Reimagining Virus Diagnostics

Novel Approaches for Virus Detection Supplemented by Biophysical Characterisation Studies of Viral Envelope Interactions

Nicolas Shiaelis
St Anne's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

HT 2025

Disclaimer

I hereby declare that the work presented in this thesis is my own, except where explicitly stated otherwise in the text. Contributions from collaborators, supervisors, or external sources have been duly acknowledged and referenced.

The following exceptions apply:

- **Lab-grown viruses:** Provided by Dr. Nicole Robb.
- **Infectious bronchitis virus (IBV):** Supplied by Erica Bickerton at the Pirbright Institute.
- **Super-resolution data analysis:** Data was analysed by Dr. Andrew McMahon.
- **Zeiss Elyra data:** Collected by Zeiss Germany technicians. Samples and DNAs were prepared and sent by Dr Nicole Robb, and I performed the analysis.
- **Clinical Samples qPCR**: Performed by the staff of the Clinical Microbiology Lab at the John Radcliffe Hospital, Oxford University Hospitals.
- **Neutron reflectometry data:** Data in Chapter 7 were provided by Dr. Luke Clifton.

Nicolas Shiaelis
2025

Academic Achievements

List of Publications

- **Shiaelis, N.**, Clifton, L. and McCluskey, A. (2024). Investigating model influence on the analytical resolution of neutron reflectometry. *arXiv*.
- **Shiaelis, N.**, Tometzki, A., Peto, L., McMahon, A., Hepp, C., Bickerton, E., ... & Robb, N. C. (2022). Virus detection and identification in minutes using single-particle imaging and deep learning. *ACS Nano*, 17(1), 697–710.
- Hepp, C., **Shiaelis, N.**, Robb, N. C., Vaughan, A., Matthews, P. C., Stoesser, N., ... & Kapanidis, A. N. (2021). Viral detection and identification in 20 min by rapid single-particle fluorescence in-situ hybridization of viral RNA. *Scientific Reports*, 11(1), 19579.
- Cording, A. C., **Shiaelis, N.**, Petridi, S., Middleton, C. A., Wilson, L. G., & Elliott, C. J. (2017). Targeted kinase inhibition relieves slowness and tremor in a Drosophila model of LRRK2 Parkinson's disease. *npj Parkinson's Disease*, 3(1), 34.

Patents & IP

- **Shiaelis, N.**, "Method and System for Detecting Characteristics of Particles or Particle Environments, and Method of Training a Machine Learning Model.", N422096GB, Filing date: 23rd August 2023.
- **Shiaelis, N.**, Robb, N., "Method of Diagnosing a Biological Entity, and Diagnostic Device.", PCT/GB2021/050990, Filing date: 27th April 2020.
- **Shiaelis, N.**, "Shiaelis Pathogen Detection software." Copyright University of Oxford.

Awards and Recognition

- **Oxford Interdisciplinary Bioscience DTP Impact Awards: Entrepreneurship (Winner)** – Oxford, UK, 2022.
- **A.G. Leventis Foundation Scholar (£17k)** – Oxford, UK, 2019–2022.
- **Oxford Interdisciplinary Bioscience DTP Impact Awards: COVID-19 Response (Commendation)** – Oxford, UK, 2021.
- **Doctoral Researcher Awards (Honorable Mention)** – London, UK, 2021.

Nicolas Shiaelis

2025

To my mum, Maria,

Your unwavering support, boundless love, and countless sacrifices have been the bedrock of my academic journey. I dedicate this work to you with immense gratitude and love.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors, Achillefs Kapanidis, Nicole Robb, and Luke Clifton, for their unwavering support, guidance, and mentorship throughout my DPhil journey. Your expertise and insights have been invaluable, and I feel truly fortunate to have had you as my mentors.

Special thanks go to the past and present members of the Kapanidis and Robb labs, especially Christof, Andrew, Stelios, Hafez, Piers, Becky, Rasched, Miri, Emma, Barak, Ollie, Jacob and Sanna for the stimulating discussions, late-night brainstorming sessions, and for fostering a collaborative and friendly research environment.

I am also deeply grateful to the Modernising Medical Microbiology (MMM) team and the clinical staff at the microbiology lab at the John Radcliffe Hospital. Not only did they provide me with the opportunity to work with clinical samples, but they also embraced me as a member of their group. Some of my fondest memories from my DPhil come from their socials, including pub trips, escape rooms, and pottery classes. It has been a privilege to collaborate with Nicole Stoesser, whose significant contributions shaped the clinical aspects of my research. Special thanks to Hermione, Conor, Leon, and Kev, and especially to Ali for her patience, kindness, and outstanding organisation of socials.

I would like to express my gratitude to my MPhys masters students, Alex, Yuanmin, and Charlie, for their dedication and contributions. Your enthusiasm, curiosity, and hard work have been a constant source of inspiration, and it has been a privilege to work alongside you.

I am thankful to the Interdisciplinary Biosciences DTP and the BBSRC for providing the essential resources and funding that supported my research endeavors, as well as, the Leventis foundation for its generous scholarship.

To Jochem, Jack, Jai, Jeremie, Jacques, Andrew, Amy, Anya and Lisa thank you for the shared moments of respite, laughter, and encouragement during the challenging times. Your camaraderie has made this journey both enjoyable and memorable.

I extend my heartfelt appreciation to my family, especially my parents, Costas and Maria, my siblings, Giota and Giorgos, my grandparents Chrysoulla and Panayiotis and last but not least my aunt Dora, for their unwavering love, encouragement, and belief in my abilities. This DPhil is dedicated to my mother, Maria, who fostered my love for science and nurtured my curiosity. Your sacrifices and support have been the foundation upon which I built my academic pursuits.

To Nami, my furry cheerleader, thank you for your quiet companionship, warm presence, and the way you always reminded me to find joy in small moments, even amidst great challenges.

Lastly, to my dear George, thank you for your endless patience, understanding, and love. Your presence has been a constant source of strength and motivation.

Nicolas Shiaelis
2025

Abstract

The development of rapid, robust, and scalable virus diagnostics is a critical global health priority, as highlighted by the COVID-19 pandemic. This thesis introduces innovative diagnostic approaches that integrate biophysical and computational techniques to enhance virus detection and characterisation. Central to this work is the calcium-mediated labelling of viral envelopes with single-stranded DNA, enabling the rapid immobilisation and fluorescent tagging of viral particles. These methods are coupled with fluorescence microscopy and machine learning to achieve fast and accurate virus identification.

In addition to fluorescence-based diagnostics, this thesis explores a diffusion-based detection method, leveraging single-particle tracking and statistical modeling to classify viral particles within seconds. This complementary approach offers ultra-fast detection capabilities, reducing diagnostic times to under one minute while maintaining high specificity and sensitivity.

Key contributions include the biophysical characterization of cation-mediated interactions, focusing on the effects of pH, ion concentration, and lipid composition on labeling efficiency. Supported lipid bilayers and virus-like particles were employed as biomimetic systems, elucidating the roles of membrane heterogeneity and surface proteins. The machine learning pipeline developed for this work distinguishes between closely related virus strains and accurately identifies SARS-CoV-2 in clinical samples, achieving high sensitivity and specificity.

By integrating these diagnostic methods, this thesis addresses limitations of established technologies and aligns with the ASSURED and REASSURED criteria for point-of-care diagnostics. These advancements not only contribute to pandemic preparedness but also deepen our understanding of virus-membrane interactions, with significant implications for the development of next-generation diagnostic and therapeutic platforms.

Contents

Outline of the Thesis

This thesis is organised as follows:

- **Chapter 1: Introduction** This chapter establishes the motivation for the research, emphasising the urgent need for advanced virus diagnostics in global health, economic stability, and pandemic preparedness. Drawing on lessons from the COVID-19 pandemic, it highlights gaps in existing technologies and sets the stage for the novel approaches introduced in this thesis. A comprehensive review of established and emerging diagnostic technologies, including PCR, ELISA, CRISPR-based tools, and biosensors, is presented. The review evaluates the strengths and limitations of current methods, aligning them with the ASSURED and REASSURED frameworks, and positions the diagnostic approaches developed in this thesis as a compelling solution to address these challenges.

- **Chapter 2: Methods** This chapter details the experimental and analytical methodologies employed throughout the research. These include fluorescence microscopy for imaging, machine learning for classification, and neutron reflectometry for studying lipid bilayer systems. Together, these methods form the technical foundation of the thesis.

- **Chapter 3: Cation-Mediated Labelling Characterisation** This chapter investigates calcium-mediated labelling of viral envelopes, focusing on how factors such as pH, DNA length, and lipid composition influence labelling efficiency. Lipid bilayers are employed specifically as model systems to study the effects of lipid composition and to understand the underlying labelling mechanism.

- **Chapter 4: Detection of Immobilized Virions in Minutes Using Deep Learning** This chapter integrates the calcium-mediated labelling method with fluorescence microscopy and deep learning. It presents a pipeline for image segmentation and machine learning-based classification, enabling rapid and accurate detection of immobilized virions.

- **Chapter 5: Clinical Validation of Detection of Immobilised Virions** This chapter validates the diagnostic framework using clinical samples. It highlights the sensitivity and specificity of the calcium-labelling method and explores viral differentiation based on charge interactions with lipid bilayers.

- **Chapter 6: Diffusion-Based Detection of Virions in Under One Minute** This chapter introduces a complementary method for ultra-fast virus detection. It leverages diffusion-based single-particle tracking and statistical models, such as PCA and KNN, to classify virions in under one minute.

- **Chapter 7: Effective Resolution of Neutron Reflectometry** This chapter examines the resolution limits of neutron reflectometry for characterizing lipid bilayers. It provides critical structural insights into the bilayer systems used in this research and connects these findings to the calcium-mediated virus detection methods.

- **Chapter 8: Conclusion and Future Directions** This chapter discusses innovative diagnostic approaches that shift specificity from molecular interactions to computational analysis, combining biophysical labeling with machine learning for rapid and scalable solutions. Key challenges, including cost and real-world implementation, are addressed alongside proposed future directions, such as algorithm refinement and expanded pathogen detection. The chapter concludes by highlighting the transformative potential of these technologies for global healthcare and pandemic preparedness.

Chapter 1

Introduction

1.1 Motivation: Why are diagnostics important?

In today's world, the significance and relevance of timely, accurate, and adaptable viral diagnostics cannot be overstated. Our global health landscape is teeming with both established and emerging viral threats [1] that have significant implications for human and animal health, economic stability, and our evolving climate situation [2].

The importance of viral diagnostics in human and animal health is twofold. On a microscale, early detection is crucial for effective treatment, as viral diseases can quickly progress to life-threatening stages if not adequately managed. On a broader scale, rapid identification of viral pathogens can prevent potential outbreaks from escalating into full-blown epidemics or pandemics, reducing morbidity and mortality rates. The 2019-2023 COVID-19 pandemic dramatically underscored the importance of this facet of disease control, with the virus claiming millions of lives and causing profound societal disruption due to a lack of early detection and mitigation strategies.

Beyond their substantial impact on health, viruses profoundly influence global economic dynamics. The COVID-19 pandemic underscored this fact, with the World Bank projecting its global economic cost could surpass $12 trillion over two years [3]. This highlights the significant fiscal burden posed by unchecked viral diseases. The economic ramifications of such diseases transcend immediate healthcare expenditures, encompassing a broad spectrum of both direct and indirect financial consequences.

Direct financial burdens arise from diagnosing and treating the disease, including testing, hospital care, medications, and the execution of public health measures such as contact tracing and disease surveillance. These costs, especially for severe viral diseases requiring intensive or prolonged care, can be quite significant. Furthermore, public health interventions needed to control viral spread compound these direct economic costs.

Indirect financial impacts often exceed the direct costs and are multifaceted in nature. They include diminished productivity due to illness-induced workforce incapacitation and business closures, disruptions in local and global supply chains inducing shortages and sparking inflation, and declines in consumer and investor confidence. Sectors that hinge on physical interactions, like tourism and retail, are especially susceptible to these impacts. Uncertainties stemming from viral diseases can trigger a decline in spending, a withdrawal of investments, and market volatility. These widespread economic repercussions underscore the pivotal role of effective viral diagnostics in safeguarding not just public health, but also the stability and resilience of the global economy.

Another vital aspect of viral detection pertains to animal health and, by extension, food security. Viruses that infect livestock, such as Avian Influenza, Swine Flu, and Foot-and-Mouth Disease, can lead to severe productivity loss, undermining agricultural economies and threatening global food supplies. Hence, swift detection and control of such viral diseases are imperative to ensure food security and the stability of farming industries. Recent examples include the highly pathogenic H5N1 avian influenza which as of June 2023 is widespread in wild birds with sporadic outbreaks in poultry flocks [4] and the outbreak of the African Swine Flu in China [5] which resulted in dozens of millions of pigs being culled and an economic loss that accounts for 0.78% of China's gross domestic product in 2019 [6].

Lastly, climate change, in its capacity to alter ecosystems and migration patterns, influences the emergence and spread of viruses. Changes in temperature and precipitation patterns can expand the geographical range of vector-borne viruses, like Dengue and Zika, posing a significant risk to populations previously unaffected. Such climate change-induced viral spread highlights the importance of developing robust and adaptable viral detection methods to keep pace with our rapidly changing world.

In summary, the importance of viral diagnostics in the modern world is driven by its profound implications for human and animal health, global economic stability, and the increasingly significant interplay with our changing climate. Given these factors, the necessity of developing novel virus detection methods that are quick, precise, and adaptable to new viral threats is both urgent and paramount.

1.1.1 Lessons from the COVID-19 Pandemic

The COVID-19 pandemic highlighted several critical lessons in the role of diagnostics for managing infectious diseases, emphasizing the need for a robust and scalable diagnostic infrastructure to ensure global health security. One of the most significant lessons was the importance of rapid, accurate, and accessible diagnostic tests. In the early stages of the pandemic, the inability to quickly scale up testing capacity and significant delays in test result turnaround times hampered efforts to control viral spread. Many regions, especially rural and under-resourced areas, struggled to access adequate testing, revealing significant vulnerabilities in both national and international health systems [7, 8].

Scalability emerged as a key factor during the pandemic, as laboratories worldwide were overwhelmed by surges in case numbers. As countries struggled to ramp up testing capacity, delays in results compromised the effectiveness of contact tracing and isolation measures. Looking ahead, future diagnostic systems must be flexible and capable of rapid scaling to accommodate sudden increases in demand. This will likely involve developing decentralized, portable diagnostic tools and ensuring that infrastructure can expand quickly during health crises[7].

Accessibility was another major challenge, particularly for rural and remote populations that lacked adequate diagnostic infrastructure. Many of these regions experienced delays in identifying cases, leading to slower public health responses. Equitable access to diagnostic tools is essential in managing future pandemics, which could be addressed by creating affordable point-of-care testing options and improving distribution systems to reach underserved areas [7].

Speed of diagnostics is equally important. The pandemic demonstrated that delays in test results, due to overwhelmed laboratories [9, 10] or logistical bottlenecks, slowed the implementation of timely public health interventions such as isolation, contact tracing, and treatment. Advances in molecular diagnostics, such as point-of-care tests that deliver results within minutes, have the potential to significantly reduce these delays in future outbreaks, enabling more rapid decision-making and containment efforts[7].

A critical takeaway was the need for integration of diagnostic data into public health surveillance systems. Fragmented data-sharing processes between laboratories, healthcare facilities, and public health authorities hindered efforts to track and respond to the virus. Going forward, diagnostic platforms must ensure seamless integration with centralized public health databases to enable real-time monitoring and rapid, coordinated responses to future outbreaks [8].

The pandemic also underscored the importance of international collaboration in diagnostics. The development and global distribution of COVID-19 tests required unprecedented levels of cooperation between countries, research institutions, and private companies. However, disparities in access, particularly for low- and middle-income countries, revealed the need for better international agreements on the equitable distribution of diagnostics and the sharing of innovations. Establishing global platforms for data sharing and collaboration will be crucial for managing future pandemics [7, 8].

In conclusion, the COVID-19 pandemic has demonstrated that without scalable, accessible, rapid, and well-integrated diagnostics, efforts to contain and mitigate infectious diseases are severely compromised. The lessons learned emphasize the need for investment in diagnostic innovation and infrastructure to ensure global preparedness for future health crises. The integration of these key principles will shape the development of future diagnostic technologies and systems, helping to build more resilient health frameworks capable of responding effectively to emerging threats [8].

1.2 Virus Diagnostics

Virus diagnostics are a vital tool in the management and control of infectious diseases, utilised in a variety of contexts by numerous professionals. They are deployed in diverse situations, such as during disease outbreaks to identify the causative agent, in routine surveillance to monitor disease prevalence, and in clinical settings to diagnose individual patients.

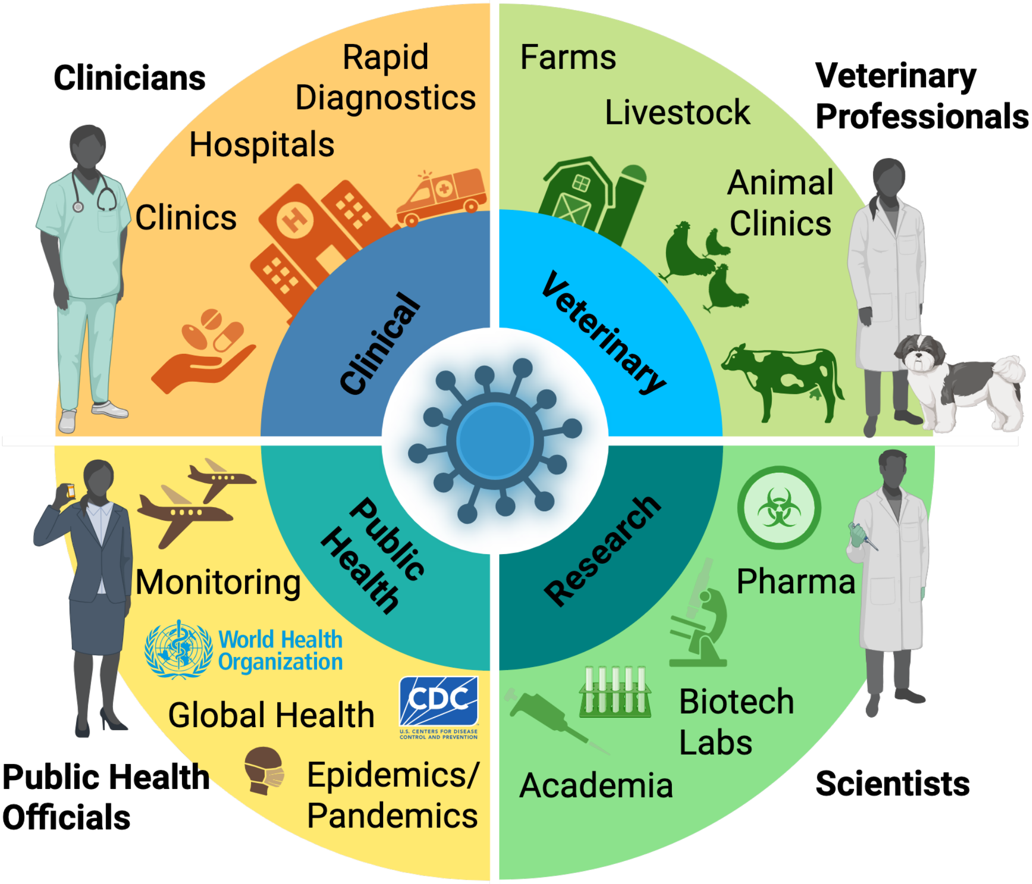


Figure 1.1: Radial schematic illustrating the applications, settings, and users of virus diagnostics. At the centre, a virus symbol represents the focus of diagnostic efforts. Surrounding this core, the first ring categorizes the primary purposes of virus diagnostics into four domains: clinical, veterinary, research, and surveillance and outbreak control. The second ring highlights the settings in which these diagnostics are applied, including hospitals, farms, academia, and public health environments, each represented with illustrative icons and contextual keywords. The outer ring depicts the end users of virus diagnostics, such as clinicians, veterinary professionals, scientists and researchers, public health authorities, and industry specialists. This figure illustrates the multidimensional role of virus diagnostics across disciplines, contexts, and professional domains. *Figure created with BioRender.com.*

1.2.1 Where Are Virus Diagnostics Used

Virus diagnostics are applied across various settings, including clinical, veterinary, research, and public health environments, depending on the specific needs for detecting, monitoring, and managing viral infections. In clinical settings, diagnostics are primarily used in hospitals, clinics, and diagnostic laboratories. These facilities rely on advanced molecular tools, such as PCR machines, and rapid point-of-care tests for the timely diagnosis of viral infections. This is particularly critical in outpatient settings [11], emergency rooms, and intensive care units [12, 13], where swift identification of viral pathogens can guide appropriate treatment and infection control measures.

In veterinary settings, virus diagnostics are crucial in animal hospitals, veterinary clinics, and farms. These diagnostics help detect viral diseases in livestock, companion animals, and wildlife, playing a key role in preventing outbreaks that could have severe economic and health implications. For example, routine viral testing supports the agricultural industry by identifying contagious diseases early and mitigating the risk of large-scale animal infections. Diagnostics are also vital for controlling zoonotic diseases, where viruses can transfer from animals to humans.

In research environments, virus diagnostics are integral to the work conducted in academic institutions, pharmaceutical companies, and biotech firms. These diagnostics enable the study of viral pathogenesis, the monitoring of antiviral drug efficacy, and vaccine development. In field research, portable diagnostic tools help detect viral infections in remote areas or among wildlife, contributing to broader disease surveillance and epidemiological studies.

Within public health settings, virus diagnostics are essential for tracking outbreaks and managing disease surveillance. Public health laboratories, government agencies, and international organizations, such as the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC), use these diagnostics to monitor the spread of infectious diseases. During epidemics and pandemics, such as the COVID-19 crisis, virus diagnostics are deployed in large-scale testing programs to detect cases early and help contain the spread of infections. They are also employed at borders, ports, and airports to screen travelers for contagious viral diseases, thereby preventing international transmission.

Lastly, in industrial and commercial settings, virus diagnostics are used during the quality control processes in vaccine production. Ensuring that live viral products are properly characterized is critical to the development of safe and effective vaccines.

1.2.2 What Are Virus Diagnostics Used For

Virus diagnostics are essential for the detection, monitoring, and management of viral infections in both clinical and public health contexts. In clinical practice, diagnostics help healthcare providers identify the specific viral pathogen responsible for an infection, enabling precise treatment decisions and timely interventions. Early detection of viruses like influenza, HIV, or hepatitis is crucial for initiating antiviral therapies, which can prevent severe disease progression. Virus diagnostics also play a role in monitoring the progression of chronic viral infections, helping clinicians adjust treatments and manage patient outcomes effectively.

In public health, virus diagnostics are vital for outbreak management and disease surveillance. During outbreaks of emerging viruses, such as COVID-19 or Ebola, diagnostics are key for identifying infected individuals, enabling isolation measures, and tracking viral spread in real time. This information helps public health authorities implement targeted interventions, such as vaccination campaigns or quarantine protocols, to prevent widespread transmission. Moreover, virus diagnostics support the ongoing surveillance of endemic viruses, helping detect new viral strains or mutations that may lead to future epidemics or pandemics, guiding vaccination strategies, and influencing the development of new therapies.

In animal health, virus diagnostics are used to control viral diseases in livestock, pets, and wildlife. These diagnostics allow veterinarians to identify infections early, which is essential for preventing the spread of contagious diseases within animal populations. This is particularly important in the agricultural sector, where viral outbreaks can lead to significant economic losses. For instance, the rapid identification of viruses like foot-and-mouth disease or avian influenza enables swift containment measures to protect herds and flocks. Additionally, virus diagnostics support the monitoring of zoonotic diseases, such as rabies and avian flu, which pose a risk to both animal and human health.

In research, virus diagnostics are indispensable for advancing our understanding of viral biology, immune responses, and pathogen evolution. Laboratories utilize diagnostics to identify viruses in experimental studies, assess the effectiveness of antiviral drugs and vaccines, and track viral mutations. These diagnostics are crucial in vaccine development, where they help monitor immune responses to experimental vaccines and evaluate their efficacy. Furthermore, diagnostics play a role in gene therapy studies, where viruses are used as vectors to deliver therapeutic genes, necessitating precise viral characterization throughout the process.

1.2.3 Who Uses Virus Diagnostics

A broad spectrum of professionals across healthcare, veterinary medicine, public health, and scientific research depend on virus diagnostics. In clinical settings, healthcare providers—including physicians, nurses, and clinical laboratory technicians—use virus diagnostics to identify viral infections in patients. The results of these tests are crucial for guiding treatment decisions, administering antiviral therapies, and managing infection control protocols. Public health officials also rely on diagnostic data to monitor viral outbreaks, conduct disease surveillance, and implement public health interventions during outbreaks or epidemics.

In veterinary medicine, veterinarians and animal health technicians use virus diagnostics to detect viral infections in animals, particularly in agriculture, where livestock diseases can have significant economic consequences. These professionals also play a role in controlling zoonotic diseases by identifying and managing viral infections in animals that have the potential to infect humans.

Researchers in academic institutions, pharmaceutical companies, and biotech firms use virus diagnostics to study viral mechanisms, track the evolution of viral pathogens, and develop new antiviral drugs and vaccines. In research fields such as virology and immunology, virus diagnostics enable the precise identification of viral strains in experimental models and clinical trials. This is essential for understanding immune responses, viral behaviors, and the development of new therapeutic approaches.

Government agencies and non-governmental organizations (NGOs) involved in global health initiatives also use virus diagnostics for monitoring emerging viruses and controlling the spread of infectious diseases. For example, organizations like the WHO and CDC use diagnostics to track viral outbreaks and inform broader epidemiological efforts to address global health threats. Finally, private companies involved in the development of diagnostic technologies and vaccines use virus diagnostics for quality control and validation throughout the product development process, ensuring that their products meet safety and efficacy standards.

1.3 Established Diagnostic Technologies

1.3.1 Polymerase Chain Reaction(PCR)

Polymerase Chain Reaction (PCR) is a widely used molecular biology technique that allows for the amplification of specific DNA sequences [14]. It is highly sensitive and specific, making it a cornerstone of modern diagnostic methods for viral detection. PCR involves the use of short DNA sequences known as primers, which are complementary to the target DNA region. The process includes repeated cycles of heating and cooling, enabling the DNA polymerase enzyme to synthesize new DNA strands from the original template [15]. The main steps in PCR are denaturation, annealing, and extension. During denaturation, the double-stranded DNA is heated to around 95°C to separate it into two single strands. In the annealing step, the temperature is lowered to 50-65°C to allow the primers to bind to the complementary sequences on the single-stranded DNA. Finally, during the extension step, the temperature is raised to approximately 72°C, the optimal temperature for DNA polymerase to synthesize new DNA strands by adding nucleotides to the primers. This cycle is repeated multiple times (typically 25-35 cycles), leading to exponential amplification of the target DNA sequence.

PCR is used for a variety of diagnostic purposes, including the detection of viral DNA [16], quantification of viral load [17], and genotyping and mutation detection. It is essential in diagnosing infections such as Hepatitis B [18], Herpes Simplex Virus [19], and Human Papillomavirus [20, 21, 22], as well as in monitoring the progression of viral infections like HIV [23, 16]. Moreover, PCR's versatility allows it to be adapted to detect a wide range of viruses, making it a highly valuable tool in the field of diagnostics. One of the main advantages of PCR is its high sensitivity and specificity, enabling the detection of minute amounts of viral DNA [24]. Additionally, quantitative capabilities of modified versions of PCR, such as quantitative PCR (qPCR), allow for the precise measurement of viral load.

However, PCR has some limitations. It requires specialized equipment such as PCR machines and thermocyclers, as well as trained personnel to properly execute the technique. Furthermore, the high sensitivity of PCR makes it prone to contamination, which can lead to false-positive results. Despite these limitations, PCR remains a fundamental technology in viral diagnostics due to its accuracy and adaptability.

1.3.1.1 Real-Time Reverse Transcriptase PCR (RT-PCR)

Real-Time Reverse Transcriptase PCR (RT-PCR) is a variation of PCR that enables the detection and quantification of RNA viruses by converting RNA into complementary DNA (cDNA) using reverse transcriptase before amplification [25]. RT-PCR involves two main steps: reverse transcription [26] and real-time PCR amplification [27]. In the reverse transcription step, RNA is reverse transcribed into cDNA using the enzyme reverse transcriptase. This cDNA is then amplified using PCR. Real-time PCR allows for the monitoring of the amplification process in real time through the use of fluorescent dyes or probes that bind to the DNA and emit fluorescence as the DNA is amplified.

RT-PCR is crucial for detecting and quantifying RNA viruses, including SARS-CoV-2, Influenza viruses, and HIV-1. It has been widely used in the diagnosis of COVID-19 by detecting viral RNA in respiratory samples, and it plays a significant role in identifying and subtyping different strains of influenza. In the case of HIV-1, RT-PCR is used to monitor viral load in patients to manage antiretroviral therapy effectively. The ability to provide immediate results during the amplification process enables rapid diagnosis, making RT-PCR an invaluable tool in clinical settings.

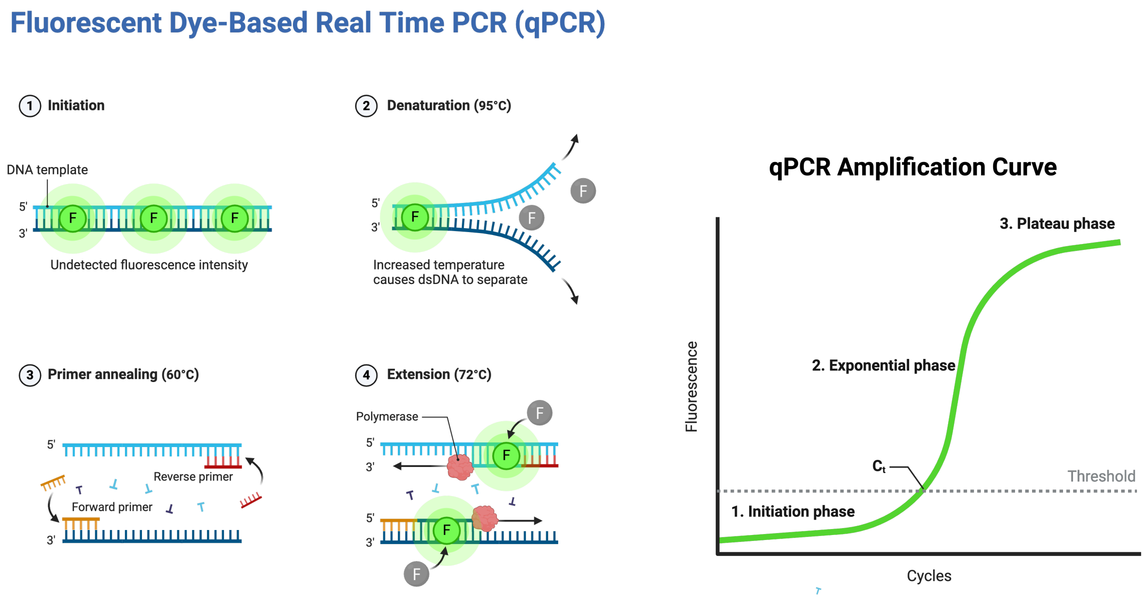
The main advantages of RT-PCR include real-time monitoring and quantification, which allows for precise measurement of viral RNA levels. This method is highly sensitive and specific, capable of detecting low levels of viral RNA, which is essential for early and accurate diagnosis. However, RT-PCR also has its limitations. The process is more complex than traditional PCR due to the additional reverse transcription step, and it is more expensive due to the need for specific reagents and real-time PCR equipment. Additionally, it requires skilled personnel to perform and interpret the results accurately.

1.3.1.2 Real-Time PCR (qPCR)

Quantitative PCR (qPCR), also known as real-time PCR, is a variation of PCR that allows for the quantification of DNA or RNA in real time as the amplification process progresses. Unlike traditional PCR, which only provides end-point analysis, qPCR enables continuous monitoring of the amplification, providing quantitative data on the amount of nucleic acid present in the sample. qPCR can be applied to both DNA (when directly amplified) and RNA (when combined with reverse transcription, referred to as RT-qPCR).

qPCR involves the use of fluorescent dyes or probes that emit fluorescence in response to the accumulation of PCR products during each cycle of amplification (see Figure 1.2a). The intensity of the fluorescence is measured at each cycle, and the data is used to generate a quantification curve(see Figure 1.2b). The cycle threshold (Ct) value, which is the point at which the fluorescence exceeds a certain threshold, is inversely proportional to the initial quantity of the target nucleic acid in the sample. This allows for the quantification of the initial amount of DNA or RNA with high precision.

qPCR is used for a variety of applications, including the quantification of viral load, gene expression analysis, and detection and quantification of pathogens. It is



(a) qPCR process showing the key steps: (1) Initiation, (2) Denaturation, (3) Primer annealing, and (4) Extension.

(b) qPCR amplification curve illustrating the phases: (1) Initiation phase, (2) Exponential phase, and (3) Plateau phase.*Figure created with BioRender.com.*

Figure 1.2: Overview of the qPCR process and its amplification dynamics. The left panel highlights the four key steps in qPCR, while the right panel shows the corresponding amplification curve and its distinct phases. *Figure created with BioRender.com.*

essential in monitoring the progression of viral infections like HIV, Hepatitis B, and SARS-CoV-2. It is also used to measure the expression levels of specific genes in research and clinical diagnostics, and it is applied in fields like microbiology, virology, and food safety to detect and quantify bacteria and viruses.

The main advantages of qPCR include its quantitative capabilities, real-time monitoring, and high sensitivity and specificity. It provides precise and quantitative data on nucleic acid levels, allowing for immediate analysis and results during the amplification process. This method is capable of detecting low levels of nucleic acids, which is essential for early and accurate diagnosis. However, qPCR also has its limitations. It requires sophisticated equipment and reagents, making it more expensive than traditional PCR. Additionally, it requires skilled personnel for accurate execution and interpretation.

1.3.2 Loop-Mediated Isothermal Amplification (LAMP)

Loop-Mediated Isothermal Amplification (LAMP) is a rapid and efficient molecular diagnostic technique for amplifying DNA at a constant temperature, typically between 60°C and 65°Cv[28]. Unlike traditional PCR, which requires thermal cycling, LAMP operates isothermally, simplifying the equipment needed and speeding up the process.

Using four to six primers to target six to eight regions of the DNA, and a strand-displacing DNA polymerase, LAMP creates loop structures that facilitate continuous amplification, producing significant DNA amounts in less than an hour. This method is highly specific due to the use of multiple primers, reducing non-specific amplification, and highly sensitive, capable of detecting low DNA levels, which is crucial for early virus detection [29].

LAMP has been used to diagnose infections like malaria [30, 31], tuberculosis [32], and viruses such as Zika [33] and mumps [34], demonstrating sensitivity comparable to RT-nested PCR. Its simplicity and cost-effectiveness, requiring only basic heating devices, make it particularly valuable in point-of-care and resource-limited settings.

However, the complex design of multiple primers and the potential for false positives due to primer-dimer formation are challenges. Additionally, result interpretation, often visual, can be subjective. More objective and quantitative readout methods, such as lateral flow dipsticks or real-time fluorescence detection, are being developed to address these issues.

1.3.3 Next-Generation Sequencing (NGS)

Next-Generation Sequencing (NGS) is a revolutionary technology that has transformed the field of genomics and molecular diagnostics [35]. Unlike traditional sequencing methods, NGS allows for the high-throughput sequencing of entire genomes [36], exomes [37], transcriptomes [38], or targeted genomic regions [39] with unprecedented speed and accuracy. This capability makes NGS an invaluable tool for comprehensive genetic analysis [40], pathogen identification [41, 42, 43], and the study of complex biological systems.

NGS involves several key steps, starting with the preparation of a DNA or RNA sample. The nucleic acids are first fragmented into smaller pieces, which are then ligated to adaptors necessary for subsequent amplification and sequencing. These fragments are then amplified using a method such as emulsion PCR or bridge PCR, generating clusters of identical sequences. Sequencing is performed by adding nucleotides that emit a fluorescent signal when incorporated into the growing DNA strand. These signals are captured and converted into sequence data through sophisticated bioinformatics algorithms [44].

One of the primary advantages of NGS is its ability to produce massive amounts of sequence data in a single run. This high-throughput capability enables the simultaneous sequencing of multiple samples, making it cost-effective and efficient for large-scale studies [45]. NGS can generate billions of base pairs of sequence data, allowing for comprehensive genomic coverage and the detection of rare variants that may be missed by other methods [44].

In the context of viral diagnostics, NGS offers several distinct advantages. It can detect and characterize a wide range of pathogens, including viruses, bacteria, fungi, and parasites, without prior knowledge of the infectious agent [46]. This makes NGS particularly useful for identifying novel or emerging pathogens and for studying mixed infections where multiple pathogens are present [47, 48]. NGS also provides detailed information on viral genomes, enabling the tracking of viral evolution, the identification of drug resistance mutations, and the study of viral transmission dynamics [49].

NGS has been instrumental in the response to viral outbreaks, such as the COVID-19 pandemic. It has been used to sequence the genome of SARS-CoV-2 [50], the virus responsible for COVID-19, providing critical information on its genetic structure, mutation patterns, and transmission routes. This information has been essential for the development of diagnostic tests, vaccines, and therapeutic strategies. NGS has also been used to monitor the emergence of new variants of SARS-CoV-2 [51, 52], guiding public health interventions and informing vaccine updates.

Despite its many advantages, NGS has some limitations. The technology requires sophisticated laboratory equipment and bioinformatics expertise, making it less accessible in resource-limited settings. The large volumes of data generated by NGS also require substantial computational resources for data analysis and storage. Additionally, the interpretation of NGS data can be complex, particularly when dealing with large and diverse datasets. Other challenges include the lack of automation in some workflows, the need for highly skilled personnel to operate and interpret the systems, and the relatively high cost associated with implementation and routine use, which may limit scalability and broader adoption.

1.3.4 Antibody Tests

Antibody tests, or serological tests, detect specific antibodies produced in response to a viral infection, such as Immunoglobulin M (IgM) and Immunoglobulin G (IgG), indicating current or past infections. Unlike nucleic acid-based tests that identify viral genetic material, antibody tests focus on the immune response [53].

The process begins with collecting a blood sample, which is then analyzed using an assay like the Enzyme-Linked Immunosorbent Assay (ELISA). In ELISA, viral antigens are attached to a solid surface, and the patient's serum is added. If antibodies are present, they bind to the antigens. A secondary antibody linked to an enzyme is introduced, and a substrate for the enzyme causes a detectable signal, usually a color change, indicating the presence of antibodies.

Antibody tests are crucial for determining whether an individual has been exposed to a virus and developed an immune response. They are useful in epidemiological studies to understand virus spread and can identify individuals who may have developed immunity. For example, during the COVID-19 pandemic, antibody tests helped estimate the prevalence of SARS-CoV-2 [54, 55] and identify potentially immune individuals.

A key advantage of antibody tests is their ability to detect past infections, unlike tests that only identify active infections. This is important for detecting asymptomatic or mild cases. However, they are not suitable for diagnosing acute infections as it takes one to two weeks post-infection for detectable antibody levels to develop. Additionally, the presence of antibodies does not necessarily confer immunity, and the duration of immunity can vary [56]. False positives can occur due to cross-reactivity with antibodies from other pathogens [57].

Despite these limitations, antibody tests are vital in public health for evaluating vaccine efficacy by measuring antibody response and assisting in diagnosing conditions related to the immune response, such as determining the presence of protective antibodies in managing Hepatitis B [58, 59].

1.3.5 Antigen Tests

Antigen tests, or rapid diagnostic tests (RDTs), detect specific viral proteins, such as nucleocapsid or spike proteins, to identify current infections, often before antibodies are present, making them useful for early diagnosis [60, 61]. Designed for quick results within 15 to 30 minutes, they are ideal for point-of-care testing and large-scale screening [62, 63].

For respiratory viruses, the testing involves swabbing the nasopharynx or throat to collect a sample that is then mixed in a solution that releases viral antigens. This solution is applied to a test strip with antibodies that bind to the antigens if present. A visible line indicates a positive result, while no line indicates a negative result.

Antigen tests are particularly useful in settings that require rapid results, such as airports, schools, workplaces, and large public events. Their speed and simplicity make them suitable for mass testing and situations where laboratory facilities are unavailable.

A major advantage of antigen tests is their speed and convenience. They don't require complex lab equipment or highly trained personnel, making them accessible and scalable, which is crucial during pandemics for quickly identifying and isolating infected individuals to reduce transmission.

However, antigen tests are generally less sensitive than PCR tests. They may not detect low levels of viral antigens, especially early in infection or in asymptomatic individuals, leading to false negatives [64, 65]. To address this, negative antigen test results are sometimes confirmed with PCR tests if clinical symptoms suggest infection.

1.3.6 Enzyme-Linked Immunosorbent Assay (ELISA)

The Enzyme-Linked Immunosorbent Assay (ELISA) is a highly sensitive and specific laboratory technique used to detect and quantify soluble substances, such as proteins, peptides, antibodies, and hormones. ELISA plays a crucial role in the diagnosis of viral infections.

In ELISA, antigens are immobilized on a solid surface, typically the wells of a microplate. A sample containing antibodies is added to these wells. If antibodies are present, they bind to the immobilised antigens. The wells are washed to remove unbound substances, and then a secondary antibody linked to an enzyme is added. This enzyme-conjugated antibody binds to the antigen-antibody complexes. After another wash, a substrate for the enzyme is introduced. The enzyme catalyzes a reaction with the substrate, producing a detectable signal, usually a color change, which is measured using a spectrophotometer. The intensity of the color is directly proportional to the amount of antibody or antigen present in the sample.

There are several variations of ELISA, including direct, indirect, sandwich, and competitive ELISA, each serving different diagnostic purposes (see Figure 1.3). ELISA is widely used to detect antibodies against viruses such as HIV [66], Hepatitis B and C [67, 68, 69], Dengue [70], and SARS-CoV-2. During the COVID-19 pandemic, ELISA was essential for identifying individuals exposed to the virus and assessing population immunity.

The major advantages of ELISA include its high sensitivity and specificity, making it effective for early diagnosis. It is versatile, capable of analyzing various sample types like blood, urine, and saliva, and can be automated for high-throughput screening. However, ELISA is time-consuming and requires careful handling to avoid errors. The need for specific reagents can increase costs, and cross-reactivity can sometimes lead to false positives, necessitating confirmatory tests.

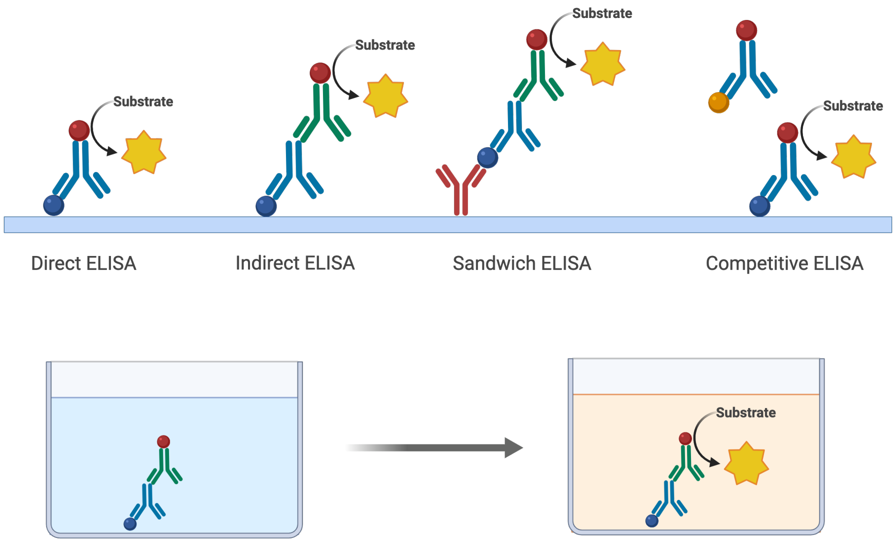


Figure 1.3: key variations and mechanisms underlying enzyme-linked immunosorbent assays (ELISA). It includes representations of direct ELISA, indirect ELISA, sandwich ELISA, and competitive ELISA, with an emphasis on the colorimetric changes observed upon substrate binding. Adapted from Thermofisher [71] and Bio-Rad [72].*Figure created with BioRender.com.*

Despite these limitations, ELISA remains a cornerstone in diagnostic virology and immunology. It provides reliable, quantitative results critical for diagnosing infections, monitoring immune responses, and conducting epidemiological studies. For example, in vaccine development, ELISA evaluates vaccine efficacy by measuring the antibody levels generated in response to the vaccine.

1.3.7 Immunofluorescence Assays (IFA)

Immunofluorescence Assays (IFA) are laboratory techniques used to detect specific antigens or antibodies in a sample using fluorescently labeled antibodies. This method is widely employed in the diagnosis of viral infections due to its ability to provide visual confirmation of the presence of target molecules.

The IFA process involves applying a sample to a microscope slide and adding a fluorescently labeled antibody that binds specifically to the target antigen or antibody. After incubation, unbound antibodies are washed away. The slide is then examined under a fluorescence microscope. If the target antigen or antibody is present, the bound fluorescent antibodies will emit light, which can be visualized and photographed.

IFA is particularly useful for detecting viral proteins in infected cells or tissues. It is used in the diagnosis of infections such as influenza, respiratory syncytial virus (RSV), and cytomegalovirus (CMV). IFA can also be employed to detect antibodies against viruses, aiding in the diagnosis of diseases like Epstein-Barr virus (EBV) and rubella.

One of the main advantages of IFA is its high specificity, as the use of fluorescently labeled antibodies ensures that only the target antigen or antibody is detected. Additionally, IFA provides a visual confirmation of the presence and localization of the target molecules within cells or tissues. This can offer valuable insights into the pathogenesis of infections.

However, IFA has some limitations. The technique requires a fluorescence microscope and trained personnel to interpret the results, which can be subjective. The process can be time-consuming and less suitable for high-throughput screening compared to other methods like ELISA. Furthermore, the quality of the results depends on the quality of the fluorescent antibodies and the sample preparation.

1.3.8 Discussion

Established diagnostic technologies such as PCR, RT-PCR, qPCR, LAMP, NGS, antibody tests, antigen tests, ELISA, and immunofluorescence assays each have unique strengths, but they also share several common traits. These methods leverage the amplification of nucleic acids or detection of immune responses, enabling sensitive and specific detection of viral infections. Most of these technologies require a certain degree of specialized equipment and trained personnel, contributing to their reliability but limiting their use in decentralized or low-resource settings.

1.3.8.1 Common Strengths Across Technologies

One of the key strengths shared by PCR, RT-PCR, qPCR, and LAMP is their high sensitivity and specificity in detecting viral genetic material, even at low concentra-

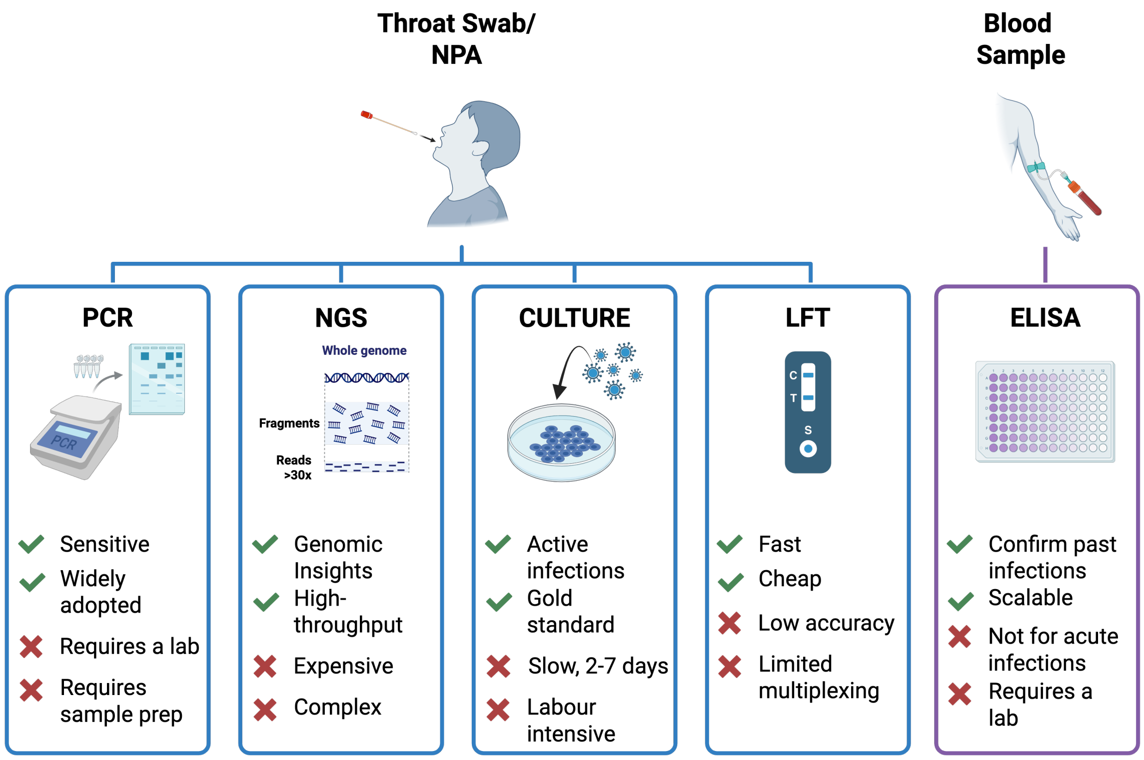


Figure 1.4: Overview of established diagnostic methods for viral infections following sample collection. This schematic illustrates typical diagnostic workflows beginning with a throat swab/nasopharangial aspirate (NPA) or blood sample. A throat swab can be processed using a range of viral detection methods, including RT-PCR, next-generation sequencing (NGS), viral culture, and lateral flow tests (LFTs). Each method is represented with associated advantages and limitations, highlighting trade-offs in sensitivity, speed and infrastructure requirements. In parallel, a blood sample may be analyzed by ELISA, commonly used to detect host antibody responses to viral infections.*Figure created with BioRender.com.*

tions. This makes them particularly valuable for early diagnosis when viral loads may be minimal. Technologies like NGS further extend this precision by providing comprehensive genomic data, which is crucial for tracking viral mutations and studying emerging variants. Antibody tests and ELISA offer another layer of diagnostic capability by detecting immune responses, making them essential for assessing past infections and population-level immunity. Antigen tests and LAMP distinguish themselves with their speed and simplicity, which makes them more adaptable for rapid testing needs in point-of-care settings. Antigen tests, for instance, are particularly valuable for large-scale screening due to their quick turnaround time and ease of use. LAMP, with its isothermal amplification process, reduces the need for complex thermal cycling equipment, offering a more accessible alternative to PCR in resource-limited environments.

1.3.8.2 Areas for Improvement

Despite their strengths, established methods have limitations that present opportunities for improvement and the need for new diagnostic technologies. For instance, PCR-based methods, including RT-PCR and qPCR, require sophisticated instruments and laboratory settings, making them less accessible in low-resource or remote areas. The high sensitivity of these methods, while beneficial, can also make them prone to contamination, leading to false positives. Similarly, the complexity of sample preparation and the need for skilled personnel limit their scalability in large testing campaigns.

Antibody tests and ELISA are powerful tools for understanding immune responses but are less suitable for detecting acute infections, as antibodies take time to develop. Their results can also be influenced by cross-reactivity, potentially leading to false positives. Antigen tests, while rapid and convenient, are generally less sensitive than PCR-based methods and may fail to detect low viral loads, especially in asymptomatic individuals. The accuracy of such tests often requires confirmation through more sensitive methods like RT-PCR.

There is considerable scope for new diagnostic technologies that can address these limitations. For example, advances in microfluidics and lab-on-a-chip technologies could make nucleic acid amplification methods more portable and faster, allowing for PCR-level sensitivity without the need for extensive laboratory infrastructure. Similarly, new point-of-care devices that integrate the high sensitivity of molecular tests with the simplicity and speed of antigen assays could transform diagnostics in both developed and resource-limited settings.

Additionally, technologies that can directly detect intact viral particles or provide insights into viral infectivity, such as next-generation imaging methods and nanoparticle-based assays, could complement existing methods. These would be particularly valuable in distinguishing active infections from residual genetic material, which PCR tests may detect long after a person is no longer infectious.

Finally, improvements in machine learning and artificial intelligence could enhance data interpretation for complex methods like NGS, making them more user-friendly and accessible to non-specialists. AI-driven analysis could also improve the precision of pattern recognition in tests like immunofluorescence assays, offering deeper insights into viral behavior and immune responses.

1.4 Up and Coming Technologies

The domain of viral diagnostics is persistently progressing, yet there remains a pressing need for new technologies. Traditional diagnostic methods often present a dichotomy between accuracy and speed: high-precision techniques typically require lab facilities and extended processing times, while point-of-care methods may compromise accuracy for the sake of rapid results. The COVID-19 pandemic underscored these challenges, with supply chain disruptions exacerbating issues in diagnostic capacities [73].

Despite the emergence of innovative technologies and methodologies, such as CRISPR-based tools and Next-Generation Sequencing, there is still a significant demand for further advancements [74]. These new techniques offer unique advantages, including increased sensitivity, specificity, and the ability to detect multiple pathogens simultaneously. However, challenges such as the need for sophisticated equipment, bioinformatics expertise, and the ability to rapidly scale up testing in response to outbreaks remain [75].

Therefore, the development of novel, efficient, and scalable diagnostic technologies remains a critical area of research. As we continue to confront new viral threats, the evolution of diagnostic technologies will be crucial in enhancing our capacity to respond effectively.

1.5 Emerging Diagnostic Technologies

1.5.1 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based Diagnostics

CRISPR-based tools have revolutionized molecular diagnostics, providing precise and versatile methods for detecting viral infections [76]. CRISPR, or Clustered Regularly Interspaced Short Palindromic Repeats, utilizes CRISPR-associated (Cas) proteins [77] to accurately target and cleave specific viral DNA or RNA sequences [78]. The core component is the Cas protein, typically Cas9 or Cas12, which is guided by a custom-designed RNA sequence (gRNA) that matches the viral target. When the Cas protein binds to the target sequence, it cleaves it. This cleavage activates a reporter molecule to generate a detectable signal, such as fluorescence or a color change, indicating the presence of the virus.

1.5.1.1 Specific High Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK)

Specific High Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK) [79] is a CRISPR-based diagnostic method designed for high sensitivity and specificity in detecting viral infections. Utilizing the CRISPR-Cas13 enzyme, SHERLOCK targets viral RNA sequences with a gRNA [80]. In virus diagnostics, SHERLOCK functions by programming Cas13 with a gRNA that matches the target viral RNA. When Cas13 binds to the target, it cleaves nearby reporter molecules, producing a detectable signal, such as fluorescence or a color change, indicating the presence of the viral RNA. The key advantage of SHERLOCK is its ability to detect low concentrations of viral RNA, which makes it ideal for early-stage infection detection [81]. Its high specificity ensures accurate targeting of viral genomes [82]. The method typically takes around 1 to 2 hours from sample preparation (lysis, RNA extraction, Reverse Transcription) to result.

1.5.1.2 DNA Endonuclease-Targeted CRISPR Trans Reporter (DETECTR)

DETECTR is an advanced CRISPR-based diagnostic platform utilizing the CRISPR-Cas12 enzyme for precise and rapid viral DNA detection[83]. It employs a guide RNA (gRNA) designed to match specific viral DNA sequences. Upon binding, the Cas12 enzyme activates and cleaves nearby reporter molecules, producing a detectable signal such as fluorescence or color change. This system offers results within about 30 to 60 minutes from sample preparation to result, making it ideal for point-of-care applications. DETECTR is highly sensitive and specific, reducing false positives and ensuring accurate diagnostics. Its versatility allows adaptation to various DNA viruses by altering the gRNA sequence. Clinically, DETECTR has been effective in detecting viruses such as Human Papillomavirus (HPV) [84] and SARS-CoV-2 [85]. The method's rapid turnaround and robust performance make it a valuable tool in both clinical and field settings, enabling timely and accurate viral detection during outbreaks. In summary, DETECTR represents a significant advancement in CRISPR-based diagnostics, offering a powerful, efficient, and adaptable approach to viral detection.

1.5.2 Nanopore Sequencing

Nanopore sequencing is an advanced technique that has significantly impacted viral diagnostics[86] by enabling real-time analysis of long DNA or RNA molecules. This method utilizes nanopores—tiny holes through which nucleic acids pass—allowing for the direct sequencing of genetic material. In nanopore sequencing, DNA or RNA molecules are threaded through a nanopore embedded in a membrane[87]. As each nucleotide passes through the pore, it causes characteristic disruptions in an ionic current, which are recorded and analyzed to determine the sequence. This real-time sequencing capability facilitates the rapid identification and characterization of viral genomes. Nanopore sequencing offers several advantages in viral diagnostics. Its speed provides near-instantaneous sequencing results, crucial for timely outbreak responses. The portability of devices like the Oxford Nanopore MinION [88] allows

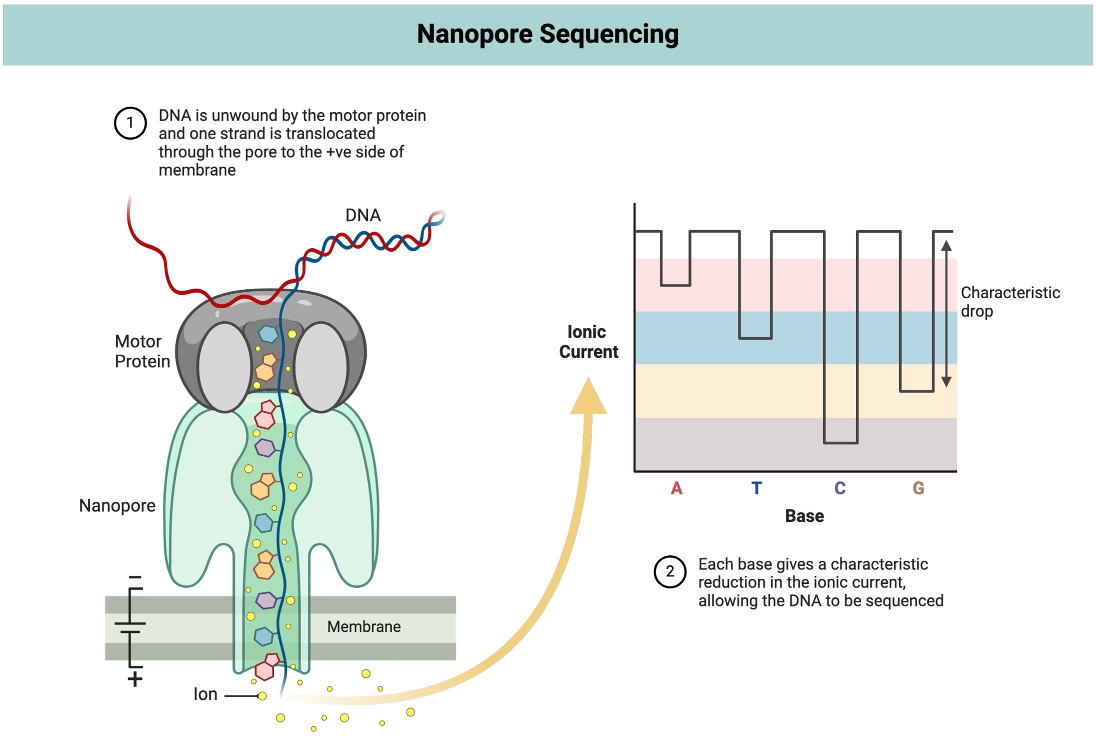


Figure 1.5: Schematic representation of nanopore sequencing. (1) DNA is unwound by a motor protein, and a single strand is translocated through the nanopore toward the positively charged side of the membrane. (2) As each nucleotide passes through the pore, it causes a characteristic reduction in the ionic current, enabling the DNA sequence to be determined in real time.*Original creators of figure Flo Glencross and Daid Ahmad Khan. Figure created with BioRender.com.*

for on-site viral surveillance and diagnostics in field settings. The ability to read long nucleotide sequences helps in assembling complex viral genomes and identifying genetic variations, while its versatility enables the sequencing of both DNA and RNA, making it adaptable to a wide range of viruses.

