bayes' model was a top contender because it required no hyperparameter tuning, unlike the Support Vector Classification model. This makes setting up and training a model trivial, especially when one would like to update the model with the addition of new data. To summarize, while most of these models (including the conventional model) performed extremely well, the Naïve Gaussian Bayes' model stood out as the top choice. This model was able to achieve perfect performance in the IgM category, and near perfect (~95%) performance on the IgG.

The outcome of this experiment was a huge success, verifying that the prototype could reliably produce high quality data on a repeated experiment with several LFIAs at each concentration. The prototype was able to resolve a 0.1 μg/mL cutoff concentration with an accuracy and specificity of about 95% when using the Naïve Gaussian Bayes' model. This is a result which is comparable to previous results using a more expensive system, which was able to reliably differentiate between 0.2 μg/mL and 0 μg/mL [19]. Furthermore, because these are machine learning models, as more LFIAs are tested using this prototype, continuous learning will be possible, further improving the model.

These results indicated that the device is working as well as it could be expected to when dealing with lab-prepared LFIAs. The next step was to attempt to get similar results when working with human factors.

### 4.4 Human Saliva THC Experiment

The final and perhaps most crucial experiment to validate the performance of this prototype is real-world data. Preparing LFIAs in a lab setting is highly controlled, if this device is to be of use to anybody it must also function when the setting is not so tightly controlled. For this reason, our final and most rigorous experiment was to test human saliva for THC.

Human experiments are especially difficult because of all the confounding factors introduced. Every person's salival makeup is different, even the same person will have different makeups depending on the time of day, hydration, diet, oral hygiene or many other factors. Beyond this, collecting saliva requires the participant to use a sponge collector, which can be used in many differing ways while still collecting saliva. That also does not account for the fact that saliva sponges are known to trap THC, further confounding results. For these reasons, and because it is relatively expensive to conduct human experiments, they are rarely done for this type of device.

This type of experiment is novel and would take the validation of this device a step beyond what has been seen in the past in engineering thesis work. If this device could be demonstrated to reliably work on humans, this would be a huge motivator to further the development of the prototype.

A total of ten participants, some participating more than once, provided a total of 17 series of saliva samples. Each sample series included saliva from before smoking cannabis and several saliva samples at regular intervals after smoking cannabis (see Chapter 3: Methods for details). An additional participant also provided a control negative series. In order to study the longitudinal trends in sample series as detected by our device and mass spectroscopy, the concentrations as measured by mass spectrometry were plotted against the thermal wave responses as measured by the device (for example, Figure 47); connecting lines were used to indicate the temporal relationship between the measurement points. The letters in each plot's legend indicate the session which that participant took part of, and the number associated

indicates the different LFIAs prepared with the same saliva sample. In most cases, two LFIAs were prepared with each saliva sample.

The trend expected from this experiment is that the initial collection has zero THC, the second collection should have the highest THC, and we would expect the concentration to decrease after that. Plotted, this ideal relationship looks like a point starting in the top left, dipping towards the bottom right, and returning on a negative slope towards zero. Figure 47 shows the most ideal collections; the initial collections are at or near zero, the second collection shows an increase in THC along with a decrease in signal, and subsequent collections return along nearly the same trend line as they first decreased.

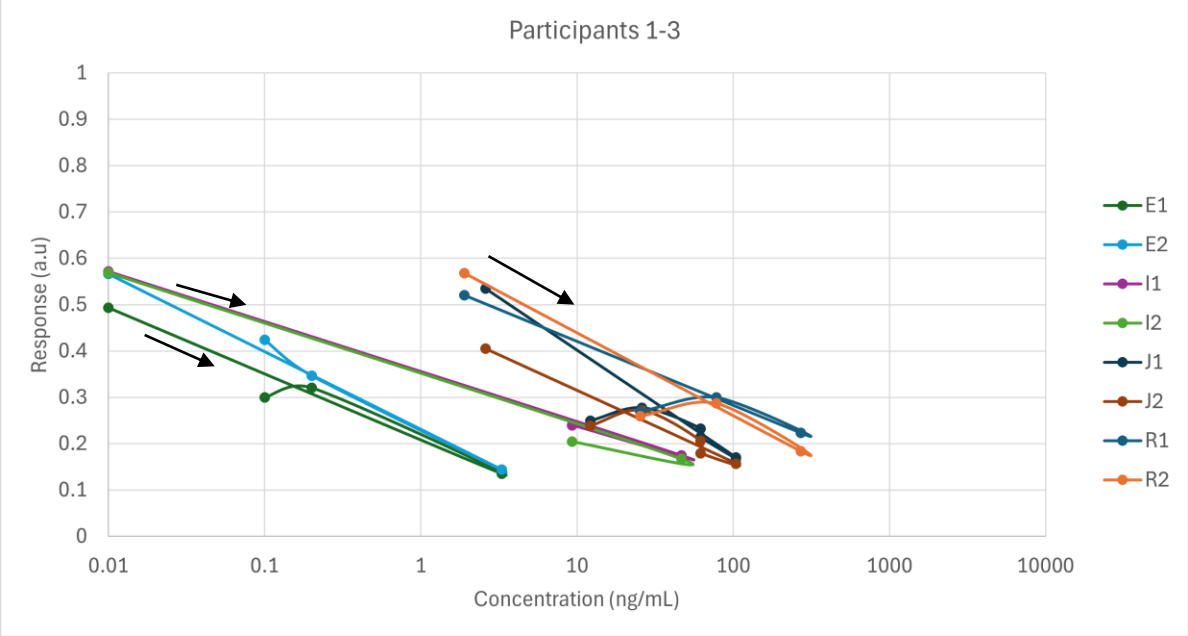



**Figure 47:** The collections produced by the first three participants; their temporal trends indicated by the black arrows. These are the most ideal collections, starting near the top left, moving downwards to the bottom right, then back along the same path they travelled.

Participant 4's collections – seen in Figure 48– are mostly ideal as well. There was a collection, series F, which was errant in that after smoking, changes in the LFIA responses were noted, however the mass spectrometry results showed no increase in THC until the fourth collection. This does not conform with our understanding of human physiology, and the strong, repeatable response of the LFIA indicated that there was likely an issue with the samples sent for testing. For this reason, the entire F series was considered an outlier.

