Information Engineering and Technology
German University in Cairo

ECG Classification using Neural Networks

Bachelor Thesis

Author: Alaa Shams Eldin
Supervisors: Prof.Dr. Tallal Elshabrawy
Prof.Dr. Maggie Mashaly
MSc.Ing. Fatma Hassan

Submission Date: 19 May, 2024

Information Engineering and Technology
German University in Cairo

ECG Classification using Neural Networks

Bachelor Thesis

Author:	Alaa Shams Eldin
Supervisors:	Prof.Dr. Tallal Elshabrawy
	Prof.Dr. Maggie Mashaly
	MSc.Ing. Fatma Hassan

Submission Date:	19 May, 2024

This is to certify that:

(i) the thesis comprises only my original work toward the Bachelor Degree

(ii) due acknowledgement has been made in the text to all other material used

Alaa Shams Eldin
19 May, 2024

Acknowledgments

First of all, I praise and thank Allah for giving me patience and strength to complete this work. I would like to thank my supervisors Prof.Dr. Tallal Elshabrawy, Prof.Dr. Maggie Mashaly and Msc.Ing. Fatma Hassan for their guidance, motivation and mentor ship across the course of this work.
I extend my appreciation to my dearest parents for their support and love throughout my life. Last but not least my friends, for their love and words of affection throughout my college years.

Abstract

This thesis proposes a monitoring system with on-device processing capabilities, using a two stage machine learning (ML) approach for anomaly detection and multi-classification. Our approach involves an initial stage featuring a Multi-Layer Perceptron (MLP), followed by Convolutional Neural Network (CNN) combined with Long Short-Term Memory (LSTM) neural network at the second stage. This design affords flexibility in deploying each stage separately: the first stage operates on-device with minimal computational requirements yet high accuracy, while the subsequent stage, more complex in nature, operates on edge devices.

Our model demonstrates superior computational efficiency compared to baseline approaches, with reduced numbers of multiplications, additions, and parameters. Specifically, our proposed MLP model has over 60% fewer parameters compared to the baseline model, and the overall computational workload is reduced by more than 70%. Evaluation on the MIT-BIH Arrhythmia Database showcases competitive performance against state-of-the-art methods. Thus, the proposed model is suitable for applications in wireless remote health monitoring systems, enabling real-time anomaly detection and classification with minimal computational overhead.

Contents

Chapter 1

Introduction

1.1 Motivation

Long-term monitoring of cardiovascular activity plays a crucial role in the prevention of cardiovascular diseases. One example is the detection of pathological conditions, e.g. anomalies, through continuous analysis of electrocardiography (ECG) signals over a period of time. Traditionally, this is done offline using a portable device called a Holter monitor [40], which records cardiac activity for 24 to 48 hours or more. However, manually analyzing these recordings can be slow and inefficient due to irregular and infrequent patterns in ECG signals.

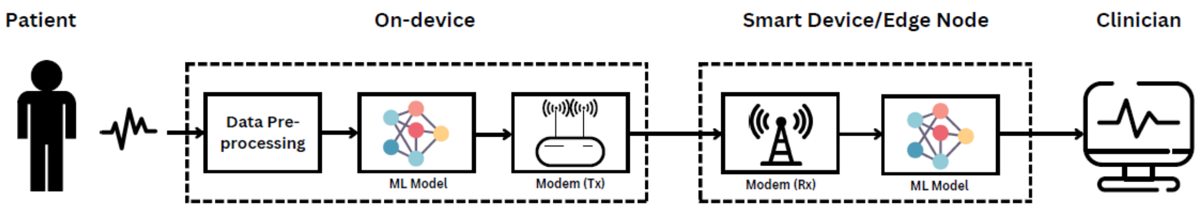


Figure 1.1: Concept of continous ambulatory cardiac monitoring system including ML operation splitting (on-deivce/edge)

To address this challenge, automated methods based on machine learning (ML) are increasingly used to assist clinicians in analyzing vast amounts of data effectively [31, 39]. The widespread availability of mobile communication technologies such as 4G and 5G networks enables seamless connectivity between patients and remote hospitals facilitating real-time and online monitoring. However, ensuring reliable wireless connectivity for conveying real-time sensor measurements remains challenging in practical settings. To enhance efficiency, communication overhead can be minimized using techniques like time-series compression [22] or asynchronous sampling [19]. Nevertheless, these methods may compromise signal quality upon reconstruction of the compressed signal at the receiver which may not be acceptable for clinical interpretation and could lead to decreased accuracy in ML-based analysis. Additionally, transmitting real-time ECG sensor data to hospitals for automated ML-based analysis may be inefficient from an information theory standpoint because relevant information (e.g. anomalies) are rare and therefore much of

the data has little value for the clinical application. Another approach that has gained popularity, particularly with the emergence of energy-efficient neuromorphic processing architectures, is to run inference models directly on the device. In this approach, only alerts corresponding to the defined applications [5] are transmitted. However, due to the limited computing capabilities of devices, these inference models are constrained, highlighting the necessity for energy-efficient solutions.
Future communication systems such as 6G will inherently incorporate artificial intelligence (AI) by embedding AI algorithms directly into the network infrastructure, including on-device. An illustrative example of such monitoring architecture is shown in Fig 1.1.
Consequently, the integration of edge and on-device intelligence becomes pivotal for future monitoring applications. Local pre-processing on sensor level close to the patient (potentially in conjunction with distributed ML functionalities) allows the split ML-operation between on-device and edge and communication can be performed highly efficient in an event-drive fashion.

1.2 Objectives

In this paper, we introduce a novel two-stage machine learning (ML) model designed for anomaly detection and classification with monitoring systems. The first stage utilizes a Multi-Layer Perceptron (MLP) model for initial binary classification on-device, followed by a CNN-LSTM model at the second stage for further classification of abnormal heartbeats. Our approach emphasizes computational efficiency achieved through reduced input sizes and parameters compared to baseline model [45]
This results in fewer multiplications, additions, and parameters. We evaluate the performance of our model against state-of-the-art methods and extend the existing framework introduced in [12], showcasing advancements in real-time anomaly detection and classification for health monitoring systems.

1.3 Outline

The organization of this thesis is as follows:

- **Chapter 1:** The motivation of this thesis, objectives, and outline are introduced.
- **Chapter 2:** A brief literature review of Electrocardiography and the role of artificial neural networks
- **Chapter 3:** A summary about related work that has contributed to the writing of this thesis.
- **Chapter 4:** The implementation of the proposed scheme is clearly explained
- **Chapter 5:** The experiments and results of the proposed model.
- **Chapter 6:** A conclusion is given with some ideas for future work on this topic.
- **Chapter 7:** list of tables and figures that is used throughout the thesis

Chapter 2

Background

2.1 Electrocardiogram (ECG)

Electrocardiography is the process of producing an electrocardiogram (ECG or EKG), which is a a recording of the heart's electrical activity through repeated cardiac cycles. It is an electrogram of the heart which is a graph of voltage versus time of the electrical activity of the heart using electrodes placed on the patient's skin [48]. The electrodes pick up the small electrical fluctuations resulting from the cardiac muscle depolarization and subsequent re-polarization of cardiac cycle with each heartbeat. The figure 2.1 [8] below shows an illustration of the cardiac cycle

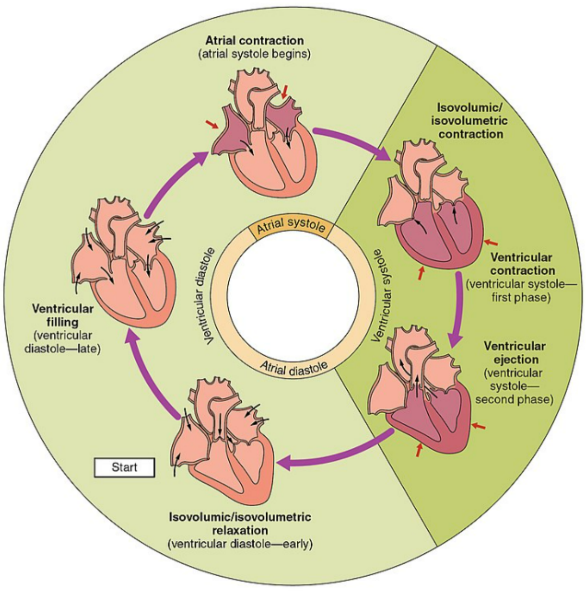


Figure 2.1: The cycle diagram depicts one heartbeat of the continuously repeating cardiac cycle

The cardiac cycle is the performance of the human heart from the beginning of one heartbeat to the beginning of the next. It comprises two distinct phases: diastole, where the heart muscle relaxes and fills with blood, followed by systole, characterized by strong contraction and blood pumping. Following the expulsion of blood, the heart then relaxes

and expands to accommodate the incoming blood from the lungs and body systems. Subsequently, it contracts again to circulate blood to the lungs and other parts of the body [47].
An ECG records these electrical impulses to show how fast the heart is beating, the rhythm of the heart beats (steady or irregular), and the timing of the electrical impulses as they move through different parts of the heart [21]. Changes in the normal ECG pattern can occur due to numerous cardiac abnormalities, which include:

- Cardiac rhythm disturbances (such as atrial fibrillation and ventricular tachycardia)
- Inadequate coronary artery blood flow (such as myocardial ischemia and myocardial infarction).
- Electrolyte disturbances, such as hypokalemia.

To capture these electrical impulses, special stickers called electrodes are placed on specific spots on the body. In a conventional 12-lead ECG, ten electrodes are placed on the patient's limbs and on the surface of the chest. The magnitude of the heart's electrical potential is then measured from twelve different angles called "leads" and is recorded over a period of time commonly lasting around ten seconds. Through this approach, both the overall magnitude and direction of the heart's electrical depolarization are systematically captured at each moment throughout the cardiac cycle.
The ECG signal shows six points of interest makred by letters of the alphabet: P, Q, R, S, T, and U as shown in Fig 2.2 [33]

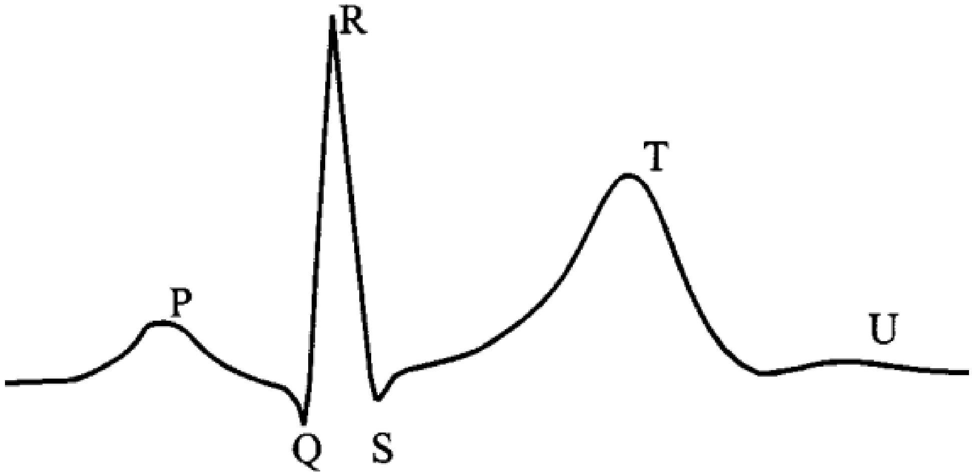


Figure 2.2: The typical ECG waveform

The size and shape of this signal contains valuable information about the nature of the disease and abnormalities affecting the heart.
The ECG consists of two phases: depolarization and re-polarization of the muscle fibers making up the heart. The depolarization phases correspond to the P-wave (atrial depolarization) and QRS-wave (ventricles depolarization). The re-polarization phases correspond to the T-wave and U-wave (ventricular re-polarization) [6]. These waves contribute to the heart beat impulse which defines electrocardiogram (ECG).

2.2 Real-time ECG monitoring

Previously, obtaining real-time ECG monitoring required the patients to visit hospitals equipped with 12-lead ECG machines [4]. These machines are typically fixed within hospital facilities. Patients would need to physically present at the hospital, where trained healthcare professionals would attach the electrodes and initiate the ECG recording process.
This approach had several limitations. Firstly, it restricted monitoring the patients' recording to the hospital premises, which in return limits the patients' mobility and hindering the detection of certain cardiac abnormalities that might occur outside the hospital premises. Secondly, it required the presence of a specialized healthcare professionals to operate the equipment and analyze the results. However, with advancements in technology and the development of portable monitoring devices such as Holter monitors, the vision for real-time ECG monitoring while the patients are in their home has been accessible. Holter monitors as shown in figure 2.3 [7] are capable of recording ECG signals over an extended period (typically 24 to 48 hours) before needing to be recharged. Moreover, modern Holter monitors often come equipped with software diagnostic tools that enable the real-time analysis of the recorder ECG data. This means that patients are no longer needed to be physically present in a hospital to undergo real-time ECG monitoring. Instead, they can wear portable Holter monitors allowing for convenient and continuous monitoring of cardiac health outside of hospital premises.