This technology has been effectively used to sequence viral genomes during outbreaks, including Ebola [89, 90], Zika [91], and SARS-CoV-2 [92, 93], providing critical information on viral evolution, transmission, and mutation patterns. Nanopore sequencing's rapid and comprehensive capabilities make it a valuable tool in the fight against viral infections, aiding in both diagnostics and epidemiological studies.

1.5.3 Surface-Enhanced Raman Spectroscopy (SERS)

Surface-Enhanced Raman Spectroscopy (SERS) is a highly sensitive analytical technique that amplifies Raman scattering signals of molecules on nanostructured metallic surfaces, such as gold or silver nanoparticles. Originally, Raman spectroscopy was discovered in 1928 by C.V. Raman [94], and SERS was first observed in the 1970s [95], revolutionizing the sensitivity of molecular detection. This enhancement allows for the detection of viruses and their biomarkers at very low concentrations by identifying unique molecular fingerprints of viral components like proteins, nucleic acids, and lipids. In virus diagnostics, SERS involves binding viral particles to a SERS-active substrate, which significantly boosts the Raman signal, enabling precise identification and quantification.

SERS offers several advantages for virus diagnostics, including rapid detection, high sensitivity that often exceeds traditional methods like PCR, and the capability for multiplexed assays to identify multiple viruses or biomarkers simultaneously. Portable SERS devices facilitate point-of-care testing, enhancing accessibility and timely diagnosis essential for outbreak control and treatment initiation. However, challenges such as ensuring consistent and reproducible SERS-active substrates, handling interference from complex biological samples, and the need for standardized protocols must be addressed. Additionally, the cost and accessibility of high-quality SERS equipment can limit its use in resource-constrained settings.

Recent advancements have improved SERS applications in virus diagnostics through the development of novel nanostructures, integration with microfluidic platforms for automated detection, and the use of machine learning algorithms to analyze complex spectra, thereby increasing diagnostic accuracy. Ongoing clinical validations are crucial for establishing SERS-based diagnostics alongside established methods, facilitating regulatory approvals and broader clinical adoption.

1.5.4 Biosensors

Biosensors are emerging tools in viral diagnostics, offering diverse mechanisms for the rapid and specific detection of viral pathogens. Electrochemical biosensors operate by measuring electrical changes that occur when a virus interacts with the sensor's recognition elements. For example, electrochemical sensors coated with anti-HIV antibodies can detect HIV particles by monitoring alterations in current or impedance upon virus binding [96, 97], while paper-based electrochemical sensor chip can detect SARS-CoV-2 antigens through electrical signals [98, 99]. Optical biosensors utilize light-based techniques to detect viruses by observing changes in properties such as absorbance, fluorescence, or refractive index. An example includes Surface Plasmon Resonance (SPR) sensors used for influenza detection, which monitor refractive index changes when viruses bind to antibodies on the sensor surface. Piezoelectric biosensors detect viruses by measuring mass or mechanical changes on the sensor surface using piezoelectric materials. Quartz Crystal Microbalance (QCM) sensors, for instance, identify Hepatitis B Virus (HBV) by detecting frequency shifts caused by viral binding to immobilized antibodies [100]. Additionally, thermal biosensors measure temperature changes resulting from biochemical reactions during virus-antibody interactions, such as microcalorimetric sensors used for influenza detection. Magnetic biosensors employ magnetic nanoparticles to capture and detect viral components through changes in magnetic properties, with Giant Magnetoresistance (GMR) sensors being used to detect HBV by measuring resistance changes from bound magnetic nanoparticles [101]. While Surface-Enhanced Raman Spectroscopy (SERS) is a notable optical biosensor, it is one of many optical methods and is not the primary focus in the broader context of biosensor applications. Overall, these various types of biosensors provide versatile approaches for virus detection, each with unique advantages that contribute to enhancing diagnostic capabilities in clinical and point-of-care settings.

1.5.5 Digital PCR (dPCR)

Digital PCR (dPCR) is a highly precise molecular technique for quantifying nucleic acids, offering unparalleled sensitivity and absolute quantification without the need for calibration curves [102]. By partitioning the sample into thousands of tiny reactions and using fluorescence detection to measure the presence of target sequences, dPCR is uniquely suited for detecting rare targets, low viral loads, and subtle genetic variations.

There are two main formats: chip-based dPCR, such as the Thermo Fisher QuantStudio Absolute Q, where partitioning is achieved through microfluidic channels or nanowells; and droplet digital PCR (ddPCR), used in systems like the Bio-Rad QX200, where samples are emulsified into thousands of nanoliter-sized droplets suspended in oil. In both cases, amplification and fluorescence detection occur in each individual partition, but the droplet-based method is generally more scalable and better suited to high-throughput workflows, while chip-based systems often offer more integrated and streamlined protocols.

Other notable platforms include the Stilla Naica and RainDance RainDrop, which use their own proprietary partitioning technologies with advanced multiplexing capabilities. Widely used for pathogen detection [103, 104], mutation analysis [105], viral load monitoring[106], and vaccine development [107], dPCR stands out for its accuracy and reproducibility in challenging diagnostic scenarios. Digital PCR is not yet widely adopted due to its high cost, complex workflows, and limited throughput compared to qPCR. Additionally, a lack of standardization and the need for specialized training further hinder its widespread use, though its precision and sensitivity make it valuable for niche applications.

1.5.6 Lab-on-a-Chip (LoC) Devices

Lab-on-a-Chip (LoC) devices integrate laboratory processes, such as sample preparation, nucleic acid amplification, and detection, onto compact microfluidic platforms, enabling rapid and precise virus diagnostics [108]. Notable examples, like the Cepheid

GeneXpert and Abbott ID NOW, automate complex workflows in self-contained cartridges, reducing sample and reagent use while offering portability for point-of-care testing. However, their adoption is limited by high development costs, scalability challenges, and the need for specialized expertise.

1.5.7 Molecular Imprinted Polymers (MIPs)

Molecular Imprinted Polymers (MIPs) are synthetic materials designed to recognize specific viral particles or biomarkers through tailored binding sites, offering advantages like stability, reusability, and cost-effectiveness compared to biological elements [109, 110]. While they hold promise for virus diagnostics and are being actively researched for applications such as SARS-CoV-2 [111] and influenza detection [112], MIPs remain largely in the experimental stage. Challenges like scaling, reproducibility, and achieving clinical-level sensitivity must be addressed before they can transition into widely adopted commercial products.

1.5.8 Single-Molecule Real-Time (SMRT) Sequencing

Single-Molecule Real-Time (SMRT) sequencing is a high-throughput DNA and RNA sequencing technology that enables real-time analysis of nucleic acids at the single-molecule level [113]. Developed by Pacific Biosciences (PacBio), SMRT sequencing uses zero-mode waveguides (ZMWs) to observe the activity of DNA polymerase as it incorporates fluorescently labelled nucleotides. This approach provides long-read sequences with high accuracy, making it ideal for detecting complex genetic variations, structural rearrangements, and epigenetic modifications.

The long-read capability of SMRT sequencing offers several advantages over short-read platforms: it enables the resolution of repetitive regions, the phasing of genetic variants, and the accurate assembly of complete genomes without the need for extensive computational stitching. This is particularly important in virology, where viral genomes may contain high levels of sequence complexity, structural diversity, and quasispecies variation.

In virus diagnostics, SMRT sequencing is valuable for characterising viral genomes, identifying mutations, and studying viral evolution [114, 115]. The commercial platform, PacBio Sequel IIe, is widely used for applications such as whole-genome sequencing of viruses like SARS-CoV-2 and HIV. Its long-read capabilities are particularly beneficial for resolving repetitive regions and structural variants in viral genomes. While highly accurate, the cost and complexity of SMRT sequencing limit its use to specialised research and clinical laboratories. Advances in cost reduction and workflow simplification aim to expand its accessibility for broader diagnostic applications.

1.5.9 Field-Effect Transistor (FET) Biosensors

Field-Effect Transistor (FET) biosensors are highly sensitive diagnostic devices that detect viral particles or biomarkers by monitoring changes in the electrical properties of a semiconductor-based platform [116]. A recognition element, such as an antibody or nucleic acid, is immobilized on the FET surface, and binding of the target viral biomolecule alters the local charge distribution. This change modulates the conductivity of the FET channel, producing an electrical signal that allows real-time, label-free detection. FET biosensors have been studied for virus diagnostics, including SARS-CoV-2 [117, 118], where graphene-based FETs have demonstrated the ability to detect spike proteins and viral RNA with high sensitivity.

While most FET biosensors remain in the research stage, advancements are bringing them closer to commercialization, with companies like Nanowear and Grolltex exploring related technologies. Their key advantages include rapid response times, high sensitivity, and potential for miniaturization into portable devices. However, challenges such as complex fabrication, environmental interference, and scalability need to be addressed for broader adoption. Ongoing research aims to enhance their robustness and integrate FET biosensors into cost-effective, point-of-care platforms for virus diagnostics.

1.5.10 Paper-based Analytical Devices (PADs)

Paper-based Analytical Devices (PADs) are cost-effective, portable platforms that use capillary action to transport fluids through porous paper, enabling chemical or biochemical reactions for virus diagnostics [119]. Commercial examples like the BinaxNOW and CareStart COVID-19 Antigen Tests detect viral antigens via lateral flow, producing visible lines as a readout. Studies also focus on multiplexed detection of viruses like dengue and Zika [120, 121]. Their affordability and simplicity make PADs ideal for resource-limited settings, though challenges in sensitivity and reagent stability remain areas for improvement.

1.5.11 Multiplexed Assays using Barcoded Nanoparticles

Multiplexed assays using barcoded nanoparticles are advanced platforms that enable the simultaneous detection of multiple viral targets within a single test [122]. These assays use nanoparticles encoded with unique identifiers, such as optical, magnetic, or chemical "barcodes," to correspond to specific viral antigens or nucleic acids [123, 124]. Upon binding of the target molecule, interactions are detected and decoded using fluorescence, spectroscopy, or similar techniques.

While much of the technology remains in the research phase, commercial platforms like Luminex xMAP Technology and Nanostring nCounter System have demonstrated the potential for multiplexed viral detection, including respiratory viruses like influenza and RSV. Such systems offer high sensitivity, specificity, and scalability for clinical and research applications. However, challenges like production complexity and the need for specialized instrumentation remain barriers to widespread adoption. Continued development aims to create cost-effective, point-of-care-compatible solutions for routine virus diagnostics.

1.5.12 Surface Plasmon Resonance (SPR)

Surface Plasmon Resonance (SPR) is an optical sensing technique that measures changes in the refractive index near a sensor surface to detect biomolecular interac-

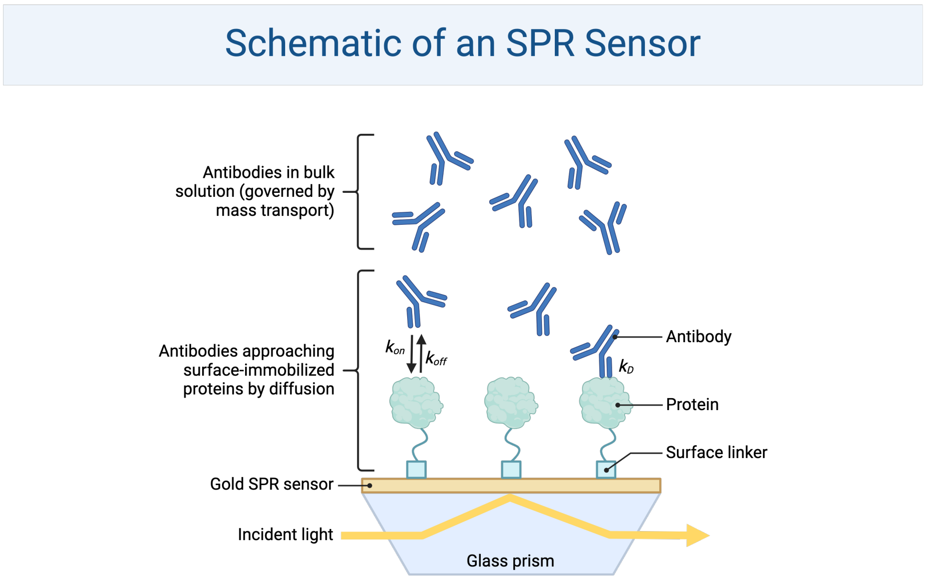


Figure 1.6: Schematic representation of an SPR sensor [125]. Antibodies in bulk solution diffuse toward the surface, where they interact with immobilized proteins on a gold SPR sensor. Incident light passes through a glass prism to excite surface plasmons, enabling the detection of biomolecular interactions in real time. *Figure created with BioRender.com.*

tions in real time. It relies on the excitation of surface plasmons—oscillating electrons at the metal-dielectric interface—by polarized light. When a biomolecule, such as a viral antigen or antibody, binds to the sensor surface, the resonance angle or intensity changes, enabling label-free and quantitative analysis of binding events [126] (see Figure 1.6).

In virus diagnostics, SPR is used to study viral antigens, antibodies, and inhibitors, providing data on binding kinetics, affinity, and specificity [127, 128]. Commercial products include Cytiva's Biacore systems, which are widely used in research and diagnostics, and Nicoya OpenSPR, a benchtop platform designed for affordable and accessible biomolecular interaction studies. While SPR offers high sensitivity and real-time analysis, its cost and need for specialized equipment limit its routine use in diagnostics. Efforts to miniaturize and integrate SPR into portable systems aim to expand its applications in point-of-care viral testing.

1.5.13 Coherent Anti-Stokes Raman Spectroscopy (CARS)

Coherent Anti-Stokes Raman Spectroscopy (CARS) is a powerful nonlinear optical technique used to detect and characterize molecules based on their vibrational modes [129]. In CARS, two laser beams (pump and Stokes) interact with a sample to generate a coherent signal at a higher frequency (anti-Stokes), which is specific to the vibrational energy levels of the molecules. This method provides highly sensitive, label-free chemical imaging with exceptional spatial and temporal resolution.

In the context of virus diagnostics, CARS has been explored for detecting viral particles, analyzing their chemical composition, and identifying infected cells [130]. Its ability to provide real-time, non-invasive analysis makes it a promising tool for studying viral structures and interactions [131]. However, the complexity of the setup, high cost of equipment, and sensitivity to environmental factors limit its current use to research applications rather than routine diagnostics. Ongoing advancements in instrumentation and data analysis are focused on improving accessibility and expanding its clinical potential.

1.5.14 Smartphone-based Methods

1.5.14.1 Cough Analysis for COVID-19 Detection

Cough analysis for COVID-19 detection leverages artificial intelligence (AI) and machine learning (ML) to identify unique acoustic features of cough sounds associated with the disease [132]. Using audio recordings, algorithms analyze parameters such as frequency, amplitude, and duration to differentiate COVID-19-related coughs from those caused by other conditions. These systems often utilize large datasets to train models, improving their accuracy and generalisability.

Several studies and prototypes have demonstrated the potential of cough analysis as a non-invasive, low-cost screening tool, particularly in resource-limited settings. Apps like Coughvid [133] and research from institutions like MIT have shown promising results in identifying COVID-19 with high sensitivity and specificity. However, challenges include ensuring robust performance across diverse populations, addressing privacy concerns, and integrating these tools into clinical workflows. Ongoing development aims to validate and standardize cough analysis for broader adoption in public health and clinical settings.

1.5.14.2 Smartphone-based Imaging for Viral Detection

Smartphone-based imaging for viral detection leverages the advanced cameras, sensors, and computational capabilities of smartphones to identify viral particles, biomarkers, or diagnostic signals[134]. These platforms typically integrate smartphone cameras with external components such as microfluidic chips, fluorescence readers, or lateral flow assays to capture and analyze diagnostic results [135]. Image processing algorithms then interpret the data to detect the presence of viral targets with high sensitivity and specificity.

Applications include detecting viral antigens, nucleic acids, or antibodies in diseases like COVID-19, influenza, and dengue. Commercially available examples include smartphone adapters for lateral flow assays, such as the Ellume COVID-19 Home Test [136, 137], which pairs with a smartphone app for result interpretation, and research prototypes that use smartphone cameras for fluorescence or colorimetric detection. While offering portability, low cost, and ease of use, challenges include ensuring consistent image quality across devices, achieving clinical-grade sensitivity, and maintaining data privacy. Continued advancements aim to enhance accuracy and expand smartphone-based imaging for point-of-care and at-home diagnostics.

1.5.14.3 Mobile Apps for Symptom Tracking and Diagnosis

Mobile apps for symptom tracking and diagnosis offer digital platforms to monitor symptoms, assess risk, and guide users toward appropriate actions, such as seeking testing or medical care [138]. In viral infections like COVID-19 and influenza, these apps gather user-reported symptoms such as fever, fatigue, and respiratory distress, often using AI or decision-tree algorithms to analyse patterns and provide risk assessments or recommendations.

Table 1.1: Comparison of emerging virus diagnostic technologies grouped by detection type. Techniques are evaluated for their detection targets, label-free capability (LF), multiplexing potential (MPX), cost, portability, readout method, and key limitations. Groupings are color-coded: blue for nucleic acid detection, yellow for antigen/antibody detection, purple for molecular signature detection, and green for symptom-based or indirect detection methods.

Technique	**Category**	**Target**	**LF**	**MPX**	**Cost**	**Portable**	**Readout**	**Limitation**
CRISPR Detection [76, 77]	Nucleic Acid	Viral RNA/DNA	▲	✓	Medium	Medium	Fluorescence	Needs amplification, false positives
Nanopore Seq. [88, 87]	Nucleic Acid	Nucleic acids	✗	✓	High	Low	Electrical signal	Cost, error rate, data processing
Digital PCR [102, 104]	Nucleic Acid	Nucleic acids	✗	✓	High	Low	Quant. fluorescence	Instrument cost, technical skill
SMRT Sequencing [113]	Nucleic Acid	Nucleic acids	✗	✓	High	Low	Optical signal	Expensive, complex
Lab-on-Chip [108]	Nucleic Acid	DNA/RNA, proteins	▲	✓	Medium	High	Mixed	Integration complexity
Biosensors [96, 98, 99]	Nucleic Acid	Antigens, antibodies, nucleic acids	▲	▲	Medium	Medium	Electrical/optical	Environment sensitivity
FET Biosensors [116]	Nucleic Acid	Viral proteins, nucleic acids	✓	▲	Med–High	High	Electrical	Fabrication complexity
PADs [119, 120]	Antigen/Ab	Viral antigens	✗	✓	Low	High	Visual (LFA)	Low sensitivity, reagent stability
Barcoded NP Assays [122, 124, 123]	Antigen/Ab	Antigens, nucleic acids	✗	✓	Med–High	Medium	Optical/fluorescence	Special equipment
SPR [125, 126]	Antigen/Ab	Binding interactions	✓	▲	High	Low	Optical shift	Cost, not POC-friendly
SERS	Molecular Sig.	Molecular vibrations	✓	▲	High	Low	Raman shift	Complex instrumentation
CARS [131]	Molecular Sig.	Molecular vibrations	✓	▲	Very High	Low	Anti-Stokes Raman	Setup, environmental sensitivity
MIP Sensors [109]	Molecular Sig.	Protein/antigen templates	✓	▲	Low	High	Binding detection	Selectivity, reproducibility
Smartphone Methods [137, 136, 134, 135]	Symptom-based	Cough, images, symptoms	✓	▲	Low	High	Audio, visual, input	Privacy, device variation

Examples include the ZOE COVID Symptom Study [139], which aggregates large-scale symptom data for public health insights, and Ada Health, an AI-powered app offering tailored symptom assessments. These tools are scalable and accessible, providing valuable resources for both individual users and public health systems. However, challenges include ensuring data privacy, improving diagnostic accuracy, and addressing biases in user-reported information. Integrating symptom tracking with wearable devices and diagnostic technologies is a key area for future development.

Evidence from large-scale deployments has demonstrated the effectiveness of mobile symptom tracking apps in supporting public health efforts. The ZOE COVID Symptom Study, for example, engaged over four million users in the UK and provided real-time data that accurately predicted regional COVID-19 prevalence and hospital admission trends [140]. Its data also influenced national health policy, such as expanding the list of recognised symptoms. Similarly, the COVID Radar app in the Netherlands showed strong correlation between user-reported symptoms and official case numbers, validating its use for early outbreak detection [141]. In Singapore, the TraceTogether app reduced contact tracing time from four days to under two, significantly improving the speed of quarantine measures [142]. These examples illustrate how symptom tracking apps, when widely adopted and properly integrated with public health systems, can offer timely insights, accelerate response measures, and improve disease surveillance at the population level.

1.6 Entrenched Systems and the Challenges of Innovation in Virus Diagnostics

The dominance of large companies in the virus diagnostics market can be understood through an analogy to transportation. Imagine a groundbreaking new mode of transportation that promises unparalleled speed, efficiency, and convenience. However, this new system comes with substantial challenges: it would require the construction of millions of miles of specialized roads, a network of fueling stations spaced every few miles, and the establishment of entirely new industries for manufacturing, maintenance, insurance, and driver training. The cost for individuals would be exorbitant, with tens of thousands of pounds needed upfront for the vehicle, alongside ongoing expenses for fuel, maintenance, and insurance.

At first glance, such a proposal might seem absurd—an unreasonably complex, expensive, and impractical solution compared to simpler alternatives. Few would endorse building such a system from scratch, let alone maintaining it. However, this system already exists; it is the car. What appears untenable in theory becomes completely rational in practice because the infrastructure for cars is so entrenched. Roads, gas stations, and repair shops are ubiquitous, and the sunk costs invested over decades make the car the default mode of transportation. It is not that the car was the simplest or most logical solution; it is that the system supporting it grew incrementally into an unshakable norm, making alternative modes of transportation difficult to imagine or adopt.

A similar dynamic exists in the virus diagnostics market. Just as the car relies on a vast, entrenched infrastructure, the market for virus diagnostics is dominated by large companies like Abbott, Roche, and Thermo Fisher, which have developed systems deeply integrated into laboratory workflows. Over decades, these companies have established extensive networks of diagnostic machines, reagents, and protocols that laboratories worldwide depend on. This infrastructure is as ubiquitous and relied upon as the roads and gas stations that support cars.

For a new diagnostic technology, entering this market is akin to proposing an entirely new mode of transportation. Laboratories would need to purchase new machines, overhaul existing workflows, and retrain staff. Supply chains for reagents and consumables would need to be rebuilt or replaced. Even if the new technology offers clear advantages, such as greater speed or sensitivity, the barriers to adoption make it a formidable challenge. Laboratories, like transportation users, are naturally reluctant to disrupt systems that are already functional and familiar.

The existing diagnostic infrastructure creates a system of inertia that resists change. The cost of switching to new technologies is high, both in terms of financial investment and time required for adaptation. Logistics are already solved under the current system, with reliable supply chains and robust workflows in place. Furthermore, familiarity breeds trust—laboratories rely on established diagnostic methods because they are proven and understood. This creates a self-reinforcing cycle where innovation is discouraged, not because it lacks merit, but because it disrupts the entrenched norm.

Breaking this cycle requires intentional effort. Just as alternative transportation systems have gradually gained traction through subsidies, pilot programs, and gradual integration into existing infrastructure, the adoption of new diagnostic technologies must be incentivized. Policies that lower the cost of transition, streamline regulatory approvals, and foster collaboration between innovators and established players can help overcome these barriers. The success of new technologies often hinges not only on their technical superiority but also on their ability to integrate into or enhance the existing infrastructure without causing major disruptions.

In conclusion, the dominance of large companies in the virus diagnostics market is supported by a deeply entrenched infrastructure that creates significant barriers to the adoption of new technologies. This dynamic parallels the transportation industry, where cars have become the default mode due to the vast system built around them. Understanding and addressing the factors that perpetuate this inertia is essential for fostering innovation in virus diagnostics, enabling laboratories to adopt more advanced and efficient technologies

1.7 Opportunities for New Technologies in Point-of-Care Diagnostics

While established diagnostic technologies dominate laboratory settings, they are not well-suited for point-of-care (POC) applications. Current solutions for POC diagnostics, such as lateral flow tests (LFTs) and loop-mediated isothermal amplification (LAMP), have certain limitations that constrain their broader utility. These technologies often require trade-offs between speed, sensitivity, and specificity, meaning they can deliver rapid results or high accuracy, but rarely both. Moreover, these technologies are reaching the limits of their potential; much like squeezing the last drops of juice from a lemon, further improvements yield diminishing returns.

This creates a unique opportunity for new technologies to enter the POC diagnostics market, both because established players have not yet cracked this segment and because the existing solutions were not designed with this context in mind. The potential is transformative: imagine knowing the exact infection you have—if any—during your consultation with a doctor, or being able to perform a diagnostic test at your local pharmacy. Even more ambitiously, envision a future where such tests could be performed from the convenience of your own home. These possibilities represent significant advancements for healthcare systems in developed countries, but they also highlight an urgent need in low-resource environments.

In many parts of the world, such as rural areas in developing countries, central laboratories and pharmacies are either scarce or entirely absent. Here, the importance of reliable, portable, and easy-to-use diagnostic tools cannot be overstated. During epidemics or pandemics, when healthcare resources are stretched thin, POC diagnostics could play a pivotal role in identifying and controlling outbreaks. Whether in the field, at makeshift clinics, or in remote areas, these tools could provide critical information to healthcare workers and patients alike, guiding timely interventions and saving lives.

The challenge is to develop technologies that meet these needs without compromising on accuracy, reliability, or accessibility. Current gaps in the market provide a clear entry point for innovators who can design diagnostics tailored specifically for POC settings. These tools must be rapid, portable, affordable, and user-friendly, while maintaining high levels of accuracy and reliability. By addressing these unmet needs, new technologies have the potential not only to disrupt the market but also to redefine how and where diagnostics are performed, bringing high-quality healthcare to even the most resource-constrained environments.

1.8 Frameworks for Point-of-Care Viral Diagnostics: ASSURED and REASSURED

Given the lack of a perfect product—or even technology—for point-of-care (POC) diagnostics, the World Health Organization (WHO) has established clear metrics to guide the development of new diagnostic tools. Known as the ASSURED [143] criteria, this framework emphasizes seven key attributes that ensure diagnostic tests are both practical and accessible across a variety of settings. These attributes include affordability, sensitivity, specificity, user-friendliness, rapidity and robustness, equipment-free operation, and deliverability to end-users (see Table 1.2 for further details). Together, they serve as a blueprint for creating tools that can effectively address the challenges of diverse healthcare environments, particularly in resource-limited settings.

Recognizing the evolving landscape of diagnostic technology and healthcare needs, the REASSURED guidelines were later introduced to expand upon the original ASSURED framework. In addition to retaining the core attributes, the REASSURED criteria add two critical elements: real-time connectivity and ease of specimen collection. Real-time connectivity allows for seamless integration of diagnostic data into digital systems, enabling enhanced disease surveillance and swift public health responses. Prioritizing ease of specimen collection addresses patient comfort while reducing technical barriers, making diagnostics more accessible and feasible for a wider range of users.

These criteria are particularly pertinent for POC viral diagnostics, where rapid and accurate results are paramount. Beyond technical performance, such tools must also be robust enough to function in varied and often challenging environments, from urban clinics to remote field settings. The focus on connectivity within the REASSURED [144] framework aligns with modern trends in digital health, paving the way for data-driven decision-making during outbreaks (see Table 1.2 for further details). By adhering to these principles, the development of new diagnostic technologies can

REASSURED	**ASSURED**	**Metric**	**Description**
✓	✓	Affordable	Cost-effective for the target population and healthcare systems
✓	✓	Sensitive	High true-positive rate to correctly identify infected individuals
✓	✓	Specific	High true-negative rate to correctly exclude non-infected individuals
✓	✓	User-friendly	Simple operation requiring minimal training
✓	✓	Rapid and Robust	Delivers results quickly and functions reliably under variable conditions
✓	✓	Equipment-free	Does not require complex instrumentation, suitable for low-resource settings
✓	✓	Deliverable	Easy to distribute and deploy to the point of need
✓		Real-time connectivity	Enables digital integration for enhanced disease surveillance and data sharing
✓		Ease of specimen collection	Minimally invasive and straightforward sample acquisition for patient comfort

Table 1.2: Consolidated ASSURED [143] and REASSURED [144] metrics for point-of-care diagnostics. Shared metrics are marked in both columns with ticks, while unique metrics are marked only in the respective framework. Grey cells highlight the ticks.

not only fill existing gaps but also enhance global health initiatives, ensuring equitable and effective healthcare delivery worldwide.

1.9 Limitations of Established Diagnostic Technologies in Meeting ASSURED and REASSURED Criteria

Point-of-care (POC) diagnostics aim to provide rapid, accurate, and accessible testing at or near the site of patient care, addressing critical needs in resource-limited settings. The ASSURED framework (Affordable, Sensitive, Specific, User-friendly, Rapid and

Robust, Equipment-free, Deliverable) and its expanded REASSURED version (adding Real-time Connectivity and Ease of Specimen Collection) offer benchmarks for evaluating the suitability of diagnostic technologies for such purposes. However, as the comparison table illustrates, none of the established diagnostic technologies fully satisfy all ASSURED and REASSURED criteria (see Table 1.2), highlighting significant gaps that must be addressed to optimise POC diagnostics.

REASSURED	**ASSURED**	**Metric**	**PCR**	**LAMP**	**NGS**	**Antibody Tests**	**Antigen Tests**
✓	✓	Affordable	×	×	×	✓	✓
✓	✓	Sensitive	✓	✓	✓	×	✓
✓	✓	Specific	✓	✓	✓	✓	×
✓	✓	User-friendly	×	×	×	✓	✓
✓	✓	Rapid and Robust	×	×	×	✓	✓
✓	✓	Equipment-free	×	×	×	✓	✓
✓	✓	Deliverable	×	×	×	✓	✓
✓		Real-time Connectivity	✓	✓	✓	×	×
✓		Ease of Specimen Collection	✓	✓	✓	✓	✓

Table 1.3: Evaluation of common diagnostic technologies against the ASSURED [143] and REASSURED [144] criteria. Each method is assessed for affordability, sensitivity, specificity, usability, and operational suitability in resource-limited settings. REASSURED extends ASSURED by incorporating real-time connectivity and ease of specimen collection.

PCR, including real-time and reverse-transcription variants, is highly sensitive and specific, making it a cornerstone of diagnostic virology. However, it requires expensive equipment, skilled personnel, and a controlled laboratory environment, which contradicts the criteria for affordability, user-friendliness, and equipment-free operation. Similarly, while LAMP addresses some of the operational challenges of PCR by eliminating the need for thermal cycling, it still depends on specialized equipment and reagents. Neither method is fully affordable or deliverable in resource-limited settings.

Next-generation sequencing (NGS) offers unparalleled sensitivity, specificity, and scalability for comprehensive genomic analysis. However, its high cost, complex workflows, and reliance on sophisticated infrastructure make it unsuitable for most POC settings. NGS technologies fail to meet the criteria for affordability, user-friendliness, equipment-free operation, and deliverability. While they offer real-time connectivity, their high technical barrier prevents broader adoption in non-specialized environments.

Antibody and antigen tests are more aligned with the ASSURED criteria. They are affordable, user-friendly, rapid, equipment-free, and deliverable, making them well-suited for POC applications. However, they are less sensitive and specific compared to molecular methods, particularly in detecting early-stage or low-burden infections. Additionally, they lack real-time connectivity, which limits their integration into centralized healthcare systems for surveillance and epidemiological monitoring.

The inability of any single technology to satisfy all ASSURED and REASSURED criteria underscores the need for continued innovation in POC diagnostics. Existing molecular methods like PCR and NGS prioritize sensitivity and specificity at the cost of accessibility, while immunoassays prioritize accessibility at the cost of diagnostic accuracy. Bridging these gaps requires leveraging emerging technologies, such as CRISPR-based diagnostics and lab-on-a-chip devices, to develop solutions that balance accuracy, affordability, and operational simplicity.

No established diagnostic technology currently fulfills all the requirements of the ASSURED and REASSURED frameworks, highlighting the trade-offs between accuracy, cost, and accessibility. Future POC diagnostic tools must integrate advancements in real-time connectivity, ease of specimen collection, and affordability without compromising sensitivity and specificity. This balanced approach will be crucial for addressing diagnostic challenges in resource-limited and decentralized settings, enabling equitable access to high-quality healthcare.

1.10 Challenges and Limitations of Emerging Diagnostic Technologies in Meeting Point-of-Care Needs

Emerging diagnostic technologies hold significant promise for transforming point-of-care (POC) diagnostics. However, most of these technologies are either not commercially available or lack sufficient clinical validation. While they demonstrate innovative approaches to addressing diagnostic challenges, their real-world applicability remains limited. Two of the most critical factors for effective POC diagnostics—portability and speed—are particularly challenging for most emerging technologies to achieve consistently (see Table 1.4). Many of these tools require specialized equipment and can only meet the under-15-minute waiting time limit in highly controlled or limited scenarios.

REASSURED	**ASSURED**	**Metric**	**CRISPR**	**Nanopore**	**SERS**	**Biosensors**	**dPCR**	**LoC**	**MIPs**	**SMRT**	**FET**	**PADs**	**Multiplexed Assays**	**SPR**	**CARS**	**Smartphone**
✓	✓	Affordable	×	×	×	×	×	×	×	×	×	✓	×	×	×	✓
✓	✓	Sensitive	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓	✓	×
✓	✓	Specific	✓	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓	✓	×
✓	✓	User-friendly	×	×	×	✓	×	×	×	×	×	✓	×	×	×	✓
✓	✓	Rapid and Robust	×	×	×	×	×	×	×	×	×	✓	×	×	×	✓
✓	✓	Equipment-free	×	×	×	×	×	×	×	×	×	✓	×	×	×	✓
✓	✓	Deliverable	×	×	×	×	×	×	×	×	×	✓	×	×	×	✓
✓		Real-time Connectivity	✓	✓	×	×	×	×	×	×	×	×	×	×	×	✓
✓		Ease of Specimen Collection	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓

Table 1.4: Comparison of emerging diagnostic technologies based on ASSURED and REASSURED metrics.

Smartphone-based methods exemplify this tension. While they leverage the portability of smartphones, their primary applications include interpreting results from pre-existing tests like lateral flow assays or performing cough analysis for respiratory diseases. These applications are not standalone diagnostic tools but rather enhancements to existing workflows. Additionally, many smartphone applications focus on symptom tracking, aggregating user-reported data to monitor disease progression or assess risk. While valuable for public health and individual management, this does not constitute diagnostic testing and fails to address the need for rapid, comprehensive diagnostics at the POC.

CRISPR-based diagnostics demonstrate exceptional sensitivity and specificity, making them promising candidates for early and accurate disease detection. However, these tools typically require specialized reagents and equipment, such as Cas proteins and advanced detection systems, which compromise their portability. While CRISPR platforms like SHERLOCK and DETECTR can produce results in under 15 minutes in limited scenarios, this speed often depends on ideal conditions that are not representative of real-world environments.

Nanopore sequencing, known for its portability and ability to perform real-time genomic analysis, offers a potential solution for decentralized diagnostics. However, the technology's reliance on trained personnel, sophisticated infrastructure, and complex workflows limits its applicability outside of specialized settings. Achieving rapid results within the 15-minute threshold is particularly challenging for nanopore sequencing, given the time required for sample preparation and analysis.

Technologies like surface-enhanced Raman spectroscopy (SERS) and field-effect transistor (FET) biosensors face similar challenges. While SERS excels in sensitivity and specificity, its dependence on nanostructured substrates and advanced instrumentation makes it difficult to scale for portable and rapid use. FET biosensors, though conceptually portable, often require precise fabrication and environmental controls to function effectively, limiting their robustness and speed in POC applications.

Digital PCR (dPCR), lab-on-a-chip (LoC) devices, and molecularly imprinted polymers (MIPs) showcase significant advancements in precision and sensitivity. However, these technologies often involve intricate workflows and specialized instruments that detract from their portability. Rapid results are typically achieved only in specific scenarios, such as when testing small, pre-prepared samples in ideal laboratory conditions.

Single-molecule real-time (SMRT) sequencing and surface plasmon resonance (SPR) further illustrate these constraints. Both are highly accurate and capable of detailed molecular analyses but are inherently tied to laboratory infrastructure and long processing times. Their potential for rapid diagnostics in decentralized settings is limited by these operational demands.

Paper-based analytical devices (PADs) and smartphone-based methods stand out for their portability and user-friendliness, aligning with several ASSURED criteria. PADs, for example, are highly portable and capable of providing results in under 15 minutes in many scenarios. However, their sensitivity and specificity are often sacrificed for simplicity and cost-effectiveness. Similarly, smartphone-based methods capitalize on portability but are typically adjunctive tools that depend on existing diagnostics rather than serving as independent solutions.

In summary, while emerging technologies offer innovative solutions, most fall short of meeting the critical requirements of portability and speed for POC diagnostics. The reliance on specialized equipment and the inability to consistently deliver results within 15 minutes remain significant barriers. Smartphone-based methods and PADs show potential in these areas but lack the diagnostic accuracy of more complex tools. Bridging these gaps will require focused innovation to integrate portability, rapid response times, and high diagnostic performance into a unified platform. This balance is essential for fulfilling the promise of POC diagnostics in diverse healthcare settings.

1.11 The Potential of Smartphones in Bridging Diagnostic Gaps

From the above discussion, it is evident that achieving sensitivity and specificity comparable to laboratory-based diagnostic tests will almost certainly require some form of specialized equipment. This reliance on advanced tools often limits portability, affordability, and accessibility, especially in resource-limited or decentralized healthcare settings. However, this challenge presents a unique opportunity to harness a device that is nearly ubiquitous in modern society: the smartphone.

Smartphones offer a compelling platform for diagnostic innovation. They combine significant computational power with advanced imaging capabilities, are inherently small and portable, and have robust connectivity features for data sharing and real-time updates. These attributes make smartphones ideal candidates for extending diagnostic capabilities through the development of extra attachments or modular add-ons. Leveraging this device effectively could revolutionize point-of-care (POC) diagnostics, bringing advanced testing capabilities to virtually any location.

The real opportunity lies in creating diagnostic technologies that can fully exploit the smartphone's potential. For this to succeed, a technology must be developed that makes the most of the smartphone's computational and imaging strengths. Such a technology should be capable of high sensitivity and specificity without requiring complex workflows or laboratory-grade equipment. Ideally, it would enable universal pathogen detection using a single, versatile reagent. Crucially, the ability to adapt to new pathogens through software updates, rather than by modifying reagents or surfaces, would mark a dramatic shift in diagnostic capabilities.

This vision represents a fundamental departure from the current paradigm of diagnostics, which relies on the specificity of biological reagents or engineered surfaces. Instead, it proposes a universal detection method, where pathogens are differentiated by computational algorithms rather than by the physical properties of the diagnostic system. This software-driven approach would allow for unprecedented flexibility, enabling rapid adaptation to emerging pathogens without requiring hardware changes or new reagents.

Such a shift would dramatically reduce the time and cost associated with updating diagnostic tools, making them more responsive to evolving public health threats. For instance, during a pandemic, new pathogens could be added to the detection system with a simple software update, rather than the laborious and expensive process of developing, manufacturing, and distributing new biological reagents. This adaptability would be transformative, enhancing global readiness for emerging diseases and reducing barriers to high-quality diagnostics.

In summary, the smartphone presents a unique opportunity to bridge the gap between laboratory-grade sensitivity and specificity and the accessibility required for POC diagnostics. By developing a technology that fully leverages the computational and imaging capabilities of smartphones, it may be possible to create a universal, software-driven diagnostic platform. This approach has the potential to redefine the field of diagnostics, enabling rapid, portable, and adaptive testing solutions that meet the needs of a rapidly changing healthcare landscape.

Chapter 2

Methods

> Scientific inquiry starts with observation. The more one can see, the more one can investigate.
>
> Martin Chalfie

2.1 Fluorescence Microscopy

Fluorescence microscopy is a powerful and versatile tool used in various scientific disciplines, particularly in the fields of biology and medicine. It is a type of light microscopy that uses fluorescence and phosphorescence instead of, or in addition to, reflection and absorption to study the properties of organic or inorganic substances.

The history of fluorescence microscopy can be traced back to the early 20th century. Sir George Gabriel Stokes discovered the concept of fluorescence itself in 1852 [145], but it wasn't until 1911 that Oskar Heimstädt [146] built the first fluorescence microscope.

One of the major milestones in the field was the development of the confocal microscope in the 1950s by Marvin Minsky [147]. This technology enabled three-dimensional imaging of specimens by using a pinhole to eliminate out-of-focus light in specimens that are thicker than the focal plane [148].

In the late 20th century, the advent of genetically encoded fluorescent proteins like Green Fluorescent Protein (GFP) [149] revolutionized fluorescence microscopy. These proteins, introduced into the cells or attached to proteins of interest genetically, allow scientists to observe processes within living cells in real-time [150].

The field of fluorescence microscopy has seen rapid advancements in the 21st century. Super-resolution microscopy techniques, such as STED (Stimulated Emission Depletion) and PALM [151]/STORM [152] (Photo-Activated Localization Microscopy/ Stochastic Optical Reconstruction Microscopy), have broken the diffraction limit, allowing scientists to visualize structures at the nanoscale [153].

In recent years, advancements in computational techniques and the development of novel fluorescent probes have further expanded the capabilities of fluorescence microscopy. Techniques like light-sheet fluorescence microscopy (LSFM [154]) and adaptive optics (AO) [155] have improved imaging speed, depth, and resolution, enabling more detailed and precise observations.

The field of fluorescence microscopy continues to evolve, with ongoing research and development promising even more sophisticated tools for scientific exploration in the future such as the recent work of Reinhardt et al [156] who introduced resolution enhancement by sequential imaging (RESI) to improve the resolution of fluorescence microscopy down to the Ångström scale.

2.1.1 How Does a Fluorescence Microscope Work?

The "fluorescence microscope" refers to any microscope that uses fluorescence to generate an image, whether it is a more simple setup like an epifluorescence microscope or a more complicated design such as a confocal microscope which uses optical sectioning to get better resolution of the fluorescent image.

The basic task of the fluorescence microscope is to let excitation light radiate the specimen and then sort out the much weaker emitted light to make up the image. A fluorescence microscope has several key components that allow it to function properly. These include:

- **Light Source**: The light source provides the initial illumination that the microscope uses to view the sample. In a fluorescence microscope, this is typically

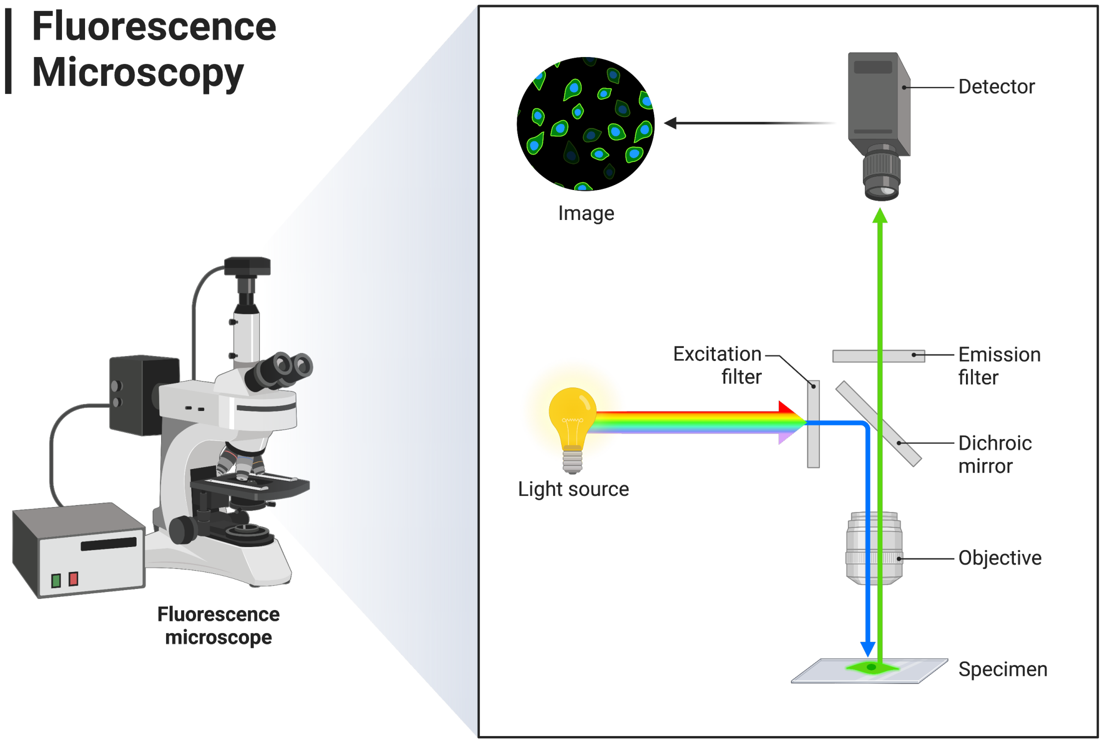


Figure 2.1: Schematic of a fluorescence microscope. Excitation light from the light source passes through an excitation filter, selecting the appropriate wavelength for fluorophore excitation. The light is then reflected by a dichroic filter and focused onto the specimen by the objective lens. Fluorescence emitted by the specimen passes back through the objective, then through the dichroic filter and an emission filter that isolates the emitted wavelength. The filtered fluorescence signal is captured by a detector, forming the final image. *Figure created with BioRender.com.*

a high-intensity light source such as a mercury-vapor lamp, xenon arc lamp, or more recently, LEDs.

- **Excitation Filter**: This is a filter that only allows light of a specific wavelength to pass through (the excitation wavelength). This light will excite the fluorophores in the sample.
- **Dichroic Mirror (or Beam Splitter)**: This is a special mirror that reflects light at certain wavelengths and allows light at other wavelengths to pass through. In a fluorescence microscope, it reflects the excitation light onto the sample, and then allows the longer wavelength emitted light to pass through.

- **Emission Filter**: After the light from the sample passes through the dichroic mirror, it goes through the emission filter. This filter only allows light of the emission wavelength to pass through, blocking out all other light.

- **Objective Lens**: The objective lens collects the fluorescence emissions and focuses them to create the image. The quality and type of the objective lens can greatly affect the resolution and quality of the image.

- **Detector**: The detector captures the image created by the emitted light. This could be an eyepiece for direct viewing, or more commonly in modern systems, a camera or photomultiplier tube that can capture and digitize the image for viewing and analysis on a computer.

- **Stage**: The stage holds the sample. It can usually be moved in all three dimensions to allow different areas of the sample to be viewed.

2.1.2 Epifluorescence (EPI) Microscopy: A Mainstay in Fluorescence Imaging

Epifluorescence (EPI) microscopy stands as one of the foundational techniques in the world of fluorescence imaging, offering a straightforward yet powerful approach to visualizing biological specimens. At its core, EPI microscopy involves illuminating the sample with a broad spectrum of light and then detecting the fluorescence emitted by the sample, all through the same objective lens. This concurrent illumination and detection mechanism distinguishes EPI from other microscopy techniques and is responsible for its widespread adoption in biological research.

The operational principle of EPI microscopy is relatively simple. A light source, typically a mercury or xenon lamp, emits a broad spectrum of light which is then passed through an excitation filter to select a specific wavelength range. This filtered light illuminates the sample, exciting any fluorescent molecules present. Upon relaxation, these molecules emit light at a longer wavelength, which is then separated from the excitation light using a dichroic mirror and further filtered by an emission filter to ensure only the desired fluorescence is detected. The resulting image provides a

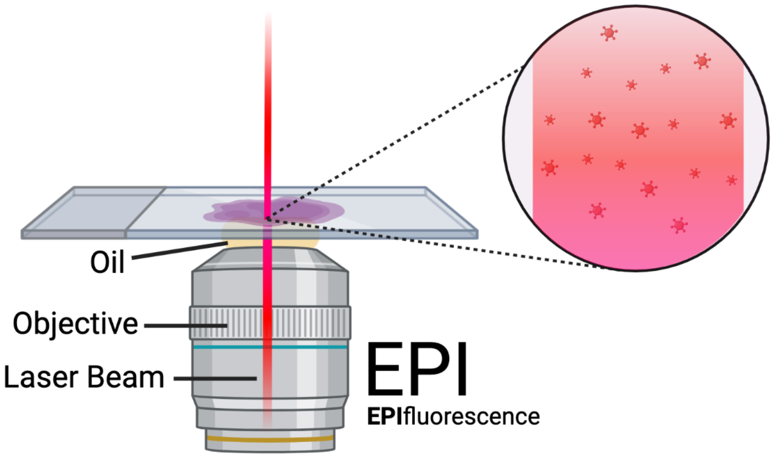


Figure 2.2: Schematic of epifluorescence (EPI) illumination in fluorescence microscopy. A laser beam is directed through the microscope's objective lens, which focuses the light through a thin layer of immersion oil and onto a glass slide containing the specimen. The excitation beam penetrates directly into the sample, uniformly illuminating the volume of interest. Fluorescently labelled viruses within the specimen are excited throughout the illuminated path and emit fluorescence, which is collected back through the same objective. This wide-field excitation mode allows visualization of all fluorophores in the illuminated volume, including out-of-focus light above and below the focal plane. *Figure created with BioRender.com.*

clear view of the distribution and intensity of fluorescent molecules within the sample [157].

While EPI microscopy offers several advantages, including ease of use, rapid imaging, and compatibility with a wide range of fluorescent probes, it is not without its limitations. The technique inherently captures fluorescence from the entire thickness of the specimen, leading to potential background noise from out-of-focus planes. However, despite this limitation, EPI microscopy remains a staple in many biological laboratories due to its versatility and the wealth of information it can provide about cellular structures, protein localization, and dynamic processes within living cells [158].

2.1.3 Highly Inclined and Laminated Optical Sheet (HILO) Microscopy: Illuminating the Near-Surface Region

In the realm of optical microscopy, the ability to selectively illuminate and visualize specific regions within a specimen is of paramount importance, especially when studying dynamic processes near the cell surface. Highly Inclined and Laminated Optical Sheet (HILO) microscopy emerges as a powerful technique tailored for this purpose, offering enhanced imaging of the near-surface region of biological samples.

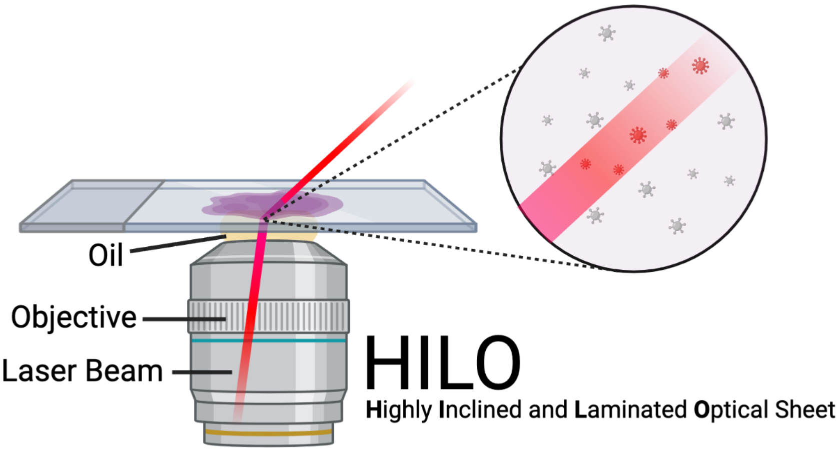


Figure 2.3: Schematic of HILO (Highly Inclined and Laminated Optical sheet) illumination in fluorescence microscopy. A laser beam is directed into the specimen at a highly inclined angle through the microscope's objective lens. After passing through the immersion oil and glass slide, the beam enters the sample at an oblique angle, creating a thin, inclined illumination sheet. This configuration selectively excites a narrow section of the sample, reducing background fluorescence from out-of-focus regions. Fluorescently labelled viruses within the inclined illumination volume are excited, and their emitted fluorescence is collected by the same objective. HILO illumination improves signal-to-noise ratio compared to wide-field epifluorescence while maintaining greater penetration depth than TIRF. *Figure created with BioRender.com.*

HILO microscopy employs a highly inclined laser beam to produce a thin, oblique plane of illumination within the specimen. This inclined beam intersects with the coverslip, creating a laminated optical sheet that illuminates a restricted, near-surface region of the sample. The result is a significant reduction in background fluorescence from regions outside this thin illuminated slice, leading to enhanced contrast and improved signal-to-noise ratio in the acquired images [159]. This is particularly advantageous when observing fast cellular dynamics or single-molecule events near the cell membrane, as it minimizes the interference from out-of-focus fluorescence.

2.1.3.1 Single Particle Tracking (SPT)

Single Particle Tracking (SPT) is a powerful technique used in fluorescence microscopy [160] to study the motion and dynamics of individual particles, such as molecules, nanoparticles, or fluorescently labeled virus particles, in various environments like a solution, on a membrane, or within a cell.

The fundamental principle behind SPT involves labeling the particles of interest with a fluorescent tag. These tags can be fluorescent proteins, small organic dyes, or even quantum dots, each chosen based on the specific requirements of the experiment. The labeled particles are then visualized using fluorescence microscopy, and the signal from each particle can be followed over time.

2.1.4 Total Internal Reflection Fluorescence (TIRF)

Total Internal Reflection Fluorescence (TIRF) microscopy [161] is an optical technique used to observe fluorescent molecules and structures near the surface or interface of a specimen with high sensitivity and spatial resolution. TIRF microscopy takes advantage of the phenomenon of total internal reflection, which occurs when light passes from a medium with a higher refractive index (e.g., glass) to a medium with a lower refractive index (e.g., liquid or air) at a certain critical angle.

When the incident light reaches or exceeds the critical angle, instead of being refracted into the specimen, it undergoes total internal reflection at the interface between the two media. This reflection creates an evanescent wave, an electromagnetic field that penetrates a few hundred nanometers into the specimen but decays exponentially with distance from the interface. The evanescent wave excites fluorophores in this thin region of the specimen near the surface, enabling selective illumination and imaging of structures in that specific region.

TIRF microscopy offers several advantages over other fluorescence microscopy techniques. Firstly, it provides excellent axial (vertical) resolution, as only fluorophores within the evanescent wave region are excited, while out-of-focus background fluorescence is greatly reduced. This leads to a high signal-to-noise ratio and improved

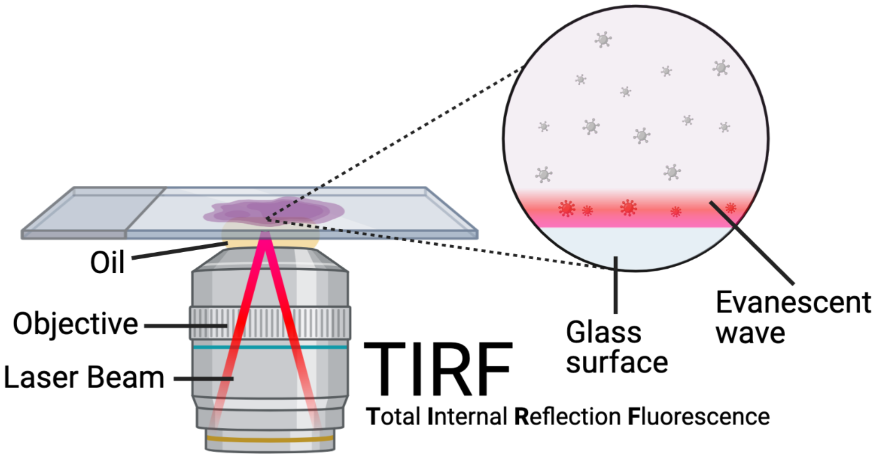


Figure 2.4: Schematic of Total Internal Reflection Fluorescence (TIRF) microscopy illumination. In TIRF microscopy, a laser beam is directed through the objective lens at an angle exceeding the critical angle for total internal reflection at the interface between the glass slide and the aqueous sample. This generates an evanescent field that penetrates only 100–200 nm into the sample above the glass surface. As a result, only fluorescently labelled viruses or molecules within this thin near-surface region are excited, while fluorescence from deeper in the sample is largely excluded. The emitted fluorescence is collected by the same objective, enabling high-contrast imaging with exceptional signal-to-noise ratio near the surface. *Figure created with BioRender.com.*

image contrast. Secondly, TIRF microscopy allows for high temporal resolution imaging, making it well-suited for capturing dynamic processes at the cell membrane or near interfaces. It is particularly useful for studying molecular interactions, membrane dynamics, and trafficking events.

The setup for TIRF microscopy involves directing a laser or other light source at an angle beyond the critical angle through a high numerical aperture objective lens. The incident light undergoes total internal reflection at the glass/sample interface, generating the evanescent wave that excites the fluorophores. The emitted fluorescence is then collected by the same objective lens and detected using a camera or other imaging system.

2.1.5 Super Resolution Microscopy: Beyond the Diffraction Limit

The study of cellular and molecular structures has been historically constrained by the diffraction limit of light, which restricts the resolution of conventional optical microscopes to approximately 200-250 nm. However, the advent of super-resolution microscopy (SRM) has revolutionized the field of optical imaging, allowing scientists to visualize biological structures at the nanoscale, far beyond the traditional diffraction barrier.

One of the pioneering techniques in SRM is Stimulated Emission Depletion (STED) microscopy. Introduced by Hell and Wichmann [162], STED employs a focused beam (the STED beam) to effectively 'switch off' fluorescence from molecules around a tiny volume, leaving only a sub-diffraction-sized volume to fluoresce. This results in images with resolutions better than the diffraction limit.

Another groundbreaking method is Photoactivated Localization Microscopy (PALM) and its related technique, Stochastic Optical Reconstruction Microscopy (STORM) [152]. These methods rely on the controlled activation and time-resolved localization of individual fluorescent molecules. By determining the centroid of each isolated molecule's image, PALM and STORM achieve resolutions down to 20 nm.

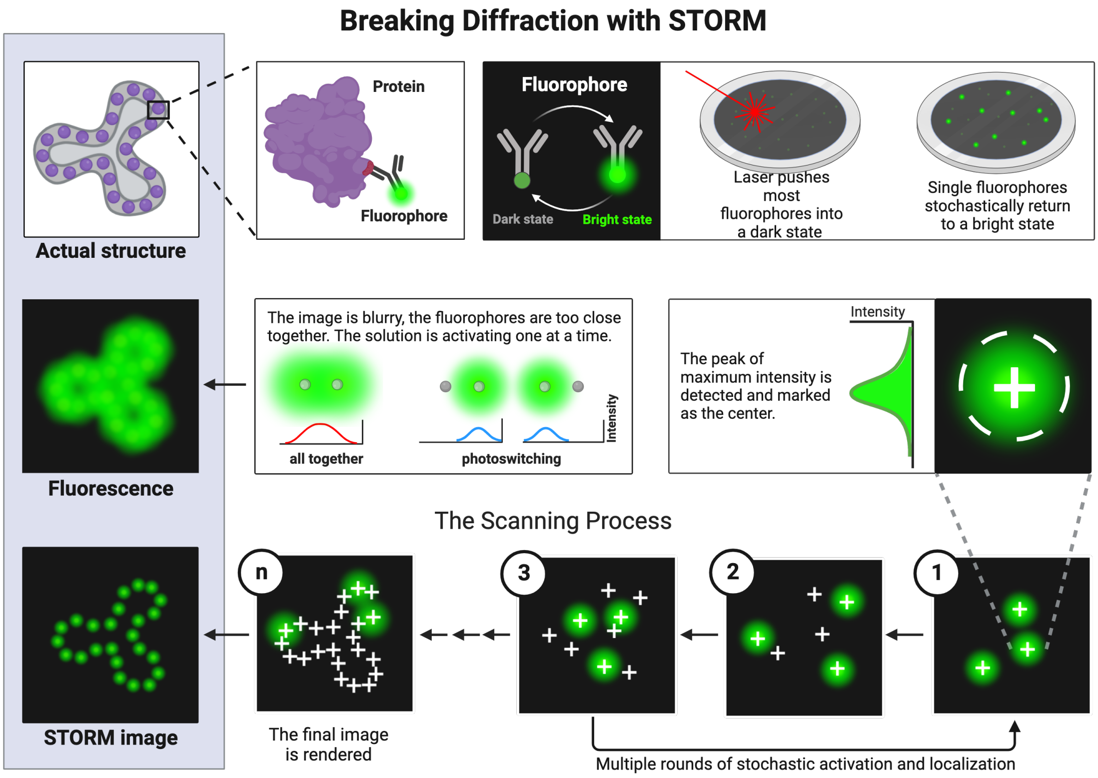
Structured Illumination Microscopy (SIM) is yet another approach that doubles the resolution of traditional microscopes by illuminating the sample with patterned light [163]. The resulting interference patterns, when processed, provide finer details than conventional imaging methods.