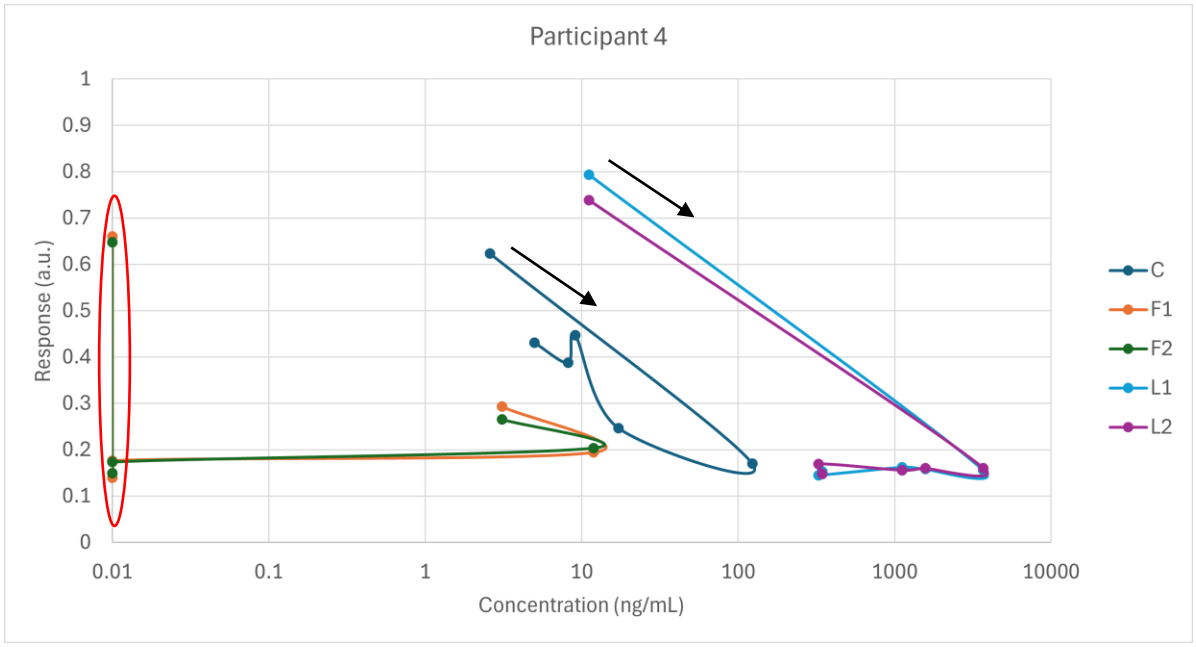



**Figure 48:** The results from Participant 4's collections; their temporal trends indicated by the black arrows. Series F is errant, indicating experimental errors and has therefore been removed.

Collections from participants 5 and 6 are more complex. Indicated in Figure 49 by red circles, there are three points which have been removed from collections K and N because subsequent collections showed an increase in THC concentration. There are other points indicated by the red arrow which are anomalous, but not outliers. These points show a thermal wave response which is higher than the initial collection, even though there is still THC in the saliva. Although this is not the trend we would expect to arise from an LFIA, these are the sorts of incongruities which must be dealt with if one hopes to have a device which can work on human data.

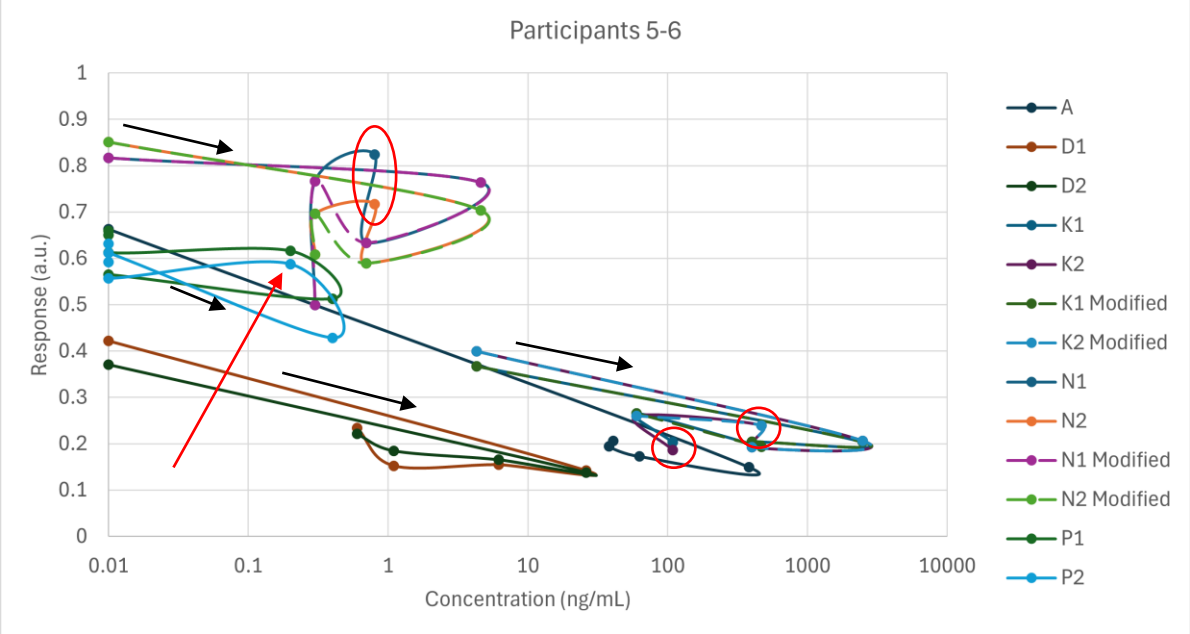



**Figure 49:** Participants 5 and 6 collection results, which largely behaved as expected. The temporal trend is indicated by the black arrows. There were several datapoints however where an increase in THC was measured, and therefore these points (circled in red) have been removed. Indicated by the arrow is a point which, while non-ideal, is not necessarily an errant point.

Participants 7 and 8 produced the most complex responses, with several large outliers from rounds G and M being removed as indicated by red circles in Figure 50. Even after these outliers are removed, the resulting datasets are still anomalous, with large changes in thermal wave response corresponding to relatively small changes in THC concentration. This, again, is not ideal, however it is still a valid collection, and therefore the rest of the points remain a part of the dataset.

