UNIVERSITY
OF SUSSEX

Doctoral Thesis

Abnormality Detection Using Molecular Communications Based Nano-scale Sensor Networks

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

in the

School of Engineering and Informatics

Author:
Sinem Nimet Solak

Supervisor:
Menguc Oner

April 20, 2022

Gratitude, my dear Serap Cakici, for constantly being my charming gardener who make my soul blossom.

To my family, the reason what I become today. The values they thought prepared me for a humane and meaningful life.

To my grandfather Nihat Kayadelen who always belived in me.

Acknowledgements

I am sincerely grateful for the guidance and support of my supervisor, Dr. Menguc Oner. Throughout the previous four years, he has been a source of encouragement and support for me. Without that, this thesis would not have been possible.

I would also like to appreciate all contributions from my second supervisors during my doctoral study; Prof. Maziar Nekovee, Prof. Falah Ali and this gratitude also extends to my colleagues; Ximiao Wang, Victor Ocheri.

I am grateful to the members of my examining committee for their time, opinions, and interest in my work.

Finally, I would like to say thank you to my parents; Osman and Bilge, my brothers; Abdullah, Adnan for their continuous support over the years. Last but not least, I would like to thank my greatest supporter Dr. Serap Cakici.

I, Sinem Nimet Solak, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

Brighton,

April 20, 2022

Sinem Nimet Solak

University of Sussex

School of Engineering and Informatics

Doctoral Thesis

Abnormality Detection Using Molecular Communications Based Nano-scale Sensor Networks

by Sinem Nimet Solak

Abstract

Abnormality detection is one of the most highly anticipated application areas of Molecular Communication (MC) based nanonetworks. This task entails sensing, detection, and reporting of abnormal changes in a fluid medium that may characterize a disease or disorder using a network of collaborating nanoscale sensors. Such distributed detection (DD) problems are of paramount interest in applications of nanonetworks. For the first time in literature, we proposed to employ sequential probability ratio test (SPRT) to decision fusion (DF). The proposed approach yields considerable gains in the average number of samples required for the decision resulting in significant improvement in decision delay, which is one of the main challenges encountered in a molecular communications based sensor network. Existing strategies for such distributed collaborative detection problems require a complete statistical characterization of the underlying communication channel between the sensors and the fusion centre (FC), with the assumption of perfectly-known or accurately estimated channel parameters. This assumption is usually impractical both due to mathematical intractability of the analytical channel models for MC except in a few ideal cases, and the slow and dispersive signal propagation characteristics that make the channel estimation a difficult task even in these ideal cases. This work, for the first time in the literature, proposes to employ a machine learning (ML) approach to this task and shows that

this approach provides the robustness and flexibility required for practical implementation. We focus on detection based on deep learning, specifically on a feed-forward neural network and a recurrent neural network structure that learn the underlying model from data. This study shows that the proposed DF strategy can perform well without any knowledge of the communication channel.

Contents

List of Figures

List of Tables

List of Abbreviations

ALLR	Average Log-likelihood Ratio
ANN	Artificial Neural Networks
APS	Average Percentage Saving
ASN	Average Sample Number
AUC	Area Under the Receiver Operating Characteristic Curve
AWGN	Additive White Gaussian Noise
BIOMEMS	Biomedical (or biological) Microelectromechanical
CSI	Channel State Information
DD	Distributed Detection
DF	Decision Fusion
DNA	Deoxyribonucleic Acid
DNN	Deep Neural Networks
EM	Electromagnetic
ER	Endoplasmic Reticulum

FC	Fusion Centre
FN	False Negative
FP	False Positive
ID3	Iterative Dichotomiser 3
IoBNT	Internet of Bio-Nano Things
IoE	Internet of Everything
IoT	Internet of Things
ISI	Inter-symbol Interference
KNN	K-Nearest-Neighbour
LLR	Log- likelihood ratio
LSTM	Long Short Term Memory
MC	Molecular Communication
MEMS	Micro-electro-mechanical Systems
ML	Machine Learning
MLL	Maximum likelihood
MLP	Multilayer Perceptrons
NEMS	Nano-electro-mechanical Systems
Systems	
NN	Neural Networks
OC	Operating Characteristics
PMF	Probability Mass Function
RNA	Ribonucleic Acid

RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
RV	Random Variables
SAPRT	Sequential Average Probability Ratio Test
SNR	Signal to Noise Ratio
SPRT	Sequential Probability Ratio Test
SVM	Support Vector Machine
TP	True Positive
TN	True Negative

Publications

Sinem Solak and Mengüç Öner. Sequential Decision Fusion for Abnormality Detection via Diffusive Molecular Communications. *IEEE Communications Letters*, 25(3):825–829, 2021.

Sinem Nimet Solak and Menguc Oner. Neural Network Based Decision Fusion for Abnormality Detection via Molecular Communications. In *2020 IEEE Workshop on Signal Processing Systems (SiPS)*, pages 1–5, 2020.

Sinem Nimet Solak and Menguc Oner. RNN Based Abnormality Detection with Nanoscale Sensor Networks Using Molecular Communications. In *Proceedings of the 7th ACM International Conference on Nanoscale Computing and Communication*, NanoCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

Introduction

1

If one day, my words are against science, choose science.

Mustafa Kemal Ataturk

1.1 Overview

Nanotechnology is the process of creating and implementing physical, chemical, and biological systems on sizes ranging from individual atoms or molecules to sub-micron dimensions, and the incorporation of the resulting nanostructures into larger systems capable of performing tasks for a wide range of engineering applications [3]. Researchers and industry aim to harness the unique features of nanomaterials and nanomanufacturing technologies for medical and healthcare applications, materials science, computing and electronics, energy applications and environmental benefits, and so on [4]. Therefore, it has been an attractive and promising research area in recent years, drawing a considerable research effort from diverse fields of science and engineering. With the progress in science and technology, fabrication of micro/nanodevices and systems has advanced for a variety of industrial, consumer, and healthcare applications. New study disciplines such as micro- and nano-electro-mechanical systems (MEMS and NEMS) have been established to extend nanotechnology into robotics. The field of biomedical (or biological) microelectromechanical systems (BIOMEMS) is primarily concentrated on mechanical compo-

nents and microfabrication methods adapted for biological applications such as biofluidic chips, biosensors, etc. [5]. Besides man-made machines, the ability to directly reuse biological structures found in living organisms or to reengineer them is expected to be particularly beneficial in biomedical applications, as well as providing the foundation for bio-inspired communications using bioengineering [6].

Many of the nanotechnology applications envisaged require the cooperation and collaboration of a number of nano-devices. In this regard, nanonetworks, i.e. networks that enable the information exchange between nano-devices such as sensors and actuators for performing tasks that require coordination and cooperation, have become an active research area.

Inspired by the largely successful paradigm of the internet of things (IoT) that enables ubiquitous and pervasive connectivity and interaction between devices on the macroscopic scale, the recently introduced concept of the Internet of Bio-Nano Things (IoBNT) is expected to enable the connectivity between nanodevices operating within biological or other systems, complementing and interfacing with the IoT [7], thus, bridging the gap between the macroscopic and microscopic worlds and elevating the pervasiveness of the connectivity of things into unprecedented levels. The IoBNT has the potential to extend the connectivity to microscopic scales, even to biological systems such as the human body or living ecosystems, with an ultimate vision of seamless integration of the macroscale domain with the microscale biochemical domain into a multiscale pervasive Internet of Everything (IoE) [8].

Different communication paradigms can be employed in nanonetworks, depending on the particular application and the type of medium where the nanonetwork is expected to be deployed. In environments where nanonetworks are to be deployed, such as biological systems, conventional synthetic methodologies for communication, such as radio frequency (RF) transmission, may be unsafe or infeasible such as in a biological system. As previously mentioned, some of the most highly anticipated applications envisioned for nanonetworks in healthcare, environmental science and many other fields require the deployment of nanonetworks within a biological system. The environment has a detrimental effect on the propagation of electromagnetic (EM) waves. Also, EM is difficult at extremely small dimensions, such as between microscale or nanoscale robots, due to limits such as the antenna size ratio to the electromagnetic signal's wavelength [9]. Moreover, the use of high-frequency EM radiation for communication may be undesirable for applications within a living organism due to health concerns. For such cases, MC may be bio-compatible alternative to the conventional EM-based wireless communication paradigm by employing dedicated molecules as information carriers, mimicking the naturally evolved communication mechanisms between biological entities at this physical scale [1].

In the Literature, many different mechanisms have been considered for signal propagation in MC, including chemotactic signalling [10], quorum sensing [11], and propagation through gap junction [12]. Specifically, we focused on the diffusive MC approach as it has garnered considerable interest from the communications research community [13–32]. Diffusive MC relies on Brownian motion for signal propagation, i.e. the net passive movement of molecules or particles from high to low concentration areas. The main advantages of this mechanism are that it requires no infrastructure and additional external energy [33]. There has been a considerable amount of research in recent years on different elements of MC systems, such as transmitter and receiver design [19], [15], [27], network layer [1] and various access protocols [1].

This thesis investigates the task of abnormality detection, i.e. the detection and reporting of abnormal changes in a fluid environment that may characterize the presence of a disease or disorder using a diffusive MC based nanoscale sensor network that is still in its early phases of development. We encounter such decision fusion (DD) problems in many potential application areas of nanonetworks, such as health monitoring, disease diagnosis, targeted drug delivery, environmental sensing and monitoring, contaminant and toxic agent detection. Some of the anomalies that may be of interest in the context of this task include abnormal changes in the concentration of an indicator molecule in the medium or in the parameters of the medium itself, such as the pH value, temperature, pressure, and viscosity [34], [35].

Using a DD approach in which a network of connected sensor nodes collaborates to distinguish between hypotheses representing the presence and absence of an abnormality, is a widely used technique for detection problems requiring a high degree of accuracy, reliability, and robustness. On macroscopic dimensions, wireless sensor networks utilizing traditional EM-based wireless communications have attracted intense research interest for decades and continue to do so [36]. However, implementing a sensor network at the nanoscale using MC involves novel and unique obstacles, because of the nature of signal propagation in the diffusive MC channel, which is significantly different from that of the well-studied conventional wireless communications channel.

Some processing can be done at the individual sensors in nanoscale sensor networks, while compressed forms of local sensor decisions are transmitted to a fusion centre (FC), where the incoming signals from various sensors are suitably combined to carry out the final global decision. We particularly focus on the problem of decision fusion (DF) in an MC based nanoscale sensor network, i.e. the task of merging the sensing decisions transmitted by the sensor nodes via the diffusive MC channel to provide a global inference regarding the presence or absence of the abnormality of interest.

1.2 Motivation and Challenges

MC draws attention with the rapid and advanced development in nanoscale communication networks. Employing non-intrusive MC based nanoscale sensor networks within the human body in order to sense and report abnormal and unexpected physical and biochemical changes in their environment has the potential to enable applications such as health monitoring, disease diagnosis and targeted drug delivery that could transform healthcare and medical treatments by fundamentally changing the ways these services are provided. Real-time health monitoring within the human body and early diagnosis of diseases or disorders well before symptoms are detectable from outside the body could lead to earlier treatment with a much higher possibility of success, contributing to society's health and well-being and significantly reducing healthcare costs by avoiding the need for invasive tests and interventions, unnecessary medication or hospital admission. This technology has the potential to transform the way the drug treatments are carried out by enabling precise targeting of the drug delivery to the tissue where a disease or disorder has been detected, allowing to lower required doses of drugs, limiting the extent of detrimental side-effects to the rest of the body and leading to more cost-effective and efficient drug treatments.

Nanoscale sensor networks can be employed for real-time detection of abnormal physical and chemical changes, presence of pollution, contaminants, toxic agents or pathogens in ecosystems, in the air, in freshwater reservoirs, water supply networks, or other physical infrastructure with an unprecedented level of pervasiveness and connectivity, changing how environmental and infrastructure monitoring is conducted, an enabling technology for smart city implementations. Furthermore, this technology also has the potential to lead to groundbreaking security-related applications, especially against chemical and biological terrorism.

In this thesis, we focus on the DF task in abnormality detection using a sensor network via MC. DD is a scheme providing some of the advantages such as better response to rapid changes, reduced communication bandwidth, ability to reconfigure in the case of sensor/link failures etc. [37], [38]. We believe the presented DD strategies lie in the heart of the nanonetworks and can overcome some limitations of MC such as slow signal propagation, limited power resources,and highly dispersive nature of the channel while improving detection performance with DF.

We have started our research with the following questions that we believe to be useful to address.

- Are there any solutions to the unique challenges of MC-based nanoscale sensor network

MC?

The signal propagation characteristics in a diffusive MC channel are highly random in nature and depend heavily on a multitude of factors. We need consider all of these factors in practical applications to build more robust and flexible systems. However, recent research in MC has been limited to the development and study of nanonetworks based on simplifiying assumptions about nanomachines and the propagation environment. For instance, ideal sensors are assumed, and simple scenarios in an unbounded environment are investigated. We should relax these simplifying approximation for the sake of performance of the networks. A critical problem in moving the field of MC forward is the development of robust strategies for creating nanoscale sensor networks that operate in the real world of practical applications. In practical scenarios where one or more of these assumptions don't hold, mathematical descriptions of the channel are evasive. Even if the channel model known, channel conditions may change over time requiring estimation of channel state information (CSI).

- How can we overcome the limitations of DD problems in nanonetworks using MC?

  One of the main characteristics of diffusive MC is the exceedingly slow signal propagation speed in the medium and the channel's high dispersion, which results in long pulse duration and latency resulting in decision delays. Thus, performing a reliable DF task with as few received samples as possible is of paramount interest and also helps to overcome the low bandwidth problem and limited power sources in nanonetworks. We propose to employ SPRT leading to significant savings in the average number of samples required for DF.

- Can ML help to overcome some of the unique challenges specific to diffusive MC based sensor networks ?

  As stated in the preceding section, the signal propagation characteristics in a diffusive MC channel are very stochastic and highly dependent on a variety of parameters. As a result, we believe that ML algorithms can be a strong tool for overcoming these difficulties, as data-driven learning algorithms would help to design of more resilient and flexible systems.

  Improving the detection performance in highly challenging and changing environments is crucial. One of the best methods would be using the prior knowledge and incorporating into algorithms. Also, different input vector dimensions are possible in practical applications. Recurrent neural networks (RNN) can provide good detection performance along

with robustness and flexibility required for practical implementations of such MC based nanoscale sensor networks.

The above issues are critical for the study of the DF in diffusive MC-based DF, and we believe that our novel contributions, as discussed in this dissertation, offer solutions to a variety of these challenges.

1.3 Roadmap

This thesis is structured as follows:

Chapter 2 is divided into five main sections. We summarize sensor networks, namely distributed sensor networks, in Section 2.2. Afterwards, in Section 2.3, we describe diffusive MC using the diffusion mechanism. The Section 2.4 summarizes the challenges in DD in nanoscale networks using the diffusive MC. Finally, the system model is discussed in Section 2.5.

Moving forward, we focus on decision-theoretic methods for DF in Chapter 3 starting with a brief overview of fixed sample Neyman Pearson tests in Section 3.1. After that, we propose to employ a SPRT based method for DF in diffusive MC based DD as well as advantages associated with the proposed test in Section 3.2. Then we provide a basic background of ML to give the communication community an idea of the field.

Then, we present neural networks (NN) method to detection task for the first in the literature to overcome the most challenging difficulties in MC affecting the detection performance in Chapter 4. We investigate different DNN techniques: Feed-forward (acyclic) NNs and Recurrent (cyclic) NNs in Section 4.2. After that, we provide the results of the proposed NN based detectors via Monte Carlo simulations using the ideal signal model and compare with the existing LLR detector in Section 4.3.2.2. We conclude the chapter by showing the performance of the ML-based detector using particle-based MC simulator AcCoRD.

In Chapter 5, we evaluate our ML-based detection algorithm using particle-based simulator AcCoRD in practical scenarios where an analytical channel model is not mathematically tractable or is too complex to be useful. In Section 5.3.1, we simulate a basic cell environment in AcCoRD to detect abnormalities using MC. In the following Section 5.3.2, we address the abnormality detection problem in a blood vessel using several nanoscale sensors and an absorbent receiver, inspired by the numerous healthcare uses of MC inside the IoBNT.

Finally, Chapter 6 integrates our conclusions and future view by proposing different ML approaches, as well as listing unresolved problems that could be an inspiration for future research.

Background

2

There's plenty of room at the bottom.

Richard P. Feynman

2.1 Overview

This study is at the intersection of many different research areas, such as nanotechnology, sensor networks, MC and ML. In this chapter, we provide an overview of the theoretical background required for the rest of the thesis starting with sensor networks, diffusive MC, challenges in DD in nanoscale sensor networks via diffusive MC, and the system model used in this thesis.

Feynman's 1959 lecture is generally regarded as the birth of nanotechnology [39]. It is a broad vision for manipulating matter at the nanoscale, including individual atoms. Nanotechnology has a lot of potential to allow breakthrough applications in a variety of technical fields which permit the miniaturization and fabrication of electronics on a nanometer scale. Nanoscale devices can be linked to execute distributed collaborative activities, which necessitates information flow between them. Communication in nanonetworks may be achieved by MC, which is recognized to be one of the most promising techniques owing to benefits such as the necessitating no infrastructure, considerable energy efficiency, and biocompatibility, among others[33].

Abnormality detection refers to the problem of finding patterns in data that do not accord

with expected behaviour. These nonconforming patterns are often referred to as anomalies, outliers, peculiarities, or contaminants in different applications [34], [40]. Information theoretic techniques using information theoretic measures such as Kolomogorov complexity, entropy, relative entropy, classification-based using algorithms like Support Vector Machines (SVM), Kernel Fisher Discriminants are some examples of conventional abnormality detection techniques [34]. Abnormality detection requires high accuracy because of its critical and crucial nature in most applications, such as fault detection in safety-critical systems, and military surveillance for enemy operations. In most of the existing applications, this accuracy is attained by employing distributed detection using multiple sensors, where the decision is performed using a network of sensors and decision devices, which communicate with each other, i.e. sensor networks [37],[38]. For conventional abnormality detection problems, the inherent redundancy possible with multiple sensors, the availability of high-speed wireless or wire-bound communication, and increased computational capability has led to a considerable research interest in distributed detection [37], [38]. Many different sensor network architectures possible [37], and many different detection strategies both at the individual sensors and/or the fusion centre can be employed [38].

This thesis focuses on the task of abnormality detection, i.e. the detection and reporting of abnormal events that may characterize the presence of a disorder in a fluid environment, employing an MC based nanoscale sensor network [17]. Such DD problems lie in the heart of the most highly anticipated applications of nanoscale networks, such as health monitoring, disease diagnosis, targeted drug delivery, environmental sensing and monitoring, contaminant and toxic agent detection, environmental remediation and many more. Depending on the application, the abnormalities of interest may be quite diverse in nature, e.g. abnormal changes in the concentration of a molecule in the medium, or abnormal changes in the properties of the medium itself, such as the pH value, temperature, viscosity.

Although the problem of abnormality detection has been studied extensively in different fields, only very few previous works exist for the abnormality detection problem in the context of diffusive MC based DD. [35] investigates the modelling and analysis of abnormality detection in bio-molecular nanonetworks for the first time, where the MC channel has been modelled as additive white Gaussian noise (AWGN), and a sub-optimal OR fusion rule is employed for the decision. [41] is similar to [35], the sensing channel is modelled as correlated Gaussian. The work in [18] focuses on a similar type of abnormality as in [41], using a more realistic channel model based on the solution of the diffusion-reaction equations in an unbounded medium, and provides a sub-optimal DF strategy. In [42], after decoding the local decisions using sub-optimal detectors based on an approximation of the log-likelihood ratio, the FC uses OR and AND logic-

based fusion rules to make the final decision.

The remainder of this chapter is organised as follows. In Section 2.2, we provide a brief overview of the sensor networks, namely distributed sensor networks. We explain diffusive MC via diffusion mechanism in the section 2.3. Challenges in DD in nanoscale networks via the diffusive MC are summarized in section 2.4. In section 2.5, we described the system model. Finally, we conclude the chapter in conclusion section.

2.2 Distributed Sensor Networks

Wireless sensor networks are composed of a large number of low-cost devices that are linked together via low-power wireless communications. It is the networking capacity that distinguishes a sensor network from a conventional collection of sensors, allowing sensor elements to cooperate, coordinate, and collaborate [43]. In traditional multi-sensor detection, all local sensors send their data to a central processor that uses standard statistical algorithms to conduct optimal detection. On the other hand, in a decentralized approach, instead of transmitting raw data to a central processor, usually called the fusion centre, they employ their processing capabilities to do simple computations locally and only transfer the required, partially processed data [37].

In this thesis, we specifically consider distributed sensor networks due to their advantages over centralized structures, such as reduced bandwidth, increased reliability and lowered cost. A distributed system architecture may also provide a faster response to sudden changes in the environment.

A variety of topological structures can be used to arrange sensor networks. The three most commonly employed topologies are the parallel (Fig 2.1), the serial (Fig 2.2) or tandem, and the tree topologies (Fig 2.3). The parallel network topology has attracted the greatest attention among the topologies studied in the literature [36], [44]. In the parallel topology, each sensor communicates directly with the fusion centre. We show the configuration of a parallel structure with M sensors in Fig 2.1.

Let the phenomenon $\mathcal{H}$ depicts the physical abnormality to be detected by the sensor network. The $\mathcal{H}_0$ and $\mathcal{H}_1$ represent the absence and the presence of the abnormality of interest, respectively. The task of the FC is to perform the DF, i.e. to decide for $\mathcal{H}_0$ or $\mathcal{H}_1$ by observing the sensor outputs received via the diffusive MC channel.

Each of the M sensors observes one or more sensing variables and produce a quantized soft output between 0 and 1 that represents its sensing decision, i.e. the output of the m'th sensor. Then, mth sensor passes the information to x_m to the fusion centre. The fusion centre makes a

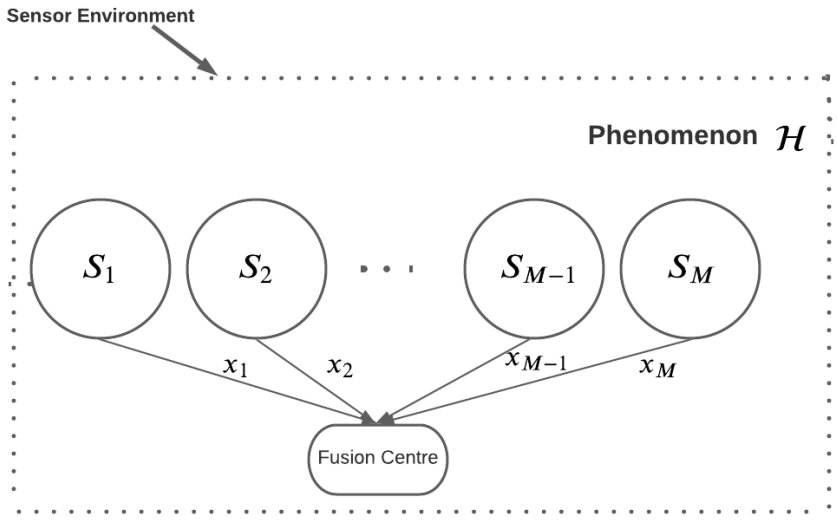


Figure 2.1: Parallel structure with fusion centre.

general decision based on received data that favours either $\mathcal{H}_0$ or $\mathcal{H}_1$.

In the serial (tandem) configuration, the $(m-1)$th sensor transmits its quantized information to the mth sensor, which creates its own quantized information based on its own observation and the quantized data received from the "previous" sensor as shown in Fig 2.2. The initial sensor in the network derives its quantized data only from its observation, and the last sensor in the network chooses which of the two potential hypotheses $\mathcal{H}_0$ or $\mathcal{H}_1$. When the conditional independence assumption,i.e. the observations at the sensors are independent, the joint density of the observations can be represented as the product of the marginal densities, is not viable, the issue becomes unsolvable.

In Tree topology (Fig 2.3), all of the sensors that are installed in the sensor field form a logical tree. A leaf node transmits data packets to its parent nodes. In response, a receiver node that receives data from a child node sends data to the receiver's parent node after combining the data with its own data [45].

For a specific application, the choice of the sensor network topology to be employed is determined by many factors, including the requirements of the application of interest, the physical extent of the sensor network, the medium in which the sensor network is to be deployed, the power limitations and the processing power capabilities of the individual nodes, and the employed mode of communications. In this thesis, we focus on a parallel sensor networks structure. The reasoning behind this choice will become apparent in the following sections, where the unique characteristics of diffusive MC and nanoscale sensor networks will be discussed in detail.

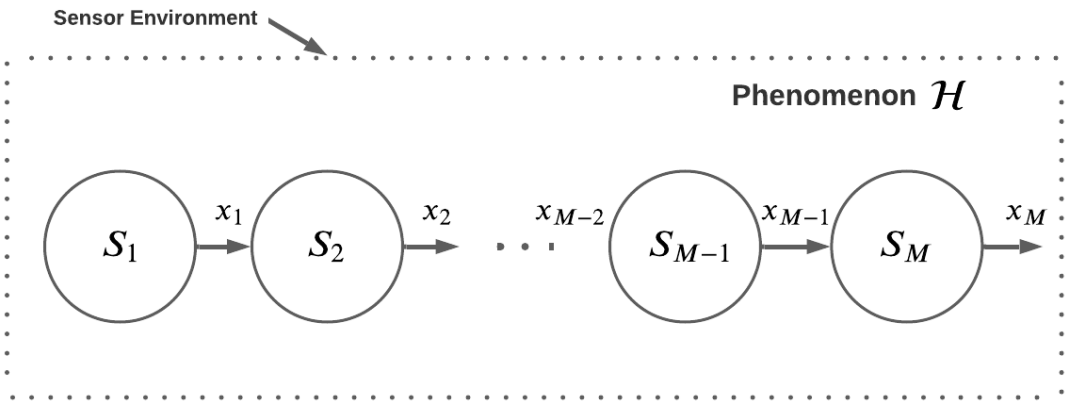


Figure 2.2: Serial structure.

2.2.1 Nanoscale Sensors

Nanoscale sensors are devices with an average diameter in the micrometre range or less and are composed of nanoscale components with individual diameters ranging between 1 and 100 nm [46]. These are small components made up of an ordered collection of molecules that are capable of performing simple processing, sensing, and/or actuation activities [47],[33].

The small size of these systems limits the sensors, actuators, and motion mechanisms, power sources, computer power, and wireless communication capacity. On a nanometer scale, the surface-to-volume ratio is reduced, surface forces become more impactful than volume-based forces. Interatomic forces or surface chemistry play a crucial impact on the mechanism. Thus, inertial forces and weight are practically negligible, but micro/nanoscale surface interatomic forces, fluid dynamics, heat transfer, surface chemistry, adhesion-based contact mechanics, and friction dominate robot mechanics. Micro/nanoscale forces have a wide variety of features when compared to macroscale forces [46]. It is important to note that these differences should also be considered in mathematical descriptions when designing nanonetworks. It would be almost impossible to take into account all of these features and effects.

There are three distinct methods for the development of nanoscale sensors. Top-down approaches produce nanomachines by downscaling current microelectronic and MEMS [48]. The bottom-up design begins with molecular components that self-assemble chemically via molecular recognition principles, organising themselves molecule by molecule [49]. A third approach, named bio-hybrid, has been proposed recently for the construction. This strategy is centred on the utilization of existing biological nano structures, such as molecular motors, bacteria-based

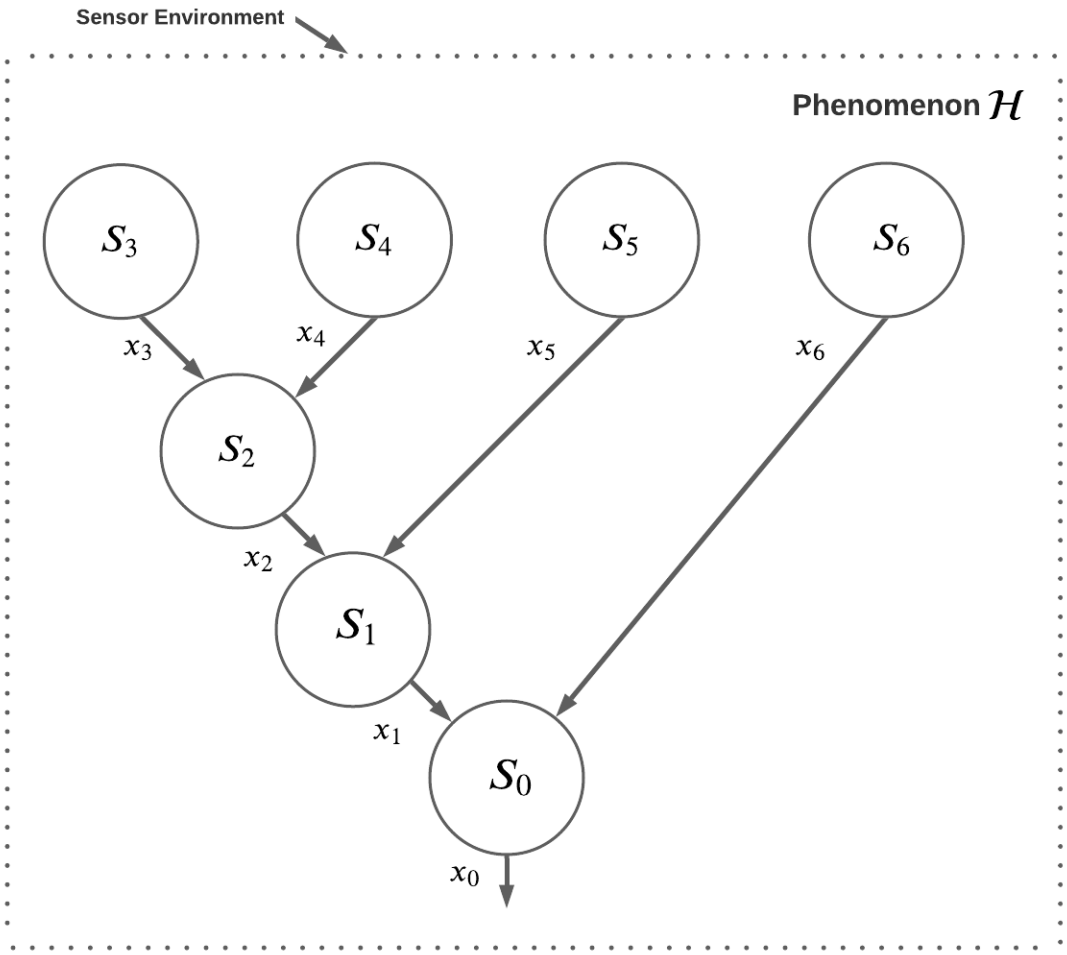


Figure 2.3: Tree structure.

microrobots, as components or models for developing new nanomachines. The characteristics of future nanomachines are currently present in a live cell, which can be regarded as a collection of self-replicating nanomachines [33], [50], [51].

2.2.2 Nanoscale Sensor Networks

Nanonetworks are made up of very small devices, nanomachines described in previous section, that connect and cooperate with one another to enable them to operate on more sophisticated tasks such as drug delivery inside the body or disease therapy [6]. Nanonetworks are attracting interest for their potential applications in a wide variety of fields. These applications can be grouped into four main idea:

- Biomedical Applications: Lab-on-a-chip, drug delivery systems, health monitoring, detection of infectious agents, etc.

- Industrial and Consumer Goods Applications: Food and water quality management systems, manufacturing intelligent materials and fabrics, etc.
- Environmental Applications: Environmental sensing and monitoring, contaminant and agent detection, etc.
- Military Applications: Nanostructured materials, situational awareness, etc.

Next, we will investigate the communication between nanosensors. Several communication paradigms have been proposed based on EM, optical, and acoustic communication [6]. Because of the characteristics and restrictions of nano dimension in environments such as in aqueous media, these paradigms may not be suitable for nanonetwork applications. Also EM waves show adverse propagation characteristics causing implementation constraints such being the ratio of the antenna's size to the wavelength of signal [52], [6].

2.3 Molecular Communication

Historically, the term 'communication' is synonymous with the term 'common'. It is derived from the Latin verb communicare, which meaning 'to share' or 'to make common,' and is linked to the Latin word for common:communis [53]. Communication systems are ubiquitous and communication is critical between biological systems and occurs at all levels of the systems, from subcellular proteins through organelles, tissues, and organs, and finally to groups of individuals [54]. At the molecular level, individual cells require communication to share information [55].

MC is a biologically inspired approach to the challenges of traditional communication for nanoscale networks [1]. In this thesis, we consider diffusion-based MC as stated in the previous section. Diffusion-based MC encodes information into some characteristic of the released molecules, such as the release time, the number, or the kind of molecules, by using specialized molecules as information carriers.

"Wireless transmission as well as nanoscale transport have existed for at least 3.5 billion years; ever since magnetic storms and lightning have lit the sky, ever since van der Waals forces have existed, and ever since the first living cells transported nucleotide units such as RNA and DNA. The point is that humans do not really "create" anything new; we leverage and fine-tune what nature provides." says Stephan J. Bush [46].

Communications engineers have worked with electromagnetic modes of communication in the past. However, as nanotechnology is becoming a promising research area for solving many

challenging problems of humanity, our understanding of communication has to change to include the restrictions of the nanoscale size, limited energy reserves, and fluid propagation environment such as those found within the human body. Finding improved ways to communicate among smaller nanoscale sensors will improve sensor coverage. Furthermore, as nanotechnology advances, the necessity for low-cost, robust, and reliable communication among nanomachines will become evident.

MC presents a promising approach by mimicking the naturally evolved communication mechanisms between biological entities at this physical scale. This approach enables nanoscale entities to communicate using molecules as information carriers.

When we imagine an abnormality detection system consisting of tiny, blood cell-sized sensors that continuously measure the environment parameters and transmits their signal results to sense abnormality which might indicate a presence of disease in the human body, we need to physically enable the communication between the distributed nanoscale sensors and the decision device under the constraints of nanoscale networks. These devices must operate in the body without disrupting healthy tissues, or being destroyed by the immune system. Microorganisms and their naturally evolved communication mechanisms have been inspiring and MC emulates these to enable information exchange between the nodes of a nano or microscale network.

The past decade has seen a rapid increase in information-theoretic analysis of MC [9]. In biological systems, various molecular signalling pathways arise (e.g., the transmission of diffusive molecules, protein-nanomachines, transport of materials by propagating over protein filaments) [26]. The information propagation process is usually modelled as discrete Brownian motion as the number of molecules becomes too large in continuous form for diffusive MC, and we focus on diffusive MC in this thesis.

2.3.1 Molecular Communication via Diffusion

Diffusive MC employs dedicated molecules as information carriers while relying on the diffusion of these molecules for signal propagation, requiring no infrastructure or additional external energy [33]. Diffusion is the transfer of particles from areas of higher concentration to areas of lower concentration by random motion [56]. A generalized model of a diffusive molecular communications system is depicted in Fig 2.4.