Lastly, Expansion Microscopy (ExM) offers a unique approach by physically expanding the sample. Tissue samples are embedded in a swellable polymer network, which, when expanded, allows for super-resolution imaging using standard microscopes [164].

In conclusion, super-resolution microscopy has ushered in a new era of biological imaging, enabling unprecedented insights into cellular structures and functions. As the field continues to evolve, it promises to unlock even more intricate details of the biological world.

2.1.5.1 Stochastic Optical Reconstruction Microscopy (STORM)

Stochastic Optical Reconstruction Microscopy (STORM) represents a significant leap in the realm of super-resolution microscopy, enabling researchers to visualize biological structures at the nanoscale, far beyond the diffraction limit of conventional optical microscopes. Developed in the mid-2000s, STORM capitalizes on the stochastic (random) activation and time-resolved localization of individual fluorescent molecules to reconstruct high-resolution images [152] (see Figure 2.5).


Breaking Diffraction with STORM

Laser pushes most fluorophores into a dark state

Single fluorophores stochastically return to a bright state

The image is blurry, the fluorophores are too close together. The solution is activating one at a time.

The peak of maximum intensity is detected and marked as the center.

The Scanning Process

The final image is rendered

Multiple rounds of stochastic activation and localization

Figure 2.5: STORM (Stochastic Optical Reconstruction Microscopy) achieves super-resolution imaging by utilizing the stochastic activation and precise localization of individual fluorophores. This method relies on the sequential activation of a sparse subset of fluorophores, their localization with nanometer precision, and the reconstruction of an image from these localized points. The figure illustrates the process, showing (A) the actual structure of the sample, (B) the image with overlapping fluorophores, and (C) the final super-resolved reconstruction, revealing fine structural details beyond the diffraction limit. Adapted from Rust et.al. [152] *Figure created with BioRender.com.*

The fundamental principle behind STORM is the ability to temporally separate the fluorescence emission of individual molecules within a densely labeled sample. This is achieved by using photoswitchable or photoactivatable fluorophores that can be toggled between a fluorescent "on" state and a non-fluorescent "off" state. During a STORM imaging session, only a sparse subset of these fluorophores is activated to the "on" state at any given time, ensuring that their emitted fluorescence spots are well-separated and can be precisely localized. By repeating this process thousands to millions of times and localizing the position of each molecule in each frame, a super-resolved image is reconstructed from the cumulative positions of all localized molecules [165].

STORM's ability to achieve resolutions on the order of 20 nm has opened new avenues in biological research, allowing for the detailed visualization of intricate cellular structures, protein complexes, and other nanoscale entities. However, the technique requires specialized fluorophores, careful sample preparation, and sophisticated imaging setups. Despite these challenges, STORM has been instrumental in advancing our understanding of cellular and molecular biology, offering unprecedented insights into the nanoscopic world that were previously unattainable with traditional microscopy methods [166].

2.2 Machine Learning

Machine Learning (ML) is a subset of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The concept of machine learning is centered around the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

The history of machine learning dates back to the mid-20th century. The term "machine learning" was first coined by Arthur Samuel in 1959, who was a pioneer in the field of artificial intelligence and computer gaming. He defined machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed" (Samuel, 1959) [167].

The first major milestone in the history of machine learning was the development of the perceptron in 1957 by Frank Rosenblatt. The perceptron was the first algorithm that could learn from its mistakes, making it a significant step towards the development of modern machine learning [168].

In the 1990s, the development of the Support Vector Machine (SVM) marked another major milestone in the field. SVMs are a set of supervised learning methods used for classification and regression analysis, and they played a crucial role in the advancement of machine learning [169].

The 2010s saw the rise of deep learning, a subset of machine learning that focuses on algorithms inspired by the structure and function of the brain called artificial neural networks. Deep learning has been responsible for many of the recent advancements in machine learning, including the development of advanced speech recognition, image recognition, and natural language processing systems [170].

Machine learning has a wide range of applications across various industries. In healthcare, machine learning algorithms are used to predict disease outcomes and personalize treatment plans. In finance, they are used for credit scoring and algorithmic trading. In transportation, machine learning is used in the development of autonomous vehicles. In e-commerce, machine learning algorithms are used for recommendation systems and customer segmentation.

2.2.1 Machine Learning in Diagnostics

Machine learning has been increasingly utilized in the field of disease diagnostics, as highlighted by several key studies. Fatima and Pasha (2017) conducted a comprehensive survey of machine learning algorithms used for diagnosing various diseases, including heart disease, diabetes, liver disease, dengue, and hepatitis. They emphasized the crucial role of machine learning in enhancing computer-aided diagnosis applications and improving the accuracy of disease perception and diagnosis [171].

In the realm of dermatology, Haenssle et al. (2018) developed a deep learning-based approach for melanoma detection using dermoscopic images. Their study compared the diagnostic performance of a deep learning convolutional neural network with 58 dermatologists. Remarkably, the network outperformed the dermatologists in terms of specificity, demonstrating the potential of machine learning in advancing diagnostic accuracy [172].

Similarly, in cardiology, Coenen et al. (2018) employed a machine-learning approach to improve the performance of coronary computed tomographic angiography (CTA) in detecting functionally obstructive coronary artery disease. Their findings showed that the machine-learning approach not only improved the diagnostic accuracy of CTA but also performed equally well as a computationally more demanding method [173].

In the field of behavioral sciences, particularly autism research, Bone et al. (2014) critically evaluated the use of machine learning. While they acknowledged the potential of machine learning in enhancing diagnostic research in autism, they also cautioned against its use in the absence of clinical domain expertise. Their study underscores the importance of integrating machine learning with clinical knowledge for effective and reliable diagnostics [174].

In summary, these studies underscore the transformative potential of machine learning in disease diagnostics across various medical fields. However, they also highlight the importance of integrating these advanced computational methods with clinical expertise to ensure reliable and effective diagnostic outcomes.

2.2.2 Machine Learning in Microscopy

Machine learning has been increasingly leveraged in the field of fluorescence microscopy, as demonstrated by several key studies. Mannam et al. (2020) explored the application of machine learning techniques to fluorescence lifetime imaging microscopy (FLIM), a powerful tool in biomedical research. They highlighted how machine learning can efficiently extract and interpret measurements from multi-dimensional FLIM datasets, thereby improving speed and accuracy over conventional methods. They also discussed the potential for further enhancements to FLIM through machine learning [175].

In a study by Koydemir et al. (2017), a field-portable and cost-effective microscopy platform was developed for the automated detection and counting of Giardia cysts in water samples. The researchers compared the sensitivity and specificity of their platform using multiple supervised classification models, demonstrating that a bootstrap aggregating approach provided the best performance for automated detection of Giardia cysts [176].

Wu and Rifkin (2015) introduced a computational pipeline capable of distinguishing the true signal of fluorescently labeled molecules from background fluorescence and noise. Their software, which uses a supervised random forest classifier, recognizes and classifies individual mRNA spots by measuring several features of local intensity maxima. Importantly, their software can estimate the error introduced by image classification, providing a unique measure of image quality and classification confidence [177].

Borgmann et al. (2016) presented a workflow that combines high-resolution fluorescence microscopy, image processing, and machine learning for the identification of the Rhesus D antigen. Their technique enables the identification of even small amounts of D antigen at the cellular level, allowing for automated population analyses and resulting in high classification test accuracies [178].

In summary, these studies underscore the transformative potential of machine learning in fluorescence microscopy, demonstrating its capacity to enhance speed, accuracy, and the overall quality of image analysis.

2.3 Overall Method Development and Stages

The methodological development presented in this thesis was structured to address the limitations of existing viral diagnostics, with a focus on achieving universality, scalability, and portability. The work followed a logical sequence, progressing through four conceptual stages, each informing the next and shaping the subsequent experimental design.

The first stage, Problem Identification, involved a critical evaluation of existing diagnostic technologies and recognition of a significant gap: the absence of a universal, reagent-agnostic, and portable method capable of meeting ASSURED/REASSURED standards. This realisation defined the strategic direction of the research. Founda-

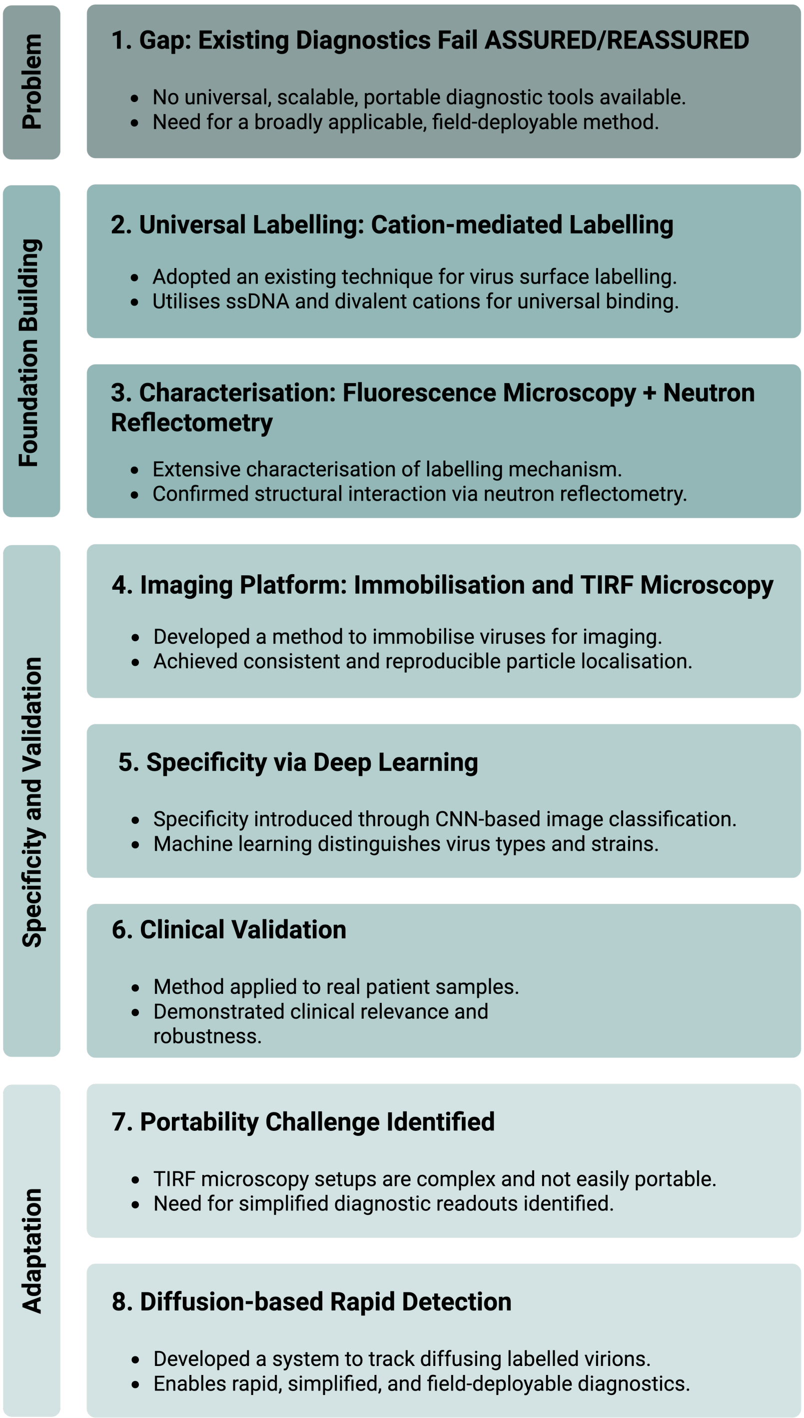


Figure 2.6: Schematic overview of method development and stages. The research progressed through four interconnected stages: identifying the diagnostic gap (Problem Identification), establishing a universal labelling and imaging platform (Foundation Building), introducing computational specificity and validating performance on clinical samples (Specificity and Validation), and adapting the methodology for field deployment through diffusion-based detection (Adaptation). Colour groupings represent conceptual themes, reflecting the logical structure underpinning the methodological development.

tion Building formed the second stage. A cation-mediated labelling approach was adopted due to its potential for universal applicability across enveloped viruses. Extensive characterisation of the labelling mechanism was performed using fluorescence microscopy and other biophysical methods, systematically assessing the effects of DNA length, ion concentration, pH, lipid composition, and surface protein content. Based on these findings, an imaging platform was developed by immobilising labelled viral particles on surfaces and imaging them with total internal reflection fluorescence (TIRF) microscopy, thereby enabling reproducible and stable acquisition of particle-level fluorescence signals.

The third stage focused on Specificity and Validation. Specificity was introduced through the development of convolutional neural networks (CNNs) capable of classifying viral types and strains based on subtle differences in fluorescence imaging features. The approach was then validated using clinical samples, demonstrating sensitivity, specificity, and robustness in real-world diagnostic settings.

The final stage, Adaptation, addressed the practical limitations of the immobilisation-based TIRF method. Recognising the challenges associated with translating this system into portable, field-deployable formats, a diffusion-based detection approach was developed. By tracking the movement of labelled virions in solution, this method has the potential to enable rapid, simplified diagnostics using optical setups more amenable to point-of-care deployment.

The overall methodological progression is summarised schematically in Figure 2.6, illustrating the sequential development from problem identification to the construction of a field-adaptable diagnostic platform. Each of the subsequent chapters corresponds to one or more elements of this framework, beginning with the characterisation of the cation-mediated labelling strategy described in the following chapter.

Chapter 3

Cation-Mediated Labelling Characterisation

We don't start from scratch; we continue the narrative.

3.1 Previous Work

The research paper "Rapid functionalisation and detection of viruses via a novel Ca2+-mediated virus-DNA interaction" by Nicole Robb, Achillefs N. Kapanidis and colleagues, which was published in Scientific Reports[179], serves as the foundational basis for the labelling technique that I will be employing throughout this project. This innovative method leverages a calcium-mediated interaction between the surface of enveloped viruses and DNA, enabling the functionalisation of virus particles with chemical groups attached to the DNA.

The technique utilises DNA modified with fluorophores for the labelling and detection of viruses, such as influenza, through single-particle tracking and particle-size determination. Remarkably, this approach enables the detection of clinical isolates of influenza in just one minute, a significant improvement over existing diagnostic tests. The method has been validated on all enveloped viruses tested, including those present in clinical samples, and does not necessitate any amplification or purification steps.

As I progress further into this project, I will be extending the foundational work to probe deeper into the underlying mechanisms of this labelling technique. I plan to explore its compatibility with other elements not yet tested, and scrutinize how it interacts with different variables such as pH and lipid composition. This exploration aims to provide a more holistic understanding of this method and its potential applications in the realm of virus detection.

While the labelling technique swiftly labels viruses, it does so indiscriminately, attaching DNA to any entity with a lipid membrane. Therefore, I aim to develop an assay or method to introduce specificity to the labelled particles, thereby enhancing the precision and utility of this technique. Moreover, a significant portion of my DPhil research will be dedicated to characterising this labelling process using biophysical methods which is the main topic of this chapter.

3.2 Methods and Materials

3.2.1 Lab-grown Viruses and Virus-like Particles

The study utilized WSN (H1N1 A/WSN/33) and recombinant Chikungunya Virus-Like Particles (VLPs) for different experimental setups. WSN was employed for microscopy experiments and was propagated in Madin–Darby canine kidney (MDCK) cells, with viral concentrations determined using a plaque assay.

Recombinant Chikungunya VLPs were used for quartz crystal microbalance (QCM) experiments. These VLPs were produced in mammalian HEK293 cells, where Chikungunya structural proteins were transiently expressed to form particles comprising the E1, E2, and capsid proteins. The VLPs lacked non-structural proteins and genomic material, making them completely non-infectious. The resultant particles were greater than 90% pure and were procured from the Native Antigen Company.

3.2.2 Single Stranded DNAs

We procured single-stranded oligonucleotides modified at the $5'$ end with ATTO647N dye from IBA, based in Germany. Our choice of oligonucleotides was guided primarily by their length and the use of a bright and photostable fluorophore. To eliminate concerns about secondary structure formation, we specifically used poly-thymidine sequences, as these do not form stable secondary structures under experimental conditions.

For our microscopy experiments, we employed single-stranded DNA of varying lengths: a 40mer (5'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTT3'), a 20mer (5'TTTTTTTTTTTTTTTTTTTT3'), a 14mer (5'TTTTTTTT TTTTTT3'), and an 8mer (5'TTTTTTTT3'). Each oligonucleotide was labeled at the $5'$ end with ATTO647N, ensuring consistent fluorescence labeling. These labeled oligonucleotides were critical for fluorescence microscopy, enabling the visualization of single molecules under our experimental conditions.

For quartz crystal microbalance (QCM) experiments, we used single-stranded DNA derived from calf thymus, procured from Sigma-Aldrich. We did not focus on a specific DNA sequence because the purpose of the experiments was to study bulk interactions between DNA and lipid bilayers, rather than sequence-specific effects. Calf thymus DNA was used as a convenient, natural source of nucleic acids for surface functionalisation and interaction studies, and was sonicated to fragment it into smaller pieces, improving diffusion and access to the bilayer surface without introducing sequence-dependent variability.

3.2.3 Lipid Mixtures

POPC (1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine), POPS (1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-L-serine, sodium salt), POPE (1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine), and Liver Total Lipid Extract (Bovine) were procured from Avanti Polar Lipids (Alabaster, AL, USA) and used without further purification.

Artificial lipid membranes were used to mimic biological membranes. Lipids (described) were supplied in high concentrations, either freeze-dried or suspended in chloroform. To prepare these lipids for use, the chloroform was removed, and different lipid types were combined before being suspended in the desired buffer. The preparation began with the cleaning of glass vials, which were washed with ethanol, dried with nitrogen, and rinsed with chloroform in a fume hood. The desired amount of lipid was withdrawn from the concentrated stock solution, calculated based on the final concentration and volume requirements, using a pre-rinsed Hamilton syringe. The chloroform was then evaporated under a gentle nitrogen gas flow to prevent the lipid solution from splashing. Once the lipid appeared hazy with no visible liquid, it was vacuum-dried for at least one hour or overnight. Finally, the dried lipid was resuspended in the desired buffer at the calculated volume and concentration. Prepared lipid mixtures were stored at $-20\,^{\circ}\mathrm{C}$ for up to six months and diluted as needed for specific experiments, such as QCM-D.

3.2.4 Quartz Crystal Microbalance with Dissipation Monitoring (QCM-d)

Quartz Crystal Microbalance with Dissipation Monitoring (QCM-D) is a technique that allows for the real-time study of surface interactions. At the heart of the technique is a thin quartz crystal disk, sandwiched between two electrodes. The piezoelectric properties of quartz cause the crystal to oscillate at its resonant frequency when an alternating voltage is applied to the electrodes.

This resonant frequency is highly sensitive to mass changes on the crystal's surface. When a substance, such as a liquid or gas, interacts with the crystal, it alters the crystal's mass and consequently its resonant frequency. This frequency shift (Δf) can be used to calculate the mass change (Δm) on the crystal surface. This relationship is described by the Sauerbrey equation, given as:

$$\Delta m = -\frac{C\Delta f}{n} \tag{3.1}$$

where C is the mass sensitivity constant for the crystal (depends on the properties of the quartz and the overtone), n is the overtone number (integer multiples of the fundamental frequency), and Δf is the change in frequency. This equation assumes the layer forming on the crystal is rigid and evenly distributed.

However, for viscoelastic layers, such as polymers or biological membranes, the Sauerbrey equation is not sufficient. This is where dissipation monitoring becomes crucial. Dissipation (D) measures the energy loss in the oscillation system per oscillation cycle, providing information about the viscoelastic properties of the layer. A higher dissipation value indicates a softer, more viscoelastic layer, while a lower value suggests a more rigid layer.

QCM-D measures both the frequency shift and the dissipation simultaneously, offering a comprehensive understanding of the layer forming on the crystal surface. For instance, a decrease in frequency (indicating a mass increase) with constant dissipation suggests the formation of a rigid layer. If the dissipation also increases, it indicates the formation of a viscoelastic layer.

One of the key advantages of QCM-D is its ability to monitor surface interactions in real-time. This enables the observation of layer formation and changes over time, providing valuable insights into the kinetics of the process. Therefore, QCM-D is a powerful tool for studying surface interactions, offering information about the mass, structure, and viscoelastic properties of layers on a surface.

3.2.5 Supported Lipid Bilayers

To create a supported lipid bilayer, lipids were first combined in glass vials at specific concentrations to achieve the desired composition. Initially, the lipids formed large vesicles, which were sonicated to reduce their size to less than 200 nm. This size reduction was visually indicated by the solution transitioning from a cloudy appearance to a transparent blue hue due to decreased light scattering by smaller vesicles. Once prepared, the lipid vesicle mixture was pumped into the flow cell (see Figure 3.1 **A**).

The initial step involved using a high concentration of NaCl to saturate the surface with vesicles. Following this, a low NaCl concentration or salt-free buffer was introduced to induce osmotic shock. This caused the vesicles to absorb water, grow in size, and rupture into fragments, which adhered to the surface and merged to form a continuous supported lipid bilayer(see Figure 3.1 **B**).

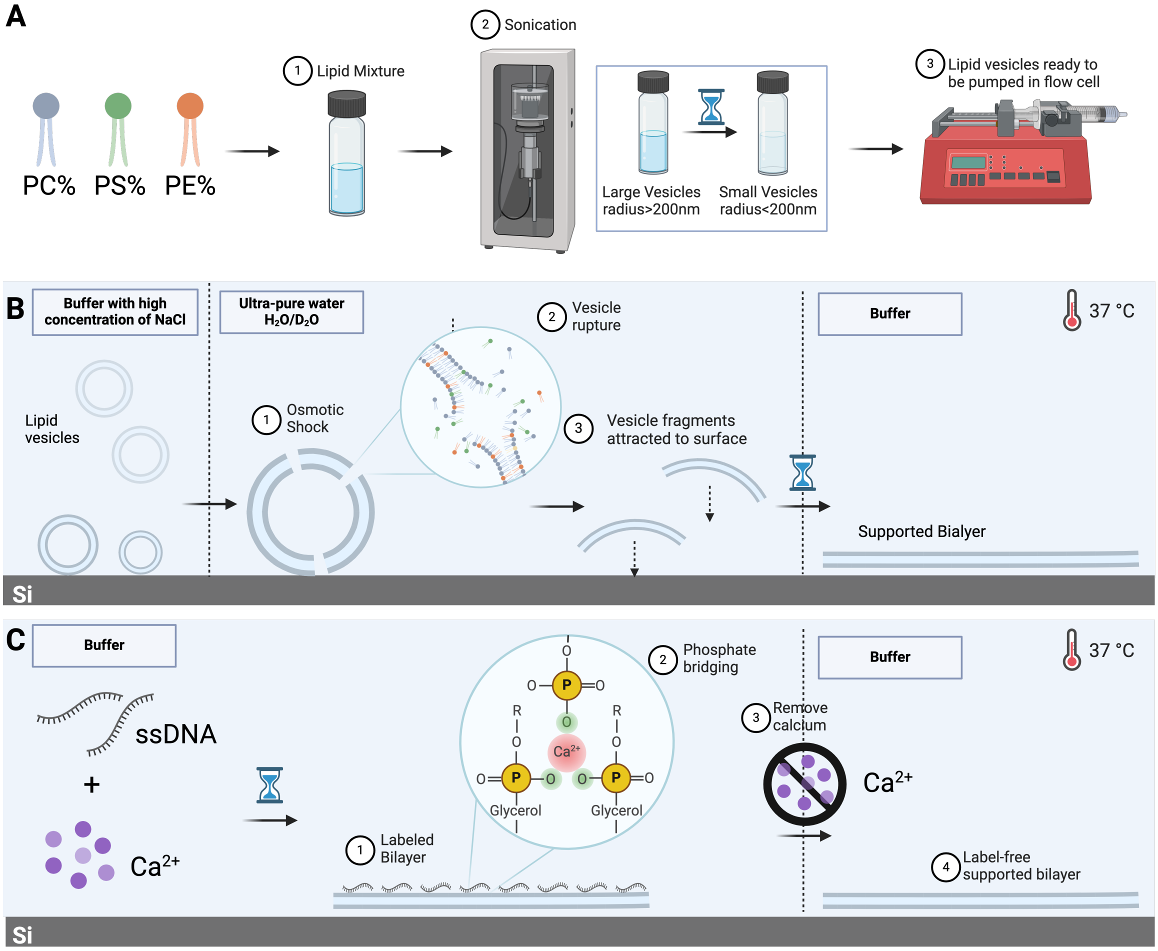


Figure 3.1: Preparation and formation of a supported lipid bilayer. (A) Lipids were mixed at desired compositions and concentrations, forming large vesicles that were sonicated to reduce their size to less than 200 nm, indicated by a shift from a cloudy to a transparent blue hue. (B) High-concentration NaCl was used to saturate the surface with vesicles, followed by low or no salt to induce osmotic shock, causing vesicle rupture and bilayer formation. (C) The supported lipid bilayer was used to study cation-mediated labeling by flowing single-stranded DNA (ssDNA) in calcium chloride solution.

The resulting bilayer was utilized for studies involving cation-mediated labeling. In this process, single-stranded DNA (ssDNA) suspended in a calcium chloride solution was pumped through the bilayer to facilitate labeling interactions (see Figure 3.1 **C**).

3.2.6 Formation of Viral Supported Bilayers

To create a supported viral bilayer, chikungunya virus-like particles (VLPs) were prepared and introduced into the system. Initially, the VLPs were suspended in a high-salt buffer (460 mM NaCl, 20 mM HEPES, 2 mM $CaCl_2$, ph 7) at a controlled temperature of 37 °C and a flow rate of 0.25 mL/min. The high NaCl concentration facilitated the deposition of VLPs onto the SiO_2 sensor surfaces.

Once the VLPs were adsorbed onto the surface, the system was transitioned to a low-salt or salt-free buffer (0 mM NaCl, 20 mM HEPES, 2 mM $CaCl_2$, pH 7). This change induced osmotic shock, causing the VLPs to swell, rupture, and release fragments of their viral membrane. These fragments adhered to the surface and spread to form a continuous viral supported bilayer (see Figure 3.2).

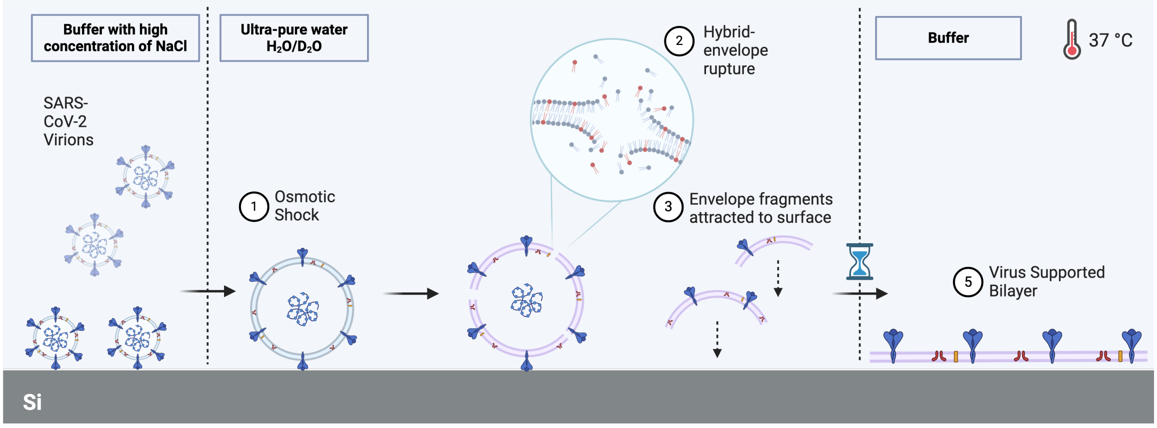


Figure 3.2: Schematic representation of viral supported bilayer formation. (1) Osmotic shock induced by transitioning from high to low salt buffer causes the viral envelope to swell. (2) The viral envelope ruptures, releasing membrane fragments. (3) The fragments adhere to the surface, forming a continuous viral supported bilayer.

The resulting bilayer was used for cation-mediated DNA labeling studies. In this step, single-stranded DNA (ssDNA) in a calcium chloride solution (e.g., 200 mM $CaCl_2$, 20 mM HEPES, pH 7) was flowed through the system. The calcium ions facilitated interactions between the ssDNA and the viral bilayer, enabling effective labelling.

3.3 Results

3.3.1 Single-stranded DNA Length Effect

This study investigates a novel mechanism for labeling diffusing virions using fluorophore-modified ssDNA and cation-mediated interactions. By exploring the effects of ssDNA length and calcium ion (Ca^{2+}) concentration at pH 8, we aimed to better understand how these factors influence binding efficiency and single-particle tracking (SPT) outcomes.

The number of tracks varied significantly with ssDNA length and Ca^{2+} concentration. The 8mer produced no detectable tracks at any Ca^{2+} concentration, suggesting it was too short for stable binding. In contrast, the 14mer, 20mer, and 40mer showed clear trends. The 14mer peaked at 1 M Ca^{2+} but fell to zero at 0.75 M, while the 20mer and 40mer also peaked at 1 M but decreased more gradually. The 40mer retained more tracks at lower Ca^{2+} concentrations, highlighting the advantages of longer ssDNA for robust binding.

We extended the original work from Robb et.al. [179], which suggested the 20mer was the minimum effective length, by showing that the minimum ssDNA length depends on the Ca^{2+} concentration. Unlike previous studies, we selected ssDNA sequences composed entirely of thymine residues, ensuring the absence of secondary structures. This demonstrates that the labeling mechanism is not driven by secondary structure but instead relies on direct interactions mediated by the cationic environment. Our results reveal that higher Ca^{2+} concentrations at pH 8 enhance labeling efficiency for all ssDNA lengths except the 8mer, highlighting how the interplay of these factors influences binding and revising the conclusions of earlier studies.

This work sheds light on the interplay between ssDNA length, Ca^{2+} concentration, and pH in cation-mediated labeling. Stronger electrostatic interactions at higher Ca^{2+} levels likely promote more stable binding, particularly for longer ssDNA. The inability of the 8mer to generate tracks even under the stated experimental conditions underscores the importance of length for this mechanism.

3.3.2 pH Effects

The pH of the experimental environment significantly influences the efficiency of ssDNA labeling of WSN virions, as demonstrated by experiments with the 14mer and 20mer ssDNA constructs at pH 8 and pH 9. The average number of tracks per Ca^{2+} concentration for each condition is presented in Figure 3.4a (14) and Figure 3.4b (20mer).

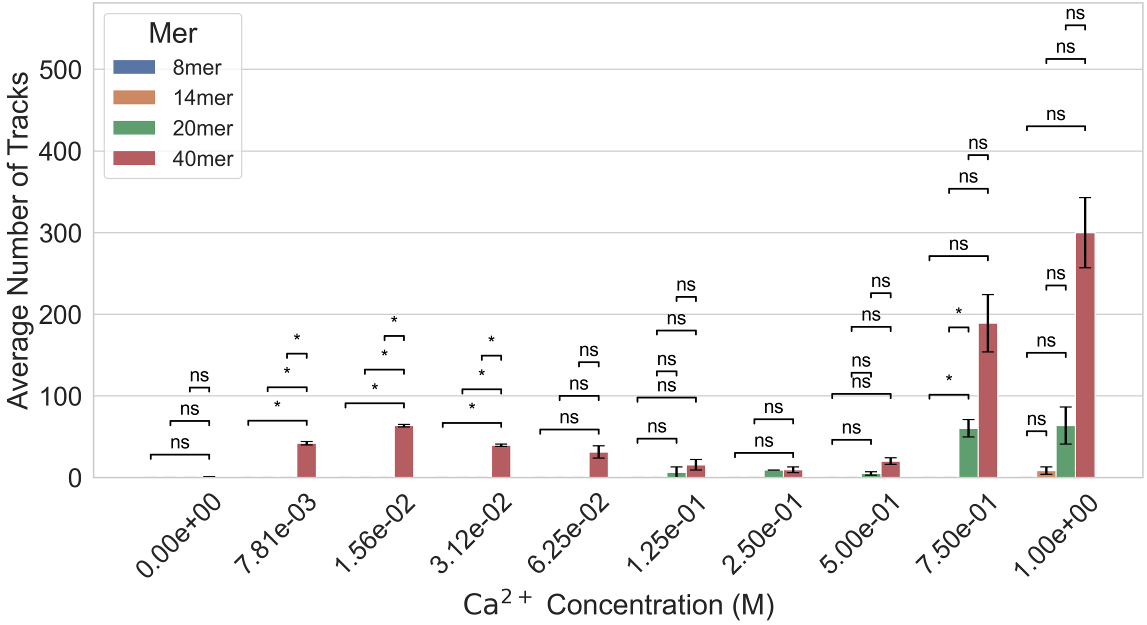
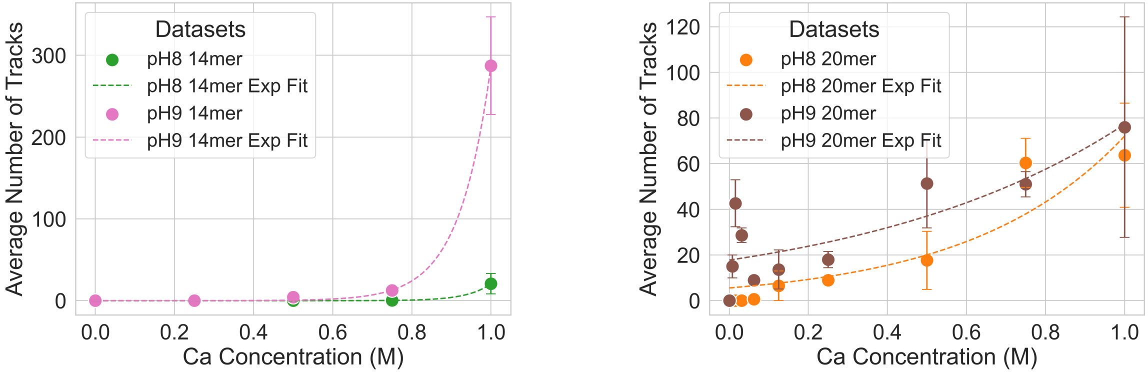


Figure 3.3: The number of tracks obtained from single-particle tracking experiments as a function of Ca^{2+} concentration for single-stranded DNA (ssDNA) lengths of 8, 14, 20, and 40 nucleotides (all thymine residues). Each condition was measured in biological triplicates. The 8mer showed no detectable tracks at any Ca^{2+} concentration, indicating insufficient binding. The 14mer exhibited a peak in tracks at 1.0 M, but decreased to zero by 0.75 M. The 20mer and 40mer also peaked at 1.0 M, with a gradual reduction in track number as Ca^{2+} concentration decreased. The longer ssDNA (40mer) displayed greater persistence of tracks at lower Ca^{2+} concentrations, highlighting the role of ssDNA length and cation-mediated interactions in labeling efficiency. Error bars represent the standard error of the mean across triplicates. Statistical comparisons between different mers at each concentration were performed using Welch's two-tailed t-test. Significant differences were found at low Ca^{2+} concentrations: at 7.81×10^{-3} M, the 40mer differed significantly from the 8mer, 14mer, and 20mer ($p = 0.030$); at 1.56×10^{-2} M, similar significant differences were observed ($p = 0.015$); and at 3.12×10^{-2} M, again the 40mer differed significantly from the shorter mers ($p = 0.024$). At higher concentrations, no statistically significant differences were found between the different ssDNA lengths ($p > 0.05$).

For the 14mer, the number of tracks was substantially higher at pH 9 compared to pH 8, indicating enhanced binding efficiency at the higher pH. At pH 8, the number of tracks peaked at 1 M Ca^{2+} and declined to zero at 0.75 M. In contrast, at pH 9, the tracks also peaked at 1 M but dropped to zero more rapidly, reaching zero at 0.25 M Ca^{2+}. This suggests that while the overall binding efficiency of the 14mer improves at pH 9, the interaction becomes less stable at lower Ca^{2+} concentrations under these conditions.

For the 20mer, the trend was similar, with generally higher track counts observed at pH 9 compared to pH 8. However, the increase in tracks at pH 9 was less pronounced than for the 14mer. At both pH values, the number of tracks peaked at 1 M Ca^{2+} and decreased as the Ca^{2+} concentration dropped, eventually reaching zero at lower concentrations. This suggests that the 20mer maintains more consistent labelling performance across pH values compared to the 14mer.

The differences observed between pH conditions can be attributed to the effect of pH on the effective charge of the virions. At higher pH, the surface charge of the virions becomes more negative due to deprotonation of acidic residues. This increased negative charge likely enhances the interaction with the positively charged calcium ions, facilitating stronger ssDNA binding. However, this interaction also becomes more dependent on maintaining high Ca^{2+} concentrations to neutralize the increased negative charge. Consequently, at pH 9, the enhanced binding efficiency is accompanied by a steeper decline in tracks as Ca^{2+} concentrations decrease.



(a) 14mer.

(b) 20mer.

Figure 3.4: Average number of tracks per calcium concentration at pH 8 and 9.

To further investigate the influence of pH on ssDNA labeling, the average maximum photocount per track was analyzed for each ssDNA length (8mer, 14mer, 20mer, and 40mer) across all Ca^{2+} concentrations at pH 8 and pH 9 (see Figure 3.5). The photocount data provide additional insights into the efficiency and effectiveness of ssDNA-mediated labeling under varying pH conditions.

The results show that the photocount increases with ssDNA length, with the 40mer exhibiting the highest average maximum photocount at pH 9, significantly surpassing the other lengths. At pH 8, however, the photocount for the 40mer decreases and becomes comparable to the values observed for the 20mer and 14mer. Between these two, the 20mer consistently produces slightly higher photocounts than the 14mer at both pH values, reflecting its enhanced binding stability. The 8mer, in contrast, shows a significantly lower photocount at pH 9 and no detectable tracks (photocount = 0) at pH 8.

The stark difference in photocounts for the 40mer between pH 9 and pH 8 highlights the importance of pH in enhancing labeling efficiency for longer ssDNA constructs. At pH 9, the increased negative charge of the virion surface promotes stronger cation-mediated interactions, allowing the 40mer to achieve superior fluorescence signals. The 20mer and 14mer show moderate sensitivity to pH changes, maintaining reasonable photocounts across both conditions, albeit with reduced efficiency at pH 8. In contrast, the 8mer's performance is highly pH-dependent, with detectable photocounts only at pH 9, further emphasizing that the minimum ssDNA length required for successful labeling is pH-sensitive.

These findings reinforce the conclusion that pH not only affects the efficiency of ssDNA labeling but also determines the minimum ssDNA length required for stable interaction. Longer ssDNA constructs, such as the 40mer, benefit the most from the increased negative charge of the virion surface at higher pH, while shorter ssDNA lengths struggle to form stable interactions, particularly at pH 8. This understanding is crucial for optimizing experimental conditions for cation-mediated labeling in single-molecule fluorescence experiments.

3.3.3 Reversible Interactions Between ssDNA and Bilayers

The initial stage of the experiment involved the formation of a supported lipid bilayer (SLB) on the SiO_2 sensor surface, which served as a biomimetic platform for studying DNA binding dynamics. The lipid mixture, consisting of 1 mg/mL 5:2:3 POPE:POPS:POPC, was introduced in a buffer containing 500 mM NaCl and 2 mM

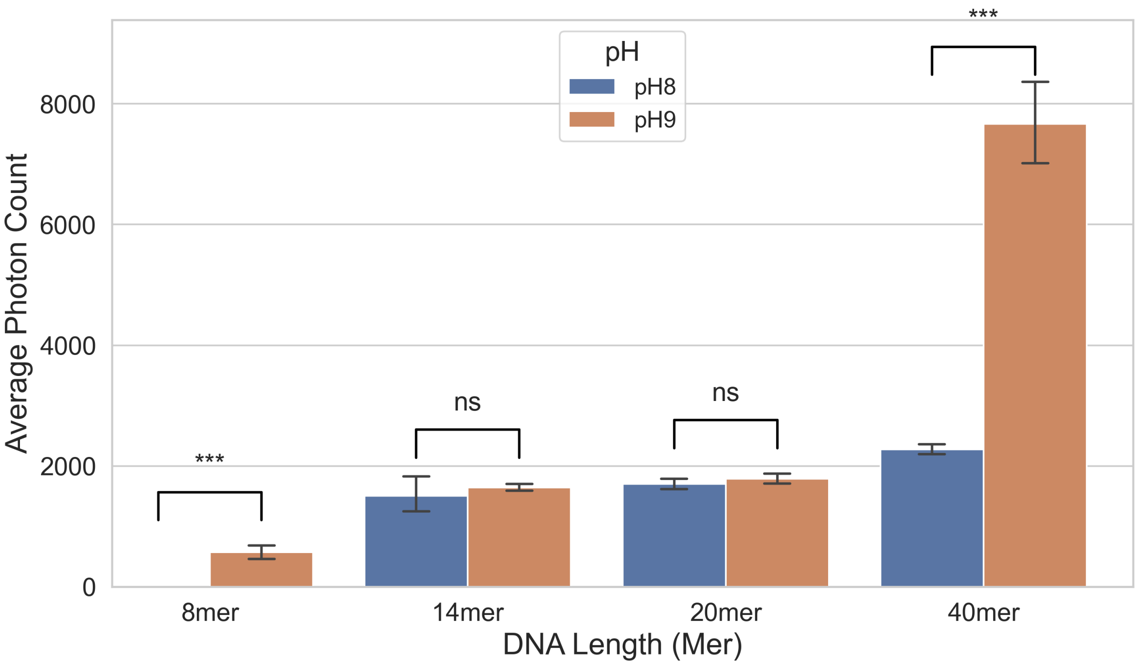


Figure 3.5: Average maximum photon counts per track for ssDNA of different lengths (8mer, 14mer, 20mer, and 40mer) at pH 8 and pH 9. For each condition, triplicate measurements were performed. Statistical significance between pH 8 and pH 9 was assessed for each DNA length using an unpaired t-test. The resulting p-values were: 8mer: 1.0×10^{-10}, 14mer: $3.545\,551 \times 10^{-1}$, 20mer: $1.522\,439 \times 10^{-1}$, and 40mer: $9.049\,814 \times 10^{-52}$. Statistical significance is indicated as ns (not significant, $p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$), or *** ($p \leq 0.001$). Horizontal brackets mark comparisons between pH 8 and pH 9 for each DNA length.

$CaCl_2$ at pH 8. This resulted in a significant decrease in frequency and an increase in dissipation, as observed in the QCM-D data. These changes indicated the adsorption of lipid vesicles onto the sensor surface, followed by their rupture and fusion to form a continuous bilayer. The frequency decrease during this process reflected the addition of mass, while the stabilization of dissipation values demonstrated the transition to a rigid and well-ordered bilayer. The presence of calcium ions was critical for this process, as they neutralized the negative charges on the lipid headgroups, reduced electrostatic repulsion, and promoted vesicle fusion. Once the SLB was established, ssDNA was introduced to assess its binding behavior. The binding of ssDNA to the bilayer was driven by electrostatic interactions and calcium-mediated bridging, resulting in concentration-dependent changes in the QCM-D signals. At 100 nM ssDNA, the largest frequency decrease and dissipation increase were observed, indicative of

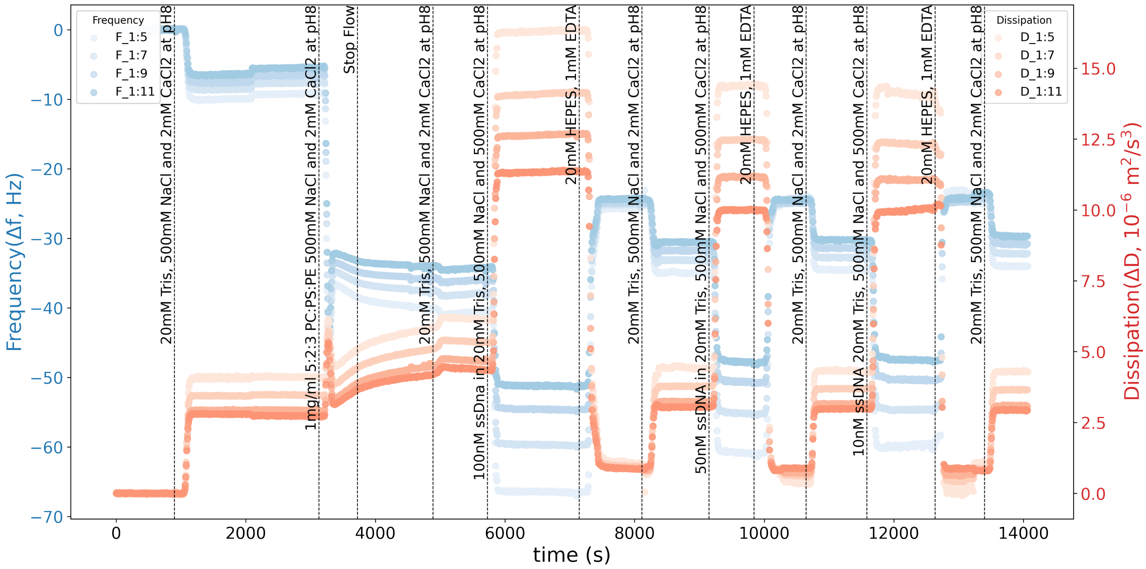


Figure 3.6: QCM-D data illustrating the formation of the supported lipid bilayer (SLB) and subsequent ssDNA interactions. The frequency (Δf) and dissipation (ΔD) shifts are shown as a function of time for the SiO_2 sensor surface. The initial decrease in frequency and dissipation corresponds to the adsorption and fusion of POPE:POPS:POPC vesicles, forming a stable bilayer. Subsequent injections of 100 nM, 50 nM, and 10 nM ssDNA result in concentration-dependent decreases in frequency and increases in dissipation, reflecting DNA binding and layer viscoelasticity. Washing steps with HEPES buffer containing 1 mM EDTA reverse the binding, restoring the system toward baseline. This figure highlights the dynamic and reversible interactions between DNA and the SLB.

robust binding and the formation of a thick, viscoelastic layer on the surface (see Figure 3.6). In contrast, the introduction of 50 nM ssDNA resulted in a smaller response, consistent with partial surface saturation, while 10 nM ssDNA elicited the smallest changes, reflecting minimal binding due to reduced DNA availability.

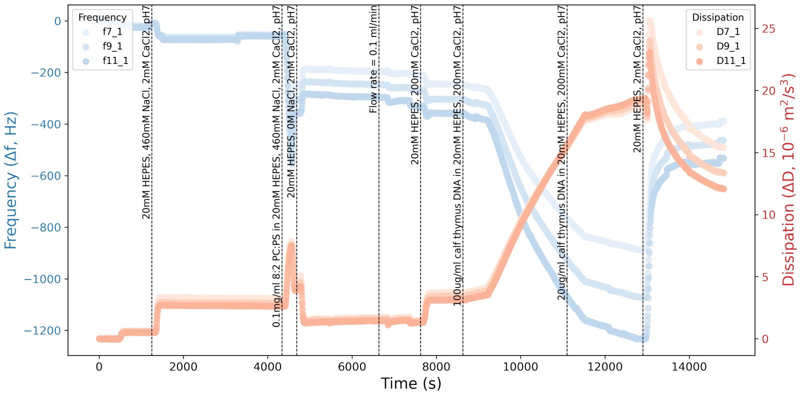
The reversibility of the DNA attachment was assessed by washing the system with 20 mM HEPES buffer containing 1 mM EDTA. The chelation of calcium ions by EDTA disrupted the interactions between DNA and the bilayer, leading to mass removal from the surface and a return of frequency and dissipation values toward baseline levels. This recovery was consistent across all DNA concentrations, demonstrating the system's ability to reverse the binding process.

These results highlight the successful formation of a supported lipid bilayer and its utility as a platform for studying DNA interactions. The observed concentration-dependent binding dynamics and the reversibility of DNA attachment provide valuable insights into the biophysical behavior or membranes and the effect of DNA binding via a cation mediated mechanism. These findings have implications for applications requiring controlled binding and reversible interactions on biomimetic surfaces.

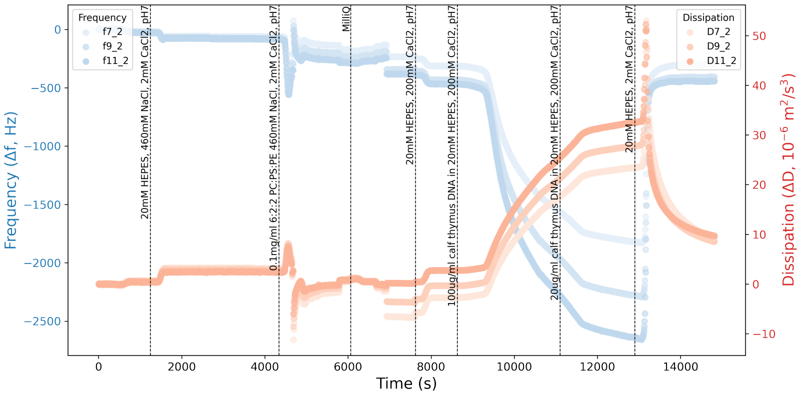
3.3.4 Lipid Composition Effects

The experiment investigated the dynamics of bilayer formation and subsequent interactions with calf thymus DNA using SiO_2 sensors under controlled conditions. Sensor 1, Sensor 2, and Sensor 3 data provided complementary insights into the system's response to buffer changes, lipid deposition, osmotic shock, and DNA addition/removal. Measurements of frequency shifts (Δf) and dissipation changes (ΔD) at multiple harmonics were employed to monitor mass adsorption, bilayer formation, and interaction events.

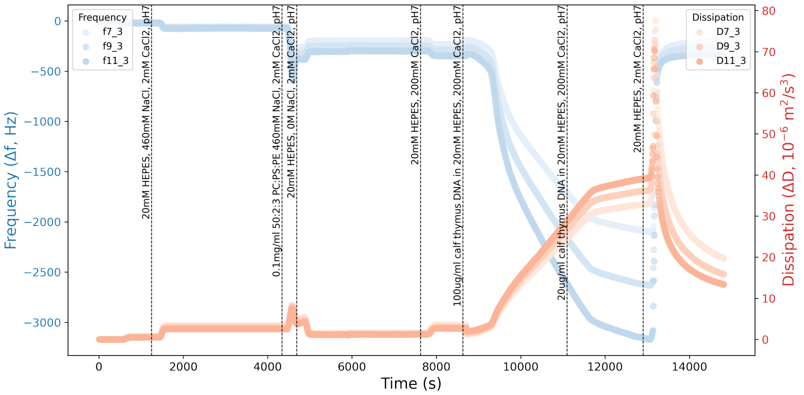
Initial measurements established baselines in MilliQ water for all sensors, followed by the introduction of a high-salt buffer (20 mM HEPES, 460 mM NaCl, 2 mM $CaCl_2$, pH 7). The subsequent addition of lipid vesicles (6:2:2 PC:PS:PE for Sensor 2, 8:2 PC:PS for Sensor 1, and 50:2:3 PC:PS:PE for Sensor 3) resulted in significant frequency decreases (Δf) and dissipation changes (ΔD) across multiple harmonics, indicative of vesicle adsorption. Osmotic shock was induced by switching to a low-salt buffer (20 mM HEPES, 0 mM NaCl, 2 mM$CaCl_2$, pH 7). This step ruptured the lipid vesicles, facilitating the transition from vesicle adsorption to bilayer formation. The release of lipids and their subsequent reorganization into bilayers was evidenced by an increase in dissipation (ΔD) and partial recovery of frequency (Δf). The differences in bilayer formation can be attributed to the distinct characteristics of the lipids used. POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), present in all sensors, is known for its zwitterionic, neutral charge, which contributes to bilayer stability and fluidity. POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine), present in all sensors, introduces a negative charge, facilitating interactions with divalent cations like calcium and enhancing the adsorption of charged molecules. POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine), more prevalent in Sensor 3, is



(a) Minimal bilayer reorganization and stable formation with 8:2 PC:PS.



(b) Intermediate responses with a balance of fluidity and rigidity with 6:2:2 PC:PS:PE.



(c) Significant reorganization and DNA binding with 50:2:3 PC:PS:PE.

Figure 3.7: Frequency and dissipation changes during lipid deposition, osmotic shock, and DNA interactions. Sensor 1 shows stability, Sensor 2 exhibits enhanced reorganization, and Sensor 3 demonstrates balanced bilayer properties influenced by lipid composition.

known for its small headgroup and hydrogen-bonding capabilities, which promote bilayer rigidity and influence the bilayer's mechanical properties.

The differences in osmotic response can be further linked to lipid composition. Sensor 2, with a higher proportion of PC and PE, showed the most significant dissipation increase, indicative of greater bilayer reorganization. Sensor 3, with its balance of POPC, POPS, and POPE, exhibited the largest frequency changes, particularly at higher harmonics, indicating strong DNA adsorption and bilayer perturbation. Sensor 1, dominated by POPC and POPS, displayed minimal changes, suggesting a more stable bilayer structure.

Following bilayer stabilization, calf thymus DNA was introduced at two concentrations: 100 μg/mL and 20 μg/mL, in a high-$CaCl_2$ buffer (20 mM HEPES, 200 mM $CaCl_2$, pH 7). The initial injection of DNA ($100\,\mu g\,mL^{-1}$) produced a pronounced frequency shift and dissipation change ($\Delta f < 0$, $\Delta D > 0$), indicating significant mass adsorption and a corresponding viscoelastic change in the bilayer. This response is consistent with DNA binding to the bilayer, potentially mediated by calcium bridging between DNA and negatively charged lipids. Subsequent rinsing steps with buffer (20 mM HEPES, 2 mM $CaCl_2$, pH 7) partially reversed these changes ($\Delta f > 0$, $\Delta D < 0$), reflecting the removal of loosely bound DNA. The lower DNA concentration (20 μg/mL) showed a smaller but qualitatively similar response, supporting a concentration-dependent binding mechanism.

Comparing the responses from the three sensors, several distinctions are apparent (see Table 3.1). Sensor 1 (8:2 PC:PS) exhibited the smallest frequency and dissipation changes, reflecting a stable bilayer with minimal reorganization or DNA adsorption (see Figure 3.7a). Sensor 2 (6:2:2 PC:PS:PE) showed moderate frequency and dissipation changes, highlighting enhanced bilayer fluidity and better DNA interaction due to PE (see Figure 3.7b). Sensor 3 (50:2:3 PC:PS:PE) displayed the largest frequency changes, particularly at higher harmonics, and moderate dissipation changes, indicating strong DNA adsorption and significant bilayer perturbation driven by POPE's influence (see Figure 3.7c).

Table 3.1: Summary of key findings from QCM-D measurements across different lipid compositions.

Sensor	**PC:PS:PE**	**Bilayer Formation (Osmotic Shock)**	**Bilayer Properties**	**DNA Binding (100 μg/mL)**	**DNA Removal (Rinse)**	**Key Observations**
1	8:2:-	Minimal $\Delta f/\Delta D$ changes; stable bilayer	Stable, low reorganization	Weak: small $\Delta f \downarrow$, $\Delta D \uparrow$	Largely reversible	Highly stable bilayer; minimal DNA interaction
2	6:2:2	Largest dissipation increase; strong reorganization	Intermediate fluidity and rigidity	Moderate $\Delta f \downarrow$, $\Delta D \uparrow$	Partially reversible	PE promotes flexibility and moderate DNA interaction
3	5:2:3	Moderate dissipation; largest Δf shifts at higher harmonics	Viscoelastic, sensitive to perturbation	Strongest $\Delta f \downarrow$, moderate $\Delta D \uparrow$	Least reversible	POPE enhances DNA adsorption and bilayer perturbation

During osmotic shock, Sensor 2 showed the most pronounced dissipation increase, reflecting significant bilayer reorganisation, while Sensor 3 exhibited strong frequency changes, demonstrating its sensitivity to DNA interactions. Sensor 1 displayed minimal changes, underscoring its stability. These differences underscore the critical role of lipid composition in modulating bilayer properties and interactions with nucleic acids. DNA removal during rinsing was more effective in Sensor 1, further supporting its lower binding affinity, while Sensors 2 and 3 retained more DNA, highlighting the influence of bilayer composition on binding strength.

The osmotic shock appeared to enhance bilayer fluidity, as evidenced by the dissipation dynamics during low-salt buffer exposure. This step likely facilitated subsequent DNA binding by optimising the bilayer's electrostatic and structural properties. DNA adsorption was characterized by a rapid frequency decrease and increased dissipation, suggesting the formation of a hydrated, viscoelastic layer. DNA removal was incomplete during rinsing, indicating strong binding affinity under high-$CaCl_2$ conditions. The observed concentration dependence of DNA binding supports the hypothesis of multivalent interactions between DNA and the bilayer. These findings highlight the interplay between osmotic conditions, bilayer structure, and DNA-binding dynamics, providing insights into the molecular mechanisms governing nucleic acid-lipid interactions.

3.3.5 Cholesterol Effects

The experimental investigation on Sensor S4 aimed to form a supported lipid bilayer (SLB) and study its interactions with DNA. This was achieved using a natural liver extract (0.1 mg/mL) in a buffer containing 20mM HEPES, 460 mM NaCl and 2 mM $CaCl_2$ at pH 7. The SLB formation was monitored using QCM-D, which showed characteristic changes in frequency and dissipation (see Figure 3.8). These observations suggested the adsorption of lipid vesicles onto the SiO_2 sensor surface, followed by their rupture and fusion into a continuous bilayer. Unlike Sensors S1-S3, which employed isolated lipid mixtures, Sensor S4 used a total lipid extract. This extract included cholesterol and a more diverse lipid composition, reflecting the natural heterogeneity of biological membranes. The presence of cholesterol likely contributed to differences in bilayer properties, such as increased rigidity and altered fluidity, as compared to bilayers formed on S1-S3.