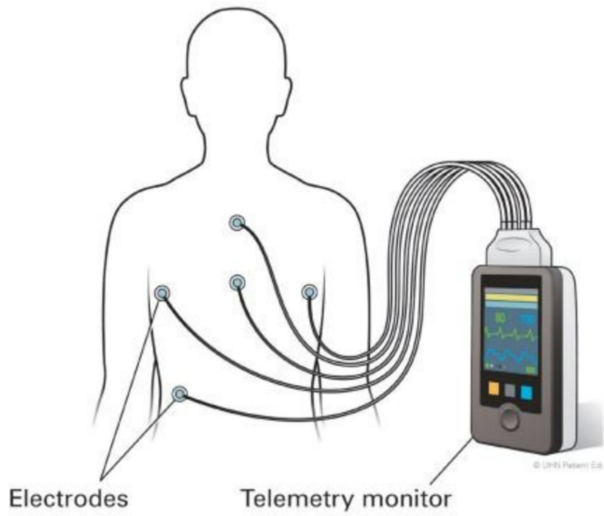


Figure 2.3: Illustration of real-time ECG monitoring

Real-time ECG monitoring opened the eyes for further improvement in technology to transmit analyzed heart beats data using wireless connectivity technologies such as BLE, Wi-Fi or internet from a patient's home to a specialized physician located at the hospital, facilitating remote monitoring and care [2].
Real-time ECG monitoring on an outpatient basis is known today as Mobile Cardiac Telemetry (MCT) [9]. Mobile Cardiac Telemetry is an outgoing form of outpatient ECG that monitors the patient 24 hours a day, for a period of up to 30 days. MCT is a relatively new form of cardiac monitoring which has been proven to be reliable and efficient method to monitor the patient's ECG information while the patient is performing his day-to-day activities and also during sleep.
MCT monitoring is one of a kind as the technology automatically detects and transmits ECG rhythm abnormalities to a remote diagnostic monitoring laboratory without any patient involvement as shown in fig 2.4 [18] .

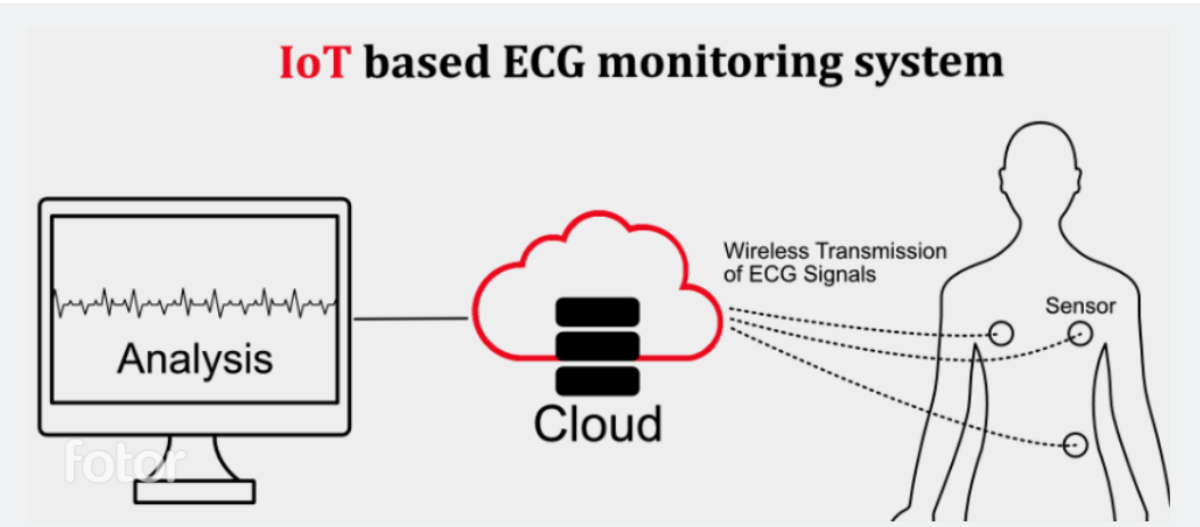
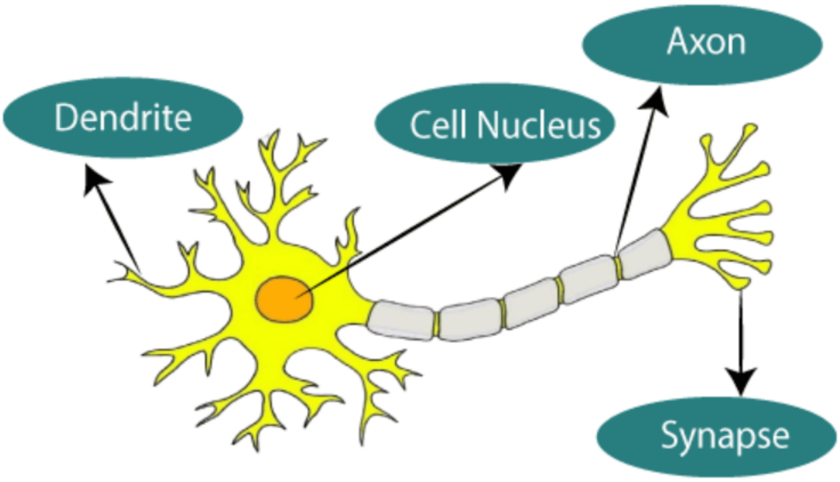


Figure 2.4: Detection and transmission of ECG data

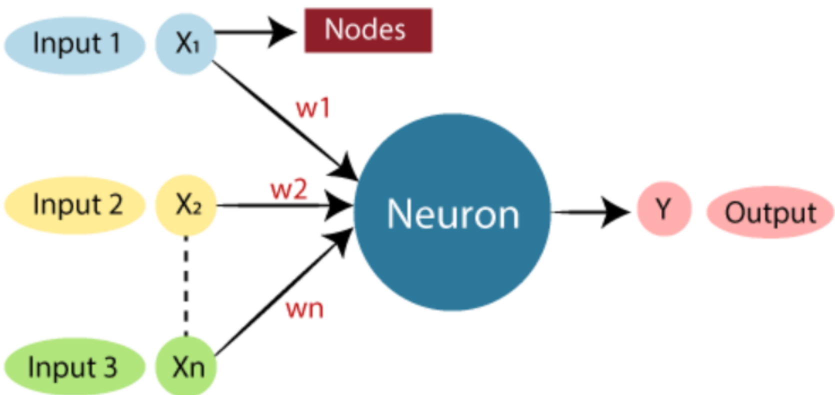
This is way advanced beyond any other forms of cardiac monitoring as this automated capability is executed in real time. The diagnostic laboratory that receives these abnormal ECG activities are well trained professional physician who can analyze the impulses and take immediate action.
Mobile Cardiac Telemetry (MCT) monitoring is now largely considered superior compared to other forms of Ambulatory Cardiac Monitoring primarily because Mobile Cardiac Telemetry (MCT) technology is capable of automatically transmitting the electrocardiograph data at the time of occurrence without patient involvement.

2.3 Artificial Neural Network (ANN)

In machine learning, a neural network also known as artificial neural network or neural net is a model inspired by the structure and design of biological neural network in human brains [49]. An ANN consists of connected units or nodes called artificial neurons, which in some way mimic the neurons in a brain. These neurons are connected by edges, which model the synapses in a brain. Each artificial neuron receives signals from connected neurons, then processes them and sends a signal to other conned neurons. This signal is a real number and the output of each neuron is computed by some non-linear function of the sum of its inputs which is called activation function. During the learning process, the strength of the signal at each connection is determined by a weight that adjusts during the learning process. Fig 2.5 [20] illustrates the typical diagram of Biological Neural network and Artificial Neural Network.



(a) Typical diagram of Biological Neural Network



(b) Typical diagram of Artificial Neural Network

Figure 2.5: Biological Neural Network vs Artificial Neural Network

The Dendrites from Biological Neural Network represent inputs in Artificial Neural Network, cell nucleus represents Nodes, synapse represents Weights, and Axon represents Output. The table 2.1 below shows the relationship between biological neural network and artificial neural network.

Table 2.1: Relationship between Biological and Artificial Neural Networks

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell nucleus	Nodes
Synapse	Weights
Axon	Output

2.3.1 Artificial Neural Networks Architecture

The Neural Network architecture is made of individual units called neurons that mimic the biological behavior of the brain as illustrated in fig 2.5b. The neuron comprises of various components including [42]:

- Input: It is the set of features that are fed into the model for the learning process. For example, the input in imagine detection can be an array of pixel values pertaining to an image.
- Weight: A pivotal role of weight in neural networks is to prioritize features that significantly aid in learning. This is achieved through scalar multiplication between input values and the weight matrix.
- Transfer function: its sole purpose is to combine multiple inputs into one output value so that the activation function can be applied. It is done by a simple summation of all inputs to the transfer function.
- Activation Function: It introduces non-linearity in perceptrons to consider varying linearity with the inputs. Unless an activation function is implemented, The output would simply result from a linear combination of input values lacking the ability to introduce non-linearity within the network.
- Bias: It's role is to shift the value produced by the activation function. It acts the constant in a linear function.

When multiple neurons are stacked together in a row, they form a layer, and multiple layers added up next to each other are called multi-layer neural network. The Fig 2.6 [32] below illustrates the main components of multi-layer neural network.

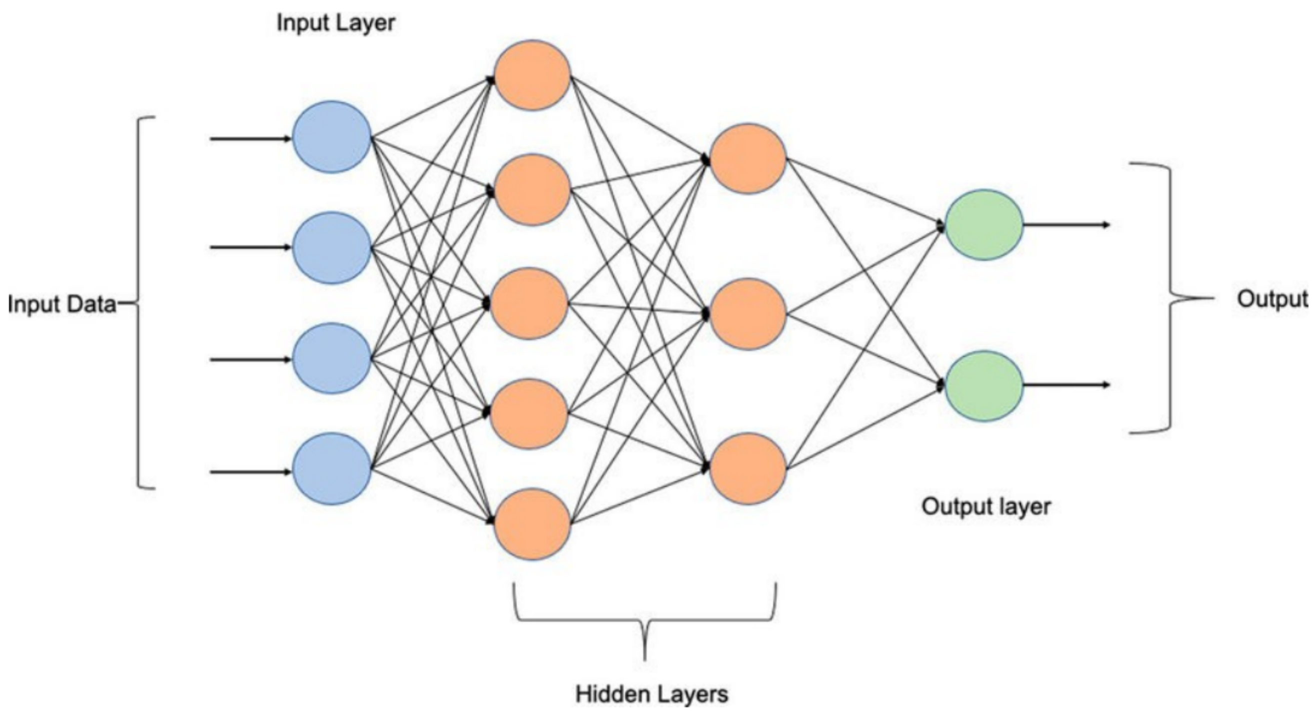


Figure 2.6: Multi-Layer Neural Network

- Input Layer: It is the data that we feed to the model. It is the only visible layer in the complete Neural Network architecture that passes the complete information from outside world without any computation.
- Hidden Layers: They are the key component of what makes deep learning what it is today. They are intermediate layers that do all the computation and extract all the features from the data. Multiple interconnected hidden layers can be utilized to detect various hidden features within the data. For instance, in image processing, initial hidden layers focus on extracting basic features such as edges, shapes, and boundaries, while subsequent layers tackle more complex tasks like recognizing complete objects such as cars, buildings, or people.
- Output Layer: It comes to a final prediction based on the model's learnings after it takes the inputs from the preceding hidden layers. Usually, a different type of activation function at the output layer to address the probability of occurrence of the output nodes.

This movement from the input layer to the output layer is called forward propagation, However there's another movement from the output layer to the input layer which is called backward propagation as shown in fig 2.7 [10] of which the model actually learns about the data. Back Propagation calculates the derivatives of the weights and biases for all layers of the model in respect to a defined loss function and uses an optimization technique called "gradient descent" to eventually find the optimum value for the weights and biases to result in the least amount of prediction errors and high accuracy gains [16].