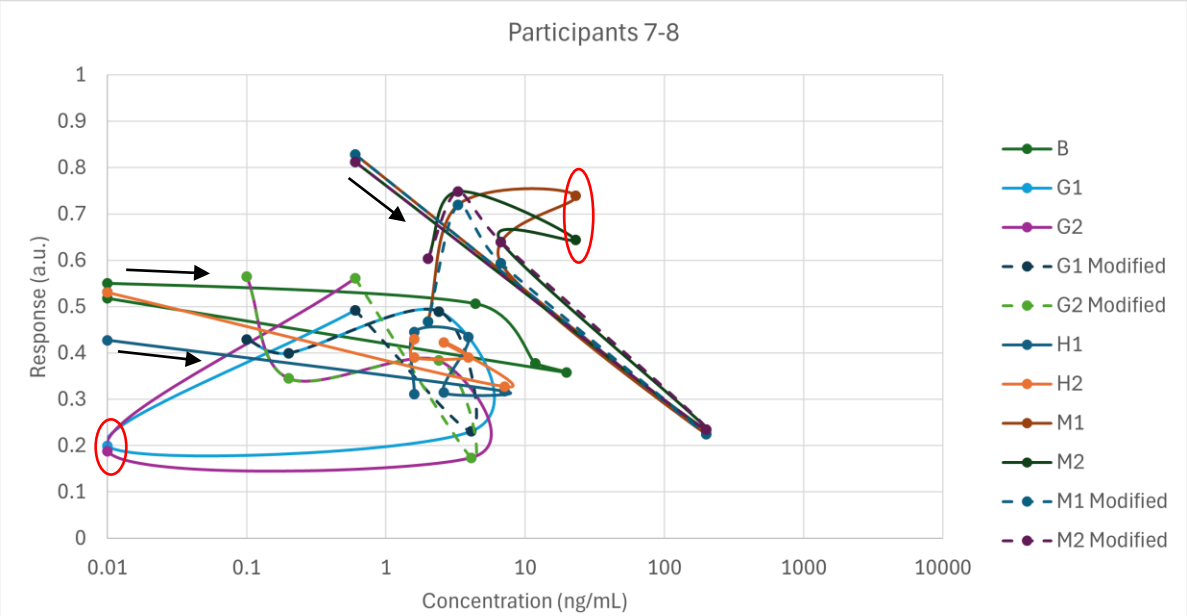



**Figure 50:** Data collected from participants 7 and 8. The black arrows indicate the temporal trends. These also contain outliers (circled in red) which have been removed because they act in a way contrary to human physiology.

Finally, participants 9 and 10 produced anomalous responses in that the THC content of their saliva barely deviated from zero even after smoking. This is apparent by the number of points fully on the left side of Figure 51. While anomalous, these responses do not violate our current understanding of the behaviour of THC in saliva, and thus are not removed as outliers.

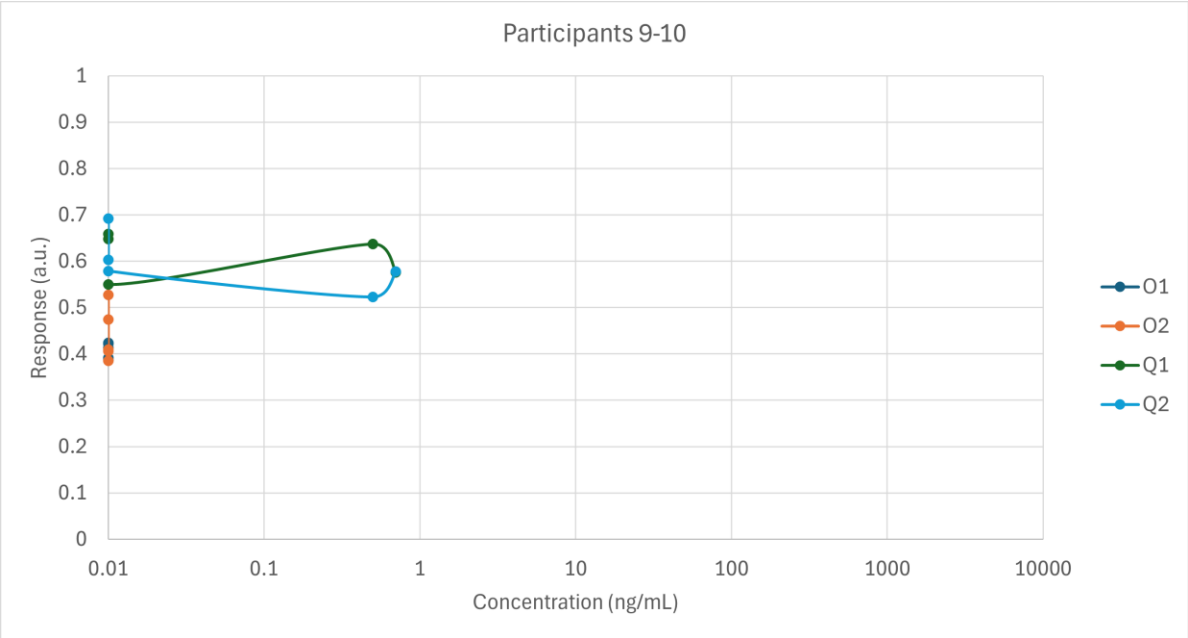



**Figure 51:** These results obtained from participants 9 and 10 are errant in that very little change in the THC concentration was detected even after smoking. While errant, these points do not qualify as outliers as they are not contrary to our understanding of THC metabolism.

Several of the datasets contained consecutive datapoints wherein the concentration as measured by mass spectrometry increased temporarily. This does not fit the current understanding of how THC concentration in saliva changes over time, so these points were considered outliers and removed for the purpose of model training. These errant points were likely produced by improper use of the saliva collectors, or by mishandling when preparing/shipping the sample to the lab to perform mass spectrometry. The trends minus the omitted points are also plotted in a dashed line. It is also important to note that participant 3's series F was omitted, as the trend exhibited was clearly impacted by some source of experimental error.

Initially, these results were compiled and processed in an identical fashion to the initial, lab-based THC experiment. The points were arranged into concentration categories with corresponding 95% confidence intervals, shown in Figure 52.