The process was marked by a substantial decrease in frequency, indicative of mass addition on the sensor surface, and a transient increase in dissipation, signifying initial viscoelastic properties. Over time, the dissipation values stabilised, indicating the formation of a rigid and well-ordered bilayer. The presence of calcium ions was critical in this context, as they mitigated electrostatic repulsion between lipid vesicles and facilitated vesicle fusion by neutralizing the negative charges on the lipid headgroups. Subsequent to SLB formation, calf thymus DNA was introduced to study its binding behaviour. The addition of 100 g/mL DNA resulted in the most pronounced frequency decrease and dissipation increase on S4, signaling the formation of a dense,

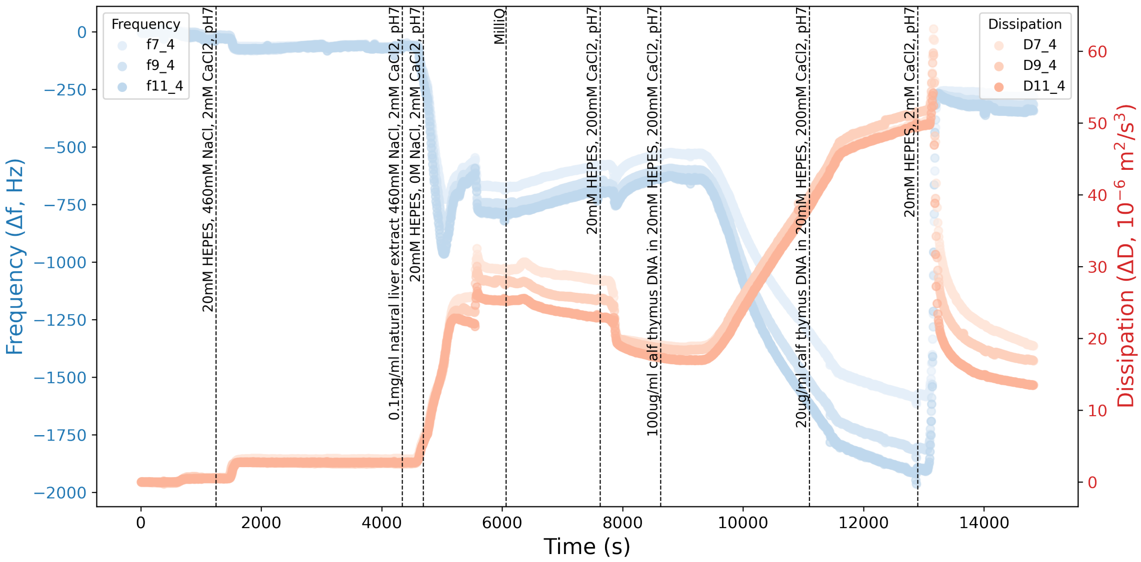


Figure 3.8: QCM-D data for Sensor S4 showing frequency and dissipation changes during SLB formation with a natural lipid extract and subsequent DNA binding. The presence of cholesterol and diverse lipid composition contributed to a rigid bilayer and enhanced viscoelastic properties during calcium-mediated DNA interactions.

viscoelastic layer on the SLB. Lower concentrations of DNA (20 g/mL) elicited a less pronounced response, consistent with partial coverage of the bilayer surface. These concentration-dependent changes highlighted the role of electrostatic interactions and calcium bridging in facilitating DNA binding.

Compared to S1-S3, Sensor S4 exhibited elevated baseline noise relative to the other sensors, particularly S1 and S3, which could have influenced the precision of measurements during the bilayer formation and DNA binding phases. The use of a total lipid extract on S4, containing cholesterol and a diverse lipid profile, introduced differences in bilayer properties, such as increased rigidity and altered interaction dynamics with DNA, compared to the isolated lipid mixtures used in S1-S3. Specifically, the presence of cholesterol likely increased the bilayer's packing density, which may have enhanced calcium-mediated bridging interactions between the bilayer and DNA. While S4 exhibited significant frequency decreases and dissipation increases during lipid deposition and DNA binding, the magnitude of these changes was generally comparable to those observed for S1-S3. However, the dissipation changes in S4 suggest a slightly more viscoelastic layer, likely due to cholesterol's contribution to lipid heterogeneity. The dissipation stabilization on S4 following SLB formation was similar to that of S1-S3, suggesting that despite the initial noise, a rigid and well-ordered bilayer was achieved. Cholesterol may have contributed to this rigidity, further stabilizing the bilayer. The responses to DNA concentrations on S4 aligned well with the trends observed on S1-S3. However, the slightly higher dissipation values during DNA binding on S4 may reflect the presence of cholesterol, which could promote a more viscoelastic interaction layer by stabilising calcium-mediated binding sites. These observations validate the utility of Sensor S4 for studies on lipid bilayers and DNA interactions, despite its noisier baseline. The presence of cholesterol and a diverse lipid composition on S4 provided additional insights into the role of natural lipid heterogeneity in biomimetic systems, particularly in enhancing the stability and dynamics of calcium-mediated DNA interactions.

3.3.6 Surface protein effects

The experiments utilizing virus-like particles (VLPs) produced in HEK293 cells revealed significant insights into the role of surface proteins, lipid heterogeneity, and environmental pH on DNA interactions. Unlike artificial lipid bilayers, which lack the complexity of biological membranes, or the total liver lipid extract, which lacks surface proteins, the VLPs employed in this study incorporate cholesterol, diverse lipid species, and a range of surface proteins, closely mimicking native viral envelopes.

The QCM-D analysis demonstrated distinct binding dynamics across experimental phases. The largest changes in dissipation and frequency were observed at pH 6, followed by pH 8, with the smallest responses recorded at pH 4 (see Figure 3.9). At pH 6, the increased dissipation indicates the formation of a highly viscoelastic layer, potentially driven by optimal conformational states of surface proteins that facilitate stronger interactions with DNA. The intermediate behavior at pH 8 is consistent with robust but less dynamic binding, where calcium-mediated electrostatic bridging remains a dominant mechanism. At pH 4, the binding interactions diminished, as evidenced by smaller changes in frequency and dissipation. This is likely due to the disruption of calcium-mediated bridges and conformational alterations in surface proteins, reducing their capacity to mediate DNA attachment. The role of surface proteins in these interactions is critical. At pH 6, surface proteins may adopt configurations that maximize electrostatic complementarity and hydrophobic interactions, promoting enhanced DNA binding. In contrast, the reduced interaction strength at pH 8 reflects a less optimized state for protein-mediated binding. The highly acidic conditions at pH 4 appear to destabilize protein structure and interfere with lipid-mediated interactions, further diminishing DNA attachment. These findings emphasize the interplay between surface protein dynamics, calcium bridging, and environmental pH in regulating DNA interactions on VLPs.

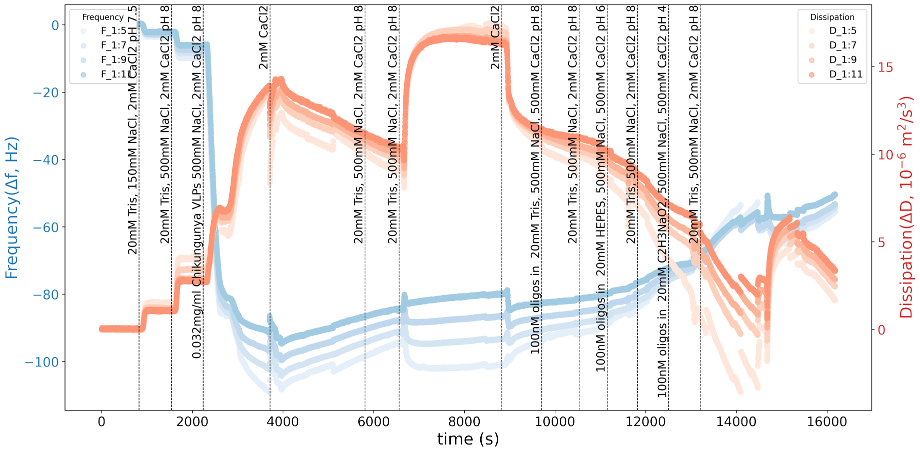


Figure 3.9: QCM-D analysis of DNA binding to a viral supported bilayer (Chikungunya VLPs) under varying pH conditions. The largest changes in frequency and dissipation were observed at pH 6, indicating the formation of a highly viscoelastic layer. Moderate responses at pH 8 suggest robust but less dynamic binding, while minimal changes at pH 4 reflect disrupted calcium-mediated interactions and reduced protein-lipid activity.

In comparison to artificial lipid bilayers and total liver extracts, the presence of surface proteins on VLPs introduces additional complexity and functionality. While artificial bilayers and liver extracts are primarily influenced by lipid composition and cholesterol content, the inclusion of surface proteins on VLPs enhances their capacity for selective and dynamic interactions. Surface proteins likely contribute to the observed differences in binding behavior across pH conditions, with their conformational states playing a pivotal role at pH 6 in facilitating enhanced interaction dynamics. This added complexity makes VLPs a more representative model for studying biomolecular interactions on native viral surfaces.

These findings provide a nuanced understanding of how VLPs, with their protein-lipid heterogeneity, behave across different pH environments. The results highlight the critical role of surface proteins in modulating DNA interactions, particularly under mildly acidic conditions, and offer valuable insights into the interplay between membrane composition, protein structure, and environmental factors in biomimetic systems and their interactions with DNA.

3.4 Discussion

This chapter explored the cation-mediated labeling of viruses using fluorophore-modified single-stranded DNA (ssDNA), focusing on the effects of ssDNA length, pH, and calcium ion (Ca^{2+}) concentration. Building on the foundational work by Robb et al. [179], we refined the understanding of the labeling mechanism and its dependence on these parameters, while identifying areas for further investigation and application.

3.4.1 Role of ssDNA Length, Ca^{2+} Concentration, and pH

Our findings confirmed that ssDNA length is a critical determinant of labeling efficiency. The 8mer was insufficient for stable interactions at pH 8 across all tested Ca^{2+} concentrations, whereas the 14mer, 20mer, and 40mer exhibited robust labeling, with track counts peaking at 1 M Ca^{2+}. The 40mer displayed superior performance, retaining significant track counts even as Ca^{2+} concentrations decreased, highlighting the benefits of longer ssDNA. The results also demonstrated that higher pH enhances labeling efficiency, with pH 9 promoting stronger cation-mediated interactions due to the increased negative charge on the virion surface. This was particularly evident in the improved performance of the 14mer and the detection of nonzero photocounts for the 8mer at pH 9, which were absent at pH 8.

The photocount analysis further supported these observations. The 40mer exhibited significantly higher photocounts at pH 9 than at pH 8, reflecting its enhanced labeling efficiency under alkaline conditions. The 20mer and 14mer maintained moderate photocounts at both pH values, while the 8mer was detectable only at pH 9. These results underscore the importance of optimizing pH and Ca^{2+} concentration for efficient ssDNA labeling, particularly for shorter constructs.

3.4.2 Insights from Viral Supported Bilayers

The viral supported bilayers produced in this study offer valuable insights not only for ssDNA labeling but also as model systems for studying the viral envelope. These bilayers, derived from virus-like particles (VLPs), mimic key features of viral membranes, including lipid composition and the presence of surface proteins. They provide a planar, biomimetic platform for probing virus-membrane interactions under controlled conditions, enabling investigations into the effects of pH, ionic strength, and binding agents on membrane dynamics.

While these supported bilayers are powerful tools, it is important to note that they represent a planar system, unlike the spherical geometry of actual virions. This structural difference is a limitation for the QCM-D experiments, as it may result in deviations from the natural curvature and spatial organization of viral envelopes. These differences could influence the binding dynamics of ssDNA and other interacting molecules. Future studies could explore methods to incorporate curvature into such systems, for instance, by using curved substrates or intact VLPs.

3.4.3 Mechanistic Insights

The use of poly-thymine ssDNA ensured that the observed interactions were not influenced by secondary structure, confirming that the labeling mechanism is driven by electrostatic and cation-mediated interactions rather than structural features of the DNA. However, the role of secondary structure in ssDNA labeling remains unexplored and warrants further investigation. For example, sequences capable of forming hairpins or other secondary structures may exhibit altered binding properties due to changes in effective length, charge distribution, or rigidity.

3.4.4 Future Directions

This work provides a solid foundation for extending the study of cation-mediated labeling to new contexts and applications. Several avenues for future research are as follows:

- **Testing a Broader Range of Viruses:** While this study focused on WSN virions, it would be valuable to test additional enveloped viruses to assess the universality of the cation-mediated labeling mechanism. Understanding how differences in lipid composition, surface protein content, and envelope architecture affect labeling efficiency could expand the utility of this method.

- **Exploring the Role of ssDNA Secondary Structure:** Future experiments could investigate how sequences capable of forming hairpins or other structures impact labeling efficiency. This could reveal whether structural properties influence binding strength, stability, or fluorescence signal.

- **Compatibility with Other Labeling Methods:** Investigating the compatibility of cation-mediated labeling with established techniques, such as fluorescence in situ hybridization (FISH) assays, could open new possibilities for multiplexed or complementary labeling strategies.

- **Optimizing for Specificity:** While this technique labels enveloped viruses indiscriminately, integrating it with additional functionalization methods or detection assays could increase specificity. For instance, combining this approach with sequence-specific hybridization or antibody-based methods might enable selective labeling of target viruses in complex samples.

- **Improving Model Systems:** Developing systems that better mimic the curvature of native viral membranes would address the limitations of planar supported bilayers. This could include using curved substrates or intact VLPs to provide a more representative model for studying virus-envelope interactions.

3.4.5 Broader Applications

The insights gained from this study have broad implications for the development of advanced diagnostic tools and experimental techniques. By refining the cation-mediated labeling process and identifying its key parameters, this work sets the stage for more efficient and robust methods for detecting and studying enveloped viruses. Additionally, the use of viral supported bilayers as model systems offers a versatile platform for probing virus-membrane interactions and could inform the design of novel therapeutics and diagnostics.

3.4.6 Conclusion

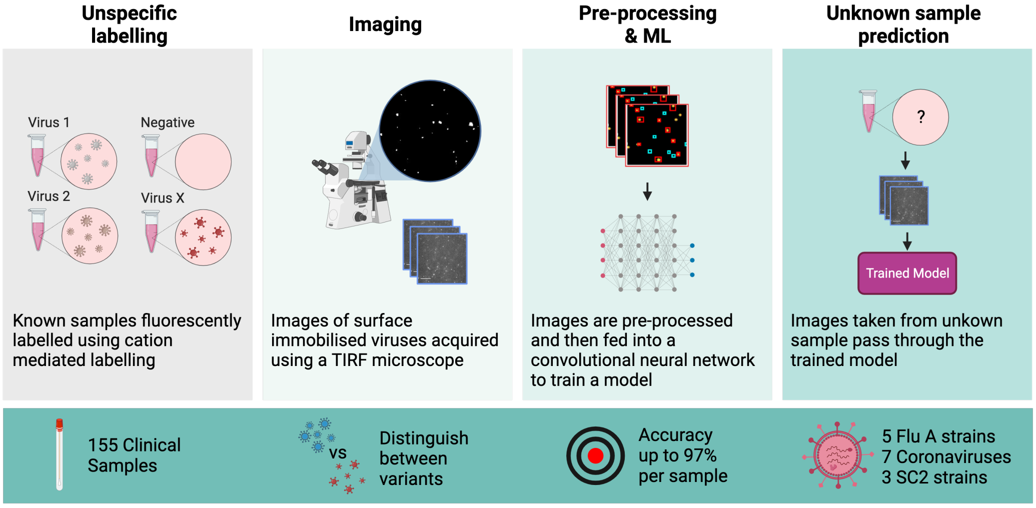
This chapter demonstrates the critical roles of ssDNA length, pH, and Ca^{2+} concentration in cation-mediated virus labeling. By extending the foundational work and introducing systematic variations in these parameters, we have provided a deeper understanding of this mechanism and identified promising directions for future research and application.

Chapter 4

Detection of Immobilised Virions in Minutes using Deep Learning

> *"Great things are done by a series of small things brought together."*
>
> Vincent Van Gogh

4.1 Graphical Abstract



Known samples fluorescently labelled using cation mediated labelling

Images of surface immobilised viruses acquired using a TIRF microscope

Images are pre-processed and then fed into a convolutional neural network to train a model

Images taken from unkown sample pass through the trained model

Figure 4.1: Cation-mediated fluorescent labelling of viruses enables TIRF imaging and machine learning analysis: Known samples are labelled and imaged, pre-processed images train a convolutional neural network, which then predicts the identity of viruses in unknown samples. *Figure created with BioRender.com. Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

4.2 Introduction

The SARS-COV-2 virus has affected a vast number of individuals since its appearance, leading to significant fatalities and causing global upheaval. The rise of multiple strains that impact transmission rates, disease severity, or vaccine effectiveness has further challenged containment efforts.

Common diagnostic methods for this virus encompass tests that detect its genetic material, its presence through antigens, and the antibodies produced in response to it. The widely recognized standard for diagnosis involves a specific genetic amplification technique. While accurate, this method can be time-consuming, necessitates specialized facilities, and can face supply constraints. Alternative amplification techniques offer quicker results without the need for complex equipment but still face similar logistical challenges. Rapid tests using colorimetric labels for immediate virus detection are available but might compromise on accuracy. Identifying specific strains primarily depends on sequencing the virus's genetic material. Given these challenges, there's a pressing demand for innovative detection techniques, especially those suitable for field deployment.

In prior research, we introduced an efficient technique to swiftly label virus particles using a combination of divalent cations, short DNA sequences, and a particle with a specific charge. This method allowed for the creation of brightly fluorescent particles for any enveloped virus examined. Our extensive evaluations confirmed the specificity and reliability of this labeling approach.

To meet the demand for innovative detection methods, we've harnessed this labeling technique to devise a diagnostic test based on fluorescence imaging. The process begins with rapid fluorescence labeling of viruses in a specimen, followed by immobilizing these labeled particles for imaging. Subsequent image analysis and machine learning techniques are employed to discern different viruses in various samples. This strategy capitalizes on the unique characteristics of different virus types and strains, which influence the distribution and density of fluorescent markers on their surfaces. Advanced computational models, like convolutional neural networks (CNNs), have proven effective in classifying and detecting virus particles based on these distinct features.

4.3 Methods and Materials

4.3.1 Lab-grown virus Strains

The study utilized several influenza strains, naamely H1N1 A/Puerto Rico/8/1934 (PR8), H3N2 A/Udorn/72 (Udorn), H1N1 A/WSN/33 (WSN), and H3N2 A/Aichi/68 (X31). The strains WSN, PR8, and Udorn were cultivated in either Madin–Darby bovine kidney (MDBK) or Madin–Darby canine kidney (MDCK) cells. In contrast, X31 was propagated in embryonated chicken eggs. After cultivation, either the cell culture supernatant or allantoic fluid was harvested, and the viral concentrations were determined using a plaque assay. The recorded titers for PR8, Udorn, WSN, and X31 were $1\times10^{8}\,\mathrm{PFU\,mL^{-1}}$, $1\times10^{7}\,\mathrm{PFU\,mL^{-1}}$, $2\times10^{6}\,\mathrm{PFU\,mL^{-1}}$, and $4.5\times10^{8}\,\mathrm{PFU\,mL^{-1}}$, respectively. Additionally, the coronavirus IBV (Beau-R strain) [180] was cultivated in embryonated chicken eggs and had a titer of $1 \times 10^{6}\,\mathrm{PFU\,mL^{-1}}$ as determined by a plaque assay. Both influenza and IBV were rendered inactive with a 2% (w/v) formaldehyde addition prior to their application in the study. As for SARS-CoV-2, it was cultivated in Vero E6 cells, with a titer of $1.05 \times 10^{6}\,\mathrm{PFU\,mL^{-1}}$, and was inactivated using a 4% (w/v) formaldehyde solution before its utilization.

4.3.2 Single Stranded DNAs

We procured single-stranded oligonucleotides modified with either “red” or “green” dyes from IBA, based in Germany. When choosing oligonucleotides, our primary considerations were their length (ensuring they exceeded 20 bases) and the type of fluorophore modification (prioritizing bright and enduring dyes). The specific sequence of the oligonucleotides was of secondary importance, as effective labeling was observed irrespective of sequence, provided the aforementioned conditions were satisfied. The DNA labeled with a “red” hue in our study was modified at its $5'$ end with ATTO647N, having the sequence 5’ACAGCACCACAGACCACCCGCGGAT

GCCGGTCCCTACGCGTCGCTGTCACGC. Conversely, the “green” labeled DNA was altered at its $3'$ end with Cy3, with the sequence 5’GGGTTTGGGTTGGGTTG GGTTGGGTTTTGGGTTTGGGTTGGGTTGGGAAAAA3’. For super-resolution imaging purposes, we used DNA modified at the $5'$ end with Alexa647, bearing the sequence 5’TCCGCTCTCACAATTCCACACATTATACGAGCCGAAGCATAAA GTGTCAAGCCT3’.

4.3.3 Sample Preparation

The preparation process for both positive and negative samples was identical. For instance, they were inactivated using the same formaldehyde concentration or labeled in a consistent buffer. Only samples that underwent identical preparation were analysed in comparison to each other. Glass slides underwent a treatment process with $0.015 mg/mL$ chitosan, a linear polysaccharide, in $0.1M$ acetic acid for a duration of 30 minutes. Subsequently, they were rinsed three times with Milli-Q water. Alternatively, they were treated with 0.01% poly-l-lysine (from Sigma) for 15 minutes. Although both agents introduced some background in the control samples, they facilitated swift immobilization of the virus, which is crucial to prevent virus clumping. Unless specified otherwise, virus stocks, usually about $10\mu L$, were diluted in either $0.23MCaCl_2$ or $SrCl_2$ (as detailed in the figure legends) combined with $1nM$ of each dye-labeled DNA, resulting in a total volume of $20\mu L$. This mixture was then applied to the slide. For the imaging of SARS-CoV-2, the cationic labeling solution was buffered using $20mM$ Tris at pH 8. The method of virus labeling with $CaCl_2$ has been documented earlier; $SrCl_2$ yields comparable outcomes. For lab-cultivated virus stocks, negative samples were produced using either virus-free Minimal Essential Media (from Gibco) or allantoic fluid sourced from uninfected eggs, substituting the virus.

4.3.4 Imaging

Three wide-field Nanoimager microscopes were utilized to capture the images. "Microscope 1" featured a Hamamatsu Flash 4 LT.1 sCMOS camera, while "Microscopes

2 and 3" were fitted with a Hamamatsu Flash4 V3 sCMOS camera. Apart from these distinctions, the systems were identical in their specifications. The imaging of the sample was conducted using total internal reflection fluorescence (TIRF) microscopy, combined with a 100x oil-immersion lens. The laser light was directed at an angle of 53° relative to the perpendicular. Short movies, consisting of 5 frames for each field of view (FOV) that spanned $75 \times 49 \mu m$, were recorded at a rate of $33 Hz$ with a $30 ms$ exposure time. The laser intensities were consistently set at $0.78 kW/cm^2$ for the red ($640 nm$) laser and $1.09 kW/cm^2$ for the green ($532 nm$) laser. To automate the process and ensure unbiased FOV selection, the entire sample underwent scanning using the microscope's multi-acquisition feature, capturing 81 FOVs within 2 minutes. An automated autofocus system meticulously managed defocusing, positioning the sample at a predetermined axial location before each FOV was illuminated by the excitation lasers. This positioning was accomplished by observing the reflection of a near-infrared laser on the glass/sample medium boundary and aligning the image with a previously stored reference image.

The data presented in Supplementary Figure 6 was gathered using a Zeiss Elyra 7 microscope, which was equipped with two pco.edge sCMOS (version 4.2 CL HS) cameras. TIRF images were captured using the alpha Plan-Apochromat $63 \times /1.46$ oil lens. For imaging Atto647N, a laser intensity of 10% from the $641 nm$ laser was employed. For capturing Cy3 images, laser intensities were set at 6% for the $561 nm$ and 3% for the $488 nm$ laser, with an exposure duration of $50 ms$.

4.3.5 Super-resolution Imaging

For capturing images at super-resolution, we prepared microscope slides with a passivation process. Initially, these slides were cleaned using acetone and then treated with a Vectabond solution (from Vector Laboratories). Subsequently, they were exposed to a mixture of NHS-PEG and Biotin-NHS-PEG (sourced from Laysan Bio) at a ratio of $80 : 1$. Just before adding the virus, the slide was treated with $0.5 mg/mL$ of neutravidin for a 10-minute duration at ambient temperature. The viruses underwent biotinylation by being treated with $1 mg/mL$ of Sulfo-NHS-LC-Biotin (sourced from

ThermoFisher) for 3 hours at a temperature of 37°C. Post this, they were labeled using 0.23 M $CaCl_2$ and $1nM$ of DNA tagged with Alexa647, resulting in a final volume of $20\mu L$, which was then applied to the slide.

After a 30-minute incubation at room temperature, the slide was rinsed three times using 1× PBS. The imaging process was conducted in a solution containing $50mM$ MEA and an enzyme-based oxygen scavenging system, which comprised $1mg/mL$ glucose oxidase, $40\mu g/mL$ catalase, and 1.0% (by weight/volume) glucose. The Nanoimager software's built-in features were used to extract super-resolution localizations, which were then further analyzed using Matlab. The DBScan method was employed to cluster these localizations, setting a minimum cluster size of 50 and an epsilon value of $30nm$. This was followed by calculating the convex hull to determine the area covered by these clustered points.

4.3.6 Data Segmentation

For each field of view (FOV) in the red channel, we converted it into a binary format using MATLAB's integrated imbinarize function, setting the adaptive filtering sensitivity to 0.5. This adaptive filtering technique determines if a pixel is part of the foreground or background based on the statistics of its surrounding pixels. The filter sensitivity is a parameter that, when increased, lowers the threshold for a pixel to be considered as foreground. To filter out free ssDNA and aggregates, we utilized the bwpropfilt function, excluding objects with an area not within the 10–100 pixels range (with 1 pixel equivalent to 117 nm). Through imaging of individual fluorophores, we observed that their area did not surpass 10 pixels, setting our lower boundary. We selected 100 pixels as the upper boundary to eliminate extensive aggregates or cellular remnants. The regionprops function was then applied to derive various attributes of each detected object, such as its area, the ratio of its semi-major to semi-minor axis, its center's coordinates, its encompassing bounding box (BBX), and the peak pixel intensity inside the BBX.

Each FOV is paired with a location image (LI) that provides an overview of the signal locations from both channels (red and green). In the LI image, signals that are colocalized are represented in yellow. Objects identified in the red FOV were cross-referenced with their corresponding signal in the related LI. Objects not originating from colocalized signals were discarded. The approved BBXs were subsequently superimposed onto the original FOV, and images containing the individual viruses within were preserved.

4.3.7 Deep Learning

We utilized the Convolutional Neural Network (CNN) with only the red channel as its input. Incorporating both channels didn't enhance the accuracy. While the images weren't normalized, the bounding boxes (BBXs) from data segmentation varied in size. However, they never exceeded 17 pixels in any direction due to size constraints. To standardize, all BBXs were padded (by adding zero-value pixels) to achieve a uniform size of 17×17 pixels.

These adjusted images served as input for our 15-layer CNN. This network was constructed using Matlab 2020b and underwent training on the computer's GPU, which boasted specifications like NVIDIA 2080Ti, 32 GB RAM, and an i7 processor. The architecture included three convolutional layers: the initial two had 2×2 kernels, while the last one utilized a 3×3 kernel. We fixed the learning rate at 0.01, maintaining a consistent learning schedule rate throughout. All hyperparameters, including a mini batch size of 1000, a maximum epoch count of 100, and a validation frequency of 20 iterations, remained unchanged for all models.

For classification, the trainNetwork function utilized the softmax function's values. It then allocated each input to one of the K distinct classes using the cross entropy function for a 1-of-K coding scheme, represented as:

$$loss = \sum_{i=1}^{N} \sum_{j=1}^{K} t_{ij} \ln(y_{ij}) \tag{4.1}$$

where N denotes the sample count, K represents the class count, t_{ij} indicates that sample i belongs to class j, and $\ln(y_{iy})$ is the output for the ith sample for the jth class. Essentially, it's the probability that the network correlates the ith input with the jth class. The optimization process employed a stochastic gradient descent with a momentum of 0.9.

4.3.8 Zeta Potential Measurements

All zeta potential measurements were conducted using the Zetasizer Nano S, employing disposable folded capillary cells (DTS1070). We maintained the temperature at 25 °C and allowed the system to equilibrate for 120 seconds. Depending on the pH levels, the samples were diluted in different solutions: 20 mM sodium acetate was used for pH 4–5, 20 mM HEPES for pH 6–7, and 20 mM Tris for pH 9. The buffer details for each pH level were input into the software to compute the viscosity and dielectric constant. For the type of material sample, a combination of 50% lipid and 50% protein was used to determine the absorption and refractive index. The software was set to "auto" mode for analysis. The software also decided the measurement duration, with a range between a minimum of 10 runs and a maximum of 100 runs. We recorded three measurements for each sample at every pH level.

4.3.9 Statistical Analysis

4.3.9.1 Confusion Matrices

Each network validation is represented through a confusion matrix, utilizing the following definitions:

- True Positive (TP): Correctly recognized positive bounding boxes (BBXs).
- False Positive (FP): Mistakenly recognized positive BBXs.
- True Negative (TN): Correctly recognized negative BBXs.
- False Negative (FN): Mistakenly recognized negative BBXs.

The term "Sensitivity" denotes the test's capability to accurately detect positive BBXs. It's derived from the ratio of true positives to the sum of true positives and false negatives:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

On the other hand, "Specificity" indicates the test's proficiency in accurately detecting negative BBXs. It's determined by the ratio of true negatives to the sum of true negatives and false positives:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4.3}$$

The trained model's accuracy in predicting BBXs is categorized into two metrics: the positive predictive value (PPV) represents the percentage of correctly predicted BBXs, while the negative predictive value (NPV) denotes the percentage of incorrectly predicted BBXs:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.4}$$

$$\text{NPV} = \frac{\text{TN}}{\text{FN} + \text{TN}} \tag{4.5}$$

The model's comprehensive validation accuracy, considering both positive and negative predictions, is expressed as:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{4.6}$$

4.3.9.2 Limit of Detection

To determine the limit of detection (LOD), we labeled and imaged escalating concentrations of IBV, diluted in allantoic fluid. After pre-processing the images, individual signals were input into the trained network. We then computed the normalized average of True Positives ($\text{TP}/\text{TP} + \text{FP}$) and its standard deviation (STD). These values were plotted against their respective concentrations in a scatter plot. This data was then fit into a linear regression model described by:

$$y = ax + b \tag{4.7}$$

In this equation, the virus concentration is the independent variable, and 'a' signifies the LOD. The final LOD value was determined using the equation $a + (3 * \text{STD}) = 6 * 10^4$ PFU/mL, which aligns with a 99.85% confidence interval under the assumption of a normal distribution. Experiments involving the influenza strain A/WSN/33 followed a similar procedure.

For determining the LOD of SARS-CoV-2, we used a similar approach, but the data was fit into a sigmoidal curve described by:

$$y = \frac{L}{1 + e^{-k(\text{Concentration} - X_0)}} + b \tag{4.8}$$

In this equation, L is the maximum normalized positive value (interpreted as the model's sensitivity), b introduces a bias to the output, adjusting its range from $[0, L]$ to $[b, L+b]$, k scales the input, and X_0 represents the point where the sigmoid outputs the value $L/2$.

4.4 Lab-Grown viruses

4.4.1 Efficient Labeling, Immobilization, and Visualization of Viral Particles

To illustrate our capability in labeling, immobilizing, and visualizing viral particles, we initially employed the infectious bronchitis virus (IBV), a known avian coronavirus (CoV). For labeling, IBV was treated with a divalent cation, specifically Sr2^+, which exhibited performance characteristics analogous to Ca2^+ (elaborated further below), in conjunction with a combination of green and red fluorescent DNAs (tagged with Cy3 or Atto647N fluorophores, see Figure 4.2 B). These labeled particles were subsequently immobilized on a chitosan-coated glass substrate and visualized through total-internal-reflection fluorescence microscopy (TIRF) (Figure 4.2 A). The fluorescence labeling was achieved expeditiously through a singular addition of the labeling mixture, followed by immediate immobilization of the viruses. The resultant images displayed particles with distinct green or red fluorescence emissions (represented as green and red particles), as well as those with colocalized green and red fluorescence

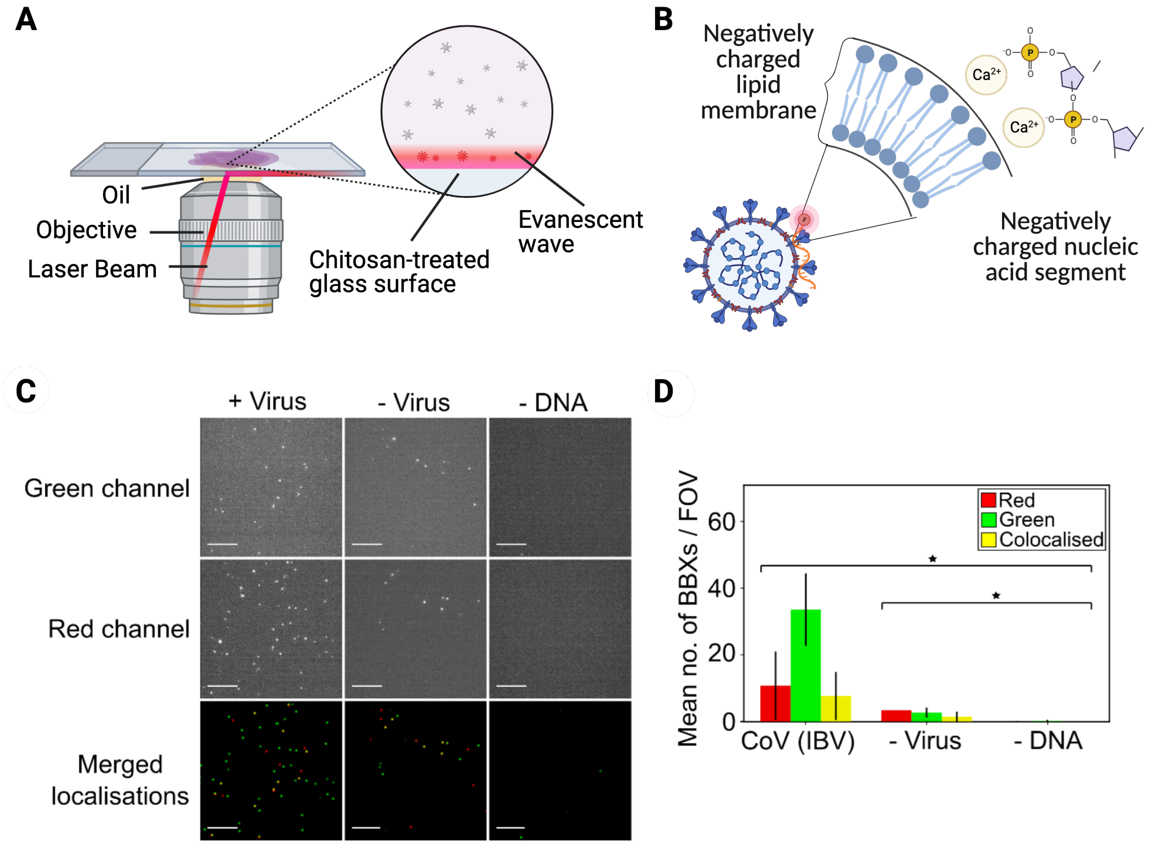


Figure 4.2: **Schematic of virus labelling and immobilisation strategy.** **A.** Labelled viruses were immobilised on a chitosan-coated glass slide and illuminated with red and green laser light on a widefield. Total Internal Reflection Fluorescence (TIRF) microscope. **B.** Positively charged cations (i.e. calcium, strontium) bridge the lipid membrane of the virus and the negatively charged phosphate groups on the ssDNA, binding fluorescently labelled ssDNA to the surface of the virus. **C** Representative fields of view (FOVs) of fluorescently labelled infectious bronchitis virus (CoV (IBV)). The virus sample, at a final concentration of $1 * 10^4$ PFU/mL, was immobilized and labelled with $0.23nM$ $CaCl_2$, $1nM$ Cy3 (green) DNA and $1nM$ Atto647N (red) DNA before being imaged. Green DNA was observed in the green channel (top panels) and red DNA in the red channel (middle panels); merged red and green localisations are shown in the lower panels. Scale bar $10\mu m$. A negative control where DNA was replaced with water is included. **D** Plot showing the mean number of BBXs per FOV for labelled CoV (IBV) and the negative controls. Error bars represent the standard deviation of 81 FOVs from one slide. Statistical significance was determined by one-way ANOVA, *P<0.0001. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

emissions (depicted as yellow particles) (Figure 1B–D). Effective viral labeling was realized using either CaCl2 (Figure 4.2 C-D) or $SrCl_2$ (Figure 4.3 b). Nonetheless, both solutions yielded a certain number of colocalised signals in the virus-negative controls, potentially due to stochastic coincidences or cation-mediated aggregation of DNAs on the substrate. Exclusion of DNAs culminated in a complete abrogation of the fluorescent signal (Figure 4.2 C, right panels).

4.4.2 Image Segmentation

Before their use in machine learning algorithms, viral images were pre-processed to separate individual image signals into bounded boxes (BBX) by segmenting the field of view (FOV) using adaptive filtering (Figure 1E). Employing a BBX-centric approach over full FOVs for classification is advantageous as it negates potential artifacts arising from variations in viral concentration or disparities in background and illumination patterns. The pristine FOVs from the red channel (Figure 4.3 1E-i) were transmuted into a binary format (Figure 4.3 1E-ii), and area filtering was employed to exclude entities with a cumulative area (i.e., width × length) less than 10 pixels ($1170nm$; singular fluorophores) or exceeding 100 pixels ($11700nm$; aggregates or cellular fragments) (Figure 4.3 1E-iii). Concurrently, to augment our viral sampling and exclude non-viral signals, the location image (depicting the green, red, and yellow signals from both channels; Figure 4.3 1E-iv) was utilized to discern colocalized signals (Figure 4.3 1E-v). This data, amalgamated with the signals discerned in the filtered binary image (Figure 4.3 1E-iii), facilitated the rejection of signals that did not satisfy the colocalization criterion (Figure 4.3 1E-vi; cyan boxes) and the retention of those that did (Figure 4.3 1E-vi–vii; red boxes).

The segmentation process was entirely automated, with each FOV being processed in approximately 2 seconds. In this experimental setup, the average count of colocalized BBXs per FOV in the presence of IBV was approximately six times greater than in its absence (Figure 4.3 1F), corroborating that our preprocessing step is adept at capturing virus-specific images.

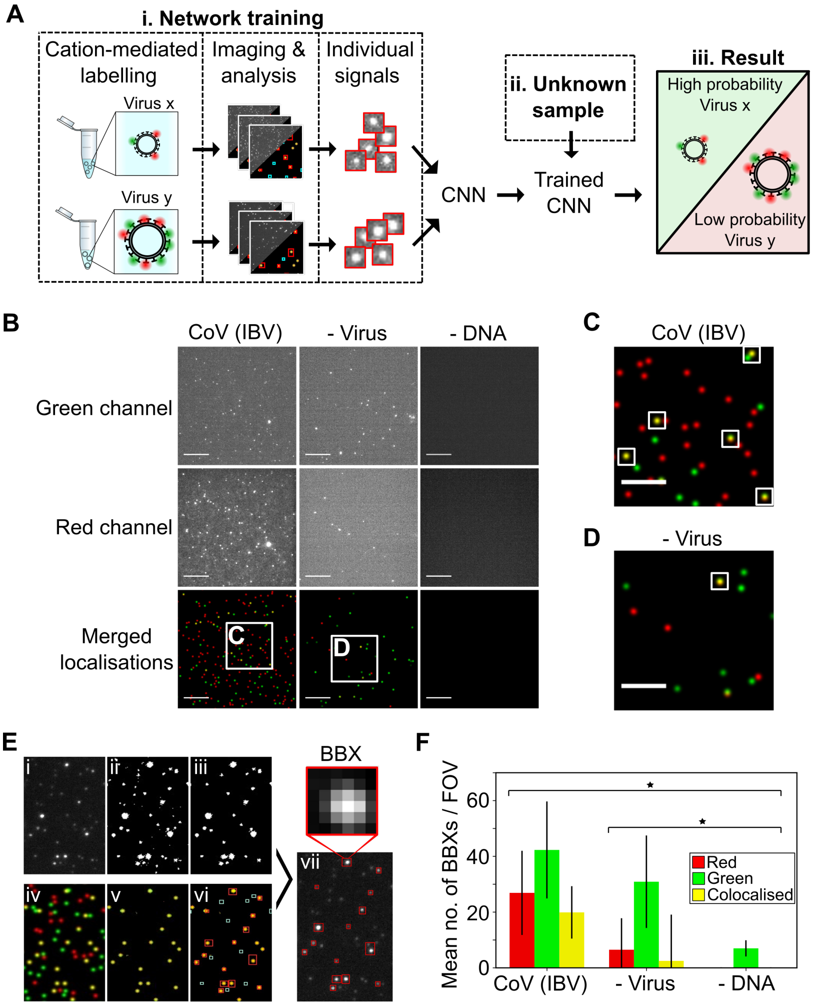


Figure 4.3: **Fluorescent labeling and imaging strategy to detect viruses. A.** Process Outline: (i) Viruses were labeled and imaged. Signals were segregated, and a CNN was trained to identify viruses using their unique attributes. (ii) Signals from unknown samples were input into the pre-trained CNN, enabling (iii) virus classification. **B.** Fields of view (FOVs) for infectious bronchitis virus (CoV (IBV)) at 1×10^4 PFU/mL, labeled with 0.23M $SrCl_2$, 1nM Cy3 (green) DNA, and 1nM Atto647N (red) DNA. Green and red DNA were visualized separately (top and middle panels), with combined localizations shown below. Scale bar: 10μm. A water control (no DNA) is also shown. **C & D.** Magnified views from B; white rectangles indicate co-localizing particles. Scale bar: 5μm. **E.** Segmentation: (i) Red-spectrum FOV. (ii) Intensity filtering produces a binary image. (iii) Area filtering retains 10–100 pixel entities. (iv) Corresponding spatial image. (v) Co-localized signals in the spatial image. (vi) Bounding boxes (BBXs) from (iii) overlaid on (v); cyan objects lacking co-localization are discarded. (vii) Co-localized entities (red) overlaid on (i). Scale bar: 10μm. **F.** Average BBX sizes for labeled CoV (IBV) vs. controls. Error bars: standard deviation across 81 FOVs from one slide. Statistical significance: one-way ANOVA, $*P = 6.01 \times 10^{˘22}$. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

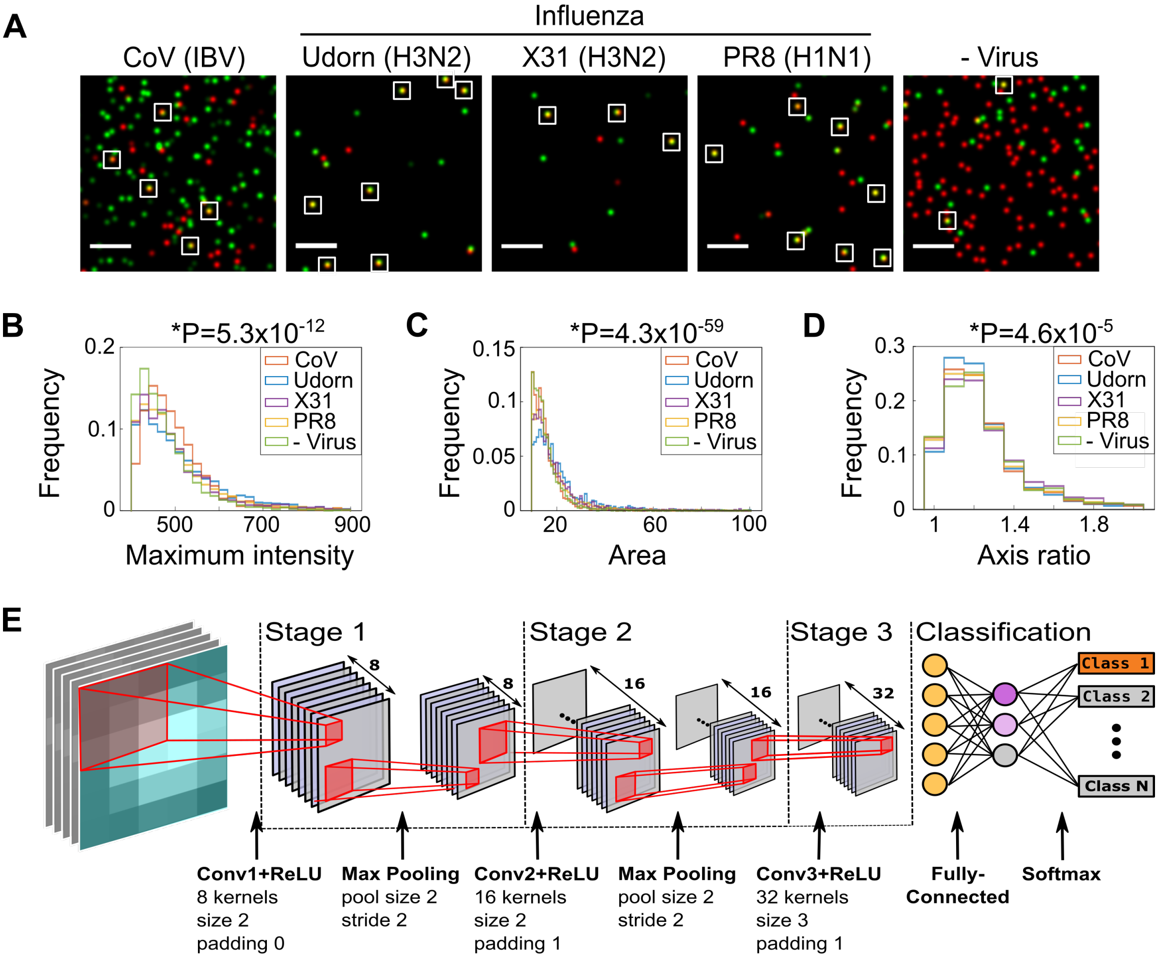


Figure 4.4: **Design of a CNN to classify imaged viruses.** **A.** Fields of view (FOVs) for CoV (IBV), H3N2 variants (Udorn, X31), H1N1 (PR8), and a virus-free control. Samples were tagged with 0.23 M SrCl$_2$, 1 nM Cy3 (green) and Atto647N (red) DNA. Colocalizations are highlighted (white rectangles). Scale bar: 10 μm. **B–D.** Frequency plots of peak pixel brightness, area, and axis ratio for 81 FOVs per virus. P-values from one-way ANOVA. **E.** 15-layer CNN: three convolutional stages, followed by a fully connected layer and softmax for classification. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

4.4.3 Positive and Negative Virus Images Can Be Distinguished Using Deep Learning

Having successfully imaged viral samples and segmented the resultant signals into BBXs, I theorized that a convolutional neural network (CNN) could discern between signals from virus-positive and virus-negative samples, as well as differentiate between different viruses. To validate this, I employed fluorescent labeling and imaging on IBV, three lab-cultivated influenza A strains—H3N2 A/Udorn/72 (Udorn), H3N2 A/Aichi/68 (X31), and H1N1 A/PR8/8/34 (PR8)—and a control devoid of viruses using virus-free cell culture media (Figure 4.4 A). These viruses, while similar in morphology, are indistinguishable to the naked eye in diffraction-limited microscope visuals of fluorescently marked particles (Figure 4.6). However, after image segmentation and analysis of the BBXs' attributes, I observed subtle but statistically relevant variations in pixel intensity, area, and the ratio of semimajor to semiminor axes within the BBXs (Figure 4.4 B-D). For instance, IBV exhibited greater brightness than influenza, while Udorn spanned a more extensive area than its counterparts.

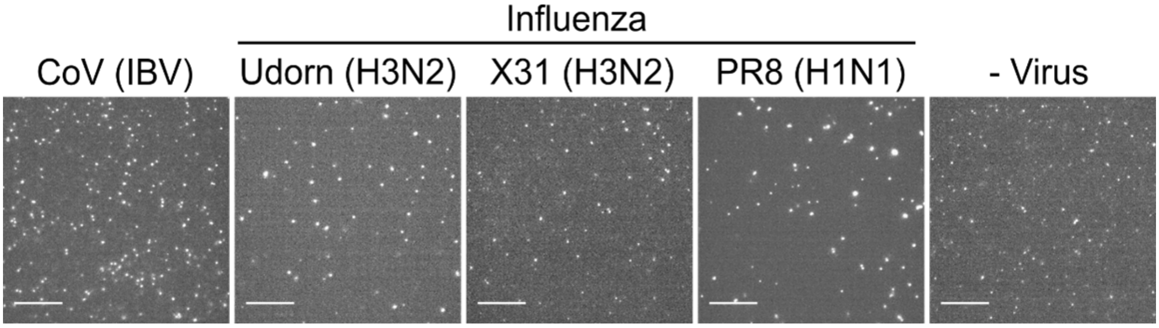


Figure 4.5: **Super-resolution images of cation-labeled virus particles. A.** Schematic of immobilization: biotinylated virus was labeled with 0.23 M $CaCl_2$ and 1 nM DNA-Alexa647, incubated on biotin-PEG slides treated with neutravidin, and imaged. **B.** Diffraction-limited TIRF FOV and corresponding super-resolved image. **C-E.** Super-resolved images of PR8, Udorn, and X31 viruses (scale bar: $10\mu m$; zoom-in: $100nm$). **F.** Histogram of virus areas from PR8 (4 FOVs), Udorn (2 FOVs), and X31 (1 FOV). **G.** Histogram of major/minor axis ratios for PR8, Udorn, and X31.

This observation was further corroborated by super-resolution imaging of cation-tagged viral entities. Utilizing fluorescence-driven super-resolution microscopy, I captured both diffraction-limited and high-definition visuals of identical fields of view. This offered a juxtaposition between the BBX-isolated signals for machine learning and their super-resolved counterparts. Biotinylated viruses were anchored on pegylated coverslips, subsequently labeled with $CaCl_2$ and a DNA tethered to a switchable Alexa647 dye. During imaging, the fluorescent outputs from the Alexa647 DNAs on the viral entities were documented, and each resulting localization was meticulously adjusted to reconstruct a super-resolved visual. Cluster analysis of these super-

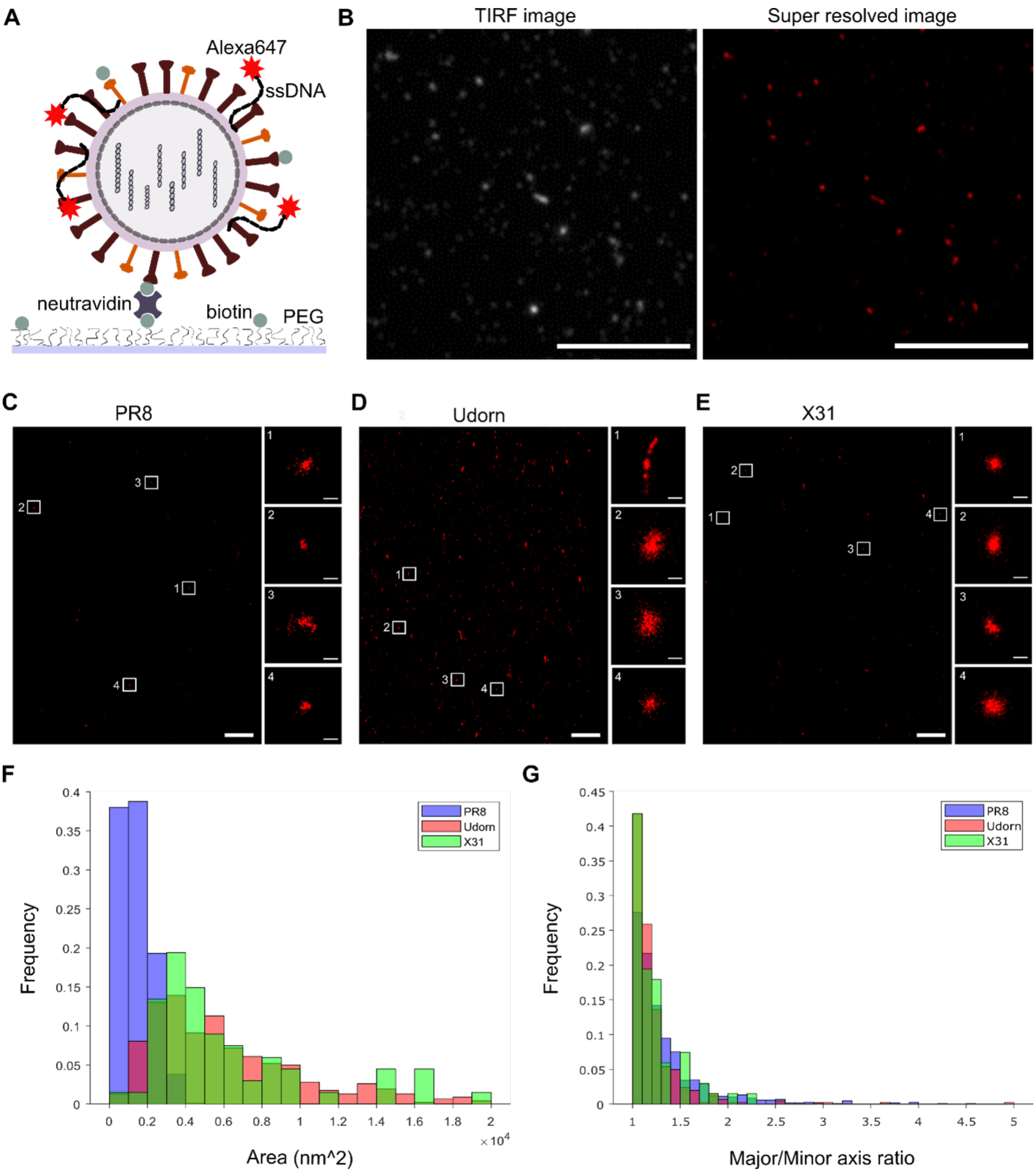


Figure 4.6: **Super resolution images of cation-labelled virus particles.** Representative fields of view (FOVs) of fluorescently labelled CoV (IBV), influenza A (Udorn, X31 and PR8) and a virus-negative control (- Virus). The samples were immobilized and labelled with 0.23 M SrCl2, 1 nM Cy3 (green) DNA and 1 nM Atto647N (red) DNA before being imaged. FOVs from the red channel are shown. Scale bar 10μm. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

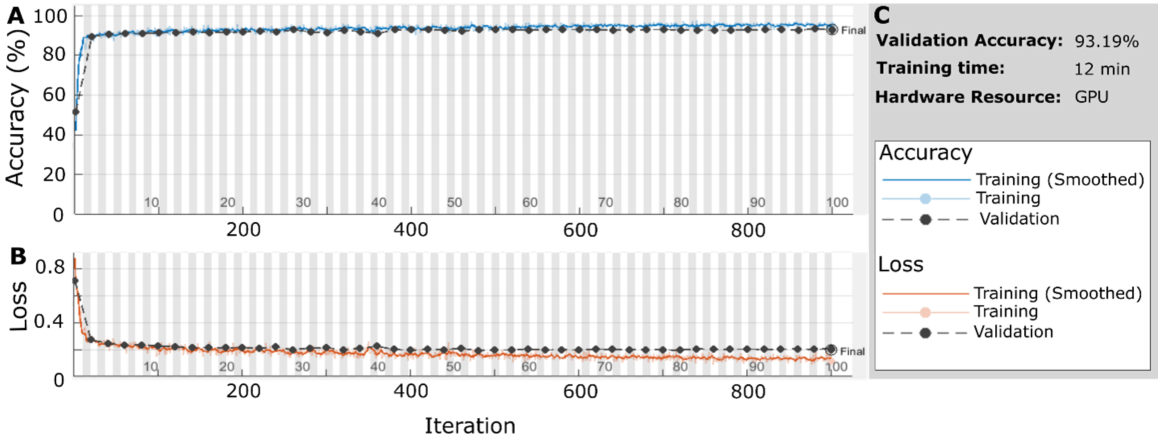


Figure 4.7: **Training Progress for CoV (IBV) and the virus-negative control (-Virus).** **A.** Graph describing the validation accuracy of the network per iteration. **B.** Graph showing the loss function value for each. **C.** Overall validation accuracy and training time. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

resolved localizations indicated that the fluorescent outputs seen in the diffraction-limited visuals of labeled specimens align with virion-sized particles. Moreover, distinct viral classes seemed to exhibit minor variations in labeling density, area, and morphology (Figure 4.5). I believe that deep learning algorithms can harness these minute disparities, along with more intricate image attributes like pixel correlations, to categorize the viruses.

For sample classification, I devised a 15-layer CNN (Figure 4.4 E, details in legend). The initial step involved imaging IBV and a virus-negative control, comprising solely SrCl2 and DNA. Both samples underwent independent imaging across three days, with four sessions each. This three-day imaging span ensured the capture of any inherent variability in the imaging process, such as minor temperature fluctuations, thereby bolstering the trained models' capability to categorize data from subsequent datasets. The extracted BBXs for each specimen were then arbitrarily split into a training set (70%) and a validation set (30%). The training set facilitated the CNN's learning to distinguish IBV from negative signals, utilizing approximately 3000 BBXs for each sample. The trained network underwent validation with the untouched 30% of the data. The inaugural data point during network validation registered at 50%

accuracy, anticipated for a wholly arbitrary categorization of entities into two groups. This was succeeded by a swift surge in validation accuracy as the network identified prominent parameters, with a more gradual ascent as iterations multiplied (Figure 4.7 A). Concurrently, there was a corresponding decline in the Loss Function (Figure 4.7 B). The entire training-validation cycle culminated in 12 minutes (Figure 4.7 C).

The network validation outcomes are depicted as a confusion matrix, a standard tool for illustrating classification task performance metrics Figure 4.8 A. Rows represent the predicted class, columns indicate the true class, and the bottom-right cell denotes the model's overall validation accuracy for each categorized particle. For the IBV vs. negative comparison, the trained network adeptly distinguished between positive and negative samples, achieving an impressive accuracy (91.4%), sensitivity (91.9%), and specificity (90.9%) (Figure 4.8 B). Importantly, these probabilities pertain to the identification of individual viral particles in the specimen, not the entire sample. Consequently, the likelihood of accurately identifying a sample teeming with hundreds or thousands of viral particles is expected to rise (discussed subsequently).

4.5 Efficient Classification of Virus Strains Across Optical Systems Using Deep Learning

We evaluated the network's ability to distinguish between different virus types and strains by training it on BBXs derived from images of IBV and various influenza strains (Udorn, X31, PR8, and H1N1 A/WSN/33). The network achieved high accuracy, with 95.5% for IBV vs. Udorn (Figure 3C) and 94.3% for IBV vs. PR8 (Figure 3D). It could also differentiate between closely related influenza strains, achieving 79.6% accuracy for WSN vs. PR8 (Figure 3E), likely reflecting the homogeneity between these H1N1 strains. Importantly, the network's classification was independent of the host cell line, as both WSN and PR8 were grown in MDCK cells. It also distinguished IBV from a pooled dataset of virus-negative controls and influenza strains with 92.2% accuracy (Figure 4.8 F), and performed well in a multi-class experiment, distinguishing IBV, PR8, and WSN with 81.9% accuracy (Figure 4.9).