The general processes of communication are as follows:

- Information is encoded into some aspect of the released molecules by the transmitter and

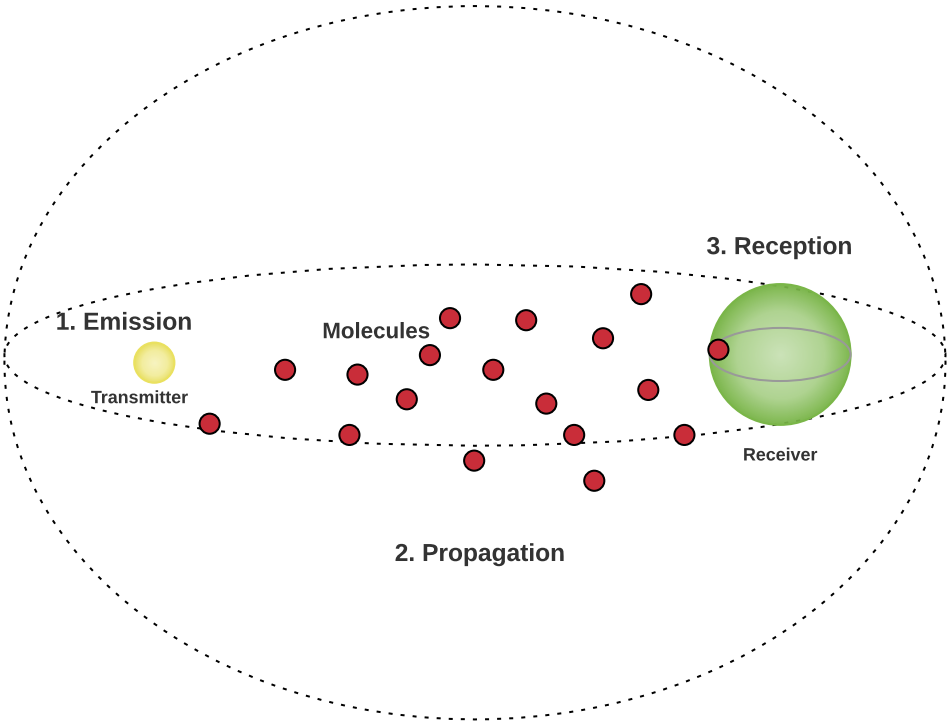


Figure 2.4: A model of molecular communication.

the information molecule is sent into the environment,

- The information molecules propagates through the environment,
- Receiver captures information molecules and decodes .

In the following sections, we first explain the mathematical modelling of diffusion. Later, the transmitter, the receiver and the channel of the model is briefly depicted.

2.3.1.1 Mathematical Modelling of Diffusion

It is necessary to have a thorough grasp of the propagation environment to do meaningful communications analysis. Characterizing the channel between a transmitter and its receiver is crucial to understand MC systems. Naturally-evolved MC pathways can be imitated synthetically, such as walk-based mechanisms, flow-based mechanisms, and diffusion-based mechanisms [57]. The one that has garnered attention from the scientific community for nano-communication networks is diffusion-based communication, which refers to the random movement of information-carrying molecules. No infrastructure or external energy is required for the diffusion of information-carrying molecules to their intended destination. This makes MC via diffusion particularly well-

suited for biocompatible short-range ad-hoc communication between nano or micro-scale machines that are limited in terms of energy. Conventional radio-wave communication methods and theoretical results are only partially applicable and must be reassessed. Molecular diffusion may be investigated at both the microscopic and macroscopic levels. The macroscopic focus is on the overall behaviour of a large number of molecules. The random movement of individual molecules, known as Brownian motion, is the subject of microscopic study. In classical diffusion, the time-varying molecule concentration is determined by solving Fick's second law with boundary and initial conditions determined by the geometry of the bounds of the environment, the receiver and transmitter models, their physical extents and geometries, and their relative positions. The most general form of Fick's second law is:

$$\nabla^2 C(\boldsymbol{r}, t) = \frac{1}{D}\frac{\partial}{\partial t} C(\boldsymbol{r}, t) \tag{2.1}$$

Where C is the concentration, r represents the spatial coordinates ($\boldsymbol{r} = [x, y, z]$ in cartesian coordinates), D is the diffusion coefficient and ∇^2 is the Laplace operator, which, in cartesian coordinates are given as $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. Environmental effects such as flow and chemical reactions in the medium require the addition of further terms to the partial differential equation above, which is then referred to as the diffusion-convection-reaction equation.

Brownian motion, i.e. the random motion caused by collisions with the fluid's molecules, leads to the dispersion of free molecules within in a fluid medium. The Brownian motion is a random walk process, whose variance depends on the diffusion constant. Einstein demonstrated in his renowned work in 1905 that the probability density function of a single particle's position under Brownian motion (microscopic) satisfies the differential equation provided by Fick's second law of diffusion (macroscopic) [56]. Einstein started with assuming the existence of a time interval δt which, could be considered as infinitesimally small on a macroscopic scale. On the other hand it was large enough that a solute molecule typically experiences many collisions with the solvent molecules in that time. He later introduced a probability density function $\phi(\xi; \delta \mathrm{t})\mathrm{d}\xi$ gives the probability that the x coordinate of a solute molecule will change during the next δt by the amount between δ and $\delta + d\delta$. Stochastic change in the position of a solute molecule is considered to represent the combined effects of the many random collisions with many smaller solvent molecules in time δt . For simplicity let us show the classical one-dimensional diffusion equation Einstein had

$$\frac{d\rho(x,t)}{dt} = D\frac{d^2\rho(x,t)}{dx^2}, \tag{2.2}$$

where D is defined by

$$D = \frac{1}{2\delta t}\int_{-\infty}^{\infty} \xi^2 \phi(\xi; \delta t) d\xi. \tag{2.3}$$

He reasoned the expected effects of the solute molecules collisions from the surrounding solvent molecules and provided a connection between the macroscopic and microscopic views. It should be noted that determining the behaviour of a particular diffusive channel requires the solution of a partial differential equation for a given geometry. This partial differential equation may contain only diffusion, or, in the presence of additional flow, an additional convection term. An additional reaction term is included in presence of chemical reactions. Unfortunately, an analytical solution to such differential equations is only available for the most simplest of the cases [56], [17], [27]. For other more complicated settings, the use of numerical methods or approximations may be required for the characterization of the channel [29], [30].

2.3.1.2 Transmitter

We model the transmitter as a point source of molecules, located at the origin. This point source can control the concentration, type and release time of molecules. For example, we encode the information in concentration by releasing a certain number of molecules for bit 1 and bit 0 depending on the data modulation model. In practice, the transmitter cannot be considered as perfect. To model this imperfection, we should use a probability distribution model.

2.3.1.3 Channel

The physical system of molecular transport between the transmitter and the receiver has to be specified for a statistical model for the molecular channel. This motion of molecules towards the receiver is categorized in three forms; walk based, flow-based or diffusion-based [9]. In walk-based mechanisms, information molecules are encased in cargo, which is subsequently pushed toward the destination by a motor protein, like as dynein or kinesin, along a pre-defined path, similar to microtubule tracks. On the other hand, in flow-based mechanisms, the propagation of molecules is impacted by an external flow. Because flow is a one-way phenomena, it cannot be used for two-way communication. Diffusion-based transport that is considered in this thesis involves molecules randomly propagating in all accessible directions via Brownian motion. Even tough this results in a larger level of uncertainty at the receiver,diffusion-based mechanism is fully passive and constantly accessible, requiring no additional energy or infrastructure, and is best suited to highly dynamic and unpredictable settings. In the presence of a drift, diffusion-based transport can also be considered, resulting in a hybrid of diffusion and flow-based processes. The diffusion-based transport mechanism is the focus of the majority of the literature [57].

2.3.1.4 Receiver

The receiver, namely the FC in this thesis, captures information molecules circulating in the environment. There are some options for a receiver nanomachine to obtain information molecules, such as using a surface permeable to the molecules, or surface channel can be employed to pass information molecules into a receiver. Another technique is exploiting surface receptors capable of interacting with a certain type of information molecule and causing reactions at the surface, which induces reactions within the receiver bio-nanomachine [1].

2.3.1.5 Transmitter and Receiver Models Used in MC Literature

There are different models of receiver in the literature, i.e. sampling receiver [36], transparent receiver [20], [58] absorbing receiver [25], [59], [60], [61] and ligand or reactive receiver [28]. For a comparison of different receiver models and their transformations into one another, see [21]. Molecule absorption is widespread in nature when diffusion is the primary mode of communication. As widely used in the MC literature, we consider an absorbing receiver in this thesis. The absorbing receiver absorbs any molecules that hit its surface, removing them from the environment. These receivers use the zero boundary condition when solving Ficks's second law with a reaction term. In [25], the initial and boundary conditions is defined.

The assumption that a transmitter is represented as a point source of molecules has been frequently used in the MC literature [1], [25], [17], [42]. Most of the first and foremost papers in the MC literature considers point transmitters. Therefore, we consider the point transmitter model in this dissertation. There are other transmitter models in the MC literature, such as a sphere transmitter is considered in [62], where the molecules are uniformly distributed within the sphere before release, and all the molecules are released at the same time. The sphere disappears and the molecules are able to diffuse freely after release. Another transmitter type is given in [30], [63]. Molecules are released from a point on a reflective sphere in this model, and the sphere serves as a barrier to the molecules' diffusion after emitting (i.e. if the released molecules hit the surface while diffusing, molecules bounce off). As a result, this transmitter model is directional (as opposed to the ideal point transmitter model, which is non-directional) and may be used to represent a physical transmitter with a real physical extent that releases molecules from a small hole on its surface. It may be used to model a molecule-releasing micro- or nanomachine, as well as a bioengineered cell that employs exocytosis to release molecules. An ion-channel based spherical transmitter is presented in [64]. In this paper, modulator model for diffusive MC is proposed that takes advantage of the regulatory mechanisms seen in natural cells ion channels.

The modulator controls the rate of molecule release from the transmitter acting like ion channels in the cell membrane.

2.3.1.6 Characteristics of Molecular Communication

In this section, the general characteristics of MC in the basic model are summarized. These characteristics are needed to be considered in designing a detector in a diffusive MC based sensor network. It is also important to benefit from the advantages of MC features in our systems. In terms of devices, signal type, signal propagation speed, range, and medium, a comparison of certain features of MC with telecommunication is given in Table 2.1. After providing a summary of these characteristics, the challenges of diffusive MC are provided in the following section.

2.3.1.7 Communication via Chemical Signals

The physical qualities or characteristics of information molecules, such as the concentration of the molecules, type of information molecules employed, their three-dimensional structure (e.g., protein), sequence information (e.g., DNA) can all be used to encode information in MC. A molecular structure may store a high density of information [65]. Furthermore, functional information can be encoded. A DNA sequence, for example, can be utilized to encode a functioning protein. As a result of gene expression, a receiving bio-nanomachine may gain additional functionality (e.g., tolerance to harmful chemicals) [2].

2.3.1.8 Biocompatibility

The communication between the nodes in an MC network does not use methods that are foreign or disturbing to biological systems, such as traditional communication systems may disturb their functioning, causing pollution or toxicity. Also, by utilizing materials and processes from biological systems, MC may be an alternative in terms of biocompatibility especially for medical

Table 2.1: Telecommunication and Molecular Communication, adopted from [1], [2]

Communication	Telecommunication	Molecular Communication
Devices	Electronic devices	Bio-nanomachines
Signal types	Optical/electrical	Chemical
Propagation speed	Speed of light	Extremely slow
Propagation range	$\mathrm{m} - \mathrm{km}$	$\mathrm{nm} - \mu\mathrm{m}$
Media	Air/cables	Aqueous (in this thesis)
Energy consumption	High	Low

and healthcare applications. Bio-nanomachines may then utilize encoding and decoding mechanisms similar to those used by biological cells in the human body to communicate with them straightforwardly [1].

2.3.1.9 Energy Efficiency, Low Heat Dissipation

MC employs methods and materials from biological systems, and thus it is predicted to be energy efficient and has a low heat dissipation. The environment in which bio-nanomachines are placed is intended to provide the chemical energy required for MC, and as such require no external energy supply [1].

2.3.1.10 Scales of Nanodevices

A nanodevice used in nanonetworks is described as a mechanical or electromechanical device in nanometer-scale with nanoscale functional components as explained in detail in section 2.2.1. These machines may be synthetically designed such as modified cells, artificial cells, electronic devices developed in the research fields of MEMS and NEMS [66], or may be nature-made biological forms such as proteins, cells, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) referred to as bionanomachines [67],[68]. These are small-scale devices made out of biological molecules that react chemically such as carbohydrates, lipids, proteins, and nucleic acids, which are abundant in living organisms. Input signals may cause a bio-nanomachine to transmit output signals, change its internal state, or adjust its functioning. For example, A DNA molecule, which stores genetic information, changes its state in response to molecular signals in the cell by turning on and off specific genes [1]. Energy reserves may be a major problem at nanoscale [69]. Nanodevices may face some practical limitations as a result of their confined processing capacities and power management limitations. These devices are supposed to perform basic tasks that necessitate novel and innovative approaches [70]. One of the areas in which MC research has great potential is the use of materials and processes from biological systems to achieve energy efficiency. Yet, in this thesis, some techniques are presented to perform even better energy efficiency. In the following chapters, these approaches are given in detail.

2.3.1.11 Uncertainties in the Channel Model

Due to the highly random nature of diffusion based propagation of information carrying molecules, the received signal in a diffusive MC system is highly random as well. The information molecules' unpredictable motion distorts the received signal, making the detection a very challenging task.

The detection methodologies existing in the literature usually rely on the presence of accurate a-priori information on system parameters at the detector, such as sensor-FC distances, diffusion coefficient etc. Also, molecules may degrade over time in the environment and fail to reach a bio-nanomachine receiver. Additionally, bio-nanomachines' reaction with information molecules is probabilistic. This highly probabilistic behaviour is one of the main points needed to be considered in designing a MC system. Also, the channel conditions may change over time which may affect the performance of the system [9].

2.3.1.12 Slow Communication Speed

The speed of MC is extremely slow and the range is limited. They differ based on the biological materials, processes, and environment. The slowest mode of MC is through free diffusion. The time t required to propagate a molecule over a distance L is denoted by the formula $t \approx \frac{L^2}{D}$ (i.e. the time required increases with the square of the distance), where D is the molecule's diffusion coefficient [71]. In the following section, more detailed explanation is provided with plots regarding the distance and diffusion coefficient. Also, the molecule's size and shape, temperature and viscosity of the medium may impact the diffusion coefficient. This characteristic of MC leading to long pulse intervals and long latency. A faster and better DF method would increase the speed of the communication network while retaining the detection performance.

2.4 Challenges in Distributed Detection via Diffusive MC

From a communications and signal processing perspective, two of the most challenging issues are investigated in this thesis. First of the issues is the slow speed of propagation and the dispersive nature of the channel that leads to high decision delay. Let us consider a simple configuration with a point transmitter and a spherical absorbing receiver with the radius $3\mu\mathrm{m}$. A single pulse of molecules is transmitted by a point transmitter and diffuses in the environment with a typical diffusion coefficient ($500\mu\mathrm{m}^2/\mathrm{s}$). The transmitter releases the information carrying molecules to FC starting at the time instant $t = 0$ and only release one pulse in this example. It has been shown in [25] that for a molecule released by a point transmitter located at a distance of r_1 from the centre of a perfectly absorbing spherical receiver of radius r_2, the probability of hitting the surface of the receiver within a time interval $[kT, (k+1)T]$ seconds after release is given as:

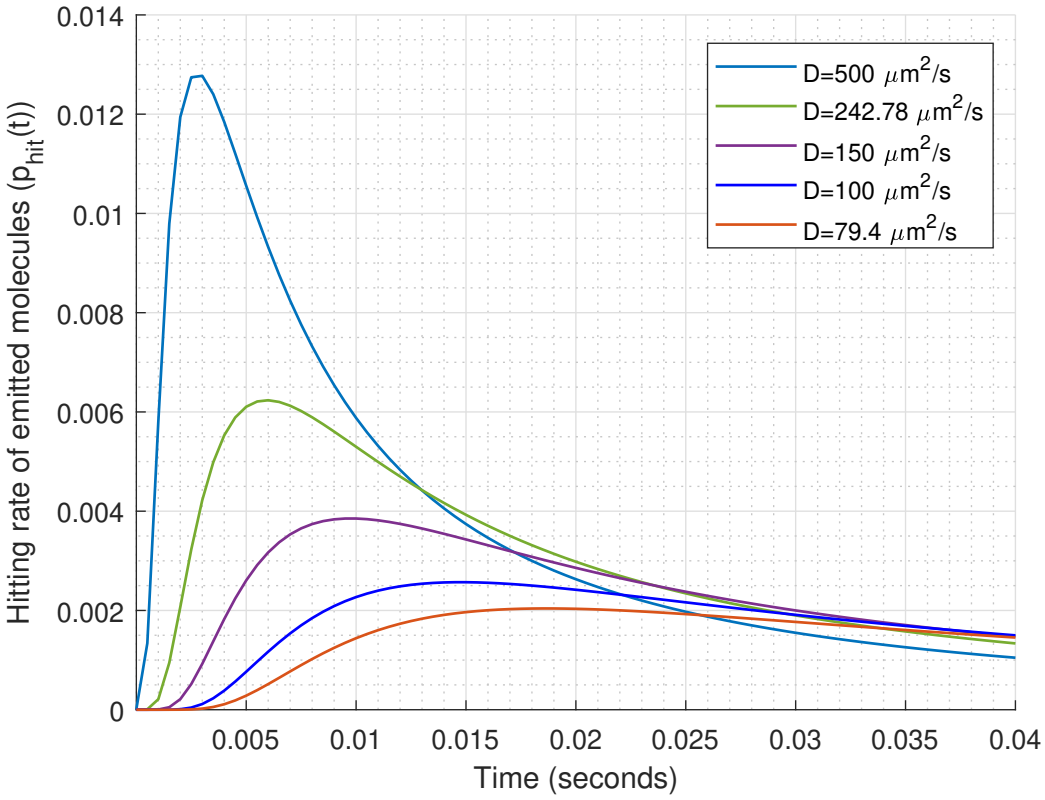


Figure 2.5: Hitting rate considering different diffusion coefficient, where distance between transmitter and receiver is 3μm.

$$p_k = \begin{cases} \frac{r_2}{r_1}\,\mathrm{erfc}\left(\frac{r_1-r_2}{\sqrt{4DT}}\right), & \text{for } k=0 \\ \frac{r_2}{r_1}\left(\mathrm{erfc}\left(\frac{r_1-r_2}{\sqrt{4(k+1)DT}}\right)-\mathrm{erfc}\left(\frac{r_1-r_2}{\sqrt{4kDT}}\right)\right), & \text{for } k\geq 1, \end{cases} \tag{2.4}$$

where T is the duration of a time slot, D is the diffusion coefficient of the transmitted molecule in the medium and erfc(·) is the complementary error function. By using the hitting probability function, the probability of a single pulse of molecules hitting the receiver is calculated. The effect of the diffusion coefficient considering five different values of D illustrated in Fig 2.5. The effect of the the distance between the transmitter and receiver (transmission distance) on the number of molecules absorbed by the receiver is shown are in Fig 2.6. We consider five different values to observe the effect of the transmission distance $T - dist$. These results illustrate the slow signal propagation speed and the highly dispersive nature of the diffusive molecular communications channel. The time required for the peak of the pulse to arrive gives an interpretation of how long the signal will take to travel from the transmitter to the receiver. Inter-symbol interference (ISI) occurs when a new pulse is emitted before the previously transmitted molecules have vanished

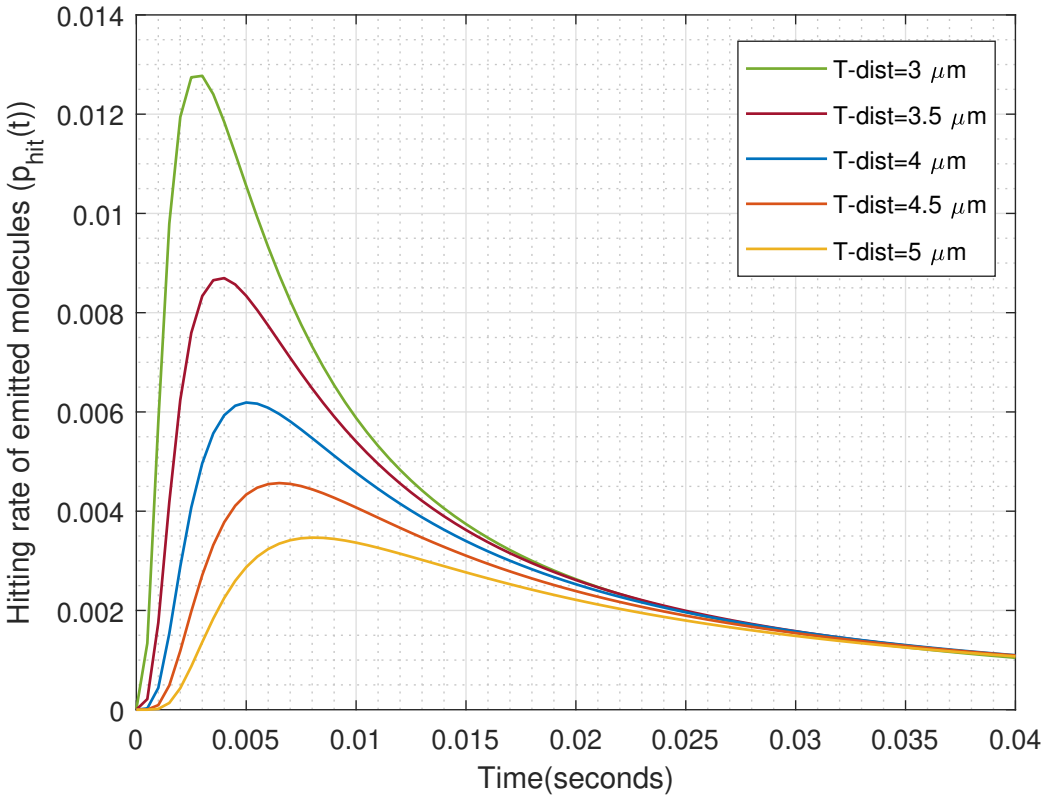


Figure 2.6: Hitting rate considering different transmission distances, where diffusion coefficient is $500\mu\mathrm{m}^2/\mathrm{s}$.

from the receiving area. As shown in figures 2.5, 2.6, the signal is an infinite-duration signal that does not vanish entirely from the receiving area in a short period of time [23]. The influence of the ISI can be minimized if the transmitter waits long enough to deliver a subsequent pulse but the information transmission rate then becomes very low.

The slow speed of signal propagation and the dispersive nature of the channel leads to the following consequences for the task of abnormality detection via a diffusive molecular communication based sensor network:

1. The sensor network layout should be chosen considering these factors. A parallel structure with a fusion centre is preferred in our scenario to overcome the issue of the slow speed of signal propagation. Transmission distance may also be crucial depending on the application in which the algorithm would be implemented. There may be cases in which a serial sensor network may be most suitable. For instance, in a case in which sensors are spread out over a wide area and multi-hop communication is preferred to achieve efficient and reliable communication performance, each sensor node also performs as a repeater, receiving a signal from another sensor node, fusing it with its own sensing data, and sending

the resulting signal to the next node. However, the serial structure would require sensor nodes capable of sensing, receiving, decoding, processing, and transmitting, resulting in a more sophisticated sensor structure. In the parallel structure described in this thesis, however, the sensors are just required to detect the abnormality and transmit a signal. The parallel structure is more suited for circumstances where the nodes are close to one another, i.e. communication is short-range, and the sensors are placed uniformly or regularly throughout the medium.

2. Due to the very slow speed of propagation, decision delay is a considerably more important concern in comparison to typical wireless sensor networks where the signal propagates at the speed of light. When compared to a wireless communications channel, high dispersivity indicates a limited bandwidth, i.e. poor channel capacity (i.e. low symbol rates), which means you can only use a few signals for detection.

We propose sequential decision fusion approaches in Chapter 3 because, when compared to an existing fixed-sample-size Neyman-Pearson benchmark test, this approach provides significant reductions in the number of samples required for decision.

The other major constraint is that the channel characteristics in practical scenarios where these sensor networks are to be implemented may contain a lot of uncertainties (parameters unknown to the fusion centre, which may change over time and are difficult to predict). In most cases, it cannot be analytically determined, since the partial differential equation characterizing the channel does have an analytical solution only in some geometrically very simple scenarios. Even the diffusion process itself may be an anomalous type of diffusion that cannot be explained by simple Brownian motion in some instances. Although the channel model exists, the channel may be dynamically changing, the network geometry may be changing (because everything is floating around in a fluid environment), there may be flow, and other activities going on. As a result, characterizing the channel is challenging, and techniques that are robust to changes in channel conditions and/or do not require an analytical model of the channel are needed. This study propose to use an ML method to this task for the first time in the literature, and shows that this technique offers the resilience and flexibility necessary for practical implementation to overcome these issues.

2.5 The System Model

A sensor network with M identical nanosensors transmitting their soft decisions via diffusive MC to a spherical FC in an unbounded $3D$ environment is considered in this dissertation as

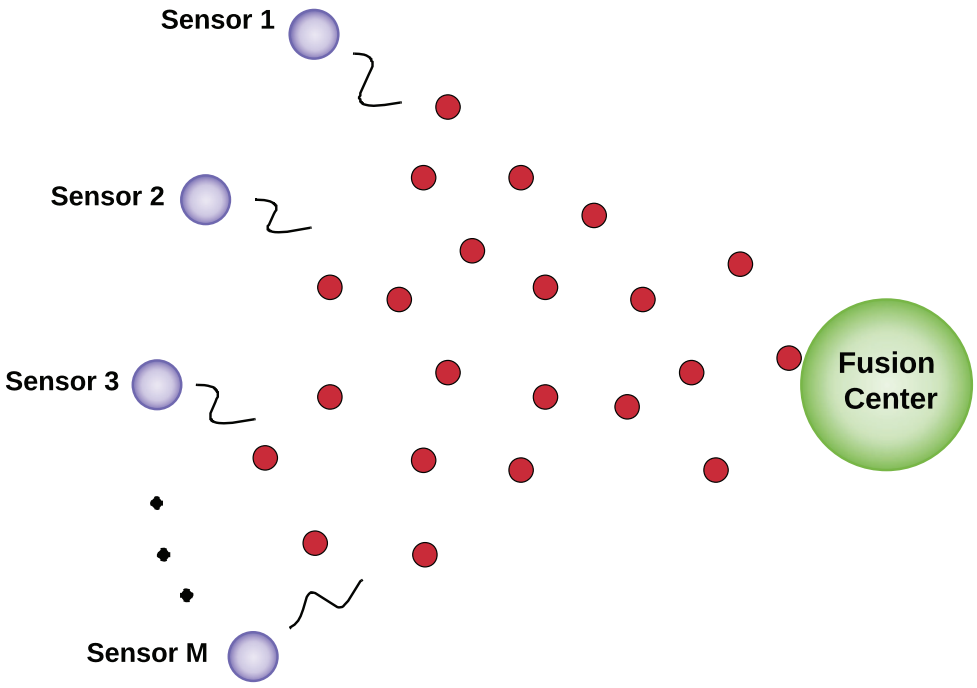


Figure 2.7: Nanonetwork consisting of M nanosensors and a fusion centre.

depicted in Fig 2.7. Let the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ represent the absence and the presence of the abnormality of interest, respectively. The task of the FC is to perform the DF, i.e. to decide for $\mathcal{H}_0$ or $\mathcal{H}_1$ by observing the sensor outputs received via the diffusive MC channel. To provide a fair comparison with the fixed sample size strategy in [17] chosen as a benchmark, we employ the same abstract sensing model and the same communication model between the sensors and the FC, which will be described in the rest of this section.

2.5.1 The Sensing Model

Due to the broad spectrum of potential applications envisioned for nanoscale sensor networks, a wide range of abnormalities representing diverse physical or biochemical phenomena may become of practical interest, which requires different sensing mechanisms. While some of the existing works have been focusing on a specific type of abnormality and sensing model (e.g. [41], [18]), others, such as [17], and [42], have employed abstract sensing models to achieve more general results. In this thesis, we chose to focus on the latter approach for the sake of generality. Each of the M nanosensors is assumed to measure one or more sensing variables, i.e. inputs, and generates a quantized soft output between 0 and 1 representing the sensing decision, $X_m \in \{0, 1/(L-1), 2/(L-1), ...1\}$ where L is the number of quantization levels

and $m \in \{1, 2, \ldots, M\}$. Clearly, this model can accommodate both hard decisions (for $L = 2$) and soft decisions (for $L > 2$) at the sensors. The uncertainties in the sensor outputs due to measurement imperfections associated with the sensing mechanism, e.g. due to sensor noise are accounted for by modelling the sensor outputs X_m as random variables (RV) with a conditional probability mass function (pmf) $q_i(x_m)$:.

$$q_i\left(x_m\right) = P\left(X_m = x_m \mid \mathcal{H}_i\right), \text{ for } i = 0, 1. \tag{2.5}$$

Here, we use distributions of the form $b_i \exp(c_i x)$ which satisfy the imperfections associated with the sensing mechanism with appropriately chosen values for the constants b_i and c_i , i.e.

$$q_0\left(x_m\right) = \frac{\exp\left(-c_0 x_m\right)}{\sum_{x \in \mathcal{X}} \exp\left(-c_0 x\right)} \tag{2.6}$$

$$q_1\left(x_m\right) = \frac{\exp\left(c_1 x_m\right)}{\sum_{x \in \mathcal{X}} \exp\left(c_1 x\right)} \tag{2.7}$$

where b_i is an normalization constant assures that $\sum_{x_m \in \mathcal{X}} \mathrm{q}_i\left(x_m\right) = 1$. Larger values for c_1 would result in implementing more reliable sensors under hypothesis $\mathcal{H}_0$, larger values for c_0 would result in more reliable the sensor decisions under $\mathcal{H}_1$.

2.5.2 The Reporting Model

Perfect point transmitters at the sensors, a perfectly absorbing spherical receiver model at the FC, and an unbounded medium for diffusion is considered as in [17]. The hitting probability function given in [25] and presented in the section 2.4 is used to calculate the fraction of molecules absorbed by the receiver as a function of time. Sensors are assumed to be point sources release information molecule impulses. Each of the M sensors transmits its output to the FC starting at the time instant $t = 0$ by releasing N consecutive pulses of $X_m A$ information-carrying molecules, each T seconds apart, where A is the maximum number of molecules that a sensor can release for each pulse. Hence, the sensor output X_m modulates the amplitude of the transmit pulse train of the corresponding sensor. In this thesis, one case is considered, where a single molecule type is employed for communication by all sensors, allowing the use of a simpler receiver in practice. The sensors are assumed to be equidistant to the FC with statistically independent sensing measurements. The received signal at the FC, Y_n, is defined as the random sequence representing the number of molecules absorbed by the FC within the time slot $[(n-1)T, nT]$. In such a case, for a given realization of the sensor outputs $X_m = x_m$ $m = 1, \ldots, M$, Y_n can be modelled as an

independent Poisson distributed random sequence with a time-varying mean:

$$Y_{n|X_m=x_m} \sim \text{Pois}\left(J + \sum_{m=1}^{M}\sum_{k=0}^{n} p_k x_m A\right), \tag{2.8}$$

where J is the expected value of the received Poisson distributed additive noise molecules[17]. Note that the independence of the sequence Y_n is easily verified, assuming regular Brownian diffusion, a large number of molecules, and a perfectly absorbing receiver that irreversibly removes all the molecules crossing across its surface. The task of the FC is to perform the DF, i.e. to decide for the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ by observing y_n, a realisation of the sequence Y_n, which will be investigated in the following sections with different methods:

- Sequential Decision Fusion
- Neural Network Based Decision Fusion.

The first proposed method, which we will be referring to as the sequential average probability ratio test (SAPRT), is based on the SPRT proposed by Wald in [72]. Because of its generality, flexibility, and practical relevance, the fixed sample size test was chosen as a benchmark for performance comparison. The suggested sequential technique reduces the average number of samples required for DF and, as a result, the decision delay, while maintaining the same average detection performance without depending on an independent and identically distributed (i.i.d.) assumption, which may result in extra decision delays in practice.

ML is one of the powerful approaches for making decisions under uncertainty, and it can be used in a variety of ways, such as predicting the future based on previous data, finding the best model to explain existing data, and so on. We use DNN type of ML in our research. DNN comprises multiple hidden layers to model complex non-linear relationships.

2.5.3 Conclusion

In this chapter, we introduce a brief overview of different research areas starting with the sensor networks overview and possible structures of sensor networks in section 2.2. Theoretical background regarding DD using sensor networks via MC is provided in 2.3. Later, we present a general overview of the features and challenges of MC, and we explain the major challenges of the DD task the diffusive MC methods in section 2.4. First, we show the slow speed of propagation and the dispersive character of the channel, which leads to significant decision delays.

Second, we point out that in practice, channel characteristics will be uncertain (unknown parameters) and may change over time. It cannot be calculated analytically in most situations, since the partial differential equation that characterizes the channel only has an analytical solution in a few basic scenarios. In the last section 2.5, the system model considered throughout the dissertation is defined providing the sensing model and the reporting model.

Chapter 3: Decision Theoretic Methods for Decision Fusion

My brain is only a receiver, in the Universe there is a core from which we obtain knowledge, strength and inspiration.

Nikola Tesla

The DD strategies existing in the literature, aforementioned in Chapter 2, approach the task of DF by employing fixed sample size tests within the conventional Neyman-Pearson framework, with the detection probability for a specific false alarm rate for a given fixed number of channel observations as the main performance criterion. However, one of the main characteristics of diffusive MC is the extremely slow signal propagation speed in the medium [33] and the highly dispersive nature of the channel, leading to long pulse intervals and long latency. Thus, DF schemes that require a large number of channel observations (i.e samples) at the FC, and/or rely on assumptions that may result in additional latency in practice, may lead to excessive decision delays. Consequently, performing a reliable DF with as few receive samples as possible is of paramount interest. This makes the use of sequential tests, which allow the use of variable observation window lengths in order to minimize the average number of observations required for decision, while retaining a prescribed detection performance, promising and efficient alternative to fixed sample size based approaches for DF investigated in the literature. In this thesis, for the first time in the literature, to employ a SPRT based method is proposed for DF in diffusive

MC based DD. Note that an SPRT based approach is employed in a point-to point MC system for increasing the robustness of data demodulation in [73], i.e. in a distinctly different context with a different purpose. The proposed approach, which we refer to as the SAPRT, is based on the SPRT proposed by Wald in [72]. The fixed sample size test in [17] is chosen as a benchmark for performance comparison due to its generality, flexibility and practical relevance. The proposed sequential approach leads to significant savings in the average number of samples required for DF, and, consequently, a considerable reduction in the decision delay, while achieving the same average detection performance without relying on an independent and i.i.d. assumption, that may lead to additional decision delays in practice. A brief overview of fixed sample Neyman Pearson tests, proposed SAPRT and performance evaluation is given in the following sections, respectively.

3.1 Decision Fusion Based on Fixed Sample Size Tests

The statistical basis for the design of detectors of signals in the presence of noise follows from the theory of the theory of statistical hypothesis testing. The classical approach based on the Neyman-Pearson theorem is one of the primary approaches to simple binary hypothesis testing.

Let us consider a statistical model $X \sim f(x \mid \theta), \theta \in \Omega$ where θ represents the parameter value of the statistical model of choice and Ω is the set of all possible values of the parameter θ. The likelihood function $\mathrm{lik}(\boldsymbol{x}; \theta)$(also known as the likelihood) defines the joint probability density of statistically independent observed data $X = x$ as a function of the parameters:

$$\mathrm{lik}(\boldsymbol{x}; \theta) = f\left(x_1 \mid \theta\right) \times \ldots \times f\left(x_n \mid \theta\right). \tag{3.1}$$

The maximum likelihood (MLL) estimates the parameter $\theta \in \Omega$ that maximizes the $\mathrm{lik}(\theta)$ function [74]. In applying the Neyman-Pearson approach to detection task using the idea of MLL, the goal is to choose among given hypotheses based on an observed data meaning a mapping from each possible data set into a decision.

