

5-2021

HEART RHYTHM CLASSIFICATION FROM STATIC AND ECG TIME-SERIES DATA USING HYBRID MULTIMODAL DEEP LEARNING

Ahmad Abdulrazaq Abdulla Alnajjar

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/all_theses



Part of the [Software Engineering Commons](#)

United Arab Emirates University

College of Information Technology

Department of Computer Science and Software Engineering

HEART RHYTHM CLASSIFICATION FROM STATIC AND ECG
TIME-SERIES DATA USING HYBRID MULTIMODAL DEEP
LEARNING

Ahmad Abdulrazaq Abdulla Alnajjar

This thesis is submitted in partial fulfilment of the requirements for the degree of
Master of Science in Software Engineering

Under the Supervision of Dr. Mohammad Mehedy Masud

May 2021

Declaration of Original Work

I, Ahmad Abdulrazaq Abdulla Alnajjar, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled “*Heart Rhythm Classification from Static and ECG Time-Series Data Using Hybrid Multimodal Deep Learning*”, hereby, solemnly declare that this thesis is my own original research work that has been done and prepared by me under the supervision of Dr. Mohammad Mehedy Masud, in the College of Information Technology at UAEU. This work has not previously been presented or published, or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature:  Date: 08 June 2021

Copyright © 2021 Ahmad Abdulrazaq Abdulla Alnajjar
All Rights Reserved

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

- 1) Advisor (Committee Chair): Mohammad Mehedy Masud

Title: Associate Professor

Department of Information Systems and Security

College of Information Technology

Signature  Date 08 June 2021

- 2) Member: M. Adel Serhani

Title: Professor

Department of Information Systems and Security

College of Information Technology

Signature  Date 08 June 2021

- 3) Member (External Examiner): Ashraf M. Elnagar

Title: Professor

Department of Computer Science

Institution: University of Sharjah, UAE

Signature  Date 13 June 2021

This Master Thesis is accepted by:

Dean of the College of Information Technology: Professor Taieb Znati

Signature Taieb Znati Date 11/08/2021

Dean of the College of Graduate Studies: Professor Ali Hassan Al Marzouqi

Signature Ali Hassan Date 11/08/2021

Copy ____ of ____

Abstract

Cardiovascular arrhythmia diseases are considered as the most common diseases that cause death around the world. Abnormal arrhythmia diseases can be identified by analyzing heart rhythm using an electrocardiogram (ECG). However, this analysis is done manually by cardiologists, which may be subjective and susceptible to different cardiologist observations and experiences, as well as to noise and irregularities in those signals. This can lead to misdiagnosis. Motivated by this challenge, an automated heart rhythm diagnosis approach from ECG signals using Deep Learning has been proposed. In order to achieve this goal, three research problems have been addressed. First, recognizing the role of each single-lead of a 12-lead ECG to classify heart rhythms. Second, understanding the importance of static data (e.g., demographics and clinical profile) in classifying heart rhythms. Third, realizing whether the static data can be combined with the ECG time series data for better classification performance. In this thesis, different deep learning models have been proposed to address these problems and satisfactory results are achieved. Therefore, using these knowledges, an effective hybrid deep learning model to classify heart rhythms has been proposed. As per knowledge obtained from relevant literature, this is the first work to identify the importance of individual lead and combined lead as well as the importance of combining static data with ECG time series data in classifying heart rhythms. Extensive experiments have been performed to evaluate this algorithms on a 12-lead ECG database that contains data from more than 10,000 individual subjects and obtained a high average of accuracy (up to 98.7%) and F1-measure (up to 98.7%). Moreover, in this thesis, the distribution of heart rhythms from the database based on heart rhythm type, gender, and age group have been analyzed, which will be valuable for further improvement of classification performance. This study will provide valuable insights and will prove to be an effective tool in automated heart rhythm classification and will assist cardiologists in effectively and accurately diagnose heart disease.

Keywords: Artificial Intelligent, Deep Learning, Multimodal, ECG, Arrhythmia, Cardiovascular Diseases.

Title and Abstract (in Arabic)

تصنيف الإيقاعات القلبية من بيانات المخطط الكهربائي للقلب والبيانات الوصفية الثابتة باستخدام الذكاء الاصطناعي ممثلاً بالتعلم العميق