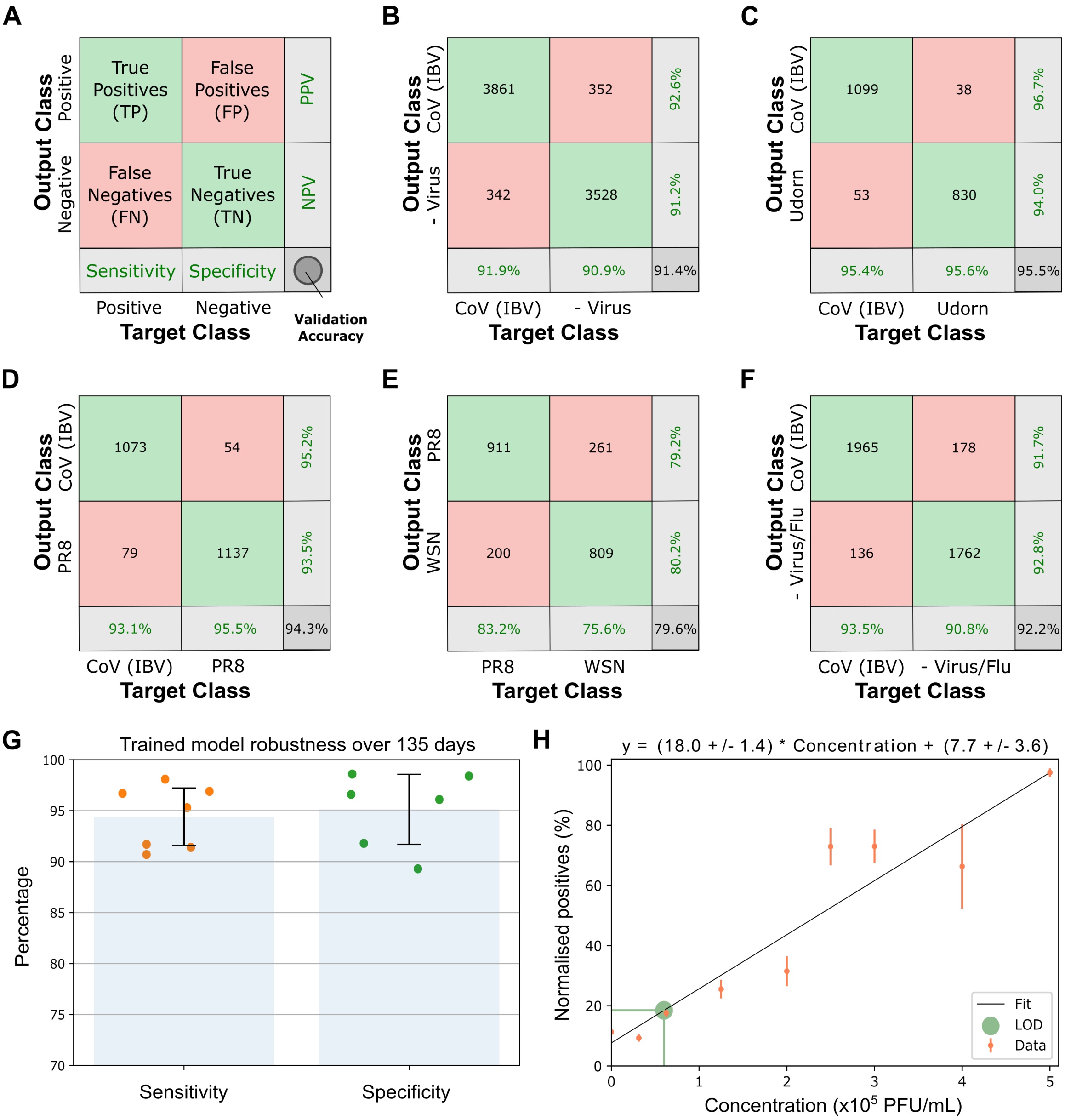


Figure 4.8: **Network validation results for laboratory-grown virus strains. A.** Confusion matrix with predicted vs. true class, showing predictive values, sensitivity, and specificity. **B.** Confusion matrix for CoV (IBV) positive and negative samples. **C, D.** Confusion matrices comparing CoV (IBV) vs. influenza strains Udorn and PR8. **E.** Confusion matrix for influenza PR8 vs. influenza WSN. **F.** Confusion matrix for CoV (IBV) vs. a pooled dataset of virus-negative control and three influenza A strains. **G.** Network stability over 135 days, with sensitivity and specificity shown for each time point. **H.** Limit of detection determined as 610^4 PFU/mL based on the linear increase in normalized positive particles with virus concentration. Error bars represent standard deviation. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

Figure 4.9: **Multi-classifier network validation for virus strains.** Confusion matrix showing the network's ability to differentiate between IBV, PR8, and WSN. 5200 BBXs per strain were split into 70% training and 30% validation datasets. The trained CNN accurately classified IBV from the two influenza strains, which were also distinguished from each other with high accuracy. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

To demonstrate broader applicability, we repeated the experiments using a different optical system (Zeiss Elyra 7 with a 63× objective vs. ONI Nanoimager with a 100× objective). The network successfully distinguished two influenza strains (WSN and Udorn) from virus-negative controls and from each other, with accuracies ranging from 74% to 78% (Figure 4.10). This confirms that virus classification is independent of imaging conditions and microscope type. Additionally, the network maintained stable performance over 135 days without retraining, with no loss in sensitivity or specificity (Figure 4.8 G).

We also estimated the limit of detection (LOD) by testing the network's ability to detect decreasing concentrations of IBV, WSN, and SARS-CoV-2 (Figure 4.8 H, Figure 4.11). Using virus samples of known titers, the network produced LODs of 6×10^4, 4.6×10^4, and 5.4×10^4 PFU/mL for IBV, WSN, and SARS-CoV-2, respectively.

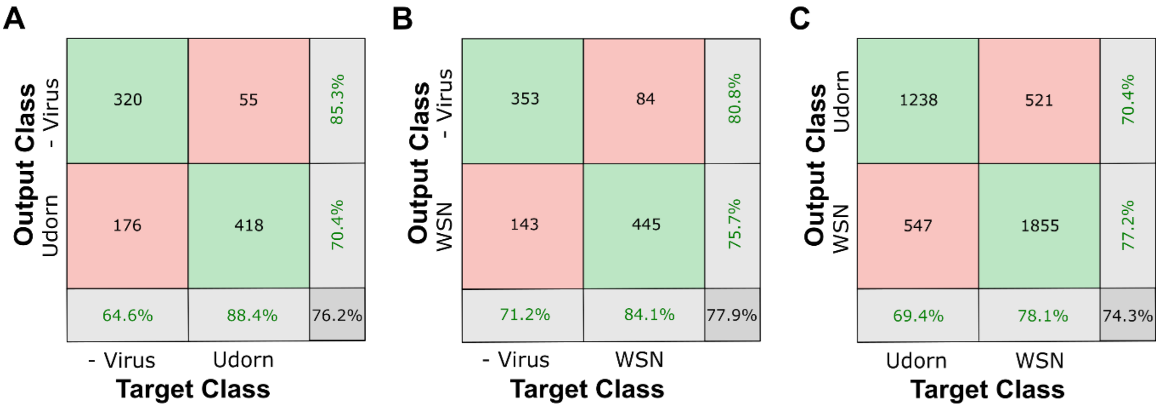


Figure 4.10: **Data taken on a Zeiss Elyra 7 shows similar results.** **A.** Confusion matrix showing that a trained network could differentiate between Udorn and negative control **B.** Confusion matrix showing that a trained network could differentiate between WSN and negative control. **C.** Confusion matrix showing that a trained network could differentiate between two strains of influenza (WSN and Udorn). *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

While this sensitivity is lower than RT-PCR (approximately 10^2 PFU/mL [181]), it remains clinically relevant, as viral loads for SARS-CoV-2 typically range from 10^4 to 10^7 copies/mL in clinical samples [182].

4.6 Discussion

Our initial proof-of-concept experiments using lab-grown viruses demonstrated that single-particle fluorescence microscopy, combined with deep learning, could effectively distinguish between virus-positive and negative samples. The CNN achieved high accuracy in differentiating avian coronavirus IBV from various influenza strains, with over 90% accuracy per particle (e.g., IBV vs. Udorn, IBV vs. PR8). These results were validated using multiple virus stocks and different microscopes, showing consistent performance. However, a fully independent validation was not possible with lab-grown viruses, as the stocks were derived from the same virus type, cultured in identical conditions. Despite this, the experiments established that our approach can differentiate between closely related viruses, even those prepared under the same conditions, as demonstrated with H3N2 strains.

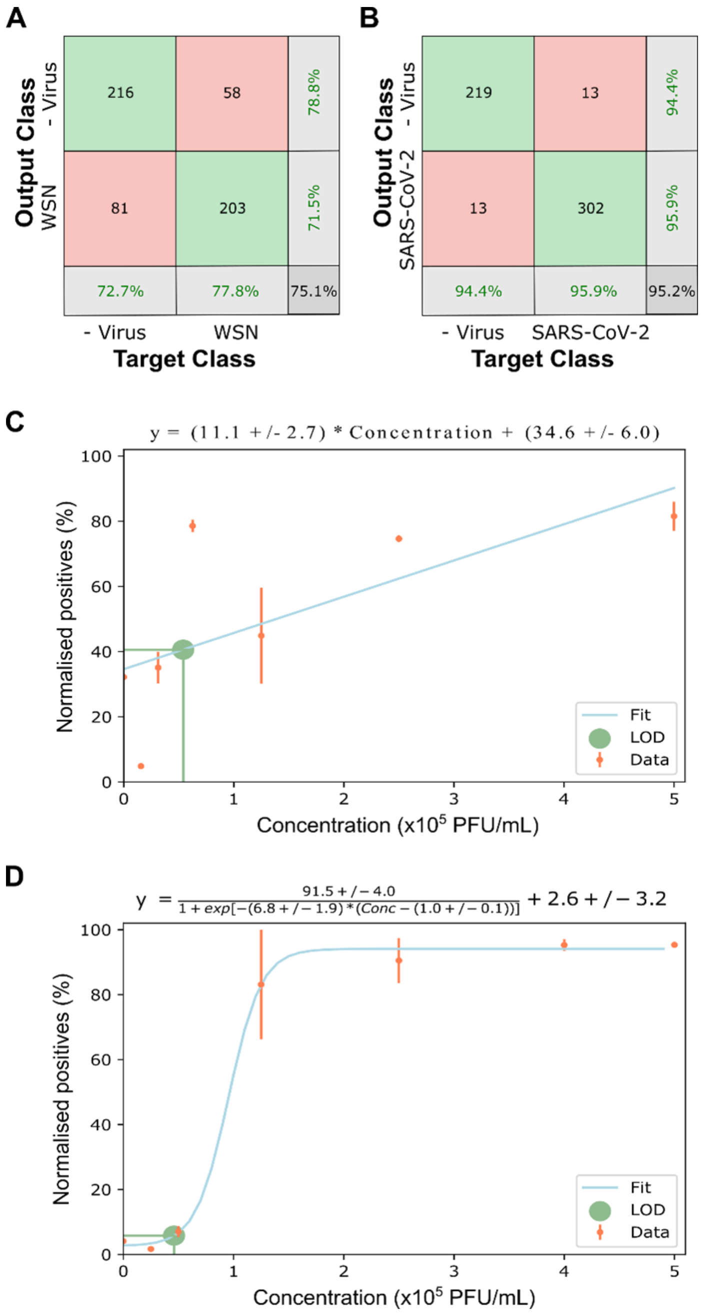


Figure 4.11: Limit of detection for influenza and SARS-CoV-2. **A.** Confusion matrix for WSN vs. virus-negative control. **B.** Confusion matrix for SARS-CoV-2 vs. virus-negative control. **C.** WSN: normalized positive particles increased with virus concentration; LOD = 5.4×10^4 PFU/mL. **D.** SARS-CoV-2: same as C; LOD = 4.6×10^4 PFU/mL. Error bars represent standard deviation. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

Chapter 5

Clinical Validation of Detection of Immobilised Virions

> Every virus tells a story, not just in its structure, but in its motion.

5.1 Methods and Materials

5.1.1 Single Stranded DNAs

Same as in Chapter 4.

5.1.2 Clinical Samples

We secured ethical clearance to utilize anonymized samples taken from the oro- or nasopharynx of patients for diagnosing influenza and other respiratory pathogens, including SARS-CoV-2. This approval was granted by the North West-Greater Manchester South Research Ethics Committee (REC reference: 19/NW/0730). Before inactivation, these specimens were preserved in Copan Universal Transport Medium (UTM). The inactivation process involved treating them with a 4% formaldehyde solution for 30 minutes at ambient temperature. However, for the 104 samples used in the subsequent validation study, a 1% formaldehyde solution was used for 5 minutes at room temperature. The presence or absence of SARS-CoV-2 in samples was confirmed using either the Public Health England 2019-nCoV real-time RT-PCR RdRp gene test or the RealStar SARS-CoV-2 RT-PCR Kit by Altona Diagnostics. For detecting other respiratory pathogens and identifying specific strains of seasonal human coronavirus (hCoV) like OC43, HKU1, or NL63, we employed the BioFire FilmArray Respiratory Panel (by Biomerieux) and the Cepheid Xpert Xpress Flu/RSV test.

In our study, we analyzed 213 clinical samples collected from patients between November 2020 and July 2021. For training our network, we captured images of samples from varied patients over a span of 3 days, ensuring different sample preparations from the same patient samples each day. We allocated 70% of the bounding boxes (BBXs) extracted from all images taken during these 3 days for training purposes, reserving the remaining 30% for validating the network. The outcomes of this validation are depicted in the provided confusion matrices, with each matrix representing a uniquely trained model. In total, 58 clinical samples were employed for both training and validation.

Subsequently, we conducted two separate evaluations of the trained networks using clinical samples that hadn't been used in either the training or validation phases. The initial test involved 51 patient samples, which included negative samples or those positive for SARS-CoV-2, Flu A, or seasonal hCoVs. The latter test utilized 104 patient samples, either negative or positive for SARS-CoV-2. Among the positive samples, 14 were identified as the original Wuhan variant, 12 as the Alpha variant (as indicated by a spike gene target failure in RT-PCR using the TaqPath COVID-19 combo kit by ThermoFisher), and 10 were confirmed as the Delta variant through RT-PCR using the Taqman SARS-CoV-2 mutation panel by ThermoFisher.

5.1.3 Sample Preparation

The preparation process for both positive and negative samples was consistent, including inactivation using the same formaldehyde concentration and labeling in a uniform buffer. Only samples prepared in the same manner were compared. For clinical samples, the treatment used was limited to chitosan and the labeling with calcium chloride. Glass slides were coated with 0.015, mg/mL chitosan, a linear polysaccharide, in 0.1, M acetic acid for 30 minutes, followed by three rinses with Milli-Q water.

Although chitosan introduced some background in control samples, it facilitated rapid virus immobilization, which is essential to prevent virus aggregation. Virus stocks, typically $10, \mu L$, were diluted in $0.23, \mathrm{M}, \mathrm{CaCl_2}$ along with $1, \mathrm{nM}$ of each dye-labeled DNA, resulting in a final volume of $20, \mu L$, which was applied to the slide. For SARS-CoV-2 imaging, the cationic labeling solution was buffered with $20, \mathrm{mM}$ Tris at pH 8.

5.1.4 Imaging

Three wide-field Nanoimager microscopes were utilized to capture the images. "Microscope 1" featured a Hamamatsu Flash 4 LT.1 sCMOS camera, while "Microscopes 2 and 3" were fitted with a Hamamatsu Flash4 V3 sCMOS camera. Apart from these distinctions, the systems were identical in their specifications. The imaging of the sample was conducted using total internal reflection fluorescence (TIRF) microscopy, combined with a 100 oil-immersion lens. The laser light was directed at an average angle of 53° relative to the perpendicular. Short movies, consisting of 5 frames for each field of view (FOV) that spanned $75 \times 49 \mu m$, were recorded at a rate of $33Hz$ with a $30ms$ exposure time. The laser intensities were consistently set at $0.78kW/cm^2$ for the red ($640nm$) laser and $1.09kW/cm^2$ for the green ($532nm$) laser. To automate the process and ensure unbiased FOV selection, the entire sample underwent scanning using the microscope's multi-acquisition feature, capturing 81 FOVs within 2 minutes. An automated autofocus system meticulously managed defocusing, positioning the sample at a predetermined axial location before each FOV was illuminated by the excitation lasers. This positioning was accomplished by observing the reflection of a near-infrared laser on the glass/sample medium boundary and aligning the image with a previously stored reference image.

5.1.5 Data Segmentation

Same as in Chapter 4.

5.1.6 Deep Learning

Same as in Chapter 4.

5.1.7 Zeta Potential Measurements

All zeta potential measurements were performed using the Zetasizer Nano S with disposable folded capillary cells (DTS1070). The temperature was maintained at 25 °C, with an equilibration time of 120 seconds. Samples were prepared in different buffers based on their pH: 20 mM sodium acetate for pH 4–5, 20 mM HEPES for pH 6–7, and 20 mM Tris for pH 9. The buffer details were used in the software to specify the dispersant properties, allowing for the calculation of viscosity and dielectric constant. For the sample material, a composition of 50% lipid and 50% protein was selected to determine the absorption and refractive index. The analysis model was set to “auto” mode, and the software automatically adjusted the measurement duration, with a minimum of 10 runs and a maximum of 100 runs. Each sample was measured three times at each pH level.

5.1.8 Statistical Analysis

5.1.8.1 Confusion Matrices

Same as in Chapter 4.

5.1.8.2 Chi-Squared Test

To transition from classifying individual bounding boxes (BBX) to determining the result for an entire clinical sample, we employed the Chi-squared test. This test factors in the total count of bounding boxes, the number of BBXs identified as positive or negative, and the model’s specificity (that is, the likelihood of correctly identifying a negative signal). By incorporating the model’s specificity and the total signal count in a sample, we adjust for variations in signal detection across samples. The test also acknowledges that a certain number of bounding boxes might be misclassified. The Chi-squared test is a statistical method that, under its null hypothesis, expects observed frequencies for a categorical variable to align with its anticipated frequencies. The test can be represented by the following equation:

$$\chi^2 = \sum i = 1n\frac{(O_i - E_i)^2}{E_i} \tag{5.1}$$

In this equation, χ^2 stands for chi-squared, Oi represents the observed value, Ei is the expected value, and n is the total number of labels. The threshold p-value can differ based on the specific model, but it's generally set to be less than 0.01, corresponding to a confidence level exceeding 99%. In this study, the null hypothesis presumes the sample is negative. It's only dismissed when the p-value falls below the threshold, at which point the sample is deemed positive.

5.2 Accurate Classification of Clinical Samples

After demonstrating our assay on laboratory-grown viruses, we proceeded to evaluate clinical samples (workflow shown in Figure 5.1 A). Throat swabs were collected from 33 patients who tested negative for viruses or positive for SARS-CoV-2, seasonal hCoVs (OC43, HKU1, or NL63), or human influenza A, as determined by RT-PCR. These samples were inactivated using formaldehyde before being labeled and immobilised (see Experimental Methods). The labeling process for viruses in clinical samples was consistent and reproducible (Figure 5.2). Images of the samples, taken over three different days, were used to train and validate the network for various paired classification tasks (e.g., SARS-CoV-2 vs. negative or SARS-CoV-2 vs. hCoVs). Details of the clinical samples used for network training and validation can be found in Appendix A, Table A.1. The network validation results were presented as confusion matrices, similar to the earlier analysis.

Our initial results with SARS-CoV-2 clinical samples showed lower validation accuracy compared to laboratory-grown virus strains (approximately 70% at the BBX level, Figure 5.3 A). However, accuracy significantly improved when labeling was performed at a higher pH (pH 8). This improvement is likely due to the higher isoelectric point (pI) of SARS-CoV-2 (around 9) compared to influenza (around 6) [183, 184,

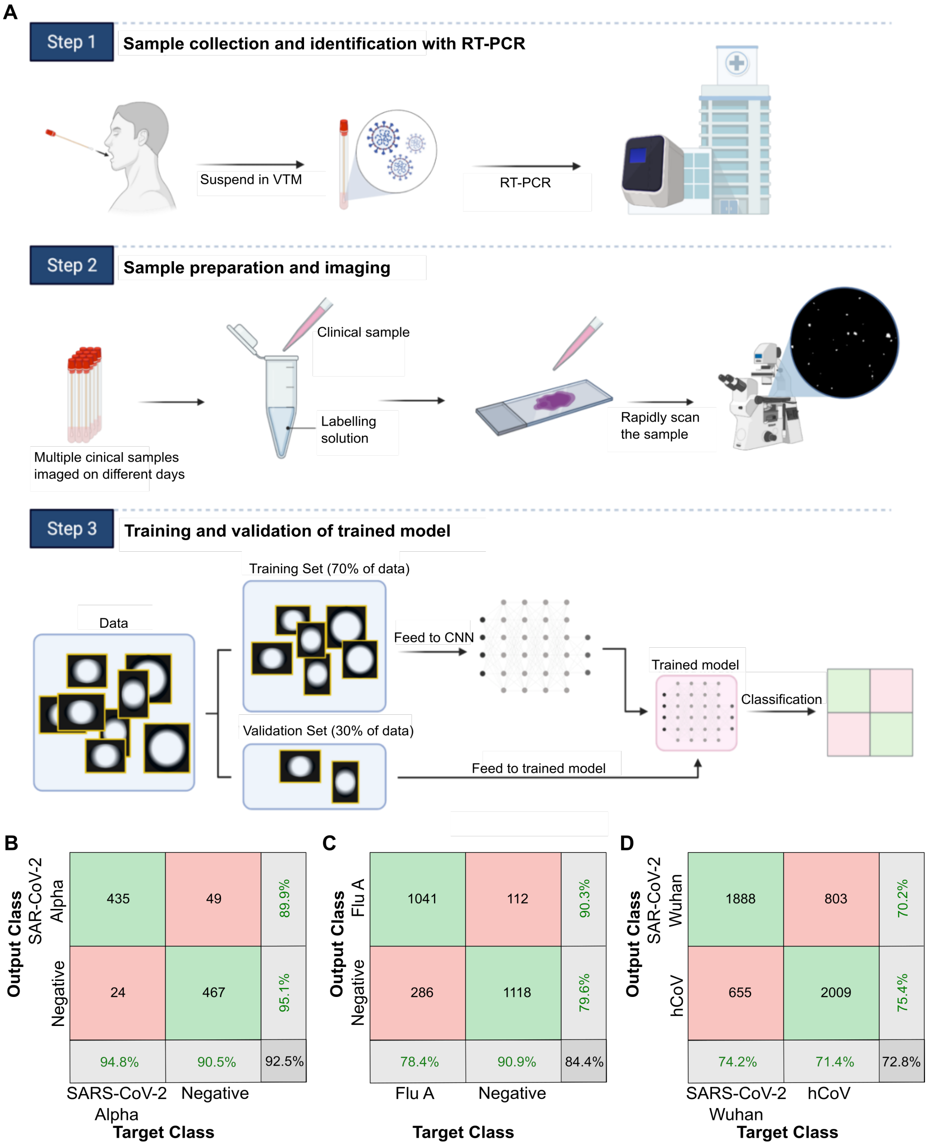


Figure 5.1: **A deep learning network distinguishes viruses in clinical samples. A.** Workflow for training and validation: Samples from 33 patients were labeled, imaged over three days, and processed into BBXs. A convolutional neural network (CNN) was trained using 70% of the BBXs, while the remaining 30% validated the model, with results shown in a confusion matrix. **B.** Confusion matrix demonstrating the network's ability to differentiate SARS-CoV-2 Wuhan strain (10 samples) from seasonal human coronaviruses (hCoVs HKU1, NL63, and OC43; 28 samples). **C.** Confusion matrix showing differentiation of influenza A (6 samples) from negative samples (17 samples). **D.** Confusion matrix displaying the network's ability to distinguish the SARS-CoV-2 Alpha variant (3 samples) from negative samples (3 samples). *Figure created with BioRender.com. See Appendix A Table A.1 for further details on the clinical samples. Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

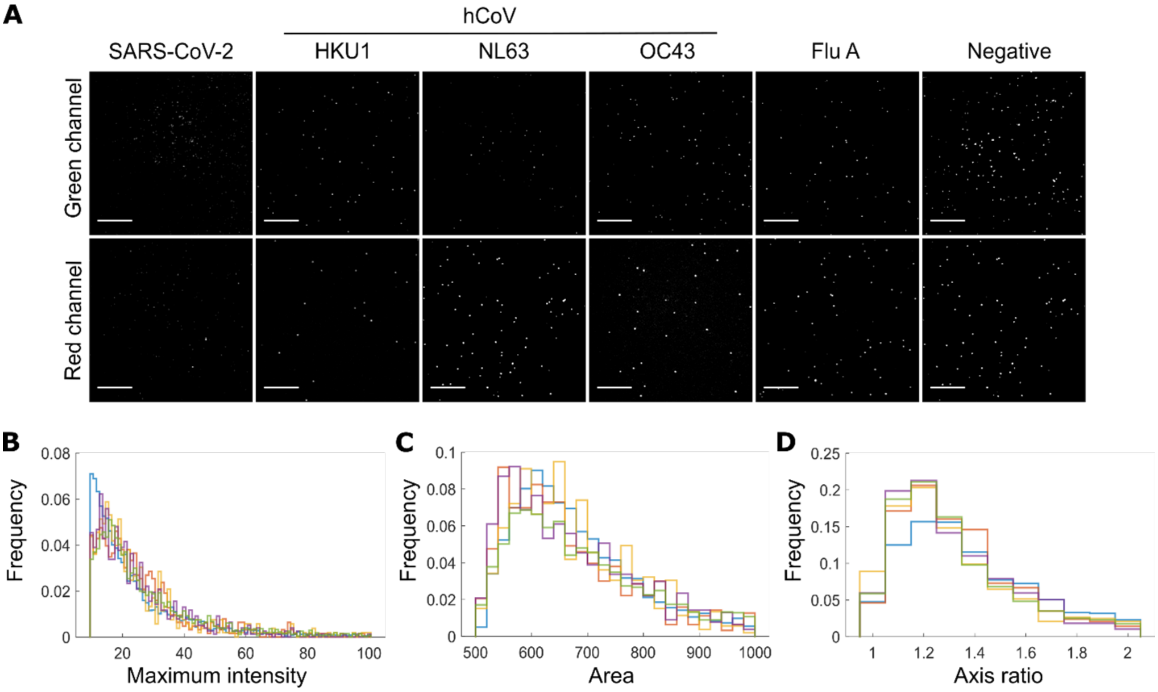


Figure 5.2: **Comparison of imaging results from clinical samples. A.** Representative fields of view (FOVs) of fluorescently labeled SARS-CoV-2, three seasonal human coronaviruses (hCoV), influenza A (Flu A), and a virus-negative control. Samples were immobilized and labeled with $0.23 M CaCl_2$, $1 nM$ Cy3 (green) DNA, and $1 nM$ Atto647N (red) DNA before imaging. Scale bar: 10 µm. **B-D.** Normalized frequency plots of maximum pixel intensity, area, and semi-major-to-semi-minor axis ratio within BBXs from five randomly selected SARS-CoV-2 clinical samples, illustrating labeling robustness and reproducibility. Data were obtained from 81 FOVs per slide for each virus, with each sample shown in a different color. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

185], making the virions more negatively charged at higher pH. This enhanced negative charge allowed for more efficient labeling with the cationic solution and better capture by the charged chitosan surface, leading to more accurate SARS-CoV-2 detection. Using the optimized protocol, the network achieved a validation accuracy of approximately 93% when distinguishing between SARS-CoV-2-positive and negative BBXs (Figure 5.1 B).

The network also demonstrated a validation accuracy of about 84% for differentiating Flu A-positive from negative BBXs (Figure 5.1 C) and around 78% for distinguishing seasonal hCoVs from negative samples (Figure 5.3 B). It could differentiate between SARS-CoV-2 and seasonal hCoVs with an accuracy of about 73% (Figure

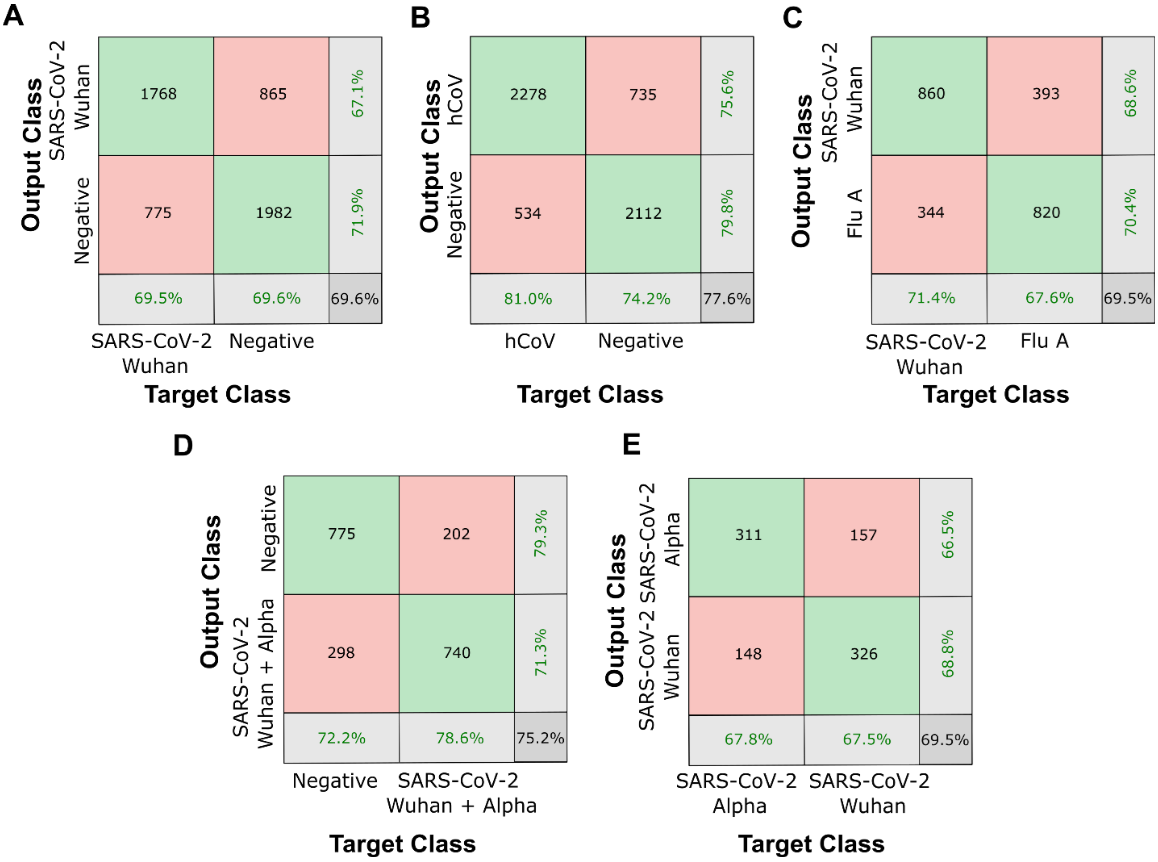


Figure 5.3: **Network validation results for clinical samples. A.** Confusion matrix for differentiating SARS-CoV-2 Wuhan strain (10 samples) from negative samples (8 samples). **B.** Matrix showing differentiation between seasonal human coronaviruses (hCoV HKU1, NL63, and OC43; 29 samples) and negative samples (9 samples). **C.** Matrix for distinguishing influenza A (6 samples) from SARS-CoV-2 Wuhan strain (12 samples). **D.** Matrix showing differentiation between SARS-CoV-2 strains (Wuhan and Alpha; 6 samples) and negative samples (4 samples). **E.** Matrix for distinguishing between the SARS-CoV-2 Wuhan strain (3 samples) and Alpha variant (3 samples). See Appendix A Table A.1 for further details on the clinical samples. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

5.1 D) and between SARS-CoV-2 and Flu A with approximately 70% accuracy (Figure 5.3 C), useful for identifying co-circulating infections. Additionally, the network achieved about 75% accuracy in distinguishing negative samples from combined data of the original Wuhan strain and the Alpha variant of SARS-CoV-2 (Figure 5.3 D), and around 70% accuracy when distinguishing between these two variants (Figure 5.3E).

5.3 Validation of Trained Networks Using Independent Clinical Samples

I then evaluated the trained network's ability to diagnose independent clinical samples not previously used in training or validation. A total of 51 new samples (from different patients) were analysed, including negative samples and those positive for SARS-CoV-2, Flu A, or seasonal hCoVs. These samples were imaged on a fourth day and classified by the trained networks within seconds. Comparing the results to RT-PCR and performing chi-squared tests where needed (Figure 5.4 A and Figure 5.5, Steps 1 and 2), we found that 50 out of 51 samples were correctly classified, achieving an overall accuracy of 98.0% (Figure 5.4 B and Appendix A, Table A.2). All 11 negative samples were accurately identified, resulting in a specificity of 100%, while 39 out of 40 positive samples were correctly classified, yielding a sensitivity of 97.5%. The single misclassified sample had significantly fewer BBXs than the others (Appendix A, Table A.2), suggesting a viral load near the detection limit, which may explain the misclassification.

In this analysis, we also examined whether different SARS-CoV-2 variants could be distinguished. BBXs from images of seven clinical samples, initially classified as positive by a SARS-CoV-2 vs. negative network, were further analyzed by a second network to identify whether they were the original Wuhan variant or the Alpha variant, which first appeared in the UK in 2020 (Figure 5.5, Steps 3 and 4). All seven samples were correctly classified (Figure 5.4 B).

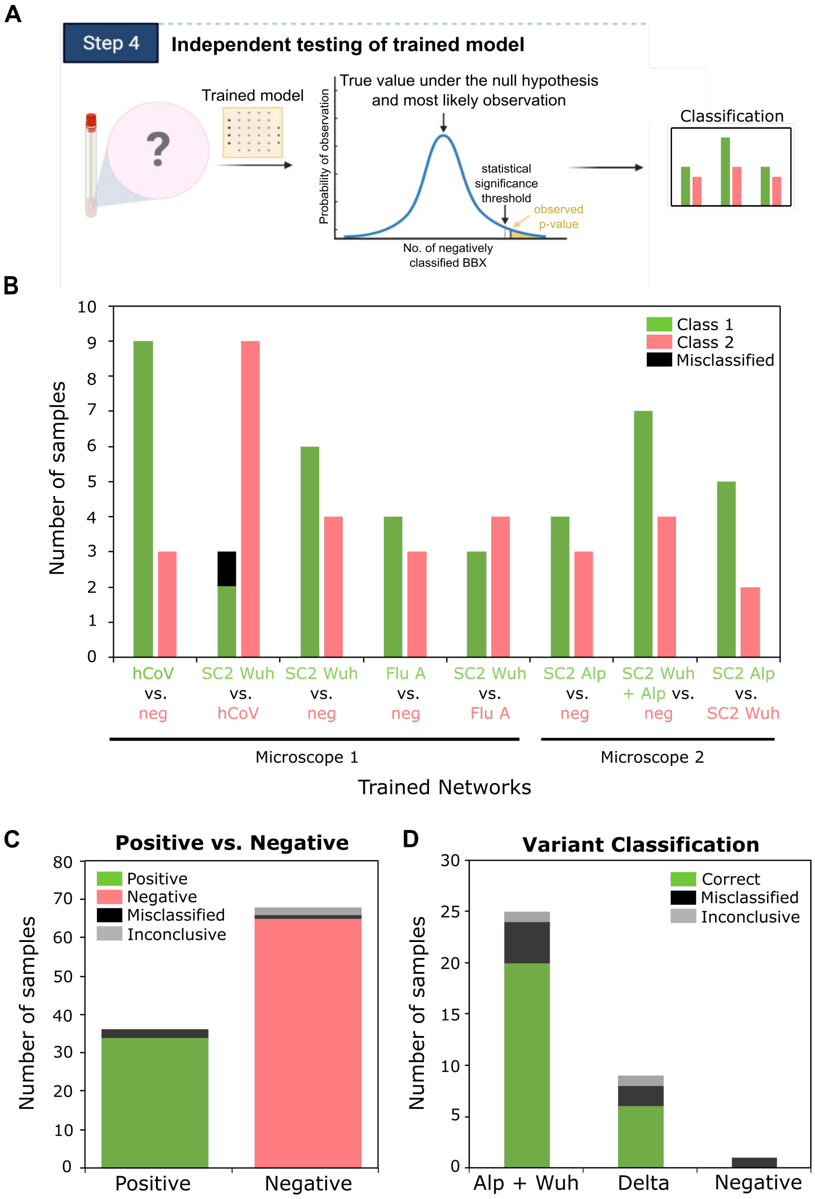


Figure 5.4: **Independent testing of the trained network with clinical samples. A.** Workflow overview: Unseen samples are imaged, processed into BBXs, and analyzed by a trained network. A chi-squared test is used to classify samples as positive or negative based on a pre-set confidence threshold. **B.** Summary of testing results for 51 new patient samples using various trained networks. Chi-squared tests were applied to classify the samples, with 50 out of 51 correctly identified, yielding an accuracy of 98.03%. **C.** Results of testing 104 samples using a single trained network (SARS-CoV-2 vs negative).**D.** Matrix showing differentiation between SARS-CoV-2 strains (Wuhan and Alpha) and negative samples. **E.** Variant classification of positive samples from C using a second network to differentiate between Wuhan/Alpha and Delta variants. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

Recognising the clinical importance of a test that can identify SARS-CoV-2 variants without sequencing, we expanded the study with 104 additional clinical samples. These included 68 virus-negative and 36 SARS-CoV-2-positive samples, determined by RT-PCR (Appendix A Table A.3). Collected between November 2020 and July 2021, the positive samples comprised 14 original Wuhan, 12 Alpha (identified by spike gene target failure in RT-PCR), and 10 Delta variants. Samples were labeled, immobilized, and imaged as usual, and the images were processed into BBXs. In the first analysis step, the 104 samples were classified as either SARS-CoV-2-positive or negative (validation results in Figure 5.6 A). Two negative samples were inconclusive due to fewer than the 5 BBXs required for the chi-squared test. Of the remaining 102 samples, all but three were correctly classified (Figure 5.4 C), resulting in an overall accuracy of 97.1%, with a sensitivity of 94.4% and specificity of 98.5%.

In the next step, the BBXs from the 35 samples identified as positive were passed through a second model trained to distinguish between Wuhan/Alpha variants and Delta SARS-CoV-2 (Figure 5.6 B). Two samples were inconclusive due to closely matched p-values (Figure 5.5, Steps 3 and 4). The single misclassified negative sample was identified as Delta, and six positive samples were misclassified (two each of Delta, Alpha, and Wuhan variants), resulting in a 77.1% accuracy, a 5.7% inconclusive rate, and a 17.1% misclassification rate (Figure 5.4 D and Appendix A, Table A.4). Although the variant classification accuracy was lower than the positive vs. negative test accuracy, this method could still be valuable for rapid variant screening when sequencing facilities are unavailable.

5.4 Mechanisms of Virus Differentiation: Charge Interactions and Deep Learning Insights

We further explored the underlying mechanisms that enable the network to distinguish between different viruses by examining the role of charge interactions and the specific patterns the network identifies in the images. This analysis involved both

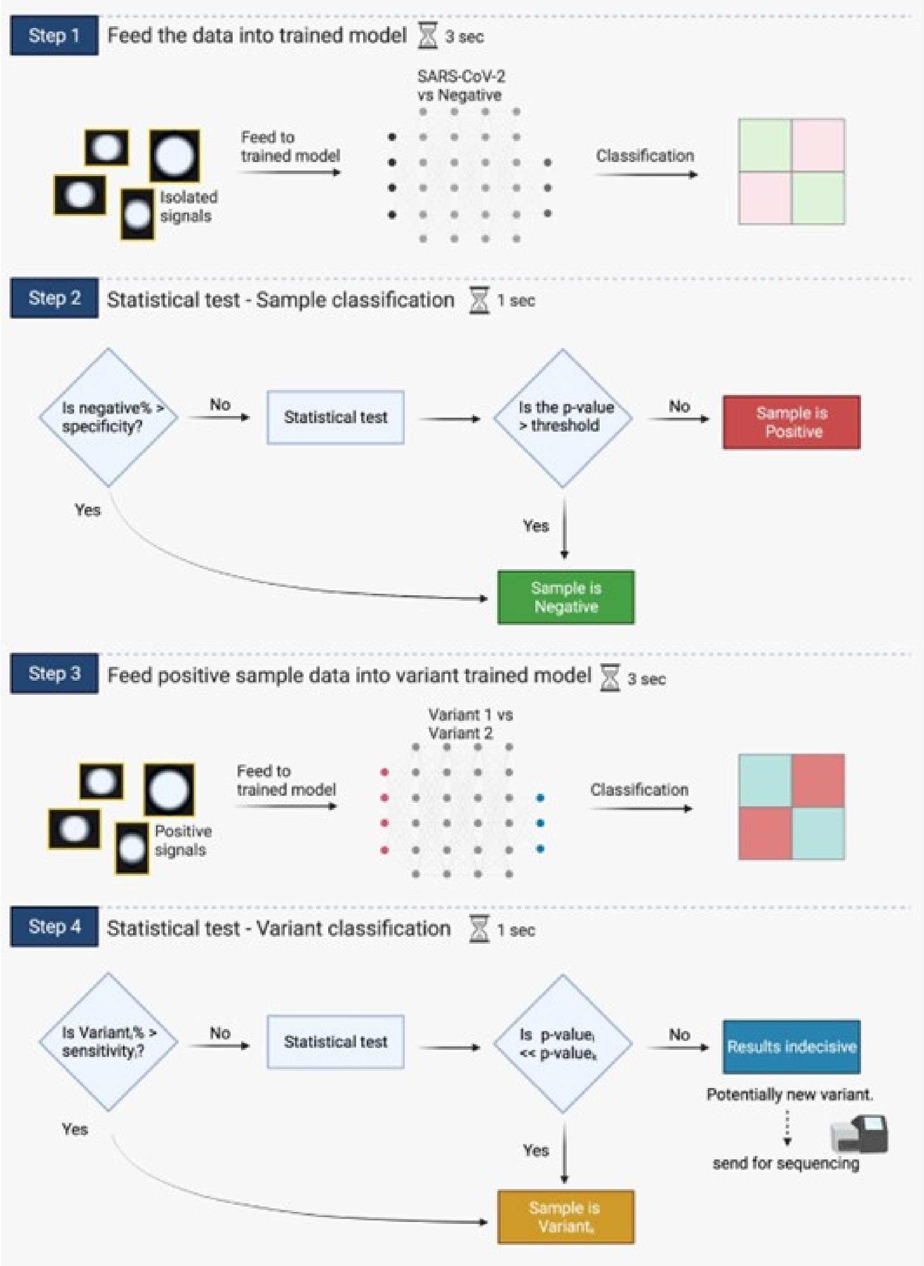


Figure 5.5: **Workflow for sample classification. Step 1**: Unseen samples are imaged, processed into BBXs, and analyzed by a trained network (e.g., SARS-CoV-2 vs. negative). **Step 2**: Samples are classified as negative if the proportion of BBXs classified as negative exceeds the network's specificity. Otherwise, a chi-squared test is used to determine if the sample is positive. **Step 3**: SARS-CoV-2 positive samples are further analyzed to identify the variant using a second trained network. **Step 4**: If the proportion of BBXs classified as a specific variant exceeds the network's sensitivity, the sample is classified as that variant. Otherwise, chi-squared tests are used for final classification, with inconclusive results sent for sequencing if needed. *Figure created with BioRender.com. Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

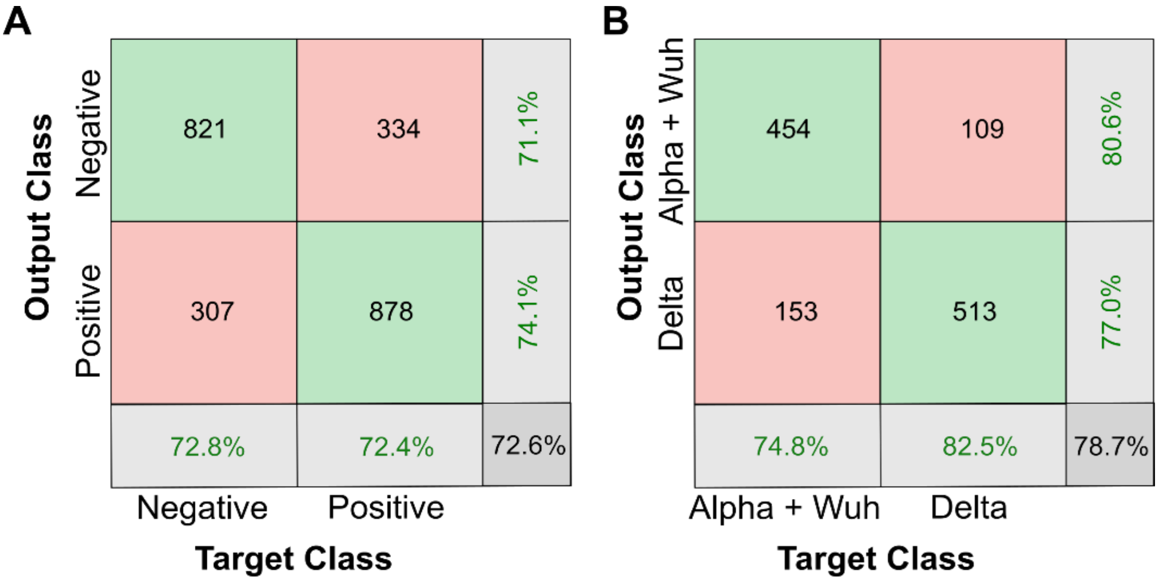


Figure 5.6: **Validation results for the trained network on clinical samples. A.** Confusion matrix demonstrating the network's ability to distinguish SARS-CoV-2 positive samples (4 clinical samples of Wuhan and Delta variants, plus additional lab-grown SARS-CoV-2 used in training) from negative samples (7 clinical samples). **B.** Confusion matrix illustrating the network's capability to differentiate between SARS-CoV-2 Wuhan strain (7 samples) and Delta variant (7 samples). See Appendix A Table A.1 for further details on the clinical samples. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

zeta potential measurements and visualizations of the features that the convolutional neural network (CNN) uses for classification (Figure 5.7).

5.4.1 Zeta Potential Measurements

The zeta potential, which measures the surface charge of particles in solution, provides insights into how viruses interact with the cationic labeling solution used in our assay. We measured the zeta potential of three influenza strains—WSN (H1N1), PR8 (H1N1), and Udorn (H3N2)—across a range of pH values. The results indicated that all strains became more negatively charged as the pH increased, reflecting the general trend of increasing surface charge negativity at higher pH levels (Figure 5.7 A). Notably, even closely related strains exhibited differences in zeta potential at the same pH. These variations suggest that charge-based interactions between the virus surface and the labeling molecules could affect the efficiency and uniformity of labeling, creating subtle but consistent differences in the fluorescence images. Such differences in labeling could contribute to the CNN's ability to distinguish between virus strains, as variations in charge can alter the overall density and distribution of fluorescent signals on the viral particles.

5.4.2 Visualization of Learned Patterns

To understand how the CNN differentiates between virus samples, we used the Matlab DeepDreamImage function, which highlights the image features that the network finds most relevant for classification. This approach involves feeding an image through the trained network and adjusting the image to amplify the activations in a specific layer, revealing the patterns the network uses to distinguish between categories.

For the network trained to classify samples as either negative or positive for SARS-CoV-2, the visualized patterns indicated a focus on high-intensity, rounded structures in the positive samples (Sup. Fig. 12B, right panel). These features likely correspond to the fluorescent signals from the labeled virus particles. In contrast, the negative samples showed patterns with much lower intensity in the center of the bounding

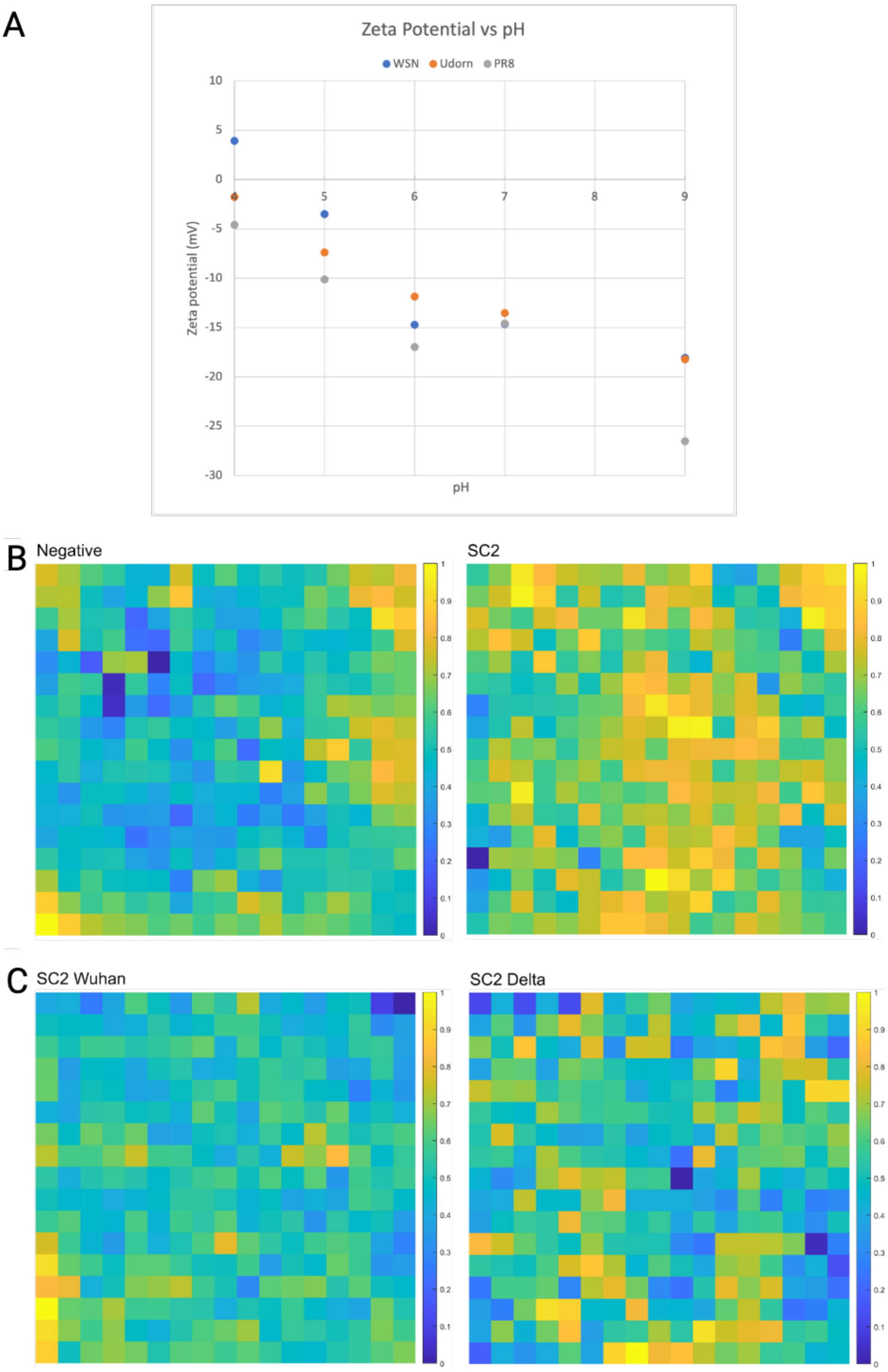


Figure 5.7: **Zeta potential measurements and visualization of patterns learned by the CNN using the Matlab DeepDreamImage function. A.** Zeta potential measurements for influenza strains WSN (H1N1), PR8 (H1N1), and Udorn (H3N2) across different pH values, with error bars showing the standard deviation from triplicate measurements. **B.** Visualized patterns identified by a network trained to distinguish between negative and SARS-CoV-2 positive samples, using image gradients to highlight key features. **C.** Patterns identified by a network trained to differentiate between the Wuhan and Delta variants of SARS-CoV-2. *Reproduced from the author's own work previously published in ACS Nano, 2023, 17 (7), 6276–6298. https://doi.org/10.1021/acsnano.2c10159. Copyright © 2023 American Chemical Society.*

boxes (BBXs), suggesting that the network can distinguish the presence of virus particles based on the brightness and structure of the labeled regions.

A similar analysis was performed for the network trained to differentiate between two SARS-CoV-2 variants—the original Wuhan strain and the Delta variant. Here, the visualized patterns revealed that the network identified differences in the intensity distribution and shapes between the two variants (Sup. Fig. 12C). These distinctions could reflect subtle changes in the virus surface structure or labeling efficiency due to mutations in the viral glycoproteins that alter the isoelectric point and, consequently, the surface charge of the particles.

5.4.3 Implications for Virus Differentiation

The combined insights from the zeta potential measurements and feature visualizations support the hypothesis that the CNN leverages subtle differences in size, shape, and labeling patterns to classify viruses accurately. As each virus strain or variant may exhibit unique electrostatic properties and structural features, these differences become encoded in the fluorescent images and are learned by the network during training. This ability to detect nuanced variations between virus particles allows the CNN to achieve high accuracy in distinguishing even closely related viruses and variants.

The capability to visualize what the network "sees" further underscores the robustness of our approach. By focusing on isolated signals from individual virus particles rather than entire fields of view, the network's classification remains unaffected by variations in virus concentration or imaging conditions. This property enhances the reliability of the test, making it suitable for diverse clinical samples, including those with low viral loads or prepared under varying conditions. Thus, these findings emphasize the potential of our method for rapid and accurate virus detection in a diagnostic setting, even in the absence of high-resolution imaging or complex sample preparation steps.

5.5 Discussion

We expanded our study to include 155 clinical samples, achieving high overall accuracies of 98.0% and 97.1% in two independent tests. The samples varied in collection methods, viral transport media volumes, and storage conditions, which could affect virus particle integrity. Despite these variations, our method showed promise as a reliable diagnostic tool. Using clinical samples allowed for truly independent validation, as these samples came from patients not involved in network training or validation. Our network accurately classified unseen samples, proving its capability to identify SARS-CoV-2 and differentiate between variants like Wuhan and Alpha. Although the variant classification accuracy was slightly lower, it demonstrated potential for rapid, sequencing-free variant detection. Further optimization, such as improving virus immobilization or using multi-classifier networks for detecting various respiratory viruses, could enhance sensitivity and broaden the diagnostic capabilities of this method.

Chapter 6

Diffusion-based Detection of Virions in under 1min

> Every virus tells a story, not just in its structure, but in its motion.

6.1 Graphical Abstract

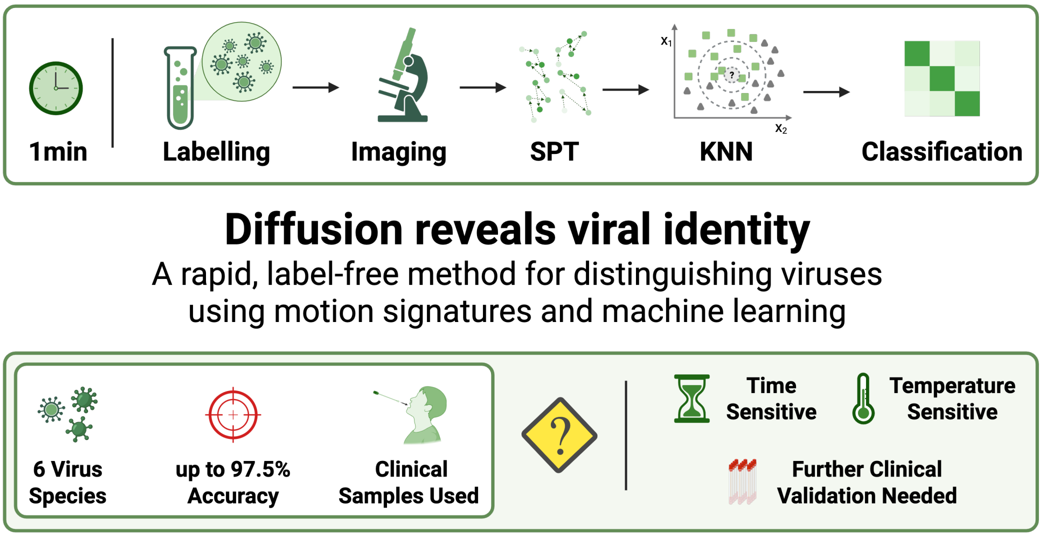


Figure 6.1: A one-minute workflow rapidly classifies viruses by leveraging diffusion-based motion signatures and machine learning. Viruses are labeled, imaged, and tracked at the single-particle level; motion patterns are then analyzed using a fine KNN classifier to distinguish viral species with high accuracy. This method achieved up to 97.5% accuracy across six virus types, including clinical samples. While powerful, the approach is sensitive to time and temperature, and further clinical validation is required for broader deployment. *Figure created with BioRender.com.*

6.2 Introduction

Accurate and rapid diagnostics are essential in the fight against infectious diseases. The ability to detect and identify viruses quickly is critical, not only during pandemics but also for routine disease management. Many current diagnostic methods rely on complex laboratory techniques, such as nucleic acid amplification or antigen-based assays, which, while effective, often fall short of the requirements for true point-of-care (POC) diagnostics. These methods typically require specialized equipment, trained personnel, and controlled laboratory environments, making them difficult to deploy in resource-limited settings or for immediate field use. Additionally, even when portable diagnostic devices are available, challenges with IT interfacing — such as difficulties in integrating devices with electronic health records, ensuring secure data transmission, and maintaining user-friendly software interfaces — can hinder the seamless use of diagnostics at the point of care. Addressing both the biological and digital challenges is therefore crucial for developing truly effective POC diagnostic solutions.

The growing demand for diagnostics that are truly POC—not adaptations of laboratory techniques but systems designed for simplicity, portability, and real-time operation—calls for innovation in both detection methodologies and analytical approaches. Current alternatives, such as lateral flow assays, provide rapid results but often trade accuracy and sensitivity for speed and ease of use. To address these limitations, it is necessary to develop diagnostic methods that combine the precision of laboratory techniques with the usability required for decentralized deployment.

In this context, we have developed a method leveraging fluorescence imaging to detect and classify diffusing virus particles. Using a combination of rapid fluorescence labeling and High-Intensity Laser Oblique (HILO) microscopy, this technique allows for the visualization of virus particles diffusing in solution. HILO microscopy provides high-contrast fluorescence imaging with minimal background, enabling the detection of even faintly labeled particles.

Once imaged, the labeled particles are analyzed using particle tracking to determine their trajectories over time. The physical and optical properties of viruses, such as size, diffusion coefficients, and fluorescence intensity, manifest as unique features in these trajectories. These features are then used to classify viruses using machine learning techniques, specifically k-Nearest Neighbors (kNN). By comparing the trajectory and fluorescence characteristics of each particle to known samples, kNN enables accurate classification with minimal computational requirements, making it suitable for POC applications.

This approach eliminates the need for immobilizing virus particles or extensive sample preparation, focusing instead on real-time, in-solution analysis. By combining advanced fluorescence labeling, particle tracking, and simple yet effective machine learning algorithms, this method achieves the speed, portability, and accuracy needed for truly point-of-care diagnostics. It represents a shift from adapting laboratory-based techniques to creating systems inherently designed for decentralized, rapid, and actionable diagnostics.

6.3 Methods and Materials

6.3.1 Lab-grown virus Strains

Same as in Chapter 4.

6.3.2 Clinical Samples

Same as in Chapter 4.

6.3.3 Single Stranded DNAs

We procured single-stranded oligonucleotides modified with fluorescent dyes from IBA, based in Germany. The DNA labeled with a “red” fluorophore in our study was modified at its $5'$ end with ATTO647N, with sequence ACAGCACCACAGACCAC CCGCGGATGCCGGTCCCTACGCGTCGCTGTCACGC. The DNA labeled with a “green” fluorophore was modified with Cy3b, modified at its $3'$ end, with sequence GGGTTTGGGTTGGGTTGGGTTGGGTTTTGGGTTTGGGTTGGGTTGGGA AAAA.

6.3.4 Sample Preparation

The preparation process for both positive and negative samples was identical. For instance, all samples were inactivated using the same formaldehyde concentration and labeled in a consistent buffer. Only samples that underwent identical preparation were analyzed in comparison to one another.

Virus samples were prepared by mixing $3\,\mu$L of virus with $15\,\mu$L of buffered $CaCl_2$ solution (0.45 M concentration) and $2\,\mu$L of 20 nM of fluorescently-labeled ssDNA, resulting in a final volume of $20\,\mu$L. This mixture enabled immediate labeling of enveloped viruses without the need for incubation time. The labeled virus solution was then placed on a glass slide for imaging.

6.3.5 Imaging

Two Nanoimager microscopes employing Highly Inclined and Laminated Optical Sheet (HILO) microscopy were used to capture images for this study. Both systems were equipped with Hamamatsu Flash4 V3 sCMOS cameras and 100× oil-immersion objectives, ensuring identical specifications. HILO microscopy was chosen for its ability to illuminate a thin optical section by directing the laser at a steeply inclined angle, creating a laminated optical sheet near the coverslip.

For each microscope, a single field of view (FOV) measuring approximately $75 \times 49\,\mu$m was imaged. Imaging consisted of continuous movies of 1000 frames, recorded at a frame rate of 33 Hz with a 30 ms exposure time, resulting in a total acquisition time of approximately 30 seconds per movie. Laser intensities were maintained at $0.78\,\mathrm{kW/cm^2}$ for the red (640 nm) laser and $1.09\,\mathrm{kW/cm^2}$ for the green (532 nm) laser to ensure consistent and reproducible illumination.