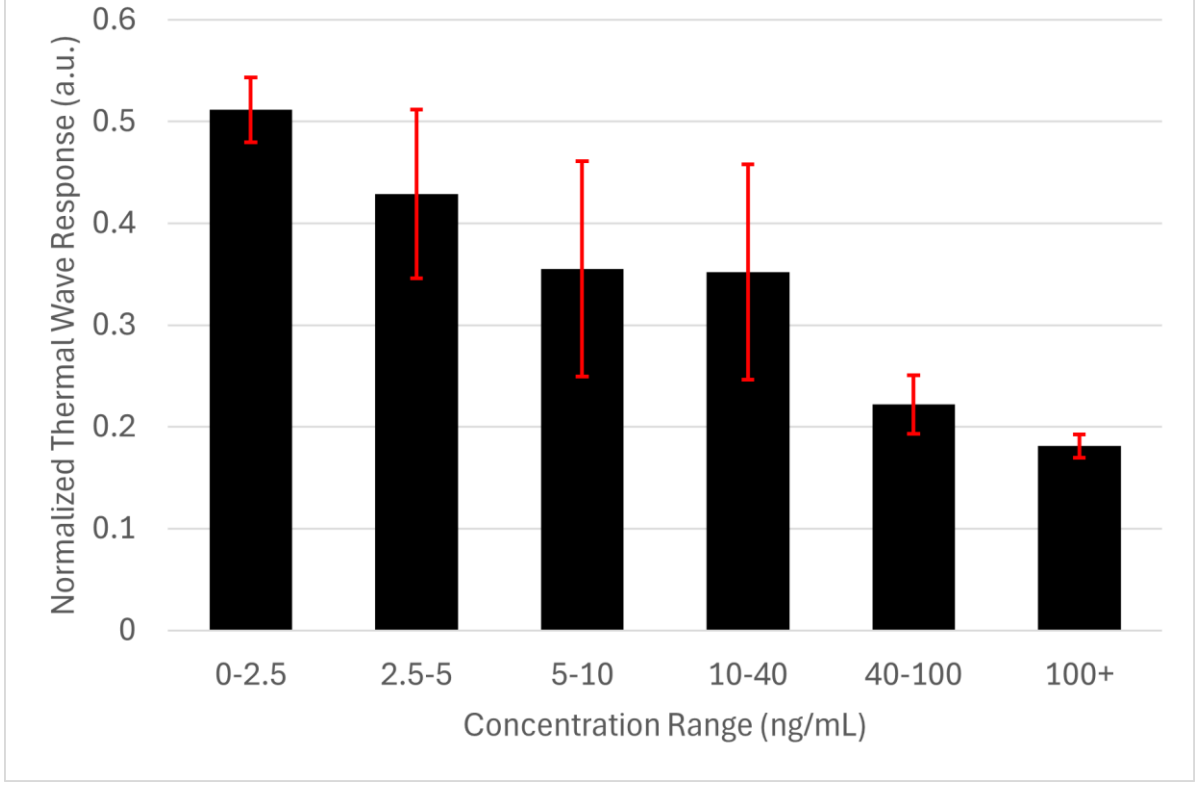



**Figure 52:** Total results of the human THC study in a bar plot with 95% confidence intervals. The LFIA trend can still be seen in the change of mean for each concentration bin, but this response is far more variable than the lab experiments

As was to be expected, human data does not cleanly follow the trend demonstrated with the lab-prepared samples. This initial result was understandable; human saliva is not a standard viscosity, may have varying concentrations of confounding solutions such as mucus, toothpaste/mouthwash, or even residuals organic matter from previous meals. While participants are not allowed to eat during the experiment and can only drink water, this does not ensure an identical saliva sample from every participant. Further confounding the process is the sponge collectors, which are used differently by each participant, even through attempts by the facilitator to standardize their use. It is clear from these data that a monotonic relationship between the thermal wave response and the actual concentration does not exist as it did in the lab-prepared experiment, as there is a large overlap between each concentration range in Figure 52.

The main confounding variable of the human experiment is the differences in saliva makeup for each participant; even the same participant will have different salival composition depending on the day. Since this cannot be controlled for, we proposed a different type of experiment to determine the presence of THC. We know that the amount of THC in one's saliva decreases over time, meaning that if THC is present, taking multiple tests over time should indicate a change in the concentration of THC. In the case where there is no THC, the signal should stay the same. By considering the difference between two tests gathered after a specified time period, we take into consideration the specific salival makeup of the participant.

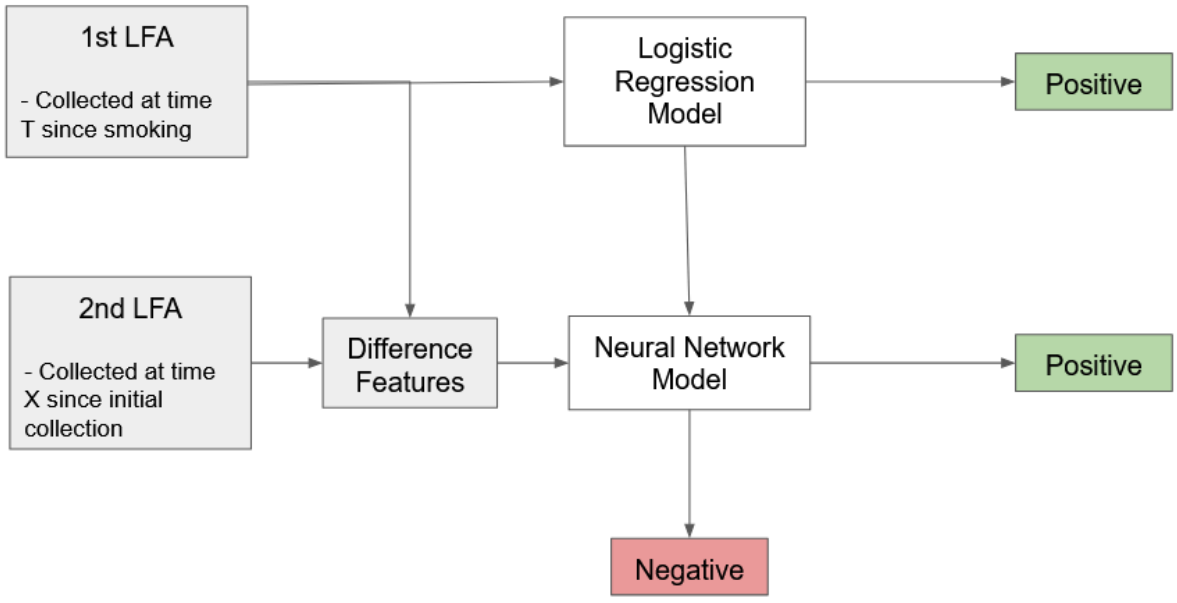



**Figure 53:** The outline of the two-stage algorithm used to interpret a pair of LFIAs.

Processing this newly paired dataset requires a new style of processing which is outlined in Figure 53. This is a two-step algorithm, utilizing two different machine learning algorithms for different purposes. The first stage is a logistic regression algorithm, which screens for LFIAs which are overwhelmingly positive. This stage is required because if an algorithm was only looking for changes in response and a participant produced two subsequent LFIAs which were both overwhelming positive, the algorithm would rightly determine there was no difference in signal, but wrongly attribute that as a negative response. The second stage is a neural network which is fed all the features of both LFIAs to determine if a significant enough change has occurred between the two. The optimal neural network was found to have three hidden layers

(20, 150, and 20 nodes respectively), and the identity activation function. The ROC curve produced by this method, both with and without outliers removed, can be seen in Figure 54.