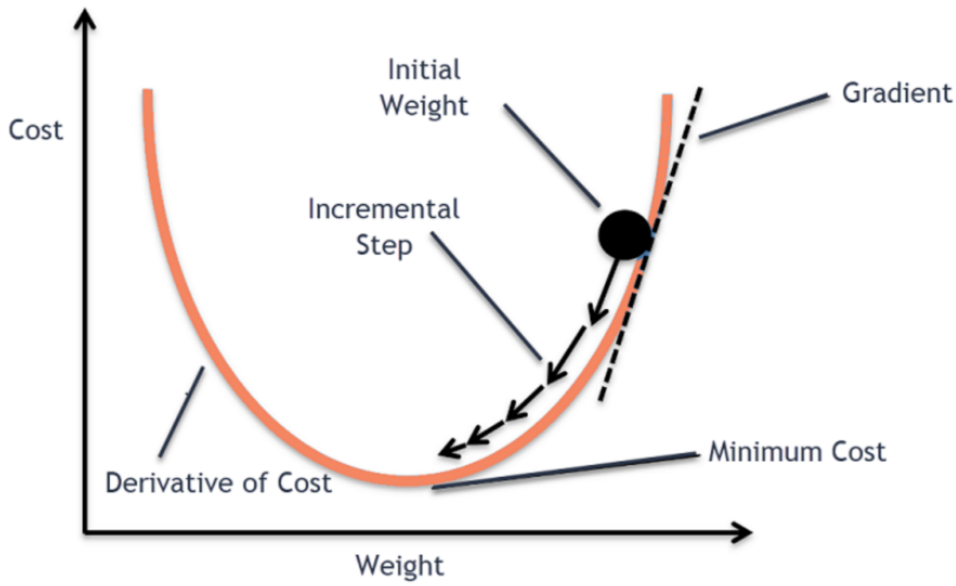


Figure 2.7: Illustration of Gradient Descent

2.3.2 Supervised vs Unsupervised Learning

Within artificial intelligence (AI) and machine learning, there are two basic approaches in distinguishing the type of learning for the machine learning model: supervised learning and unsupervised learning. The main difference is that one uses labeled data to ease prediction of outcomes, while the other does contain labeled data. However, there are some nuances between the two approaches and key areas in which one outperforms the other [11].
Supervised learning is a methodology in machine learning characterized by its reliance on labeled datasets. These datasets are designed to train or supervise algorithms into classifying data or predicting outcomes accurately. Through the use of labeled inputs and corresponding outputs, the model can assess its performance and iteratively improve its predictive capabilities.
Supervised learning can be separated into two types of categories, which include:

- Classification: Classification problems use an algorithm to accurately assign test data into specific categories, such as whether an email is span or not, or whether a medical image shows a tumor or not. Classification algorithms learn a function that maps the input features to a probability distribution function over the output categorized classes [17]. Linear classifiers, support vector machines, decision trees, random forest and deep neural networks are all common types of classification algorithms.

- Regression: Another type of supervised learning method that uses an algorithm to understand the relationship between dependent and independent variables. They are very handy in predicting numerical values based on different data points. Regression algorithms learn a function that maps from the input features to the output value. Linear regression, logistic regression and polynomial regression are some popular regression algorithms.

On the other hand, Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. The goal of unsupervised learning is to discover patterns and relationships in the data without any explicit guidance.
Unsupervised learning models are categorized into three main categories, which include:

- Clustering: It is a data mining technique for grouping unlabeled data based on their similarities or differences. For example, The K-means clustering algorithm categorizes data points with similarities into groups, using the parameter K to determine the size and detail of these groupings. This method finds utility in tasks such as market segmentation and image compression.

- Association: It is another type of unsupervised learning that uses different rules to find relationships between the variables of the given dataset. These methods are frequently used for market basket analysis and recommendation engines, such as people that buy X also tend to buy Y.

- Dimensionality Reduction: It is a method employed in machine learning when datasets have an excessively large number of features or dimensions. Its sole purpose is to decrease the number of data inputs to a manageable level while maintaining the integrity of the data. This technique is frequently applied during the data preprocessing phase, such as when auto-encoders are utilized to eliminate noise from visual data, thus enhancing image quality.

Fig 2.8 [36] illustrates an example between supervised learning and unsupervised learning.

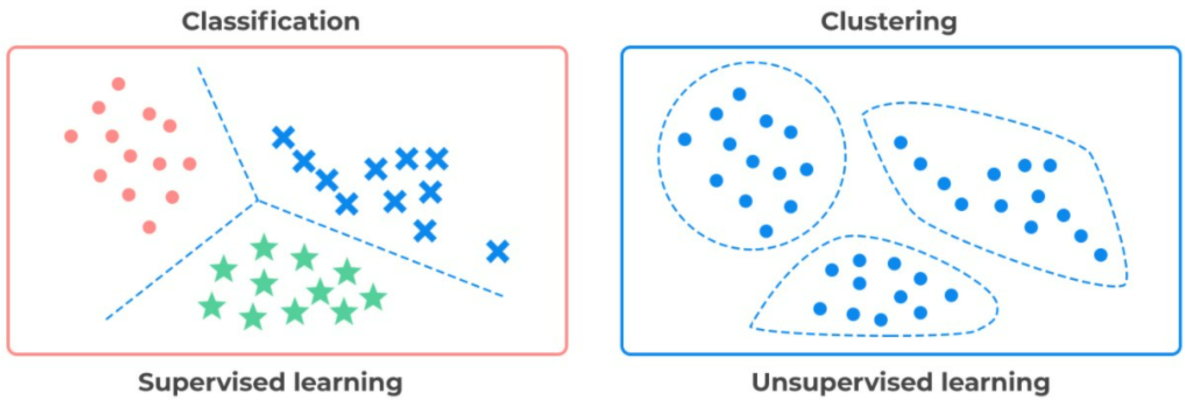


Figure 2.8: Supervised and Unsupervised Learning

2.3.3 Types of Artificial Neural Networks Architecture

Neural networks are an efficient way to solve machine learning problems and can be used in various situations. Neural networks offer precision and accuracy. Finding the correct neural network for each project can increase efficiently [46]. Here are some types of artificial neural networks architectures that has been used in this thesis. They include the following [25]:

- Multi-layer Perceptron (MLP) : also known as classical neural networks is a neural network architecture designed to estimate any continuous functions and can solve non-linear spearable problems. Multi-layer Perceptron consists of 3 main layers, which are named input layer, hidden layer, and output layer as shown in fig 2.9 [25]

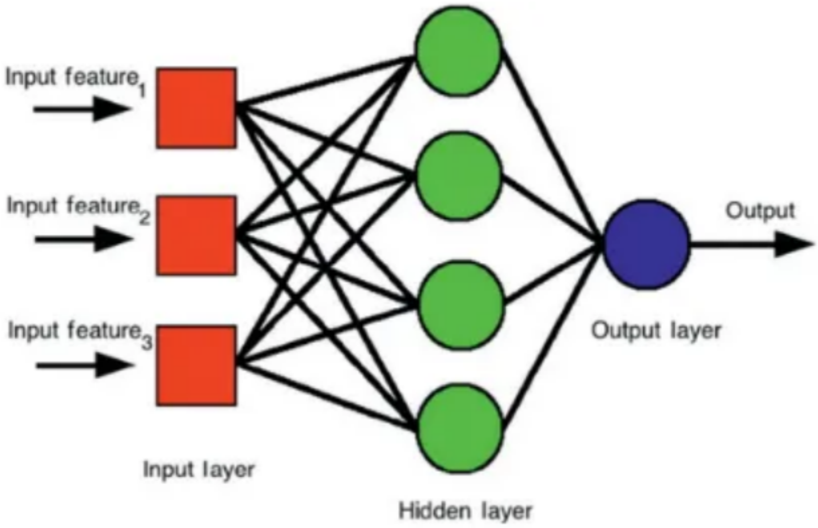


Figure 2.9: Multi-layer perceptron

  The input layer receives input signals and passes them to the hidden layer, which acts as the primary computation unit in the MLP. It receives input from the input layer and forwards it to the output layer, or vice versa during backpropagation. The output layer then makes predictions or classifications based on the input data received from the hidden layers.
  The multi-layer perceptron is widely used in several cases, such as pattern classification, recognition, prediction and finally approximation.

- Convolutional Neural Network (CNN): They are neural network architectures that are designed to solve compute vision problems such as image classification or image detection. This neural network consists of 3 main layers, which are named the convolutional layer, pooling layer, and fully connected layer as shown in fig 2.10 [25].
  The convolutional layer serves as a fundamental component within CNN architecture, comprising multiple layers depending on the complexity of the problem to be solved. In this layer, the initial connection isn't made directly to every pixel in the input image; rather, it's linked solely to receptive features within the image. This first layer primarily focuses on detecting basic features like colors, lines, and

edges. Additionally, a pooling layer reduces the input image's size from the convolutional layer, enhancing computational efficiency. Finally, the fully connected layer establishes connections between every neuron in one layer to all neurons in another layer.

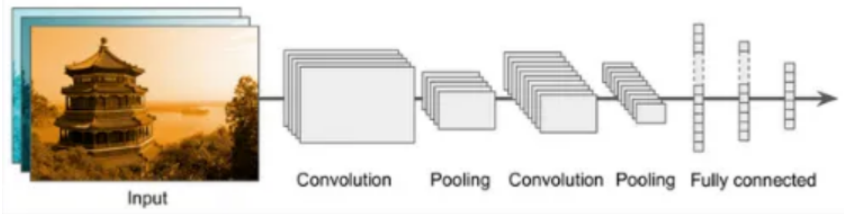


Figure 2.10: Convolutional Neural Network

There are many applications of convolutional neural networks that are part of uor daily lives, such as object detection in newer self-driving cars, face recognition in airports and image analysis in the medical field which is used in this thesis.

- Recurrent Neural Networks (RNN): They are neural network architecture that works very well in handling sequential data such as natural language processing (NLP) and time series data. Generally, the structure of recurrent neural networks is almost the same as multi-layer perceptron, the difference is that there are recurrent layers as shown in fig 2.11 [25].

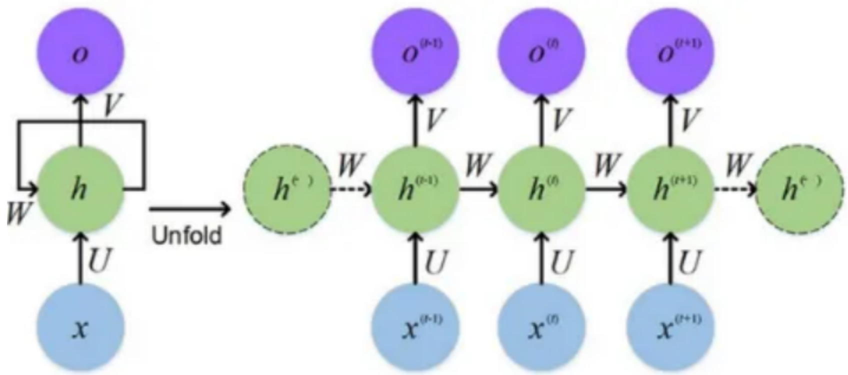


Figure 2.11: Recurrent Neural Network

The recurrent layer takes input, generates output, and feeds this output back into itself. However, when dealing with long sequences, recurrent neural networks can face issues with unstable gradients. To address this challenge, we replace the neuron cells in recurrent neural networks with specialized Long Short-Term Memory (LSTM) cells and Gated Recurrent Unit (GRU) cells, which are designed to handle such problems.
There are many applications of recurrent neural networks that are part of our daily lives, such as machine translation, text summarization, document generation and chatbots.

2.3.4 The use of AI in ECG analysis

Artificial intelligence (AI) has significantly enhanced electrocardiography (ECG), presenting a promising avenue for the accurate diagnosis and treatment of cardiovascular conditions [30]. While conventional ECG is widely used due to its affordability and accessibility, However the ECG can be interpreted differently by humans depending on the interpreter's level of training and expertise posing challenges for diagnosis. Leveraging artificial intelligence particularly deep learning convolutional neural networks (CNN) to analyze individual ECG leads continuously has resulted in fully automated models capable of interpreting ECGs similar to human experts, if not more accurately and consistently.
These artificial intelligence algorithms serve as effective non-invasive biomarkers for cardiovascular diseases, because they can identify difficult patterns and signals in the ECG that may not be readily apparent to expert physicians.
The integration of AI into ECG analysis offers numerous advantages, including swift, accurate and precise detection of conditions like arrhythmias, silent cardiac illnesses, and left ventricular failure. Moreover, AI-driven ECG analysis aids healthcare professionals interpretation, diagnosis, risk assessment, and illness management. Furthermore, AI-enhanced ECGs have demonstrated efficacy in enhancing the identification of heart failure and other cardiovascular disorders, particularly in emergency department settings, facilitating prompt and accurate treatment decisions.
Although the use of artificial intelligence in ECG classification has huge impact over the industry it also suffers from several limitations and obstacles. The effective implementation of AI-powered ECG analysis is limited by several issues such as systematic bias. These biases are based on age, gender, and race can result from unbalanced datasets, which the AI models actually learn from before being implemented into real life use. Potentially, disregarding age-related, physiological differences and genetic variations can have a huge decline in ECG readings accuracy. To guarantee a safe and successful deployment in clinical facilities and real-time ECG monitoring at homes, the use of AI in cardiology must be done with a thorough understanding of the algorithms and their limitations.

2.4 Neuromorphic Computing

As Moore's law reaches its limits and Dennard scaling ceases to provide performance gains, the computing industry is turning its attention to alternative technologies to sustain ongoing advancements in performance. Neuromorphic computing represents one such emerging technology [37].
Neuromorphic computers are defined as computing systems that diverge from the traditional Von Neumann architecture. Instead, they draw inspiration from the structure and function of biological brains, which utilizes neurons and synapses as their fundamental components. In contrast, Von Neumann computers are composed of separate CPUs and memory units, where data and instructions are stored. On the other hand, In a neuromorphic computer integrate both processing and memory functions withtin the neurons and synapses themselves. Fig 2.12 [43] illustrates the difference between Von Neumann Processor and Neuromorphic Processor.