Manual focus adjustments were performed at the start of imaging to position the sample accurately within the optimal illumination plane. The experimental design focused on a single FOV for each microscope, enabling detailed observation and tracking of diffusing particles over time. The use of HILO microscopy, with its highly inclined illumination, provided exceptional contrast and resolution, making it ideal for studying the dynamic behavior of particles in solution across two independent systems.

Despite its advantages, fluorescence microscopy also presents several challenges and limitations. Photobleaching, where fluorescent dyes lose their ability to emit light over time under continuous illumination, can limit the duration of imaging experiments and affect quantitative measurements. Additionally, phototoxicity may occur, especially at high laser intensities, potentially altering the behavior of live samples. Achieving consistent focus and minimizing background fluorescence are also critical for maintaining image quality, particularly in dynamic experiments like particle tracking. Furthermore, the requirement for specialized fluorescent labeling and the potential for dye-related artifacts must be carefully considered when interpreting results.

6.3.6 Single Particle Tracking

In this study, single particle tracking (SPT) was employed to analyze the dynamic behavior of individual virus particles imaged using Highly Inclined and Laminated Optical Sheet (HILO) microscopy. SPT involves identifying and connecting the positions of individual fluorescent particles over multiple frames, forming continuous trajectories that reveal their motion.

The positions of virus particles in each frame were extracted from the videos using the Nanoimager software (NimOS). These positions were then linked into trajectories based on a maximum step distance threshold. The maximum step distance was calculated using the Stokes-Einstein relation for diffusion:

$$D = \frac{k_B T}{6\pi\eta r}, \tag{6.1}$$

where D is the diffusion coefficient, k_B is Boltzmann's constant, T is the temperature, η is the fluid viscosity, and r is the effective radius of the particle. The mean squared displacement ($\langle x^2 \rangle$) for a particle diffusing over time interval Δt was then calculated as:

$$\langle x^2 \rangle = 4D\Delta t. \tag{6.2}$$

From this, the maximum step size L for tracking in 2D was determined as:

$$L = \sqrt{4\langle x^2 \rangle}. \tag{6.3}$$

An exclusion radius was applied within NimOS to differentiate between true viral trajectories and spurious connections due to noise or free dye molecules. This ensured that only biologically relevant tracks were retained for analysis.

Using this method, trajectories were generated for each labeled virus particle, allowing for the calculation of key parameters such as diffusion coefficients and trajectory lengths. These trajectories were later used as input features for machine learning algorithms, such as k-Nearest Neighbors (kNN), to classify virus particles based on their dynamic properties. The SPT method enabled robust and high-resolution tracking of diffusing particles, providing critical insights into their behavior in solution.

6.3.7 Statistical Analysis

6.3.7.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was employed to perform dimensionality reduction by identifying the key dimensions that capture the majority of the variance in the dataset. PCA begins by defining a hyperplane that minimizes the squared distances between the data points and their projections onto the plane. The axis that maximizes the variance of the dataset was determined first, followed by subsequent orthogonal axes that account for progressively less variance.

The variance of each principal component (PC) was calculated using the equation:

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{n-1}, \tag{6.4}$$

where x_i represents the value of the i-th data point along a particular dimension, $\bar{x}$ is the mean of all data points along that dimension, and n is the total number of data points.

The proportion of the dataset's variance explained by each PC, termed the variance ratio, was then computed as:

$$\text{Variance ratio} = \frac{\text{Variance of PC}_i}{\sum \text{Variance of all PCs}}. \tag{6.5}$$

To ensure the retention of at least 95% of the dataset's variance, the variance ratios were summed cumulatively, starting from the largest component, until the threshold was reached. The PCs selected in this manner defined the reduced-dimensionality space. This approach avoids arbitrary selection of dimensionality and ensures that the reduced representation preserves the maximum possible variance.

6.3.7.2 K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) is a supervised machine learning algorithm commonly used for classification. It operates under the assumption that similar inputs have similar outputs, meaning that data points close to one another in feature space are likely to belong to the same category.

For a dataset D and a test point x, the KNN algorithm identifies a set of the k nearest neighbors to x, denoted as S_x, where $S_x \subseteq D$. This set is defined such that the distance from x to any point y in S_x satisfies:

$$\text{dist}(x, y) \leq \max\left(\text{dist}(x, x_i)\right), \tag{6.6}$$

where x_i are the k nearest neighbors of x in D, and dist is a distance metric, typically Euclidean distance.

The algorithm assigns a label to the test point x based on the most frequent label among the k neighbors. Formally, the classification rule is expressed as:

$$h(x) = \text{mode}\left(\{y_i : (x_i, y_i) \in S_x\}\right), \tag{6.7}$$

where $h(x)$ is the predicted label for x, y_i is the label of neighbor x_i, and mode selects the label with the highest frequency among the neighbors.

In this project, the KNN classifier was implemented with $k = 1$ and $k = 3$ to evaluate performance. When $k = 1$, the test point takes the label of its single nearest neighbor. For $k = 3$, the test point is classified based on a majority vote among its three nearest neighbors.

The KNN algorithm assumes that the feature space is adequately populated to allow for meaningful comparisons. However, at higher dimensions, the distance between data points increases, leading to challenges in classification accuracy, a phenomenon known as the "curse of dimensionality." To mitigate this, preprocessing steps, such as normalization and dimensionality reduction, were performed to improve the performance of the KNN classifier.

6.3.7.3 Confusion Matrices

Same as is in Chapter 4.

6.4 Track Pre-Processing

To ensure sufficient data for training the machine learning model, preprocessing steps were applied to the extracted tracks to expand the dataset. A minimum track length of 10 steps was enforced to filter out incomplete or noisy trajectories, with shorter tracks being discarded. Tracks exceeding this minimum length were split into smaller sub-tracks to increase the dataset size while preserving relevant information. Each original track, T_i, containing n_i steps, was divided into overlapping sub-tracks, v_i, each containing 10 steps (10 x-coordinates and 10 y-coordinates), such that:

$$v_i = \{(i, \ldots, i+9)\}, \quad \text{for } i \in [1, n_i - 9]. \tag{6.8}$$

This overlapping approach ensured that sub-tracks retained the trajectory information essential for classification while significantly increasing the dataset size. Smaller step sizes, such as 5 or 7, were tested but resulted in minor accuracy drops, likely due to insufficient trajectory information. Conversely, larger step sizes, such as 20, produced fewer sub-tracks and reduced model performance due to the smaller training set size. The choice of a 10-step sub-track length balanced dataset expansion and model accuracy, as it retained sufficient trajectory information for effective feature extraction.

Splitting tracks into sub-tracks also provided several benefits beyond dataset expansion. Virus particle trajectories often exhibit variations in dynamics, such as changes in diffusion coefficients or directional movement, which are particularly relevant in solution where viruses tend to aggregate. Dividing tracks into shorter sub-tracks allowed these local variations to be captured and analyzed separately, providing more detailed insights into particle behavior and aggregation dynamics. Additionally, long tracks can accumulate noise or artifacts over time, which may affect the accuracy of downstream analyses. By sub-dividing tracks, the impact of noise was limited to individual segments, as each sub-track represented a shorter, more stable time interval. Furthermore, splitting tracks created a larger and more diverse dataset, improving the generalization capability of the machine learning model and reducing the risk of overfitting to specific trajectory patterns.

6.5 Training and Validation of Models using Lab Grown samples

6.5.1 The Data

The dataset used in this study was generated over four days with triplicates for each sample, ensuring robust and representative data. Each experiment produced tracks corresponding to the motion of individual virus particles or negative control samples under different conditions. These tracks were extracted from fluorescence microscopy videos and subjected to preprocessing as described above. This approach not only increased the number of usable tracks but also captured localised particle dynamics.

The preprocessing resulted in varying track counts depending on the virus type or medium. Table 6.1 summarizes the track counts for each virus and negative medium after preprocessing. Virus samples, such as PR8, Udorn, WSN, X31, IBV, and SC2, produced track counts ranging from 3,543 (SC2) to 15,584 (Udorn). Negative control samples (ALLA, DMEM, and MEM) yielded larger datasets, with MEM producing the largest dataset of 30,751 tracks.

Sample	**Track Count**
PR8	11,138
Udorn	15,584
WSN	13,603
X31	12,228
IBV	11,895
SC2	3,543
ALLA	15,956
DMEM	13,474
MEM	30,751

Table 6.1: Track counts for virus and negative medium samples after preprocessing.

To create balanced datasets for machine learning, tracks from these samples were combined into distinct training sets tailored for specific comparisons. Table 6.2 outlines the dataset sizes for virus vs negative sample comparisons and virus vs virus comparisons. For comparisons involving more than two classes (e.g., MEM vs WSN vs PR8), the dataset was classified as a multiclassifier. For virus vs virus tests involving SC2, only 3,000 tracks per sample were used to maintain balance, while 11,000 tracks were used for other virus-only comparisons.

Comparison	**Dataset Type**	**Track Count**
ALLA vs X31 vs IBV	Multiclassifier	33,000
DMEM vs Udorn	Virus vs Negative Sample	22,000
MEM vs WSN vs PR8 vs SC2	Multiclassifier	12,000
MEM vs WSN	Virus vs Negative Sample	22,000
MEM vs WSN vs PR8	Multiclassifier	33,000
PR8 vs SC2	Virus vs Virus	6,000
All Flu A	Virus vs Virus	44,000
All Flu A + SC2	Virus vs Virus	15,000
All Flu A + IBV	Virus vs Virus	55,000
All Viruses	Virus vs Virus	18,000

Table 6.2: Dataset sizes for virus vs negative sample and virus vs virus comparisons.

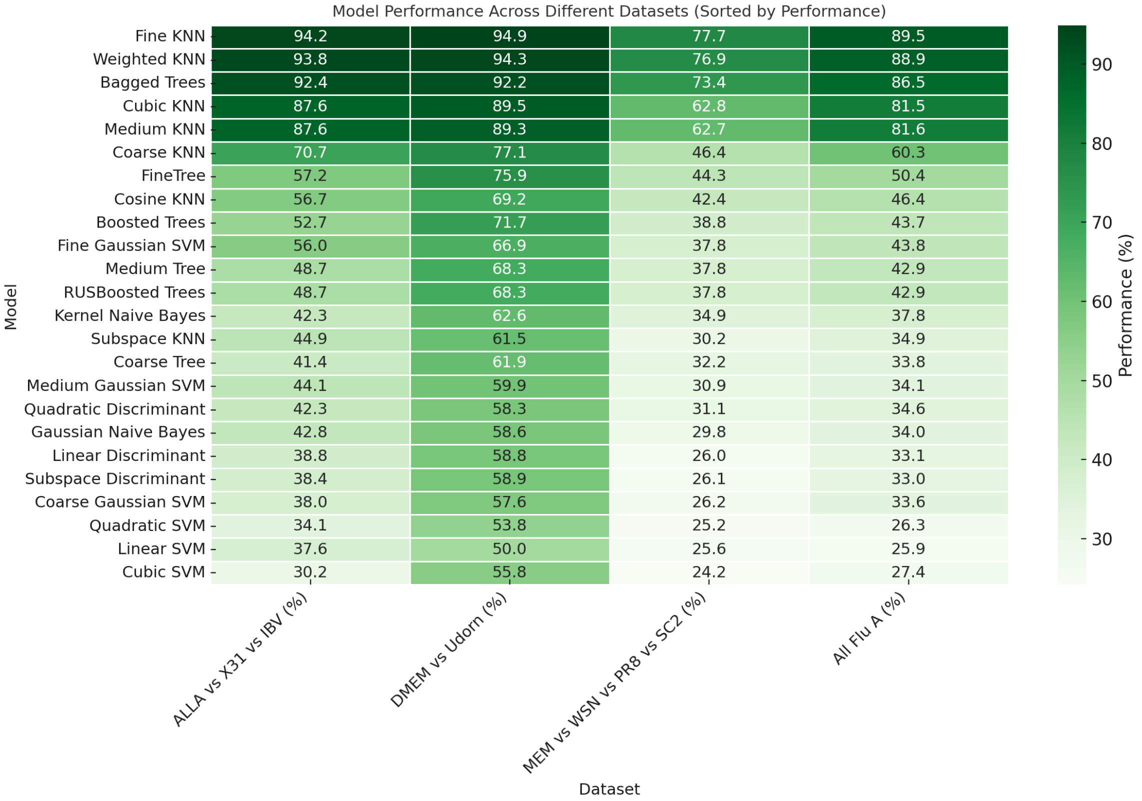
This careful preparation ensured that datasets were balanced and representative of the experimental conditions. The resulting datasets facilitated robust training and testing of machine learning models, supporting accurate classification and analysis of virus particle dynamics.

6.5.2 Model Selection

To evaluate and compare the performance of various machine learning algorithms, MATLAB's Machine Learning Toolbox was utilized. The dataset was split into two subsets: 70% of the data was used for training, while the remaining 30% was reserved for validation. This split ensured sufficient data for training while maintaining a robust test set to assess the generalizability of the models.

The training process involved testing all available models within the toolbox to identify the algorithm best suited for the classification tasks. These models included decision trees (FineTree, Medium Tree, Coarse Tree), support vector machines (SVMs) with linear, quadratic, cubic, and Gaussian kernels, k-nearest neighbors (KNN) with varying configurations, discriminant analysis (linear and quadratic), naive Bayes classifiers (Gaussian and kernel), and ensemble methods such as bagged trees and boosted trees. Model performance was evaluated on the validation set using percentage accuracy as the primary metric. By testing all models under identical data splits and preprocessing conditions, this approach ensured a fair comparison and enabled the selection of the most effective algorithms for classifying virus particle dynamics. The heatmap (Figure 6.2) illustrates the performance of various machine learning models across different datasets. Among the tested models, k-nearest neighbors (KNN) classifiers consistently outperformed other approaches, with Fine KNN and Weighted KNN achieving the highest accuracies. Specifically, Fine KNN achieved accuracies of 94.2%, 94.9%, 77.7%, and 89.5% on the ALLA vs X31 vs IBV, DMEM vs Udorn, MEM vs WSN vs PR8 vs SC2, and All Flu A datasets, respectively. Similarly, Weighted KNN demonstrated comparable performance, underscoring its robustness across diverse dataset conditions.

Ensemble methods, particularly Bagged Trees, also performed well, achieving high accuracy (92.4%, 92.2%, 73.4%, and 86.5% on the respective datasets). These meth-



Model	ALLA vs X31 vs IBV (%)	DMEM vs Udorn (%)	MEM vs WSN vs PR8 vs SC2 (%)	All Flu A (%)
Fine KNN	94.2	94.9	77.7	89.5
Weighted KNN	93.8	94.3	76.9	88.9
Bagged Trees	92.4	92.2	73.4	86.5
Cubic KNN	87.6	89.5	62.8	81.5
Medium KNN	87.6	89.3	62.7	81.6
Coarse KNN	70.7	77.1	46.4	60.3
FineTree	57.2	75.9	44.3	50.4
Cosine KNN	56.7	69.2	42.4	46.4
Boosted Trees	52.7	71.7	38.8	43.7
Fine Gaussian SVM	56.0	66.9	37.8	43.8
Medium Tree	48.7	68.3	37.8	42.9
RUSBoosted Trees	48.7	68.3	37.8	42.9
Kernel Naive Bayes	42.3	62.6	34.9	37.8
Subspace KNN	44.9	61.5	30.2	34.9
Coarse Tree	41.4	61.9	32.2	33.8
Medium Gaussian SVM	44.1	59.9	30.9	34.1
Quadratic Discriminant	42.3	58.3	31.1	34.6
Gaussian Naive Bayes	42.8	58.6	29.8	34.0
Linear Discriminant	38.8	58.8	26.0	33.1
Subspace Discriminant	38.4	58.9	26.1	33.0
Coarse Gaussian SVM	38.0	57.6	26.2	33.6
Quadratic SVM	34.1	53.8	25.2	26.3
Linear SVM	37.6	50.0	25.6	25.9
Cubic SVM	30.2	55.8	24.2	27.4

Figure 6.2: The heatmap shows the performance of machine learning models across datasets, sorted by average performance. Rows represent models (e.g., FineTree, Medium KNN), and columns correspond to datasets: ALLA vs X31 vs IBV, DMEM vs Udorn, MEM vs WSN vs PR8 vs SC2, and All Flu A. Each cell displays the model's percentage accuracy. The color gradient from light to dark green highlights performance, with darker shades indicating higher accuracy (e.g., above 80%). Annotations provide exact values, enabling direct comparison. The visualization identifies models with consistently strong performance across datasets and specific strengths in individual comparisons.

ods benefit from combining predictions from multiple models to reduce variance and improve generalization. In contrast, simpler models such as linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) exhibited lower performance, particularly on more complex datasets like MEM vs WSN vs PR8 vs SC2, indicating their limitations in handling high-dimensional or non-linear data.

The preprocessing approach played a pivotal role in enhancing model accuracy. Splitting tracks into sub-tracks significantly expanded the dataset, providing machine learning models with sufficient examples for training. This was particularly beneficial for models like KNN, which rely on proximity-based decision-making and require a diverse training dataset for optimal performance. Sub-track segmentation not only increased data availability but also captured local dynamics, such as changes in diffusion coefficients or directional motion, which are critical for accurate classification. Moreover, this process reduced the impact of noise by segmenting longer tracks into shorter, more stable intervals, ensuring that essential trajectory features were preserved.

Given the consistent superiority of k-nearest neighbors (KNN) classifiers, particularly Fine KNN and Weighted KNN, these models were selected for subsequent analyses. Their high performance across all datasets, coupled with their ability to effectively utilize the expanded and preprocessed dataset, makes them ideal for classifying virus particle dynamics. Future steps leverage the strengths of KNN models to extract meaningful insights and improve classification accuracy further.

6.5.3 PCA: Effect on Accuracy and Prediction Speed

The classification performance of influenza A virus subtypes was evaluated using Fine KNN, both with and without the application of principal component analysis (PCA). The results are summarized in Figure 6.3 and Table 6.3.

The confusion matrices in Figure 6.3 highlight the trade-offs between dimensionality reduction and classification accuracy. Without PCA (A), the model achieves high accuracy, with true positive rates exceeding 95% for all subtypes and minimal misclassification. For instance, PR8 is correctly classified 96.9% of the time, with only

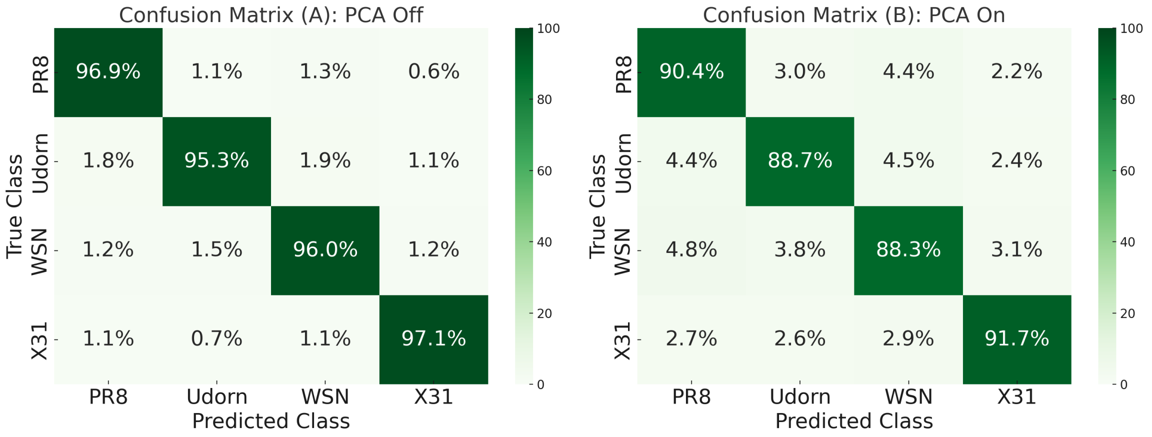


Figure 6.3: Confusion matrices for influenza A virus classification. (A) PCA Off, overall accuracy 96.3%; (B) PCA On, overall accuracy 89.8%. Values represent percentages of predictions.

minor confusion with Udorn, WSN, and X31 (1.1%, 1.3%, and 0.6%, respectively). Similarly, X31 has a true positive rate of 97.1%.

When PCA is applied (Figure 6.3 B), the overall accuracy drops to 89.8%, and misclassification rates increase. For example, PR8's true positive rate decreases to 90.4%, with increased confusion with other subtypes, such as WSN (4.4%) and Udorn (3.0%). These results demonstrate that while PCA accelerates prediction speeds, it does so at the expense of critical information necessary for distinguishing closely related virus subtypes.

Dataset	**Accuracy (%)**	**Prediction Speed (obj/s)**	**PCA**
PR8 vs Udorn vs WSN vs X31 (All FluA viruses)	96.3	3,000	OFF
PR8 vs Udorn vs WSN vs X31 (All FluA viruses)	89.8	150,000	ON
PR8 vs Udorn vs WSN vs X31 vs IBV (FluA vs IBV)	95.4	2,300	OFF
PR8 vs Udorn vs WSN vs X31 vs IBV (FluA vs IBV)	87.3	160,000	ON
All viruses (FluA vs IBV vs SC2)	83.3	7,800	OFF
All viruses (FluA vs IBV vs SC2)	77.2	150,000	ON

Table 6.3: Performance of Fine KNN with and without PCA for virus classification tasks.

Table 6.3 further underscores the trade-offs. Prediction speeds are significantly improved with PCA, increasing from approximately 3,000 objects/s without PCA to 150,000 objects/s with PCA for influenza A viruses. However, this speed comes at the cost of reduced accuracy, highlighting the challenge of balancing computational efficiency and model performance.

To address this trade-off, expanding the training dataset can help compensate for the loss of information introduced by PCA. This strategy ensures that the classifier remains robust and accurate, even when dimensionality reduction techniques are applied. Consequently, the choice to use PCA should consider the specific requirements of the task, such as the need for high prediction speeds or maximum accuracy.

6.5.4 Positive vs Negative

Distinguishing viral samples from negative controls is a critical step in developing diagnostic tools, as it ensures that only relevant viral signals are captured and false positives are minimized. Effective classification between virus and negative samples enables a reliable foundation for downstream analysis, including virus strain differentiation. The accuracy of this distinction is crucial for both clinical diagnostics and research applications, where precise identification of viral presence can inform treatment decisions and improve understanding of viral behavior. Figure 6.4 presents confusion matrices summarizing the performance of the Fine KNN model across four datasets: DMEM vs Udorn, MEM vs WSN, MEM vs WSN vs PR8, and ALLA vs IBV vs X31. The overall average accuracies achieved in these datasets demonstrate the model's robustness in separating viral and negative samples. Specifically, the average accuracies were 94.3% for DMEM vs Udorn, 91.9% for MEM vs WSN, 88.1% for MEM vs WSN vs PR8, and 94.2% for ALLA vs IBV vs X31.

The high accuracy across all datasets underscores the reliability of the Fine KNN model for this task. Binary classifications, such as DMEM vs Udorn and MEM vs WSN, performed exceptionally well, benefiting from simpler decision boundaries. Multiclass scenarios, like MEM vs WSN vs PR8 and ALLA vs IBV vs X31, presented greater complexity but still achieved competitive performance. These results highlight the model's suitability for distinguishing viral from negative samples, providing a strong basis for further analyses, including virus strain classification.

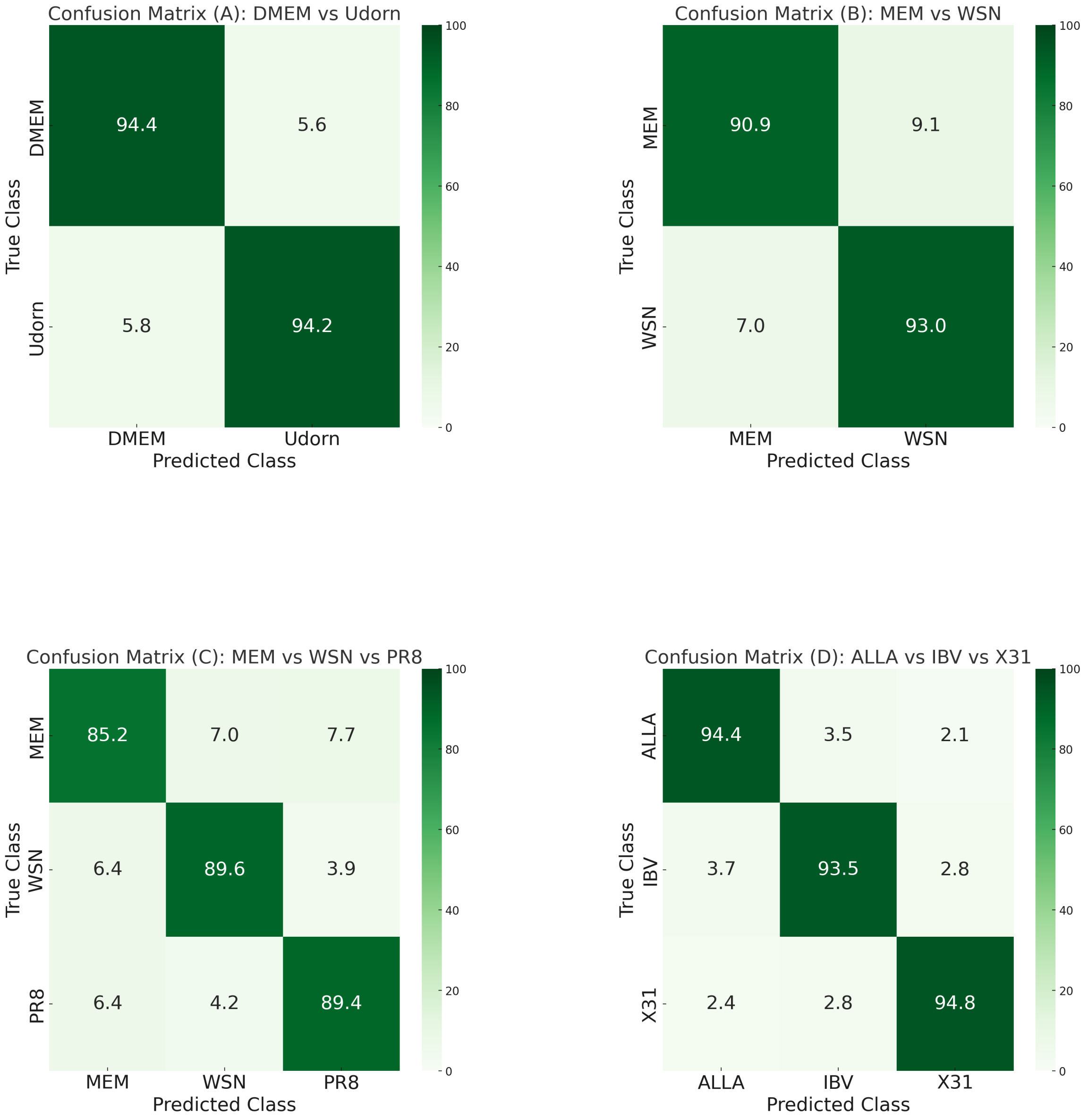


Figure 6.4: Confusion matrices summarizing the performance of the Fine KNN model for classifying viral and negative samples across four datasets. Each matrix displays the average classification accuracy as percentages for DMEM vs Udorn **(A)**, MEM vs WSN **(B)**, MEM vs WSN vs PR8 **(C)**, and ALLA vs IBV vs X31 **(D)**. The diagonal entries represent true positives, while off-diagonal entries indicate misclassifications. A consistent green color gradient is used, where darker shades correspond to higher accuracy.

6.5.5 Virus Multiclassifiers

The performance of classification models for virus vs. virus datasets was evaluated using confusion matrices (Figure 6.5). Three multiclassifiers were analyzed: (1) PR8, Udorn, WSN, and X31 (**Virus A**); (2) IBV, PR8, Udorn, WSN, and X31 (**Virus B**); and (3) PR8, SC2, Udorn, WSN, and X31 (**Virus C**). These datasets were constructed to test the models' ability to differentiate between closely related virus types and subtypes.

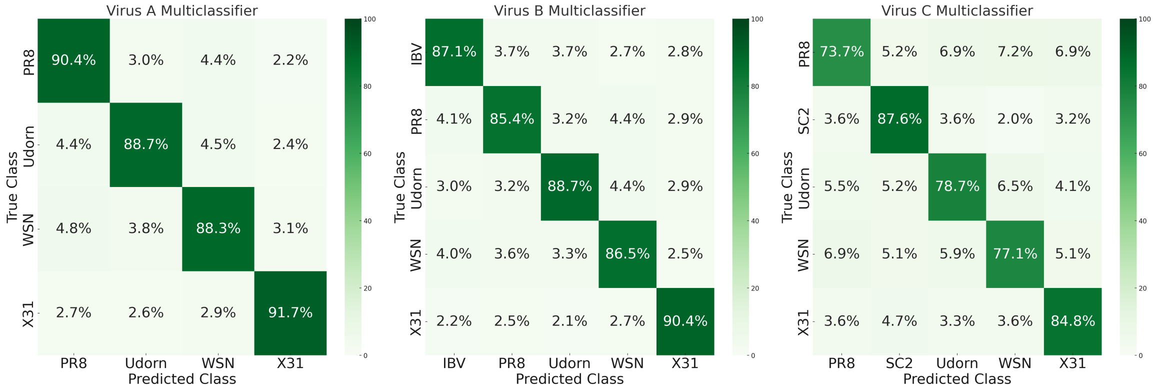


Figure 6.5: **Confusion matrices for virus vs. virus classification**. (Left) Virus A multiclassifier: PR8, Udorn, WSN, and X31. (Middle) Virus B multiclassifier: IBV, PR8, Udorn, WSN, and X31. (Right) Virus C multiclassifier: PR8, SC2, Udorn, WSN, and X31.

In the Virus A multiclassifier, each class consisted of 11,000 sub-tracks, which allowed the model to achieve high accuracy. The true positive rates exceeded 90% for most classes, with X31 having the highest accuracy (91.7%) and PR8 following closely (90.4%). Misclassification rates were minimal, with minor confusion between PR8 and WSN (4.4%) and between WSN and Udorn (3.8%). These results demonstrate the model's robustness in distinguishing influenza A subtypes when sufficient data is provided.

The Virus B multiclassifier, which introduced IBV alongside influenza A subtypes, maintained strong performance with 11,000 sub-tracks per class. IBV achieved a true positive rate of 87.1%, and the influenza A subtypes showed consistent accuracy, with X31 (90.4%) and WSN (86.5%) performing particularly well. However, the inclusion of IBV increased complexity, leading to slightly higher misclassification rates, such as between IBV and PR8 (4.1%) and between PR8 and WSN (4.4%). These results highlight the need for more nuanced differentiation as dataset complexity increases.

The Virus C multiclassifier included SC2, and each class consisted of only 3,000 sub-tracks. The reduced data volume likely contributed to the lower overall accuracy compared to the other multiclassifiers. SC2 achieved the highest true positive rate (87.6%), indicating the model's ability to effectively identify a distinct virus type. However, PR8 showed a reduced accuracy of 73.7%, with increased misclassification rates between PR8 and Udorn (5.5%) and PR8 and WSN (6.9%). The influenza A subtypes also exhibited moderate performance, with true positive rates ranging from 77.1% (WSN) to 84.8% (X31). These results emphasize the impact of data volume on classification performance, as models trained with fewer sub-tracks per class may lack the robustness seen in larger datasets.

Figure 6.5 provides a visualization of the confusion matrices, illustrating the classification performance for each multiclassifier. The lower accuracy observed in the Virus C multiclassifier, particularly for influenza A subtypes, underscores the importance of ensuring sufficient data availability for each class to maintain reliable predictions.

These findings highlight the critical role of data volume in machine learning model performance, especially in distinguishing between viruses with overlapping characteristics. Future experiments should prioritize ensuring sufficient sub-track counts for all classes to enhance model robustness and accuracy.

6.6 Testing Unseen Datasets

Testing and Comparison of Trained Models

The trained machine learning models were evaluated using independent validation sets, ensuring robust assessment of their classification performance. The results, visualized in the confusion matrices and sensitivity/specificity tables (Figure 6.6), reveal the strengths and limitations of the models across various virus classification tasks. The first classification task focused on influenza A viruses (PR8, Udorn, WSN,

Confusion Matrix 1

	PR8	Udorn	WSN	X31
PR8	123	189	107	34
Udorn	6	4061	91	31
WSN	7	192	2343	28
X31	2	142	62	1135

Confusion Matrix 2

	IBV	PR8	SC2	Udorn	WSN	X31
IBV	6913	449	11	649	577	293
PR8	442	6230	26	630	593	318
SC2	331	338	468	410	399	302
Udorn	468	324	13	9733	502	274
WSN	401	495	12	705	8085	371
X31	340	302	13	457	447	7670

Sensitivity & Specificity 1

	PR8	Udorn	WSN	X31
Sensitivity (%)	27.2	96.9	91.2	84.6
Specificity (%)	99.8	88.0	95.7	98.7

Sensitivity & Specificity 2

	IBV	PR8	SC2	Udorn	WSN	X31
Sensitivity (%)	77.7	75.6	20.8	86.0	80.3	83.1
Specificity (%)	95.2	95.4	99.8	92.6	93.7	96.2

Confusion Matrix 3

	IBV	PR8	Udorn	WSN	X31
IBV	788	4	137	79	30
PR8	43	115	152	87	24
Udorn	26	5	4034	91	22
WSN	23	8	164	2288	31
X31	15	6	97	58	1121

Confusion Matrix 4

	PR8	SC2	Udorn	WSN	X31
PR8	6528	17	594	630	357
SC2	353	479	481	444	309
Udorn	479	14	10207	572	293
WSN	456	17	782	8408	360
X31	322	16	520	549	7909

Sensitivity & Specificity 3

	IBV	PR8	Udorn	WSN	X31
Sensitivity (%)	75.9	27.3	96.6	91.0	86.4
Specificity (%)	98.7	99.7	89.6	95.5	98.7

Sensitivity & Specificity 4

	PR8	SC2	Udorn	WSN	X31
Sensitivity (%)	80.3	23.2	88.3	83.9	84.9
Specificity (%)	95.1	99.8	92.0	92.9	95.8

Figure 6.6: The confusion matrices and sensitivity/specificity tables illustrate the classification performance across different virus datasets. The total number of tracks per matrix varied: Figure 1 (8,553 tracks) classified influenza A viruses (PR8, Udorn, WSN, X31), with PR8 having 453 tracks, Udorn 4,189, WSN 2,570, and X31 1,341. Figure 2 (49,991 tracks) included IBV, influenza A viruses, and SC2, with track counts of IBV (8,892), PR8 (8,239), SC2 (2,248), Udorn (11,314), WSN (10,069), and X31 (9,229). Figure 3 (9,448 tracks) focused on IBV and influenza A viruses, with IBV (1,038), PR8 (421), Udorn (4,178), WSN (2,514), and X31 (1,297). Finally, Figure 4 (41,096 tracks) classified PR8 (8,126), SC2 (2,066), Udorn (11,565), WSN (10,023), and X31 (9,316). Reduced data volumes for SC2 in Figures 2 and 4 contributed to lowerer classification accuracy.

and X31). With a total of 8,553 tracks, distributed as PR8 (453 tracks), Udorn (4,189 tracks), WSN (2,570 tracks), and X31 (1,341 tracks), the models achieved high accuracy. X31 exhibited the highest specificity, exceeding 97%, indicating minimal misclassification as other viruses. Sensitivity values were consistently high across all viruses, highlighting the models' robustness when sufficient training data was available.

Expanding the classification to include IBV and SC2 (Figure 6.6 CM2) increased the dataset to 49,991 tracks, with IBV (8,892 tracks), PR8 (8,239 tracks), SC2 (2,248 tracks), Udorn (11,314 tracks), WSN (10,069 tracks), and X31 (9,229 tracks). The inclusion of SC2, which had a significantly smaller dataset compared to other viruses, introduced challenges, resulting in lower overall accuracy for SC2. However, IBV maintained robust sensitivity and specificity, reflecting its distinct signal characteristics.

The classification of IBV and influenza A viruses (PR8, Udorn, WSN, X31) in Figure 6.6 CM3 utilized 9,448 tracks, distributed as IBV (1,038 tracks), PR8 (421 tracks), Udorn (4,178 tracks), WSN (2,514 tracks), and X31 (1,297 tracks). Despite the smaller dataset size compared to Figure 6.6 CM2, the models demonstrated strong performance, particularly for IBV, which achieved a sensitivity of over 87%.

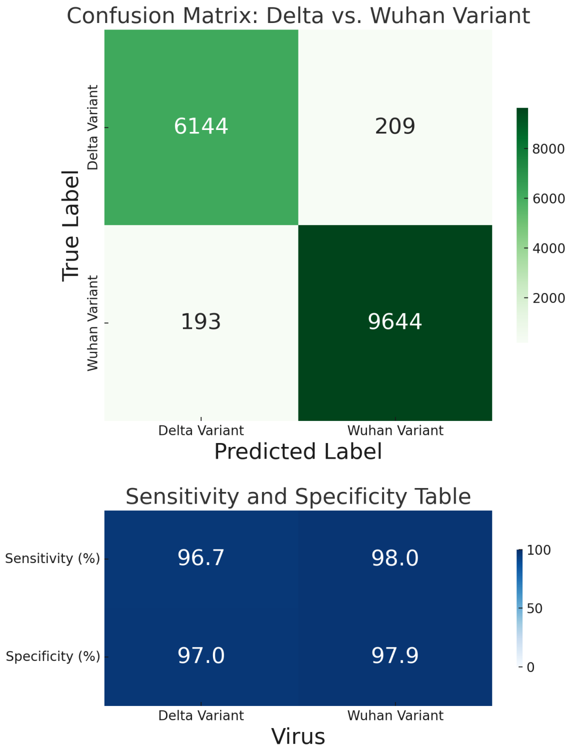
In the final task (Figure 6.6 CM4), PR8, SC2, Udorn, WSN, and X31 were classified using 41,096 tracks, with PR8 (8,126 tracks), SC2 (2,066 tracks), Udorn (11,565 tracks), WSN (10,023 tracks), and X31 (9,316 tracks). Similar to Figure 6.6 CM2, the smaller dataset for SC2 contributed to reduced sensitivity and specificity, underscoring the importance of balanced datasets for optimal classification performance.

Overall, the models performed well across all tasks, with the confusion matrices and sensitivity/specificity tables revealing high accuracy for viruses with sufficient training data. The reduced performance for SC2 highlights the challenges posed by smaller datasets, emphasizing the need for data augmentation or advanced ensemble techniques to maintain robustness in more complex multiclass scenarios.

6.7 A Preliminary Clinical Sample Classifier

The fine KNN algorithm was tested for its ability to differentiate between two strains of SARS-CoV-2: the Wuhan variant and the Delta variant, using clinical samples. Negative control samples, which contained no viruses, exhibited no tracks, resulting in a specificity of 100%. For differentiating between the two strains, the sensitivity to Delta matched the specificity for the Wuhan variant, and the positive predictive value (PPV) for Delta was equivalent to the negative predictive value (NPV) for the Wuhan variant.

A total of 64,761 sub-tracks from 7,095 distinct tracks were used in this analysis. The fine KNN algorithm achieved an overall accuracy of 97.5% (Figure 6.7). This accuracy is calculated per sub-track, meaning individual samples consisting of multiple sub-tracks would yield even higher accuracy. The workflow for this method



True Label \ Predicted Label	Delta Variant	Wuhan Variant
Delta Variant	6144	209
Wuhan Variant	193	9644

	Delta Variant	Wuhan Variant
Sensitivity (%)	96.7	98.0
Specificity (%)	97.0	97.9

Figure 6.7: Confusion matrix and sensitivity/specificity table for classifying Delta and Wuhan SARS-CoV-2 variants in clinical samples. A total of 10 Wuhan samples and 9 Delta samples were used. Correct predictions are shaded green in the confusion matrix, while the table highlights sensitivity and specificity percentages with a blue gradient (0–100%).

includes 20 seconds for sample preparation, instantaneous viral labeling, 30 seconds for imaging, and seconds to process the data through the trained model. The entire process is completed in approximately one minute. While these preliminary results are promising, they were obtained with a limited number of samples. A larger validation study is required to confirm the model's robustness and generalisability across diverse clinical datasets.

6.8 Discussion

This project demonstrates a novel approach to virus classification and strain differentiation using fluorescence imaging and machine learning, specifically fine KNN. The method relies on the principle that viruses, like charged particles, diffuse in an electrolyte solution, with their motion influenced by surface charge. This phenomenon provides a unique basis for distinguishing viral strains, as differences in surface charge affect their diffusion coefficients. By leveraging these differences, the method achieved high accuracy in multiple classification tasks, including influenza A strains (PR8, Udorn, WSN, and X31) and broader tasks involving IBV and SARS-CoV-2. For example, an accuracy of 97.5% was achieved when distinguishing between Delta and Wuhan variants of SARS-CoV-2, a particularly significant result given the similarity between these strains.

The method's strength lies in its ability to capture subtle differences in diffusion-driven motion, which reflect variations in surface charge and particle size. This is particularly valuable for differentiating closely related viral strains, where traditional methods like genomic sequencing are typically required. Despite these promising results, the approach has several limitations. As the method relies on diffusion, it is highly sensitive to environmental conditions, especially temperature. Fluctuations in temperature can directly affect diffusion rates, thereby influencing classification accuracy. Ensuring precise temperature control during sample preparation and imaging is essential for consistent results. Additionally, viral particles aggregate rapidly in solution, which alters their diffusion properties and can reduce accuracy. To mitigate this, rapid and consistent processing is critical, as any delays in the workflow could impact performance.

Another challenge is data availability, particularly for multiclass classifications involving SARS-CoV-2. For these tasks, fewer sub-tracks per class were available compared to other datasets, resulting in reduced sensitivity and specificity. Furthermore, the study relied on controlled experimental conditions and a limited number of clinical samples. Expanding the validation to diverse, real-world datasets is necessary to confirm the method's robustness and generalizability.

Several improvements could enhance this approach. Implementing precise temperature regulation during experiments would address variability due to temperature-dependent diffusion changes. Automating sample preparation and imaging steps could minimize delays and reduce the impact of aggregation. Combining diffusion-based motion with additional features, such as fluorescence intensity or particle shape, could further improve classification accuracy and robustness. Addressing data imbalances through synthetic data generation or ensemble models could also enhance performance for underrepresented classes like SARS-CoV-2. Finally, conducting validation studies with larger and more diverse datasets would be essential for establishing the clinical applicability of this method.

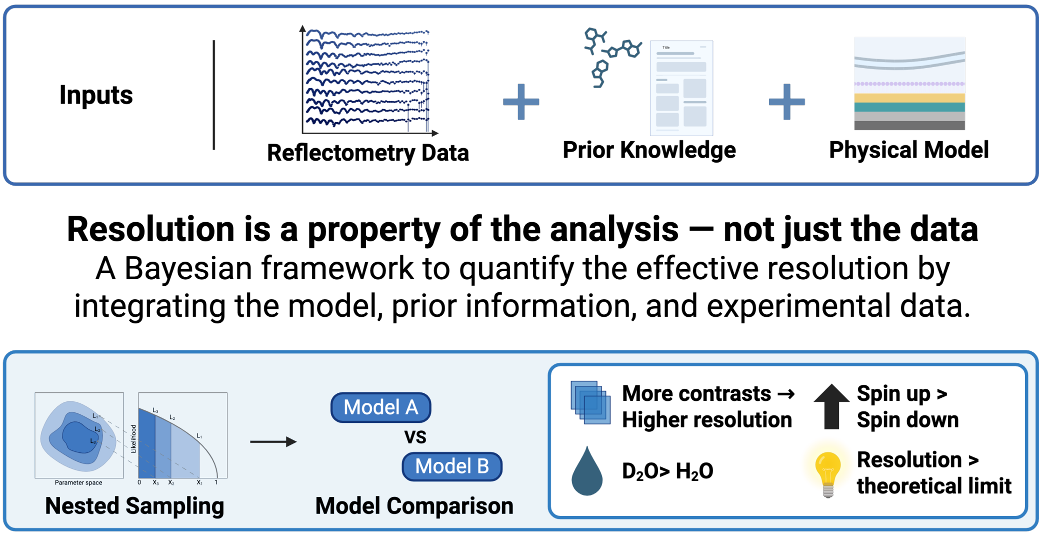
In conclusion, this project highlights a rapid and innovative approach to virus classification based on diffusion properties influenced by surface charge. The fine KNN algorithm demonstrated remarkable accuracy, particularly for datasets with sufficient and balanced training data. By offering a fast and cost-effective alternative to traditional diagnostics, such as genomic sequencing, this method has the potential to transform clinical diagnostics and epidemiological monitoring. However, further optimization and extensive validation are required to address environmental sensitivities, data constraints, and generalizability, paving the way for the widespread adoption of this promising diagnostic tool.

Chapter 7

Effective Resolution of Neutron reflectometry

> *"I am afraid neutrons will not be of any use to any one."*
>
> Sir James Chadwick

7.1 Graphical Abstract



Resolution is a property of the analysis — not just the data

A Bayesian framework to quantify the effective resolution by integrating the model, prior information, and experimental data.

Figure 7.1: A Bayesian framework integrates reflectometry data, prior knowledge, and a physical model to quantify effective analytical resolution. Resolution is determined by the full analysis—not just the data—and is evaluated through nested sampling and Bayes factor–based model selection. Increasing the number of contrasts improves resolution, with $D_2O$and spin-up Permalloy outperforming H_2O and spin-down. Notably, the effective resolution achieved exceeds the theoretical limit, highlighting the strength of model-informed Bayesian analysis in neutron reflectometry. *Figure created with BioRender.com.*

7.2 Introduction

In the elastic scattering regime, it is commonly accepted that the minimum resolvable length scale, L_{min}, is defined by the maximum measured scattering vector, Q_{max},

$$L_{\mathrm{min}} = \frac{2\pi}{Q_{\mathrm{max}}}. \tag{7.1}$$

This limit arises from the observation that the differential cross section for elastic scattering is related to the scattering length density (SLD) profile through a Fourier transform, the magnitude of which will decay to zero at scattering vectors of $2\pi/L_{\mathrm{min}}$. The observation, itself, is dependent on the Born approximation [186], which assumes that there is a single scattering event between the probing radiation and atoms which scatter them.

Neutron and X-ray reflectometry are elastic scattering techniques that are important in the study of interfacial structures, such as model membranes [187, 188], surfactant monolayers [189], and magnetic multilayers [190]. As with other reciprocal space measurements, the analysis of reflectivity data presents a significant challenge due to the phase problem [191]. Frequently, it is assumed that reflectivity measurements are limited in resolvable resolution by the same Eqn. 7.1. It is well documented [192, 193], however, that the Born approximation does not hold for reflectometry measurements. The reason that the Born approximation does not hold for reflectometry appears to be a combination of multiple factors, typically related to the scattering geometry of the measurement.

Measurements governed by the Born approximation can be analysed by an inverse Fourier transform, e.g., Fourier inversion of small angle scattering [194]. The same approach has been applied to reflectometry data [195, 196, 197], overlooking the collapse of the Born approximation already discussed. The more common approach to reflectivity data analysis is a model-dependent approach [198], where a model system is proposed and model reflectivity calculated, using the optical matrix formalism, and the model is then refined against the experimental data.

It is still regularly suggested that the maximum Q-value measured limits the length scale probed by reflectometry, despite evidence that the Born approximation does not apply in the reflection geometry. In this work, we show, analytically, that this length scale limitation does not hold in the analysis of neutron reflectometry measurements when a model-dependent approach is used. To achieve this, we establish an analysis resolution metric, through the robust analysis of neutron reflectivity data by nested sampling [199]. This metric quantifies the minimum length scale, for a given model parameter, that can be observed.

We concentrate on a floating bilayer system [200], a widely used and well-understood model system for studying lipid membranes. It is shown that the analytical resolution, instead of depending on the maximum Q-value, is related to both the data and the analytical model. In particular, we investigated the effect of isotopic/spin contrasts and model assumptions on the analytical resolution. By providing a better understanding of the influence of the collected data and the analytical model on the resolution, we believe this work can inform the development of more comprehensive models for interpreting neutron reflectivity data. Specifically, our analytical methodology, which demonstrates the enhanced resolution achievable through model-dependent analysis, can guide the refinement of these models. Furthermore, our findings can aid in the design of experiments by offering insights into how data collection and model assumptions impact the minimum resolvable length scale.

7.3 Materials and Methods

7.3.1 Materials

DPPC (1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine) was obtained from Avanti polar lipids (Alabaster, AL, USA) and used without further purification. The self-assembled monolayer material (TAAA-SAM; $HS-CH_2-(CONH)_{15}-CH_2(OH)_2$) was obtained from Prochimia surfaces (Gdansk, Poland). Deuterium oxide (D_2O), HEPES buffer salts, and all other chemicals were sourced from Sigma-Aldrich or Fisher Scientific (Loughborough, UK). Silicon substrates were obtained from Crystran (Poole, UK).

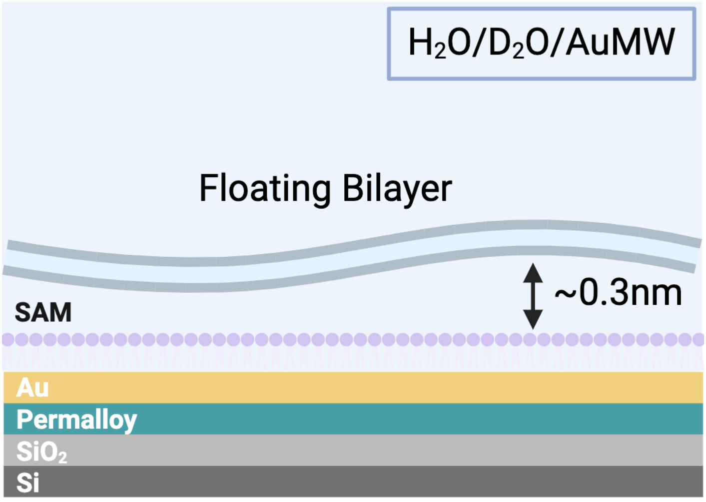


Figure 7.2: Schematic of the sample that was measured, showing the construction of the-floating bilayer system.

7.3.2 Sample preparation

In this work, we focus on a floating bilayer system, which has been extensively studied [200, 201, 202]. The floating bilayer system was prepared as in the work of Clifton *et al.* [203]. The resulting sample consisted of a substrate of silicon/silicon oxide with a permalloy and a gold layer. On top of this substrate, the self-assembled moonolayer of the TAAA-SAM material was deposited. The floating bilayer was then formed adjacent to this by Langmuir-Blodgett/Langmuir-Schaefer deposition (Fig. 7.2). The system was measured under solution in custom-built solid/liquid flow cells.

7.3.3 Neutron reflectometry measurements

Polarised neutron reflectometry (PNR) measurements where conducted at the POLREF reflectometer at the ISIS Neutron and Muon Source. This instrument measures the reflection of a white neutron beam and examines the reflection of a single neutron spin state from the sample. Polarisation was achieved using a polarizing mirror and a spin flipper. Reflectivity data was gathered across a Q_z range of approximately (0.01, 0.3) Å^{-1} (at 2.5 % resolution) using glancing angles of (0.25, 0.5, 1.25, and 2.5)° with neutron wavelengths of (2, 12) Å. The total illuminated sample length was $60mm$.

The floating bilayer system was analyzed using six distinct isotopic/spin con-

	Sample		**Permalloy spin state**		**Water**		
Contrast	**Substrate**	**Substrate + bilayer**	**po**	**mo**	**H2O**	**D2O**	**AuMW**
H_2O/po	●		●		●		
H_2O/mo	●			●	●		
D_2O/po	●		●			●	
D_2O/mo	●			●		●	
H_2O/po/h-DPPC		●	●		●		
H_2O/mo/h-DPPC		●		●	●		
D_2O/po/h-DPPC		●	●			●	
D_2O/mo/h-DPPC		●		●		●	
AuMW/po/h-DPPC		●	●				●
AuMW/mo/h-DPPC		●		●			●

Table 7.1: The isotopic/spin contrasts that gave rise to the ten neutron reflectometry measurements that make up the data set investigated herein.

trasts, and an additional four contrasts were measured for the substrate. This was accomplished by adjusting the magnetic orientation of the permalloy layer with an external magnetic field, which allowed the neutron spin to align either parallel (p) or anti-parallel (m) to the field. Furthermore, the surrounding medium was varied using either H_2O,D_2O, or gold-contrast matched water (AuMW) to achieve different contrast conditions (Tab. 7.1). A liquid chromatography pump (Knauer Smartline 1000) was connected to the liquid cell inlet for programmable control of the H2O/D2O solution mixture in the sample cell.

7.3.4 Analytical model

The model-dependent analysis of neutron reflectivity data is performed by applying the Abelès optical matrix formalism [204] to a structure of layers with given scattering length densities and thicknesses with interfacial roughnesses between the layers [205]. Here, we parameterise an analytical model that describes the floating bilayer system that consists of the layers shown in Figure 7.2. Table 7.2 gives the parameter prior distributions for those that were allowed to vary.

The substrate layers shown in Figure 7.2 were described with thicknesses (with the exception of the semi-infinite silicon layer) and scattering length densities and

	Parameter	Prior Range	Prior distribution	Mean	Std
1	SiO2 SLD	[3.41, 7.11]	Truncated normal	4	2
2	SiO2 thickness	[12.00, 18.00]	Truncated normal	15	1
3	SiO2 roughness	[12.00, 18.00]	Truncated normal	15	1
4	Permalloy spin up SLD	[9.00, 12.00]	Uniform	-	-
5	Permalloy spin down SLD	[6.00, 10.00]	Uniform	-	-
6	Permalloy thickness	[100.00, 200.00]	Uniform	-	-
7	Permalloy roughness	[5.00, 11.00]	Uniform	-	-
8	Au thickness	[100.00, 200.00]	Uniform	-	-
9	Au roughness	[4.00, 8.00]	Uniform	-	-
10	SAM Area Per Molecule (APM)	[15.00, 30.00]	Truncated normal	23	1.5
11	SAM tail thickness	[15.00, 20.00]	Uniform	-	-
12	SAM head thickness	[7.00, 11.00]	Truncated normal	9	0.7
13	SAM hydrtation	[0.00, 1.00]	Truncated normal	0.5	1.5
14	Water Interlayer thickness	[0.10, 5.00]	Truncated normal	0.5	1
15	Bilayer tail volume	[600.00, 1000.00]	Uniform	-	-
16	Bilayer head volume	[300.00, 380.00]	Uniform	-	-
17	Bilayer defect hydration	[0.00, 1.00]	Truncated normal	0.1	0.5
18	Bilayer head hydration	[0.10, 1.00]	Truncated normal	0.5	1

Table 7.2: A list of all the parameters used in the model with the corresponding prior range, and prior distribution.

interfacial roughnesses between each. The SAM layer is split into separate heads and tails layers, and the scattering length densities, β, for these layers were constrained based on the area per molecule (APM) of the SAM with the following relation

$$\beta = \frac{b}{d\text{APM}} + \beta_{\text{sol}}(1 - \phi), \tag{7.2}$$

where, b is the sum of the scattering lengths of the atoms that make up either the head or the tail, d is the thickness of the head or tail layer, ϕ is the volume fraction of the head or tail material in the solvent of scattering length density β_{sol}. By having the same APM for both the head and tail, the chemical constraint that there are the same numbers of head and tail groups is achieved. The DPPC bilayer is described using two pairs of similar components, placed such that the tail layers are adjacent. However, unlike the SAM layer, the components are described in terms of the molecular volumes for the head and tail groups. This means that in order to introduce the chemical constraint to ensure an equal number of head and tail groups the tail thickness, d_{tail} is constrained as follows,

$$d_{\mathrm{tail}} = \frac{d_{\mathrm{head}} \phi_{\mathrm{h}} V_{\mathrm{tail}}}{\phi_{\mathrm{tail}} V_{\mathrm{head}}}, \tag{7.3}$$

where, $d$head, ϕ_{head}, and V_{head} are the thickness, volume fraction, and molecular volume of the head group respectively, and similarly for the tail group parameters. Between the SAM and bilayer, there is a narrow interlayer of water and above the bilayer a semi-infinite water layer.

In addition, interfacial roughness, σ, is present at each of the three interfaces (solvent-head, head-tail, and tail-air), and is modeled using an error function. Notably, in the study by Campbell et al, the roughness was assumed to be conformal, such that it does not vary between interfaces. This assumption is reasonable in the case of a monolayer of a single lipid type which is the case for our system.

We have also used the area per molecule assumption, a common simplifying assumption used in the study of monolayers and thin films. It assumes that all molecules in the monolayer or thin film have the same cross-sectional area and that the total area of the monolayer or thin film is simply the product of the number of molecules and their individual cross-sectional area.

The area per molecule assumption is a useful simplification because it allows for straightforward calculations of the thickness and packing density of the monolayer or thin film from measurements of its total area and the number of molecules present.

7.3.5 Bayesian model selection framework

Bayesian model selection is a statistical framework used to compare and select the most suitable model from multiple options based on how well they fit the observed data. This process involves evaluating the relative likelihoods of different models, considering their complexity, and then choosing the model with the highest posterior probability.

This selection process is based on Bayes' theorem, which provides a way to calculate the posterior distribution - the probability of the model parameters given the observed data. The theorem combines the likelihood of the data given the model parameters and the prior probability of the parameters.

The evidence for the data given in our model, which is integral over all possible parameter combinations, is a crucial part of this calculation. This evidence can be efficiently estimated using nested sampling[199], a Monte Carlo method. Nested sampling[206] provides a measure of how well the model fits the data, taking into account its complexity.

We then use Bayes factors[207], a popular method for comparing the relative likelihoods of different models in the Bayesian model selection framework. The Bayes factor between the two models is the ratio of their evidence, calculated on the same dataset. It is important to note that in the case where our *a priori* belief in the two models is equal, the evidence ratio, called the Bayes factor, completely specifies the relative probability of the two models.

Considering a case where we have two competing models M_1 and M_2, both of which describe our system. Given some data $\mathbf{D}$ the Bayes factor is defined as [208],

$$BayesFactor(BF) = \frac{P(M_1|D,I)}{P(M_2|D,I)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_2}\frac{P(M_1|I)}{P(M_2|I)}. \tag{7.4}$$

We select a parameter of interest, in this case the bilayer tail thickness, and run the nested sampling algorithm for a range of possible values, resulting in the evidence values associated with every parameter value. It should be noted that two models can only be compared for the same data (Fig. 7.3 ③). In this work we are working with log evidence values and every evidence value is normalised by the highest evidence for the particular set of data. This results in the logarithmic Bayes factors Fig. 7.3 ③).

$$\ln{(BF_{i,max})} = \ln{(\frac{\mathcal{Z}_i}{\mathcal{Z}_{max}})} = \ln{(\mathcal{Z}_i)} - \ln{(\mathcal{Z}_{max})} \tag{7.5}$$

This allows us to define the analytical resolution of the technique as the number of consecutive parameter values with a Bayes factor smaller than 4.61 (which is the most conservative value we can choose, see Table 7.3). This definition allows us to

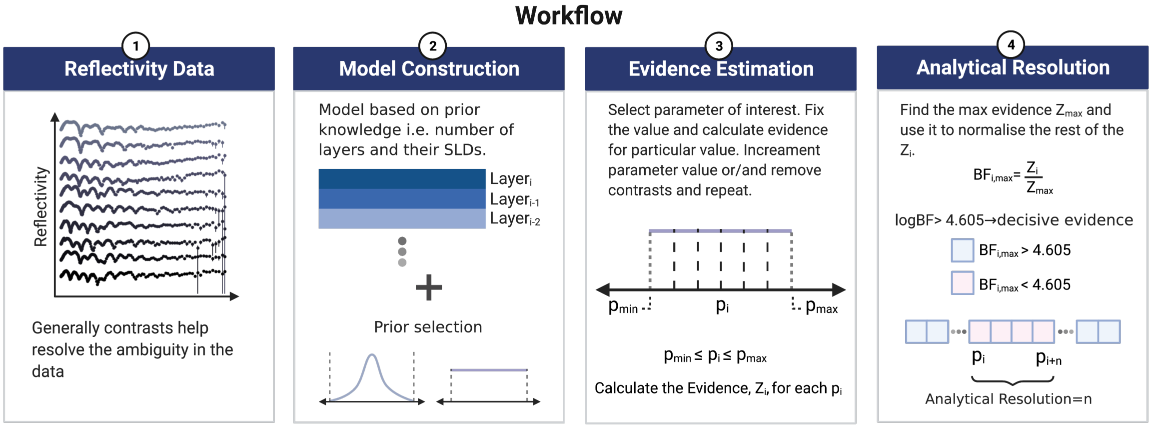


Figure 7.3: **Analytical Resolution Estimation Workflow.**
① Optimally good data set of a well-understood system, with H_2O, D_2O, $AuMW$ and spin contrasts.
② Code model creating layers with SLD, thickness and roughness value-ranges found in the literature, using all or part of the contrasts available.
③ Estimate the evidence for incremental values of the parameter of interest.
④ Analytical resolution calculation based on bayes factors.