Two hypotheses are tested $\mathcal{H}_0 : \theta = \theta_0, \mathcal{H}_1 : \theta = \theta_1$ for the binary hypothesis test considered in this thesis. In case of our detection problem defined in previous chapter, hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$ represent the presence and the absence of the abnormality of interest, respectively. When the data is observed, we can compare two hypotheses by considering the ratio $\mathrm{lik}\left(\boldsymbol{x}; \theta_0\right) / \mathrm{lik}\left(\boldsymbol{x}; \theta_1\right)$ to understand the probability to be observed under two hypotheses. A predefined threshold can be used to compare the probability of observed data belonging to the hypotheses. As a result, two types of errors are presented in the task. $P\left(\mathcal{H}_0; \mathcal{H}_1\right)$ represents the probability of deciding

$\mathcal{H}_0$ when $\mathcal{H}_1$ is true and $P(\mathcal{H}_1;\mathcal{H}_0)$ is vice versa. Trying to minimize the chosen type of error while fixing the other type of error is a typical approach as it is not possible to reduce both of them. The Neyman-Pearson criterion defines the optimal detector as the solution to the following optimization problem: $\max\{P_D\}$, such that $P_F \leq \alpha$, where P_D is probability of detection, P_F is probability of false alarm, and α is a predefined value considering the decision rule Γ_τ. In case of our detection problem, the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true defined as P_F and the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_1$ is true defined as P_D. The detector that maximizes the P_D for a given probability of P_F considering two hypotheses is given [75]

$$L(\boldsymbol{x}) = \frac{\text{lik}(\boldsymbol{x};\text{H}_1)}{\text{lik}(\boldsymbol{x};H_0)} \tag{3.2}$$

where the γ is defined

$$P_F = \int_{\{\mathbf{x}:L(\mathbf{x})>\gamma\}} \text{lik}(\boldsymbol{x};H_0)\, d\mathbf{x} = \alpha. \tag{3.3}$$

As stated in the section 2.5.2, existing DD strategies in the literature approaches the task of DF by employing fixed sample size tests within the conventional Neyman-Pearson framework. In the theory of testing hypotheses, the number of observations, i.e. the size of the sample on which the test is based, is usually treated as a constant for any particular problem [75]. In the following section, the benchmark fixed sample size method is summarized.

3.1.1 Average Log-likelihood Ratio Test

Detection rules for this method is derived as comparing the log- likelihood ratio (LLR) with a threshold denoted by γ [17]. If $\overline{\text{LLR}}(\boldsymbol{y}) \lessgtr \gamma$, detector selects $\mathcal{H}_0$, otherwise selects $\mathcal{H}_1$. Let $\boldsymbol{y}$ be the realization of RV $\boldsymbol{Y}$ in equation (2.8). Using equation (2.8) the conditional pmf of an observation is given as:

$$P(Y_n = y_n \mid X = x, \mathcal{H}_i) = \frac{e^{-J-x\sum_{k=0}^{n} p_k A}\left(J + x\sum_{k=0}^{n} p_k A\right)^{y_n}}{y_n!}, \tag{3.4}$$

where $X = \sum_{m=1}^{M} X_m$ is the sum of the sensor outputs and $x \in \mathcal{X} = \{0, 1/(L-1), 2/(L-1), \ldots, M\}$ is a realization of X.

When deriving the DF rules, the benchmark fixed sample size method in [17] approximates the $\sum_{k=0}^{n} p_k A$ in (3.4) with its limit as $n \to \infty$, with the assumption that a large number of pulses are transmitted, and number of the samples received at the FC at the beginning of an observation window are discarded until a steady state is achieved. While these assumptions make Y_n stationary and independent and identically distributed (i.i.d.), simplifying the derivation and the analysis of the DF rule, they imply the presence of an additional decision delay beyond the

fixed observation window length parameter N that does not account for the discarded samples, and also an inefficient utilisation of the molecules available for transmission. Since our primary focus is the decision delay, we chose to employ the exact non stationary conditional distribution in the derivation of the average likelihood function of the observations. Let $\mathbf{Y}^{(l)}$ represent the random vector containing l consecutive samples of Y_n, i.e. $\mathbf{Y}^{(l)} = [Y_1, Y_2, \ldots, Y_l]^T$, corresponding to an observation window length of $(l-1)T$. Using (3.4) and the independence of sequence Y_n, the conditional pmf of $\mathbf{Y}^{(l)}$ for a given realization of $X = x$ and the hypothesis $(\mathcal{H}_i)$ can be expressed as:

$$P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid X = x, \mathcal{H}_i\right) = \exp\left(-Jl - \sum_{n=1}^{l} xB_n\right) \prod_{n=1}^{l} \frac{(J + xB_n)^{y_n}}{y_n!}, \tag{3.5}$$

where $\mathbf{y}^{(l)} = [y_1, y_2, \ldots, y_l]^T$ a realization of $\mathbf{Y}^{(l)}$ and $B_n = \sum_{k=0}^{n} p_k A$. Clearly, during detection, it is not possible for the FC to have any a-priori information on the current realization of X. Thus, X is treated as a nuisance parameter by averaging equation (3.5) over its conditional pmf.

$$P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid \mathcal{H}_i\right) = \sum_{x \in \mathcal{X}} P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid X = x\right) P\left(X = x \mid \mathcal{H}_i\right). \tag{3.6}$$

Hence, the average likelihood function (over X) of a received signal vector $\boldsymbol{y}(l)$ of length l is calculated as:

$$\begin{aligned} P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid \mathcal{H}_i\right) = \sum_{x \in \mathcal{X}} \Bigg(& Q_i^{(M)}(x) \exp\left(-Jl - \sum_{n=1}^{l} xB_n\right) \\ & \times \prod_{n=1}^{l} \frac{(J + xB_n)^{y_n}}{y_n!} \Bigg), \end{aligned} \tag{3.7}$$

where $Q_i^{(M)}(x)$ is the conditional pmf of X, the aggregated sensor output and can be calculated as:

$$Q_i^{(M)}(x) = P\left(X = x \mid \mathcal{H}_i\right) = \underbrace{q_i(x) * q_i(x) * \ldots q_i(x)}_{M-1 \text{ times}}, \tag{3.8}$$

i.e. by convolving $q_i(x)$ in (2.5) $M-1$ times with itself, due due to the statistical independence of the individual sensor outputs X_m. Hence, the average log-likelihood ratio (ALLR) of $\boldsymbol{y}^{(l)}$ for this binary detection problem, is given as:

$$\tilde{\Lambda}_Y(l) = \log\left\{\frac{P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid \mathcal{H}_1\right)}{P\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid \mathcal{H}_0\right)}\right\}, \tag{3.9}$$

with $\mathrm{P}\left(\mathbf{Y}^{(l)} = \mathbf{y}^{(l)} \mid \mathcal{H}_i\right)$ calculated as in 3.7

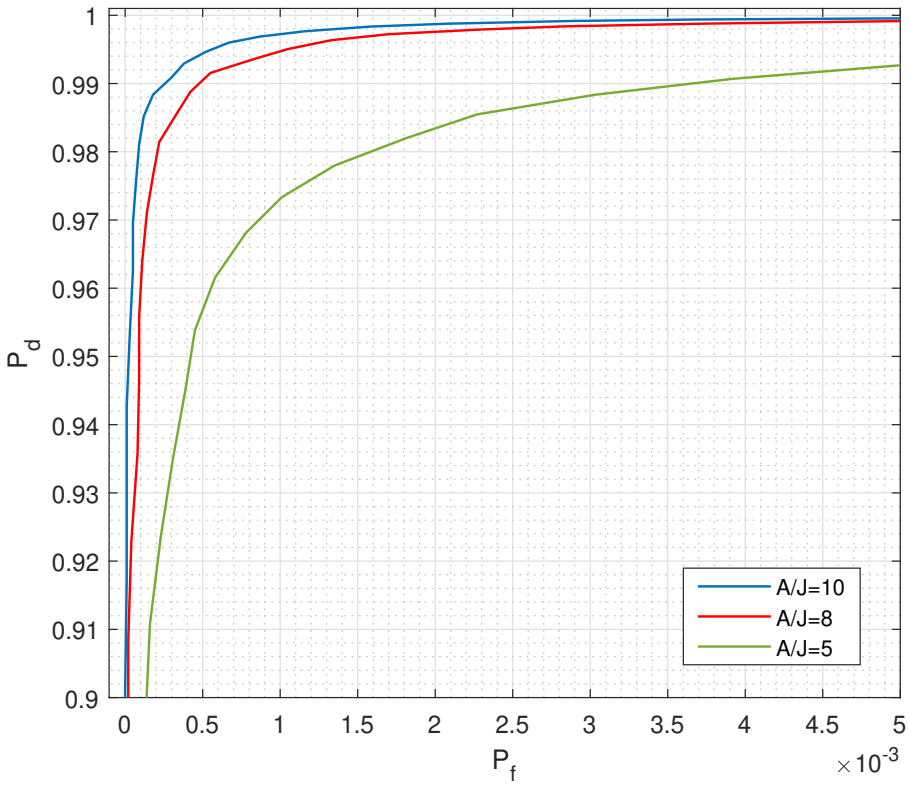


Figure 3.1: The ROC curve of fixed sample size test based DF with $M = 4$, $T = 500\mu s$, $(c_0, c_1) = (6.5, 7.5)$, and $N = 10$ under varying A/J ratios.

Before moving to SPRT details, it would be helpful to provide benchmark test results for the same simulation setup. Parameters used in the Monte Carlo simulations are given in Table 3.1. The simulations have been performed with $M = 4$, $T = 500$ μs, and $(c_0, c_1) = (6.5, 7.5)$. Fig 3.1 displays receiver operating characteristic (ROC) curve (P_d vs. P_f) for different ratio A/J referring signal to ratio (SNR) of the fixed sample size test in [17] chosen as a benchmark for Monte Carlo simulations. Please note that the test requires $N = 10$ to achieve this detection performance.

3.2 Sequential Test for Decision Fusion

The most important feature of the sequential test is that the number of observations required by the sequential test depends on the outcome of the observations. The number of the observation required is not deterministic, but a random variable [72], [76].

Sequential tests are equipped with a stopping rule that decides, at each time epoch, whether to wait and collect one more sample or to terminate and chose one of the hypotheses, and a

decision rule that decides for $\mathcal{H}_0$ or $\mathcal{H}_1$ based on the samples available up to the stopping time.

A decision rule is given for making one of the following three decisions at any stage of the detection procedure: (1) to accept the hypothesis $\mathcal{H}_1$, (2) is to reject the hypothesis $\mathcal{H}_0$, (3) to continue the detection procedure by making an additional observation. Thus, such a test procedure is carried out sequentially. Defining $\Lambda(l)$ as LLR of a vector of l samples of the received signal, $l = 1, 2, \ldots$, the SPRT proposed by Wald in [72] is given as:

$$\tau = \inf(l > 0 : \Lambda(l) \notin (S, U)) \tag{3.10}$$

$$\Gamma_\tau \triangleq \begin{cases} \text{Chose } \mathcal{H}_0, & \text{if } \Lambda(\tau) \leq S \\ \text{Chose } \mathcal{H}_1, & \text{if } \Lambda(\tau) \geq U \end{cases} \tag{3.11}$$

where τ is the stopping time of the test and Γ_τ is the decision rule. In other words, at each time epoch l, the running LLR $\Lambda(l)$ is compared with a lower and upper threshold S and U, $S < U$. If $\Lambda(l)$ remains within the interval $(S < U)$, the test decides to collect one more sample, updates the LLR and repeats the procedure for $l + 1$. The stopping time τ of the SPRT is defined as the time instant where the value of the LLR exits the interval $(S < U)$ for the first time. At the stopping time $l = \tau$ the test terminates, and decides for $\mathcal{H}_0$ if $\Lambda(\tau) \leq S$, and for $\mathcal{H}_1$ if $\Lambda(\tau) \geq U$. Clearly, the stopping time τ is a random variable, since its value depends on the random input sequence, and its expected value $E\tau$ characterizes the average sample number (ASN) of the test. Wald has shown that, for a binary hypothesis test with i.i.d observations, the SPRT minimizes the ASN over both hypotheses for a given false alarm probability $P_f = P\left\{\hat{\mathcal{H}} = \mathcal{H}_1 \mid \mathcal{H}_0\right\}$ and probability of detection $P_d = P\left\{\hat{\mathcal{H}} = \mathcal{H}_1 \mid \mathcal{H}_1\right\}$ pair [77]. For the DF scenario considered in this thesis, we propose to use the ALLR $\tilde{\Lambda}_Y(l)$ of the observation sequence given in (3.9) in the SPRT described in equations (3.10) and (3.11), resulting in a test which we will refer to as the sequential average probability ratio test (SAPRT), where, in contrast to the usual LLR based SPRT model, the running ALLR function $\tilde{\Lambda}_Y(l)$ of the observations cannot be expressed as a running sum of the log likelihood ratios of the individual samples due to the averaging operation over the pmf of the nuisance parameter X performed in (3.6), despite the fact that Y_n is an independent (albeit not i.i.d.) sequence.

The performance of a SPRT is characterized by its operating characteristics (OC) and the ASN. While Wald's approximation [72] provides expressions for bounds of S and U in terms of the P_f and P_d, which can be used to find approximate values for S and U, exact calculations of these parameters, and analytical derivation of the OC and the ASN is only possible for some special cases. [78] shows that OC and ASN functions obey the Fredholm integral equations of the second kind for the i.i.d. case which may be evaluated numerically, (and for some simple cases,

analytically). For the case of independent but not identically distributed (i.e. non-stationary) observations, [79] demonstrates a methodology for numerically approximating the OC and the ASN for some simple input distributions, which relies on recursively solving the governing integral equations. For the case of the SAPRT in the DF scenario considered in this thesis, the observation sequence Y_n is not i.i.d., and neither can the average LLR $\tilde{\Lambda}_Y(l)$ be expressed as a running sum of independent random variables as the LLR functions in [78], [79] and [72] as discussed above. Thus, an analytical derivation of the OC and ASN functions for this case remains mathematically intractable. However, the simulation results provided in the next section show that the proposed SAPRT based DF provides significant improvements in the average number of samples required for detection, and, subsequently, considerably less average decision delay compared to its fixed-sample-size counterpart in [17].

3.3 Performance Evaluation

In this section, the performance of the proposed SAPRT based sequential DF strategy is evaluated via both Monte Carlo simulations using the ideal signal model in equation (2.8) and equation (3.4), and particle based simulations. The particle based molecular communication simulator AcCoRD has been employed in the signal generation which is a sandbox reaction-diffusion solver designed by Dr. Adam Noel from the University of Warwick for the study of molecular communication systems [31].

AcCORD uses a hybrid of microscopic and mesoscopic simulation models that enables scalability via user control of local accuracy. Each molecule in microscopic regions is examined separately, and the evolution of each is recorded using a global microscopic time step g_{ts}. Mesoscopic areas keep track of the number of each type of molecule in each subvolume. The simulation's overall workflow can be divided into three main stages: preparation of the configuration file for running AcCoRD, execution of the software, and post-processing with MATLAB. A configuration file contains information about the setting's geometry, which is the physical space in which molecules may travel or shape, as well as information about the region's model, which may be microscopic or mesoscopic. Additionally, the configuration file defines actors. They serve as the simulation's interface by allowing for the input of molecules or the observation of molecules as output. Chemical properties such as absorption and desorption should be defined in the configuration file. All post-processing utilities were developed in MATLAB. Particle-based simulator AcCoRD is investigated further in Chapter 5 to simulate different scenarios. Because there are cases where there are no analytical models or difficult to define. We evaluate our NN algorithms

with the data produced with the AcCoRD simulator in Chapter 5 to evaluate DF task for such cases.

System parameters used in the simulation setup are presented in Table 3.1. Simulations are the average of 4×10^5 realizations and $g_{ts} = 5$ μs. In all cases, the diffusion coefficient $D = 500$ $\mu\text{m}^2/\text{s}$, in the same order of magnitude as the diffusion coefficients of small to medium sized biomolecules in blood plasma [80]. The size of the FC, $r_2 = 3$ μm has been chosen within the same range as a bacterial cell and the sensor-FC distance is $r_1 = 6$ μm. For the sensors, the number of quantization levels for the soft decision is chosen as $L = 4$ and the sensing imperfections are modeled with the following conditonal pmfs;

$$q_0\left(x_m\right) = \frac{\exp\left(-c_0 x_m\right)}{\sum_{x\in\mathcal{S}} \exp\left(-c_0 x\right)} \tag{3.12}$$

$$q_1\left(x_m\right) = \frac{\exp\left(c_1 x_m\right)}{\sum_{x\in\mathcal{S}} \exp\left(c_1 x\right)} \tag{3.13}$$

The coefficents c_0 and c_1 determine the sensing uncertainty of the individual sensors. The higher the coefficients, the less uncertainty there is in the sensing decisions under each hypothesis, and vice versa, allowing for the modelling of a wide range of sensor situations. We use the ratio A/J, i.e. the ratio of the maximum number of molecules available for a pulse to the expected number of noise molecules received at each time slot, as our SNR ratio. The efficiency of the proposed sequential test compared to the benchmark fixed sample size test in terms of decision delay is measured by the quantity Average Percentage Saving (APS) [81], which quantifies the saving in the average number of samples required for the decision achieved by the proposed test relative

Table 3.1: List of parameters used in the simulation setup.

Parameter	**Configuration 1**
A (The maximum number of molecules that a sensor can release for each pulse)	40
T (Time slot duration)	500, 600, 700 μs
N (Number of time slots)	10
M (Number or nano-sensors)	2, 3, 4, 5
L (Number of quantization levels)	4
r_1 (Sensor-FC distance)	6 μm
r_2 (FC radius)	3 μm
D (Diffusion coefficient)	500 $\mu\text{m}^2/\text{s}$

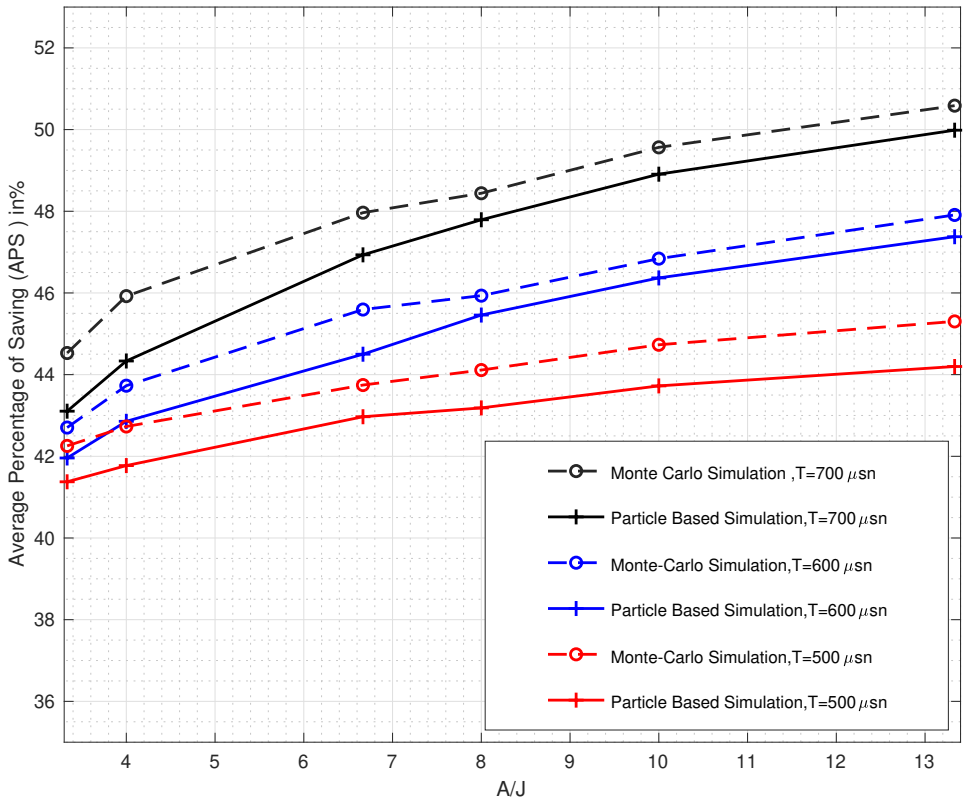


Figure 3.2: APS vs. A/J of the proposed SAPRT based DF, compared to the fixed sample size test for $M = 4$, $P_d = 0.999$ with $P_f = 0.001, T = 500, 600$ and 700 μs, $(c_0, c_1) = (6.5, 7.5)$, $N = 10$.

to the benchmark fixed sample size test in [17], i.e:

$$\text{APS} = \frac{N - \text{ASN}}{N} \times 100\% \tag{3.14}$$

where N is the sample size that the benchmark test requires to reach a decision for a given (P_f , P_d) pair, and the ASN is the average number of samples required by the proposed SAPRT to achieve the same detection performance, under the same conditions.

Fig 3.2 displays the APS of the SAPRT compared to the fixed sample size test in [17] chosen as a benchmark vs. A/J, both for the Monte Carlo simulations based on the ideal model, and for particle based simulations. The simulations have been performed with $M = 4$ sensors, $P_d = 0.999$, $P_f = 0.001$, and $T = 500, 600$ and 700 μs, $(c_0, c_1) = (6.5, 7.5)$, $N = 10$. Clearly, the APS results for the ideal Monte Carlo and particle-based simulations agree well, ca. within one percentage point across the board, where the particle-based case slightly under performs due to the likelihood mismatch between the ideal model and the signal generated by the particle based simulations, caused by the slightly lower signal mean achieved in the latter. Compared to the fixed sample size benchmark, the SAPRT achieves a considerable reduction in the average

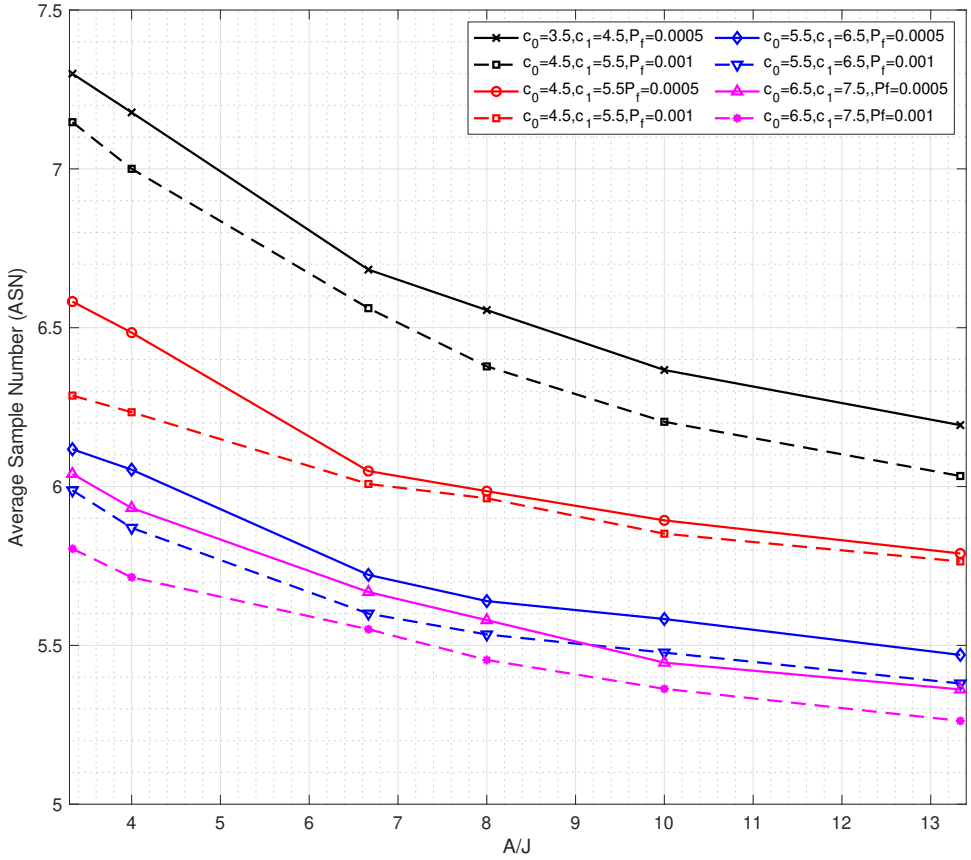


Figure 3.3: The ASN of the SAPRT based DF for $M = 4$, $P_d = 0.999$ with $P_f = 0.001$ and $P_f = 0.0005$, with $M = 4$, $T = 600$ μs under varying levels of sensor uncertainty. Only particle based simulations are considered.

number of samples required for detection, leading to a significant decrease in the average decision delay. The results show that, as expected, the APS increases both with the ratio A/J and T.

In the rest of this chapter, particle based diffusion simulations have been employed exclusively in the results. Fig 3.3 illustrates the effect of the sensor uncertainties on the performance of the SAPRT in terms of the average sample size ASN (in samples) required to achieve $P_d = 0.999$ with $P_f = 0.001$ and $P_f = 0.0005$ respectively, for the same network with $M = 4, T = 600$ μs. Here, four (c_0, c_1) pairs are chosen to model different sensing conditions, from excellent to moderate, in that order: $(c_0, c_1) = (6.5, 7.5), (5.5, 6.5), (4.5, 5.5)$ and $(3.5, 4.5)$. As expected, the SAPRT requires more samples to decide in order to achieve the required performance, as the sensor uncertainty increases. Furthermore, increasing P_f leads to a decrease in the ASN in all cases, which is also within expectations (see [81] for details).

Finally, Fig 3.4 displays the effect of the number of sensors on the ASN for $T = 600$ μs,

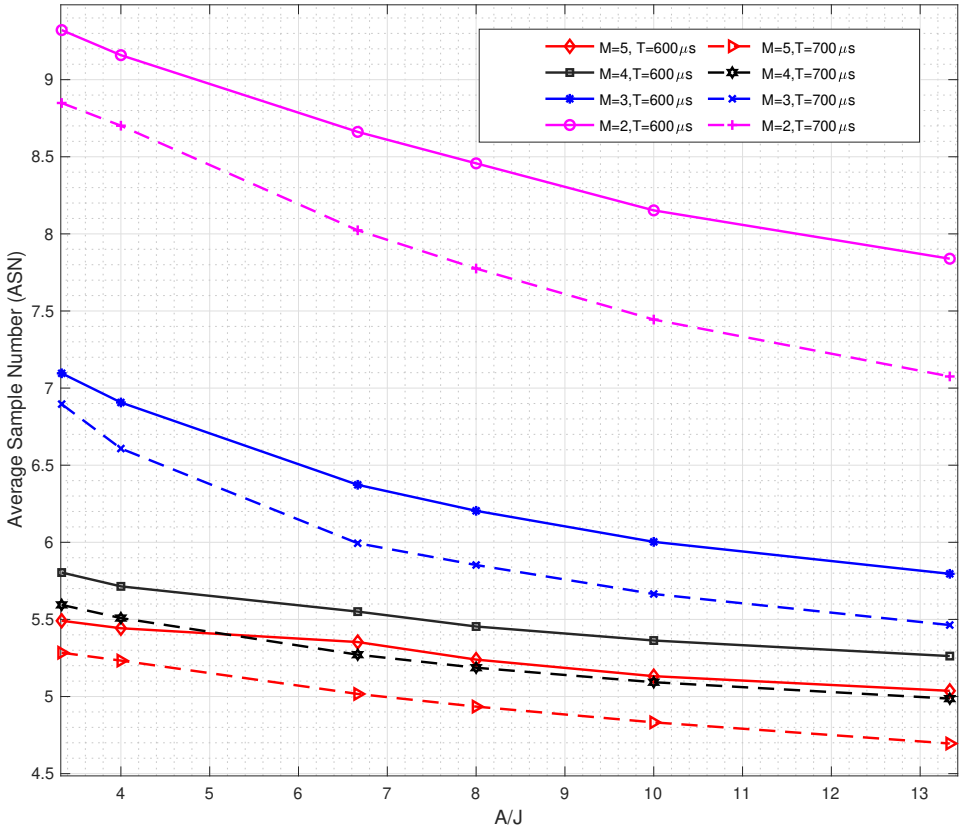


Figure 3.4: The ASN of the proposed SAPRT based DF for $P_d = 0.999$ with $P_f = 0.001$, $T = 600$ μs, and 700 μs, $(c_0, c_1) = (6.5, 7.5)$, $M = 2, 3, 4, 5$. Only particle based simulations are considered.

and 700 μs, $P_d = 0.999$ with $P_f = 0.001$ and $M = 2, 3, 4, 5$ respectively, where the detection performance increases (i.e. the ASN decreases) with increasing number of sensors M.

3.4 Conclusion

The use of a SPRT based test for the DF in a DD problem employing an MC based nanoscale sensor network is investigated in this chapter. The results show that the proposed SAPRT achieves considerable savings in the number of samples required for decision compared to an existing fixed-sample-size Neyman-Pearson benchmark test based on a maximum likelihood approach, while attaining the same detection performance. Furthermore, the proposed method does not rely on a simplifying approximation that, in practice, may lead to additional decision delays. This significant reduction in the decision delay makes the proposed strategy especially suitable for MC based DD problems, where the decision delay may become a major performance parameter. The proposed methodology is general, in the sense that it can be employed under any

type of diffusion dynamics (i,e, flow, reactions, anomalous diffusion, etc.), as long as the complete likelihood function of the receive signal is available at the FC, which, however, requires the knowledge of the all relevant system parameters. Note that, for practical cases, where some of the system parameters are unknown, and have to be estimated, the performance of the proposed DF methodology provides an upper performance bound.

Machine Learning for Decision Fusion

4

Reserve your right to think, for even to think wrongly is better than not to think at all.

Hypatia

In this thesis, we propose to use ML methods to detection task for the first time in the literature and show that this technique offers the resilience and flexibility necessary for practical implementation to overcome the main issues in DD in nanoscale sensor networks via diffusive MC.

One of the most challenging difficulties in MC and signal processing is the slow pace of propagation and the dispersive character of the channel, which results in significant decision delays. Secondly, channel characteristics will be highly unpredictable (unknown parameters) in practice. In most circumstances, it cannot be computed analytically since the partial differential equation that characterizes the channel has an analytical solution only in select geometrically very simple scenarios as stated in section 2.4.

We apply DNN techniques, namely two different NN structures, the FF- NN and the RNN. The RNN-based technique has proven to be better suitable for the suggested task, as expected, due to the sequential pattern of the data encountered in MC and its temporal structure, which the RNN can exploit. In both algorithms, high performance is possible alongside the resilience and flexibility necessary for practical implementation.

4.1 Machine Learning Basics

The definition of ML by Murphy [82] is

"*A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.*"

ML is one of the best promising methods to perform decision making under uncertainty which comes in many forms such as prediction about the future learning from past data, selecting the best model explaining available data etc. The past decades have seen rapid growth of different ML techniques. Specifically, deep learning has attracted researchers as a consequence of the advent of big data. Deep learning has emerged as a sub-field of Artificial Neural Networks (ANN) which is inspired by biological systems' information processing and distributed communication neurons. ANNs, simply called NNs, consist of interconnected nodes, referred to as neurons, each of which performs non-linear activation or transfer function. DNN comprises of NNs with multiple layers to model complex non-linear relationships. In recent years, DNNs have attracted widespread attention by outperforming alternative machine learning methods [83], [84]. This section is not intended to provide a comprehensive overview of ML but rather focus on a brief overview of NN techniques. Readers are referred to [85] for a detailed overview on the basic concepts of machine learning algorithms.

ML algorithms are discussed in this chapter whose goal is to predict the response variable given a sufficient amount of labelled data; this is referred to as supervised machine learning. However, machine learning algorithms are available in a variety of configurations, including reinforcement and unsupervised learning. Before going into more details of machine learning, basic terminologies of the field is explained in the following section.

A perceptron is a basic unit (an artificial neuron) that performs specific calculations to discover features or business intelligence in input data. A single layer perceptron is a one-layer neural network, while a multi-layer perceptron (feedforward NNs) consists of two or more layers referred as hidden layers. The networks comprise of four major components: input values, weights and bias, net sum, and an activation function as shown in Fig 4.1. The predicted variable $\hat{\boldsymbol{y}}$ is modelled as a composite of nonlinear functions of input $\boldsymbol{x}$ where $g(.)$ is the activation function, $\boldsymbol{w}$ are the weights and b_0 is the bias term.

$$\hat{y} = g\left(b_0 + \sum_{i=1}^{m} x_i w_i\right) \tag{4.1}$$

The process commences by multiplying all of the input values by their weights. The weighted sum is then calculated by multiplying all of the multiplied values and bias together. The weighted total

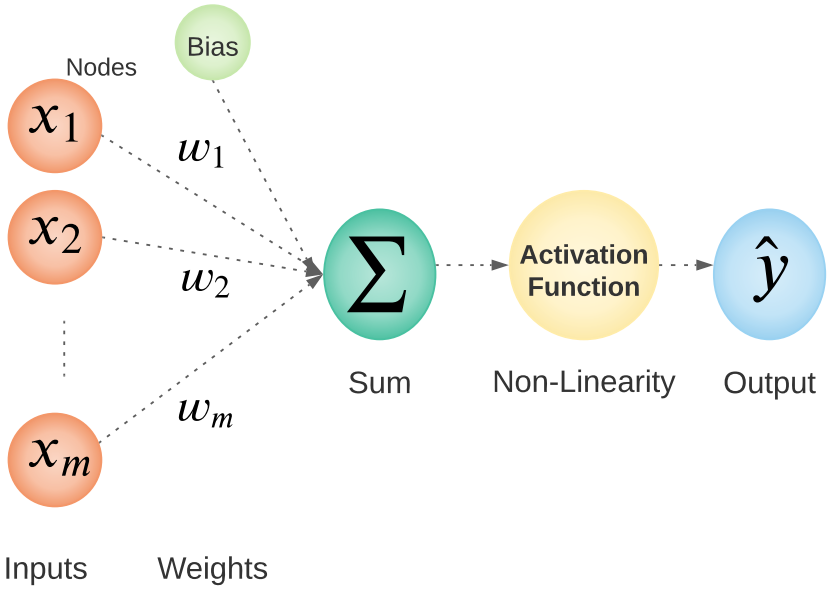


Figure 4.1: Schematic illustration of a perceptron.

is then applied to the activation function, yielding the output of the perceptron. The activation function is critical in ensuring that the output is mapped between required values.

4.1.1 Supervised Learning

Supervised Learning method enables the solution of statistical problems using examples of inputs and desired outputs. In contrast to conventional hypothesis testing, it is usually used where the underlying distributions are uncertain and are described using sample examples.

Although there are many approaches to automatic learning, the majority of popular approaches fall into the gradient-based learning category [86]. In the simplest cases, the learning procedure entails determining the value of the parameter that produces the least error rate that is defined as the percentage of instances where the prediction is incorrect. Classification accuracy is achieved by first applying a model to predict each sample in a test dataset. The predictions are then compared to the known labels for the test set examples. Accuracy is then calculated as the proportion of accurately predicted cases in the test set divided by all predictions made on the test set. The error rate, on the other hand, can be calculated as the total number of wrong predictions divided by the total number of predictions on the test set. An off-line training procedure is shown in Fig 4.2.

Each input $\boldsymbol{x}$ in the network is weighted with an optimum value of $\boldsymbol{w}$; the sum of weighted

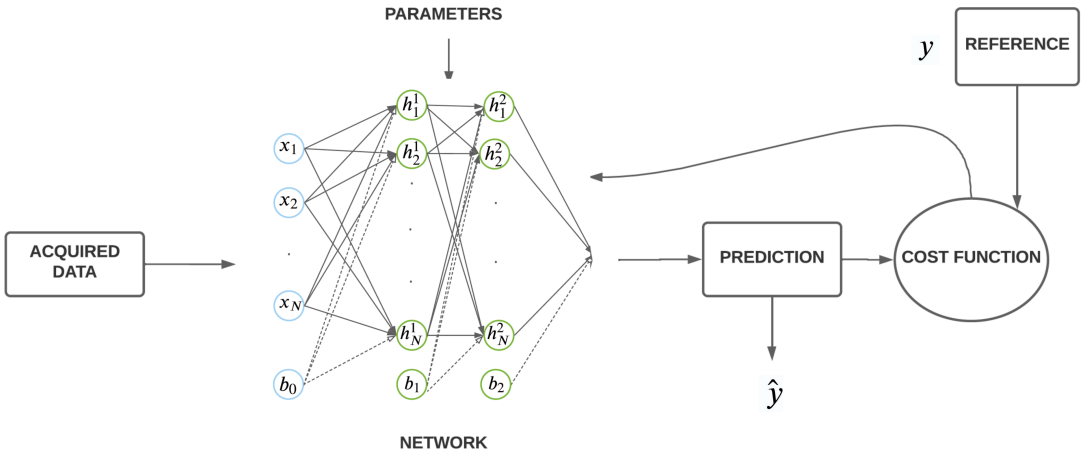


Figure 4.2: Schematic illustration of the training process.

inputs and the bias b creates the input of a activation function $h_i^{(j)}$ that represents the function of neuron i in layer j in hidden layers. From the input layer to the output layer, the transformation and nonlinear activation are calculated layer by layer.