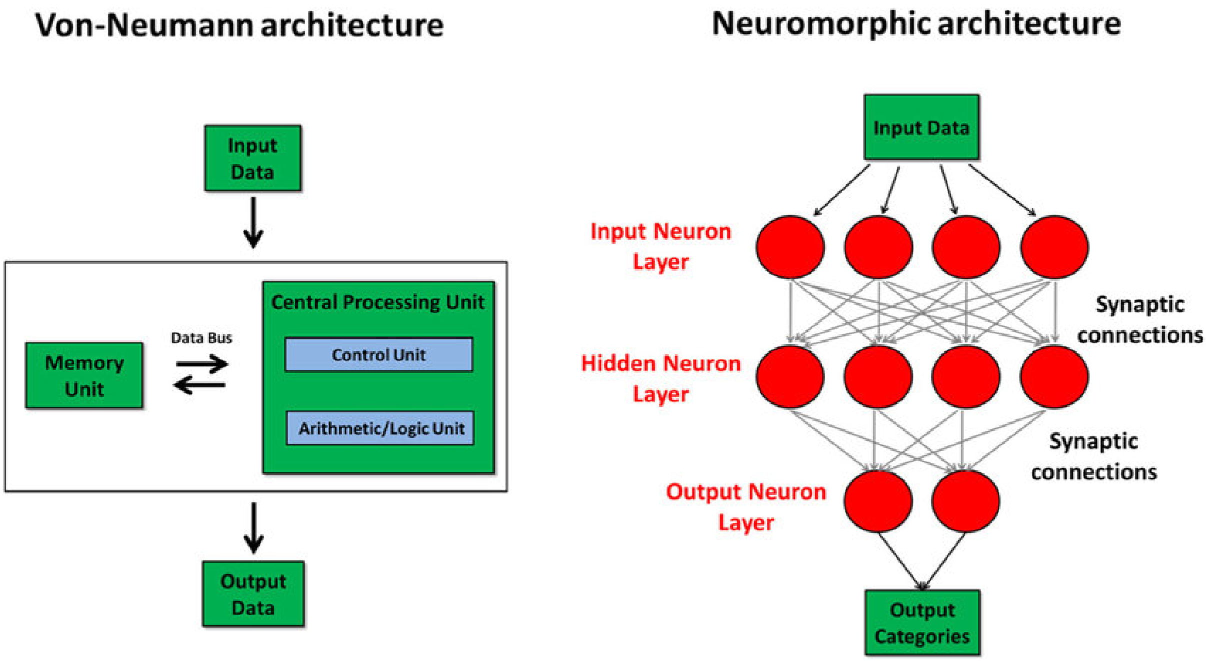


Figure 2.12: Von Neumann vs Neuromorphic Processors

Programs in neuromorphic computers are characterized by the configuration of neural networks and their parameters, rather than explicit instructions as in a Von Neumann computer. Furthermore, while Von Neumann computers encode information using numerical values represented in binary form, neuromorphic computeres receive spikes as input, where the associated time at which they occur, their magnitude and their shape can be used to encode numerical information. While it is possible to convert binary values into spikes and vice versa, the precise methodology for this conversion remains an active area of research within neuromorphic computing.

One of the key advantages of neuromorphic systems is that they are high energy efficient, as they can perform complex computations using relatively little power. This makes them particularly suitable for tasks demanding real-time processing and minimal power usage, such as robotics, autonomous systems, and sensor networks.

Some examples of real-life applications of neuromorphic processors include:

- Natural Language processing
- Brain-computer interfaces
- Signal processing and pattern recognition
- Computer vision and image processing
- Autonomous robots and drones

Chapter 3

Related Work

3.1 Real-time ECG Monitoring

This paper [45] has made a significant contribution to the development of my thesis by providing foundational insights into ECG Monitoring solutions.
In this paper, They briefed about how a patient can identify their heart problem as early as possible with intelligent wearable devices that can carry out data to health-care professional for further analysis. They introduced a low-power and energy-efficient ECG monitor scheme with two-stage end-to-end neural network diagnosis-based adaptive compression. Their workflow included pre-processing the data of Denoising, R-Peak Detection and Heart Beat segmentation. After, pre-processing is done they introduced a proposed two-stage neural network (TSNN) which includes a binary classifier (MLP) for classification of heartbeat segments either be normal or abnormal then a multi-classifier (CNN) for further classification of the abnormality type if detected. They also proposed a unique way of biased training to remove any obstacles that may limit the classification accuracy by modifying the loss function for back propagation. Then, they compared their model's performance with state-of-the-art methods through performance metrics like accuracy, specificity, sensitivity and predictivity. They also compared their results in terms of Number of Multiplications and Additions to estimate their power consumption levels. My thesis aims to extend and refine the context of more power-efficient neural network models with much higher-accuracy.
In this paper [53], the authors also introduced a two-stage neural network model. They classified the heartbeats according to the Association for the Advancement of Medical Instrumentation (AAMI). The binary classifier which is a one dimensional CNN model classifies the heart beat to either be normal or abnormal, then a four-classes CNN model is designed to further classify the abnormal ECG signals. This paper didn't include the normal type in the multi-classifier which can account to the faulty detection in the binary-classifier for the normal heartbeat segments. They also introduced the conversion of the four-classes CNN to a spiking neural network (SNN) to further reduced the power consumption. This conversion is not implemented in our thesis as it results in lower-accuracy numbers which doesn't meet the standards.
In this paper [15], the authors proposed an efficient way to build and train a deep SNN for ECG classification by constructing a counterpart structure of deep ANN, transferring the trained parameters, and replacing the activation functions with leaky integrate-and-fire

(LIF) neurons. The dataset of the paper is from 2017 PhysioNet/CinC Challenge which contains 8,528 single-lead ECG records with varying lengths. Their proposed ANN constituted of 14-layer deep convolution layers with average pooling layers included which will all be converted to a 14-layer SNN network to further reduced the power-consumption. They compared their results with state-of-the-art methods in terms of accuracy, precision, recall and F1. They concluded that ReLu activation function provides the most stable ANN-to-SNN transformation in terms of performance change.

3.2 AI in ECG

In this paper [23], a workflow for the design of small, energy-efficient recurrent convolutional neural network (RCNN) architecture for atrial fibrillation (AFIB) detection is proposed. The model consists of a small segmented CNN in combination with an optimal energy classifier. The architectural decisions are made by using the energy consumption as a metric in an equally important way as the accuracy. The optimization steps are focused on the software which can be embedded afterwards on a physical chip. The authors used the dataset from Bundesministerium fÃ¼r Bildung und Forschung (BMBF) which consists of 16,000 ECG signals, 8,000 with AFIB and 8,000 control cases of sinus rhythm. The new architecture of (RCNN) is combining the output of the current segment with information by the previous segments to make decisions. Using segments in such a recurrent manner allows them to have a temporal dynamic behavior which indeed reflects on the power-consumption of the model.This paper is concentrated on the transfer of the model on a simulated chip. The results show reduced power-consumption but low accurate results.
In this paper [34], the authors used the dataset from Massachusetts Institute of Technology arrhythmia database (MIT-BIH) to train their models on them. The authors compared three different types of artificial neural networks which are multi-layer perceptron (MLP), convolutional neural network (CNN), and spking neural network (SNN). They implemented the models on a center processing unit (CPU) platform, and deployed on a PYNQ-Z2 FPGA board to validate the model using a Jupyter Notebook. The results showed that CNN consumed more time for training and consumed most power compared to the other models but achieved the highest results in accuracy which was 95%. the MLP model showed accuracy only of 76% and the SNN was the most efficient ECG classification algorithm, which classified all four ECG arrhythmias with 90% accuracy while maintaining a low-to-average power consumption numbers.
In this paper [38], the authors implemented their dataset from the MIT-BIH database. They implemented different artificial neural network algorithms like CNN, LSTM, SVM, KNN, etc to perform the tasks of prediction, detection and classification. They implemented Fourier transform (FT) and wavelets transform (WT) as their pre-processing technique before feeding the data into the models. They compared their results with state-of-the-art models through accuracy, precision, recall and f1-score. Their results were outstanding compared to different papers. For the classification, SVM provided the best results for accuracy (around 99.83%); for prediction KNN provides the best results for accuracy (around 99.3%); for detection, CNN provides the best results for accuracy (around 99.23%).

In this Survey [26], the authors presented a comprehensive survey on the AF screening from a single lead ECG wave Which was motivated by the growing popularization of wearable devices that can collect single lead ECG conveniently and efficiently.
This Survey systematically surveys state-of-the-art methods for screening atrial fibrillation from a single lead ECG wave. The authors examined different types of pre-processing including: Noise Removal, Imbalance of dataset, Normalization and Feature extraction. The techniques for data pre-processing are reviewed and the most common and powerful features are listed, which are capable of providing a guideline for researchers aiming at developing AF detection algorithms. Then, they summarized different types of artificial neural network techniques for classification and detection of heart beats.

Chapter 4

Methodology

4.1 MIT-BIH Dataset

The ECG dataset utilized in our system model to train the neural network and assess its performance in our research is the MIT-BIH Arrhythmia Dataset [28]. This ECG dataset consists of 48 half-hour two-channel ECG recordings, obtained from 47 participants. The recordings were digitized at 360 samples per second per channel with an 11-bit resolution over a 10 mV range.

These recordings capture a wide range of arrhythmias, including premature ventricular contractions (PVCs), atrial fibrillation (AF), and supraventricular tachycardia (SVT), among others. Expert cardiologists have carefully annotated every beat in the recordings, offering labels indicating normal sinus rhythm as well as different types of arrhythmias. Moreover, these annotations contain precise details about beat onset and offset times, facilitating thorough analysis of ECG waveforms.

In our system model, we adopt the five commonly used labeling standard proposed by Association for the Advancement of Medical Instrumentation (AAMI) [3, 27] standard in 1998, These labels are classified into: N (normal Beat), F (Fustion Beat), SVEB (Supraventricular Ectopic Beat), VEB (Ventricular Ectopic Beat) and Q (unclassified beat). Table 4.1 summarizes the correspondence of AAMI and MIT-BIH standard [53].

Table 4.1: Classes recommended by AAMI

AAMI Classes	Symbol	MIT-BIH Classes
Normal	N	N,I,R
Supraventricular Ectopic Beat	SVEB	e,j,A,a,J,S
Ventricular Ectopic Beat	VEB	V,E
Fusion Beat	F	F
Unknown Beat	Q	/,f,Q

MIT-BIH dataset contains 90587 normal heartbeats, 2781 supraventricular ectopic beats, 7235 ventricular ectopic beats, 802 fusion beats and 8038 unclassified beats. Figure 4.1 illustrates the distribution of the five classes in the dataset.

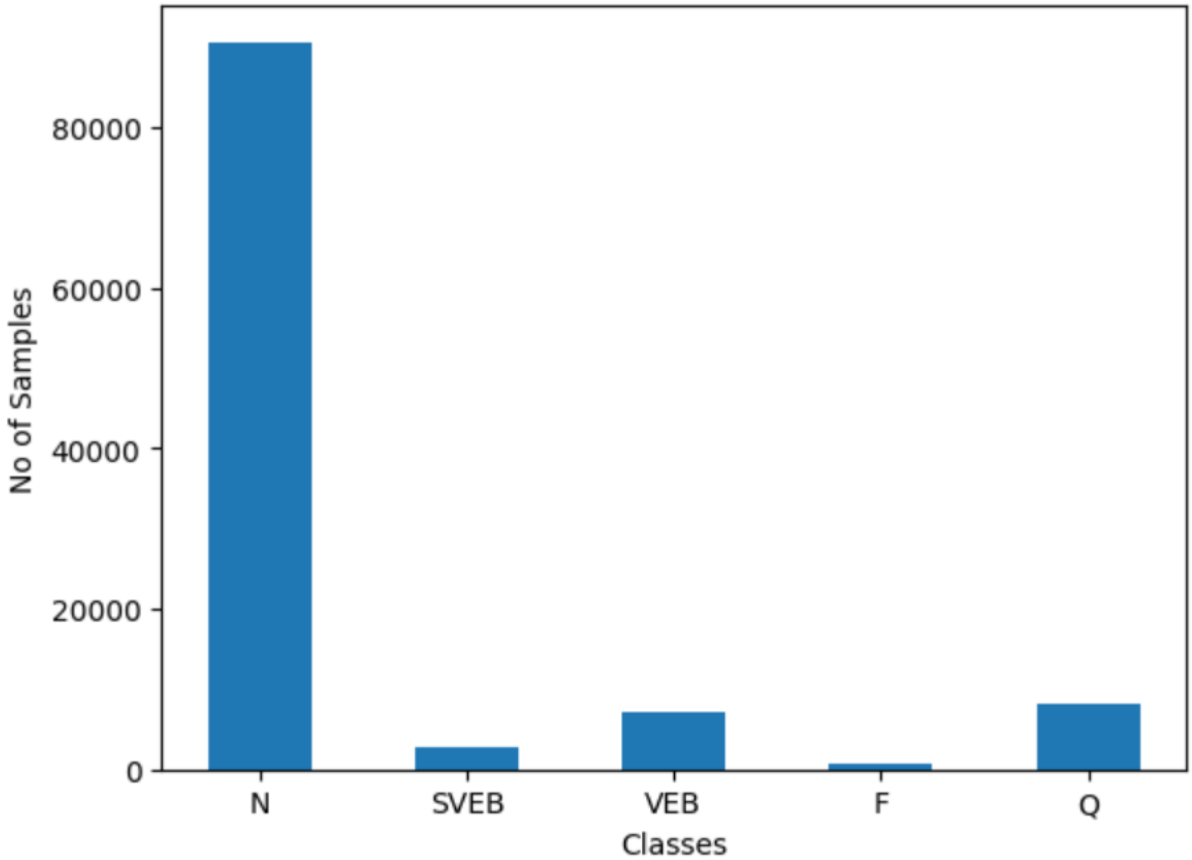


Figure 4.1: Distribution of MIT-BIH Five Heartbeat Classes

The classification of ECG beats (N,SVEB,VEB,F,Q) provides us with a valuable schematic for interpreting cardiac rhythm abnormalities. Each beat annotation offers valuable insights for training and assessing classification models. We rely on this dataset as our primary source of data for this study. Figure 4.2 illustrates the characteristics of the five mentioned heartbeat types.