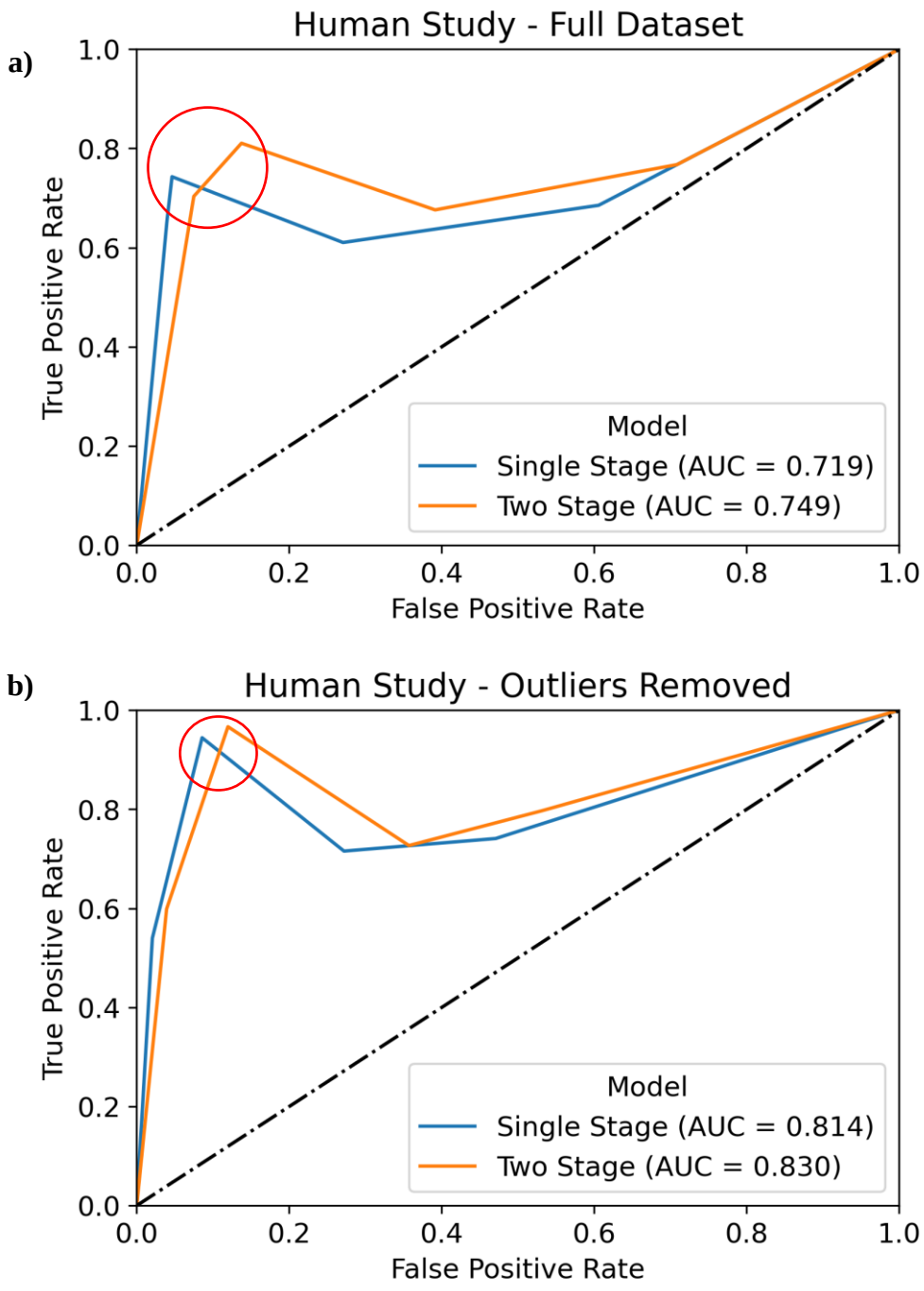



**Figure 54: a)** The two model designs after being trained on the full dataset and **b)** the two models after being trained on the dataset with outliers removed. Circles in red are the points of highest performance, which in all cases occurred at 11.8 ng/mL.

The removal of outliers has a great impact on the outcome of both the single and two stage models, increasing the AUC of the ROCs by about 0.10 and 0.08 respectively. The two-stage model performed better on both datasets, increasing the AUC on the curated dataset by a factor of about 0.015. These graphs show a clear peak performance, as indicated by the red circles, at a cutoff concentration of 11.8 ng/mL; for comparison, the manufacturer recommended cutoff concentration for this LFIA is 40 ng/mL. We also notice there is a large dip in the curve after the first major peak indicating a decrease in the true and false positive rates as the model's cutoff concentration increases further. This trend makes sense, as the LFIA responses become more variable as the concentration decreases far above the cutoff concentration which it was designed for.

**Table 7:** Comparison of the different models and datasets for a cutoff concentration of 11.8 ng/mL

| | | **Model Comparison at 11.8 ng/mL Cutoff Concentration** | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Model** | **Sensitivity** | **Specificity** | **Accuracy** | **PPV** | **NPV** |
| **Full** | **One-Stage** | 0.74 | 0.95 | 0.90 | 0.79 | 0.90 |
| | **Two-Stage** | 0.81 | 0.86 | 0.85 | 0.63 | 0.93 |
| **Outliers Removed** | **One-Stage** | 0.94 | 0.91 | 0.95 | 0.94 | 0.98 |
| | **Two-Stage** | 0.97 | 0.88 | 0.93 | 0.78 | 0.99 |

Table 7 shows the various performance criteria of the different models at the optimal cutoff concentration of 11.8 ng/mL, with the best performance of each criteria highlighted. This table makes it apparent that removing outliers clearly increased the performance of both models, as they both have very competitive values. It also demonstrates that the one and two stage models perform very similarly in most cases, the only significant difference being the positive predictive value, which was higher for the single-stage model. Even though the performance of these models is close, the two-stage model allows for more flexibility. For example, in this experiment the time difference between the first and second LFIA was 30 minutes, but longer wait times could be used which would further increase the difference in LFIA responses, and therefore the model's ability to detect a change. While both models perform at a similar level, only the two-stage model can be greatly improved upon.

**Table 8:** The different performance criteria of the two-stage model trained with the curated dataset at different cutoff concentrations

| | **Two Stage Model – Outliers Removed** | | | | |
|---|---|---|---|---|---|
| **Cutoff (ng/mL)** | **Sensitivity** | **Specificity** | **Accuracy** | **PPV** | **NPV** |
| **0.3** | 0.80 | 0.47 | 0.71 | 0.71 | 0.57 |
| **2.6** | 0.73 | 0.64 | 0.74 | 0.67 | 0.69 |
| **11.8** | 0.97 | 0.88 | 0.93 | 0.78 | 0.99 |
| **104.3** | 0.60 | 0.96 | 0.88 | 0.99 | 0.90 |

Table 8 contains the performance parameters for each cutoff concentration by the two-stage model trained on the dataset not containing outliers, highlighted is the cutoff which performs the best - 11.8 ng/mL. Increasing the cutoff concentration 104.3 ng/mL decreases the sensitivity of the system, meaning more positive results will be missed. Decreasing the cutoff to 2.6 ng/mL decreases both the sensitivity and specificity to unacceptable levels for a screening device. For these reasons, we have determined that the optimal cutoff concentration achievable by this device is 11.8 ng/mL.

This experiment demonstrated that the conjunction of this prototype and a sufficient machine learning model could reliably and significantly decrease the LOD of an LFIA from 40 ng/mL to almost a quarter of that, even when using collected human factors instead of a lab-prepared method. This is a step of validation on human factors which has been seldom attempted with these sorts of devices - let alone one so inexpensive – and for the results to be not only significant, but repeatable is remarkable.