Table 7.3: Interpretation of Bayes factor, $BF_{i,max}$ between the i^{th} and highest evidence model.

$BF_{i,max}$	$\ln(BF_{i,max})$	Interpretation
> 100	> 4.61	Decisive evidence for p_{max}
$30 - 100$	$3.40 - 4.61$	Very strong evidence for p_{max}
$10 - 30$	$2.30 - 3.40$	Strong evidence for p_{max}
$3 - 10$	$1.10 - 2.30$	Anecdotal evidence for p_{max}
$1 - 3$	$0 - 1.10$	Not worth more than a mention

quantify the analytical resolution and it is a metric of the effect of the chosen model (external information about the system) on the resulting resolution that accounts for both the data and the model.

7.4 Results

7.4.1 Neutron reflectometry data analysis

Figure 7.4 (A) presents a comparison between the experimental data and the reflectometry profiles obtained for the tail thickness parameter value with the highest

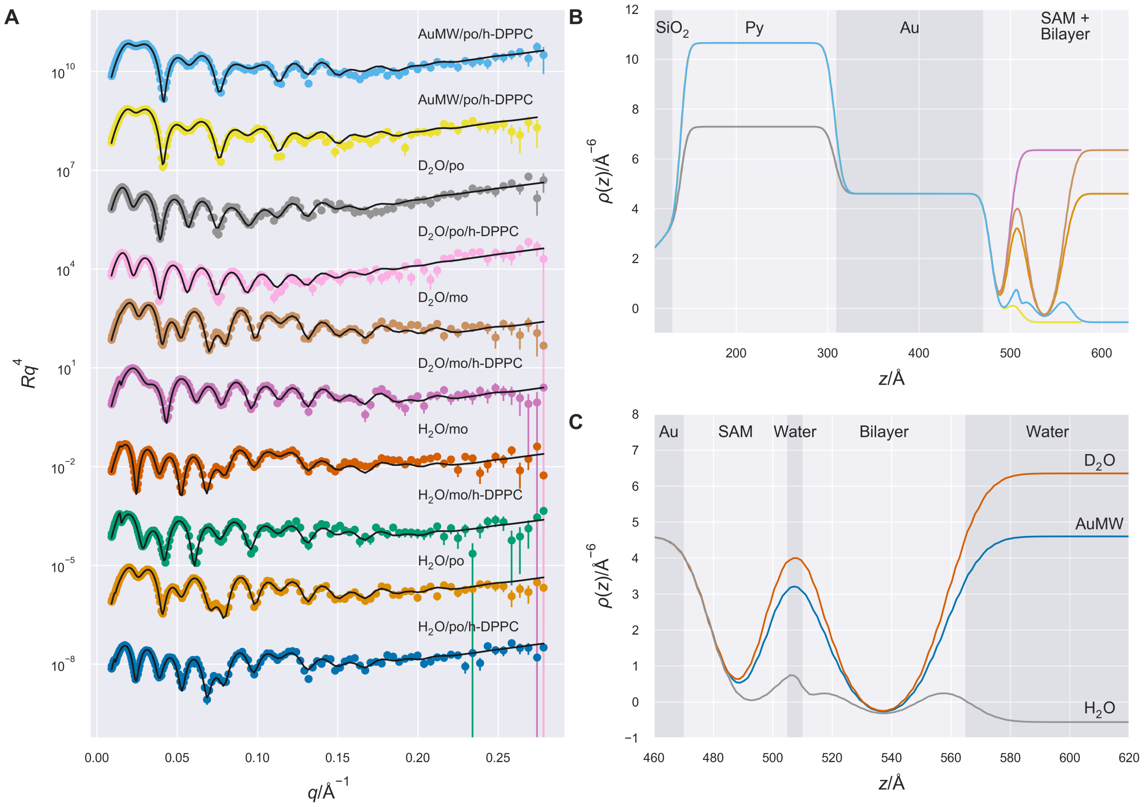


Figure 7.4: Reflectivity curve and corresponding SLD profiles. **(A)** The experimental (coloured lines) reflectometry and the median values for the model with the greatest evidence (black lines) for the tail thickness parameter. The different contrasts are offset by an order of magnitude in reflected intensity. **(B)** SLD profile for the substrate and the floating bilayer system. **(C)** Zoomed in SLD profile for the SAM and floating bilayer.

evidence when all contrasts are utilised in the analysis. Across all contrasts, a clear agreement is evident between the experimental data and the model predictions. This observation confirms the validity of the assumptions discussed in Section 2.4.

7.4.2 Effective resolution for a Single Dataset

Our investigation into the effective resolution of neutron reflectometry for a floating bilayer system illustrates the potential for empirical data to surpass the boundaries dictated by theoretical approximations. As can be seen in figure 7.5 for the tail parameter for a single contrast the worst effective resolution achieved is 10 Å and for the head parameter 7 Å which in both cases is better than the theoretical limit. For reference the literature value for the tail thickness is 18 Å and for the head thickness

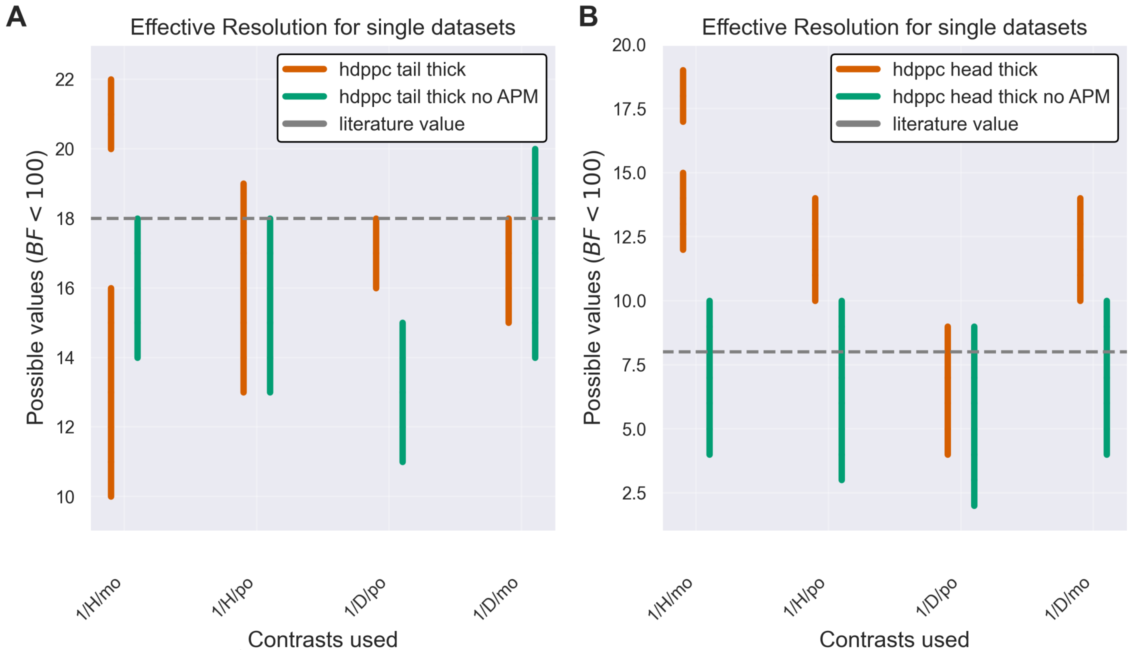


Figure 7.5: Plots of the Bayes factors below the decisive evidence value for single contrasts. **(A)** Indecisive evidence plot for the bilayer tail thickness and **(B)** the bilayer head thickness with and without the APM constraint.

is 8 Å[209].

The complex nature of the system under study, the occurrence of multiple scattering events, and the application of model fitting to the experimental data all contribute to an effective resolution that exceeds the limit suggested by the Born approximation (see equation 7.1).

7.4.3 Effect of increasing contrasts on the effective resolution

The analytical resolution for both the tail and head thickness parameters (see Figure 7.6) improves for an increasing number of contrasts. This makes sense intuitively as with more isotropic contrasts the parameter space gets smaller. This is true for both the APM and not APM restricted model.

7.4.4 Effect of removing APM assumption from the model

For the tail thickness parameter (Figure 7.6 A) the average value for most contrast combinations is higher when the APM assumption is coded into the model. On the other hand, the analytical resolution is worse when the APM assumption is used. This

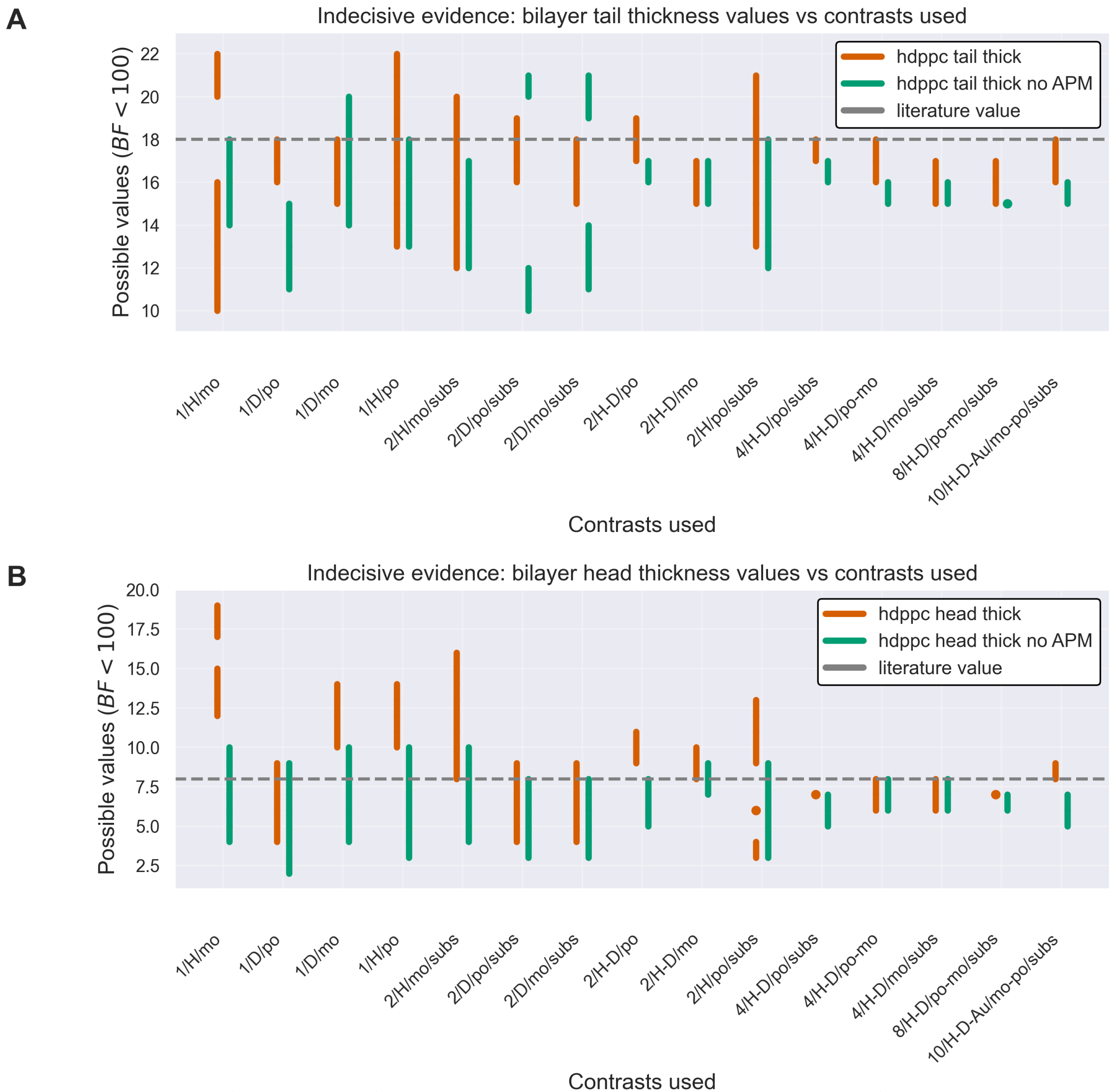


Figure 7.6: Plots of the Bayes factors below the decisive evidence value for each combination of contrasts. **(A)** Indecisive evidence plot for the bilayer tail thickness and **(B)** the bilayer head thickness with and without the APM constraint.

		Sample		Permalloy spin state		Water		
Name	**Num. contrasts**	**Substrate**	**Substrate + bilayer**	**Up**	**Down**	**H2O**	**D2O**	**AuMW**
10/H-D-Au/po-mo/subs	10	•	•	•	•	•	•	•
8/H-D/po-mo/subs	8	•	•	•	•	•	•	
4/H-D/po/subs	4	•	•	•		•	•	
4/H-D/mo/subs	4	•	•		•	•	•	
4/H-D/po-mo	4		•	•	•	•	•	
2/H/po/subs	2	•	•	•		•		
2/D/po/subs	2	•	•	•			•	
2/H-D/po	2		•	•		•	•	
2/H/mo/subs	2	•	•		•	•		
2/D/mo/subs	2	•	•		•		•	
2/H-D/mo	2		•		•	•	•	
1/H/po	1		•	•		•		
1/D/po	1		•	•			•	
1/H/mo	1		•		•	•		
1/D/mo	1		•		•		•	

Table 7.4: Details of the combinations of contrasts used in the analysis.

could be because the head thickness is calculated from the APM and tail thickness parameter when the APM assumption is used. Since the tail thickness is a well-encoded parameter whereas the head thickness is a poorly encoded parameter calculating one of the two combines their uncertainties. This is supported by Figure 7.6 B, where we can see that the analytical resolution is better when the APM assumption is used. For both parameters, the average values estimated by the model are closer to the values reported in the literature[209] when the APM assumption is used.

7.4.5 Comparison of H2O vs D2O contrasts

The analytical resolution for both the tail and head thickness parameters (see Figure 7.7 A and C) is generally better for D_2O than H_2O whereas when the APM assumption is not used the resolution is broadly similar for both D_2O and H_2O contrasts. The possible values for D_2O and H_2O are different likely due to the difference in hydration parameters for the two cases.

7.4.6 Comparison of spin up vs down Permalloy

The analytical resolution for both the tail and the head thickness parameters (see Figure 7.8 A and C) is better when the Permalloy is in the spin-up state. When the

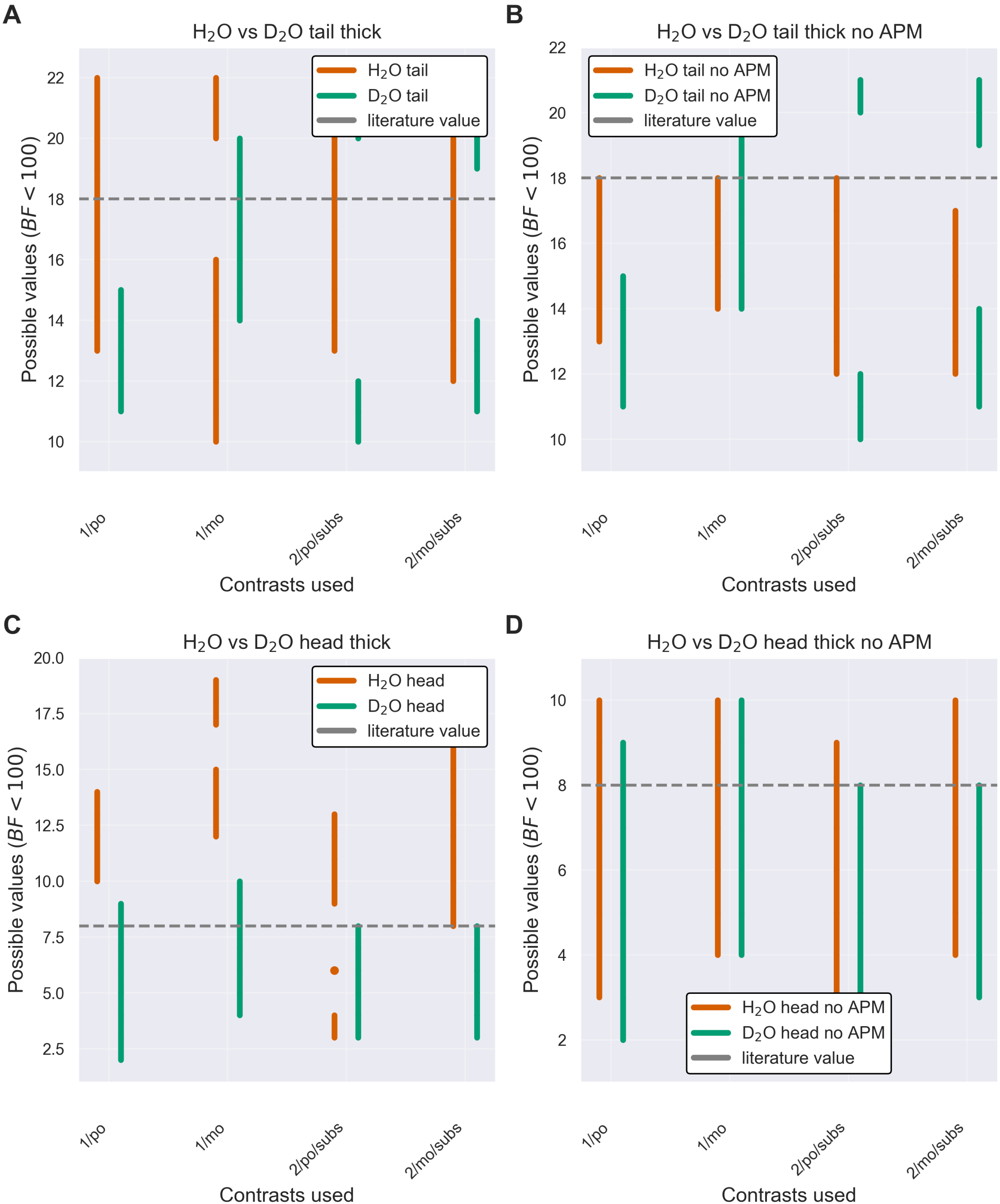


Figure 7.7: Comparison plots of the Bayes factors below the decisive evidence value for H_2O vs D_2O contrasts only. Indecisive evidence plot for the **(A)** bilayer tail thickness with the APM constraint and **(B)** without, **(C)** the bilayer head thickness with the APM constraint and **(D)** without.

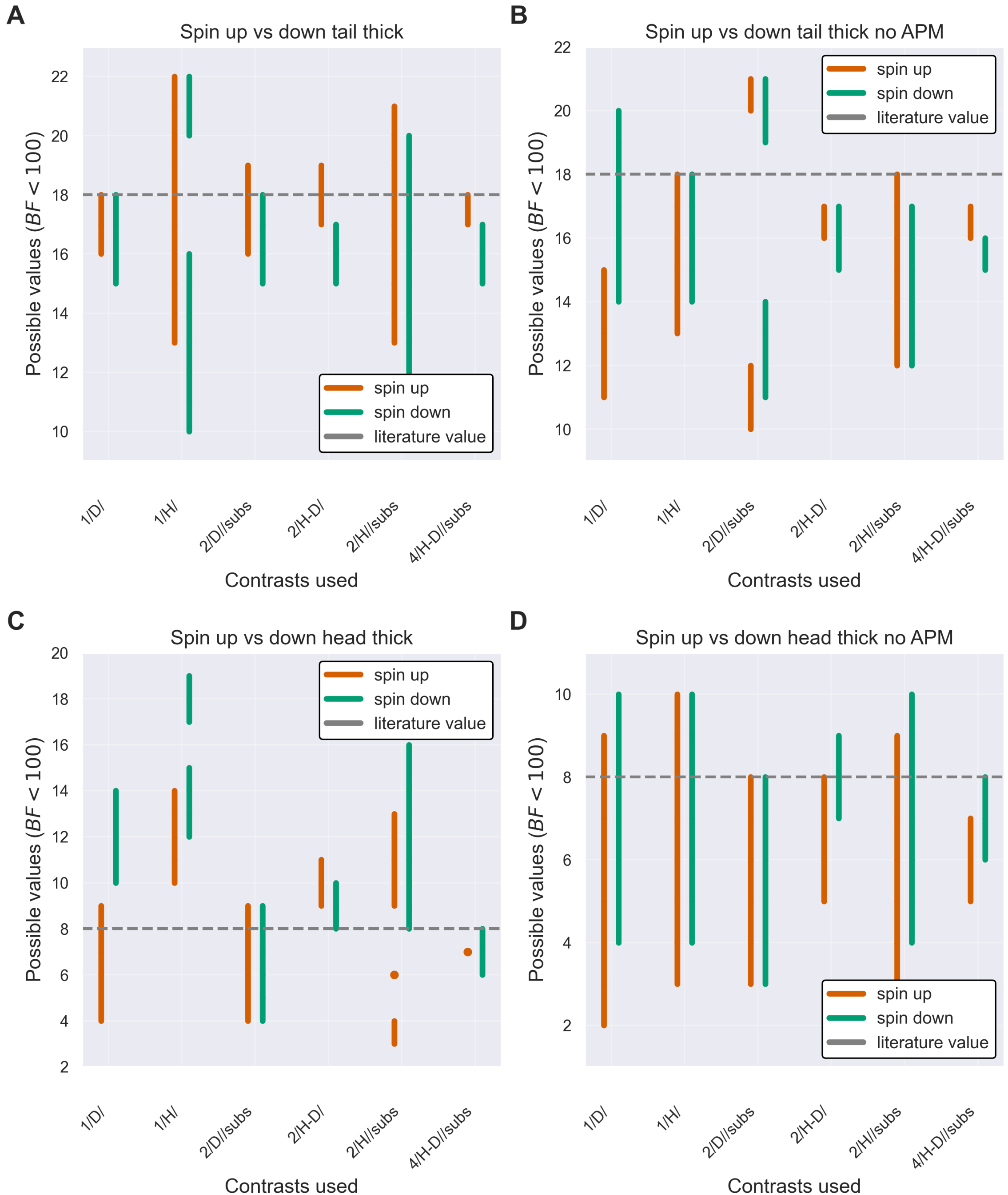


Figure 7.8: Comparison plots of the Bayes factors below the decisive evidence value for the permalloy spin up vs down contrasts only. Indecisive evidence plot for the **(A)** bilayer tail thickness with the APM constraint and **(B)** without, **(C)** the bilayer head thickness with the APM constraint and **(D)** without.

Permalloy is in the spin-up state the SLD difference between the adjacent SLD layers is greater than when the Permalloy is in the spin-down state. This implies that a greater difference in the SLDs of the substrate layers positively impacts the resolution of the parameters for the system under study. On the other hand, when the APM constraint is used (see Figure 7.8 B and D) there is no general trend of improved resolution for either spin state.

7.5 Discussion

The findings of our study demonstrate the potential and complexities of applying Bayesian analysis in neutron reflectometry data analysis to understand the analytical resolution for a specific experimental setup and sample. Bayesian analysis enables a robust comparison of different models by quantifying the evidence supporting each one. This is achieved using Bayes factors, which provide a comprehensive measure of the relative probabilities of different models, given an equal a priori belief in the two models. By following the workflow established in this work we can get the analytical resolution for the specific data, model and prior information.

Our work confirms the significant role of contrasts in the analysis process. We found that the analytical resolution for both the tail and head thickness parameters improves as the number of contrasts increases. This intuitively makes sense, as more contrasts effectively reduce the parameter space, leading to more precise estimation.

The Average Per Molecule (APM) assumption introduced an intriguing dynamic to the analytical resolution. While the APM assumption appears to worsen the resolution for the tail thickness parameter, it conversely improves the resolution for the head thickness parameter. This counterintuitive relationship can be attributed to the combination of uncertainties when one parameter is calculated from the other. Our findings suggest that careful consideration must be given to such model assumptions, as they can significantly impact the resolution.

We also found that the choice of contrast has a substantial effect on analytical resolution. Specifically, D2O contrasts generally provided better analytical resolution than H2O contrasts. This finding underscores the importance of careful contrast selection in neutron reflectometry studies and encourages further investigation into the properties and behaviour of these contrasts.

Finally, our study revealed a complex interplay between the permalloy spin states and the analytical resolution. Generally, the resolution was better when the Permalloy was in the spin-up state, suggesting that a greater difference in the scattering length densities of the substrate layers can positively impact the resolution. However, when the APM constraint was not used, no general trend of improved resolution was observed for either spin state. This finding invites further exploration into the interrelationship between spin states and model constraints, and how they collectively influence the resolution.

Perhaps the key takeaway of this work is that the analytical resolution in neutron reflectometry is dependent on the analysis. Model assumptions, multiple contrasts, substrate choice, and prior information all have an impact on the analytical resolution which as demonstrated above can be better than the common knowledge 7.1.

In conclusion, our study illuminates the intricate factors that can impact the effectiveness and accuracy of Bayesian analysis in neutron reflectometry data analysis. By shedding light on the implications of contrast selection, model assumptions, and spin states, our findings can inform more accurate and reliable model selection in future studies. It is anticipated that continued exploration of these factors will significantly advance our understanding and application of neutron reflectometry.

Chapter 8

Conclusion and Future Directions

8.1 Discussion and Implications

The diagnostic approaches developed in this thesis highlight a transformative shift in virus detection and characterization. Traditional methods usually rely on specific molecular interactions to identify pathogens, requiring centralized laboratories, specialized equipment, and trained personnel. These limitations hinder timely medical decision-making, particularly during acute conditions like viral outbreaks or pandemics. In contrast, the methods presented here propose a novel paradigm: labeling all components within a sample indiscriminately and achieving specificity through advanced computational analysis. This approach offers a scalable, adaptable, and efficient alternative to conventional diagnostics.

Central to this thesis is the refinement and validation of a cation-mediated labeling technique, enabling the attachment of fluorescently labeled single-stranded DNA to the lipid membranes of enveloped viruses. This method demonstrated versatility across diverse experimental conditions, including different viral strains, pH levels, and ion concentrations. It was also effective with both clinical and laboratory-grown samples. When combined with fluorescence microscopy, the labeling process facilitated rapid immobilization and visualization of viral particles, streamlining single-particle analysis.

Machine learning, particularly convolutional neural networks (CNNs), played a crucial role in enhancing specificity post-labeling. These models accurately distinguished virus-positive from virus-negative samples and differentiated closely related viral strains, independent of imaging conditions and optical setups. By shifting specificity from molecular interactions to computational analysis, this approach redefines diagnostic frameworks, making them more flexible and less dependent on complex molecular assays.

While fluorescence microscopy, particularly Total Internal Reflection Fluorescence (TIRF), was essential for high-resolution imaging in this research, its cost, complexity, and reliance on precision optics limit its feasibility for point-of-care (POC) applications. To address this, a diffusion-based detection approach was explored as an alternative. This method leverages particle mobility and single-particle tracking, eliminating the need for expensive imaging systems. It aligns closely with the requirements of scalable, field-deployable diagnostics and holds promise for integration with low-cost platforms such as smartphone-based microscopy. However, further research is needed to determine whether this approach can achieve the same sensitivity and specificity as fluorescence-based methods in clinical settings.

In comparison to existing diagnostic tools, such as RT-PCR and lateral flow assays, the methods proposed here offer unique advantages. RT-PCR, while highly sensitive and specific, requires centralized infrastructure and lengthy workflows, making it impractical for rapid, decentralized testing. Lateral flow assays, though affordable and portable, often lack the sensitivity needed for early-stage infections. The proposed paradigms—indiscriminate labeling combined with machine learning—strike a balance between adaptability and performance. However, the current detection limit of $\sim 10^4$ PFU/mL, though clinically relevant, remains higher than RT-PCR's sensitivity, highlighting an area for optimization.

Beyond clinical diagnostics, the approaches developed in this thesis have potential applications in environmental monitoring, food safety testing, and biomanufacturing quality control. The diffusion-based detection method, in particular, offers significant potential for democratizing diagnostics, enabling low-cost, scalable solutions that could support citizen science initiatives and decentralized pathogen surveillance.

Despite their promise, these methods face challenges in transitioning from laboratory settings to real-world applications. Fluorescence-based systems, while effective, must be simplified to reduce cost and complexity. Diffusion-based detection, while inherently low-cost, requires further validation to ensure its robustness in diverse clinical and environmental samples. Addressing these limitations will be critical to achieving the scalability and affordability necessary for global deployment.

One notable limitation of this work is the focus on single-virus detection without explicitly considering the challenge of co-infections, where multiple viral species may be present simultaneously in a sample. In real-world clinical scenarios, co-infections are common and can significantly impact both the accuracy and interpretation of diagnostic results. The indiscriminate labeling strategy combined with machine learning classification has demonstrated high specificity in controlled conditions; however, it remains to be determined how well these models can distinguish and correctly classify multiple co-circulating viruses within the same sample. Future studies will need to evaluate the system's performance in complex clinical samples with mixed infections to ensure diagnostic accuracy and to avoid potential misclassification or underreporting of infections.

By shifting diagnostic specificity from molecular interactions to computational analysis, this work lays the foundation for a new generation of diagnostics that are rapid, adaptable, and accessible. These methods not only address the limitations of traditional diagnostics but also open new avenues for scalable, decentralized, and impactful healthcare solutions. With continued research and development, the approaches described here have the potential to transform diagnostics across clinical, environmental, and industrial domains.

8.2 Future Directions

8.2.1 Refining and Expanding Algorithms

Enhancing the diagnostic accuracy and adaptability of the machine learning models developed in this thesis requires optimizing computational approaches for fluorescence-based detection systems. Lightweight algorithms, such as streamlined convolutional neural networks (CNNs), should be refined to maintain high accuracy while efficiently operating on low-cost hardware. Advanced techniques, such as attention-based mechanisms, can enhance the system's ability to focus on diagnostically relevant features, improving classification even in noisy or incomplete datasets.

Incorporating temporal data, such as changes in fluorescence intensity or diffusion patterns over time, offers another avenue for enhancing diagnostic specificity. Leveraging pathogen behaviors dynamically, without additional hardware complexity, preserves the system's simplicity and cost-effectiveness while improving its ability to differentiate pathogens in challenging samples.

Rather than integrating fluorescence microscopy data with expensive complementary modalities, such as Raman spectroscopy or mass spectrometry, future efforts should maximize the utility of fluorescence data alone. Innovations in preprocessing and augmentation—such as noise reduction, feature extraction, and synthetic dataset generation—could enhance model training while keeping the system affordable and portable for point-of-care applications.

8.2.2 Optimizing Diffusion-Based Detection for Smartphones

The diffusion-based detection method holds significant potential for integration with smartphone-based microscopy platforms. Research should focus on developing compact, low-cost imaging setups that leverage modern smartphone cameras. Innovations in lens miniaturization, such as clip-on optical modules, can improve imaging quality while minimizing costs.

Real-time data processing and analysis can be achieved through app-based software pipelines. These applications could feature optimized particle tracking algorithms and machine learning models tailored to smartphone processors. Cloud-based analysis is another avenue to handle computationally intensive tasks, ensuring rapid and reliable diagnostic results in resource-limited settings.

8.2.3 Expanding Pathogen Range

Current methods are primarily tailored for enveloped viruses. To broaden their utility, future research should adapt the cation-mediated labeling strategy to accommodate non-enveloped viruses, bacteria, and fungi. Complementary labeling agents, such as functionalized nanoparticles, could enable interactions with diverse pathogen surface chemistries.

Expanding multiplexing capabilities to detect multiple pathogens within a single sample would be invaluable, particularly in clinical settings where co-infections are common. Additionally, addressing challenges in detecting pathogens in complex matrices—such as blood, urine, or environmental samples—will require further improvements in signal-to-noise ratios and specificity.

8.2.4 Addressing Cost and Sustainability

Ensuring affordability and sustainability is crucial for scaling these diagnostic technologies. Optimizing the use of fluorescent labeling agents and exploring cost-effective alternatives will help reduce reagent expenses while maintaining sensitivity and signal stability. Developing renewable or recyclable reagents could further minimize the environmental impact.

Simplifying hardware to focus on essential components—such as compact excitation sources, affordable detectors, and efficient optical designs—will reduce costs while preserving functionality. Emphasizing modularity for scalability (e.g., adjustable field of view or sample handling capacities) rather than unnecessary dual-purpose features will also support high-throughput capabilities.

Collaboration with manufacturers and supply chain experts can optimize production and distribution, leveraging economies of scale to standardize components and lower costs. These efforts will balance performance, affordability, and accessibility, ensuring feasibility in decentralized and resource-limited settings.

8.2.5 Validation and Clinical Trials

Large-scale validation studies are vital to demonstrating the reliability and utility of these methods in real-world settings. Key performance metrics—such as sensitivity, specificity, reproducibility, and detection limits—should be evaluated across a diverse range of clinical and environmental samples, including challenging matrices like blood, urine, and saliva.

Collaborations with healthcare providers and diagnostic companies will facilitate clinical trials and regulatory approval processes. Engaging with regulatory bodies, such as the FDA and EMA, early in the development process can streamline approvals. Compliance with international standards must be prioritized to meet safety and quality requirements.

8.2.6 Designing for Scalability and Field Deployment

Diagnostic systems should be designed to meet the demands of large-scale testing, particularly during pandemics. Modular systems capable of processing high sample volumes quickly and efficiently will be essential. Portability and ease of use should also be prioritized to enable deployment in remote or resource-limited settings.

The diffusion-based detection method, in particular, is well-suited for scalability due to its reliance on simple hardware. Optimizing this method for high-throughput screening, potentially incorporating automated sample handling and imaging workflows, could further enhance its utility. Ensuring functionality under varying environmental conditions, such as temperature and humidity fluctuations, will be crucial for global deployment.

8.2.7 Exploring New Applications

While the focus of this thesis was on pathogen diagnostics, the methods developed have broader potential applications. These include environmental monitoring, such as detecting waterborne or airborne pathogens; food safety testing for microbial contamination; and real-time quality control in biomanufacturing. The diffusion-based approach, in particular, could empower citizen science initiatives and democratize pathogen surveillance.

8.2.8 Lessons Learned and Broader Implications

This work illustrates that shifting diagnostic specificity from molecular interactions to computational analysis offers a transformative framework for pathogen detection. The combination of biophysical labeling with advanced software highlights the importance of interdisciplinary innovation in addressing global health challenges.

The potential integration of diffusion-based detection with smartphone platforms represents a pivotal step toward democratizing diagnostics, making advanced tools accessible to underserved populations. These developments underscore the need for affordable, scalable, and adaptable diagnostic technologies to meet the demands of emerging diseases and pandemics. By addressing current limitations and pursuing these future directions, the findings of this thesis have the potential to revolutionize diagnostics across clinical, environmental, and industrial domains.

8.3 Conclusion

This thesis introduces a foundation for next-generation diagnostic tools that are rapid, scalable, and adaptable. While transitioning these methods from laboratory-based settings to practical point-of-care applications presents challenges, the research highlights clear pathways for overcoming them. By addressing these obstacles and advancing the proposed innovations, these technologies hold the potential to transform global diagnostics, improving healthcare outcomes and strengthening preparedness for future health crises.

Appendix A

Clinical Samples Tables

Table A.1: **Overview of clinical samples used for training and validation of the network**. Each sample was imaged across three different days (labeled as -1, -2, -3, etc.). For network training, 70% of the BBXs generated over these days were randomly selected, while the remaining 30% were reserved for validation (results shown in the confusion matrices). Abbreviations: hCoV – seasonal human coronaviruses (including NL63, OC43, and HKU1), Neg – negative, SC2 – SARS-CoV-2 (original Wuhan variant), Flu A – Influenza A. ‘Microscope 1’ was a Nanoimager with a Hamamatsu Flash 4 LT.1 sCMOS camera, while ‘Microscopes 2 and 3’ were equipped with a Hamamatsu Flash4 V3 sCMOS camera; otherwise, all systems were identical.

Sample ID	**RT-PCR**	**Total BBXs**	**Sample ID**	**RT-PCR**	**Total BBXs**
hCoV vs Negative (Microscope 1)					
T-Neg-1	Neg	1507	T-hCov-3-3	HKU1	148
T-Neg-2	Neg	562	T-hCov-4	NL63	377
T-Neg-3	Neg	451	T-hCov-4-1	NL63	80
T-Neg-4	Neg	378	T-hCov-4-2	NL63	279
T-Neg-1-1	Neg	5173	T-hCov-5	NL63	500
T-Neg-2-1	Neg	122	T-hCov-5-1	NL63	513
T-Neg-3-1	Neg	568	T-hCov-5-2	NL63	823
T-Neg-4-1	Neg	158	T-hCov-6	NL63	453
T-Neg-4-2	Neg	571	T-hCov-6-1	NL63	207
T-hCov-1	HKU1	229	T-hCov-6-2	NL63	292
T-hCov-1-1	HKU1	290	T-hCov-7	OC43	268
T-hCov-1-2	HKU1	104	T-hCov-7-1	OC43	131
T-hCov-1-3	HKU1	203	T-hCov-7-2	OC43	254
T-hCov-2	HKU1	169	T-hCov-8	OC43	897
T-hCov-2-1	HKU1	92	T-hCov-8-1	OC43	569
T-hCov-2-2	HKU1	136	T-hCov-8-2	OC43	342
T-hCov-3	HKU1	330	T-hCov-9	OC43	500
T-hCov-3-1	HKU1	918	T-hCov-9-1	OC43	145
T-hCov-3-2	HKU1	129			
Total BBXs in dataset		**18863**			
SARS-CoV-2 Wuhan vs hCoV (Microscope 1)					
T-SC2-1	SC2	696	T-hCov-3-2	HKU1	129
T-SC2-1-1	SC2	3320	T-hCov-3-3	HKU1	148
T-SC-2	SC2	370	T-hCov-4	NL63	377
T-SC-2-1	SC2	223	T-hCov-4-1	NL63	80
T-SC-2-2	SC2	131	T-hCov-4-2	NL63	279
T-SC-3	SC2	662	T-hCov-5	NL63	500

Continued on next page

Sample ID	RT-PCR	Total BBXs	Sample ID	RT-PCR	Total BBXs
T-SC-3-1	SC2	214	T-hCov-5-1	NL63	513
T-SC-4	SC2	861	T-hCov-5-2	NL63	823
T-SC-4-1	SC2	218	T-hCov-6	NL63	453
T-SC-4-2	SC2	1783	T-hCov-6-1	NL63	207
T-hCov-1	HKU1	229	T-hCov-6-2	NL63	292
T-hCov-1-1	HKU1	290	T-hCov-7	OC43	268
T-hCov-1-2	HKU1	104	T-hCov-7-1	OC43	131
T-hCov-1-3	HKU1	203	T-hCov-7-2	OC43	254
T-hCov-2	HKU1	169	T-hCov-8	OC43	897
T-hCov-2-1	HKU1	92	T-hCov-8-1	OC43	569
T-hCov-2-2	HKU1	136	T-hCov-8-2	OC43	342
T-hCov-3	HKU1	330	T-hCov-9	OC43	500
T-hCov-3-1	HKU1	918	T-hCov-9-1	OC43	145
Total BBXs in dataset		**17851**			
SARS-CoV-2 Wuhan vs Negative (Microscope 1)					
T-SC2-1	SC2	696	T-SC-4-2	SC2	1783
T-SC2-1-1	SC2	3320	T-Neg-1	Neg	1507
T-SC-2	SC2	370	T-Neg-2	Neg	562
T-SC-2-1	SC2	223	T-Neg-3	Neg	451
T-SC-2-2	SC2	131	T-Neg-4	Neg	378
T-SC-3	SC2	662	T-Neg-1-1	Neg	5173
T-SC-3-1	SC2	214	T-Neg-2-1	Neg	122
T-SC-4	SC2	861	T-Neg-3-1	Neg	568
T-SC-4-1	SC2	218	T-Neg-4-1	Neg	729
Total BBXs in dataset		**17968**			
Flu A vs Negative (Microscope 1)					
T-Neg-5	Neg	514	T-Neg-4	Neg	378
T-Neg-5-1	Neg	337	T-Neg-4-1	Neg	158
T-Neg-5-2	Neg	471	T-Neg-4-2	Neg	571
T-Neg-2	Neg	652	T-Neg-4-3	Neg	154
T-Neg-2-1	Neg	77	T-Neg-4-4	Neg	233
T-Neg-2-2	Neg	45	T-FluA-1	Flu A	1782
T-Neg-2-3	Neg	242	T-FluA-1-1	Flu A	289
T-Neg-3	Neg	451	T-FluA-1-2	Flu A	246
T-Neg-3-1	Neg	568	T-FluA-2	Flu A	61
T-Neg-3-2	Neg	398	T-FluA-2-1	Flu A	658
T-Neg-3-3	Neg	150	T-FluA-3	Flu A	1177
T-Neg-3-4	Neg	374			
Total BBXs in dataset		**10461**			
SARS-COV-2 Wuhan vs Flu A (Microscope 1)					
T-FluA-5	Flu A	289	T-SC-2-1	SC2	223
T-FluA-5-1	Flu A	246	T-SC-2-3	SC2	318
T-FluA-5-2	Flu A	2416	T-SC-4-3	SC2	122
T-FluA-6	Flu A	188	T-SC-4-4	SC2	200
T-FluA-6-1	Flu A	748	T-SC-4-5	SC2	546
T-FluA-6-2	Flu A	158	T-SC-5	SC2	170
T-SC2-1-2	SC2	248	T-SC-5-1	SC2	260
T-SC2-1-3	SC2	209	T-SC-5-2	SC2	630
T-SC2-1-4	SC2	236	T-SC-5-3	SC2	823
Total BBXs in dataset		**8060**			
SARS-CoV-2 Alpha vs SARS-CoV-2 Wuhan (Microscope 2)					
T-SC2-6	B.1.1.7	631	T-SC2-9	SC2	483
T-SC2-7	B.1.1.7	663	T-SC2-10	SC2	511
T-SC2-8	B.1.1.7	235	T-SC2-11	SC2	616
Total BBXs in dataset		**3139**			
SARS-CoV-2 (Wuhan and Alpha) vs Negative (Microscope 2)					

Continued on next page

Sample ID	RT-PCR	Total BBXs	Sample ID	RT-PCR	Total BBXs
T-SC2-6	B.1.17	631	T-SC2-11	SC2	616
T-SC2-12	B.1.1.7	127	T-Neg-5	Neg	655
T-SC2-13	B.1.1.7	771	T-Neg-6	Neg	1318
T-SC2-9	SC2	483	T-Neg-7	Neg	595
T-SC2-10	SC2	511	T-Neg-8	Neg	1008
Total BBXs in dataset		**6715**			
SARS-CoV-2 Alpha vs Negative (Microscope 2)					
T-SC-2-6	B.1.1.7	631	T-Neg-5	Neg	655
T-SC-2-7	B.1.1.7	663	T-Neg-7	Neg	469
T-SC-2-8	B.1.1.7	235	T-Neg-9	Neg	595
Total BBXs in dataset		**3248**			
SARS-CoV-2 vs Negative (Microscope 3)					
NEG-J-2	Negative	792	NEG-J-30	Negative	310
NEG-J-3	Negative	135	SC2-WT-02	SARS-CoV-2	483
NEG-J-4	Negative	182	SC2-WT-07	SARS-CoV-2	511
NEG-J-5	Negative	451	SC2-WT-13	SARS-CoV-2	616
NEG-J-6	Negative	672	SC2-D-2	SARS-CoV-2	500
NEG-J-12	Negative	1218	SC2 lab grown	N/A	582
Total BBXs in dataset		**6452**			
Wuhan + Alpha vs Delta variant (Microscope 3)					
SC2-A-01	SARS-CoV-2	360	SC2-D-3	SARS-CoV-2	31
SC2-A-03	SARS-CoV-2	188	SC2-D-4	SARS-CoV-2	9
SC2-A-09	SARS-CoV-2	1022	SC2-D-9	SARS-CoV-2	266
SC2-WT-04	SARS-CoV-2	491	SC2-D-10	SARS-CoV-2	81
SC2-WT-05	SARS-CoV-2	572	SC2-D-12	SARS-CoV-2	455
SC2-WT-11	SARS-CoV-2	460	SC2-D-13	SARS-CoV-2	236
SC2-D-1	SARS-CoV-2	1680	SC2-D-29	SARS-CoV-2	536
Total BBXs in dataset		**6387**			

Table A.2: **Results of independent testing of the network on previously unseen samples (51 samples).**Abbreviations: hCoV – seasonal human coronaviruses (including NL63, OC43, and HKU1), Neg – negative, na – statistical test not required when the proportion of BBXs classified as negative by the network (# negatives/total # of BBXs) exceeds the model's specificity, SC2 – SARS-CoV-2, Flu A – Influenza A, SC2-S0 – SARS-CoV-2 spike gene target failure in RT-PCR, indicative of the Alpha variant. Samples that were incorrectly classified are highlighted in turquoise.

Sample ID	RT-PCR	Total BBXs			P-value	Result
'Microscope 1' hCoV Neg Thresh: 0.01			**hCoV**	**Neg**		
hCoV-1	NL63	6467	4271	2196	5.18E-10	hCoV
hCoV-2	NL63	2672	1307	1365	5.18E-21	hCoV
hCoV-3	NL63	227	111	116	0.006	hCoV
hCoV-4	OC43	2239	1706	533	9.25E-268	hCoV
hCoV-5	OC43	653	442	211	2.80E-47	hCoV
hCoV-6	OC43	840	555	285	1.13E-53	hCoV
hCoV-7	HKU1	1342	1087	255	2.09E-206	hCoV
hCoV-8	HKU1	1898	1069	829	9.69E-48	hCoV
hCoV-9	HKU1	6745	3797	2948	2.82E-164	hCoV
NEG-1	Neg	514	198	316	0.49	Neg
NEG-2	Neg	398	141	257	na	Neg
NEG-3	Neg	154	29	125	na	Neg

Continued on next page

Sample ID	RT-PCR	Total BBXs	hCoV	Neg	P-value	Result
'Microscope 1' SC2 hCoV Thresh: 0.01			**SC2**	**hCoV**		
SC2-1	SARS-CoV-2	5589	3653	1936	5.18E-10	SC2
SC2-2	SARS-CoV-2	1762	1312	450	5.14E-21	SC2
SC2-3	SARS-CoV-2	284	86	198	0.917	hCoV
hCoV-7	HKU1	1342	159	1183	na	hCoV
hCoV-8	HKU1	1898	593	1305	0.237	hCoV
hCoV-10	HKU1	3182	654	2528	na	hCoV
hCoV-11	OC43	264	8	256	na	hCoV
hCoV-12	OC43	479	21	458	na	hCoV
hCoV-13	OC43	7966	2245	5721	na	hCoV
hCoV-14	NL63	305	12	293	na	hCoV
hCoV-15	NL63	968	89	879	na	hCoV
hCoV-16	NL63	2269	349	1920	na	hCoV
'Microscope 1' SC2 Neg Thresh: 0.01			**SC2**	**Neg**		
SC2-4	SARS-CoV-2	200	150	50	7.93E-21	SC2
SC2-5	SARS-CoV-2	560	503	57	1.07E-114	SC2
SC2-6	SARS-CoV-2	529	496	33	7.58E-127	SC2
SC2-7	SARS-CoV-2	173	148	25	1.11E-30	SC2
SC2-8	SARS-CoV-2	150	135	15	2.89E-32	SC2
SC2-9	SARS-CoV-2	8095	3575	4520	6.91E-4	SC2
Neg-4	Neg	158	28	130	na	Neg
Neg-5	Neg	568	240	328	0.982	Neg
Neg-6	Neg	77	16	61	na	Neg
Neg-7	Neg	337	132	215	na	Neg
'Microscope 1' Flu A Neg Thresh: 0.01			**Flu A**	**Neg**		
Flu A-1	Flu A	311	263	48	2.08E-08	Flu A
Flu A-2	Flu A	195	163	32	3.46E-05	Flu A
Flu A-3	Flu A	24383	24231	152	0.00	Flu A
Flu A-4	Flu A	212	185	27	4.13E-08	Flu A
Neg-8	Neg	1507	41	1466	na	Neg
Neg-9	Neg	374	259	115	0.22	Neg
Neg-10	Neg	354	235	119	0.14	Neg
'Microscope 1' SC2 Flu A Thresh: 0.01			**SC2**	**Flu A**		
SC2-10	SARS-CoV-2	1783	1551	232	0.00	SC2
SC2-11	SARS-CoV-2	1340	818	522	2.54E-81	SC2
SC2-12	SARS-CoV-2	13915	5413	8502	3.40E-21	SC2
Flu A-1	Flu A	311	106	205	na	Flu A
Flu A-2	Flu A	195	53	142	0.270	Flu A
Flu A-3	Flu A	24383	2871	21512	na	Flu A
Flu A-4	Flu A	212	51	161	na	Flu A
'Microscope 2' B.1.1.7 Neg Thresh: 0.01			**SC2-S0**	**Neg**		
SC2-13	SC2-S0	4979	2538	2441	0.00	B.1.1.7
SC2-14	SC2-S0	631	602	29	0.00	B.1.1.7
SC2-15	SC2-S0	127	117	10	2.18E-61	B.1.1.7
SC2-16	SC2-S0	771	738	33	0.00	B.1.1.7
Neg-8	Neg	4166	22	4144	na	Neg
Neg-9	Neg	2888	130	2758	na	Neg
Neg-10	Neg	3369	947	2422	0.147	Neg
'Microscope 2' SC2 Negative Threshold: 0.01			**SC2**	**Neg**		
SC2-17	SC2-S0	5024	2309	2715	0.000	SC2
SC2-14	SC2-S0	631	563	68	0.000	SC2
SC2-15	SC2-S0	127	97	30	0.000	SC2
SC2-16	SC2-S0	771	709	62	0.000	SC2
SC2-18	SC2-S0	2492	2254	238	0.000	SC2
SC2-19	SC2	4058	1302	2756	0.004	SC2
SC2-20	SC2	1609	1410	199	0.000	SC2
Neg-8	Neg	4166	33	4133	na	Neg
Neg-9	Neg	2888	164	2724	na	Neg

Continued on next page

Sample ID	RT-PCR	Total BBXs	hCoV	Neg	P-value	Result
Neg-10	Neg	2422	32	2390	na	Neg
Neg-11	Neg	2435	741	1694	0.64	Neg

Table A.3: **Results of independent testing of the network on previously unseen samples (104 total)**. Abbreviations: Neg – negative; NA – statistical test not required as the proportion of BBXs classified as negative by the trained network (i.e., # negatives/total # of BBXs) exceeds the model's specificity; SC2 – SARS-CoV-2. Samples with fewer than 5 BBXs required for the chi-squared test are highlighted in yellow, while incorrectly classified samples are highlighted in turquoise.

Sample	Number of NEGatives	Number of Positives	NEGative Normalised	Positive Normalised	p-value	Verdict
NEG-2-1	40	1	0.98	0.02	NA	NEG
NEG-2-2	363	71	0.84	0.16	NA	NEG
NEG-2-3	666	154	0.81	0.19	NA	NEG
NEG-2-4	898	228	0.80	0.20	NA	NEG
NEG-2-5	581	135	0.81	0.19	NA	NEG
NEG-2-6	1027	363	0.74	0.26	NA	NEG
NEG-2-7	641	250	0.72	0.28	NA	NEG
NEG-2-8	1272	491	0.72	0.28	NA	NEG
NEG-2-9	202	33	0.86	0.14	NA	NEG
NEG-J-10	662	97	0.87	0.13	NA	NEG
NEG-J-11	125	29	0.81	0.19	NA	NEG
NEG-J-12	913	428	0.68	0.32	0.47	NEG
NEG-J-13	694	294	0.70	0.30	NA	NEG
NEG-J-14	1	3	0.25	0.75	INC	INC
NEG-J-15	611	286	0.68	0.32	0.57	NEG
NEG-J-16	200	48	0.81	0.19	NA	NEG
NEG-J-17	51	14	0.78	0.22	NA	NEG
NEG-J-18	142	14	0.91	0.09	NA	NEG
NEG-J-19	185	31	0.86	0.14	NA	NEG
NEG-J-2	777	145	0.84	0.16	NA	NEG
NEG-J-20	771	392	0.66	0.34	0.05	NEG
NEG-J-21	490	144	0.77	0.23	NA	NEG
NEG-J-22	40	13	0.75	0.25	NA	NEG
NEG-J-23	185	33	0.85	0.15	NA	NEG
NEG-J-24	425	155	0.73	0.27	NA	NEG
NEG-J-25	335	121	0.73	0.27	NA	NEG
NEG-J-26	432	135	0.76	0.24	NA	NEG
NEG-J-27	305	72	0.81	0.19	NA	NEG
NEG-J-28	186	57	0.77	0.23	NA	NEG
NEG-J-29	749	508	0.60	0.40	5.34×10^{-13}	POS
NEG-J-3	232	42	0.85	0.15	NA	NEG
NEG-J-30	266	44	0.86	0.14	NA	NEG
NEG-J-31	39	7	0.85	0.15	NA	NEG
NEG-J-32	105	14	0.88	0.12	NA	NEG
NEG-J-33	241	75	0.76	0.24	NA	NEG
NEG-J-34	1148	445	0.72	0.28	NA	NEG
NEG-J-36	788	185	0.81	0.19	NA	NEG
NEG-J-37	40	6	0.87	0.13	NA	NEG
NEG-J-38	450	71	0.86	0.14	NA	NEG
NEG-J-39	1138	357	0.76	0.24	NA	NEG
NEG-J-4	151	31	0.83	0.17	NA	NEG
NEG-J-40	777	249	0.76	0.24	NA	NEG
NEG-J-41	62	11	0.85	0.15	NA	NEG
NEG-J-42	681	304	0.69	0.31	NA	NEG
NEG-J-43	1	0	1.00	0.00	INC	INC
NEG-J-44	12	2	0.86	0.14	NA	NEG
NEG-J-45	7	0	1.00	0.00	NA	NEG
NEG-J-46	29	1	0.97	0.03	NA	NEG
NEG-J-47	30	3	0.91	0.09	NA	NEG

Continued on next page

Sample	**Number of NEGatives**	**Number of Positives**	**NEGative Normalised**	**Positive Normalised**	**p-value**	**Verdict**
NEG-J-48	12	0	1.00	0.00	NA	NEG
NEG-J-49	23	1	0.96	0.04	NA	NEG
NEG-J-5	939	139	0.87	0.13	NA	NEG
NEG-J-50	14	0	1.00	0.00	NA	NEG
NEG-J-51	22	0	1.00	0.00	NA	NEG
NEG-J-52	8	0	1.00	0.00	NA	NEG
NEG-J-53	38	1	0.97	0.03	NA	NEG
NEG-J-55	26	0	1.00	0.00	NA	NEG
NEG-J-56	15	1	0.94	0.06	NA	NEG
NEG-J-57	16	2	0.89	0.11	NA	NEG
NEG-J-58	25	4	0.86	0.14	NA	NEG
NEG-J-59	770	298	0.72	0.28	NA	NEG
NEG-J-6	808	124	0.87	0.13	NA	NEG
NEG-J-60	213	54	0.80	0.20	NA	NEG
NEG-J-61	930	377	0.71	0.29	NA	NEG
NEG-J-62	1035	384	0.73	0.27	NA	NEG
Neg-J-7	1202	181	0.87	0.13	NA	NEG
Neg-J-8	454	117	0.80	0.20	NA	NEG
Neg-J-9	652	116	0.85	0.15	NA	NEG
SC2-S0-1	838	1654	0.34	0.66	0	POS
SC2-S0-10	169	144	0.54	0.46	9.44×10^{-9}	POS
SC2-S0-11	873	494	0.64	0.36	4.01×10^{-5}	POS
SC2-S0-12	30	38	0.44	0.56	9.14×10^{-6}	POS
SC2-S0-2	361	4663	0.07	0.93	0	POS
SC2-S0-3	1408	806	0.64	0.36	3.83×10^{-8}	POS
SC2-S0-4	318	172	0.65	0.35	0.05	NEG
SC2-S0-5	105	826	0.11	0.89	0	POS
SC2-S0-6	73	4045	0.02	0.98	0	POS
SC2-S0-7	28	99	0.22	0.78	2.61×10^{-30}	POS
SC2-S0-8	212	559	0.27	0.73	4.81×10^{-137}	POS
SC2-S0-9	13	3252	0.00	1.00	0	POS
SC2-D-1	9	8342	0.00	1.00	0	POS
SC2-D-10	243	81	0.75	0.25	NA	NEG
SC2-D-12	708	455	0.61	0.39	2.10×10^{-9}	POS
SC2-D-13	424	236	0.64	0.36	0.01	POS
SC2-D-25	211	126	0.63	0.37	0.01	POS
SC2-D-28	739	771	0.49	0.51	9.79×10^{-64}	POS
SC2-D-29	295	1052	0.22	0.78	9.54×10^{-306}	POS
SC2-D-3	246	945	0.21	0.79	5.65×10^{-285}	POS
SC2-D-4	156	2159	0.07	0.93	0	POS
SC2-D-9	82	266	0.24	0.76	5.04×10^{-75}	POS
SC2-WT-1	163	3895	0.04	0.96	0	POS
SC2-WT-10	189	1411	0.12	0.88	0	POS
SC2-WT-11	360	1311	0.22	0.78	0	POS
SC2-WT-12	439	1862	0.19	0.81	0	POS
SC2-WT-13	556	2252	0.20	0.80	0	POS
SC2-WT-14	9	4286	0.00	1.00	0	POS
SC2-WT-2	460	1149	0.29	0.71	4.10×10^{-269}	POS
SC2-WT-3	158	1123	0.12	0.88	0	POS
SC2-WT-4	376	830	0.31	0.69	2.02×10^{-177}	POS
SC2-WT-5	303	1740	0.15	0.85	0	POS
SC2-WT-6	68	1264	0.05	0.95	0	POS
SC2-WT-7	41	2736	0.01	0.99	0	POS
SC2-WT-8	423	437	0.49	0.51	3.35×10^{-36}	POS
SC2-WT-9	4	2126	0.00	1.00	0	POS

Table A.4: **Results of variant classification (35 samples)**. Abbreviations: Neg – negative; NA – statistical test not required if the proportion of BBXs classified as positive by the trained network (i.e., # positives/total # of BBXs) for either variant exceeds the sensitivity for that variant, in which case the sample is classified as that variant; SC2 – SARS-CoV-2. Samples with p-values that differ by less than three orders of magnitude (inconclusive) are highlighted in yellow, while incorrectly classified samples are highlighted in turquoise.