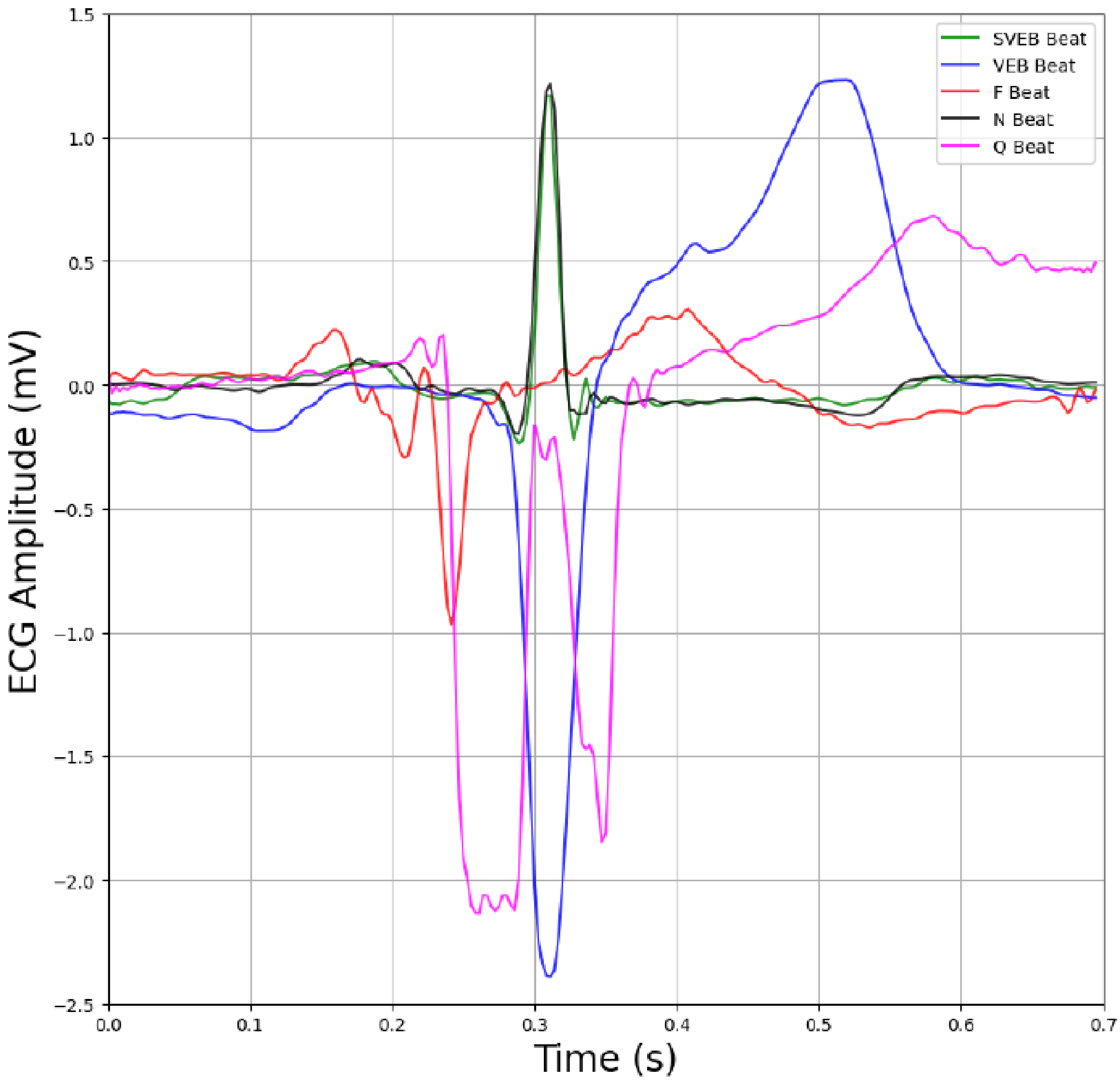


Figure 4.2: Example of Different Heartbeat Classes in the MIT-BIH Dataset

4.2 Preprocessing

Before feeding the data into a deep learning model, it's crucial to preprocess it to ensure that the model can comprehend and effectively learn from it. The MIT-BIH dataset is a clean dataset without any empty or Not a Number (NAN) entries. Nan entries are not understood by a machine learning model and thus should be removed before training a model.

The ECG data is preprocessed for training the ML model as follows; Firstly, noise reduction is applied to the ECG MIT-BIH recordings using wavelet transform, then R-peaks detection and identification from annotations as they contain the most valuable information of the heartbeats wave signal, after that Segments of ECG data are extracted around each R-peak (specifically 144 samples before and 180 samples after). These segments are then resampled to a fixed length of 251 samples.

4.2.1 Encoding

Deep learning models can only deal with numeric values, so we need non-numeric values to be encoded in some way in order to be understood by the model. In this study, encoding was done by One-hot encoding as it's the most common algorithm for addressing non-numeric data due to its simplicity [35].
One-hot encoding involves converting each classification label value into a new classification column and assigning a binary value of 1 or 0 to these columns. All the values in these columns are zero except for the position where the classification labels are present, which is set to 1 as shown in figure 4.3 [1]. Different encoding methods were experimented within this study, and after evaluation, one-hot encoding was selected due to its superior performance as it provided the best results.

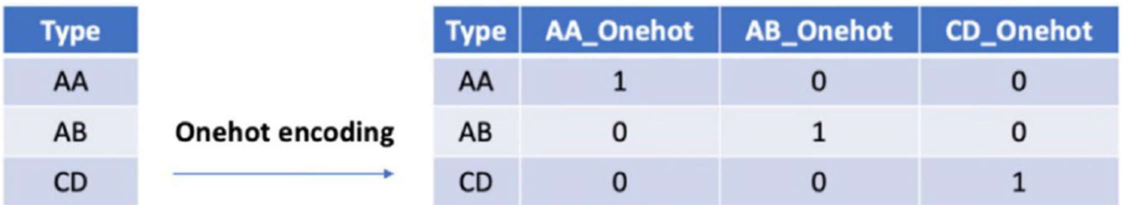


Figure 4.3: One-hot encoding Example

4.2.2 Class balancing

As shows from figure 4.1, the MIT-BIH suffers from the class imbalance problem. This appears in how the normal class comprises 82.77% , while the fusion class comprises only 0.73% of the dataset. This problem creates a false high accuracy because the model focuses solely on classifying the majority class, neglecting the minority ones. This is called an accuracy paradox [14] which is achieving high accuracies that does not reflect the performance of the model. This problem can be solved by using oversampling techniques. In this paper, various oversampling techniques has been tested. However, after careful consideration, we opted to employ a specific method known as SMOTE (Synthetic Minority Over-sampling Technique). The figure 4.4 illustrates how SMOTE addressed the imbalance issue, with each class comprising over 90,000 samples.

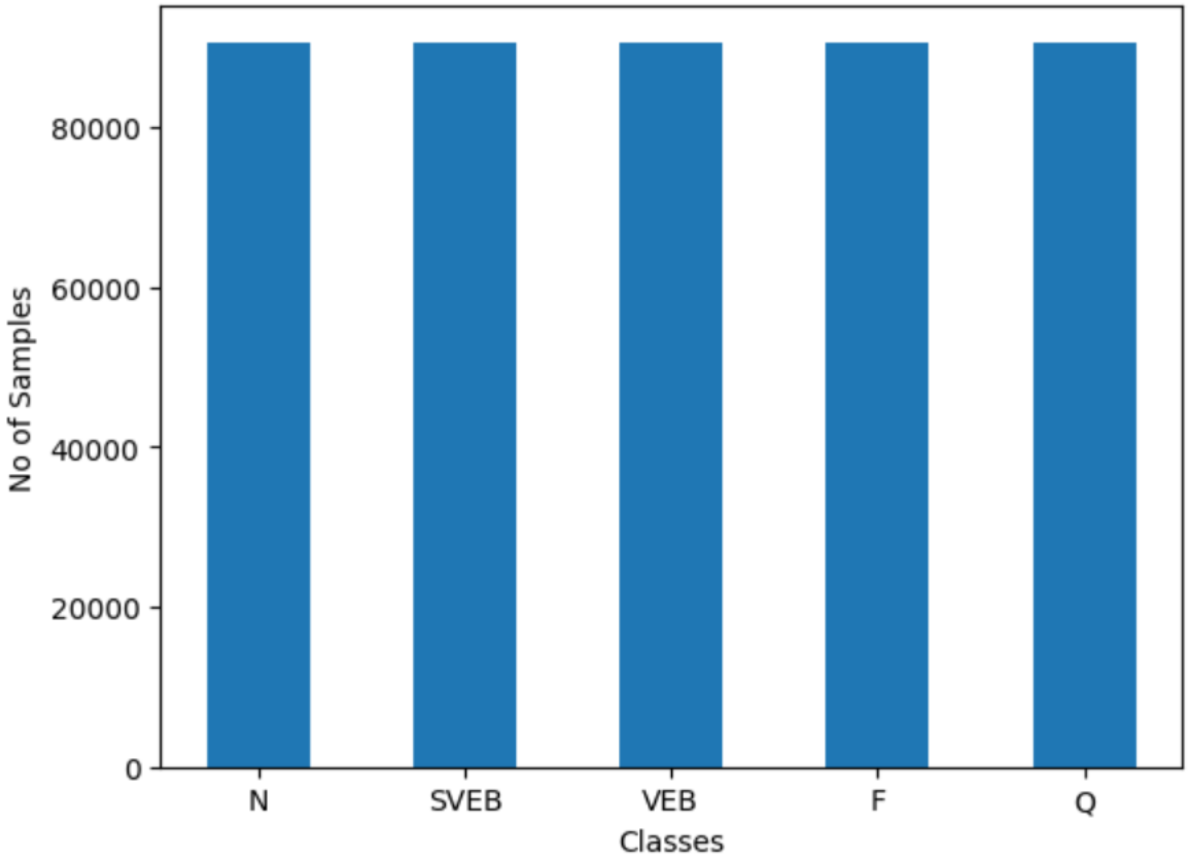


Figure 4.4: Distribution of samples in each class after oversampling

4.3 Proposed Two-Stage Architecture

In this paper, an energy-efficient low-power ECG two-stage neural network (TSNN) architecture is proposed inspired by the model presented in this paper [45]. Figure 4.5 represents the proposed system model.
After the ECG Signal has passed the preprocessing part which ends at the segmentation of the heartbeat signal, the ECG signal goes through a 5-category TSNN Classifier which consists of two parts: the first part is an anomaly detection model using multi-layer-perceptron (MLP) to classify normal and abnormal beats, then whenever an abnormal beat is detected, it is classified using a multi-classifier model consisting of Convolution Neural Network (CNN) combined with Long-Short-Term-Memory (LSTM) to further detect the type of abnormality in the signal.

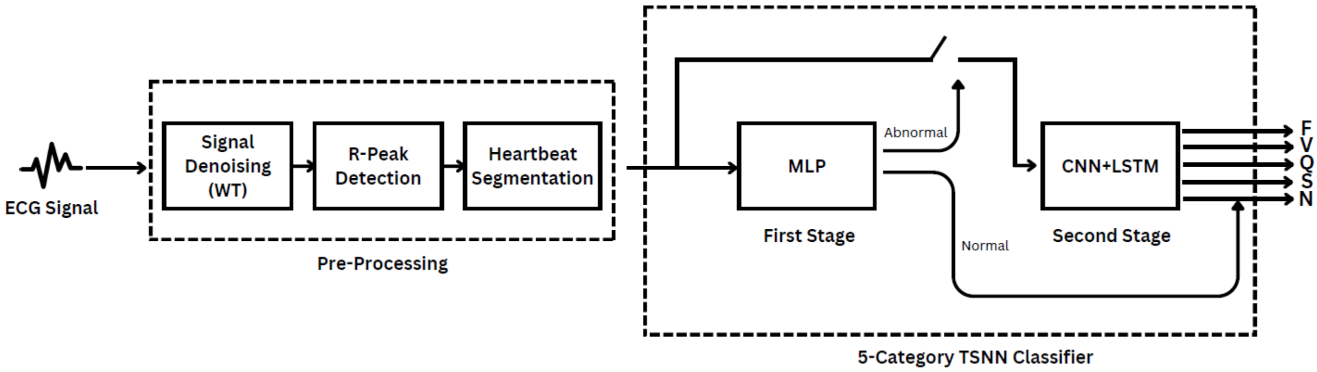


Figure 4.5: Proposed Model [Overall Architecture]

4.3.1 Proposed Biased Training

One challenge faced by the TSNN is that the accuracy of the MLP stage can impose a limitation on the overall classification accuracy of the TSNN. There are two scenarios to take into account. In the MLP stage, if the normal heart beat is mistakenly classified as abnormal, it still has the chance to be correctly by the multi-classifier stage. However, if an abnormal heartbeat is incorrectly classified as normal during the MLP stage, it will not proceed to the CNN stage, resulting in classification inaccuracies. To minimize the accuracy loss due to this dilemma, a biased training method is introduced. The basic idea is that we want to make the number of samples in the abnormal class more frequent than the number of samples in the normal class so that less abnormal heart beats are classified as normal heart beats while allowing more normal heart beats to be classified as abnormal, which will eventually get corrected by the multi-classifier.