# Chapter 5: Conclusion

## 5.1 Summary

This thesis began with an assessment of different single element sensors and lighting components. These where then situated in a tabletop testing jig, and the geometric elements of the system were optimized by testing a representative number of lab-prepared LFIAs. Simultaneously, collection and post-processing software and algorithms were developed to interpret the experimental results. Once the system had been optimized – including the breakthrough that was the transmission mode – a prototype handheld version was designed. A set of PCBs were iteratively designed and assembled, as well as the mechanical housing for all components and the LFIAs to be interrogated.

After the prototype was completed, two sets of serial dilutions (THC, and COVID-19 IgG + IgM) were prepared. Two sets of different LFIAs were then labelled, prepared, and doped with the diluted concentrations. In parallel with this process, human trails were being performed; cannabis was consumed (by participants), saliva samples collected and prepared for transport after being split and used to prepare multiple LFIAs before being shipped and tested by mass spectrometry. These LFIAs, lab-prepared and from human trials, then each had to be interrogated multiple times by the prototype devices and the data saved for post-processing.

The resulting data then had to undergo a post-processing algorithm. A conventional algorithm was first introduced, but later several different machine learning techniques and models would be trained using these collected data. The hyperparameters of each individual model were then optimized using an exhaustive search method. These models were then all compared, and the best ones selected to represent the results. Additionally, a custom two-stage model was then prepared for processing the human data once it was determined that methods used to classify the lab-prepared LFIAs would be unsuitable for the complexities introduced by human factors.

All this effort leaves us with a device that can demonstrably increased the limit of detection of the different LFIAs that it has been tested on thus far. The device can connect to the IoT through Wi-Fi or Bluetooth connections, is handheld and battery-powered, and is inexpensive enough that if it were developed into a commercially viable version, it would be affordable enough to become a household device.

## 5.2 Future Direction

Development of this device is far from over as it continues its lifecycle from a validated prototype towards a consumer electronic product. This transition will force the next iteration to be designed for manufacturing – a factor which was unable to be adequately addressed through this process so far. The five PCBs will be simplified into a single, or pair, of PCBs which can be mass manufactured rather than assembled by hand. The case will also decrease in size and become more ergonomic to hold. The Arduino which currently serves as the main computing component will be swapped out for a less powerful but smaller surface-mount processor, further increasing the battery life of the overall system.

Other than changes for manufacturing, other versions of this device can easily be developed to handle multiplexed LFIAs by simply adding more LEDs and sensors. It was also imagined that a single LED which can output more than one wavelength could be used to interrogate the same line multiple times, changing the wavelength each time and comparing the difference in response to further increase the accuracy. Alternate modulation methods, such as utilizing a chirp or a phase-shifted keying wave were experimented but ultimately were not found to greatly increase the performance, however further investigation may find use for these methods.

The final and most productive area for improvement of this system lies in improving the response of LFIAs themselves. The advent of a clear backing on the THC LFIAs was a great boon to the performance of this device already, but there is still room for improvement. As it is now, LFIAs are designed for visual interpretation, and a part of that is designing it such that the control line is always visible when the test is completed properly. Designing an LFIA with this device in mind however does not necessarily require those design constraints. For example, an LFIA which replaces the control line with effectively a second test line could greatly decrease the variability of the outcome of the LFIA response. Any decrease in the variability of the LFIA response would greatly increase the performance of the system.

## 5.3 Impact on Society

Over the course of the COVID-19 pandemic, billions of rapid screening LFIAs were used by individuals. While these tests did assist individuals in making decisions about exposure and isolation, these tests are essentially analog devices, and there was no effective method for compiling the results at the scale they were being distributed. Interpreting these tests using a device which could not only standardize but digitize the result of every test makes available an immense wealth of data which would assist epidemiologists and policy makers in making decisions, as well as assessing the efficacy of previous decisions. This device is also inexpensive enough that in the case of another pandemic, distributing these devices per household may be an achievable goal.

This is one example of the impact a device such as this could have on the world, but it is far from the only benefit. All the industries which require testing of fluid samples can benefit from lowered limit of detection for their rapid, onsite testing, especially if it is being performed by a device which can readily connect with existing data tracking technologies.

In all, the LFIA market is slated to be a $97 billion market by 2030 [16]. With this device, every one of those $97 billion worth of LFIAs can be more effective, more accurate, standardized, and digitized.

## 5.4 Conclusion

In conclusion, this prototype has been much success in the validation experiments performed so far. We have demonstrated that the device is inexpensive, handheld, and battery powered. We have also demonstrated through several tests that the device can indeed greatly reduce the LOD of both sandwich and competitive style LFIAs, performing well on off-the-shelf LFIAs, as well as LFIAs created custom for the device. The device was also validated on humans, demonstrating a significant decrease in the LOD compared to nominal LOD even when considering all the complexities present in testing in real-world situations. Overall, this prototype is ready for the next stages in its development, moving it closer to an end-user device that one day could change the way that many at-home tests are performed, or even possible.

# Chapter 6: References


1. Jain, R.; Thakur, A.; Kaur, P.; Kim, K.-H.; Devi, P. Advances in imaging-assisted sensing techniques for heavy metals in water: Trends, challenges, and opportunities. *TrAC Trends in Analytical Chemistry* **2020**, *123*, 115758, doi:https://doi.org/10.1016/j.trac.2019.115758.
2. López_Marzo, A.M.; Pons, J.; Blake, D.A.; Merkoçi, A. High sensitive gold-nanoparticle based lateral flow Immunodevice for Cd2+ detection in drinking waters. *Biosensors and Bioelectronics* **2013**, *47*, 190-198, doi:https://doi.org/10.1016/j.bios.2013.02.031.
3. San Vicente de la Riva, B.; Costa-Fernández, J.M.; Pereiro, R.; Sanz-Medel, A. Spectrafluorimetric method for the rapid screening of toxic heavy metals in water samples. *Analytica Chimica Acta* **2002**, *451*, 203-210, doi:https://doi.org/10.1016/S0003-2670(01)01411-8.
4. Johnson, S.; Harrison, K.; Turner, A.D. Application of rapid test kits for the determination of Diarrhetic Shellfish Poisoning (DSP) toxins in bivalve molluscs from Great Britain. *Toxicon* **2016**, *111*, 121-129, doi:https://doi.org/10.1016/j.toxicon.2016.01.052.
5. Rastogi, S.; Dwivedi, P.D.; Khanna, S.K.; Das, M. Detection of Aflatoxin M1 contamination in milk and infant milk products from Indian markets by ELISA. *Food Control* **2004**, *15*, 287-290, doi:https://doi.org/10.1016/S0956-7135(03)00078-1.
6. Grothaus, G.D.; Bandla, M.; Currier, T.; Giroux, R.; Jenkins, G.R.; Lipp, M.; Shan, G.; Stave, J.W.; Pantella, V. Immunoassay as an Analytical Tool in Agricultural Biotechnology. *Journal of AOAC INTERNATIONAL* **2006**, *89*, 913-928, doi:10.1093/jaoac/89.4.913.
7. Kozel, T.R.; Burnham-Marusich, A.R. Point-of-Care Testing for Infectious Diseases: Past, Present, and Future. *Journal of Clinical Microbiology* **2017**, *55*, 2313-2320, doi:doi:10.1128/jcm.00476-17.
8. Chen, H.; Liu, K.; Li, Z.; Wang, P. Point of care testing for infectious diseases. *Clinica Chimica Acta* **2019**, *493*, 138-147, doi:https://doi.org/10.1016/j.cca.2019.03.008.
9. Wu, J.-L.; Tseng, W.-P.; Lin, C.-H.; Lee, T.-F.; Chung, M.-Y.; Huang, C.-H.; Chen, S.-Y.; Hsueh, P.-R.; Chen, S.-C. Four point-of-care lateral flow immunoassays for diagnosis