In machine learning terminology, there are two sorts of parameters: model parameters and hyper-parameters. A model parameter is a configuration variable particular to the model, the value of which can be inferred from the data, such as the weights in a NN, the coefficients in a logistic regression. On the other hand, a hyper-parameter is an external setting to the model, the value of which cannot be estimated from the data. The optimal value for a model hyperparameter cannot be known. We can apply general rules, copy values from other networks we designed for similar problems, or utilize trial and error to get the optimal solution. Using a grid search or a random search, these parameters can be tuned to determine the model parameters that provide the best reliable estimates. Some examples of the hyperparameters are:

- Number of hidden layers and neurons: Between the algorithm's input and output are hidden layers in which the function applies weights to the inputs and directs them through an activation function as the output.

- Learning rate : It sets the step size for each learning iteration while aiming to minimize the loss calculated by loss function.

- Number of epochs: This hyperparameter specifies the number of full runs over the training data.

- Batch size: This states the number of training samples that must be processed before the

model's internal parameters are changed.

- Activation Function : In a neural network, the activation function specifies how the weighted sum of the input is turned into an output from a neuron or neurons within a layer.

The discrepancy between the output of the network and the desired output is measured by a cost function (The loss function calculates the error for a particular training example; the cost function is the sum of all training set's loss functions). In its most basic form, the learning problem entails determining the value that minimizes the cost function. By calculating the effect of minor changes in parameter values on the loss function, this function can be minimized. The gradient of the loss function with respect to the parameters is employed for quantifying the effect of these changes. This determines the error rate that is back-propagated [87] through the network to obtain a new set of updated parameters. This process called *back-propagation* is repeated until the training has completed. Once all the parameters are learned, data previously unseen can be predicted efficiently.

The general formula for the loss function for a given ground truth $\boldsymbol{y}$ and predicted output $\hat{\boldsymbol{y}}$ in supervised learning is stated as

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \min_{f} \sum_{i=1}^{N} \eta\left(\mathbf{y_i}, f\left(\mathbf{x_i}; \mathbf{w}\right)\right) \tag{4.2}$$

where $f(.)$ is the system that we want to extract information from, η denotes the underlying loss function that quantifies the penalty of deviating from the ground-truth. Two main types of loss function are;

- *Cross-entropy* for classification problems:
  Cross-entropy measures the difference between two probability distributions p and q for a given random variable or set of incidents. The cross-entropy of a distribution q relative to a distribution p is defined as follows:

  $$H(p, q) = -\mathrm{E}_p[\log q] \tag{4.3}$$

  where $E_p[\cdot]$ is the expected value function with respect to the distribution p.

- *Mean-square-error* for regression problems:
  The loss is calculated as the mean of the squared discrepancies between true and predicted values

  $$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} \left(y_i - \hat{y}_i\right)^2 \tag{4.4}$$

  where y_i is the true value and $\hat{y}_i$ is the predicted value.

There are other examples of loss functions can be applied to empower the algorithm depending on the application, such as *Hinge* loss, *squared-Hinge* loss, *Chebyshev* loss, [88].

4.1.2 Reinforcement Learning

The system communicates with its environment through the production of actions that have an impact on the state of the environment, resulting in earning scalar rewards (or penalties). The system's objective is to learn to behave in such a way that it maximizes future rewards (or minimizes future punishments) over its lifetime.

4.1.3 Unsupervised learning

This is a process in which the system receives inputs but does not receive supervised goal outputs or rewards from its environment. In other words, unsupervised learning is a form of self-organized learning that enables the discovery of previously unknown patterns in a data set in the absence of pre-defined labels.

4.2 Deep Neural Networks

Deep learning is a subset of ML approaches based on ANN and representation learning. This form of ML technique called DNN employs multiple layers of neurons to extract high-level information from the input data. Learning can take place in a supervised, semi-supervised, or unsupervised environment.

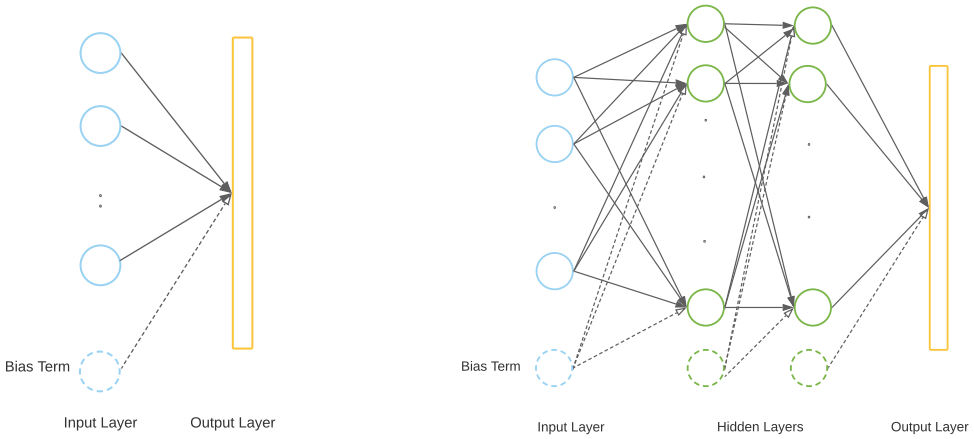
In this thesis two main DNN techniques have been considered:

1. Feed-forward (acyclic) NNs
2. Recurrent (cyclic) NNs

In the following sections, both DNN techniques are explained briefly.

4.2.1 Deep Feed Forward Neural Networks

Deep Feed-forward NNs, also called multilayer perceptrons (MLPs), are the most well-known deep learning models (Fig 4.3b). These models are referred as feed-forward since information flows from input to the function being evaluated, through the intermediate computations, and



(a) Perceptron (b) FF-NN

Figure 4.3: Different deed-forward neural network architectures.

finally to the output. There are no feedback connections in which the model's outputs are fed back into the model. The single layer of neural networks is called Perceptron (Fig 4.3a). The network aims to map an input x to a category y in a classification problem which is the approach of abnormality detection scheme in this thesis. Feed-forward NNs are referred to as networks because they are usually represented as a set of several functions. The predicted output $\hat{\mathcal{Y}}$ is modelled as a composite of nonlinear functions of input as

$$\hat{\mathcal{Y}} = f(\mathbf{x}; \mathbf{w}) = f_{\mathbf{w}_k} \circ f_{\mathbf{w}_{k-1}} \circ \ldots \circ f_{\mathbf{w}_0}(\mathbf{x}), \tag{4.5}$$

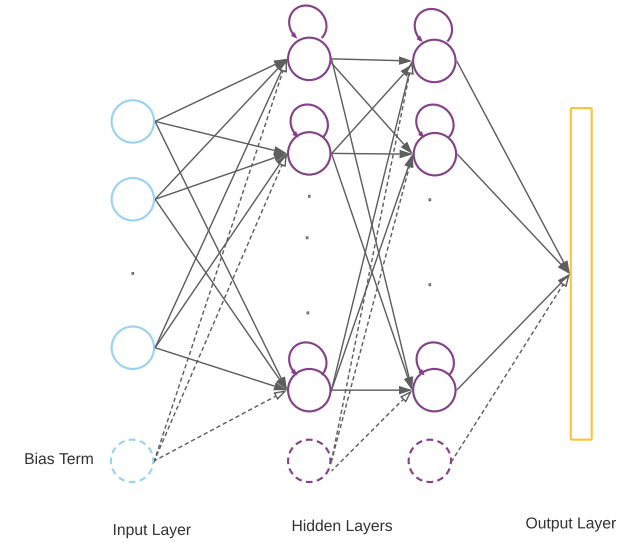
where each composite is described as a function of previous variables (neurons) in each layer and their related parameters (weights)

$$f_{\mathbf{w}_{k-1}}(\mathbf{k}) = p(\mathbf{w}_{\mathbf{k-1}}^T \mathbf{k} + b_{k-1}), \tag{4.6}$$

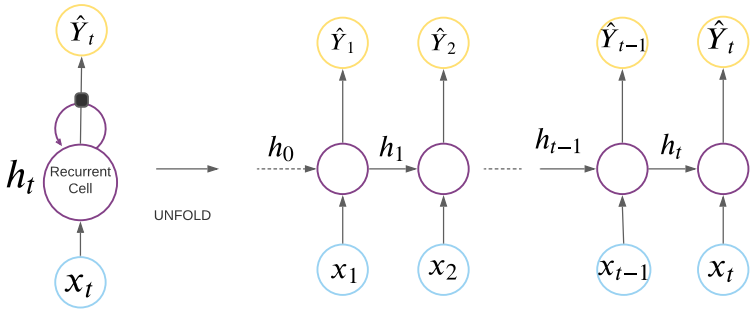
where p is an activation function, $\mathbf{k}$ is the hidden variables (neurons), $\mathbf{w}_{\mathbf{k}-1}$ are the learned parameters, b_{k-1} is the bias term. The number of the layers indicates the model's depth. Due to the fact that the training data does not reveal the desired output on any of these layers, they are referred to as hidden layers. The output layer is the final layer of the network that is responsible to produce the final result.

4.2.2 Recurrent Neural Networks

When FF-NNs are expanded to provide feedback connections, they are called RNN [86]. RNN is a type of neural network that is used to process time-series data and sequential data such as text, video, language, and genomes [89]. The basic architecture is depicted in Fig 4.4. Transitioning



(a) RNN



(b) RNN cell

Figure 4.4: Recurrent neural network (a) architecture and (b) cell.

from FF-NN to RNN is conceptually straightforward. Traditionally, feed-forward networks have been used to map fixed-size inputs to fixed-size outputs. On the contrary, RNN operates naturally on variable-length input sequences and map to variable-length output sequences, such as mapping from an image to various sentences that define it. This functionality is accomplished through the gradual exchange of parameters and transformations [90]. Also, RNN can handle far longer sequences than networks that do not use sequence-based specialization.

When a RNN is equipped to perform a task that allows it to forecast the future from the past, it learns to use hidden state $\boldsymbol{h}$ as a lossy summary of the task-relevant aspects of the previous sequence of inputs up to t. In Fig 4.4b both folded and unfolded representations are shown. Folded representation contains one neuron with a backward arrow representing the delay of a single time step as shown in the left of the figure. While unfolded version contains separate neurons for each variable for each time step as in the right of the figure. Let $\mathbf{x} = [x_1, x_2, ..., x_N]$

be the sequence of examinations of inputs where N is the number of time steps. Simple RNN maps the sequence of inputs to a sequence of hidden states $\mathbf{h} = [h_1, h_2, ..., h_N]$ though a set of parameters θ.

$$\mathbf{h}_t = g\left(\mathbf{h}_{t-1}, \mathbf{x}_t; \boldsymbol{\theta}\right) \tag{4.7}$$

where g should not be interpreted as a simple activation function, e.g., tanh or sigmoid. For instance, in this thesis we prefer a special architecture of RNN called Long short-term memory (LSTM), where g is a composition of several gates and nonlinearities. The range of contextual information that ordinary RNN can acquire is fairly limited in practice because of the issue is that as an input data cycles across the network's recurrent connections, its influence on the hidden layer, and therefore on the network output, either decays or explodes exponentially. This law is known as the vanishing gradient problem in the literature. LSTM is an RNN architecture that was created expressly to overcome the vanishing gradient problem [86]. Rather than a unit that just applies element wise nonlinearity to the affine translation of inputs and recurrent units, LSTM have memory cells and corresponding gate units to apply internal recurrence in addition to the RNN's outer reference. Because LSTMs are effective at capturing long-term temporal dependencies, they have been utilized to improve the state of the art for a wide range of tough tasks. This encompasses handwriting recognition and generation, language modeling and translation, acoustic modeling of speech, voice synthesis, protein secondary structure prediction, audio and video data analysis, and so on.

RNN, by applying equation (4.7) multiple times, lets us build rich, complex models that can be trained in an end-to-end approach. To begin, inputs are feed into an input layer, resulting in a series of hidden states; these hidden states are then used as inputs to another RNN, and so forth. RNN can be generalized as

$$\hat{\mathcal{Y}}_t = f(\mathbf{x_t}; \mathbf{w}) = g_y\left(\mathbf{w}_{hy}\mathbf{h}_t + \mathbf{b}_y\right). \tag{4.8}$$

where $\hat{\mathcal{Y}}_t$ is the output of the network, $\mathbf{w}$ and $\mathbf{b}$ are coefficients that are shared temporally and g_y is the activation function. The fact that we are doing the same task at each step is reflected in parameter sharing; therefore, we do not need to relearn the rules at each point in the network.

4.2.3 Evaluation of the Algorithm

The major constraint in machine learning is that the algorithm must perform well on new, previously unknown inputs not only those used to train our model. The capacity to do well when confronted by previously unobserved data is referred to as *generalization* [91]. What distinguishes

machine learning from optimization is acquiring low generalization error which is defined as the error's expected value on new input. The prediction error can be classified into two main components: bias and variance. Bias error is mainly due to the improper choice of model complexity usually approximating complex problems by a simple model. In comparison, the variance error occurs as a result of the training set's small size [92]. The purpose of machine learning is to create models that generalize well. However, the major problem is the overfitting problem, i.e. when a model learns too much detail and noise in the training data (the known data). Since one can only control what one can see, it's important to be able to consistently quantify the model's generalizability. The performance of the model must be evaluated on the new-unseen data that's why it is strongly recommended that the data set be subdivided into independent training, test, and validation sets. The model is trained using the training data and is evaluated using the test data. Once the model is ready for release, it can be evaluated on the validation data one final time. The reason behind testing the model on two separate data after training is tuning its configuration, namely its parameters. During the tuning process, feedback signal from the test data is used to improve the performance. Therefore tuning is also a part of the learning process and can result in overfitting to test data set. In the following section, a brief overview of classification metrics is presented.

4.2.3.1 Classification Metrics

Rather than predicting classes directly, it may be more flexible to predict the probability of an observation belonging to each class in a classification problem. This flexibility stems from the way probabilities can be interpreted using a number of different thresholds, which enables the model's user to trade off concerns about the model's errors, such as the number of false positives versus false negatives. For instance, a default approach would be to use a threshold of 0.5, which indicates that a probability in the range $[0.0, 0.49]$ is a negative outcome $\mathcal{H}_0$ and a probability in the range $[0.5, 1.0]$ is a positive outcome $\mathcal{H}_1$. This threshold can be modified to fine-tune the model's actions for a given challenge. ROC Curves and Precision-Recall curves are two diagnostic tools that aid in the interpretation of probabilistic predictions for binary (two-class) classification problems. ROC analysis has been extensively used in signal processing and communications to evaluate the effectiveness of a detection algorithm [40]. The following metrics can be computed given a confusion matrix (Table 4.1) where true positive (TP), false positive (FP), false negative (FN), true negative (TN) rates are given.

Table 4.1: Decision Matrix

	True Condition		
Prediction	Positive	Negative	Total
Positive	TP	FP	T_+
Negative	FN	TN	T_-
Total	D_+	D_-	

Accuracy and Error rate: Accuracy is commonly used to evaluate a classifier performance which is the ratio of the number of correct predictions, both correct positive and correct negative, to the total number of input samples. Accuracy rate performs well when the number of samples of each class is evenly distributed. But, for problems with unbalanced categorization, accuracy is an ineffective performance metric. Precision and recall metrics provide an alternative to classification accuracy for this kind of problems.

$$\text{Accuracy } = \frac{TP + TN}{TP + TN + FP + FN}, \text{ ErrorRate } = 1 \text{ - Accuracy} \tag{4.9}$$

Precision: It is the ratio of cases that the classifier correctly classified as positive (it is used as the probability of detection (P_d) in previous sections referring to Neyman-Pearson detection framework); it is also referred to as positive predictive value. Therefore, precision measures the accuracy of the minority class. This term is defined as;

$$\text{Precision } = \frac{TP}{TP + FP} = \frac{TP}{T_+} \tag{4.10}$$

Recall: It is the ratio of positive cases that the classifier labels as positive, it is known as TP rate. Unlike precision, which only considers the accurate positive predictions out of all positive predictions, recall considers the positive predictions that were missed. Recall states some indication of the positive class's coverage. This term is defined as;

$$\text{Recall } = \frac{TP}{TP + FN} = \frac{TP}{D_+} \tag{4.11}$$

Specificity: It is the ratio of negative cases that the classifier did label as negative (it is used as the probability of false alarm (P_f) in previous sections referring to Neyman-Pearson detection framework), it is also referred to as TN rate. This term is defined as

$$\text{Specificity } = \frac{TN}{TN + FP} = \frac{TN}{D_-} \tag{4.12}$$

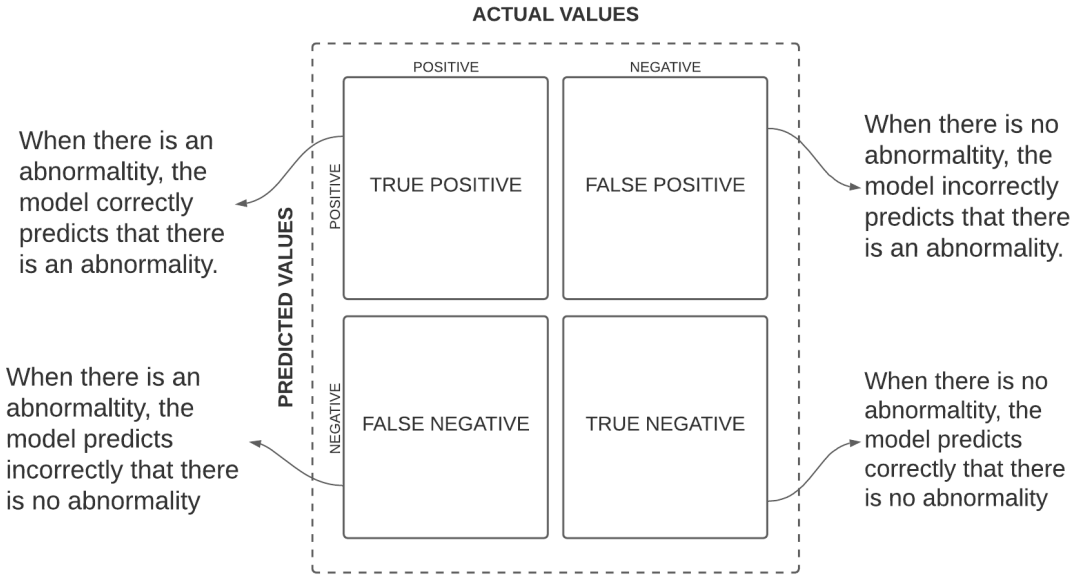


Figure 4.5: Confusion Matrix.

F-Measure: It is the harmonic mean of precision and recall and provides additional insight into the classifier's functionality. As a result, it is more effective than accuracy, especially in cases of class imbalance.

$$F_\beta = \left(1 + \beta^2\right) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \tag{4.13}$$

where the balance between precision and recall is controlled by β. 1 or 2 are the common values of the score. The better the total performance, the higher the F-score.

4.2.3.2 Confusion Matrix

A confusion matrix is a way of analyzing a classification algorithm's performance (Fig 4.5). If there are an uneven amount of observations in each class or if the dataset has more than two classes, classification accuracy alone might be deceptive as explained in previous section. The number of true and wrong predictions is summed and split down by class using count values then filled into the matrix. TP and FN values can be used to measure the accuracy and error. The values of recall, specificity, F-measure can be calculated with the rates given in the matrix. Confusion matrix gives a summary to understand what the classification model is getting right and where it is going wrong.

4.2.3.3 Receiver Operating Characteristic (ROC)

As previously mentioned, ROC analysis is a highly effective and fundamental technique for evaluating model results. It is useful to explain the ROC analysis in machine learning terminology with its specific terms. It is known as a plot of the model's sensitivity, or TP rate (P_d), against its specificity, or FP rate (P_f), as the x-coordinate, where both TP rate and FP rate are computed at each possible threshold.

4.3 Decision Fusion Based on Neural Networks

The signal propagation characteristics in a diffusive MC channel are highly random in nature and depend heavily on a multitude of factors such as the geometry (i.e. the physical bounds) of the fluid medium of propagation, its chemical properties, the chemical characteristics and the physical extent of the transmitters and the receivers, the geometry of the network (i.e. the relative positions of the transmitters and the receivers), and the environmental conditions, such as flow, temperature, viscosity, physical obstacles, etc. Existing decision theory based strategies for abnormality detection via diffusive MC based nanoscale sensor networks require a complete statistical characterization of the underlying communication channel between the sensors and the FC, with the assumption of perfectly known or accurately estimated channel parameters. This assumption is usually impractical both due to the mathematical intractability of the analytical channel models for MC except in a few ideal cases, and the slow and dispersive signal propagation characteristics that make the channel estimation a difficult task even in these ideal cases. This was also the case in Chapter 3, where we have employed decision theoretical methods for DF.

In this thesis, for the first time in the literature we propose a detector based on deep learning, specifically on aN FF-NN and an RNN structure that learn the underlying model from data. This study shows that the proposed decision fusion strategy can perform well without any knowledge of the communication channel that learn from the data providing a more pragmatic and practical approach to DF as the living organisms do in nature. Algorithm can perform well where an analytical channel model is not mathematically tractable, or too complex to be of practical use, and for cases where a channel model exists but an accurate estimation of its parameters is impractical. This approach provides the robustness and flexibility required for practical implementation. Recent works such as [93] propose the use of an end to end approach for demodulation of sequences of data symbols in point to point optical and molecular communications systems employing RNN. A deep learning-based end-to-end approach is proposed to DF in a diffusive MC

based nanoscale sensor network. In particular, the use of FF-NN and RNN structures is investigated to perform the DF. The algorithm proposed in [17] which is based on an approximation of the LLR of the sensor outputs received at the FC is employed as a benchmark for the performance evaluation. We show that our approach can be used to design DF algorithms achieving higher probability of detection and show more robust detection performance under different channel conditions and/or without any CSI and the statistical characteristics of the sensing model of the participating sensor nodes. In the following section, FF-NN based and RNN based approaches are presented and compared. Both approaches outperform the simplified-LLR test while RNN based approach performs better than FF-NN based one. Accordingly, RNN based approach has been further investigated in the rest of the thesis.

4.3.1 Decision Fusion based on a Feed-Forward Neural Network

To provide a fair comparison between the proposed machine learning-based methodology and the existing model-based approaches, the same general and abstract sensing model is employed at the participating sensor nodes and the same communication model between the sensors and the FC as the method proposed in section 2.5. Performing the DF at the FC by using a NN based ML approach through supervised learning is proposed. As the first phase of the approach, the detector is trained by using a training dataset. Once the training phase is completed offline, it is no longer a part of the detection procedure. A general FF-NN architecture is shown in Fig 4.3b. An architecture comprises of a FF-NN with one input layer consisting of 16 neurons, one fully connected hidden layer consisting of 16 neurons, and one dense layer as an output layer is considered.

4.3.2 Decision Fusion based on a Recurrent Neural Network

As explained in previous sections, RNN processes an input sequence one element at a time, keeping a state vector in their hidden units that indirectly contains details about the history of all the sequence's previous elements [94]. RNN is an extremely efficient dynamic system, but training them has proven difficult because back-propagated gradients expand or shrink with each time stage, causing them to burst or vanishing over a long period [95]. LSTM was developed to solve the problem of vanishing gradients, and it has since become one of the most common RNN architectures [96] as explained in previous sections. For this reason, one of the special kind of RNN, LSTM architecture is considered in this thesis. Specifically, one of the RNN applications called the many-to-one model is used as shown in Fig 4.6. A RNN with one LSTM layer consisting of

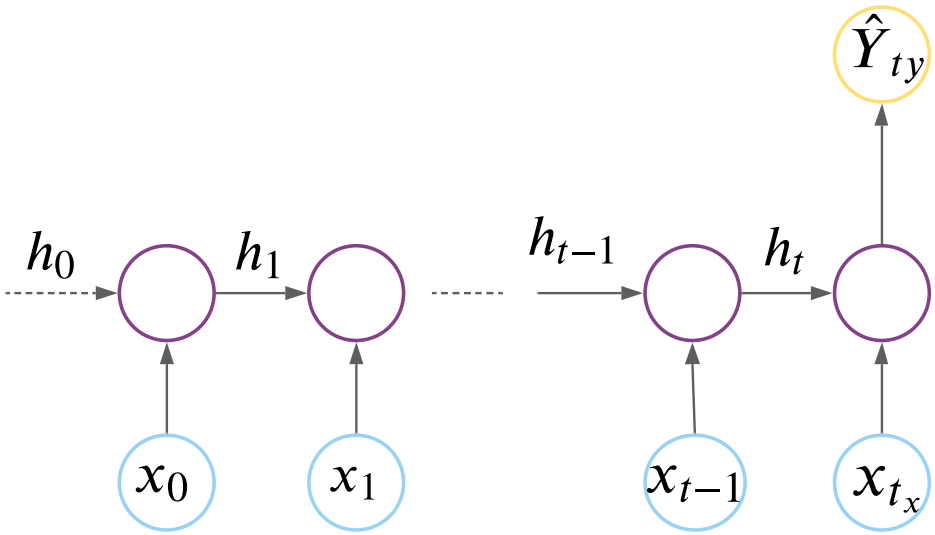


Figure 4.6: Recurrent LSTM neural network many-to-one architecture.

16 neurons and one dense layer as an output layer is applied.

4.3.2.1 Training the Neural Network Detectors

The training data can be generated by various means, e.g. from in vivo measurements, from in vitro experimental data, from particle-based diffusion/reaction simulations of the sensor network, or by generating the corresponding signal sequences using the existing statistical system and channel models, such as the one described in Chapter 2. This makes the proposed approach applicable even if no tractable analytical description is available for the statistical characteristics of the channel. For the sake of a fair comparison between existing model-based approaches based on the Poisson channel model, we have chosen to generate the training data set employing the Poisson model from Chapter 2, using various values for the parameter set $\zeta_{NN} = [J, r_1, r_2, D]$, (see Table 4.2). This makes the resulting DF algorithms robust to possible changes in the system parameters while allowing DF without requiring parameter estimation, provided that the system parameters remain within the boundaries of the training set. Let $\mathbf{y}^{(w)} = [y_1^{(w)}, y_2^{(w)}, ..., y_N^{(w)}]$ represent a known signal sequence of length N observed at the FC for the corresponding known hypothesis $\mathcal{H}_i^{(w)}$. The training data consists of W samples of $(\mathbf{y}^{(w)}, \mathcal{H}_i^{w})$ pairs.

$$\{(\mathbf{y}^{(1)}, \mathcal{H}_i^{(1)}), (\mathbf{y}^{(2)}, \mathcal{H}_i^{(2)}), \dots, (\mathbf{y}^{(W)}, \mathcal{H}_i^{(W)})\} \tag{4.14}$$

The data set consisting of $100k$ samples is used to train the proposed FF-NN and RNN detectors that map the received signal $\mathbf{y}^{(w)}$ to a hypothesis $\hat{\mathcal{H}} \in \{\mathcal{H}_0, \mathcal{H}_1\}$. During the training, the

known $(\mathbf{y}^{(w)}, \mathcal{H}_i^{(w)})$ pairs are used to find the optimal set of parameters for the corresponding NN structure. Both for the FF-NN and RNN training, we apply back-propagation with binary cross-entropy to minimize the loss between the actual hypothesis $\mathcal{H}_i$, and estimated hypothesis $\hat{\mathcal{H}}$. Adam optimizer is applied with a learning rate of 10^{-3} to calculate the FF-NN parameters without over-learning [97]. This optimizer maintains the learning rate for each network weight and makes use of the average of the second moments of the gradients. The number of epochs used during training is 100 and the batch size is 10. For the RNN, the RMSprop optimizer is employed with a learning rate of 10^{-3} which is the gradient descent algorithm with momentum which helps to calculate the RNN parameters without over-learning [98]. For the RNN, the same number of training epochs and the same batch size has been employed as for the FF-NN case.

4.3.2.2 Performance Evaluation

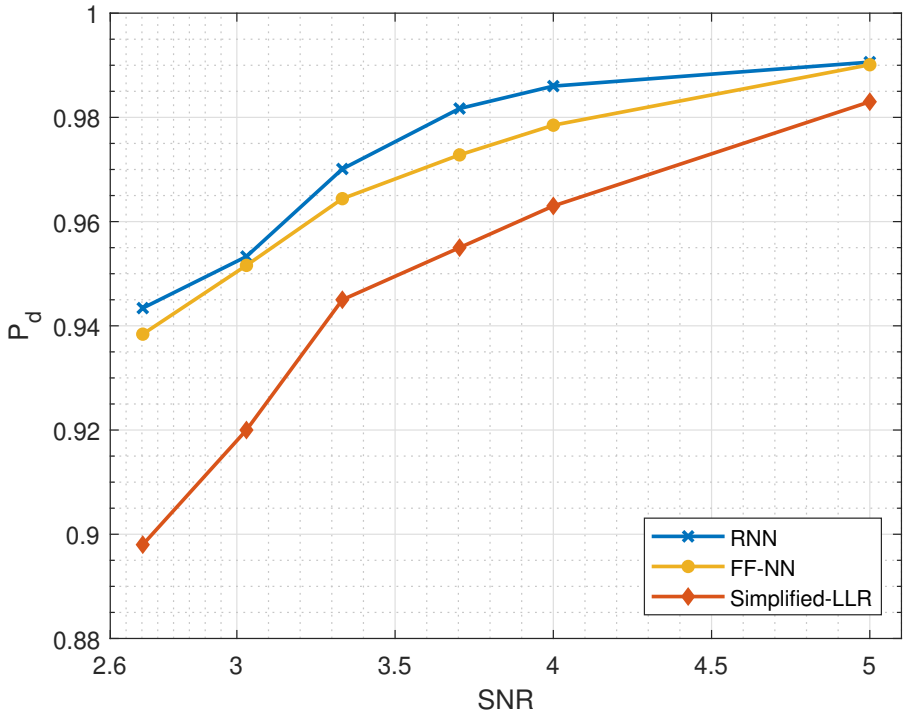
In the following, the performance of the proposed NN based detectors via simulations and the existing LLR detector from [18] used as a benchmark is proposed. The detection probability is employed $P_d = P(\hat{\mathcal{H}} = \mathcal{H}_1|\mathcal{H}_1)$ and the false alarm rate $P_f = P(\hat{\mathcal{H}} = \mathcal{H}_1|\mathcal{H}_0)$ as our main performance criteria where $\mathcal{H}_0$ and $\mathcal{H}_1$ represent the absence and the presence of the abnormality of interest, respectively. System parameters used in the configuration-1 simulation setup are presented in Table 4.2. In the training and test data, we use $N = 4$, and J varies depending on the *SNR* level, which we define as $SNR = A/J$. The conditional pmf's of the sensors used on the simulations are given as:

$$g_0(x_m) = \frac{\exp(-6.5x_m)}{\sum_{x_m \in \mathcal{X}} \exp(-6.5x_m)}, \tag{4.15}$$

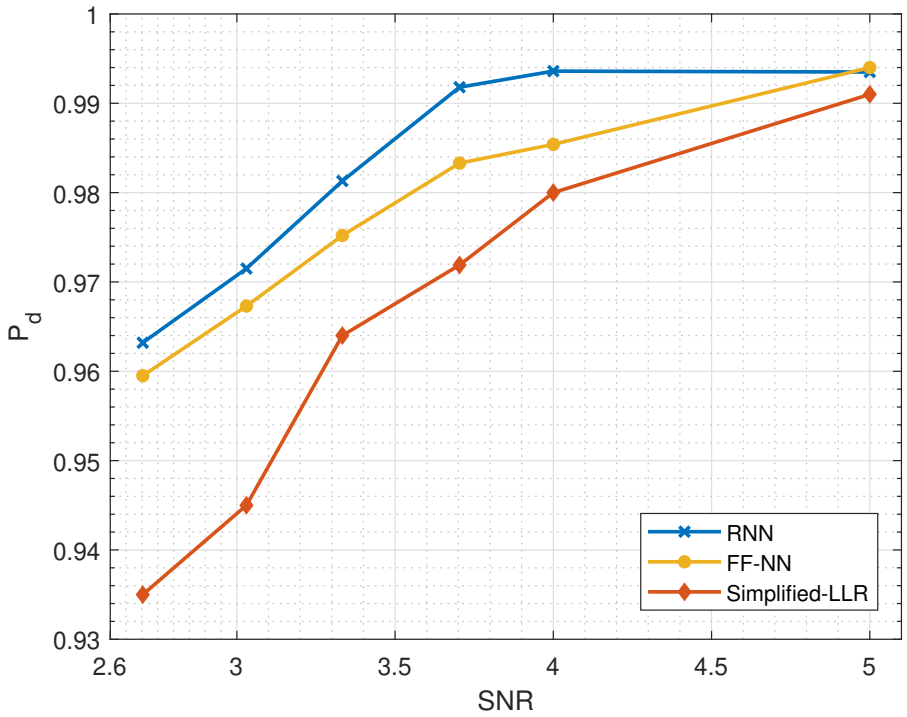
$$g_1(x_m) = \frac{\exp(7.5x_m)}{\sum_{x_m \in \mathcal{X}} \exp(7.5x_m)}. \tag{4.16}$$

Table 4.2: List of parameters used in the training and the simulations

Parameter	Configuration 1
A (The maximum number of molecules that a sensor can release for each pulse)	100
T (Time slot duration)	70 ms
N (Number of time slots)	4
Number or sensors in the nanonetwork	4
L (Number of quantization levels)	4
r_1 (Distance between the nanosensors and the receiver)	4,5,6,7,8,9 μm
r_2 (The receiver radius)	4 μm
$A/J(SNR)$ levels	2-5
D (The molecules diffusion coefficient)	50, 79.4 $\mu m^2/s$



(a) P_d vs SNR (P_f=0.05).



(b) P_d vs SNR (P_f=0.075).

Figure 4.7: P_d vs SNR for $P_f = 0.05$ (a) and $P_f = 0.075$ (b).

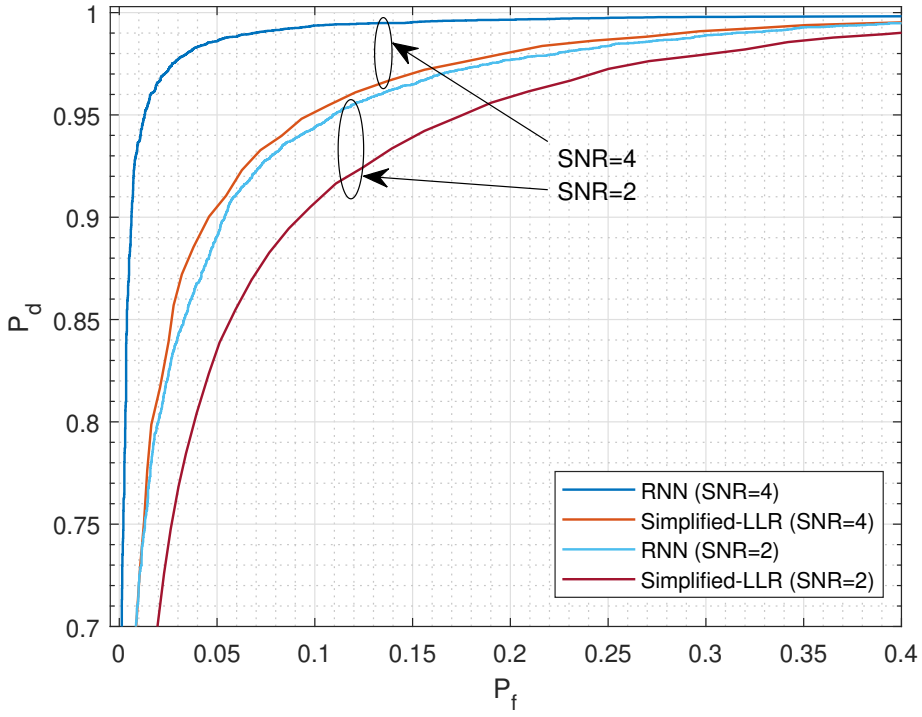


Figure 4.8: ROC curve of RNN and simplified- LLR detector for different *SNR* levels.

for $x_m \in \mathcal{X} = \{0, 1/(L-1), 2/(L-1), ..., 1\}$ with $L = 4$. Figs 4.7 (a) and (b) exhibit the detection performance of the proposed RNN and FF-NN detectors, and of the existing LLR detector from [17] for a fixed false alarm rate $P_f = 0.05$ and 0.075 respectively. Clearly, both of the proposed NN based detectors outperform the LLR detector, which is due to the fact that the LLR employs only an approximation of the likelihood function, whereas the proposed detectors both learn from actual data resulting in a better approximation to the impractical optimal detector. It can be seen from Fig 4.7 that, between both of the proposed detectors, the RNN approach outperforms the FF-NN. Therefore, in the rest of the thesis, we focus on the evaluation of the RNN detector. This is due to the well-known capability of this type of neural networks for taking into account the temporal structure of the data, which makes RNN especially suitable for this DF task. Fig 4.8 compares the ROC of the RNN and the LLR detectors for two different *SNR* levels. As in the previous case, the RNN detector performs better than the LLR detector in each case. Note that the proposed detector is robust enough to changes in noise level since the training data set includes data with different *SNR* levels.