4.3.2 Binary Classifier Model: MLP

In the first stage of the overall model, an MLP model is used to classify the heart beats into normal and abnormal beats. MLP models prove to have medium-to-low complexity with accurate classification of ECG signals which is emphasised by this review [13].
The MLP model is chosen for binary classification as it's the most optimal choice for our application as it compromises between accuracy and power consumption. The proposed MLP model, as shown in Table 4.2 , has a total of 30,905 trainable parameters, which is well-suited for deployment in energy-efficient, battery-operated devices.
For the biased training of this model, it is fed with ECG binary labeled data which contains 70% abnormal heart beats compared to normal beats which is 30%. Thus, increasing the number of normal heart beats to be classified into abnormal heart beats as shown in fig 4.6 so that it will be corrected in the second stage using CNN-LSTM model. This model achieves training and validation accuracy of 99.92% and loss of 0.0076 over 100 epochs as shown in the learning curves in fig 4.7

Table 4.2: MLP Model [First Stage]

Layer (type)	Neurons	Parameters
Input Layer	251	0
Hidden Layer	102	25,704
Hidden Layer	50	5,150
Output Layer	1	51
Total Parameters		30,905

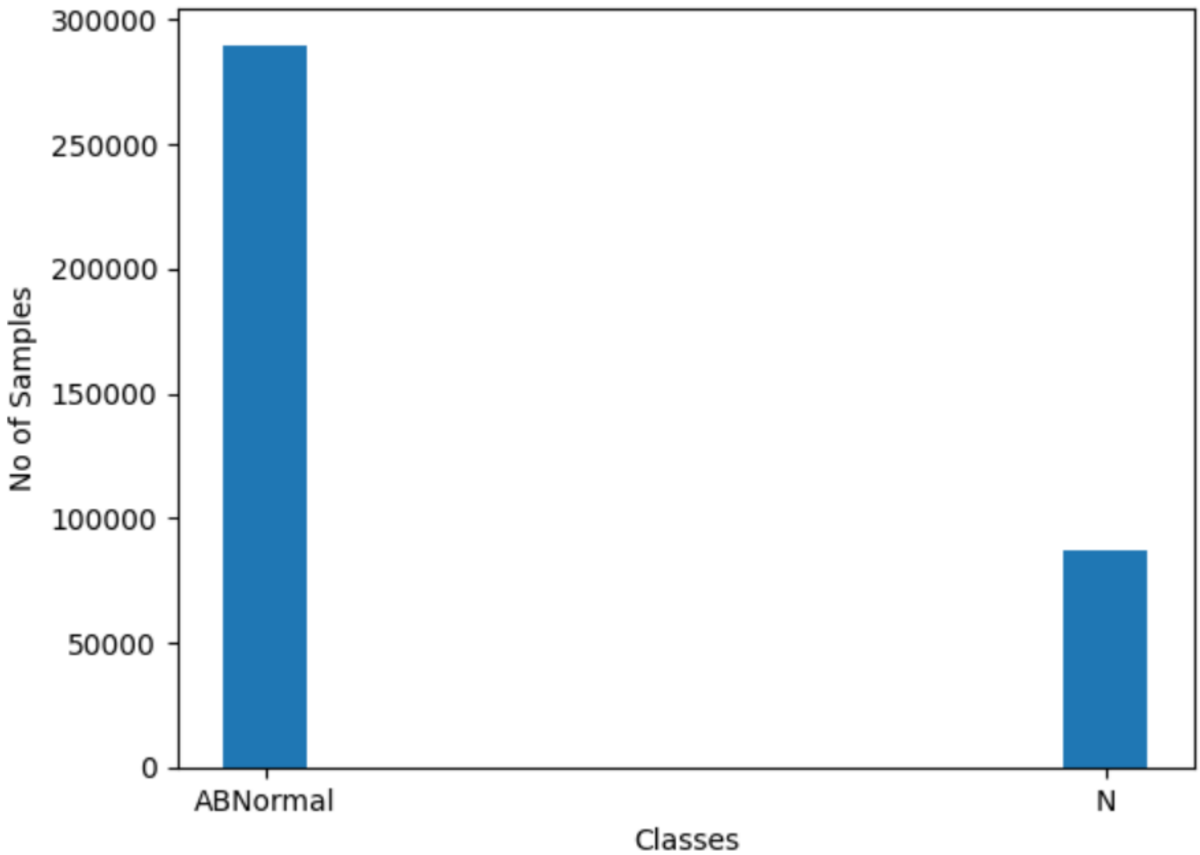
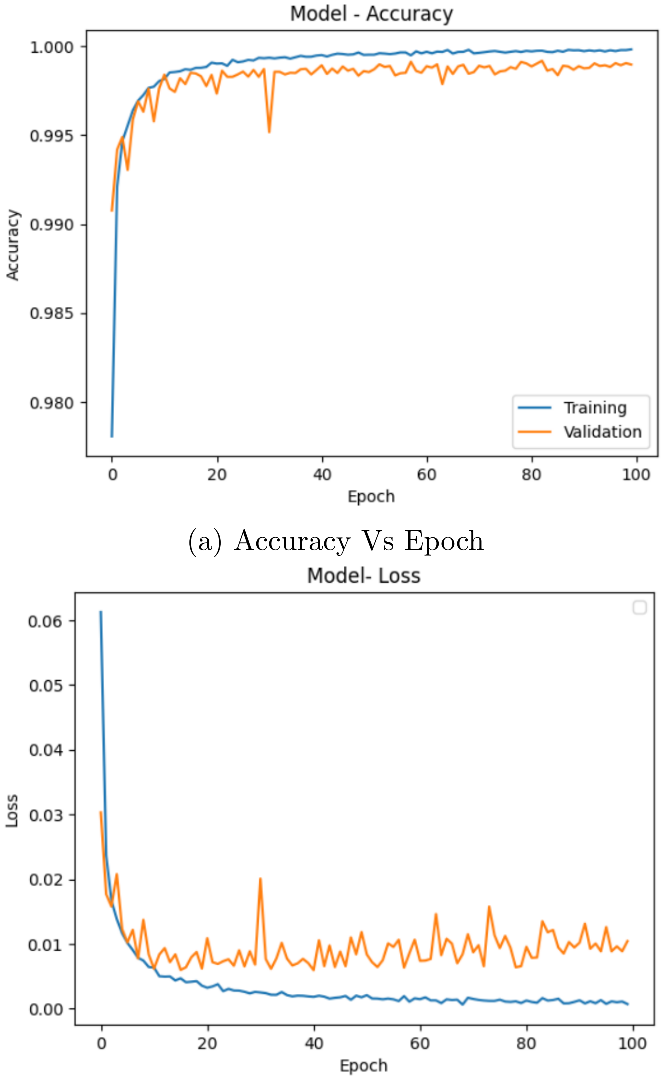


Figure 4.6: Distribution of samples in each class for Binary Model



(b) Loss Vs Epoch

Figure 4.7: Training and validation accuracy and loss of MLP first stage against Epoch

4.3.3 Multi-Classifier Model: CNN-LSTM

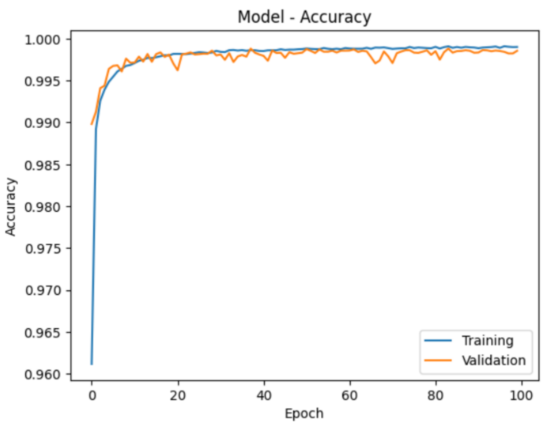
In ths second stage of the proposed model, a CNN-LSTM model is introduced to classify the ECG heartbeat segments into five classes. According to this study [13], CNN has shown outstanding performance in classifying different types of arrhythmia, however the downside that it has high computational complexity. On the other hand, LSTM models proves to have lower accuracy than CNN ones, yet the benefit is that they have medium-to-low computational complexity.

Accordingly, we proposed a CNN-LSTM combined model to achieve high classification accuracy while maintaining low computational power. Table 4.3 shows the detailed structure of the model.

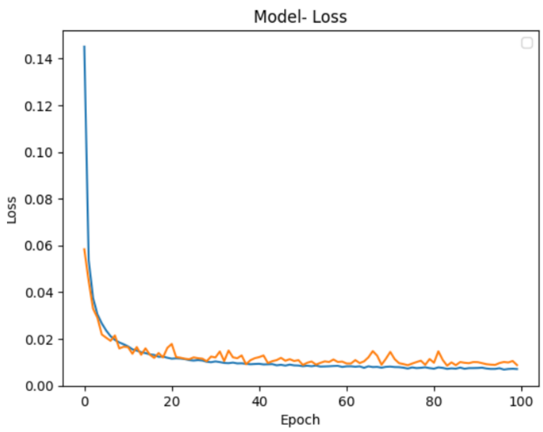
Table 4.3: Multi-Class Model: "CNN-LSTM"

Layer	Output Shape	Description
Input Layer	(251,1)	Input layer
Conv1D	(251,40) (125,40) (125,40)	Convolutional layer Max pooling Dropout
Conv1D	(125,60) (41,60) (41,60)	Convolutional layer Max pooling Dropout
Conv1D	(41,80) (10,80) (10,80)	Convolutional layer Max pooling Dropout
Bidirectional	(100,1)	Bidirectional LSTM
Dense	(5,1)	Dense layer
Total Parameters		79,385

In our proposed model, the second stage of CNN-LSTM model is only activated when an abnormal beat is detected from the first stage. For the training of this stage, the model was fed with balanced ECG heartbeat sampled data of 251 points per sample. The model comprises a total of 79,385 trainable parameters and attains a test accuracy of 99.88 after being trained for 100 epochs as shown in figure 4.8.



(a) Accuracy Vs Epoch



(b) Loss Vs Epoch

Figure 4.8: Training and validation accuracy and loss of CNN-LSTM second stage against Epoch

4.3.4 Performance Metrics

A commonly used tool when evaluating the performance and effectiveness of a supervised machine learning models is the confusion matrix. The confusion matrix is a matrix that summarizes the performance of a machine learning model based on a set of test data. It's a means of displaying the number of accurate and inaccurate instances. It's used to measure the performance of machine learning models, which aim to predict a categorical label for each input [17]. The confusion matrix is a table that shows information about the predicated classes vs the actual-true classes [44, 1].Some of the confustion matrix terminologies are:

- True positives (TP): The data instances correctly predicted to be positive
- True negatives (TN): The data instances correctly predicted to be negative
- False Positive (FP): The data instances wrongly predicted to be positive
- False Negative (FN): The data instances wrongly predicted to be negative

We use these terminologies from the confusion matrix to calculate parameters for evaluating accuracy and performance which include the following:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \tag{4.1}$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4.2}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{4.3}$$

$$\text{Predictivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4.4}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.5}$$

These parameters are calculated from the confusion matrix of each stage independently as shown in fig 5.1, and have contributed to the results in shown in Table 5.10

The first metric used and the most intuitive one is accuracy. Accuracy is used to measure the performance of the model. It's the ratoin of total correct instances to the total instances, which can be calculated as Eq. 4.1. However, when data is unbalanced, accuracy does not provide the most optimal evaluation of the model. Thus, we use additional metrics to evaluate the performance of the model. Those metrics are Sensitivity, Specificity, Predictivity and F1-score. Sensitivity also known as Recall evaluates how well a classification model can detect all relevant instances within a dataset. It's calculated by dividing the number of true positive (TP) instances by the total of true positive and

false negative (FN) instances, which can be calculated as Eq. 4.2. Specificity is another important metric, it measures the ability of the model to accurately identify negative instances. It is also known as the True Negative Rate, which can be calculated as Eq. 4.3. Predictivity also known as Precision assesses the accuracy of a model's positive predictions. It's determined by dividing the number of true positive predictions by the total number of positive predictions made by the model, which can be calculated as Eq. 4.4. Finally, F1-Score is the harmonic mean of the precision and recall, which is calculated as Eq. 4.5. [17]
According to these metrics, the objective is to maximize accuracy, sensitivity, specificity, predictivity and f1-score.

4.4 Integer Quantization

Integer quantization is a method used to optimize models by converting 32-bit floating-point numbers to the nearest 8-bit fixed-point numbers. This helps reduce model size and boosts inference speed, particularly beneficial for low-power devices like micro-controllers. Additionally, integer quantization is essential for integer-only accelerators like the Edge TPU [41].
In this thesis, we implemented post-training integer quantization to reduce the model's total computational complexity and reduce its size which means the conversion happened only after the model was fully trained and the weights was tweaked optimally to focus on computational accuracy and to reduce loss.
The full Integer Quantization was done for both the binary classifier model as well as the multi-classifier model. In the binary classifier model the quantization effect reduced the total model size by 91.23 % which is a huge upgrade for the reduction of the model's complexity but in return the accuracy dropped to only 70.26 %. On other other hand, the multi-classifier model was also affected by the integer quantization which reduced its total size by 89.01 % and the accuracy also dropped to 53.35 %.
Because the binary and multi models exhibited excessively low accuracy, the notion of implementing full integer quantization was abandoned in this thesis as the results doesn't meet the requirements to actively detect and classify heartbeat segments reliably.

Chapter 5

Experiments and Results

In this section, we discuss the setup for our experiments from various libraries and platforms then we represent the results of our proposed energy-efficient ECG classification model and compare them with state-of-the-art models in terms of accuracy, sensitivity, specificity, predictivity, and computational efficiency.