of COVID-19 and for assessing dynamics of antibody responses to SARS-CoV-2. *Journal of Infection* **2020**, *81*, 435-442, doi:https://doi.org/10.1016/j.jinf.2020.06.023.

10. Khelifa, L.; Hu, Y.; Jiang, N.; Yetisen, A.K. Lateral flow assays for hormone detection. *Lab on a Chip* **2022**, *22*, 2451-2475, doi:10.1039/D1LC00960E.
11. El-Aneed, A.; Cohen, A.; Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* **2009**, *44*, 210-230, doi:10.1080/05704920902717872.
12. Aydin, S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides* **2015**, *72*, 4-15, doi:https://doi.org/10.1016/j.peptides.2015.04.012.
13. Zhu, H.; Zhang, H.; Xu, Y.; Laššáková, S.; Korabečná, M.; Neužil, P. PCR Past, Present and Future. *BioTechniques* **2020**, *69*, 317-325, doi:10.2144/btn-2020-0057.
14. Di Nardo, F.; Chiarello, M.; Cavalera, S.; Baggiani, C.; Anfossi, L. Ten Years of Lateral Flow Immunoassay Technique Applications: Trends, Challenges and Future Perspectives. *Sensors* **2021**, *21*, 5185.
15. Wang, K.; Qin, W.; Hou, Y.; Xiao, K.; Yan, W. The application of lateral flow immunoassay in point of care testing: a review. *Nano Biomed. Eng* **2016**, *8*, 172-183.
16. Lateral Flow Assays Market by Application - Global Forecast to (2021 - 2030). Available online: https://www.alliedmarketresearch.com/rapid-tests-market (accessed on September 16, 2024).
17. Vollmer, M.; Möllmann, K.-P. Fundamentals of Infrared Thermal Imaging. In *Infrared Thermal Imaging*; 2017; pp. 1-106.
18. Thapa, D.; Samadi, N.; Patel, N.; Tabatabaei, N. Thermographic detection and quantification of THC in oral fluid at unprecedented low concentrations. *Biomed. Opt. Express* **2020**, *11*, 2178-2190, doi:10.1364/BOE.388990.
19. Thapa, D.; Samadi, N.; Tabatabaei, N. Handheld Thermo-Photonic Device for Rapid, Low-Cost, and On-Site Detection and Quantification of Anti-SARS-CoV-2 Antibody. *IEEE sensors journal* **2021**, *21*, 18504-18511, doi:10.1109/jsen.2021.3089016.
20. Bahadır, E.B.; Sezgintürk, M.K. Lateral flow assays: Principles, designs and labels. *TrAC Trends in Analytical Chemistry* **2016**, *82*, 286-306, doi:https://doi.org/10.1016/j.trac.2016.06.006.

21. Koczula, K.M.; Gallotta, A. Lateral flow assays. *Essays in biochemistry* **2016**, *60*, 111-120, doi:10.1042/ebc20150012.
22. Hayden, D.; Anacleto, S.; Archonta, D.-E.; Khalil, N.; Pennella, A.; Qureshi, S.; Séguin, A.; Tabatabaei, N. Arduino-Based Sensing Platform for Rapid, Low-Cost, and High-Sensitivity Detection and Quantification of Analytes in Fluidic Samples. *Engineering Proceedings* **2022**, *27*, doi:10.3390/ecsa-9-13277.
23. Zhang, D.; Huang, L.; Liu, B.; Ni, H.; Sun, L.; Su, E.; Chen, H.; Gu, Z.; Zhao, X. Quantitative and ultrasensitive detection of multiplex cardiac biomarkers in lateral flow assay with core-shell SERS nanotags. *Biosensors and Bioelectronics* **2018**, *106*, 204-211, doi:https://doi.org/10.1016/j.bios.2018.01.062.
24. Mak, W.C.; Beni, V.; Turner, A.P.F. Lateral-flow technology: From visual to instrumental. *TrAC Trends in Analytical Chemistry* **2016**, *79*, 297-305, doi:https://doi.org/10.1016/j.trac.2015.10.017.
25. Urusov, A.E.; Zherdev, A.V.; Dzantiev, B.B. Towards Lateral Flow Quantitative Assays: Detection Approaches. *Biosensors* **2019**, *9*, 89.
26. Nguyen, V.-T.; Song, S.; Park, S.; Joo, C. Recent advances in high-sensitivity detection methods for paper-based lateral-flow assay. *Biosensors and Bioelectronics* **2020**, *152*, 112015, doi:https://doi.org/10.1016/j.bios.2020.112015.
27. Park, J. An Optimized Colorimetric Readout Method for Lateral Flow Immunoassays. *Sensors* **2018**, *18*, 4084.
28. Foysal, K.H.; Seo, S.E.; Kim, M.J.; Kwon, O.S.; Chong, J.W. Analyte Quantity Detection from Lateral Flow Assay Using a Smartphone. *Sensors* **2019**, *19*, 4812.
29. Serebrennikova, K.; Samsonova, J.; Osipov, A. Hierarchical Nanogold Labels to Improve the Sensitivity of Lateral Flow Immunoassay. *Nano-Micro Letters* **2017**, *10*, 24, doi:10.1007/s40820-017-0180-2.
30. Plouffe, B.D.; Murthy, S.K. Fluorescence-based lateral flow assays for rapid oral fluid roadside detection of cannabis use. *ELECTROPHORESIS* **2017**, *38*, 501-506, doi:https://doi.org/10.1002/elps.201600075.
31. Taranova, N.A.; Berlina, A.N.; Zherdev, A.V.; Dzantiev, B.B. 'Traffic light' immunochromatographic test based on multicolor quantum dots for the simultaneous

detection of several antibiotics in milk. *Biosensors and Bioelectronics* **2015**, *63*, 255-261, doi:https://doi.org/10.1016/j.bios.2014.07.049.