Next, the effect of the sensor to FC distance r_1 is studied, which is a crucial parameter for the channel characteristics, on the detection performance of the proposed RNN detector, and compare it to that of the benchmark LLR. Three different values for the P_f is considered for $SNR = 2$. Figs 4.9, 4.10, 4.11 clearly shows that the proposed RNN detector for DF outperforms

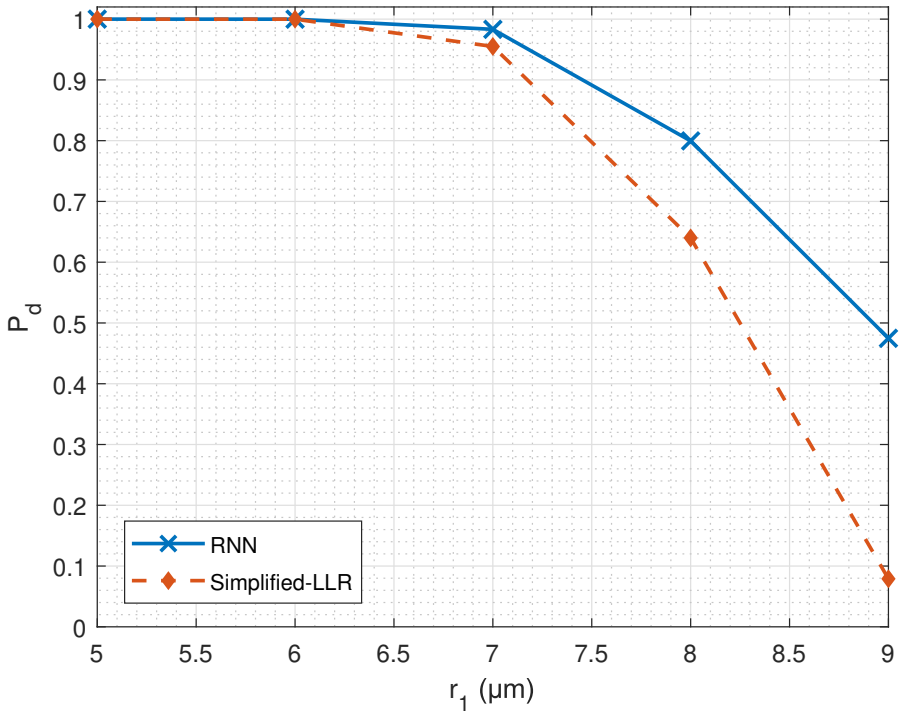


Figure 4.9: The P_d performance of the RNN detector considering different r_1 values for $P_f = 0.05$ ($SNR = 2$).

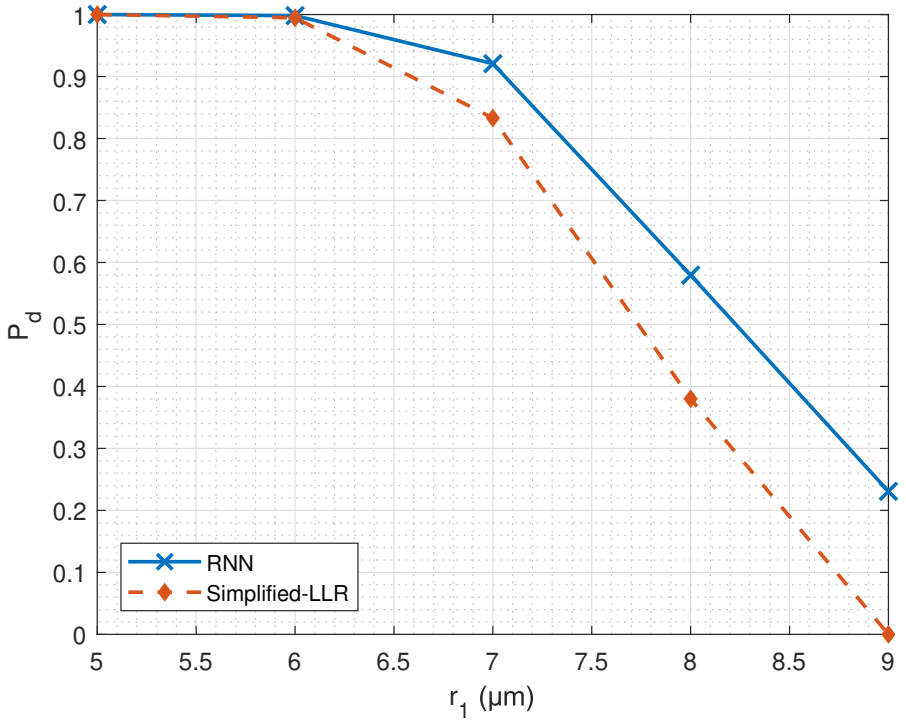


Figure 4.10: The P_d performance of the RNN detector considering different r_1 values for $P_f = 0.01$ ($SNR = 2$).

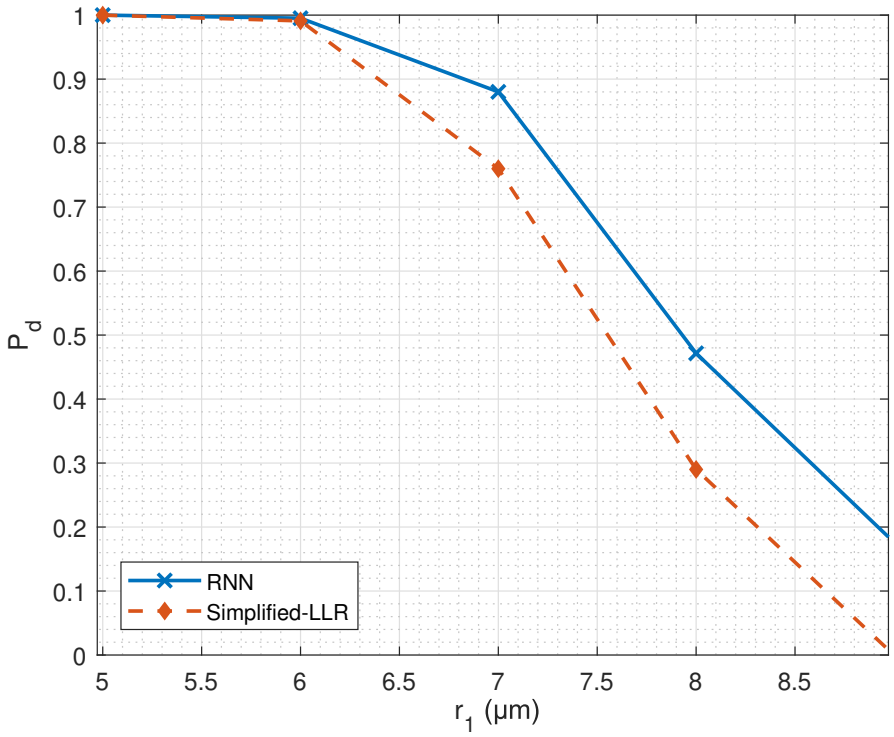


Figure 4.11: The P_d performance of the RNN detector considering different r_1 values for $P_f = 0.005$ ($SNR = 2$).

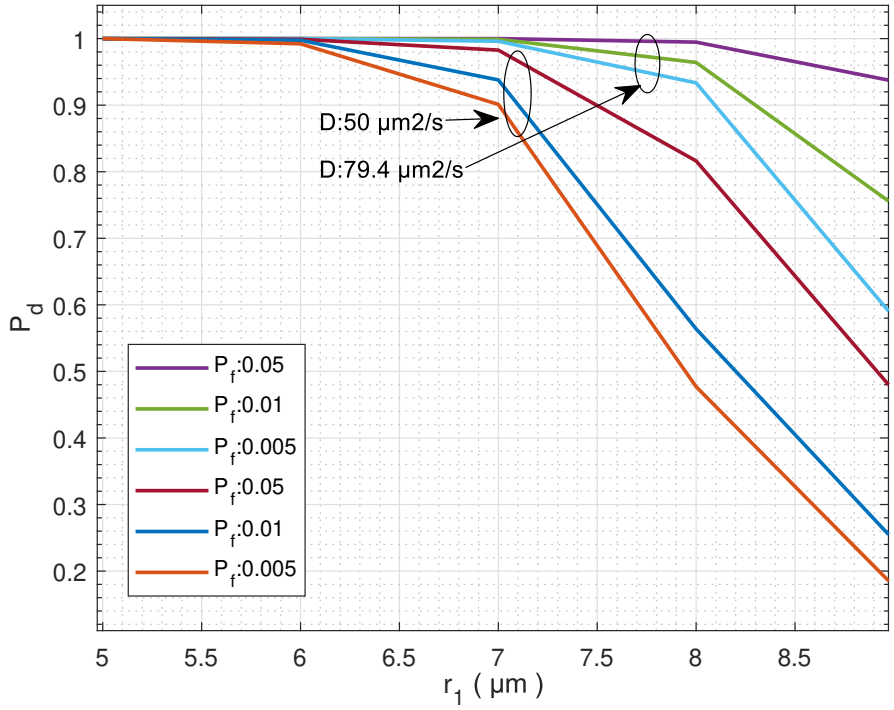


Figure 4.12: RNN performance for different D values ($SNR = 2$).

the LLR, without requiring any prior knowledge, or estimation, of the parameter r_1 for $P_f = 0.05, 0.01, 0.005$.

A similar effect is observed in Fig 4.12, where the effect of another key factor in signal propagation is investigated, which is the diffusion coefficient, *D*. The performance of RNN detector is presented for $D = 50\ \mu\mathrm{m}^2/\mathrm{s}$ and $D = 79.4\ \mu\mathrm{m}^2/\mathrm{s}$ representing the diffusion coefficient of ionic calcium in cytoplasm and the diffusion coefficient of human insulin hormone-like molecules in a blood-like fluid, respectively [99], [100].

The simulation results indicate that the RNN detector performs well and outperforms the LLR detector for different CSI parameters, *SNR*, r_1, *D*. Note that in MC these parameters may change rapidly due to the nature of the environment. Therefore, robustness is a key feature in MC applications besides good detection performance. Training the RNN detector using a data set containing data produced under different channel conditions provides robustness to the detector that is a key requirement in actual practical applications. As expected, the RNN based approach has proven to be more suitable for the proposed task due to the sequential nature of the data encountered in MC and its temporal structure, which the RNN can make use of. In the following section, RNN based approach is investigated further with different scenarios and particle-based simulations.

4.3.3 Further Analysis of Recurrent Neural Networks

In the existing literature on MC based sensor networks [17, 41, 101], the distances between the FC and the NSs are assumed to be identical and constant. While this assumption considerably simplifies the derivation and the analysis of the detectors proposed in these studies, it is rather unrealistic in practice, due to the fact that the sensors and the FC in such a network float freely within a 3D fluid medium. Hence, during the operation of a network in practice, both the sensors and FC may drift randomly in the fluid, either away from- or towards each other, which will result in non-identical sensor-FC distances that also change over time [59]. We investigate the algorithm further in this section to take this effect into account. The distance r_m, between the m'th NS and the surface of FC for $m = 1, ..., M$ is modelled as independent and identically distributed (i.i.d.) Gaussian RVs with mean μ_{r_m} and variance $\sigma^2_{r_m}$, where the particle-based molecular communication simulator AcCoRD [31] has been employed in the signal generation. The nanonetwork considered in this section is shown in Fig 4.13. In each detection cycle, the distance of the m'th NS from the FC, r_m, is determined by drawing a new realisation of the RV $R_m \sim N(\mu_{r_m}, \sigma^2_{r_m})$. Without loss of generality, the Poisson channel model for training and

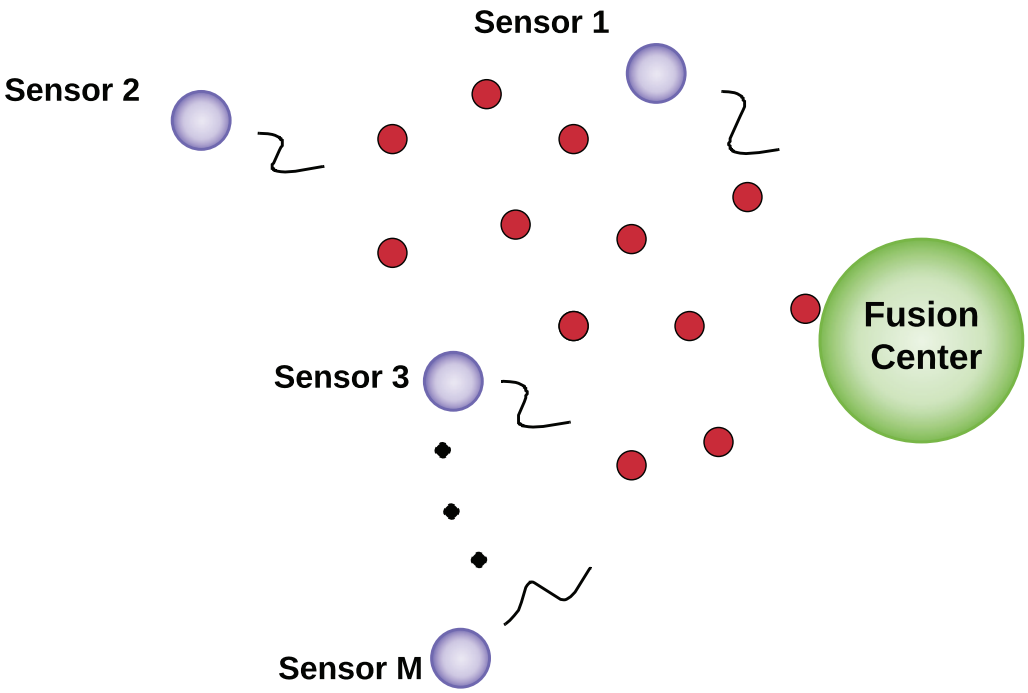


Figure 4.13: Nanonetwork consisting of M nanosensors and a fusion Centre with non-identical sensor-FC distances.

evaluation of the RNN-based algorithm is considered, which is a commonly employed model in diffusive MC [24]. This approach enables a fair comparison between the proposed methods and the LLR algorithm in [17] employing this channel model, which is used as a benchmark. In this section, an RNN with one LSTM layer consisting of 32 neurons and one Dense layer as an output layer is applied. Observations from previous time steps are summarized as the state of RNN, which is represented by a hidden state, h. In the training phase of the algorithm as explained in the section 4.3.2.1 , the data-set is divided into two subsets as 80% for training and 20% for validation. Training data-set is employed to train the proposed RNN detector that estimates the

Table 4.3: List of parameters used in the training and the simulations

Parameter	**Configuration 2**
A (The maximum number of molecules that a sensor can release for each pulse)	500
T (Time slot duration)	70 ms
N (Number of time slots)	6
Number or sensors in the nanonetwork	4
L (Number of quantization levels)	4
μ_{r_m}(mean of the distance between the nanosensors and the receiver)	6,7,8,9,10 μm
$\sigma^2_{r_m}$ (Variance of the distance between the nanosensors and the receiver)	1 μm^2
r_r (The receiver radius)	1 μm
SNR levels	3-20
D (The molecules diffusion coefficient)	50, 79.4 μm^2/s

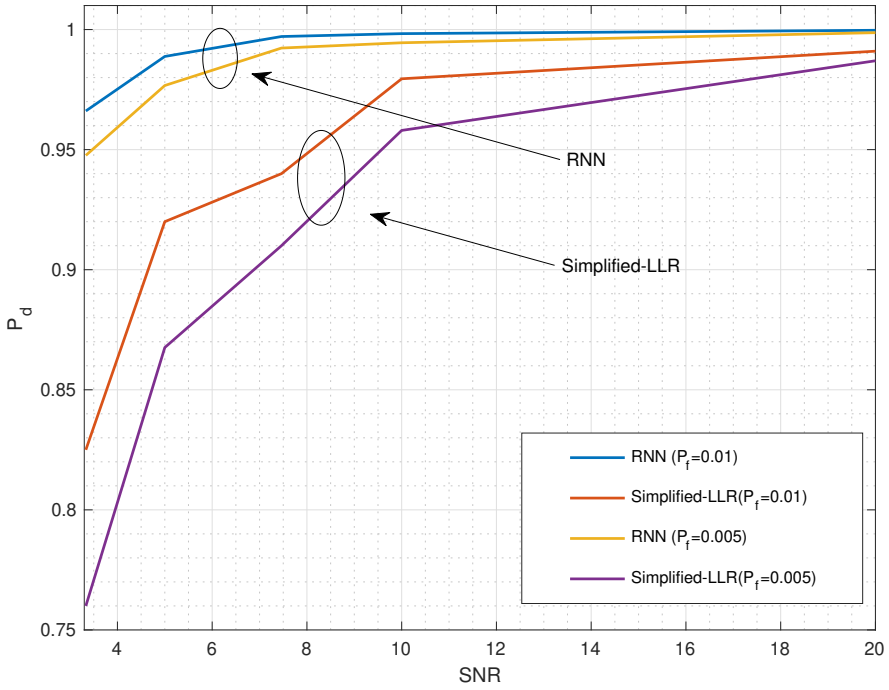


Figure 4.14: The P_d performance of the RNN and simplified-LLR detector considering different SNR levels (μ_{r_m}=8 , $\sigma^2_{r_m}$=1 μm^2).

desired hypothesis for given inputs by finding the optimal set of parameters of the algorithm. The validation data-set is then used to validate the trained algorithm presenting the preliminary performance results. For the performance evaluation of the algorithm, a test data-set is generated, where each of the hypotheses is distributed evenly over the data. The training data-set includes various values for the parameter set $\zeta_{NN} = [J, \mu_{r_m}, \sigma^2_{r_m}, D]$, (see Table 4.3). This makes the resulting DF algorithms robust to possible changes in the system parameters while allowing DF without requiring parameter estimation, provided that the system parameters remain within the boundaries of the training set. In Fig 4.14, the detection performance of the proposed RNN detector and of simplified-LLR detector from [17] for a fixed false alarm rate $P_f = 0.01$ and 0.005 are plotted. In both cases, the proposed detector outperforms the LLR detector, as the LLR employs only an approximation of the likelihood function, whereas the proposed detectors both learn from actual data, which results in a better approximation to the impractical optimal detector. The ROC of the proposed RNN and the LLR detectors for two different *SNR* levels are presented in Fig 4.15. As in the previous case, the RNN detector performs better than LLR detector in each case. Next, the effect of the expected value of the distance between sensors and FC, μ_{r_m}, on the detection performance of the proposed RNN detector is evaluated with a fixed false alarm rate $P_f = 0.01, 0.005$ and two different *SNR* levels to show the P_d performance of detector considering different μ_{r_m} which is the mean of r_m. Figs 4.16 clearly shows that the proposed RNN

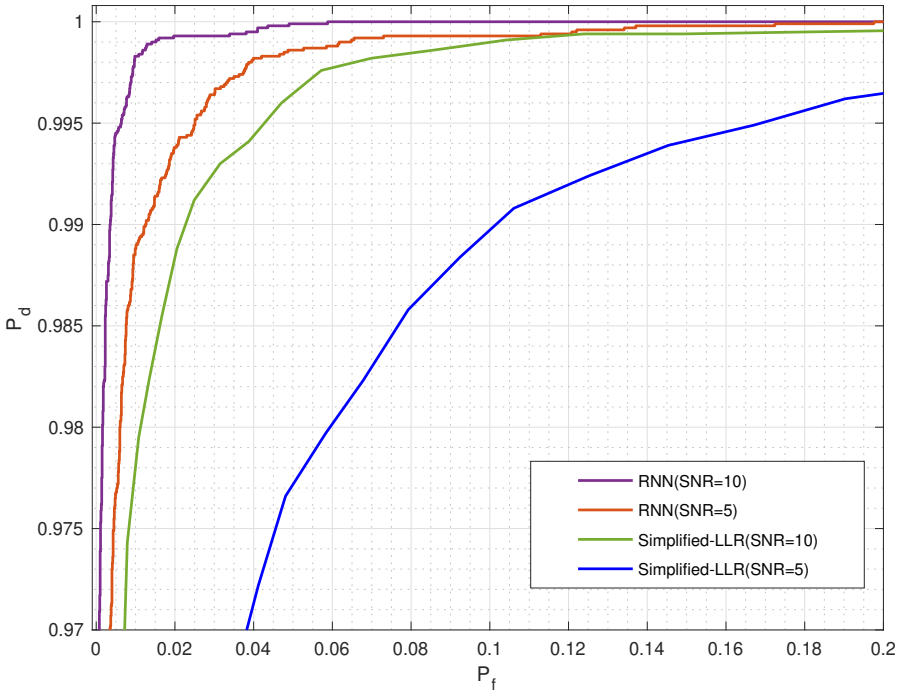
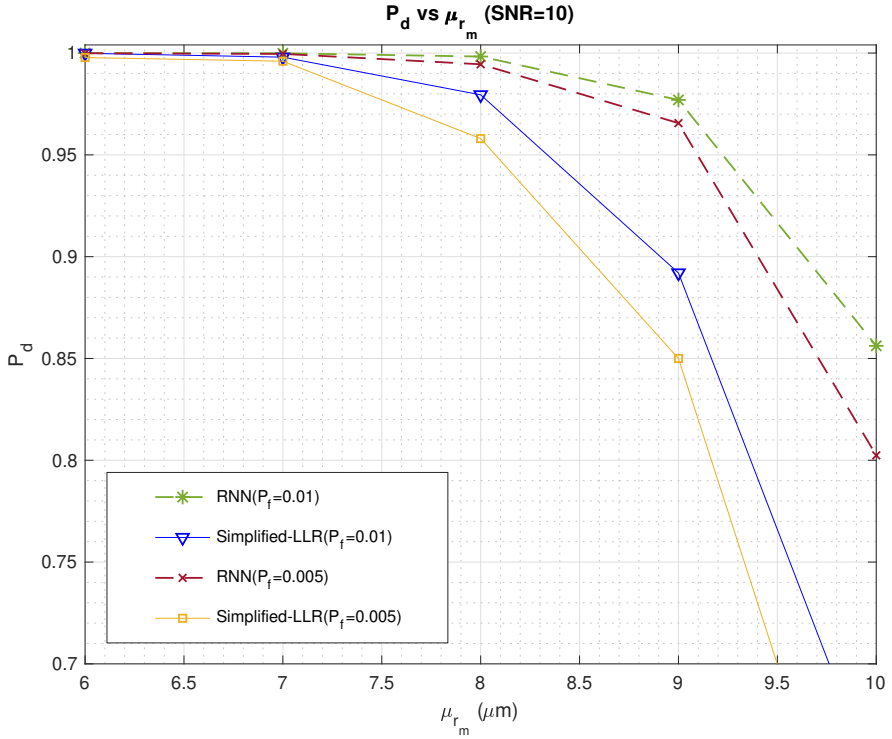


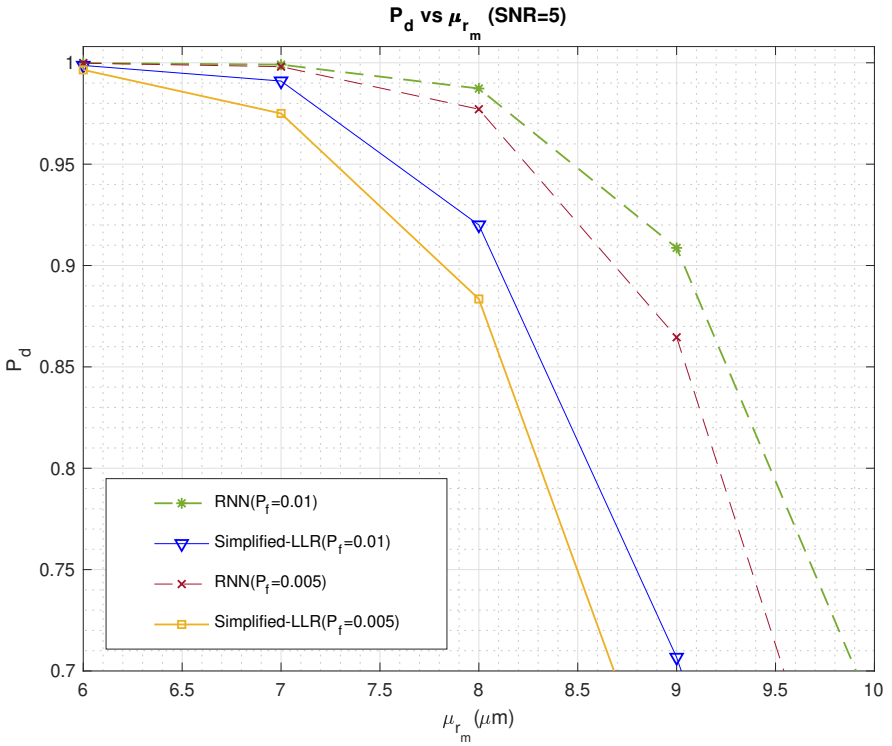
Figure 4.15: ROC curve of RNN and simplified-LLR detector for different *SNR* levels(r_r=1 μm, μ_{r_m}=8 μm, $\sigma^2_{r_m}$=1 μm^2).

method for DF outperforms the LLR for both *SNR* levels. In Fig 4.17, the P_d performance of RNN is presented considering different $\sigma^2_{r_m}$ values, the variance of r_m. It can be seen from these results that RNN performs well without requiring any prior knowledge, or estimation of the parameters μ_{r_m} and $\sigma^2_{r_m}$. A similar effect is observed in Figs 4.18, where the effect of another key factor in signal propagation is analysed, which is the diffusion coefficient, *D*. The performance of the RNN detector is presented for *D*=50 $\mu\text{m}^2/\text{s}$ and *D*=79.4 $\mu\text{m}^2/\text{s}$ for $P_f = 0.01, 0.005$, and 0.001 representing the diffusion coefficient of gamma globulin, one of the antibodies in human blood, and the diffusion coefficient of human insulin hormone-like molecules in a blood-like fluid, respectively [102], [103]. It can be seen from the figures that the proposed detector provides good performance in each case.

The simulation results indicate that the RNN detector performs well and outperforms the LLR detector for different CSI parameters, *SNR*, μ_{r_m}, $\sigma^2_{r_m}$ and *D*. Note that in MC these parameters may change rapidly due to the nature of the environment. Therefore, robustness is a key feature in MC applications besides good detection performance. Clearly, training the RNN detector using data produced under different channel conditions provides robustness to the detector that is a key requirement in actual practical applications.



(a) Probability of detection vs μ_{r_m} ($SNR = 10$).



(b) Probability of detection vs μ_{r_m} ($SNR = 5$).

Figure 4.16: The P_d performance of the RNN detector and LLR detector considering different μ_{r_m} values with $\sigma^2_{r_m}$=1 μm^2.

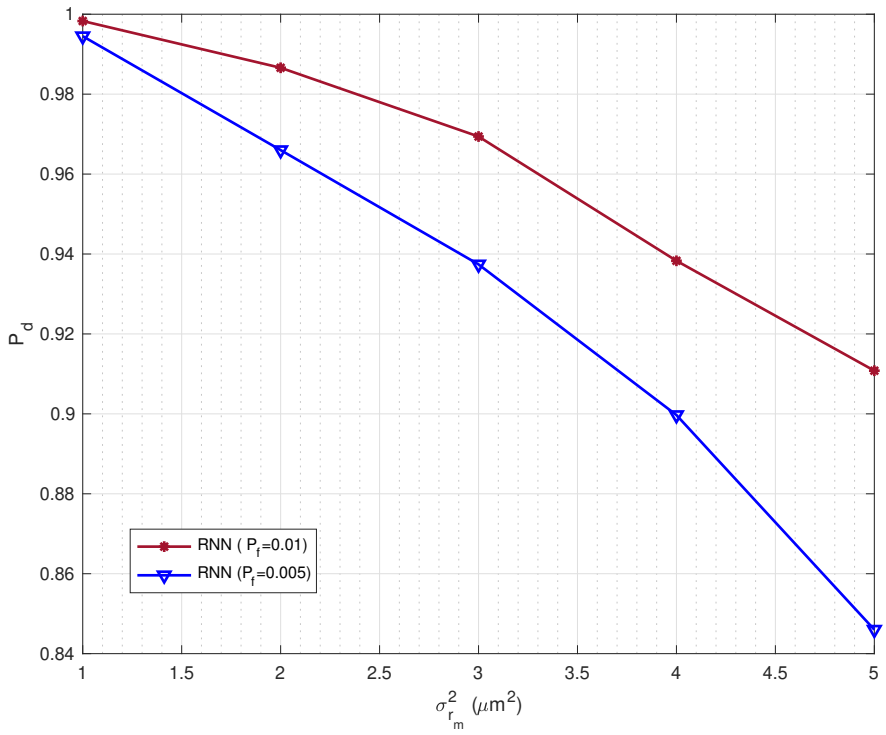
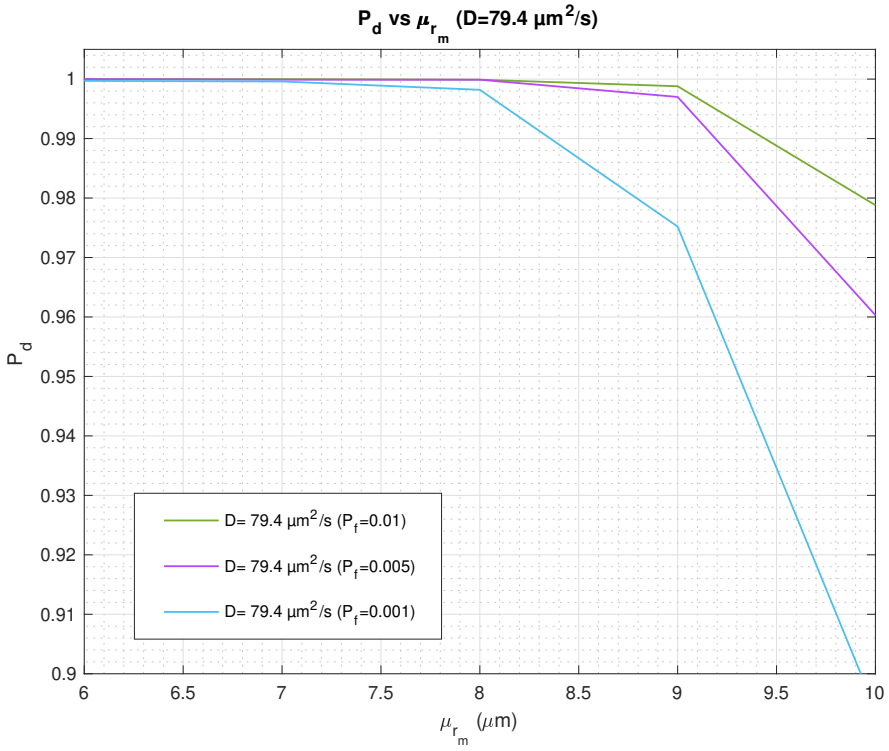


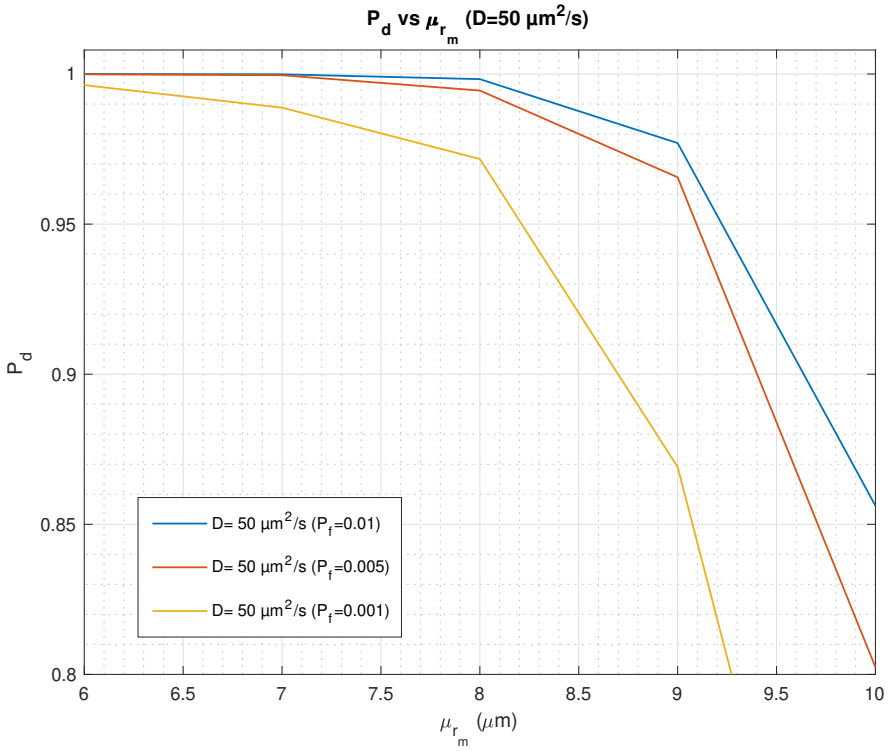
Figure 4.17: The P_d performance of the RNN detector for different P_f considering different $\sigma^2_{r_m}$ values (μ_{r_m}=8 μm, $SNR = 10$).

4.4 Conclusion

In this chapter, a novel ML-based technique, namely DNN to design DF algorithms for detecting abnormalities in a fluid environment using MC is developed. The use of RNN structures for abnormality detection using a network of collaborating NSs is proposed. The proposed RNN based approach is suitable for the proposed task considering the extremely complex and dynamic nature of the data encountered in MC. The RNN structure makes use of information from the history of all past time steps which is more realistic considering the nature of the detection task. The proposed detectors can be trained directly with data obtained from mathematical models, or from in vivo measurements, in vitro experiments, particle-based diffusion simulators, etc., without requiring an analytical statistical description of the channel model. Hence, this approach leads to DF methods that don't require any analytical channel model, instantaneous CSI or knowledge of other system parameters such as the statistical characteristics of the sensors, even the number of sensors. The presented results indicate that the novel ML-based DF approach proposed in this dissertation provides good detection performance along with robustness and flexibility required for practical implementations of such MC based nanoscale sensor networks.



(a) Probability of detection vs μ_{r_m} ($D = 79.4\ \mu\text{m}^2$).



(b) Probability of detection vs μ_{r_m} ($D = 50\ \mu\text{m}^2$).

Figure 4.18: RNN performance for different diffusion coefficient (D) values for different P_f considering different μ_{r_m} values ($\sigma^2_{r_m}$=1 μm^2, $SNR = 10$).

Decision Fusion in Practical Geometries

5

I am one of those who think, like Nobel, that humanity will draw more good than evil from new discoveries.

Marie Curie

With the increasing interest in designing new alternative approaches like MC to conventional (i.e. radio frequency) wireless communication technologies, researchers need to model realistic scenarios where the usual approach of analytical modelling of the channel characteristics hits its limitations. In Chapter 4, a novel machine learning-based technique is developed for designing DF algorithms for identifying abnormalities in a fluid environment via MC. Given the exceedingly complex and dynamic nature of the data encountered in MC, the proposed RNN-based technique is appropriate for the required task as presented in section 4.3.2.2. We consider systems where the analytical channel model exists, and in the formulation of the analytical models for diffusive MC channels proposed in the existing literature, a series of simplifying assumptions is used. In practical scenarios, there will be circumstances when an analytical channel model is not mathematically tractable or is too complex to be useful. In such cases, the use of decision-theory based methods will be impractical. Therefore, in this chapter, we investigate several sensor network architectures and scenarios that are difficult or impossible to represent analytically to explore the ML approaches, namely NNs further. The NNs require information to learn the environment and perform detection task, hence realization of signals under these channel conditions of the scenarios discussed above are acquired. Particle based simulators are preferred in such situations

to obtain the data that represents the scenarios most properly.