5.1 Experiment setup

The models were implemented using Tensorflow and Keras in a Jupyter Notebook platform. The hardware of the experiments is a Windows 10 with 12th Gen Intel(R) Core(TM) i7-12700H (CPU) and an Nvidia GeForce RTX 3060 (GPU). They python version used is 3.10.8. The dataset from the MIT-BIH is first divided into train and test dataset where the train occupies 80% and the test occupies 20% before feeding any data to both models. Then, the 80% train data is then shared among the binary classifier model and the multi-classifier model to be divided again 80% for training and 20% for testing results. After both models have done their training and testing, they both get tested again for the whole 20% that was divided early on to calculate the total overall accuracy of both models to simulate real-life working conditions.

5.2 Ablation Study

In this section, we discuss about the different models that have been tested for both the binary classifier model and the multi-classifier model in terms of their performance and computational cost.

5.2.1 Binary Classifier Model

In the binary classifier model three major models have been implemented and configured for the task of classification of heartbeats into normal and abnormal. These three models include:

Convolutional Neural Network (CNN):

Block	**Layer**	**Layer Size**	**Activation**	**Kernal**	**Stride**
Input	Input layer	251	-	-	-
Hidden Block 1	1D CNN layer	8	ReLu	4	1
	1D MaxPooling layer	2	-	-	1
Hidden Block 2	1D CNN layer	16	ReLu	6	1
	1D MaxPooling layer	2	-	-	1
Hidden Block 3	1D CNN layer	8	ReLu	6	1
	1D MaxPooling layer	2	-	-	
Hidden Block 4	1D CNN layer	16	ReLu	8	1
	1D MaxPooling layer	2	-	-	
Hidden Block 5	1D CNN layer	8	ReLu	8	1
	1D MaxPooling layer	2	-	-	
Hidden Block 6	1D CNN layer	16	ReLu	10	1
	1D MaxPooling layer	2	-	-	
Hidden Block 7	Dense layer	256	ReLu	-	-
Hidden Block 8	Dense layer	32	ReLu	-	-
Output	Output layer	1	Sigmoid	-	-

Table 5.1: CNN model Layers

Table 5.2: CNN Performance Metrics

Methods	Values
Accuracy	99.60 %
Loss	0.0309
No. of Mul.	854,224
No. of Add.	855,641
No. of Parameters	773,585
Precision	99.61 %
Sensitivity	99.86 %
Specificity	98.71 %
F1-Score	99.73 %

Multi-Layer Perceptron(MLP):

Table 5.3: MLP model Layers

Block	Layers	Number of Neurons	Activation
Input block	Input Layer	251	-
Hidden block 1	Dense layer	102	ReLu
Hidden block 2	Dense layer	50	ReLu
Output block	Output layer	1	Sigmoid

Table 5.4: MLP Performance Metrics

Methods	Values
Accuracy	99.91 %
Loss	0.0079
No. of Mul.	30,752
No. of Add.	30,905
No. of Parameters	30,905
Precision	99.90 %
Sensitivity	99.98 %
Specificity	99.69 %
F1-Score	99.94 %

The k-nearest neighbors (KNN): Where the number of n-neighbors is equal to three

Table 5.5: KNN Performance Metrics

Methods	Values
Accuracy	99.88 %
No. of Mul.	87,331,434
No. of Add.	87,331,183
Precision	99.74 %
Sensitivity	99.76 %
Specificity	99.92 %
F1-Score	99.75 %

The tables 5.1, 5.2, 5.3 and 5.4 illustrate the CNN and MLP model layers and their respective performances while table 5.5 illustrates the KNN model performance.
We concluded that the three models perform nearly identical to each other in terms of accuracy but the number of parameters, multiplications and additions in the KNN algorithm and CNN model are much bigger than the MLP model. Thus, enabling the MLP algorithm to be more suitable for binary classification.

5.2.2 Multi-Classifier Model

In the Multi classifier model two major models have been implemented and configured for the task of classification of the type of abnormality in the heartbeat. These two models include:

Convolutional Neural Network (CNN):

Block	**Layer**	**Layer Size**	**Activation**	**Kernal**	**Stride**
Input	Input layer	251	-	-	-
Hidden Block 1	1D CNN layer	32	ReLu	4	1
	1D MaxPooling layer	2	-	-	1
Hidden Block 2	1D CNN layer	16	ReLu	6	1
	1D MaxPooling layer	2	-	-	1
Hidden Block 3	1D CNN layer	32	ReLu	3	1
	1D MaxPooling layer	2	-	-	
Hidden Block 4	1D CNN layer	16	ReLu	5	1
	1D MaxPooling layer	2	-	-	
Hidden Block 5	Dense layer	128	ReLu	-	-
Hidden Block 6	Dense layer	64	ReLu	-	-
Output	Output layer	4	Softmax	-	-

Table 5.6: CNN model Layers

Table 5.7: CNN Performance Metrics

Methods	Values
Accuracy	99.42 %
Loss	0.0387
No. of Mul.	340,624
No. of Add.	341,572
No. of Parameters	268,452
Precision	99.40 %
Sensitivity	99.40 %
Specificity	99.80 %
F1-Score	99.40 %

CNN-LSTM:

Block	Layer	Layer Size	Activation	Kernal	Stride
Input	Input layer	251	-	-	-
	1D CNN layer	40	Tanh	2	1
Hidden Block 1	1D MaxPooling layer	2	-	-	1
	Dropout layer	0.1	-	-	-
	1D CNN layer	60	Tanh	3	1
Hidden Block 2	1D MaxPooling layer	3	-	-	1
	Dropout layer	0.1	-	-	-
	1D CNN layer	80	Tanh	4	1
Hidden Block 3	1D MaxPooling layer	4	-	-	
	Dropout laye	0.1	-	-	-
Hidden Block 4	Forward BLSTM layer	50	Tanh	-	-
	Backward BLSTM layer	50	Tanh	-	-
Output	Output layer	4	Softmax	-	-

Table 5.8: CNN-LSTM model Layers

Table 5.9: CNN-LSTM Performance Metrics

Methods	Values
Accuracy	99.88 %
Loss	0.0092
No. of Mul.	56,200
No. of Add.	56,622
No. of Parameters	79,385
Precision	99.88 %
Sensitivity	99.88 %
Specificity	99.97 %
F1-Score	99.88 %

Tables 5.6 and 5.7 illustrates the CNN model's layers and computational cost, while the tables From the two models we concluded that the CNN-LSTM model shows superior performance while maintaining relatively low number of parameters for computational efficiency.

5.3 Model Results

Table 5.10 illustrates the performance metrics between the binary classifier model (MLP) and the multi-classifier model (CNN-LSTM).

Table 5.10: Performance Comparison of MLP and CNN-LSTM Models

Metric	MLP Model	CNN-LSTM Model
Accuracy (%)	99.91 %	99.88%
Sensitivity (%)	99.98%	99.88%
Specificity (%)	99.69%	99.97%
Predictivity (%)	99.90%	99.88%
F1 (%)	99.94%	99.88%

The confusion matrix of the deployed machine learning models (MLP + CNN-LSTM) of which the parameters of accuracy, sensitivity, specificity, predictivity, f1-score are calculated from are illustrated in fig 5.1

5.4 Performance Comparison

Table 5.11 presents a comparative analysis of our proposed approach against several state-of-the-art methods concerning classification performance. We assessed the performance metrics on both ventricular ectopic beats (VEB) and supraventricular ectopic beats (SVEB), as well as overall accuracy. Our model demonstrates strong performance across all metrics, surpassing several established methods in accuracy, sensitivity, and specificity.

Table 5.11: Comparison Between Our Work and State-of-the-Art Methods

Method	VEB				SVEB				Overall Accuracy (%)
	Acc	Sen	Spe	Ppr	Acc	Sen	Spe	Ppr	
TBME2015 [24]	99.0	93.9	98.9	90.6	97.6	60.3	99.2	63.5	96.5
MEMEA2018 [29]	-	97.0	-	97.0	-	96.0	-	75.0	97.6
EMBC2018 [52]	99.6	98.8	99.6	95.5	99.1	92.7	99.3	80.2	-
ACCESS2019 [50]	99.9	99.6	99.9	98.9	99.3	99.1	99.8	94.1	98.6
TSNN w/ BT2019 [45]	99.1	91.8	99.6	95.3	99.4	79.5	99.9	96.3	98.4
RTSNN2021 [51]	99.7	97.2	99.8	97.5	99.6	86.6	99.9	97.3	99.1
Proposed TSNN	99.9	99.6	99.9	99.9	99.9	99.8	99.9	99.9	99.8

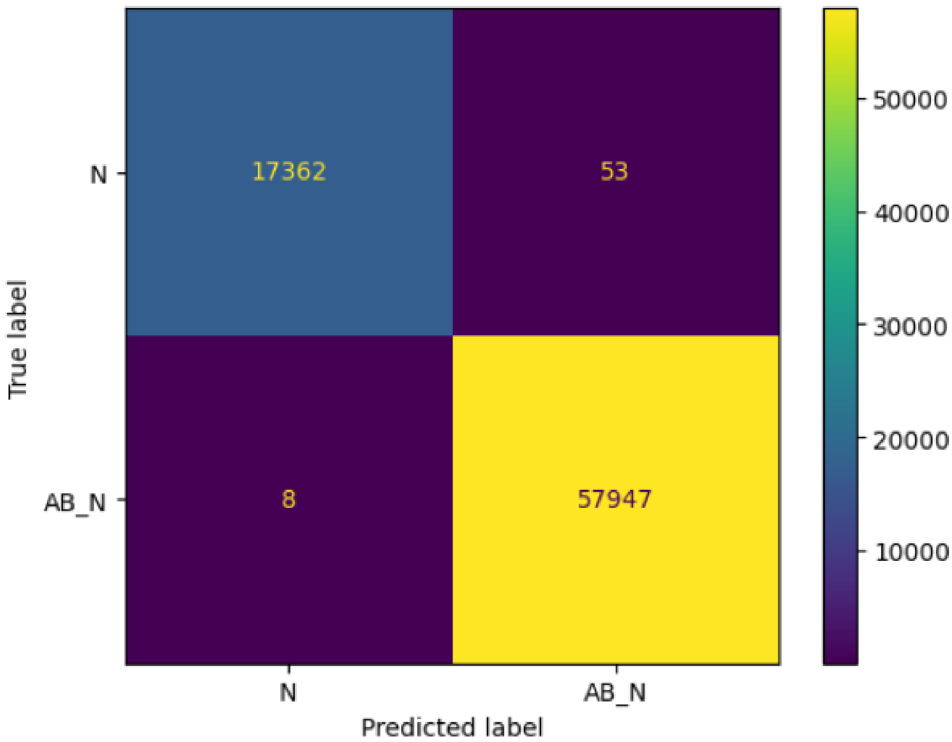
5.5 Computational Efficiency

To evaluate the computational efficienty of our model, we conducted a comparison with serveral baseline models introduced in [45], focusing on the number of multiplications, additions, and parameters. Table **??** represents a summary of these metrics, considering a scenario where 90% of the heartbeats are normal and only 10% are abnormal. Our model demonstrates promising computational efficiency, showing significantly reduced computational demands compared to traditional methods. This improvement was achieved in the MLP model by reducing the input size (e.g. ECG segments) and augmenting the dataset using the SMOTE up-sampling technique during the training stage. In the CNN-LSTM stage, efficiency enhancements were realized by integrating a CNN model with LSTM, resulting in reduced computational requirements without compromising accuracy and precision.
Overall, our findings highlight the effectiveness and computational efficiency of the proposed model for energy-efficient ECG classification.

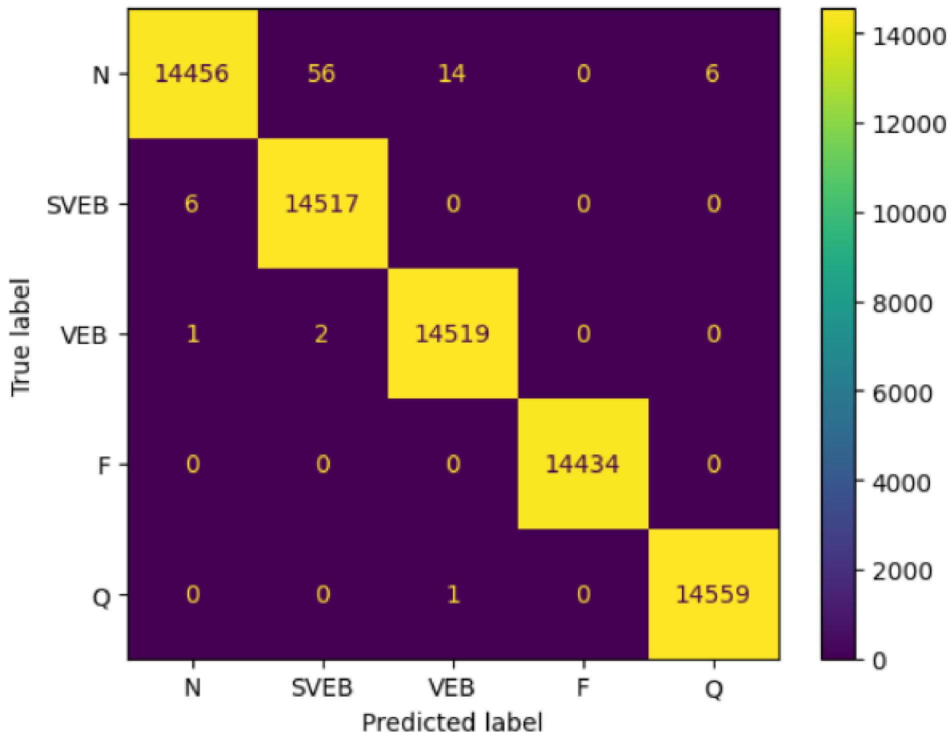
Table 5.12: Comparison of Number of Multiplications, Additions, and Parameters

Metric	TSNN2019 [45]			Proposed		
	MLP	CNN	Total Effective	MLP	CNN+LSTM	Total Effective
No. of Mul.	90,050*	667,600*	156,810*	30,752	56,200	36,372
No. of Add	90,301*	669,269*	157,228*	30,905	56,622	36,567
No. of Parameters	90,352	107,685	198,037	30,905	79,385	110,290

*All numbers are estimated based on the model introduced in the paper [45].