32. Fang, C.-C.; Chou, C.-C.; Yang, Y.-Q.; Wei-Kai, T.; Wang, Y.-T.; Chan, Y.-H. Multiplexed Detection of Tumor Markers with Multicolor Polymer Dot-Based Immunochromatography Test Strip. *Analytical Chemistry* **2018**, *90*, 2134-2140, doi:10.1021/acs.analchem.7b04411.
33. Das, R.S.; Agrawal, Y.K. Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy* **2011**, *57*, 163-176, doi:https://doi.org/10.1016/j.vibspec.2011.08.003.
34. Fu, X.; Cheng, Z.; Yu, J.; Choo, P.; Chen, L.; Choo, J. A SERS-based lateral flow assay biosensor for highly sensitive detection of HIV-1 DNA. *Biosensors and Bioelectronics* **2016**, *78*, 530-537, doi:https://doi.org/10.1016/j.bios.2015.11.099.
35. Xiao, M.; Xie, K.; Dong, X.; Wang, L.; Huang, C.; Xu, F.; Xiao, W.; Jin, M.; Huang, B.; Tang, Y. Ultrasensitive detection of avian influenza A (H7N9) virus using surface-enhanced Raman scattering-based lateral flow immunoassay strips. *Analytica Chimica Acta* **2019**, *1053*, 139-147, doi:https://doi.org/10.1016/j.aca.2018.11.056.
36. Homola, J. Present and future of surface plasmon resonance biosensors. *Analytical and Bioanalytical Chemistry* **2003**, *377*, 528-539, doi:10.1007/s00216-003-2101-0.
37. Amendola, V.; Pilot, R.; Frasconi, M.; Maragò, O.M.; Iatì, M.A. Surface plasmon resonance in gold nanoparticles: a review. *Journal of Physics: Condensed Matter* **2017**, *29*, 203002, doi:10.1088/1361-648X/aa60f3.
38. Govorov, A.O.; Richardson, H.H. Generating heat with metal nanoparticles. *Nano Today* **2007**, *2*, 30-38, doi:https://doi.org/10.1016/S1748-0132(07)70017-8.
39. Qin, Z.; Chan, W.C.W.; Boulware, D.R.; Akkin, T.; Butler, E.K.; Bischof, J.C. Significantly Improved Analytical Sensitivity of Lateral Flow Immunoassays by Using Thermal Contrast. *Angewandte Chemie International Edition* **2012**, *51*, 4358-4361, doi:https://doi.org/10.1002/anie.201200997.
40. Wang, Y.; Qin, Z.; Boulware, D.R.; Pritt, B.S.; Sloan, L.M.; González, I.J.; Bell, D.; Rees-Channer, R.R.; Chiodini, P.; Chan, W.C.W.; et al. Thermal Contrast Amplification Reader Yielding 8-Fold Analytical Improvement for Disease Detection with Lateral Flow

Assays. *Analytical Chemistry* **2016**, *88*, 11774-11782, doi:10.1021/acs.analchem.6b03406.

41. Almond, D.P.; Patel, P. *Photothermal science and techniques*; Springer Science & Business Media: 1996; Volume 10.
42. Hellen, A.; Matvienko, A.; Mandelis, A.; Finer, Y.; Amaechi, B.T. Optothermophysical properties of demineralized human dental enamel determined using photothermally generated diffuse photon density and thermal-wave fields. *Appl. Opt.* **2010**, *49*, 6938-6951, doi:10.1364/AO.49.006938.
43. Samadi, N. Active Thermography Using Cellphone Attachment Infrared Camera. York University, 2020.
44. Mitchell, T.M. *Machine Learning*; McGraw-Hill: 1997; p. 414.
45. Mertz, D. *Cleaning data for effective data science : doing the other 80% of the work with Python, R, and command-line tools*; Packt Publishing: Birmingham, England ;, 2021.
46. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: 2017.
47. Taunk, K.; De, S.; Verma, S.; Swetapadma, A. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. In Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 15-17 May 2019, 2019; pp. 1255-1260.
48. Salcedo-Sanz, S.; Rojo-Álvarez, J.L.; Martínez-Ramón, M.; Camps-Valls, G. Support vector machines in engineering: an overview. *WIREs Data Mining and Knowledge Discovery* **2014**, *4*, 234-267, doi:https://doi.org/10.1002/widm.1125.
49. *Kernel Methods in Computational Biology*; The MIT Press: 2004.
50. Hastie, T.; Tibshirani, R.; Friedman, J. Additive Models, Trees, and Related Methods. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Hastie, T., Tibshirani, R., Friedman, J., Eds.; Springer New York: New York, NY, 2009; pp. 295-336.
51. Quinlan, J.R. *C4.5 : programs for machine learning*; Morgan Kaufmann Publishers: San Mateo, Calif, 1993.
52. Zhang, H. The optimality of naive Bayes. *Aa* **2004**, *1*, 3.

53. Svozil, D.; Kvasnicka, V.; Pospichal, J.í. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems* **1997**, *39*, 43-62, doi:https://doi.org/10.1016/S0169-7439(97)00061-0.
54. Zou, J.; Han, Y.; So, S.-S. Overview of Artificial Neural Networks. In *Artificial Neural Networks: Methods and Applications*, Livingstone, D.J., Ed.; Humana Press: Totowa, NJ, 2009; pp. 14-22.
55. Wythoff, B.J. Backpropagation neural networks: A tutorial. *Chemometrics and Intelligent Laboratory Systems* **1993**, *18*, 115-155, doi:https://doi.org/10.1016/0169-7439(93)80052-J.
56. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.-R. Efficient BackProp. In *Neural Networks: Tricks of the Trade: Second Edition*, Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 9-48.
57. Bottou, L. Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nımes* **1991**, *91*, 12.
58. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations* **2014**.
59. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* **1995**, *16*, 1190-1208, doi:10.1137/0916069.
60. Marti, D.; Aasbjerg, R.N.; Andersen, P.E.; Hansen, A.K. MCmatlab: an open-source, user-friendly, MATLAB-integrated three-dimensional Monte Carlo light transport solver with heat diffusion and tissue damage. *Journal of biomedical optics* **2018**, *23*, 1-6, doi:10.1117/1.jbo.23.12.121622.
61. Bertsimas, D.; Tsitsiklis, J. Simulated Annealing. *Statistical Science* **1993**, *8*, 10-15, 16.
62. Azkur, A.K.; Akdis, M.; Azkur, D.; Sokolowska, M.; van de Veen, W.; Brüggen, M.-C.; O'Mahony, L.; Gao, Y.; Nadeau, K.; Akdis, C.A. Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. *Allergy* **2020**, *75*, 1564-1581, doi:https://doi.org/10.1111/all.14364.
63. Hassani, H. A Brief Introduction to Singular Spectrum Analysis. 2009.
64. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Physical Review E* **2004**, *69*, 066138, doi:10.1103/PhysRevE.69.066138.

65. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* **2002**, *46*, 389-422, doi:10.1023/A:1012487302797.