5.1 Introduction to Particle Based Diffusion Solvers

With these limitations, simulation tools for diffusive MC have been developed to overcome the issues of the gap between theoretical models and experiments of these models in MC. Wet lab testing is required to validate these models for many of the applications. The simulation data is the best data available as the wet-lab data is not widely available. Also, communication engineers rarely have access to wet labs, and these studies can only be carried out with the help of bioengineers, biologists, and chemical engineers [9]. Diffusion solvers seem to be the best option until experimental studies become widely available. Molecular communication appears to be more amenable to closed-form analysis and optimization than conventional communication; as a result, simulations are a significant tool for replacing physical experiments in order to collect the data needed for training and for assessing the system's performance [1].

Particle based diffusion simulators must be able to track the behaviour of information-carrying particles in a realistic environment. There are different models for molecular behaviour such as *molecular dynamics* models (like used in LAMMPS [104]), *continuum* models (like used in COMSOL Multiphysics [105]). In this thesis, we employ the particle based diffusion simulator AcCoRD, which is a reaction-diffusion solver designed for the study of molecular communications systems [106]. This simulator has been chosen because it is designed as a generic reaction-diffusion solver (i.e. sandbox simulator) and can release molecules according to the modulation of a data sequence. It uses a hybrid of microscopic and mesoscopic simulation models that enables scalability via user control of local accuracy. AcCoRD is developed in C as an open-source command-line tool and includes utilities to process simulation output in MATLAB. For the comparison of other simulation tools with AcCoRD see [31].

The simulation's overall workflow can be divided into three parts: preparing a configuration file to run the AcCoRD, running the program, and post-processing with MATLAB. A configuration file contains information about the geometry of the setting, which is the physical space in which molecules can move or be created, as well as information about the model that defines the region, which can be microscopic or mesoscopic. Each molecule in microscopic regions is observed individually, and its evolution is tracked using a global microscopic time step g_{ts}. Mesoscopic regions keep track of how many of each molecule type is present in each subvolume. Actors are also specified in the configuration file. They act as a simulation's interface, allowing molecules to be input or observed as output. Chemical features like absorption and desorption

are also be included in the configuration file. The AcCoRD simulator is compiled as a single executable file. The executable can be called directly from a command-line interface. All post-processing utilities were developed in MATLAB and simulation output can be loaded with an import function [107].

5.2 Methodology

In this chapter, we use the AcCoRD simulator to simulate the different sensor network structures that are difficult or impossible to be represented analytically. Simulation data is the best accessible data for training and evaluating the machine learning based algorithms at the moment because wet-lab data is not widely available for MC applications. In Chapter 3, we evaluate our SAPRT based sequential DF strategy via both for the Monte Carlo simulations based on the ideal model, and for particle-based simulator AcCoRD. The objective of the setups presented in this chapter is to evaluate the functionality of the entire proposed RNN approach with the only available realistic data produced in particle-based simulation considering scenarios for practical applications. After, the simulation process, MATLAB based post-processing is applied to pass the data to proposed detection algorithms developed with Python [108]. After acquiring the simulation data and preparing it for the training process, a strong optimization technique called grid-search optimization is applied to the NN algorithm firstly. It is most frequently used to fine-tune hyper-parameters in machine learning models [109], [110]. This technique automates the 'trial-and-error' process by allowing to find the optimal neural network hyper-parameters from a list of neural network hyper-parameter alternatives that are given. They can be chosen from all parameters for a given estimator. We can investigate a variety of hyper-parameters in this chapter but we choose three main of them that are generally the most effective ones; the type of optimizer, number of epochs, i.e. the number of full runs over the training data, number of batches, i.e the number of training samples that must be processed before the model's internal parameters are changed.

5.3 Investigated Sensor Network Setups

We focus on two different systems with different configurations defined in detail in the following sections. Namely, we simulate a simplified animal cell environment and a simplified blood vessel environment in AcCoRD. Chemical signals are utilised in nature for inter-cellular and intra-cellular communication at the micro and nanoscales. In the literature, MC is commonly

studied at intra-cellular level both using experiments and simulators [9]. It is crucial to investigate communication mechanisms in the complex environment of cells considering that cells as the fundamental structural and functional units of life. Recently, considerable research efforts ([42], [41], [111], [112]) have been devoted to detecting abnormalities such as tumours, cancer, and so on. Abnormality detection using diffusive MC is studied in a basic blood vessel environment in this chapter with recent developments and the motivation of these studies.

We can investigate a variety of system parameters in this chapter and there are many possible combinations of joint parameter investigation, but we take into consideration the main parameters in the training phase, such as sensor-FC distances and noise level. These system parameters may change over time in the MC environment affecting the detection performance significantly. All of the simulation findings shown in this section were averaged across 10^4 separate simulations for each case unless it is specified otherwise.

5.3.1 Abnormality Detection Within a Cell-like Environment

The processes of MC occur at all levels of biological systems, including the molecule, cell, tissue, and organ levels. MC within a cell (up to the size of a cell about $100\ \mu\mathrm{m}$) called intracellular MC. At the intracellular level, several sub-cellular bio-nanomachines within a cell communicate to sustain the life of the cell [1]. The interaction may be directly through physical contact or indirectly with diffusive molecules. There's no denying that the live cell has a complicated structure, and that this structure is at the root of the cell's complex operations. Considering the limitations in the simulations of MC, we are only able to simulate a simplified version of a cell environment. A cell is made up of three parts: the cell membrane, the nucleus, and the cytoplasm, which sits between the two. The cytoplasm contains elaborate arrangements of many tiny yet unique structures known as organelles. We need to consider the diffusion coefficient of these structures in cytoplasm. In our simulations, we only define obstacles which represent other agents in the cell environment such as small organelles like lysosomes as we are limited defining all kind of structures in the cell environment in the MC simulation environments. A cell membrane surrounds every cell in the body. The cell membrane is responsible for separating external and intracellular substances. We define a bounded environment as the cell environment as realistic diffusive environments are usually constrained. Researchers would be able to include the nucleus and other agents in the cell environments in light of new developments of MC simulators.

In System 1, we investigate abnormality detection using MC within a cell-like environment. A simple representation of the cell environment and the AcCoRD setup is presented in Fig 5.1

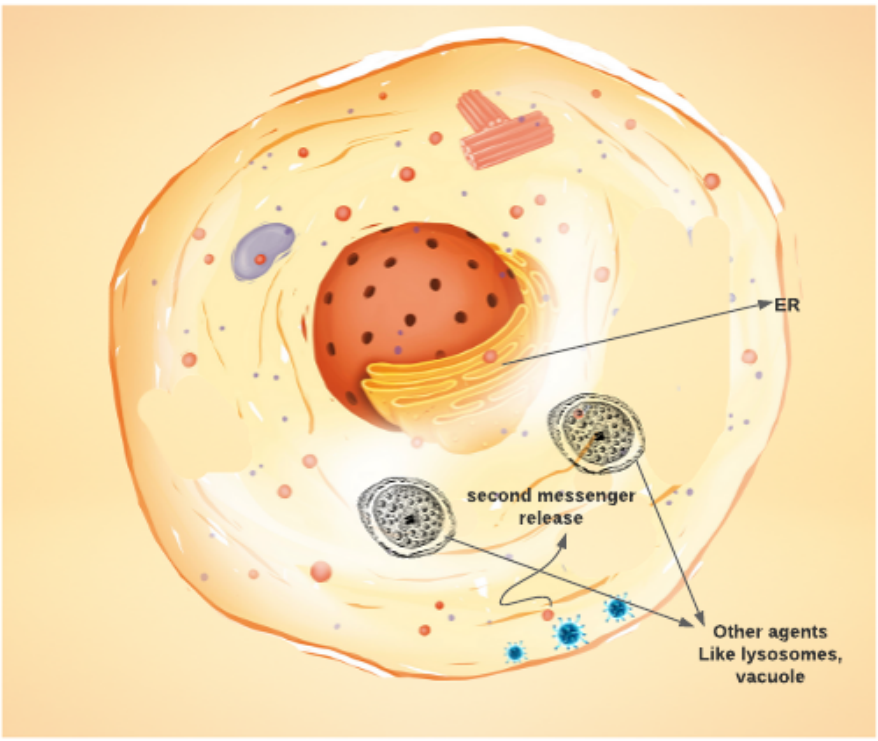


Figure 5.1: Simulation setup mimicking a basic animal cell environment.

and Fig 5.2 respectively.

When selecting parameters for numerical analysis and simulation, it is critical to have a feeling of realistic parameter values. A typical animal cell has a diameter of $5-20\ \mu\mathrm{m}$ [55], [113]. The receiver is chosen as the ER (endoplasmic reticulum) in the cell with the size of $2-3\ \mu\mathrm{m}$. The ER is considered as receiver because The ER is the cell's biggest organelle and is involved in protein synthesis and transport, protein folding, lipid and steroid production, glucose metabolism, and calcium storage [114]. The obstacles represent other agents in the cell environment such as small organelles like lysosomes with the size of $0.5\ \mu\mathrm{m}$, A is the maximum number of molecules for each pulse, is chosen according to the number of small messenger molecules, such as AMP, diffuse to act on signalling process. These molecules called second messengers, such as molecules cyclic AMP, cyclic GMP, inositol triphosphate, diacylglycerol, and calcium ([55], Ch. 16). The diffusion coefficient of the second messengers in the cytosol is in the range of $100-500\ \mu\mathrm{m}^2/\mathrm{s}$ [115]. Diffusion timestep is chosen as $5\ \mu\mathrm{m}$ in the signal generation using particle-based simulator AcCoRD. The symbol duration values are set such that they are sufficiently longer than the period when the proportion of absorbed molecules reaches its highest value to have fewer leftover molecules causing ISI. The simulation parameters are summarized in Table 5.1.

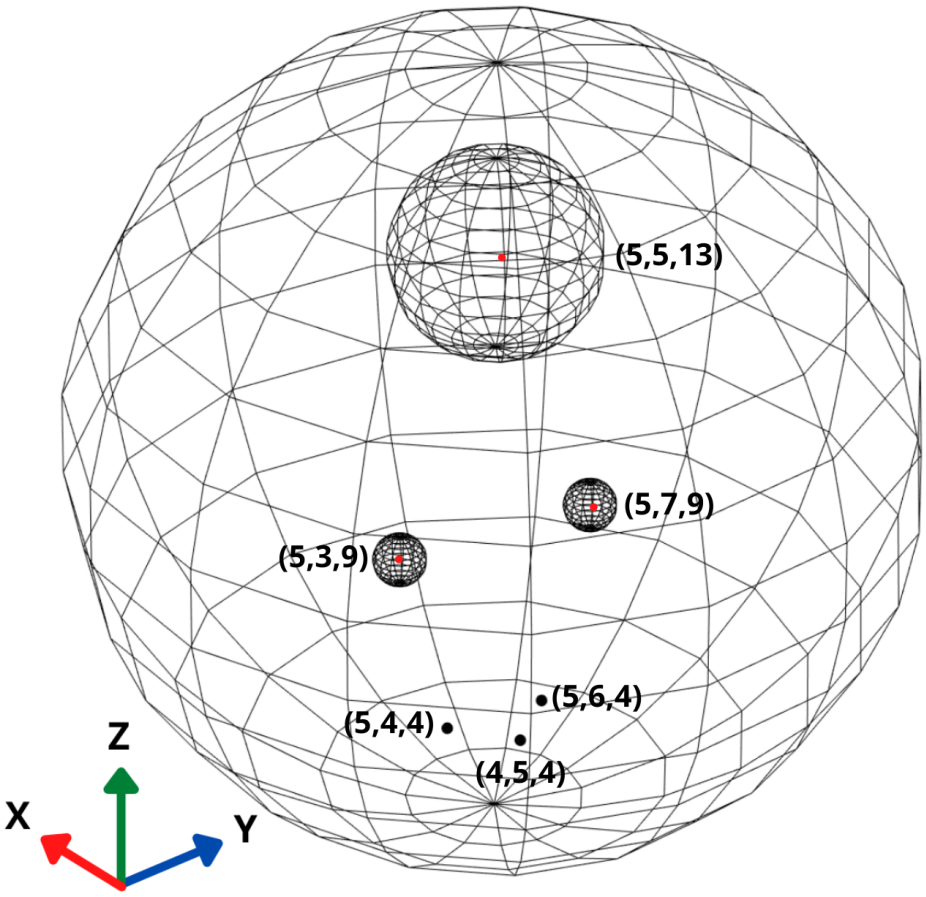


Figure 5.2: Representation of a basic animal cell in AcCoRD.

We consider two different scenarios to simulate detection of an abnormality via MC in the cell environment. In the first configuration, the distances between sensors and FC are assumed to be equal represented with $r_1 = 6, 7, 8, 9, 10, 11$ μm, where in the second configuration, sensors are positioned in a line along the y-axis 1 μm distant each other. There are three sensors in

Table 5.1: List of parameters used in the simulations of System-1

Parameter	Configuration-1	Configuration-2
A (The maximum number of molecules that a sensor can release for each pulse)	400	400
g_{ts} (Global microscopic time step)	5 μs	5 μs
T (Time slot duration)	1 ms	2 ms
N (Number of time slots)	10	50
M (Number or nano-sensors)	3	3
L (Number of quantization levels)	4	4
Sensor-FC distances	equal	unequal
r_c (Radius of cell)	8 μm	7 μm
r_1 (Sensor-FC distance)	6:1:10 μm	different for each sensor
r_2 (FC radius)	2 μm	1 μm
r_o (Obstacles radius)	0.5 μm	0.5 μm
D (Diffusion coefficient)	500 μm^2/s	100 μm^2/s

the configuration, and the distance between FC and the middle sensor is $r_1 = 6, 7, 8, 9, 10, 11$ μm, while the first one 1 μm distant in the -y-axis from the middle sensor, and the third one is 1 μm distant in the y-axis from the middle sensor in all iterations as shown in Fig 5.3. The placements of receiver, obstacles and transmitter can be seen from the figures of representation of cells. A region's position is determined by its anchor coordinate, which is the centre of a sphere or the bottom corner $[x, y, z]$ coordinates of a box in AcCoRD. The coordinate of the centre of cells in both configurations is $[5, 5, 10]$. The diffusion coefficient is chosen higher than the second configuration meaning that molecules diffuse faster in the cell. It would be important to evaluate the algorithm with different diffusion coefficients to consider different factors affecting the diffusion coefficient such as the molecule size and shape, temperature and viscosity of the medium. To show that our algorithm performs well even in the cases that some parameters are unknown or including them in the analytical channel model almost is impossible, we have chosen the sensor-FC distances different from each other in the second scenario.

The process of acquiring the training and test data is as follows: after fixing the sensors and FC locations, measurements are conducted in AcCoRD, which means that the number of absorbed molecules in the FC are recorded for each time step. For the hypothesis $\mathcal{H}_1$ denoting the presence of the abnormality, we add noise J, which is the expected value of the received Poisson distributed additive noise molecule. For the hypothesis $\mathcal{H}_0$, which indicates the absence of the abnormality, only noise J is applied. Each of the assumptions is spread uniformly over the data, implying that we obtain the same amount of independent simulations for each hypotheses.

A NN model is typically trained with the stochastic gradient descent optimization technique, and weights are updated with the backpropagation of the error algorithm. The gradient descent algorithm attempts to adjust the weights thereby the next iteration decreases the error, implying that the optimization process is reducing the error gradient. A loss function must be chosen to calculate the model's error throughout the optimization phase. By maximizing a likelihood function obtained from the training data, a maximum likelihood framework is used to discover the most fitting values for the parameters. In classification problems, we can represent the task as predicting the likelihood of belonging to each class. The accuracy and loss functions are used to consider all optimization process factors such as overfitting, underfitting, and convergence helping to assess the learning process. We train the RNN algorithm with the training dataset that contains different parameters to demonstrate that our RNN algorithm leads to more robust and flexible detection requiring no knowledge of the channel. A RNN with one LSTM layer consisting of 16 neurons and one dense layer as an output layer is used in this algorithm.

The dataset is separated into two subsets during the training phase of the algorithm: 80% for

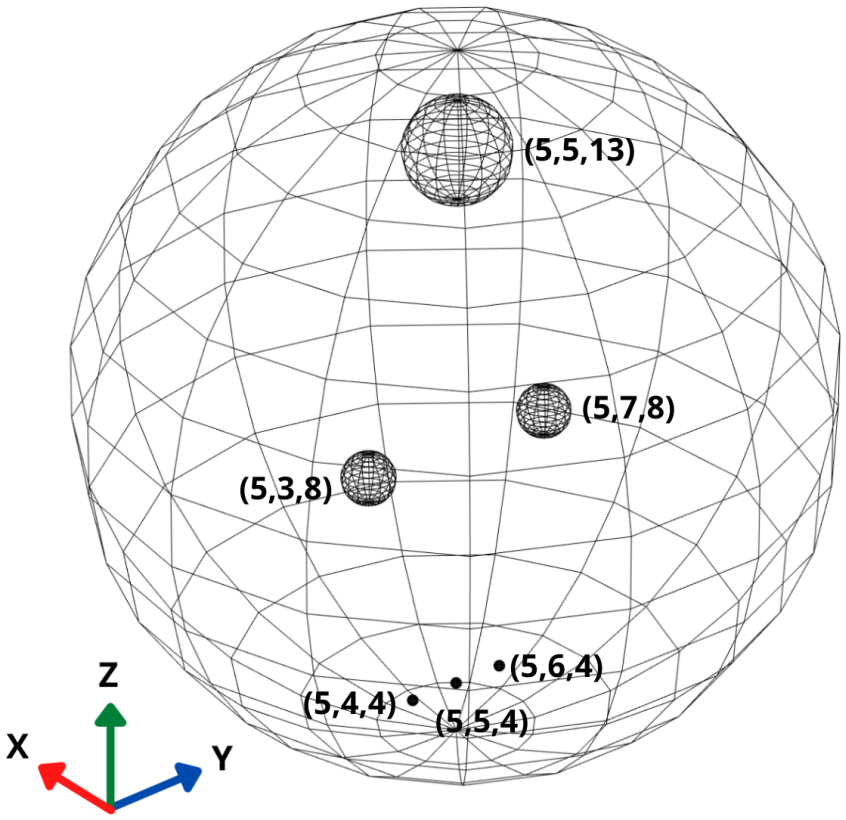
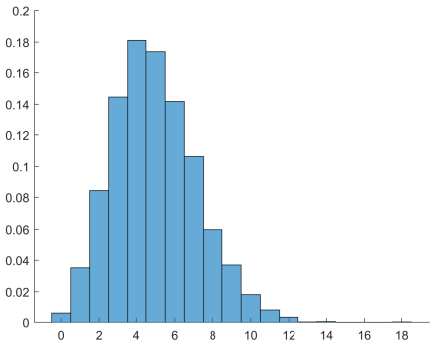


Figure 5.3: Second AcCoRD setup to simulate MC in the cell environment.

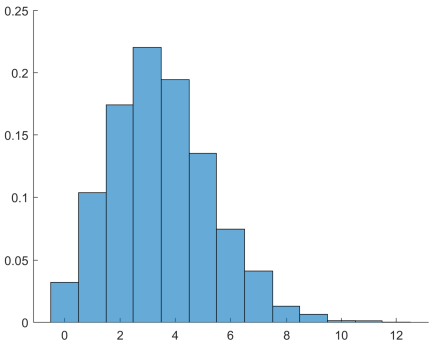
training and 20% for validation. The training data-set is used to train the proposed RNN detector, which estimates the desired hypothesis for given inputs by finding the best set of algorithm parameters. The validation dataset is then utilized to test the trained algorithm. To assess the algorithm's performance, we create a test dataset in which each of the assumptions is evenly distributed over the data. Also, it is crucial to evaluate the detector with the test dataset, because the validation dataset is involved in the tuning process as explained in detail in section 4.2.3.

The performance evaluation of the RNN-based detector is presented according to classification metrics explained in section 4.2.3.1. We first apply the grid-search optimization technique to find the best combination of hyper-parameters. The best values to get the best results is highlighted in bold in Table 5.2 for each setup. After finding the best parameter values with the grid-search technique, we continue to evaluate the RNN algorithm by plotting the confusion matrix and ROC curve using different threshold values.

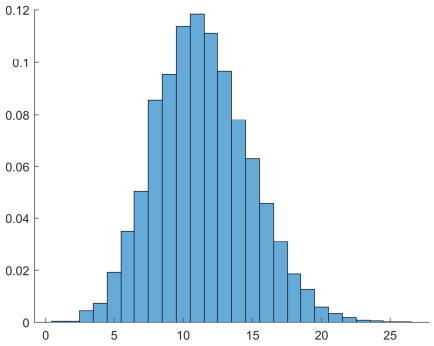
First training dataset consists signals for $r_1 = 6, 7, 8, 9, 10\ \mu\text{m}$ for $J = 5$. Other parameters are chosen as in Table 5.1 from the configuration-1. The histograms of the number of absorbed molecules at the FC acquired in the final time slot is presented in Fig 5.4. These histogram graphs



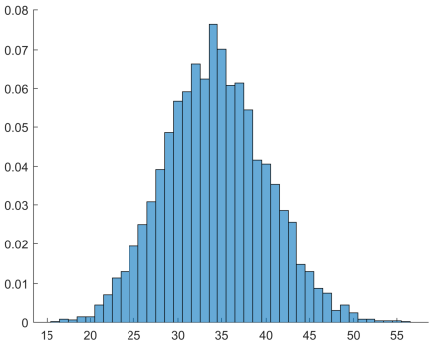
(a) Histogram for $J = 5$



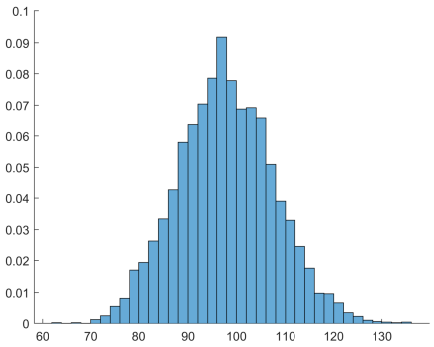
(b) Histogram for $r_1 = 10\ \mu$m



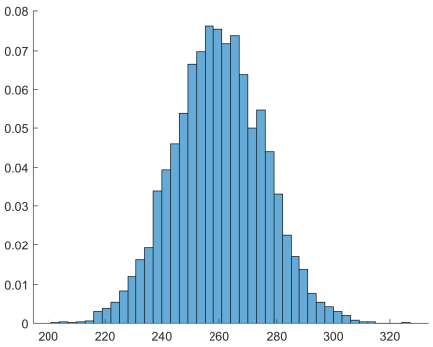
(c) Histogram for $r_1 = 9\ \mu$m



(d) Histogram for $r_1 = 8\ \mu$m



(e) Histogram for $r_1 = 7\ \mu$m



(f) Histogram for $r_1 = 6\ \mu$m

Figure 5.4: Histograms of number of absorbed molecules for system-1, configuration 1.

are important to have a grasp of signal to noise ratio. The number of molecules observed at the FC is displayed for each scenario.

Overall performance is evaluated with the test data, which is randomly chosen from the general dataset generated with the particle-based simulator AcCoRD considering the combinations of five different sensor-FC distances. The results representing the overall performance of the algorithm trained with all sensor-FC distances are illustrated in Figs 5.5, 5.6. The report shows that

RNN algorithm performs well for changing sensor-FC distances r_1 with the $TP = 0.872$ and $TN = 0.977$. The classification report is shown in Table 5.3. Precision, recall and f1-score are utilized generally in imbalanced datasets because a 99% percent accuracy can be meaningless in such dataset. In this thesis, we try to balance datasets for each hypothesis to be able to compare our results with existing studies. However, we believe that it would provide a useful insight to show the classification matrix as there would be cases where unbalanced dataset may be subject to research.

We evaluate the algorithm further for each sensor-FC distance r_1 by assessing the algorithm with the unseen data. The previously trained algorithm performs well for $r_1 = 6, 7, 8, 9, 10$ μm for $J = 5$ considering various P_f . P_d for each sensor-FC distance considering $P_f = 0.05, 0.01$ is illustrated in Fig 5.7.

It is also important to evaluate the RNN-based detector for changing noise levels. In the following results, it is shown that our proposed ML-based algorithm can adapt to changing channel conditions. Second training dataset consists signals from different *SNR* (A/J) levels (*SNR*$= 8, 10, 16, 25, 40$) for $r_1 = 8$ μm . Other parameters are chosen as in Table 5.1 from the configuration-1. The results representing the overall performance of the algorithm trained with all noise levels are illustrated in Figs 5.8, 5.9 for this case. Overall performance is calculated with the test data which is randomly chosen from the general dataset including signals from all different *SNR* levels. The classification report is demonstrated in Table 5.4. It can be seen from the results that the performance of the RNN algorithm is high with the $TP = 0.995$ and $TN = 0.996$.

We investigate the algorithm further for each *SNR* level by evaluating the algorithm with the unseen data (each validation data has one noise level). The previously trained algorithm performs

Table 5.2: List of parameters used in grid-search optimization algorithm

	Optimizer	Number of epochs	Number of batches
Configuration-1	**RMSprop**, Adam	50, **100**, 150	5, 10, **20**
Configuration-2	RMSprop, **Adam**	**50**, 100, 150	5, 10, **20**

Table 5.3: Classification report System-1, Configuration-1

Hypothesis	Precision	Recall	f1-score
$\mathcal{H}_0$	0.8917	0.9625	0.9257
$\mathcal{H}_1$	0.9600	0.8849	0.9209

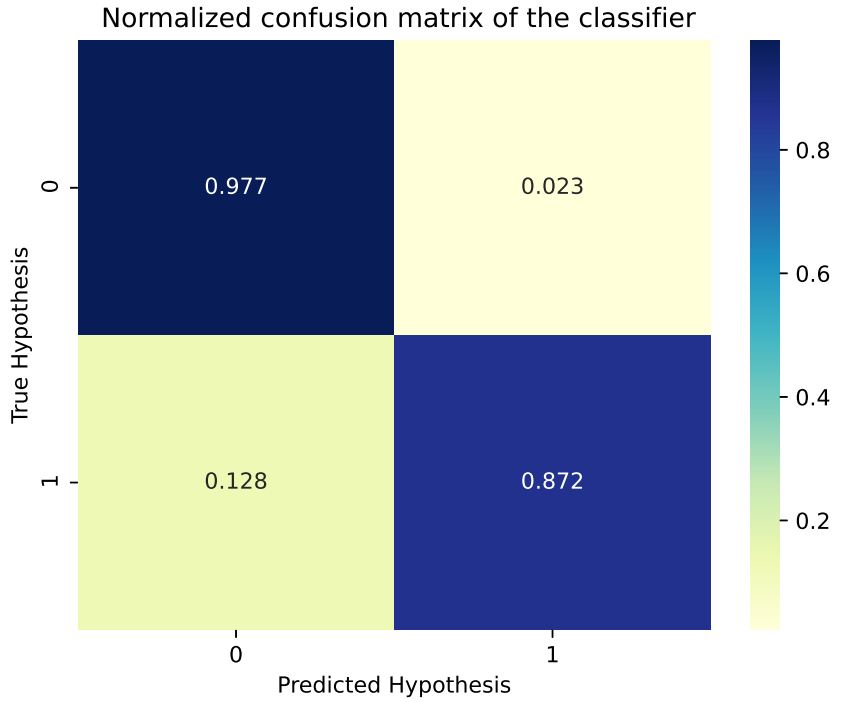


Figure 5.5: Normalized confusion matrix of the proposed RNN algorithm trained with the dataset consisting signals for $r_1 = 6, 7, 8, 9, 10$ μm, with $M = 3$, $T = 1$ ms, $J = 5$, and $N = 10$.

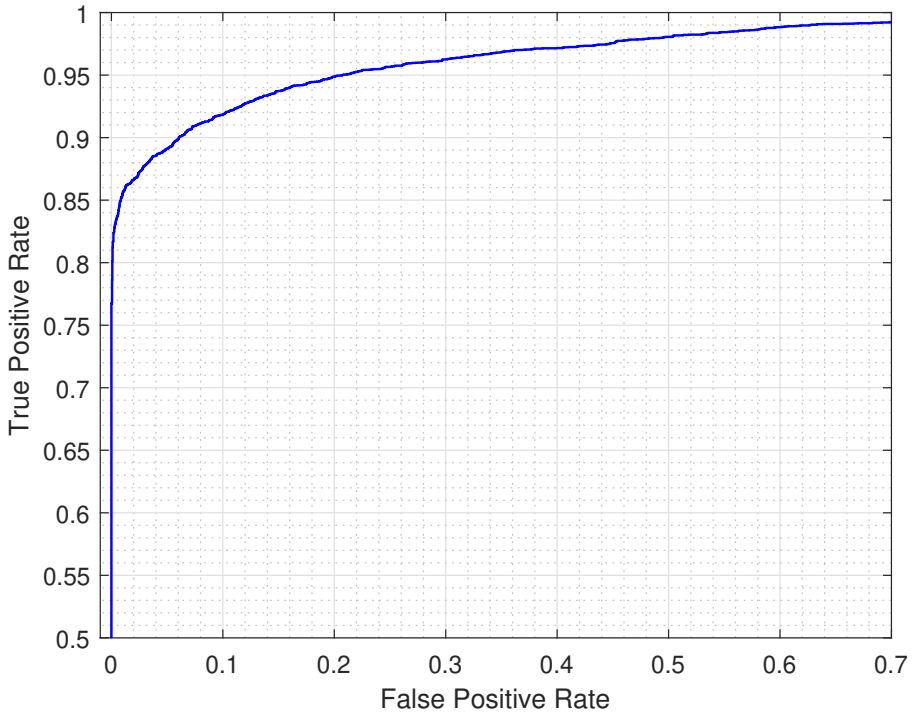


Figure 5.6: ROC Curve of the proposed RNN algorithm trained with the dataset consisting signals for $r_1 = 6, 7, 8, 9, 10$ μm, with $M = 3$, $T = 1$ ms, $J = 5$, and $N = 10$.

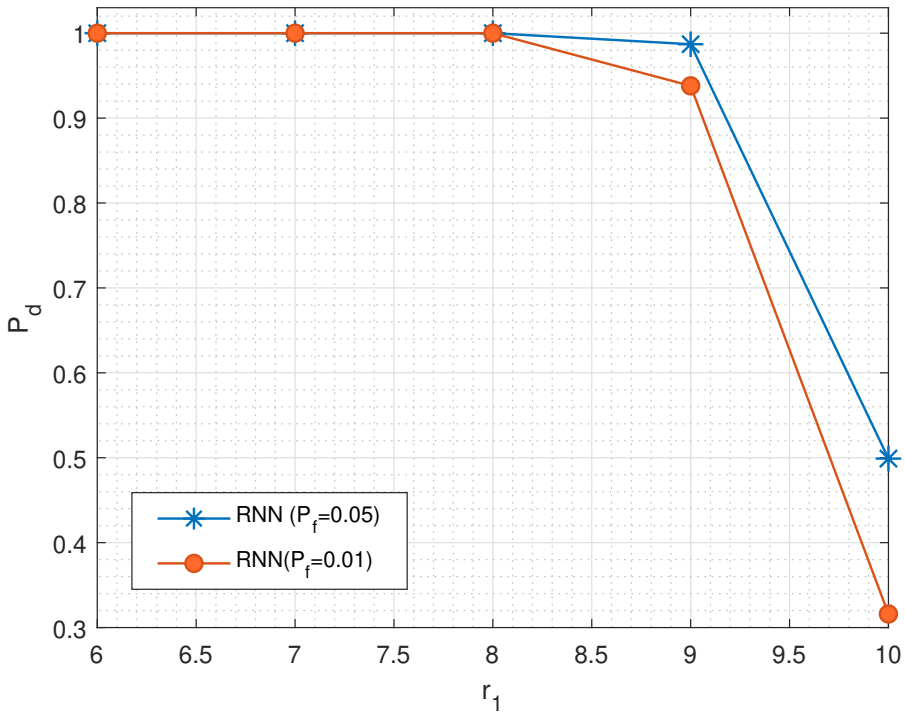


Figure 5.7: P_d of the proposed RNN-based detector for $M = 3$, $T = 1$ ms, $J = 5$, $N = 10$ considering different sensor-FC distances for $P_f = 0.05, 0.01$.

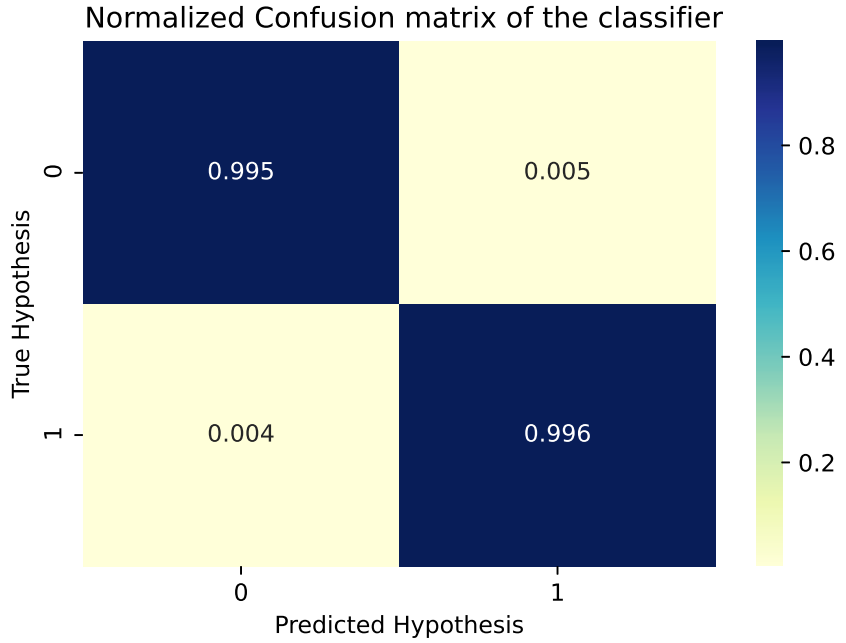


Figure 5.8: Normalized confusion matrix of the proposed RNN algorithm trained with the dataset consisting signals with different *SNR*$= 8, 10, 16, 25, 40$ for $r_1 = 8$ μm, $M = 3$, $T = 1$ ms, and $N = 10$.

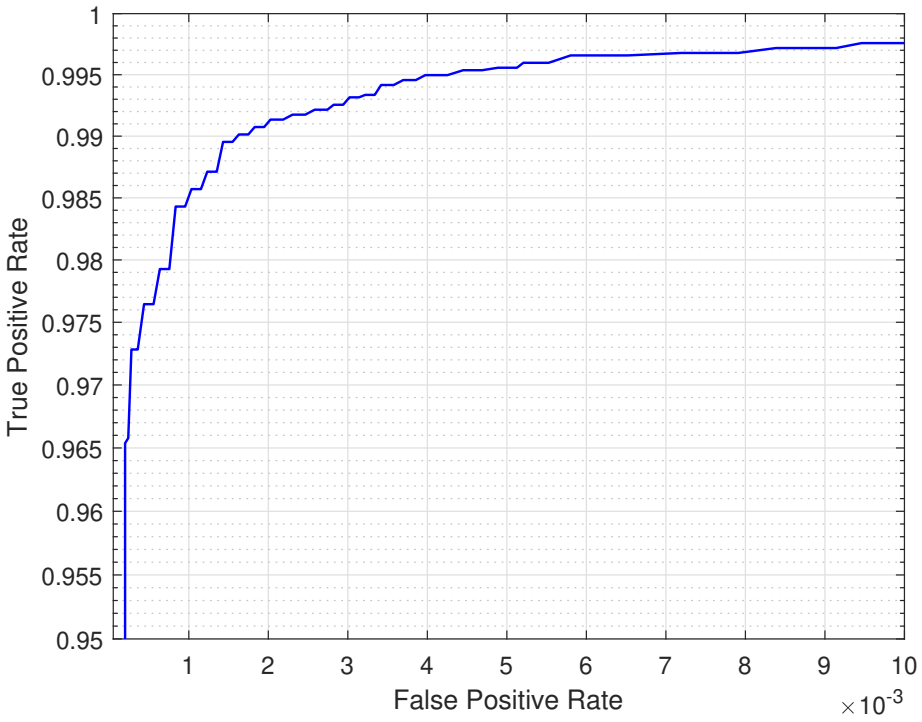


Figure 5.9: ROC Curve of the proposed RNN-based detector trained with the dataset consisting signals with different *SNR*$= 8, 10, 16, 25, 40$ with $r_1 = 8$ μm, $M = 3$, $T = 1$ ms, and $N = 10$.

well for different *SNR* levels considering various P_f . P_d for different *SNR* levels considering $P_f = 0.05, 0.01, 0.005, 0.001$ is illustrated in Fig 5.10.