(a) Confusion matrix of MLP



(b) Confusion matrix of CNN-LSTM

Figure 5.1: Confusion Matrices of the Deployed ML Models

Chapter 6

Conclusion

6.1 Summary

In this thesis, we introduce an innovative and energy-efficient system for classifying Electrocardiogram (ECG) signals using advanced Machine Learning (ML) techniques. Our system consists of a carefully designed two-stage neural network architecture aimed at improving the analysis of long-term cardiovascular activity, especially through ECG signals.

Our system's core is a binary classifier designed to detect anomalies which is operating seamlessly on the device. This initial stage is followed by a multi-classifier for detailed classification of the anomaly's type at the edge. By integrating on-device processing and edge computing, we aim to reduce computational load while enabling real-time monitoring which is a significant advancement in healthcare technology.

Evaluating our model against the MIT-BIH [28] Arrhythmia Database highlights its competitive performance compared to existing methods. The combination of on-device processing and edge computing not only enables real-time analysis but also reduces computational complexity, making our model efficient and effective.

The proposed approach demonstrates promising results in accuracy, sensitivity, specificity, and computational efficiency, making it suitable for practical real-life applications.

6.2 Future Work

Future research endeavors will concentrate on enhancing the system for use in resource-constrained environment such as wearable devices. This includes reducing computational complexity and minimizing memory usage to ensure efficient functioning on low-power devices. Additionally, exploring integration with upcoming communication technologies like 6G networks will be pursued to facilitate seamless connectivity and real-time monitoring features.

List of Tables

List of Figures

Bibliography

[1] Ahmed Abdelkhalek and Mohamed Mashaly. "Addressing the class imbalance problem in network intrusion detection systems using data resampling and Deep Learning". In: *The Journal of Supercomputing* 79.10 (2023), pp. 10611–10644. DOI: 10.1007/s11227-023-05073-x.

[2] Anuhya Ardeti and et al. "Development of real-time ECG monitoring and unsupervised learning classification framework for cardiovascular diagnosis". In: *Biomedical Signal Processing and Control* 88 (2024), p. 105553. DOI: 10.1016/j.bspc.

[3] N Association for the Advancement of Medical Instrumentation et al. "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms". In: *ANSI/AAMI EC38* 1998 (1998), p. 46.

[4] Ahmed Badr and et al. "XBeats: A real-time electrocardiogram monitoring and Analysis System". In: *Signals* 3.2 (2022), pp. 189–208. DOI: 10.3390/signals3020013.

[5] F.C. Bauer, D.R. Muir, and G. Indiveri. "Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor". In: *IEEE Transactions on Biomedical Circuits and Systems* 13.6 (2019), pp. 1575–1582. DOI: 10.1109/tbcas.2019.2953001.

[6] L. Biel et al. "ECG analysis: a new approach in human identification". In: *IEEE Transactions on Instrumentation and Measurement* 50.3 (2001), pp. 808–812. DOI: 10.1109/19.930458.

[7] L. Brewer. *Mobile Cardiac Telemetry Systems Market with Tremendous Growth by 2029*. 2023. URL: https://www.linkedin.com/pulse/mobile-cardiac-telemetry-systems-market-tremendous-growth-brewer-2p6mf/ (visited on 05/14/2024).

[8] *Cardiac Cycle*. https://www.wikiwand.com/en/Cardiac_cycle. Accessed: 14 May 2024.

[9] CardiacMonitoring.com. *Real Time ECG Monitoring*. 2022. URL: https://cardiacmonitoring.com/mobile-cardiac-telemetry/real-time-ecg-monitoring/#:~:text=Real%20Time%20ECG%20Monitoring%20on,of%20up%20to%2030%20days.

[10] Crypto1. *Gradient Descent Algorithm: How Does it Work in Machine Learning?* https://www.analyticsvidhya.com/blog/2020/10/how-does-the-gradient-descent-algorithm-work-in-machine-learning/. Accessed: 02 May 2024. 2024.

[11] J. Delua. *Supervised vs. Unsupervised Learning: What's the Difference?* https://www.ibm.com/blog/supervised-vs-unsupervised-learning/. Accessed: 02 May 2024. 2021.

[12] J. Dommel, F. Hassan, and Z. Utkovski. “An energy-efficient and robust transmission scheme for IoT-based physiological activity monitoring”. In: *2022 IEEE Globecom Workshops (GC Wkshps)*. [Preprint]. 2022. DOI: 10.1109/gcwkshps56602.2022.10008531.

[13] Z. Ebrahimi et al. “A review on deep learning methods for ECG Arrhythmia Classification”. In: *Expert Systems with Applications: X* 7 (2020), p. 100033. DOI: 10.1016/j.eswax.2020.100033.

[14] W. Elmasry, A. Akbulut, and A. H. Zaim. “Empirical study on multiclass classification-based network intrusion detection”. In: *Computational Intelligence* 35.4 (2019), pp. 919–954. DOI: 10.1111/coin.12220.

[15] Y. Feng et al. “Building and training a deep spiking neural network for ECG classification”. In: *Biomedical Signal Processing and Control* 77 (2022), p. 103749. DOI: 10.1016/j.bspc.2022.103749.

[16] A. G. Ganie and S. Dadvandipour. “From Big Data to Smart Data: A Sample Gradient Descent Approach for Machine Learning”. In: *Journal of Big Data* 10.1 (2023). DOI: 10.1186/s40537-023-00839-9.

[17] GeeksforGeeks. *Confusion matrix in machine learning*. 2024. URL: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/ (visited on 04/27/2024).

[18] V. Gupta. *IoT Based ECG Monitoring System for Remote Healthcare*. PsiBorg Technologies Pvt. Ltd. 2023. URL: https://psiborg.in/iot-based-ecg-monitoring-system-for-remote-healthcare/.

[19] Qisong Hu et al. “Asynchronous communication for wireless sensors using Ultra Wideband Impulse Radio”. In: *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*. [Preprint]. 2015. DOI: 10.1109/mwscas.2015.7282170.

[20] javatpoint. *Artificial Neural Network tutorial - javatpoint*. Accessed: May 2, 2024. URL: https://www.javatpoint.com/artificial-neural-network (visited on 05/02/2024).

[21] Johns Hopkins Medicine. *Electrocardiogram*. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electrocardiogram. Accessed on: April 30, 2024. 2019.

[22] S.B. Kale and D.H. Gawali. “Review of ECG compression techniques and implementations”. In: *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*. [Preprint]. 2016. DOI: 10.1109/icgtspicc.2016.7955376.

[23] N. Katsaouni et al. “Energy efficient convolutional neural networks for arrhythmia detection”. In: *Array* 13 (2022), p. 100127. DOI: 10.1016/j.array.2022.100127.

[24] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. “Real-time patient-specific ECG classification by 1-D Convolutional neural networks”. In: *IEEE Transactions on Biomedical Engineering* 63.3 (2016), pp. 664–675. DOI: 10.1109/TBME.2015.2468589.

[25] D. Kurniawan. “Several types of artificial neural networks architecture that you should know”. In: *Medium* (2023). Accessed: 02 May 2024. URL: https://dedekurniawann.medium.com/several-types-of-artificial-neural-networks-architecture-that-you-should-know-c169a5e22ec7.

[26] Y. Liu et al. “Survey on atrial fibrillation detection from a single-lead ECG Wave for Internet of Medical Things”. In: *Computer Communications* 178 (2021), pp. 245–258. DOI: 10.1016/j.comcom.2021.08.002.

[27] Eduardo JosÃ© da S. Luz et al. “ECG-based heartbeat classification for arrhythmia detection: A survey”. In: *Computer Methods and Programs in Biomedicine* 127 (2016), pp. 144–164. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2015.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0169260715003314.

[28] G. B. Moody and R. G. Mark. “The impact of the MIT-BiH Arrhythmia Database”. In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 45–50. DOI: 10.1109/51.932724.

[29] B. Murugesan and et al. “ECGNet: Deep Network for Arrhythmia Classification”. In: *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE. 2018. DOI: 10.1109/MeMeA.2018.8438739.

[30] M. A. Muzammil et al. “Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases”. In: *Journal of Electrocardiology* 83 (2024), pp. 30–40. DOI: 10.1016/j.jelectrocard.2024.01.006.

[31] K. Nezamabadi et al. “Unsupervised ECG analysis: A Review”. In: *IEEE Reviews in Biomedical Engineering* 16 (2023), pp. 208–224. DOI: 10.1109/rbme.2022.3154893.

[32] K. NG. *Multi-layer Perceptron*. 2023. URL: https://www.linkedin.com/pulse/multi-layer-perceptron-kishore-ng-wvrcc/ (visited on 05/14/2024).

[33] A. Rana and K.K. Kim. “A novel spiking neural network for ECG Signal Classification”. In: *Journal of Sensor Science and Technology* 30.1 (2021), pp. 20–24. DOI: 10.46670/jsst.2021.30.1.20.

[34] A. Rana and K.K. Kim. “Comparison of artificial neural networks for low-power ECG-Classification System”. In: *Journal of Sensor Science and Technology* 29.1 (2020), pp. 19–26. DOI: 10.5369/jsst.2019.29.1.19.

[35] Pedro Rodríguez et al. “Beyond one-hot encoding: Lower dimensional target embedding”. In: *Image and Vision Computing* 75 (2018), pp. 21–31. DOI: 10.1016/j.imavis.2018.04.004.

[36] S. Saini. *Supervised vs. Unsupervised Learning: Whatâs the Difference?* 2021. URL: https://www.linkedin.com/pulse/supervised-vs-unsupervised-learning-whats-difference-smriti-saini/ (visited on 05/14/2024).

[37] Catherine D. Schuman, Subhasish R. Kulkarni, Moin Parsa, et al. “Opportunities for neuromorphic computing algorithms and applications”. In: *Nat Comput Sci* 2 (2022), pp. 10–19. DOI: 10.1038/s43588-021-00184-y.

[38] H. Serhal et al. “Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG”. In: *Computers in Biology and Medicine* 142 (2022), p. 105168. DOI: 10.1016/j.compbiomed.2021.105168.

[39] N. Strodthoff et al. “Deep learning for ECG analysis: Benchmarks and insights from PTB-XL”. In: *IEEE Journal of Biomedical and Health Informatics* 25.5 (2021), pp. 1519–1528. DOI: 10.1109/jbhi.2020.3022989.

[40] L. Su, S. Borov, and B. Zrenner. “12-lead Holter Electrocardiography”. In: *Herzschrittmachertherapie + Elektrophysiologie* 24.2 (2013), pp. 92–96. DOI: 10.1007/s00399-013-0268-4.

[41] TensorFlow. *Post-training integer quantization*.

[42] V7. *The Essential Guide to Neural Network Architectures*. Accessed: May 2, 2024. URL: https://www.v7labs.com/blog/neural-network-architectures-guide (visited on 05/02/2024).

[43] Javier del Valle and et al. “Challenges in materials and devices for resistive-switching-based neuromorphic computing”. In: *Journal of Applied Physics* 124.21 (2018). DOI: 10.1063/1.5047800.

[44] J. Veeramreddy and R.P. Vaddella. “Anomaly-based network intrusion detection through Assessing Feature Association Impact Scale”. In: *International Journal of Information and Computer Security* 8.3 (2016), p. 241. DOI: 10.1504/ijics.2016.079185.

[45] N. Wang et al. “Energy-efficient intelligent ECG monitoring for wearable devices”. In: *IEEE Transactions on Biomedical Circuits and Systems* 13.5 (2019), pp. 1112–1121. DOI: 10.1109/TBCAS.2019.2930215.

[46] *What are Neural Network Architectures?* https://h2o.ai/wiki/neural-network-architectures/. Accessed: 02 May 2024.

[47] Wikipedia Contributors. *Cardiac cycle*. https://en.wikipedia.org/wiki/Cardiac_cycle. Accessed: April 30, 2024. 2019.

[48] Wikipedia Contributors. *Electrocardiography*. https://en.wikipedia.org/wiki/Electrocardiography. Accessed on: April 30, 2024. 2019.

[49] Wikipedia contributors. *Neural network (machine learning)*. https://en.wikipedia.org/wiki/Neural_network_(machine_learning). Accessed: May 2, 2024. 2024.

[50] Yifan Xia and Yubo Xie. “A novel wearable electrocardiogram classification system using convolutional neural networks and active learning”. In: *IEEE Access* 7 (2019), pp. 7989–8001. DOI: 10.1109/ACCESS.2019.2890865.

[51] Jiahao Xiao and et al. “ULECGNet: An ultra-lightweight end-to-end ECG classification neural network”. In: *IEEE Journal of Biomedical and Health Informatics* 26.1 (2022), pp. 206–217. DOI: 10.1109/JBHI.2021.3090421.

[52] Peng Xie and et al. “Bidirectional recurrent neural network and Convolutional Neural Network (BIRCNN) for ECG Beat Classification”. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2018. DOI: 10.1109/EMBC.2018.8512752.

[53] Z. Yan, J. Zhou, and W.-F. Wong. “Energy efficient ECG classification with spiking neural network”. In: *Biomedical Signal Processing and Control* 63 (2021), p. 102170. DOI: 10.1016/j.bspc.2020.102170.