Sample	**Variant**	**Total**	**Negative**	**Positive**	**p-value**	**p-value 2**	**Category**
NEG-J-29	Negative	508	325	183	9.58E-09	1.30E-21	Delta
SC2-D-01	Delta	8342	409	7933	0.00	NA	Delta
SC2-D-03	Delta	31	30	1	NA	1.47E-09	Wuhan/Alpha
SC2-D-04	Delta	9	4	5	0.03	0.93	Inconclusive
SC2-D-09	Delta	266	133	133	4.67E-21	0.02	Delta
SC2-D-12	Delta	455	223	232	1.59E-37	0.01	Delta
SC2-D-13	Delta	236	128	108	1.76E-13	0.00	Delta
SC2-D-25	Delta	337	109	228	4.27E-73	NA	Delta
SC2-D-28	Delta	771	549	222	0.01	2.27E-56	Wuhan/Alpha
SC2-D-29	Delta	1052	184	868	0.00	NA	Delta
SC2-S0-01	Wuhan/Alpha	1654	1248	406	NA	1.38E-156	Wuhan/Alpha
SC2-S0-02	Wuhan/Alpha	4663	4513	150	NA	0.00	Wuhan/Alpha
SC2-S0-03	Wuhan/Alpha	806	672	134	NA	1.36E-118	Wuhan/Alpha
SC2-S0-05	Wuhan/Alpha	526	476	50	NA	2.75E-107	Wuhan/Alpha
SC2-S0-06	Wuhan/Alpha	4045	3530	515	NA	0.00	Wuhan/Alpha
SC2-S0-07	Wuhan/Alpha	99	89	10	NA	4.28E-21	Wuhan/Alpha
SC2-S0-08	Wuhan/Alpha	559	529	30	NA	2.99E-134	Wuhan/Alpha
SC2-S0-09	Wuhan/Alpha	3252	3051	201	NA	0.00	Wuhan/Alpha
SC2-S0-10	Wuhan/Alpha	144	72	72	4.26E-12	0.09	Delta
SC2-S0-11	Wuhan/Alpha	494	147	347	2.68E-119	NA	Delta
SC2-S0-12	Wuhan/Alpha	38	22	16	0.01	0.06	Inconclusive
SC2-WT-01	Wuhan/Alpha	3895	3740	155	NA	0.00	Wuhan/Alpha
SC2-WT-02	Wuhan/Alpha	1149	868	281	NA	5.50E-110	Wuhan/Alpha
SC2-WT-03	Wuhan/Alpha	1123	1025	98	NA	3.45E-234	Wuhan/Alpha
SC2-WT-04	Wuhan/Alpha	830	770	60	NA	1.92E-184	Wuhan/Alpha
SC2-WT-05	Wuhan/Alpha	1740	1541	199	NA	0.00	Wuhan/Alpha
SC2-WT-06	Wuhan/Alpha	1264	1190	74	NA	2.51E-295	Wuhan/Alpha
SC2-WT-07	Wuhan/Alpha	2736	2516	220	NA	0.00	Wuhan/Alpha
SC2-WT-08	Wuhan/Alpha	860	501	359	8.32E-30	1.61E-19	Delta
SC2-WT-09	Wuhan/Alpha	2126	574	1552	0.00	NA	Delta
SC2-WT-10	Wuhan/Alpha	1411	1317	94	NA	0.00	Wuhan/Alpha
SC2-WT-11	Wuhan/Alpha	1311	1221	90	NA	2.49E-294	Wuhan/Alpha
SC2-WT-12	Wuhan/Alpha	1862	1590	272	NA	7.56E-299	Wuhan/Alpha
SC2-WT-13	Wuhan/Alpha	2252	1895	357	NA	0.00	Wuhan/Alpha
SC2-WT-14	Wuhan/Alpha	4286	3696	590	NA	0.00	Wuhan/Alpha

Appendix A

Ethical Approvals

This appendix contains copies of the ethical approval letter and subsequent amendments relating to the clinical samples used in this study.

NHS
Health Research Authority

London - Queen Square Research Ethics Committee
HRA NRES Centre Bristol
3rd floor, block B
Whitefriars
Lewins Mead
Bristol
BS1 2NT

16 August 2021

Professor Derrick Crook
Consultant Clinical Microbiologist
University of Oxford/Oxford University Hospitals NHS Foundation Trust
Microbiology Department
John Radcliffe Hospital, Level 7
Headley Way
Oxford OX3 9DU

Dear Professor Crook

Study title:	**Rapid extraction of pathogen nucleic acid directly from clinical samples to investigate a whole genome sequencing-based diagnostic solution, including predicting antimicrobial resistance for bacterial infections**
REC reference:	**17/LO/1420**
Amendment number:	**Substantial Amendment 1**
Amendment date:	**05 July 2021**
IRAS project ID:	**228657**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Completed Amendment Tool [Whole-genome sequencing-based rapid diagnostics]	Substantial Amendment 1	05 July 2021
Research protocol or project proposal [Whole-genome sequencing-based rapid diagnostics]	2	16 July 2021

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

Working with NHS Care Organisations

Sponsors should ensure that they notify the R&D office for the relevant NHS care organisation of this amendment in line with the terms detailed in the categorisation email issued by the lead nation for the study.

Amendments related to COVID-19

We will update your research summary for the above study on the research summaries section of our website. During this public health emergency, it is vital that everyone can promptly identify all relevant research related to COVID-19 that is taking place globally. If you have not already done so, please register your study on a public registry as soon as possible and provide the HRA with the registration detail, which will be posted alongside other information relating to your project.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

HRA Learning

We are pleased to welcome researchers and research staff to our HRA Learning Events and online learning opportunities– see details at: https://www.hra.nhs.uk/planning-and-improving-research/learning/

IRAS Project ID - 228657: **Please quote this number on all correspondence**

Yours sincerely

Wai Yeung

PP – Ms Danielle Wilson
Chair

E-mail: queensquare.rec@hra.nhs.uk

Enclosures:	*List of names and professions of members who took part in the review*
Copy to:	*Ms Heather House, Oxford University Hospitals NHS Foundation Trust*

London - Queen Square Research Ethics Committee

Attendance at Sub-Committee of the REC

Committee Members:

Name	*Profession*	*Present*	*Notes*
Ms Lisa Marie Doherty	Oncology Advanced Nurse Practitioner	Yes	
Ms Danielle Wilson	Clinical Research Operational Manager	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Mr Wai Yeung	Approvals Administrator

Amendment Tool

v1.5 25 Mar 2021

For office use

QC: No

Section 1: Project information

Short project title*:	Whole-genome sequencing-based rapid diagnostics
IRAS project ID* (or REC reference if no IRAS project ID is available):	228657
Sponsor amendment reference number*:	Substantial Amendment 1
Sponsor amendment date* (enter as DD/MM/YY):	05 July 2021
Briefly summarise in lay language the main changes proposed in this amendment. Explain the purpose of the changes and their significance for the study. If the amendment significantly alters the research design or methodology, or could otherwise affect the scientific value of the study, supporting scientific information should be given (or enclosed separately). Indicate whether or not additional scientific critique has been obtained (note: this field will adapt to the amount of text entered)*:	Current ethical approval allows for development of pathogen detection in clinical samples using long-read DNA sequencing technology only. The proposed changes will extend the methods underpinning the development of rapid diagnostic tools for the diagnosis of infection by also including: 1) detection of pathogens and antimicrobial resistance potential using any nucleic acid-based approch, including long read, short read, DNA and RNA sequencing approaches; 2) High resolution microscopy (nano-imaging) and computer learning image analysis to diagnose infection direct from clinical samples; 3) Detection of proteins either as a direct signature of infection (e.g. viral antigens) or as an indirect signature of infection (e.g. by evaluating human responses to infection, such as antibody); 4) optimising diagnostic workflows by applying AI/machine learning approaches to evaluating images generated as part of routine microbiological diagnostics (e.g. images of culture plates generated as part of automated culture-based workflows). The rationale for these inclusions is outlined in the supporting amended protocol. Short and long study titles are being updated to reflect the above changes

Project type (select):	● Specific study ○ Research tissue bank ○ Research database			
Has the study been reviewed by a UKECA-recognised Research Ethics Committee (REC) prior to this amendment?:	● Yes		○ No	
What type of UKECA-recognised Research Ethics Committee (REC) review is applicable? (select):	● NHS/HSC REC ○ Ministry of Defence (MoDREC)			
Is all or part of this amendment being resubmitted to the Research Ethics Committee (REC) as a **modified amendment** (i.e. a substantial amendment previously given an unfavourable opinion)?	○ Yes		● No	
Where is the NHS/HSC Research Ethics Committee (REC) that reviewed the study based?:	England	Wales	Scotland	Northern Ireland
	●	○	○	○
Was the study a clinical trial of an investigational medicinal product (CTIMP) OR does the amendment make it one?:	○ Yes		● No	
Was the study a clinical investigation or other study of a medical device OR does the amendment make it one?:	○ Yes		● No	
Did the study involve the administration of radioactive substances, therefore requiring ARSAC review, OR does the amendment introduce this?:	○ Yes		● No	
Did the study involve the use of research exposures to ionising radiation (not involving the administration of radioactive substances) OR does the amendment introduce this?:	○ Yes		● No	
Did the study involve adults lacking capacity OR does the amendment introduce this?:	○ Yes		● No	
Did the study involve access to confidential patient information outside the direct care team without consent OR does the amendment introduce this?:	○ Yes		● No	
Did the study involve prisoners OR does the amendment introduce this?:	○ Yes		● No	
Did the study involve children OR does the amendment introduce this?:	● Yes		○ No	
Did the study involve NHS/HSC organisations prior to this amendment?:	● Yes		○ No	
Did the study involve non-NHS/HSC organisations OR does the amendment introduce them?:	○ Yes		● No	
	England	Wales	Scotland	Northern Ireland
Lead nation for the study:	●	○	○	○
Which nations had participating NHS/HSC organisations prior to this amendment?	☑	☐	☐	☐
Which nations will have participating NHS/HSC organisations after this amendment?	☑	☐	☐	☐

Section 2: Summary of change(s)

Please note: Each change being made as part of the amendment must be entered separately. For example, if an amendment to a clinical trial of an investigational medicinal product (CTIMP) involves an update to the Investigator's Brochure (IB), affecting the Reference Safety Information (RSI) and so the information documents to be given to participants, these should be entered into the amendment tool as three separate changes. A list of all possible changes is available on the "Glossary of Amendment Options" tab. To add another change, tick the "Add another change" box.

Change 1	
Area of change (select)*:	Study Design
Specific change (select - only available when area of change is selected first)*:	Participant numbers - Significant change to sample size
Further information (free text - note that this field will adapt to the amount of text entered):	This ammendment will increase the numbers of collected clinical samples to develop additional approaches to the rapid diagnosis of infection in samples by using an expanded portfolio of nucleic-acid/sequencing-based approaches, by performing high-resolution imaging diagnostics of pathogens in clinical samples,and through testing for protein-based targets. We have also requested permission to collect images (not samples) generated as part of automated laboratory workflows for the purposes of developing AI/machine learning based approaches to diagnostics.

Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☑

Change 2	
Area of change (select)*:	Study Design
Specific change (select - only available when area of change is selected first)*:	Background information - Change that affects scientific value of study
Further information (free text - note that this field will adapt to the amount of text entered):	Recent research has highlighted the strong potential for the use of non long-read sequencing based methods in rapidly diagnosing infection, such as amplicon sequencing based methods, high-resolution nano-imaging for the detection and diagnosis of pathogens from clinical samples, and directly targeting pathogen proteins or human antibodies generated in response to infection. These ammendments will allow the study to explore this potential on samples previously approved for long read DNA sequencing only.

Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☑

Change 3	
Area of change (select)*:	Study Design
Specific change (select - only available when area of change is selected first)*:	Extension to study duration that will not have any additional resource implications for participating organisations - Please specify in the free text below
Further information (free text - note that this field will adapt to the amount of text entered):	The planned work to optimise work on long-read sequencing-based methods has been set back by the SARS-CoV-2 pandemic; however, our research work related to SARS-CoV-2 has highlighted the potential of a number of different techniques and approaches that could be used for the rapid diagnosis of infection that we wish to explore through this amendment. Given that we are increasing the number of modalities to evaluate, and have lost time on the formerly approved work, we would like to request a 22 month extension to the end of January 2024. Current funding for the originally approved project has been extended from March 2022 to August 2022, and since then we have secured several other grants supporting the work on sequencing and non-sequencing modalities (John Fell Fund, Oxford Martin award) that would cover the requested extension.

Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☑

Change 4	
Area of change (select)*:	Researchers
Specific change (select - only available when area of change is selected first)*:	PI - New PI, or temporary arrangements to cover the absence of a PI

Further information (free text - note that this field will adapt to the amount of text entered):	In order to support the investigation of the new modalities for infectious diseases diagnostics proposed in the amendment, we have added three new research team members to the project as Investigators, namely: Prof David Eyre (sequencing, database integration and analysis), Dr Phil Fowler (artifical intelligence, machine learning, image analysis), and Dr Nicole Stoesser (sequencing, nanoimaging).			
Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☑

Change 5				
Area of change (select)*:	Study Design			
Specific change (select - only available when area of change is selected first)*:	Other significant change to study design that can be implemented within existing resource in place at participating organisations - Please specify in the free text below			
Further information (free text - note that this field will adapt to the amount of text entered):	The extraction of data from laboratory and electronic patient records is currently done via an existing NHS data warehouse where data are processed to ensure that they do not contain identifiable information. In addition to data that is already being extracted for the purposes of this study, we would like to collect the following items: (i) non-antibiotic medication that may impact infection risk; (ii) details from structured infection control and infection clinician reviews; (iii) details of patient vital signs, radiology and point of care test results to enable an evaluation of severity.			
Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☑

Change 6				
Area of change (select)*:	Administrative details for the project			
Specific change (select - only available when area of change is selected first)*:	Project identification (e.g. change of title, reference numbers)			
Further information (free text - note that this field will adapt to the amount of text entered):	The study long title has been changed to: "Direct-from-sample approaches to infection diagnostics using whole genome sequencing, imaging, and protein signatures to diagnose infectious pathogens and determine antimicrobial resistance" and the short title has been changed to: "Rapid direct-from-sample approaches to diagnosing infection"s			
Applicability:	England	Wales	Scotland	Northern Ireland
Where are the participating NHS/HSC organisations located that will be affected by this change?*:	☑	☐	☐	☐
Will all participating NHS/HSC organisations be affected by this change, or only some? (**please note** that this answer may affect the categorisation for the change):	● All		○ Some	

Add another change: ☐

Section 3: Declaration(s) and lock for submission

Declaration by the Sponsor or authorised delegate

- I confirm that the Sponsor takes responsibility for the completed amendment tool
- I confirm that I have been formally authorised by the Sponsor to complete the amendment tool on their behalf

Name [first name and surname]:*	Nigel Wellman
Email address:*	nigel.wellman@admin.ox.ac.uk

Lock for submission

Please note: This button will only become available when all mandatory (*) fields have been completed. When the button is available, clicking it will generate a locked PDF copy of the completed amendment tool which must be included in the amendment submission. Please ensure that the amendment tool is completed correctly before locking it for submission.

Lock for submission

After locking the tool, proceed to submit the amendment online. The "Submission Guidance" tab provides further information about the next steps for the amendment.

Section 4: Review bodies for the amendment

Please note: This section is for **information only**. Details in this section will complete automatically based on the options selected in Sections 1 and 2.

	Review bodies																		
	UK wide:						England and Wales:				Scotland:				Northern Ireland:				
	REC	Competent Authority MHRA - Medicines	Competent Authority MHRA - Devices	ARSAC	Radiation Assurance	UKSW Governance	REC (MCA)	CAG	HMPPS	HRA and HCRW Approval	REC (AWIA)	PBPP	SPS (RAEC)	National coordinating function	HSC REC	HSC Data Guardians	Prisons	National coordinating function	Category:
Change 1:	Y					(Y)				(Y)									A
Change 2:	Y					(Y)				(Y)									A
Change 3:	N					(Y)				(Y)									C
Change 4:	N					(Y)				(Y)									A
Change 5:	Y					Y				Y									C
Change 6:	(Y)					Y				(Y)									C
Overall reviews for the amendment:																			
Full review:	Y					Y				Y									
Notification only:	N					N				N									
Overall amendment type:	Substantial																		
Overall Category:	A																		

For national coordinating function office use:	
Update HARP:	This amendment may involve an update to contact details, project end date, or other project details. Ensure that HARP is updated with the current details. If this is the only change, no further study-wide review is required.

Bibliography

[1] World Health Organization. *Prioritizing diseases for research and development in emergency contexts*. `https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts`. 2023.

[2] Colin J. Carlson et al. "Climate change increases cross-species viral transmission risk". In: *Nature* 607.7919 (2022), pp. 555–562. DOI: `10.1038/s41586-022-04788-w`. URL: `https://doi.org/10.1038/s41586-022-04788-w`.

[3] Jan. 2022. URL: `https://www.reuters.com/business/imf-sees-cost-covid-pandemic-rising-beyond-125-trillion-estimate-2022-01-20/`.

[4] Centers for Disease Control and Prevention. `https://www.cdc.gov/flu/avianflu/avian-flu-summary.htm`. May 2023.

[5] Dapeng Tao et al. "One year of African swine fever outbreak in China". In: *Acta Tropica* 211 (2020), p. 105602. ISSN: 0001-706X. DOI: `https://doi.org/10.1016/j.actatropica.2020.105602`. URL: `https://www.sciencedirect.com/science/article/pii/S0001706X2030591X`.

[6] Shibing You et al. "African swine fever outbreaks in China led to gross domestic product and economic losses". In: *Nature Food* 2.10 (2021), pp. 802–808. DOI: `10.1038/s43016-021-00362-1`. URL: `https://doi.org/10.1038/s43016-021-00362-1`.

[7] Rosanna W Peeling and Samuel K Sia. "Lessons from COVID-19 for improving diagnostic access in future pandemics". In: *Lab on a Chip* 23.5 (2023), pp. 1376–1388.

[8] Center for Health Security. *The Path Forward: Building on Lessons Learned from the COVID-19 Pandemic*. Accessed: 2024-10-25. 2024. URL: `https://centerforhealthsecurity.org/our-work/testimonies-briefings/the-path-forward-building-on-lessons-learned-from-the-covid-19-pandemic`.

[9] Jedidajah Otte. "Covid tests sent to Italy and Germany as UK labs are overwhelmed". In: *The Guardian* (Sept. 2020). Accessed: 2025-04-16. URL: `https://www.theguardian.com/world/2020/sep/13/covid-tests-sent-to-italy-and-germany-as-uk-labs-are-overwhelmed`.

[10] House of Commons Library. *Coronavirus: Testing for Covid-19*. `https://researchbriefings.files.parliament.uk/documents/CBP-8897/CBP-8897.pdf`. Briefing Paper Number CBP 8897. Oct. 2020.

[11] Almudena Burillo and Emilio Bouza. "Use of rapid diagnostic techniques in ICU patients with infections". In: *BMC infectious diseases* 14 (2014), pp. 1–12.

[12] Tilmann Schober et al. "Clinical outcomes of rapid respiratory virus testing in emergency departments: a systematic review and meta-analysis". In: *JAMA internal medicine* 184.5 (2024), pp. 528–536.

[13] Tristan W Clark et al. "Rapid multiplex PCR for respiratory viruses reduces time to result and improves clinical care: results of a systematic review and meta-analysis". In: *Journal of Infection* 86.5 (2023), pp. 462–475.

[14] Lilit Garibyan and Nidhi Avashia. "Polymerase Chain Reaction". In: *Journal of Investigative Dermatology* 133.3 (2013), pp. 1–4. DOI: `10.1038/jid.2013.1`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4102308/`.

[15] Heinz Ulrich Weier and Joe W Gray. "A programmable system to perform the polymerase chain reaction". In: *Dna* 7.6 (1988), pp. 441–447.

[16] F. Watzinger, K. Ebner, and T. Lion. "Detection and monitoring of virus infections by real-time PCR". In: *Molecular Aspects of Medicine* 27 (2006), pp. 254–298. DOI: `10.1016/j.mam.2005.12.001`.

[17] Hubert GM Niesters. "Quantitation of viral load using real-time amplification techniques". In: *Methods* 25.4 (2001), pp. 419–429.

[18] G. Albertoni, Manoel João Batista Castelo Girão, and N. Schor. "Mini review: current molecular methods for the detection and quantification of hepatitis B virus, hepatitis C virus, and human immunodeficiency virus type 1." In: *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 25 (2014), pp. 145–9. DOI: `10.1016/j.ijid.2014.04.007`.

[19] I. Engelmann et al. "Rapid quantitative PCR assays for the simultaneous detection of herpes simplex virus, varicella zoster virus, cytomegalovirus, Epstein-Barr virus, and human herpesvirus 6 DNA in blood and other clinical specimens". In: *Journal of Medical Virology* 80 (2008). DOI: `10.1002/jmv.21095`.

[20] M. Jacobs et al. "A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings". In: *Journal of Clinical Microbiology* 35 (1997), pp. 791–795. DOI: `10.1128/jcm.35.3.791-795.1997`.

[21] L. van Doorn et al. "Highly Effective Detection of Human Papillomavirus 16 and 18 DNA by a Testing Algorithm Combining Broad-Spectrum and Type-Specific PCR". In: *Journal of Clinical Microbiology* 44 (2006), pp. 3292–3298. DOI: `10.1128/JCM.00539-06`.

[22] R. A. Hubbard. “Human papillomavirus testing methods.” In: *Archives of pathology & laboratory medicine* 127 8 (2009), pp. 940–5. DOI: `10.1043/1543-2165(2003)127<940:HPTM>2.0.CO;2`.

[23] M. Bertine et al. “New Highly Sensitive Real-Time PCR Assay for HIV-2 Group A and Group B DNA Quantification”. In: *Journal of Clinical Microbiology* 55 (2017), pp. 2850–2857. DOI: `10.1128/JCM.00755-17`.

[24] Minzhe Shen et al. “Recent advances and perspectives of nucleic acid detection for coronavirus”. In: *Journal of Pharmaceutical Analysis* 10 (2020), pp. 97–101. DOI: `10.1016/j.jpha.2020.02.010`.

[25] Michael A Frohman, Michael K Dush, and Gail R Martin. “Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer.” In: *Proceedings of the National Academy of Sciences* 85.23 (1988), pp. 8998–9002.

[26] Howard M Temin and David Baltimore. “RNA-directed DNA synthesis and RNA tumor viruses”. In: *Advances in virus research* 17 (1972), pp. 129–186.

[27] Russell Higuchi et al. “Simultaneous amplification and detection of specific DNA sequences”. In: *Bio/technology* 10.4 (1992), pp. 413–417.

[28] T. Notomi et al. “Loop-mediated isothermal amplification of DNA.” In: *Nucleic acids research* 28 12 (2000), E63. DOI: `10.1093/NAR/28.12.E63`.

[29] Y. Mori and T. Notomi. “Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases”. In: *Journal of Infection and Chemotherapy* 15 (2009), pp. 62–69. DOI: `10.1007/s10156-009-0669-9`.

[30] Rashad Abdul-Ghani, Abdulsalam M. AL-Mekhlafi, and P. Karanis. “Loop-mediated isothermal amplification (LAMP) for malarial parasites of humans: would it come to clinical reality as a point-of-care test?” In: *Acta tropica* 122 3 (2012), pp. 233–40. DOI: `10.1016/j.actatropica.2012.02.004`.

[31] S. Picot, M. Cucherat, and A. Bienvenu. “Systematic review and meta-analysis of diagnostic accuracy of Loop-mediated isothermal amplification (LAMP) methods compared to microscopy, PCR, and rapid diagnostic tests, for malaria diagnosis.” In: *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* (2020). DOI: `10.1016/j.ijid.2020.07.009`.

[32] P. Shete et al. “Diagnostic accuracy of TB-LAMP for pulmonary tuberculosis: a systematic review and meta-analysis”. In: *BMC Infectious Diseases* 19 (2019). DOI: `10.1186/s12879-019-3881-y`.

[33] Severino Jefferson Ribeiro da Silva, Keith Pardee, and L. Pena. “Loop-Mediated Isothermal Amplification (LAMP) for the Diagnosis of Zika Virus: A Review”. In: *Viruses* 12 (2019). DOI: `10.3390/v12010019`.

[34] T. Okafuji et al. “Rapid Diagnostic Method for Detection of Mumps Virus Genome by Loop-Mediated Isothermal Amplification”. In: *Journal of Clinical Microbiology* 43 (2005), pp. 1625–1631. DOI: `10.1128/JCM.43.4.1625-1631.2005`.

[35] J. Davey et al. “Genome-wide genetic marker discovery and genotyping using next-generation sequencing”. In: *Nature Reviews Genetics* 12 (2011), pp. 499–510. DOI: `10.1038/nrg3012`.

[36] Xuehui Huang et al. “High-throughput genotyping by whole-genome resequencing.” In: *Genome research* 19 6 (2009), pp. 1068–76. DOI: `10.1101/gr.089516.108`.

[37] A. Kawalia et al. “Leveraging the Power of High Performance Computing for Next Generation Sequencing Data Analysis: Tricks and Twists from a High Throughput Exome Workflow”. In: *PLoS ONE* 10 (2015). DOI: `10.1371/journal.pone.0126321`.

[38] Spyros Oikonomopoulos et al. “Methodologies for Transcript Profiling Using Long-Read Technologies”. In: *Frontiers in Genetics* 11 (2020). DOI: `10.3389/fgene.2020.00606`.

[39] F. Bewicke-Copley et al. “Applications and analysis of targeted genomic sequencing in cancer studies”. In: *Computational and Structural Biotechnology Journal* 17 (2019), pp. 1348–1359. DOI: `10.1016/j.csbj.2019.10.004`.

[40] Junfeng Xia et al. “NGS catalog: A database of next generation sequencing studies in humans”. In: *Human Mutation* 33 (2012). DOI: `10.1002/humu.22096`.

[41] Marc Lecuit and M. Eloit. “The potential of whole genome NGS for infectious disease diagnosis”. In: *Expert Review of Molecular Diagnostics* 15 (2015), pp. 1517–1519. DOI: `10.1586/14737159.2015.1111140`.

[42] S. Martínez-Puchol et al. “Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens.” In: *The Science of the total environment* 713 (2020), p. 136604. DOI: `10.1016/j.scitotenv.2020.136604`.

[43] Qingfa Wu et al. “Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms.” In: *Annual review of phytopathology* 53 (2015), pp. 425–44. DOI: `10.1146/annurev-phyto-080614-120030`.

[44] A. Prodan et al. “Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing”. In: *PLoS ONE* 15 (2020). DOI: `10.1371/journal.pone.0227434`.

[45] Wendy Weijia Soon, M. Hariharan, and M. Snyder. “High-throughput sequencing for biology and medicine”. In: *Molecular Systems Biology* 9 (2013), pp. 640–640. DOI: `10.1038/msb.2012.61`.

[46] W. Gu et al. "Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids". In: *Nature Medicine* 27 (2020), pp. 115–124. DOI: `10.1038/s41591-020-1105-z`.

[47] L. Cummings et al. "Clinical Next Generation Sequencing Outperforms Standard Microbiological Culture for Characterizing Polymicrobial Samples." In: *Clinical chemistry* 62 11 (2016), pp. 1465–1473. DOI: `10.1373/CLINCHEM.2016.258806`.

[48] Eman Anis et al. "Evaluation of Targeted Next-Generation Sequencing for Detection of Bovine Pathogens in Clinical Samples". In: *Journal of Clinical Microbiology* 56 (2018). DOI: `10.1128/JCM.00399-18`.

[49] I-Na Lu, C. Muller, and Feng Q He. "Applying next-generation sequencing to unravel the mutational landscape in viral quasispecies". In: *Virus Research* 283 (2020), pp. 197963–197963. DOI: `10.1016/j.virusres.2020.197963`.

[50] C. Charre et al. "Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation". In: *Virus Evolution* 6 (2020). DOI: `10.1093/ve/veaa075`.

[51] Takahiko Koyama, D. Platt, and L. Parida. "Variant analysis of SARS-CoV-2 genomes". In: *Bulletin of the World Health Organization* 98 (2020), pp. 495–504. DOI: `10.2471/BLT.20.253591`.

[52] A. Nasereddin et al. "Identification of SARS-CoV-2 Variants of Concern Using Amplicon Next-Generation Sequencing". In: *Microbiology Spectrum* 10 (2022). DOI: `10.1128/spectrum.00736-22`.

[53] A. Cassedy, A. Parle-McDermott, and R. O'Kennedy. "Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods". In: *Frontiers in Molecular Biosciences* 8 (2021). DOI: `10.3389/fmolb.2021.637559`.

[54] Mayara Lisboa Bastos et al. "Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis". In: *bmj* 370 (2020).

[55] Zhengtu Li et al. "Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis". In: *Journal of Medical Virology* 92 (2020), pp. 1518–1524. DOI: `10.1002/jmv.25727`.

[56] Ian J. Amanna, N. Carlson, and M. Slifka. "Duration of humoral immunity to common viral and vaccine antigens." In: *The New England journal of medicine* 357 19 (2007), pp. 1903–15. DOI: `10.1056/NEJMOA066092`.

[57] Y. Lustig et al. "Potential antigenic cross-reactivity between SARS-CoV-2 and Dengue viruses". In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* (2020). DOI: `10.1093/cid/ciaa1207`.

[58] Qiao Wang et al. "A Combination of Human Broadly Neutralizing Antibodies against Hepatitis B Virus HBsAg with Distinct Epitopes Suppresses Escape Mutations." In: *Cell host & microbe* (2020). DOI: `10.1016/j.chom.2020.05.010`.

[59] S. Lemon et al. "IgM antibody to hepatitis B core antigen as a diagnostic parameter of acute infection with hepatitis B virus." In: *The Journal of infectious diseases* 143 6 (1981), pp. 803–9. DOI: `10.1093/INFDIS/143.6.803`.

[60] Tulio J Lopera et al. "The usefulness of antigen testing in predicting contagiousness in COVID-19". In: *Microbiology spectrum* 10.2 (2022), e01962–21.

[61] Irene Bosch et al. "Rapid antigen tests for dengue virus serotypes and Zika virus in patient serum". In: *Science translational medicine* 9.409 (2017), eaan1589.

[62] J. Dinnes et al. "Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection". In: *The Cochrane Database of Systematic Reviews* 2020 (2020). DOI: `10.1002/14651858.cd013705`.

[63] S. Chevaliez and J. Pawlotsky. "New virological tools for screening, diagnosis and monitoring of hepatitis B and C in resource-limited settings." In: *Journal of hepatology* 69 4 (2018), pp. 916–926. DOI: `10.1016/j.jhep.2018.05.017`.

[64] Yuan-Hung Wang et al. "Evaluation of the diagnostic accuracy of COVID-19 antigen tests: a systematic review and meta-analysis". In: *Journal of the Chinese Medical Association* 84.11 (2021), pp. 1028–1037.

[65] Caroline Chartrand et al. "Diagnostic accuracy of rapid antigen detection tests for respiratory syncytial virus infection: systematic review and meta-analysis". In: *Journal of clinical microbiology* 53.12 (2015), pp. 3738–3749.

[66] J. Moore et al. "An enzyme-linked immunosorbent assay for antibodies to the envelope glycoproteins of divergent strains of HIV-1." In: *AIDS* 3 3 (1989), pp. 155–63. DOI: `10.1097/00002030-198903000-00006`.

[67] S. Usuda et al. "Serological detection of hepatitis B virus genotypes by ELISA with monoclonal antibodies to type-specific epitopes in the preS2-region product." In: *Journal of virological methods* 80 1 (1999), pp. 97–112. DOI: `10.1016/S0166-0934(99)00039-7`.

[68] Se-ho Kim. "ELISA for Quantitative Determination of Hepatitis B Virus Surface Antigen". In: *Immune Network* 17 (2017), pp. 451–459. DOI: `10.4110/in.2017.17.6.451`.

[69] Girish J. Kotwal et al. "Detection of acute hepatitis C virus infection by ELISA using a synthetic peptide comprising a structural epitope." In: *Proceedings of the National Academy of Sciences of the United States of America* 89 (1992), pp. 4486–4489. DOI: `10.1073/PNAS.89.10.4486`.

[70] D. Vaughn et al. "Rapid serologic diagnosis of dengue virus infection using a commercial capture ELISA that distinguishes primary and secondary infections." In: *The American journal of tropical medicine and hygiene* 60 4 (1999), pp. 693–8. DOI: `10.4269/AJTMH.1999.60.693`.

[71] Thermo Fisher Scientific. *Overview of ELISA*. `https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html`. Accessed: 2025-01-12. 2025.

[72] Bio-Rad Antibodies. *An Introduction to ELISA*. `https://www.bio-rad-antibodies.com/an-introduction-to-elisa.html`. Accessed: 2025-01-12. 2025.

[73] Behnam M et al. *COVID-19: Overcoming supply shortages for diagnostic testing*. 2020. URL: `https://www.mckinsey.com/industries/life-sciences/our-insights/covid-19-overcoming-supply-shortages-for-diagnostic-testing`.

[74] Naru Zhang et al. "Recent advances in the detection of respiratory virus infection in humans". In: *Journal of medical virology* 92.4 (2020), pp. 408–417.

[75] Michael J Mina and Kristian G Andersen. "COVID-19 testing: One size does not fit all". In: *Science* 371.6525 (2021), pp. 126–127.

[76] Michael M Kaminski et al. "CRISPR-based diagnostics". In: *Nature Biomedical Engineering* 5.7 (2021), pp. 643–656.

[77] Frank Hille et al. "The Biology of CRISPR-Cas: Backward and Forward". In: *Cell* 172 (2018), pp. 1239–1259. DOI: `10.1016/j.cell.2017.11.032`.

[78] Fei Wang et al. "Advances in CRISPR-Cas systems for RNA targeting, tracking and editing." In: *Biotechnology advances* (2019). DOI: `10.1016/j.biotechadv.2019.03.016`.

[79] Jonathan S Gootenberg et al. "Nucleic acid detection with CRISPR-Cas13a/C2c2". In: *Science* 356.6336 (2017), pp. 438–442.

[80] Max J Kellner et al. "SHERLOCK: nucleic acid detection with CRISPR nucleases". In: *Nature protocols* 14.10 (2019), pp. 2986–3012.

[81] Anastasiya Kostyusheva et al. "CRISPR-Cas systems for diagnosing infectious diseases". In: *Methods* 203 (2022), pp. 431–446.

[82] Mujahed I Mustafa and Abdelrafie M Makhawi. "SHERLOCK and DETECTR: CRISPR-Cas systems as potential rapid diagnostic tools for emerging infectious diseases". In: *Journal of Clinical Microbiology* 59.3 (2021), pp. 10–1128.

[83] Janice S Chen et al. "CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity". In: *Science* 360.6387 (2018), pp. 436–439.

[84] Jen-Hui Tsou, Qixin Leng, and Feng Jiang. "A CRISPR test for detection of circulating nuclei acids". In: *Translational oncology* 12.12 (2019), pp. 1566–1573.

[85] Yangyang Sun et al. "One-tube SARS-CoV-2 detection platform based on RT-RPA and CRISPR/Cas12a". In: *Journal of translational medicine* 19 (2021), pp. 1–10.

[86] A. Greninger et al. "Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis". In: *Genome Medicine* 7 (2015). DOI: `10.1186/s13073-015-0220-9`.

[87] David Deamer, Mark Akeson, and Daniel Branton. "Three decades of nanopore sequencing". In: *Nature biotechnology* 34.5 (2016), pp. 518–524.

[88] Andrea D Tyler et al. “Evaluation of Oxford Nanopore’s MinION sequencing device for microbial whole genome sequencing applications”. In: *Scientific reports* 8.1 (2018), p. 10931.

[89] Thomas Hoenen et al. “Nanopore sequencing as a rapidly deployable Ebola outbreak tool”. In: *Emerging infectious diseases* 22.2 (2016), p. 331.

[90] Joshua Quick et al. “Real-time, portable genome sequencing for Ebola surveillance”. In: *Nature* 530.7589 (2016), pp. 228–232.

[91] Jaqueline Goes De Jesus et al. “Acute vector-borne viral infection: Zika and MinION surveillance”. In: *Microbiology spectrum* 7.4 (2019), pp. 10–1128.

[92] Ming Wang et al. “Nanopore targeted sequencing for the accurate and comprehensive detection of SARS-CoV-2 and other respiratory viruses”. In: *Small* 16.32 (2020), p. 2002169.

[93] Adrian A Pater et al. “High throughput nanopore sequencing of SARS-CoV-2 viral genomes from patient samples”. In: *Journal of biological methods* 8.COVID 19 Spec Iss (2021).

[94] Rajinder Singh and Falk Riess. “Sir CV Raman and the story of the Nobel prize”. In: *Current Science* 75.9 (1998), pp. 965–971.

[95] George C Schatz, Matthew A Young, and Richard P Van Duyne. “Electromagnetic mechanism of SERS”. In: *Surface-enhanced Raman scattering: physics and applications*. Springer, 2006, pp. 19–45.

[96] Li-Mei C Yang et al. “Direct electrical transduction of antibody binding to a covalent virus layer using electrochemical impedance.” In: *Analytical chemistry* 80 15 (2008), pp. 5695–705. DOI: `10.1021/ac8008109`.

[97] Jin-Ho Lee, Byeung-Keun Oh, and Jeong-woo Choi. “Electrochemical sensor based on direct electron transfer of HIV-1 virus at Au nanoparticle modified ITO electrode.” In: *Biosensors & bioelectronics* 49 (2013), pp. 531–5. DOI: `10.1016/j.bios.2013.06.010`.

[98] M. Alafeef et al. “Rapid, Ultrasensitive, and Quantitative Detection of SARS-CoV-2 Using Antisense Oligonucleotides Directed Electrochemical Biosensor Chip”. In: *ACS Nano* (2020). DOI: `10.1021/acsnano.0c06392`.

[99] B. Mojsoska et al. “Rapid SARS-CoV-2 Detection Using Electrochemical Immunosensor”. In: *Sensors (Basel, Switzerland)* 21 (2021). DOI: `10.3390/s21020390`.

[100] Fedor N Dultsev and Andrei V Tronin. “Rapid sensing of hepatitis B virus using QCM in the thickness shear mode”. In: *Sensors and Actuators B: Chemical* 216 (2015), pp. 1–5.

[101] Kai Wu et al. “Giant magnetoresistance biosensors in biomedical applications”. In: *ACS applied materials & interfaces* 14.8 (2022), pp. 9945–9969.

[102] Bert Vogelstein and Kenneth W Kinzler. “Digital pcr”. In: *Proceedings of the National Academy of Sciences* 96.16 (1999), pp. 9236–9241.

[103] Tao Suo et al. "ddPCR: a more accurate tool for SARS-CoV-2 detection in low viral load specimens". In: *Emerging microbes & infections* 9.1 (2020), pp. 1259–1268.

[104] Amir Asri Kojabad et al. "Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives". In: *Journal of Medical Virology* 93.7 (2021), pp. 4182–4197.

[105] Alexandra S Whale et al. "Detection of rare drug resistance mutations by digital PCR in a human influenza a virus model system and clinical samples". In: *Journal of clinical microbiology* 54.2 (2016), pp. 392–400.

[106] Diego R Hijano et al. "Clinical correlation of influenza and respiratory syncytial virus load measured by digital PCR". In: *PloS one* 14.9 (2019), e0220908.

[107] Alexander J Veach et al. "Digital droplet PCR for influenza vaccine development". In: *Procedia in Vaccinology* 9 (2015), pp. 96–103.

[108] Shivangi Sachdeva, Ronald W Davis, and Amit K Saha. "Microfluidic point-of-care testing: commercial landscape and future directions". In: *Frontiers in Bioengineering and Biotechnology* 8 (2021), p. 602659.

[109] Ghazaleh Jamalipour Soufi, Siavash Iravani, and Rajender S Varma. "Molecularly imprinted polymers for the detection of viruses: Challenges and opportunities". In: *Analyst* 146.10 (2021), pp. 3087–3100.

[110] Aijaz Ahmad Malik, Chanin Nantasenamat, and Theeraphon Piacham. "Molecularly imprinted polymer for human viral pathogen detection". In: *Materials Science and Engineering: C* 77 (2017), pp. 1341–1348.

[111] Akinrinade George Ayankojo et al. "Molecularly imprinted polymer based electrochemical sensor for quantitative detection of SARS-CoV-2 spike protein". In: *Sensors and Actuators B: Chemical* 353 (2022), p. 131160.

[112] Thipvaree Wangchareansak et al. "Influenza A virus molecularly imprinted polymers and their application in virus sub-type classification". In: *Journal of Materials Chemistry B* 1.16 (2013), pp. 2190–2197.

[113] Simon Ardui et al. "Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics". In: *Nucleic acids research* 46.5 (2018), pp. 2159–2168.

[114] Jun Yan et al. "Single-molecule sequencing assists genome assembly improvement and structural variation inference". In: *Molecular Plant* 9.7 (2016), pp. 1085–1087.

[115] Sébastien Lhomme et al. "Prediction of SARS-CoV-2 variant lineages using the S1-encoding region sequence obtained by PacBio single-molecule real-time sequencing". In: *Viruses* 13.12 (2021), p. 2544.

[116] Deniz Sadighbayan, Mohammad Hasanzadeh, and Ebrahim Ghafar-Zadeh. "Biosensing based on field-effect transistors (FET): Recent progress and challenges". In: *TrAC Trends in Analytical Chemistry* 133 (2020), p. 116067.

[117] Mazin A Zamzami et al. “Carbon nanotube field-effect transistor (CNT-FET)-based biosensor for rapid detection of SARS-CoV-2 (COVID-19) surface spike protein S1”. In: *Bioelectrochemistry* 143 (2022), p. 107982.

[118] Giwan Seo et al. “Rapid detection of COVID-19 causative virus (SARS-CoV-2) in human nasopharyngeal swab specimens using field-effect transistor-based biosensor”. In: *ACS nano* 14.4 (2020), pp. 5135–5142.

[119] Andres W Martinez et al. *Diagnostics for the developing world: microfluidic paper-based analytical devices*. 2010.

[120] Muhammad Hatta Prabowo et al. “Dengue NS1 detection in pediatric serum using microfluidic paper-based analytical devices”. In: *Analytical and bioanalytical chemistry* 412 (2020), pp. 2915–2925.

[121] Bhagwan S Batule, Youngung Seok, and Min-Gon Kim. “based nucleic acid testing system for simple and early diagnosis of mosquito-borne RNA viruses from human serum”. In: *Biosensors and Bioelectronics* 151 (2020), p. 111998.

[122] Savka I Stoeva et al. “Multiplexed DNA detection with biobarcoded nanoparticle probes”. In: *Angewandte Chemie* 118.20 (2006), pp. 3381–3384.

[123] Ding Sheng-zi Zhang et al. “Dual-encoded microbeads through a host–guest structure: enormous, flexible, and accurate barcodes for multiplexed assays”. In: *Advanced Functional Materials* 26.34 (2016), pp. 6146–6157.

[124] Xiaodong Lin et al. “Nanozyme-based bio-barcode assay for high sensitive and logic-controlled specific detection of multiple DNAs”. In: *Biosensors and Bioelectronics* 94 (2017), pp. 471–477.

[125] Affinité Instruments. *SPR Sensorgram Explained.* `https://www.affiniteinstruments.com/post/spr-sensorgram-explained`. Accessed: 2025-01-12. 2025.

[126] Anand M Shrivastav, Uroš Cvelbar, and Ibrahim Abdulhalim. “A comprehensive review on plasmonic-based biosensors used in viral diagnostics”. In: *Communications biology* 4.1 (2021), p. 70.

[127] Svitlana Zagorodnya et al. “Application of SPR Analysis for Detection of Specific Antibodies in Human Blood Serum”. In: *3rd International Electronic Conference on Sensors and Applications*. Multidisciplinary Digital Publishing Institute. 2016.

[128] Paula Gomes and David Andreu. “Direct kinetic assay of interactions between small peptides and immobilized antibodies using a surface plasmon resonance biosensor”. In: *Journal of immunological methods* 259.1-2 (2002), pp. 217–230.

[129] RF Begley, AB Harvey, and Robert L Byer. “Coherent anti-Stokes Raman spectroscopy”. In: *Applied Physics Letters* 25.7 (1974), pp. 387–390.

[130] Xiaolin Nan et al. “Intracellular imaging of HCV RNA and cellular lipids by using simultaneous two-photon fluorescence and coherent anti-Stokes Raman scattering microscopies”. In: *ChemBioChem* 7.12 (2006), pp. 1895–1897.

[131] Tanveer A Tabish, Roger J Narayan, and Mohan Edirisinghe. "Rapid and label-free detection of COVID-19 using coherent anti-Stokes Raman scattering microscopy". In: *MRS communications* 10.4 (2020), pp. 566–572.

[132] K. Santosh et al. "A systematic review on cough sound analysis for Covid-19 diagnosis and screening: is my cough sound COVID-19?" In: *PeerJ Computer Science* 8 (2022). DOI: `10.7717/peerj-cs.958`.

[133] Lara Orlandic, Tomas Teijeiro, and David Atienza. "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms". In: *Scientific Data* 8.1 (2021), p. 156.

[134] Yoshihiro Minagawa et al. "Mobile imaging platform for digital influenza virus counting." In: *Lab on a chip* (2019). DOI: `10.1039/c9lc00370c`.

[135] A. Shokr et al. "Mobile Health (mHealth) Viral Diagnostics Enabled with Adaptive Adversarial Learning." In: *ACS nano* (2020). DOI: `10.1021/acsnano.0c06807`.

[136] M. Greenleaf et al. "Pivot, persevere, or perish: how Ellume Health overcame development and regulatory obstacles to become the first authorized over-the-counter COVID-19 test in the United States." In: *Lab on a chip* (2023). DOI: `10.1039/d3lc00118k`.

[137] R. Rubin. "COVID-19 Testing Moves Out of the Clinic and Into the Home." In: *JAMA* (2021). DOI: `10.1001/jama.2021.15679`.

[138] Suzanna E. Schmeelk et al. "Monitoring Symptoms of COVID-19: Review of Mobile Apps". In: *JMIR mHealth and uHealth* 10 (2022). DOI: `10.2196/36065`.

[139] Alessia Visconti et al. "Cutaneous manifestations of SARS-CoV-2 infection during the Delta and Omicron waves in 348 691 UK users of the UK ZOE COVID Study app". In: *British Journal of Dermatology* 187.6 (2022), pp. 900–908.

[140] Beatrice Kennedy et al. "App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID Symptom Study Sweden". In: *Nature communications* 13.1 (2022), p. 2110.

[141] Willian J van Dijk et al. "COVID RADAR app: description and validation of population surveillance of symptoms and behavior in relation to COVID-19". In: *PLoS One* 16.6 (2021), e0253566.

[142] Hallam Stevens and Monamie Bhadra Haines. "Tracetogether: pandemic response, democracy, and technology". In: *East Asian Science, Technology and Society: An International Journal* 14.3 (2020), pp. 523–532.

[143] World Health Organization. *ASSURED criteria: A guide to aid the selection of diagnostic tests.* Accessed: [Insert Access Date Here]. 2016. URL: `https://apps.who.int/iris/rest/bitstreams/1133796/retrieve`.

[144] Cindy D. Chin, Vicky Linder, and Samuel K. Sia. “REASSURED Multiplex Diagnostics: A Critical Review and Forecast”. In: *Biosensors* 12.2 (2022). Accessed: [Insert Access Date Here], p. 124. DOI: `10.3390/bios12020124`. URL: `https://www.mdpi.com/2079-6374/12/2/124`.

[145] George Gabriel Stokes. “XXX. On the change of refrangibility of light”. In: *Philosophical transactions of the Royal Society of London* 142 (1852), pp. 463–562.

[146] Oskar Heimstädt. “Das fluoreszenzmikroskop”. In: (1911).

[147] Marvin Minsky. “Memoir on inventing the confocal scanning microscope”. In: *Scanning* 10.4 (1988), pp. 128–138.

[148] WB Amos and JG White. “How the confocal laser scanning microscope entered biological research”. In: *Biology of the Cell* 95.6 (2003), pp. 335–342.

[149] Roger Y Tsien. “The green fluorescent protein”. In: *Annual review of biochemistry* 67.1 (1998), pp. 509–544.

[150] Nadya I Tarasova et al. “Visualization of G protein-coupled receptor trafficking with the aid of the green fluorescent protein: endocytosis and recycling of cholecystokinin receptor type A”. In: *Journal of Biological Chemistry* 272.23 (1997), pp. 14817–14824.

[151] Eric Betzig et al. “Imaging intracellular fluorescent proteins at nanometer resolution”. In: *science* 313.5793 (2006), pp. 1642–1645.

[152] Michael J Rust, Mark Bates, and Xiaowei Zhuang. “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)”. In: *Nature methods* 3.10 (2006), pp. 793–796.

[153] Stefan W Hell. “Far-field optical nanoscopy”. In: *science* 316.5828 (2007), pp. 1153–1158.

[154] Peter A Santi. “Light sheet fluorescence microscopy: a review”. In: *Journal of Histochemistry & Cytochemistry* 59.2 (2011), pp. 129–138.

[155] Martin Booth et al. “Aberrations and adaptive optics in super-resolution microscopy”. In: *Microscopy* 64.4 (2015), pp. 251–261.

[156] Susanne CM Reinhardt et al. “Ångström-resolution fluorescence microscopy”. In: *Nature* 617.7962 (2023), pp. 711–716.

[157] James Pawley. *Handbook of biological confocal microscopy*. Vol. 236. Springer Science & Business Media, 2006.

[158] KR Spring and S Inoué. *Video microscopy: the fundamentals*. 1997.

[159] Makio Tokunaga, Naoko Imamoto, and Kumiko Sakata-Sogawa. “Highly inclined thin illumination enables clear single-molecule imaging in cells”. In: *Nature methods* 5.2 (2008), pp. 159–161.

[160] Hendrik Deschout and Kevin Braeckmans. “Single Particle Tracking”. In: *Encyclopedia of Biophysics*. Ed. by Gordon C. K. Roberts. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 2326–2327. DOI: `10.1007/978-3-642-16712-6_821`. URL: `https://doi.org/10.1007/978-3-642-16712-6_821`.

[161] Kenneth N. Fish. “Total Internal Reflection Fluorescence (TIRF) Microscopy”. In: *Current Protocols in Cytometry* 50.1 (2009), pp. 12.18.1–12.18.13. DOI: `https://doi.org/10.1002/0471142956.cy1218s50`. eprint: `https://currentprotocols.onlinelibrary.wiley.com/doi/pdf/10.1002/0471142956.cy1218s50`. URL: `https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471142956.cy1218s50`.

[162] Stefan W Hell and Jan Wichmann. “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy”. In: *Optics letters* 19.11 (1994), pp. 780–782.

[163] Mats GL Gustafsson. “Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy”. In: *Journal of microscopy* 198.2 (2000), pp. 82–87.

[164] Fei Chen, Paul W Tillberg, and Edward S Boyden. “Expansion microscopy”. In: *Science* 347.6221 (2015), pp. 543–548.

[165] Bo Huang et al. “Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy”. In: *Science* 319.5864 (2008), pp. 810–813.

[166] Adish Dani et al. “Superresolution imaging of chemical synapses in the brain”. In: *Neuron* 68.5 (2010), pp. 843–856.

[167] Arthur L Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of research and development* 3.3 (1959), pp. 210–229.

[168] Frank Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” In: *Psychological review* 65.6 (1958), p. 386.

[169] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine learning* 20 (1995), pp. 273–297.

[170] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.

[171] Meherwar Fatima, Maruf Pasha, et al. “Survey of machine learning algorithms for disease diagnostic”. In: *Journal of Intelligent Learning Systems and Applications* 9.01 (2017), p. 1.

[172] Holger A Haenssle et al. “Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists”. In: *Annals of oncology* 29.8 (2018), pp. 1836–1842.

[173] Adriaan Coenen et al. “Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography–based fractional flow reserve: result from the MACHINE consortium”. In: *Circulation: Cardiovascular Imaging* 11.6 (2018), e007217.

[174] Daniel Bone et al. "Applying machine learning to facilitate autism diagnostics: pitfalls and promises". In: *Journal of autism and developmental disorders* 45 (2015), pp. 1121–1136.

[175] Varun Mannam et al. "Machine learning for faster and smarter fluorescence lifetime imaging microscopy". In: *Journal of Physics: Photonics* 2.4 (2020), p. 042005.

[176] Hatice Ceylan Koydemir et al. "Comparison of supervised machine learning algorithms for waterborne pathogen detection using mobile phone fluorescence microscopy". In: *Nanophotonics* 6.4 (2017), pp. 731–741.

[177] Allison Chia-Yi Wu and Scott A Rifkin. "Aro: a machine learning approach to identifying single molecules and estimating classification error in fluorescence microscopy images". In: *BMC bioinformatics* 16.1 (2015), pp. 1–8.

[178] Daniela M Borgmann et al. "Single molecule fluorescence microscopy and machine learning for rhesus D antigen classification". In: *Scientific Reports* 6.1 (2016), p. 32317.

[179] Nicole C Robb et al. "Rapid functionalisation and detection of viruses via a novel Ca2+-mediated virus-DNA interaction". In: *Scientific reports* 9.1 (2019), p. 16219.

[180] Rosa Casais et al. "Reverse genetics system for the avian coronavirus infectious bronchitis virus". In: *Journal of virology* 75.24 (2001), pp. 12359–12369.

[181] Ramy Arnaout et al. "The limit of detection matters: the case for benchmarking severe acute respiratory syndrome coronavirus 2 testing". In: *Clinical Infectious Diseases* 73.9 (2021), e3042–e3046.

[182] Yang Pan et al. "Viral load of SARS-CoV-2 in clinical samples". In: *The Lancet infectious diseases* 20.4 (2020), pp. 411–412.

[183] Annoor Awadasseid et al. "Initial success in the identification and management of the coronavirus disease 2019 (COVID-19) indicates human-to-human transmission in Wuhan, China". In: *International journal of biological sciences* 16.11 (2020), p. 1846.

[184] B Michen and T Graule. "Isoelectric points of viruses". In: *Journal of applied microbiology* 109.2 (2010), pp. 388–397.

[185] Christin Scheller et al. "Physicochemical properties of SARS-CoV-2 for drug targeting, virus inactivation and attenuation, vaccine formulation and quality control". In: *Electrophoresis* 41.13-14 (2020), pp. 1137–1151.

[186] Max Born. "Quantenmechanik der stoßvorgänge". In: *Zeitschrift für physik* 38.11-12 (1926), pp. 803–827.

[187] Luke A Clifton et al. "An accurate in vitro model of the E. coli envelope". In: *Angewandte Chemie (International ed. in English)* 54.41 (Oct. 5, 2015), pp. 11952–11955. DOI: `10.1002/anie.201504287`. URL: `https://pubmed.ncbi.nlm.nih.gov/26331292`.

[188] Arwel V. Hughes et al. “Physical Properties of Bacterial Outer Membrane Models: Neutron Reflectometry & Molecular Simulation”. In: *Biophysical Journal* 116.6 (Mar. 19, 2019), pp. 1095–1104. DOI: `10.1016/j.bpj.2019.02.001`. URL: `https://doi.org/10.1016/j.bpj.2019.02.001`.

[189] Andrew R. McCluskey et al. “Bayesian determination of the effect of a deep eutectic solvent on the structure of lipid monolayers”. In: *Physical Chemistry Chemical Physics* 21.11 (2019), pp. 6133–6141. DOI: `10.1039/c9cp00203k`. URL: `https://doi.org/10.1039/c9cp00203k`.

[190] Peng Chen et al. “Tailoring the magnetic exchange interaction in MnBi2Te4 superlattices via the intercalation of ferromagnetic layers”. In: *Nature Electronics* (Dec. 2022). DOI: `10.1038/s41928-022-00880-1`. URL: `https://doi.org/10.1038/s41928-022-00880-1`.

[191] C. F. Majkrzak and N. F. Berk. “Exact determination of the phase in neutron reflectometry”. In: *Phys. Rev. B* 52 (15 Oct. 1995), pp. 10827–10830. DOI: `10.1103/PhysRevB.52.10827`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.52.10827`.

[192] D. S. Sivia. *Elementary Scattering Theory: For X-ray and Neutron Users*. Oxford University Press, 2011. ISBN: 9780199228676.

[193] V. F. Sears. “Generalized distorted-wave Born approximation for neutron reflection”. In: *Physical Review B* 48.23 (Dec. 1993), pp. 17477–17485. DOI: `10.1103/physrevb.48.17477`. URL: `https://doi.org/10.1103/physrevb.48.17477`.

[194] O. Glatter. “A new method for the evaluation of small-angle scattering data”. In: *Journal of Applied Crystallography* 10.5 (Oct. 1977), pp. 415–421. DOI: `10.1107/s0021889877013879`. URL: `https://doi.org/10.1107/s0021889877013879`.

[195] C.F. Majkrzak et al. “Phase determination and inversion in specular neutron reflectometry”. In: *Physica B: Condensed Matter* 248.1-4 (June 1998), pp. 338–342. DOI: `10.1016/s0921-4526(98)00260-9`. URL: `https://doi.org/10.1016/s0921-4526(98)00260-9`.

[196] Ming Li, M. O. Möller, and G. Landwehr. “Fitting of x-ray or neutron specular reflectivity of multilayers by Fourier analysis”. In: *Journal of Applied Physics* 80.5 (Sept. 1996), pp. 2788–2790. DOI: `10.1063/1.363196`. URL: `https://doi.org/10.1063/1.363196`.

[197] F Bridou and B Pardo. “Grazing x-ray reflectometry data processing by fourier transform”. In: *Journal of X-Ray Science and Technology* 4.3 (Sept. 1994), pp. 200–216. DOI: `10.1016/s0895-3996(05)80058-8`. URL: `https://doi.org/10.1016/s0895-3996(05)80058-8`.

[198] Andrew R McCluskey et al. “A general approach to maximise information density in neutron reflectometry analysis”. In: *Machine Learning: Science and Technology* 1.3 (July 2020), p. 035002. DOI: `10.1088/2632-2153/ab94c4`. URL: `https://doi.org/10.1088/2632-2153/ab94c4`.

[199] John Skilling. "Nested Sampling". In: *AIP Conference Proceedings* 735.1 (2004), pp. 395–405. DOI: `10.1063/1.1835238`. eprint: `https://aip.scitation.org/doi/pdf/10.1063/1.1835238`. URL: `https://aip.scitation.org/doi/abs/10.1063/1.1835238`.

[200] Luke A. Clifton et al. "Self-Assembled Fluid Phase Floating Membranes with Tunable Water Interlayers". In: *Langmuir* 35.42 (Oct. 22, 2019), pp. 13735–13744. DOI: `10.1021/acs.langmuir.9b02350`. URL: `https://doi.org/10.1021/acs.langmuir.9b02350`.

[201] G Fragneto et al. "A fluid floating bilayer". In: *Europhysics Letters* 53.1 (2001), p. 100.

[202] J Daillant et al. "Structure and fluctuations of a single floating lipid bilayer". In: *Proceedings of the National Academy of Sciences* 102.33 (2005), pp. 11639–11644.

[203] Luke A. Clifton et al. "An Accurate In Vitro Model of the E. coli Envelope". In: *Angewandte Chemie International Edition* 54.41 (2015), pp. 11952–11955. DOI: `https://doi.org/10.1002/anie.201504287`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201504287`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201504287`.

[204] F. Abelès. "La théorie générale des couches minces". In: *Journal de Physique et le Radium* 11.7 (1950), pp. 307–309. DOI: `10.1051/jphysrad:01950001107030700`. URL: `https://doi.org/10.1051/jphysrad:01950001107030700`.

[205] L. Névot and P. Croce. "Caractérisation des surfaces par réflexion rasante de rayons X. Application à l'étude du polissage de quelques verres silicates". In: *Revue de Physique Appliquée* 15.3 (1980), pp. 761–779. DOI: `10.1051/rphysap:01980001503076100`. URL: `https://doi.org/10.1051/rphysap:01980001503076100`.

[206] Greg Ashton et al. "Nested sampling for physical scientists". In: *Nature Reviews Methods Primers* 2.1 (2022), p. 39.

[207] Harold Jeffreys. *The theory of probability.* OuP Oxford, 1998.

[208] D. S. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial.* Oxford science publications. OUP Oxford, 2006. ISBN: 9780191546709. URL: `https://books.google.co.uk/books?id=Kxx8CwAAQBAJ`.

[209] Hanna P Vacklin et al. "Composition of supported model membranes determined by neutron reflection". In: *Langmuir* 21.7 (2005), pp. 2827–2837.