In the following, we present the results of the second configuration of the first system shown in Table 5.1. In this scenario, we average across 10^3 separate simulations for each of the hypotheses. Please note that it is one-tenth of the first scenario because it is critical to execute a reliable DF using as few receive samples as achievable, especially for the applications where the power resources are limited while training process of the NN algorithms.

We proceed to analyze the RNN method by visualizing the confusion matrix and ROC curve. The model is learned on the dataset acquired from the AcCoRD simulator for different range of sensor-FC distances ranging from $r_1 = 6, 7, 8, 9, 10, 11$ μm for the middle sensor while the first sensor is 1 μm distant in the -y-axis from the middle sensor, and the third one is 1 μm distant

Table 5.4: Classification report System-1, Configuration-1 considering different SNR levels

Hypothesis	Precision	Recall	f1-score
$\mathcal{H}_0$	0.9920	0.9982	0.9951
$\mathcal{H}_1$	0.9982	0.9920	0.9951

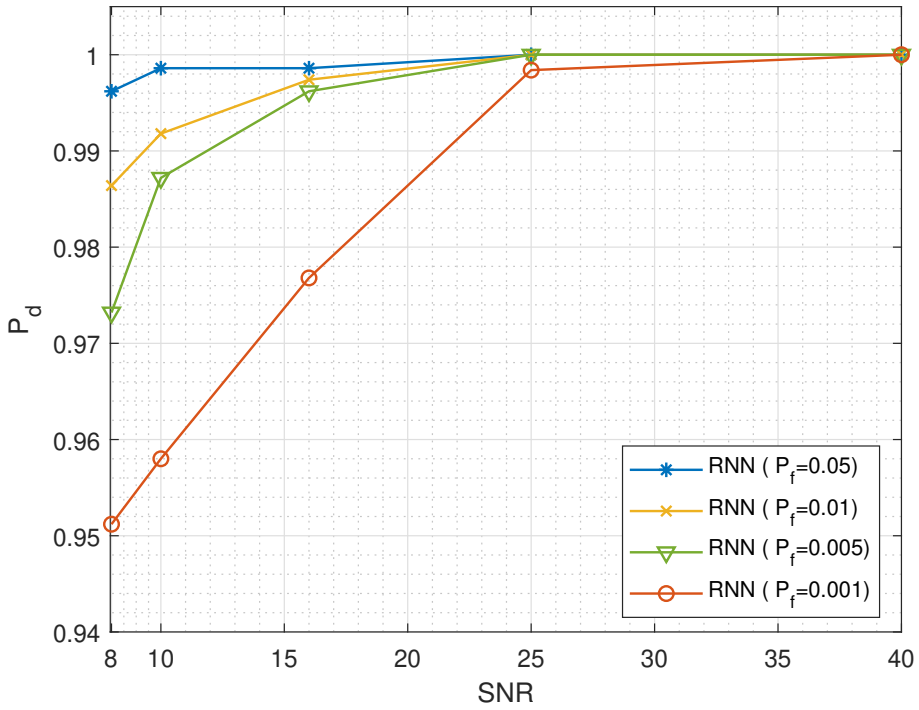
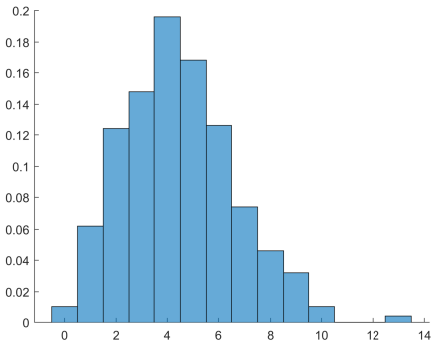


Figure 5.10: P_d of the proposed RNN-based detector for different *SNR* levels with $M = 3$, $T = 1$ ms, $J = 5$, and $N = 10$ considering $P_f = 0.05, 0.01, 0.005, 0.001$.

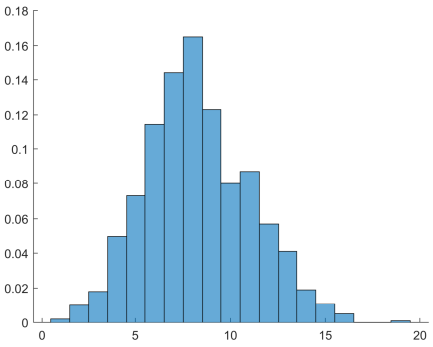
in the y-axis in all iterations as shown in Fig 5.3. The training dataset consists of signals from different sensor-FC distances for $J = 5$. Other parameters are chosen as in Table 5.1 from the configuration-2. The histogram of the number of absorbed molecules at the FC acquired in the final time slot is presented in Fig 5.11. We apply the grid-search optimization technique to find the best combination of hyper-parameters. The best values to get the best results is highlighted in bold in Table 5.2 for the second configuration of System-1.

The results representing the overall performance of the algorithm trained with different range of sensor-FC distances are demonstrated in Figs 5.12, 5.13. Overall performance is calculated with the test data which is randomly chosen from the general dataset including signals from all sensor-FC distances. The classification report of the RNN-based detector is given in Table 5.5. It can be seen from the results that the RNN-based detector performs well with the $TP = 0.989$ and $TN = 0.949$.

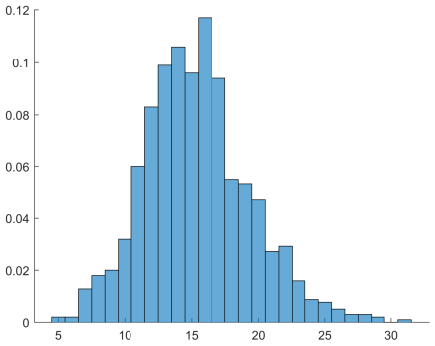
By assessing the method using unseen data, we examine the RNN detector further for different *SNR* levels (each validation data has one noise level). We train the algorithm with the training data contains signals for different *SNR* levels for $r_1 = 10$ μm for the middle sensor, while the first one 1 μm distant in the -y-axis from the middle sensor, and the third one is 1 μm distant in the y-axis in all iterations. For varied *SNR* levels considering $Pf = 0.01, 0.05$, the previously



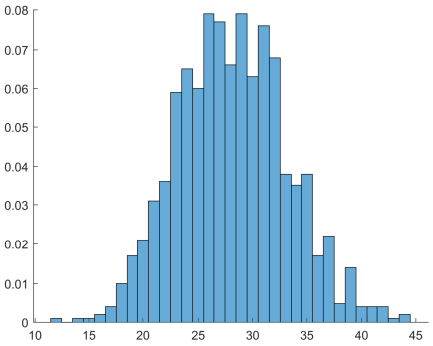
(a) Histogram for $r_1 = 11\ \mu$m



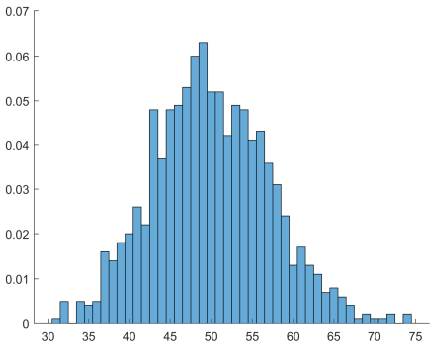
(b) Histogram for $r_1 = 10\ \mu$m



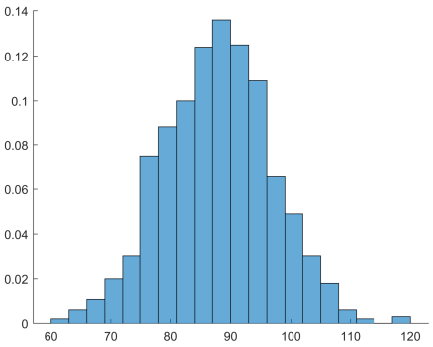
(c) Histogram for $r_1 = 9\ \mu$m



(d) Histogram for $r_1 = 8\ \mu$m



(e) Histogram for $r_1 = 7\ \mu$m



(f) Histogram for $r_1 = 6\ \mu$m

Figure 5.11: Histograms of number of absorbed molecules for System-1, Configuration 2.

trained algorithm functions well as can be seen in Fig 5.14.

The scenarios we investigate in this chapter can not be described using analytical models. However, we compare our RNN based detector with the simplified-LLR detector proposed in the paper [17] with both Monte Carlo simulations using the ideal signal model and particle-based simulations in previous chapters. As we stated in previous chapters, a series of simplifying assumptions is employed in the derivation of analytical models for diffusive MC channels suggested

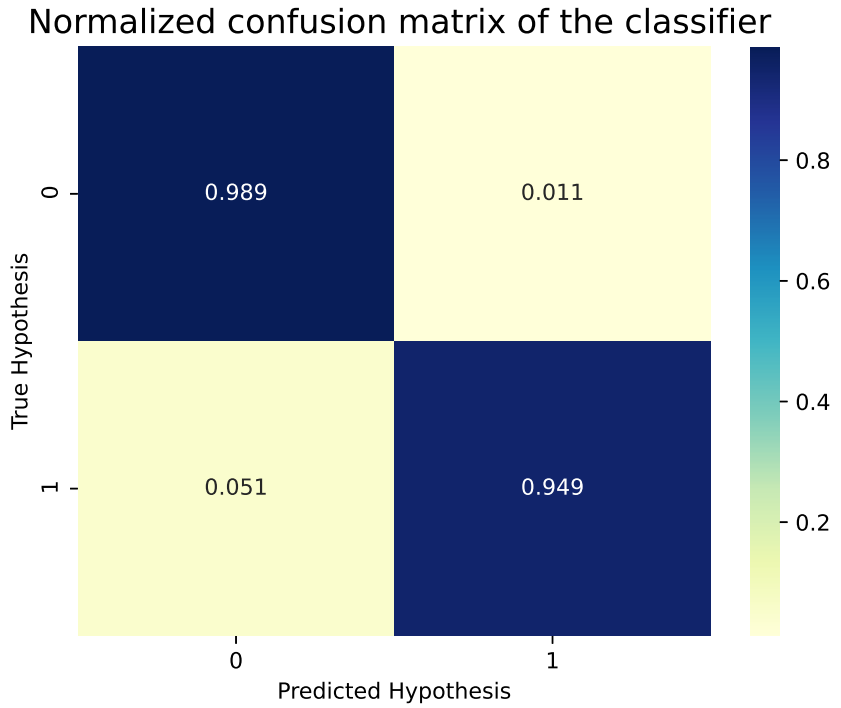


Figure 5.12: Normalized confusion matrix of the proposed RNN algorithm trained with the dataset consisting signals with different range of sensor-FC distances $r_1 = 6, 7, 8, 9, 10, 11$ μm for the middle sensor, with $M = 3$, $T = 2$ ms, $J = 5$, and $N = 50$.

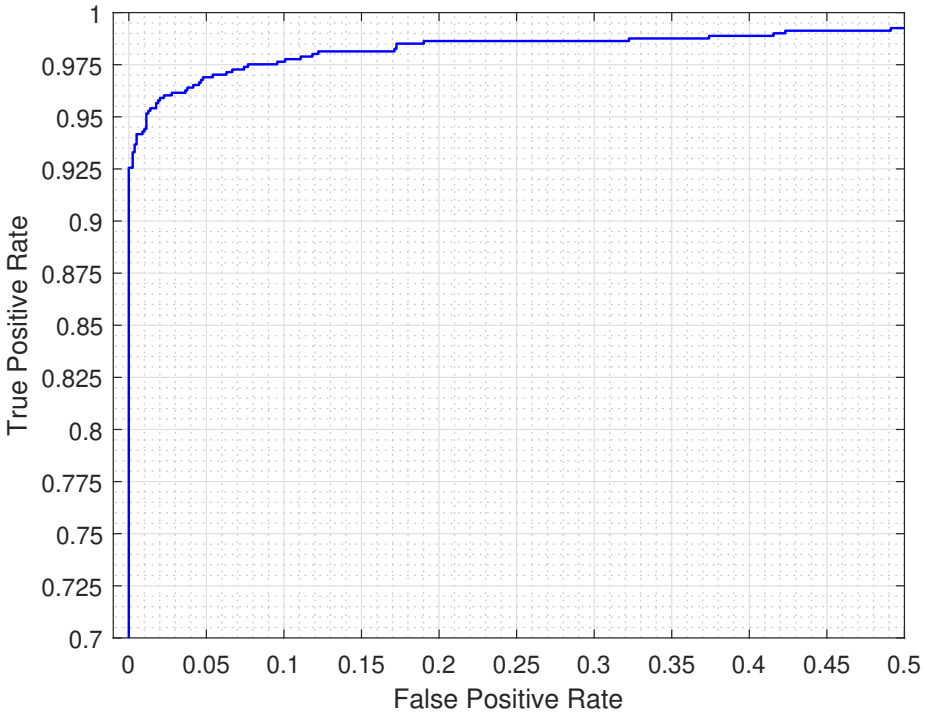


Figure 5.13: ROC Curve of the proposed RNN-based detector trained with the dataset consisting signals with different range of sensor-FC distances $r_1 = 6, 7, 8, 9, 10, 11$ μm for the middle sensor, with $M = 3$, $T = 2$ ms, $J = 5$, and $N = 50$.

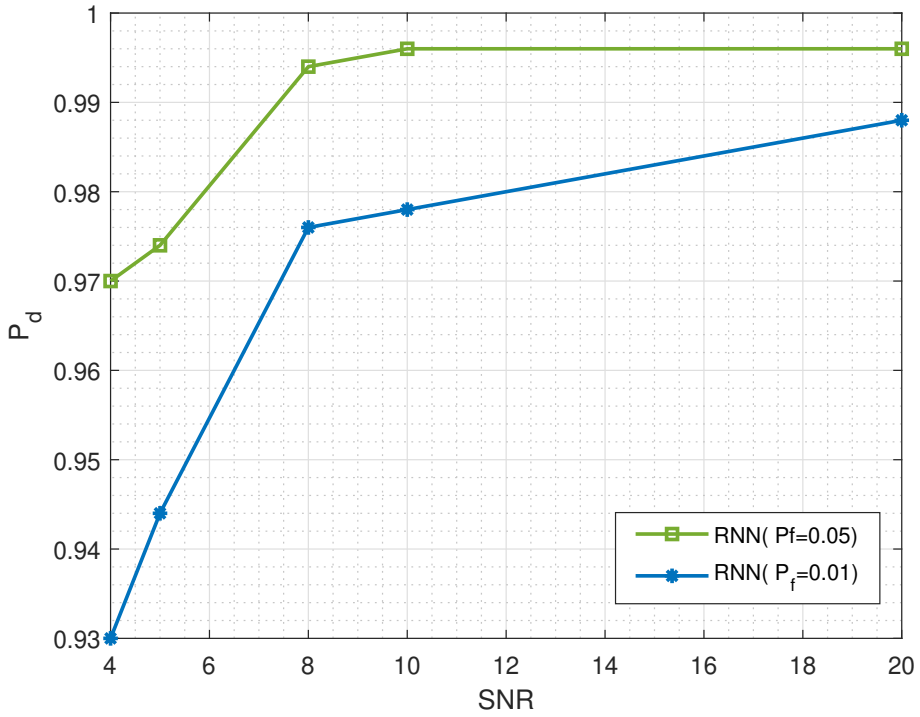


Figure 5.14: P_d of the proposed RNN-based detector for different *SNR* levels with $M = 3$, $T = 2$ ms, $J = 5$, and $N = 50$ considering $P_f = 0.05, 0.01$.

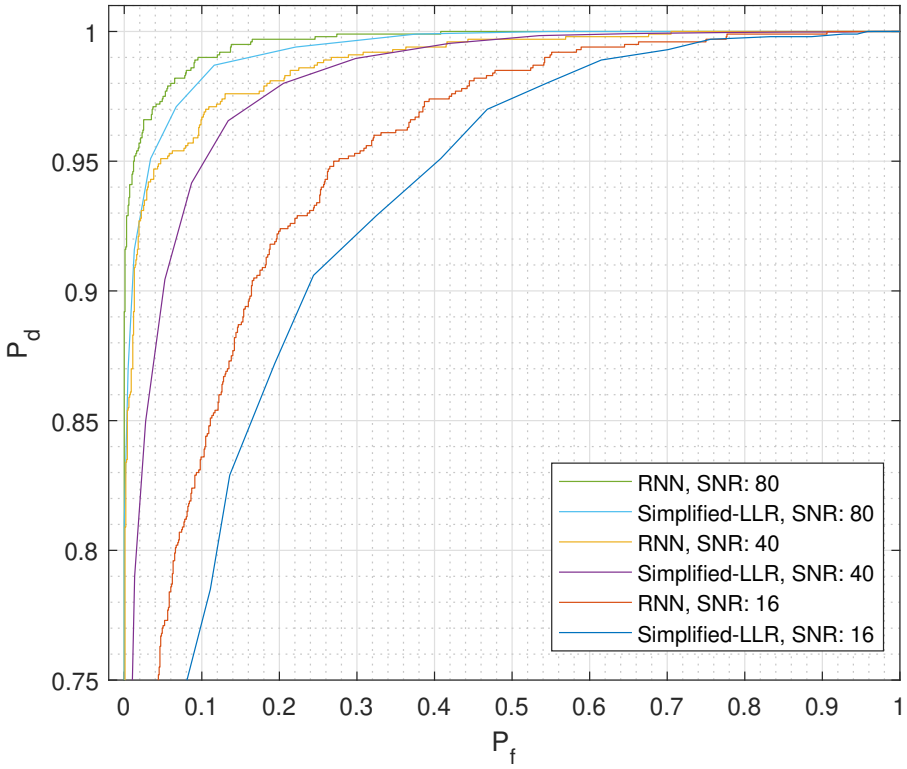


Figure 5.15: ROC curves of RNN and simplified- LLR detector for different *SNR* levels System-1, Configuration-2.

in the literature. Also, they assume an unbounded medium whose physical and chemical properties remain constant. A more pragmatic and practical approach to DF as the living organisms do in nature is required. We model the cell environment and a blood vessel using particle-based simulator AcCoRD where the analytical model of the channel is not valid.

The model-based method proposed in [17] conducts the DF at the FC using a simplified version of the LLR test as an approximation of (2.8). While MLL methods may be used to perform optimal DF at the FC, this approach needs previous knowledge of a set of channel characteristics and sensor distributions, which may be difficult owing to the high number of potentially unknown parameters. Still, it would be useful to compare the proposed RNN based detector and the simplified-LLR test using the same dataset produced with particle-based simulator AcCoRD. The *SNR* (A/J) of the data has to be known in the simplified LLR test while RNN-based detector can function with data with different *SNR* values as the training data contains this information. The ROC curves of RNN and simplified -LLR detector is illustrated in Fig 5.15. This figure can be interpreted as RNN based detector not only performs well but also can be used in applications where the channel conditions change over time. This dataset consists of signals from different *SNR* levels ($SNR = 16, 40, 80$) for $r_1 = 8\ \mu\mathrm{m}$ for the second sensor to FC distance while the first one $1\ \mu\mathrm{m}$ distant in the -y-axis from the middle sensor, and the third one is $1\ \mu\mathrm{m}$ distant in the y-axis in all iterations as shown in Fig 5.3. Other parameters are chosen from the configuration-2 in table 5.1. Please note that the training data consist signals for $SNR = 16, 40, 80$ to show the flexibility and robustness of RNN based DF strategy, while simplified-LLR test applied individually for each *SNR* level as the test is obliged to know the *SNR* information.

5.3.2 Abnormality Detection Within a Blood Vessel Environment

In this section, inspired by the numerous healthcare applications of MC within the IoBNT, we aim to address the problem of blood vessel abnormality detection using multiple nanoscale sensors and an absorbing receiver. Certain scientific efforts have been focused to the identification of abnormalities such as tumors and cancer [42], [41], [111], [112]. We investigate the abnormality detection in blood vessels, which can be utilized in various applications. With the ad-

Table 5.5: Classification report of System-1, Configuration-2

Hypothesis	Precision	Recall	f1-score
$\mathcal{H}_0$	0.9404	0.9937	0.9663
$\mathcal{H}_1$	0.9934	0.9380	0.9649

vancement of nanorobots to detect critical diseases in the cardiocirculatory system [50], [116], we believe that the detection performance would be critical. For example, we can assume that the nanosensors, pushed by the blood flow, can detect cells with the CD47 protein, one of the critical cancer-biomarker, or other biomarkers on their surface, as it has been linked to a variety of malignancies. Nanosensors release information carrier molecules to the bloodstream. These molecules propagate through the blood and a receiver with absorbing boundaries absorbs and counts them to make the final decision. According to this decision, an engineered liposome can release the drug directly to the tumour site [112].

The cardiovascular system has five different types of blood vessels: the arteries, which transport blood away from the heart; the arterioles and the capillaries, which facilitate the exchange of water and chemicals between the blood and the tissues; the venules and the veins carry blood from the capillaries back to the heart. The aorta's diameter ranges from $2-25$ mm while the diameter of an artery ranges from $2-4$ mm and arterioles from $2-30$ μm. The veins have a diameter of $0.5-5$ mm while venules' diameters may range from $1-29$ μm. Capillaries are the body's tiniest blood vessels, linking the smallest arteries to the smallest veins. The term "microcirculation" refers to these vessels. They have a diameter of $0.5-10$ μm. Flow properties of the blood vary depending on the type of blood vessels [117]. We consider blood vessels to be at a substantial distance from the heart so that the bloodstream may be modelled without turbulence. In this thesis, we consider a uniform flow with constant velocity v.

To model the above scenario, we consider cooperative abnormality detection with several nanosensors and an FC inside a blood vessel, i.e. three-dimensional (3-D) flow-induced rectangular pipe form, where the propagation length is greater than the width dimensions. For the simulation in AcCoRD, the parameters are set similar to recent studies like to simulate the blood vessels of a human body [26], [42]. The illustration of a blood vessel with 100 μm length, 10 μm height, 10 μm width and reflective boundaries is shown in Fig 5.16 [42]. Diffusion coefficient and the flow velocity are set similar to the values defined in the paper [26].

The parameters of the first scenario are: the diffusion coefficient $D = 248.78$ $\mu\text{m}^2/\text{s}$, the number of time slots $N = 9$, the number of nano-sensors $M = 3$ with distances $r_{1a} = 70$ μm, $r_{1b} = 65$ μm, $r_{1c} = 60$ μm from the FC with radius of $r_{fc} = 3$ μm and the other distances are shown in Fig 5.17. Each of the M sensors transmits its output to the FC starting at the time instant $t = 0$ by releasing N consecutive pulses transmitted of 400 information-carrying molecules. The list of the parameters used in the simulations is given in Table 5.6. The histogram of absorbed number of molecules at the FC acquired in the final time slot for the case $r_{1a} = 70$ μm, $r_{1b} = 65$ μm, $r_{1c} = 60$ μm from the FC is presented in Fig 5.18-(a).

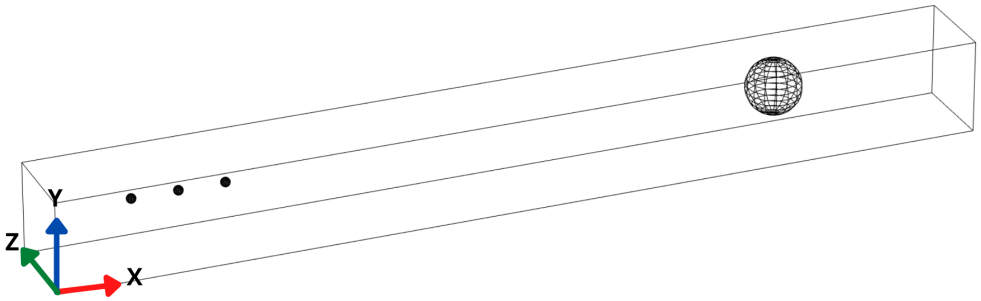


Figure 5.16: Representation of a blood vessel in AcCoRD.

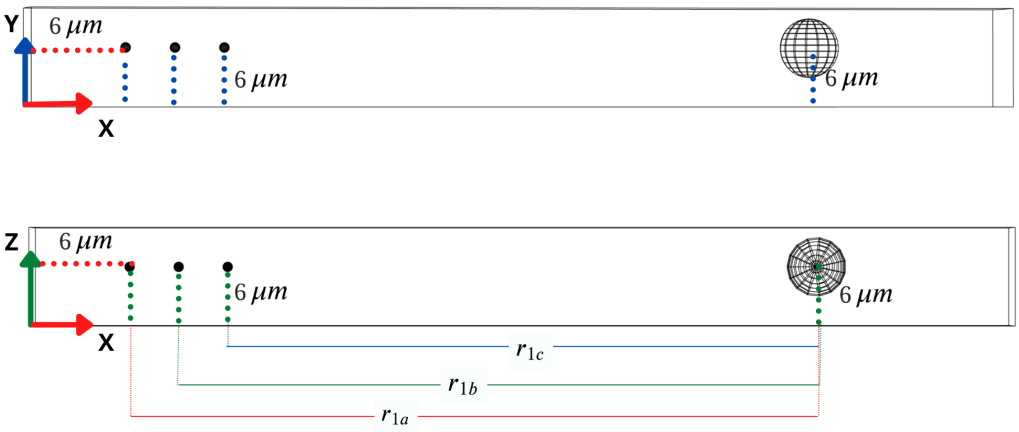


Figure 5.17: Representation of blood vessel from different angles.

To begin, we use the grid-search optimization approach to determine the optimum combination of hyper-parameters. In Table 5.7, the optimal values for producing the desired outcomes are underlined in bold for the configuration-1. We proceed to analyze the RNN algorithm after finding the optimum parameter values using the grid-search approach by showing the confusion matrix and ROC curve using different threshold settings. The training dataset consists signals from different *SNR* levels for $J = 10, 20, 50, 100, 150$. The results representing the overall performance of the RNN algorithm trained with different *SNR* levels for $J = 10, 20, 50, 100, 150$ are illustrated in Figs 5.19, 5.20.

We examine the RNN detector further for different *SNR* levels by evaluating the approach with unseen data (each validation data has one noise level). The previously trained method performs well for varying *SNR* levels considering $Pf = 0.01, 0, 05, 0.005, 0.001$ as shown in Fig 5.21.

Another dataset is prepared with a different configuration for System-2 to evaluate the performance of the detector for changing sensor-FC distances. In this configuration the noise level

is constant and it is chosen as $J = 5$ while five different sensor-FC distances are considered. The FC's location is fixed but transmitters's location are changing as shown in the parameters column of Table 5.9. After having acquired the data and defined the network model, the algorithm is ready to be trained and tested. The training data consist of 5.10^4 separate simulations containing all measurements for configuration-2. The training dataset consists of signals from different sensor-FC distances for $J = 5$. Other parameters are chosen the same as in Table 5.6. Histograms of the number of absorbed molecules for different sensor-FC distances acquired in the final time slot are plotted in Fig 5.18. We employ the grid-search optimization approach to determine the optimum combination of hyper-parameters. In Table 5.7, the optimal values for producing the desired outcomes are underlined in bold for configuration-2.

Overall performance of the algorithm trained with different range of sensor-FC distances is shown in Figs 5.22, 5.23 tested with the data which is randomly chosen from the general dataset including signals from all sensor-FC distances. It can be seen from the results that the performance of the RNN algorithm is high with the $TP = 0.996$ and $TN = 0.968$. The classification report is given in the Table 5.8.

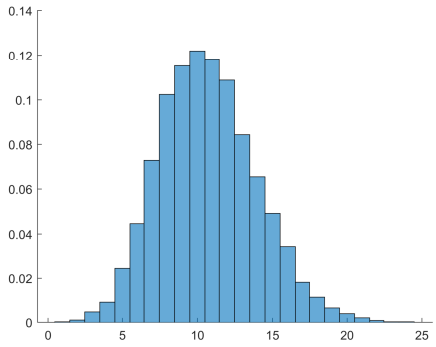
Furthermore, we show the performance of the RNN-based detector for different sensor-FC

Table 5.6: List of parameters used in the simulations of System-2

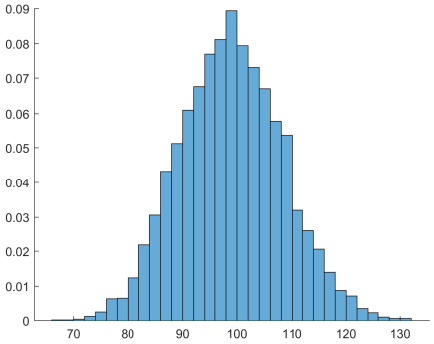
Parameter	**Configuration**
A (The maximum number of molecules that a sensor can release for each pulse)	400
g_{ts} (Global microscopic time step)	50 μs
T (Time slot duration)	15 ms
N (Number of time slots)	9
M (Number or nano-sensors)	3
L (Number of quantization levels)	4
l_v (Length of blood vessel)	100 μm
w_v (Width of blood vessel)	10 μm
h_v (height of blood vessel)	10 μm
r_{fc} (FC radius)	3 μm
v (flow velocity along the length of blood vessel)	0.04 cm/s
D (Diffusion coefficient)	248.78 $\mu m^2/s$

Table 5.7: List of parameters used in grid-search optimization algorithm

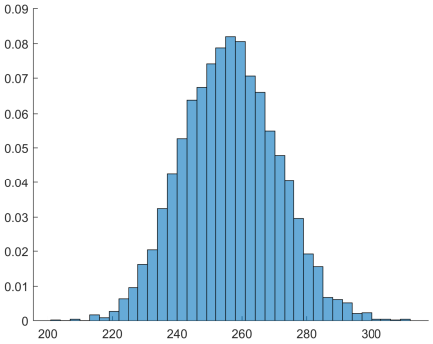
	Optimizer	Number of epochs	Number of batches
Configuration-1	**Adam**, RMSprop	50, **100**, 150	5, **10**, 20
Configuration-2	Adam, **RMSprop**	50, **100**, 150	5, **10**, 20



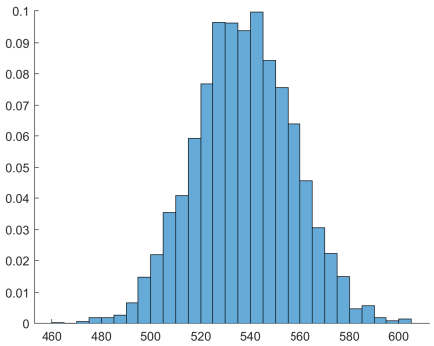
(a) Histogram for r_{1a} = 70μm, r_{1b} = 65 μm, r_{1c} = 60 μm



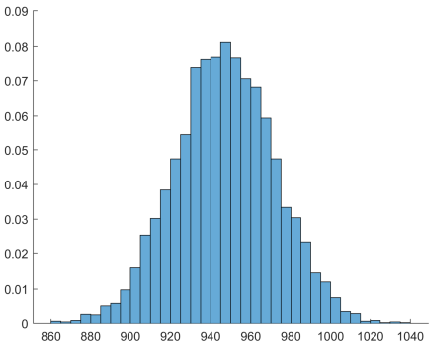
(b) Histogram forr_{1a} = 65 μm, r_{1b} = 60 μm, r_{1c} = 55 μm



(c) Histogram for r_{1a} = 60 μm, r_{1b} = 55 μm, r_{1c} = 50 μm



(d) Histogram for r_{1a} = 55 μm, r_{1b} = 50 μm, r_{1c} = 45 μm



(e) Histogram for r_{1a} = 50 μm, r_{1b} = 45 μm, r_{1c} = 40 μm

Figure 5.18: Histograms for System-2, Configuration 2 for each sensor-FC distances.

distances (validation data has is acquired for each sensor-FC distance). The previously trained algorithm is validated with unseen validation data. Classification report of each case is shown in Table 5.9.

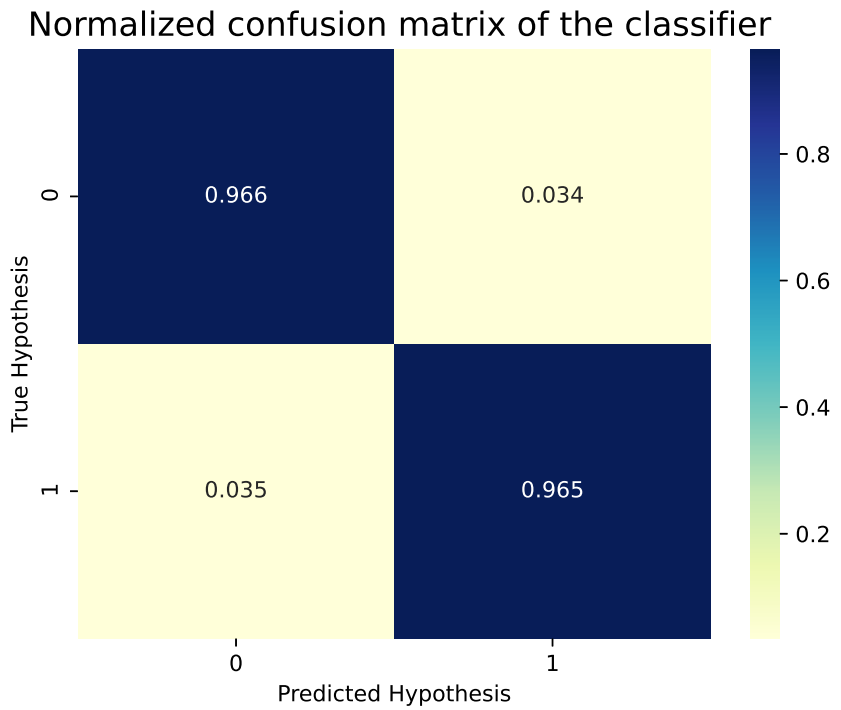


Figure 5.19: Normalized confusion matrix of the proposed RNN algorithm trained with the dataset consisting signals from *SNR* levels for $J = 10, 20, 50, 100, 150$, with $r_{1a} = 70$ μm, $r_{1b} = 65$ μm, $r_{1c} = 60$ μm distanced from the FC, and $N = 9$.

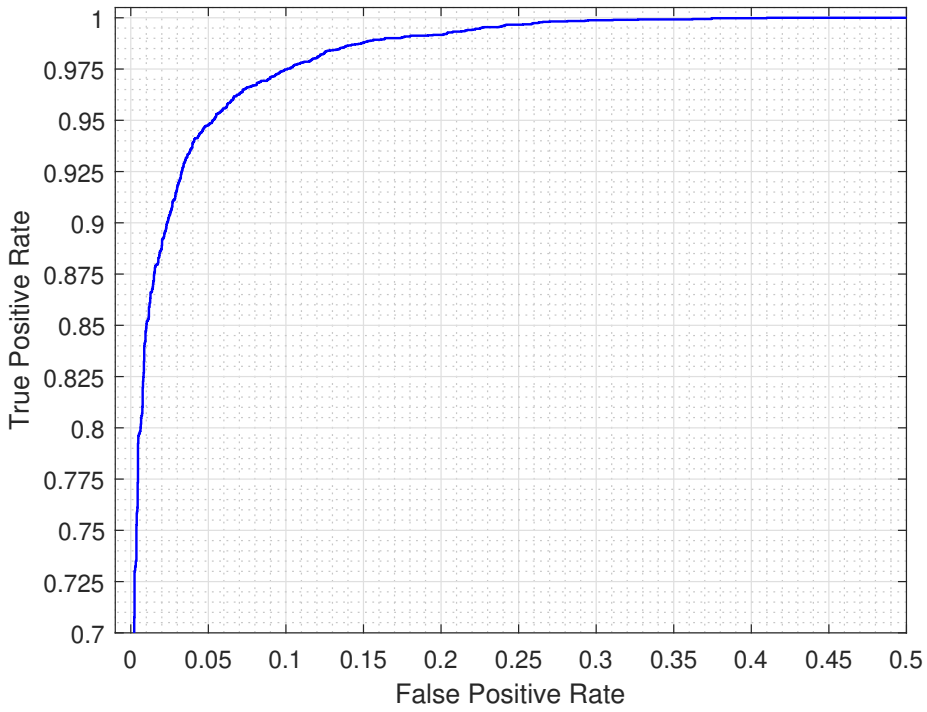


Figure 5.20: ROC Curve of the proposed RNN algorithm trained with the dataset consisting signals from *SNR* levels for $J = 10, 20, 50, 100, 150$, with $r_{1a} = 70$ μm, $r_{1b} = 65$ μm, $r_{1c} = 60$ μm distanced from the FC, and $N = 9$.

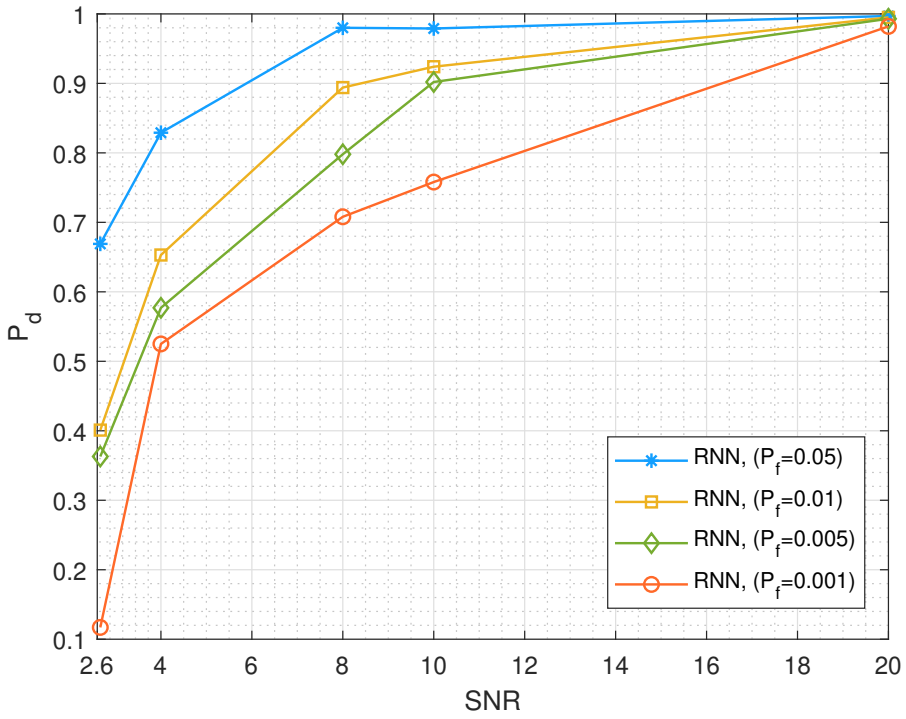


Figure 5.21: P_d values for different *SNR* levels for System-2 for $Pf = 0.01, 0, 05, 0.005, 0.001$.

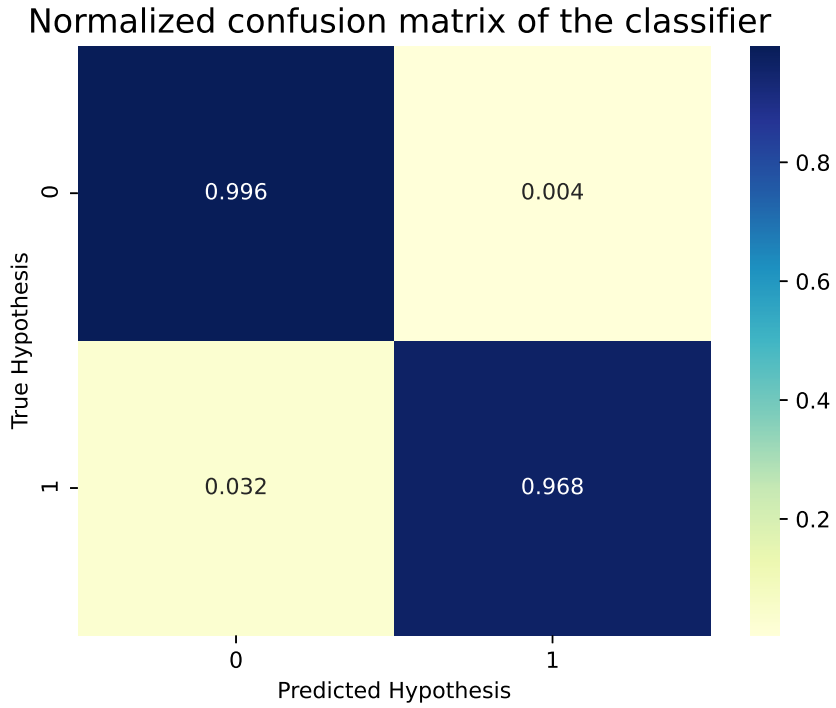


Figure 5.22: Normalized confusion Matrix.

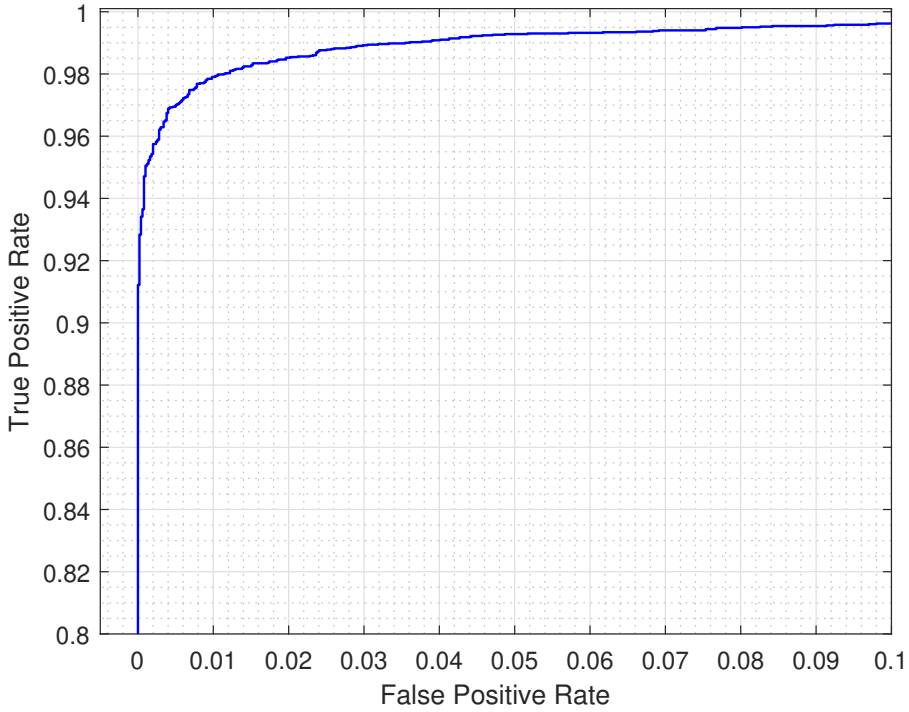


Figure 5.23: ROC Curve of the proposed RNN algorithm trained with the dataset consisting signals from five different sensor-FC distances with $J = 5$, and $N = 9$.

5.3.3 Conclusion

In this chapter, we look at two biological systems with various settings that will be helpful for healthcare applications. In AcCoRD, we model a cell environment and a blood vessel to detect abnormalities in these environments. A unique machine learning-based approach for designing DF algorithms for identifying anomalies in a fluid environment utilizing MC is developed in this chapter using the particle-based simulator AcCoRD for generating data that is required for training of the algorithms in the investigated scenarios. We propose to employ RNN structures for anomaly detection in a network of collaborating NSs. Given the very complex and dynamic nature of the data encountered in MC, the suggested RNN-based method is appropriate for the intended job. We demonstrate that our NN-based approach for designing DF algorithms for

Table 5.8: Classification report of System-2, Configuration-2

Hypothesis	Precision	Recall	f1-score
$\mathcal{H}_0$	0.9744	0.9916	0.9663
$\mathcal{H}_1$	0.9915	0.9741	0.9827

abnormality detection using a diffusive MC based sensor network perform well with the dataset acquired via simulator enabling us to simulate realistic scenarios. We believe that due to wet-lab data is not generally available, simulation data is the best available. We demonstrate that the proposed algorithm performs well for the cases mimicking the cell environment and the blood vessel considering different configurations. We show that the proposed approach is well suited for the applications requiring robustness and flexibility. The results show that the proposed approach is suitable for practical applications performing better than simplified LLR. Machine learning based DF techniques are developed that do not require any analytical channel model, instantaneous CSI, or knowledge of other system factors such as the statistical features of the sensors, or even the number of sensors and obstacles in the environment.

Table 5.9: Classification report of System-2 with validation data

Parameters	Hypothesis	Precision	Recall	f1-score
r_{1a} = 70 μm, r_{1b} = 65 μm, r_{1c} = 60 μm	0	1.000	0.9840	0.9919
	1	0.9843	1.000	0.9921
r_{1a} = 65 μm, r_{1b} = 60 μm, r_{1c} = 55 μm	0	1.000	0.9850	0.9924
	1	0.9852	1.000	0.9926
r_{1a} = 60 μm, r_{1b} = 55 μm, r_{1c} = 50 μm	0	1.000	0.9880	0.9940
	1	0.9881	1.000	0.9940
r_{1a} = 55 μm, r_{1b} = 50 μm, r_{1c} = 45 μm	0	1.000	0.9920	0.9960
	1	0.9921	1.000	0.9960
r_{1a} = 50 μm, r_{1b} = 45 μm, r_{1c} = 40 μm	0	1.000	0.9840	0.9924
	1	0.9852	1.000	0.9926

Conclusions and Future Work

6

The eyes of others our prisons; their thoughts our cages.

Virginia Woolf

This chapter concludes the dissertation and proposes opportunities for future research. Section 6.1 summarizes our findings and draws broad conclusions. Section 6.2 contains comments on the direction of this field, including topics for future relevant research.

6.1 Conclusions

In this section, we review the key results from each chapter and then outline the dissertation's conclusions.

The signal propagation characteristics in a diffusive MC channel are highly random in nature and depend heavily on a multitude of factors. We need consider all of these factors in practical applications. However, recent research in molecular communication has been limited to the development and study of nanonetworks based on basic assumptions about bionanomachines and their environment. A critical problem in moving the field of molecular communication forward is the development of robust DF strategies for creating nanoscale sensor networks that operate in the real world of practical applications.

One of the most important tasks of the most highly anticipated applications of nanoscale

networks, such as health monitoring, disease diagnosis, targeted drug delivery, environmental sensing and monitoring, contaminant and toxic agent detection, environmental remediation, and many others, are DD problems. In comparison to the existing literature in wireless sensor networks-based DD, MC-based DD research is still in its development. We believe that our research give valuable insights into the abnormality detection with sensor network literature.

The extremely slow signal propagation speed in the medium and the highly dispersive nature of the channel, resulting in long pulse intervals and significant latency, are two of the key features of diffusive MC. In this context, we have presented a novel concept for the decision fusion in diffusive MC based DD. For the first time in the literature, we propose to employ a sequential test to this critical task which we refer to as SAPRT. Our results presented in Chapter 3 confirm that the suggested approach results in significant gains in the average number of samples required for the decision compared to fixed- sample size alternatives. We observe savings of up to 80% reducing the average decision delay significantly. We obtain the same average detection performance without making an i.i.d assumption, which may result in extra decision delays in reality.

Over the last years, DNNs have gotten a lot of attention for their unique ability to solve problems even when there isn't an explicit mathematical model. Existing approaches for DD problems can only be used on the assumption that the channel model and its parameters are known at the receiver side which is usually impractical. It is due to mathematically intractable channel models for MC and channel estimation requirements, which are challenging even in ideal conditions. In this thesis, we propose for the first time in the literature to use machine learning techniques to design robust and reliable DF algorithms for practical applications. We investigate two different NN structures for DF, the FF-NN and RNN. While both type of DF algorithms perform well, the RNN-based solution has shown to be better suited for the given job, as predicted, due to the sequential pattern of the data encountered in MC and its temporal structure, which the RNN can exploit. NN detectors can be trained directly using data collected from mathematical models, simulators or, alternatively, from in vivo measurements and in vitro experiments. Analytical channel models, instantaneous CSI or knowledge of other system parameters are not required for our proposed approach. The results show that the innovative ML-based DF method described in this dissertation provides high detection performance as well as the robustness and flexibility necessary for realistic implementations of such MC-based nanoscale sensor networks.

In systems where the underlying physical models of the channel are unclear or incorrect, these algorithms have a lot of potential. Our results demonstrates that deep learning based detectors has a great potential in the future designing promising communication systems as shown

in Chapter 4. The overall detection performance can be improved with proposed approach while providing resilience to changing channel conditions. A great advantage of the approach is that no CSI is needed to perform the DF task. NN detectors are trained directly using data collected from mathematical models in the section 2 and section 3 to demonstrate the results. In Chapter 5, the performance of the proposed NN-based DF method is assessed using particle-based simulations using the particle-based molecular communication simulator AcCoRD in the signal generation. Whereas mathematical models have limits, we simulate practical circumstances in AcCoRD . Because wet-lab data is not generally available for diffusive MC, simulation data is the best alternative. In AcCoRD, we primarily model a cell environment and a blood vessel which we believe to be useful for applications in healthcare , medicine, and many more. We show that NN-detectors provides good detection performance in changing channel conditions and can adapt itself during the detection process.

6.2 Future Work

The following are some intriguing open challenges that can be investigated more in the future. First, hyper-parameter tuning with different parameters and different parameter configuration can be studied further to discover better performing detectors for DF task.

Much of what is we refer to as ”machine learning” in this research is actually ”supervised” machine learning, which is reliant on manual human feedback. We utilize ground-truth labels for the training of the algorithm and let it predict the true labels for unseen data. On the other hand, unsupervised learning does not require a training data to be labelled. It doesn't require a training process and can be fast for classification problem.

Other deep learning architecture can be applied to DF task is reinforcement learning which is, along with supervised and unsupervised learning, one of the three main machine learning paradigms.

Another approach would be combining NN and other techniques such as Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3), K-Nearest-Neighbour(KNN). The combination of these different technique may perform relatively higher than the conventional models like in the studies [118], [119].

A learning model supplied with enough high-quality data is more likely to produce correct outcomes. Gathering enough data to improve the models' accuracy might be hard practically in the MC applications. This would be mainly because of the restrictions of the nanoscale size,

limited energy reserves. A type of ML technique, namely few-shot learning algorithms where a classifier must generalize quickly after seeing only a few instances from each class can be applied to DF task using MC. This technique have been successfully used to find patterns in data and generate successful predictions in such limited-data and challenging environments. Even further, an LSTM-based *meta-learner* model proposed in [120] to learn the particular optimization strategy employed to train another learner neural network classifier in the few-shot regime.

The experimental verification of diffusive MC models is critical for the progression of the state-of-the-art towards practical applications. Other approach can be enhancing the existing diffusive MC models to use more realistic physical phenomena, allowing them to be easily compared to experimental data. Further development in particle based simulation techniques will help realizing the full potential of molecular communication to implement more realistic scenarios for practical applications until wet-lab based implementations become extensive. We may develop the system model, and keep drawing inspiration for network design from biological mechanisms for adaptability. An advanced simulator should scale well to big scenarios with a high background concentration because there would be cases that we must compute the motion and interactions of every single particle in the simulation environment. As a result, in order to build higher-level protocols for complicated scenarios, there may be a need for a higher-level simulator with improved scalability. Defining heterogeneous environments where the diffusion coefficient can be different at various locations in the propagation channel can be an important improvement. Most MC research assumed laminar flow from transmitter to receiver but time-varying flows and turbulent flows should be explored as part of future studies.

It takes a long time to establish oneself as a scientist. The most risky part is figuring out what a good scientist is and then taking the first steps down that rocky path. It then turned into a journey, which will hopefully take me home one day.

Bibliography

[1] Tadashi Nakano, Andrew W Eckford, and Tokuko Haraguchi. *Molecular communication*. Cambridge University Press, first edition, 2013.

[2] Tadashi Nakano, Michael J Moore, Fang Wei, Athanasios V Vasilakos, and Jianwei Shuai. Molecular communication and networking: Opportunities and challenges. *IEEE transactions on nanobioscience*, 11(2):135–148, 2012.

[3] Bharat Bhushan and Baumann. *Springer handbook of nanotechnology*, volume 2. Springer, second edition, 2007.

[4] Charles P Poole Jr and Frank J Owens. *Introduction to Nanotechnology*. Wiley India Pvt. Limited, first edition, 2007.

[5] Amy C Richards Grayson, Rebecca S Shawgo, Audrey M Johnson, Nolan T Flynn, Yawen Li, Michael J Cima, and Robert Langer. A biomems review: Mems technology for physiologically integrated devices. *Proceedings of the IEEE*, 92(1):6–21, 2004.

[6] Ian F Akyildiz, Josep Miquel Jornet, and Massimiliano Pierobon. Nanonetworks: A new frontier in communications. *Communications of the ACM*, 54(11):84–89, 2011.

[7] Mainor Alberto Cruz Alvarado and Patricia Alejandra Bazán. Understanding the Internet of Nano Things: overview, trends, and challenges. *e-Ciencias de la Información*, 9, 2018.

[8] Ian F Akyildiz, Max Pierobon, Sasi Balasubramaniam, and Y Koucheryavy. The internet of bio-nano things. *IEEE Communications Magazine*, 53(3):32–40, 2015.

[9] Nariman Farsad, H. Birkan Yilmaz, Andrew Eckford, Chan Byoung Chae, and Weisi Guo. A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys and Tutorials*, 18(3):1887–1919, 2016.

[10] John S Parkinson and Eric C Kofoid. Communication modules in bacterial signaling proteins. *Annual review of genetics*, 26(1):71–112, 1992.

[11] Bonnie L Bassler. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current opinion in microbiology*, 2(6):582–587, 1999.

[12] Junichi Suzuki Tadashi Nakano, T Suda, T Koujin, H Tokuko, and Y Hiraoka. Bio-inspired models of network, information and computing systems. volume 2, pages 139–146. Springer, 2012.

[13] Vahid Jamali, Arman Ahmadzadeh, Christophe Jardin, Heinrich Sticht, and Robert Schober. Channel Estimation for Diffusive Molecular Communications. *IEEE Transactions on Communications*, 64(10):4238–4252, 2016.

[14] Vahid Jamali, Nariman Farsad, Robert Schober, and Andrea Goldsmith. Non-Coherent Detection for Diffusive Molecular Communication Systems. *IEEE Transactions on Communications*, 66(6):2515–2531, 2018.

[15] Vahid Jamali, Nariman Farsad, Robert Schober, and Andrea Goldsmith. Non-coherent detection for diffusive molecular communication systems. *I.E.E.E. transactions on communications.*, 66(6), 2018-6.

[16] Ignacio Llatser, Albert Cabellos-Aparicio, Massimiliano Pierobon, and Eduard Alarcon. Detection techniques for diffusion-based molecular communication. *IEEE Journal on Selected Areas in Communications*, 31(12):726–734, 2013.

[17] Reza Mosayebi, Vahid Jamali, Nafiseh Ghoroghchian, Robert Schober, Masoumeh Nasiri-Kenari, and Mahdieh Mehrabi. Cooperative abnormality detection via diffusive molecular communications. *IEEE transactions on nanobioscience*, 16(8):828–842, 2017.

[18] Reza Mosayebi, Wayan Wicke, Vahid Jamali, Arman Ahmadzadeh, Robert Schober, and Masoumeh Nasiri-Kenari. Advanced Target Detection via Molecular Communication. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, dec 2018.

[19] Adam Noel, Karen C. Cheung, and Robert Schober. Optimal receiver design for diffusive molecular communication with flow and additive noise. *IEEE Transactions on Nanobioscience*, 13(3):350–362, 2014.

[20] Adam Noel. *Modeling and analysis of diffusive molecular communication systems.* PhD thesis, University of British Columbia, 2015.

[21] Adam Noel, Yansha Deng, Dimitrios Makrakis, and Abdelhakim Hafid. Active versus passive: Receiver model transforms for diffusive molecular communication. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2016.

[22] Shirin Salehi, Naghmeh S. Moayedian, and Eduard Alarcon. Diffusion-Based Molecular Communication Channel in Presence of a Probabilistic Absorber: Single Receptor Model and Congestion Analysis. *IEEE Transactions on Nanobioscience*, 18(1):84–92, 2019.

[23] Burcu Tepekule, Ali E. Pusane, H. Birkan Yilmaz, Chan Byoung Chae, and Tuna Tugcu. ISI Mitigation Techniques in Molecular Communication. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, 1(2):202–216, 2015.

[24] H Birkan Yilmaz and Chan-Byoung Chae. Arrival modelling for molecular communication via diffusion. *Electronics Letters*, 50(23):1667–1669, 2014.

[25] H Birkan Yilmaz, Akif Cem Heren, Tuna Tugcu, and Chan-Byoung Chae. Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Communications Letters*, 18(6):929–932, 2014.

[26] Nooshin Tavakkoli, Paeiz Azmi, and Nader Mokari. Performance evaluation and optimal detection of relay-assisted diffusion-based molecular communication with drift. *IEEE Transactions on NanoBioscience*, 16(1):34–42, 2017.

[27] Arman Ahmadzadeh, Hamidreza Arjmandi, Andreas Burkovski, and Robert Schober. Comprehensive reactive receiver modeling for diffusive molecular communication systems: Reversible binding, molecule degradation, and finite number of receptors. *IEEE transactions on nanobioscience*, 15(7):713–727, 2016.

[28] Murat Kuscu and Ozgur B Akan. Maximum likelihood detection with ligand receptors for diffusion-based molecular communications in internet of bio-nano things. *IEEE transactions on nanobioscience*, 17(1):44–54, 2018.

[29] Murat Kuscu and Ozgur B Akan. Modeling convection-diffusion-reaction systems for microfluidic molecular communications with surface-based receivers in internet of bio-nano things. *PloS one*, 13(2):e0192202, 2018.

[30] Gaye Genc, Yunus Emre Kara, Tuna Tugcu, and Ali Emre Pusane. Reception modeling of sphere-to-sphere molecular communication via diffusion. *Nano communication networks*, 16:69–80, 2018.

[31] Adam Noel, Karen C. Cheung, Robert Schober, Dimitrios Makrakis, and Abdelhakim Hafid. Simulating with AcCoRD: Actor-based Communication via Reaction–Diffusion. *Nano Communication Networks*, 11:44–75, 2017.

[32] Dogu Arifler and Dizem Arifler. Monte carlo analysis of molecule absorption probabilities in diffusion-based nanoscale communication systems with multiple receivers. *IEEE transactions on nanobioscience*, 16(3):157–165, 2017.

[33] Ian F Akyildiz, Fernando Brunetti, and Cristina Blázquez. Nanonetworks: A new communication paradigm. *Computer Networks*, 52(12):2260–2279, 2008.

[34] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[35] Siavash Ghavami, Farshad Lahouti, and Ali Masoudi-Nejad. Modeling and analysis of abnormality detection in biomolecular nano-networks. *Nano Communication Networks*, 3(4):229–241, 2012.

[36] Pramod K Varshney. *Distributed detection and data fusion*. Springer Science & Business Media, second edition, 2012.

[37] Ramanarayanan Viswanathan and Pramod K. Varshney. Distributed detection with multiple sensors: Part I - fundamentals. *Proceedings of the IEEE*, 85(1):54–63, 1997.

[38] R.S. Blum, S.A. Kassam, and H.V. Poor. Distributed detection with multiple sensors ii. advanced topics. *Proceedings of the IEEE*, 85(1):64–79, 1997.

[39] Richard P Feynman et al. There's plenty of room at the bottom. *California Institute of Technology, Engineering and Science magazine*, 1960.

[40] H Vincent Poor. *An introduction to signal detection and estimation*. Springer Science & Business Media, third edition, 2013.

[41] Siavash Ghavami and Farshad Lahouti. Abnormality Detection in Correlated Gaussian Molecular Nano-Networks: Design and Analysis. *IEEE Transactions on Nanobioscience*, 16(3):189–202, 2017.

[42] Neeraj Varshney, Adarsh Patel, Yansha Deng, Werner Haselmayr, Pramod K Varshney, and Arumugam Nallanathan. Abnormality detection inside blood vessels with mobile nanomachines. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 4(3):189–194, 2018.

[43] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.

[44] Ananthram Swami, Qing Zhao, Yao-Win Hong, and Lang Tong. *Wireless sensor networks: signal processing and communications perspectives*. John Wiley & Sons, first edition, 2007.

[45] Shamneesh Sharma, Dinesh Kumar, and Keshav Kishore. Wireless sensor networks-a review on topologies and node architecture. *International Journal of Computer Sciences and Engineering*, 1(2):19–25, 2013.

[46] Stephen F Bush. *Nanoscale communication networks*. Artech House, first edition, 2010.

[47] Tatsuya Suda, Michael Moore, Tadashi Nakano, Ryota Egashira, Akihiro Enomoto, Satoshi Hiyama, and Yuki Moritani. Exploratory research on molecular communication between nanomachines. In *Genetic and evolutionary computation conference (GECCO), Late breaking papers*, volume 25, page 29. Citeseer, 2005.

[48] Parvez Iqbal, Jon A Preece, and Paula M Mendes. Nanotechnology: The "top-down" and "bottom-up" approaches. *Supramolecular chemistry: from molecules to nanomaterials*, 2012.

[49] K Eric Drexler. *Nanosystems: molecular machinery, manufacturing, and computation*. John Wiley & Sons, Inc., first edition, 1992.

[50] Sung Jun Park, Seung-Hwan Park, Sunghoon Cho, Deok-Mi Kim, Yeonkyung Lee, Seong Young Ko, Yeongjin Hong, Hyon E Choy, Jung-Joon Min, Jong-Oh Park, et al. New paradigm for tumor theranostic methodology using bacteria-based microrobot. *Scientific reports*, 3(1):1–8, 2013.

[51] George M Whitesides. The once and future nanomachine. *Scientific American*, 285(3):78–83, 2001.

[52] S Hiyama, Y Moritani, Tatsuya Suda, R Egashira, A Enomoto, M Moore, and T Nakano. Molecular communication. *Journal-Institute of Electronics Information and Communication Engineers*, 89(2):162, 2006.

[53] Karl Erik Rosengren. *Communication: an introduction*. Sage, first edition, 1999.

[54] Uche Chude-Okonkwo, Reza Malekian, and BT Maharaj. Communication between living and non-living systems: The basis for advanced targeted nanomedicine. In *Advanced Targeted Nanomedicine*, pages 19–37. Springer, 2019.

[55] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, second edition, 2015.

[56] Daniel Thomas Gillespie and Effrosyni Seitaridou. *Simple Brownian diffusion: an introduction to the standard theoretical models*. Oxford University Press, first edition, 2013.

[57] Amin Gohari, Mahtab Mirmohseni, and Masoumeh Nasiri-Kenari. Information theory of molecular communication: Directions and challenges. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(2):120–142, 2016.

[58] Xu Bao, Yanfei Zhu, and Wence Zhang. Channel characteristics for molecular communication via diffusion with a spherical boundary. *IEEE Wireless Communications Letters*, 8(3):957–960, 2019.

[59] Yansha Deng, Adam Noel, Weisi Guo, Arumugam Nallanathan, and Maged Elkashlan. 3d stochastic geometry model for large-scale molecular communication systems. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2016.

[60] Yiran Wang, Adam Noel, and Nan Yang. A new simulation algorithm for absorbing receiver in molecular communication. In *2018 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*, pages 1–4. IEEE, 2018.

[61] Ali Akkaya, H Birkan Yilmaz, Chan-Byoung Chae, and Tuna Tugcu. Effect of receptor density and size on signal reception in molecular communication via diffusion with an absorbing receiver. *IEEE Communications Letters*, 19(2):155–158, 2014.

[62] Adam Noel, Dimitrios Makrakis, and Abdelhakim Hafid. Channel impulse responses in diffusive molecular communication with spherical transmitters. *arXiv preprint arXiv:1604.04684*, 2016.

[63] H Birkan Yilmaz, Gee-Yong Suk, and Chan-Byoung Chae. Chemical propagation pattern for molecular communications. *IEEE Wireless Communications Letters*, 6(2):226–229, 2017.

[64] Hamidreza Arjmandi, Arman Ahmadzadeh, Robert Schober, and Masoumeh Nasiri Kenari. Ion channel based bio-synthetic modulator for diffusive molecular communication. *IEEE transactions on nanobioscience*, 15(5):418–432, 2016.

[65] Michael L Simpson, Gary S Sayler, James T Fleming, and Bruce Applegate. Whole-cell biocomputing. *Trends in biotechnology*, 19(8):317–323, 2001.

[66] Jack W Judy. Microelectromechanical systems (mems): fabrication, design and applications. *Smart materials and Structures*, 10(6):1115, 2001.

[67] Aviv Regev and Ehud Shapiro. Cellular abstractions: Cells as computation. *Nature*, 419(6905):343–343, 2002.

[68] Dennis Bray. Protein molecules as computational elements in living cells. *Nature*, 376(6538):307–312, 1995.

[69] Mehmet Şükrü Kuran, H Birkan Yilmaz, Tuna Tugcu, and Bilge Özerman. Energy model for communication via diffusion in nanonetworks. *Nano Communication Networks*, 1(2):86–95, 2010.

[70] Akram Galal and Xavier Hesselbach. Nano-networks communication architecture: Modeling and functions. *Nano Communication Networks*, 17:45–62, 2018.

[71] Rob Phillips, Jane Kondev, Julie Theriot, Hernan G Garcia, and Nigel Orme. *Physical biology of the cell*. Garland Science, second edition, 2012.

[72] Abraham Wald. *Sequential analysis*. Courier Corporation, second edition, 2004.

[73] Tze-Yang Tung and Urbashi Mitra. Synchronization error robust transceivers for molecular communication. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 5(3):207–221, 2019.

[74] Ryan Martin. Stat 411–lecture notes 03 likelihood and maximum likelihood estimation.

[75] Steven M Kay. *Fundamentals of statistical signal processing*. Prentice Hall PTR, first edition, 1993.

[76] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, first edition, 2014.

[77] Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.

[78] Michele Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice hall Englewood Cliffs, first edition, 1993.

[79] Yu Liu, Yongxin Gao, and X. Rong Li. Operating characteristic and average sample number of binary and multi-hypothesis sequential probability ratio test. *IEEE Transactions on Signal Processing*, 64(12):3167–3179, 2016.

[80] Maili Liu, Jeremy K Nicholson, John A Parkinson, and John C Lindon. Measurement of biomolecular diffusion coefficients in blood plasma using two-dimensional 1h- 1h diffusion-edited total-correlation nmr spectroscopy. *Analytical chemistry*, 69(8):1504–1509, 1997.

[81] Zakkula Govindarajulu. *Sequential statistics*. World Scientific, first edition, 2004.

[82] Adriano Cavalcanti, Bijan Shirinzadeh, Robert A Freitas, and Tad Hogg. Nanorobot architecture for medical target identification. *Nanotechnology*, 19(1):015103, 2007.

[83] Jürgen Schmidhuber. Deep Learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[84] Martin Anthony and Peter L Bartlett. *Neural network learning: Theoretical foundations*. cambridge university press, second edition, 2009.

[85] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[86] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, first edition, 2016. http://www.deeplearningbook.org.

[87] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.

[88] Katarzyna Janocha and Wojciech Marian Czarnecki. On loss functions for deep neural networks in classification. *arXiv preprint arXiv:1702.05659*, 2017.

[89] Navin Kumar Manaswi. Rnn and lstm. In *Deep Learning with Applications Using Python*, pages 115–126. Springer, 2018.

[90] Robert DiPietro and Gregory D Hager. Deep learning: Rnns and lstm. In *Handbook of medical image computing and computer assisted intervention*, pages 503–519. Elsevier, 2020.

[91] Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nathan Srebro. Exploring generalization in deep learning. *arXiv preprint arXiv:1706.08947*, 2017.

[92] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.

[93] Nariman Farsad and Andrea Goldsmith. Neural network detection of data sequences in communication systems. *IEEE Transactions on Signal Processing*, 66(21):5663–5678, 2018.

[94] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[95] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[96] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[97] Adam optimizer. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam/. Accessed: 2021-08-24.

[98] Rmsprop optimizer. https://keras.io/api/optimizers/rmsprop/. Accessed: 2021-09-01.

[99] Brian S Donahue and RF Abercrombie. Free diffusion coefficient of ionic calcium in cytoplasm. *Cell calcium*, 8(6):437–448, 1987.

[100] Henry John Valentine Tyrrell and KR Harris. *Diffusion in liquids: a theoretical and experimental study*. Butterworth-Heinemann, second edition, 2013.

[101] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[102] Christie J Geankoplis. *Transport processes and separation process principles:(includes unit operations)*. Prentice Hall Professional Technical Reference, first edition, 2003.

[103] Yaşar Demirel. 6 - diffusion. In Yaşar Demirel, editor, *Nonequilibrium Thermodynamics (Second Edition)*, pages 319–362. Elsevier Science B.V., Amsterdam, second edition, 2007.

[104] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

[105] Comsol. https://www.comsol.com/. Accessed: 2021-06-24.

[106] Accord. https://warwick.ac.uk/fac/sci/eng/people/adam_noel/software/. Accessed: 2021-06-24.

[107] MATLAB. *9.9.0. 1467703 (R2020b)*. The MathWorks Inc., Natick, Massachusetts, 2020.

[108] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, second edition, 2009.

[109] Matthias Feurer and Frank Hutter. Hyperparameter optimization. In *Automated Machine Learning*, pages 3–33. Springer, Cham, 2019.

[110] Marc Claesen and Bart De Moor. Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*, 2015.

[111] Reza Mosayebi, Arman Ahmadzadeh, Wayan Wicke, Vahid Jamali, Robert Schober, and Masoumeh Nasiri-Kenari. Early cancer detection in blood vessels using mobile nanosensors. *IEEE Transactions on Nanobioscience*, 18(2):103–116, 2019.

[112] Luca Felicetti, Mauro Femminella, and Gianluca Reali. A molecular communication system in blood vessels for the detection of hyperviscosity syndrome. pages 1–6, 2019.

[113] Geoffrey M Cooper, Robert E Hausman, and Robert E Hausman. *The cell: a molecular approach*, volume 4. ASM press Washington, DC, first edition, 2007.

[114] Dianne S Schwarz and Michael D Blower. The endoplasmic reticulum: structure, function and response to cellular signaling. *Cellular and molecular life sciences*, 73(1):79–94, 2016.

[115] NL Allbritton, T Meyer, and L Stryer. Range of messenger action of calcium ion and inositol 1,4,5-trisphosphate. *Science*, 258(5089):1812–1815, 1992.

[116] Satoshi Hiyama, Yuki Moritani, and Tatsuya Suda. Molecular transport system in molecular communication. *NTT DOCOMO Technical Journal*, 10(3):49–53, 2008.

[117] Gerard J Tortora. *Principles of human anatomy*. HarperCollins College, first edition, 1995.

[118] Abien Fred M. Agarap. A neural network architecture combining gated recurrent unit (gru) and support vector machine (svm) for intrusion detection in network traffic data. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, ICMLC 2018, page 26–30, New York, NY, USA, 2018. Association for Computing Machinery.

[119] Yuan Yao, Gian Luca Marcialis, Massimiliano Pontil, Paolo Frasconi, and Fabio Roli. Combining flat and structured representations for fingerprint classification with recursive neural networks and support vector machines. *Pattern Recognition*, 36(2):397–406, 2003.

[120] S achin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. *International Conference on Learning Representations (ICLR)*, 2017.