الملخص

تعتبر النوبات القلبية وأمراض القلب من أكثر الأمراض شيوعاً، والتي تسبب الوفاة في العالم. يمكن التعرف على إيقاعات القلب الغير طبيعية عن طريق جهاز التخطيط الكهربائي للقلب (ECG). حيث يتم إجراء هذا التشخيص بواسطة أطباء القلب، ومع ذلك، قد يكون التشخيص الذاتي يختلف من اخصائي لآخر رغم الخبرة، لأن اشارات المخطط غير خالية من الضوضاء والترددات المختلفة مما يؤدي الى التشخيص الخطأ. تجنباً لذلك، ولأهمية هذا الموضوع، تم توظيف الذكاء الاصطناعي ممثلاً بالتعلم العميق في تشخيص أمراض القلب من خلال النشاط الكهربائي للقلب المكون من 12 صورة (12-lead) التي تُنتج عن طريق الجهاز الكهربائي للقلب بواسطة أجهزة الاستشعار التي توضع حول جسم المريض. لقد تم إحراز تقدم كبير في هذا الصدد، وأجريت تجارب شاملة في تقييم الخوارزميات على قاعدة بيانات مكونة من (12-lead) تحتوي على أكثر من عشرة آلاف مريض. هذا العمل يسعى للإجابة على ثلاث أسئلة بحثية: أولاً، ما هي أهمية ودور كل lead في تصنيف الإيقاعات القلبية. ثانياً، فهم أهمية البيانات الوصفية الثابتة (مثل البيانات الطبية السريرية والبيانات الديموغرافية) في تصنيف الإيقاعات، ثالثاً، ما إذا كان يمكن دمج البيانات الوصفية الثابتة مع بيانات المخطط الكهربائي الزمنية لتخطيط القلب للحصول على نتائج أفضل في التصنيف. يعتبر هذا العمل البحثي الأول من نوعه في أهمية تحديد الاستفادة من ال-lead الفردي والمشارك بالإضافة الى أهمية البيانات الوصفية الثابتة مع الزمنية في تصنيف إيقاعات القلب. باستخدام هذه المعرفة التي تم الحصول عليها، وبناءً على مخرجات التجارب، تم التوصل الى نسبة مئوية عالية في التصنيف (تصل الى 98.7%) ونسبة ال-F1-measure (تصل الى 98.7%). علاوة على ذلك، تم القيام أيضاً في هذا العمل البحثي بتوزيع وتصنيف إيقاعات القلب من قاعدة البيانات بناءً على نوع كل إيقاع والفئة العمرية، والتي ستكون خطوة مهمة في تحسين أداء التصنيف. هذه الدراسة ستوفر نظرة ثاقبة، وستثبت انها فعالة في التصنيف الآلي لإيقاعات القلب والتي ستساعد أطباء القلب في التشخيص الفعال لأمراض القلب.

مفاهيم البحث الرئيسية: الذكاء الاصطناعي، التعلم العميق، تخطيط القلب، أمراض القلب، إيقاعات القلب الغير طبيعية.

Acknowledgements

All praise to Allah (S.W.T) who kindly favored me with his kindness, grace, and mercy to complete my higher studies and preparing for the master thesis. I would like to extend my warm thanks to all my master teachers and my committee for their effort, patience, and encouragement in my research and education career at the College of Information Technology. I extend my heartfelt thanks to my advisor Dr. Mohammad Masud for his support, moral guidance, assistance, and motivation in my career and accomplish and complete of the master thesis. I would also like to express thanks and grateful to my program coordinator Dr. Salah Bouktif for his support and guidance. My sincere gratitude to Prof. M. Adel Serhani for his great expertise, feedback, and advice in my research. Also grateful to Ms. Maryam Al Mandhri for her contribution to the success of my academic career.

Special thanks and greetings to my generous father, mother, brothers, and family for their continuous inspiration and the moral supports they have provided to complete my graduate studies.

Additionally, special thanks are extended to all my friends and colleagues. May Allah protect and take care of you all.

Dedication

*To my beautiful country (UAE),
lovely Grandmother, beloved Father, Mother,
Brothers, Family, and Teachers*

Table of Contents

Title	i
Declaration of Original Work	ii
Copyright	iii
Approval of the Master Thesis	iv
Abstract	vi
Title and Abstract (in Arabic)	vii
Acknowledgements	viii
Dedication	ix
Table of Contents	x
List of Tables	xiii
List of Figures	xiv
List of Abbreviations	xv
Chapter 1: Introduction	1
1.1 Overview	1
1.2 Motivation	2
1.3 Problem statement	3
1.4 Contribution	4
Chapter 2: Background – ECG and Deep Learning	7
2.1 Electrocardiogram	7
2.2 Neural networks	9
2.3 Deep learning	12
2.3.1 Convolutional Neural Network	13
2.3.2 Recurrent Neural Network	14
Chapter 3: Literature Review	16
3.1 Application of DL models for heart disease detection in general	16
3.2 Heart rhythm classification using DL from 12-lead ECG data	21
Chapter 4: Proposed Approach	24
4.1 2D-Convolutional Neural Network architecture	25
4.1.1 2D-CNN architecture for combined lead data	26
4.1.2 2D-CNN architecture for single lead data	28

4.2 Multi-layer perceptron architecture for only static data	29
4.3 Proposed Hybrid 1D-CNN with-Bidirectional-GRU-and Bidirectional LSTM architecture.....	30
4.4 Multimodal deep learning (Proposed Hybrid combined with MLP).....	32
Chapter 5: Dataset	35
5.1 Analysis of data distribution.....	37
5.1.1 Based on gender	38
5.1.2 Based on age group.....	39
5.2 Preprocessing and data preparation	43
Chapter 6: Experiments and Results	45
6.1 Experimental setup	46
6.1.1 Hardware and software	46
6.1.2 Parameter setting.....	46
6.1.3 Overfitting avoidance.....	48
6.2 Evaluation setup	48
6.3 Experiments using 2D-CNN of single-lead and combined lead classification	50
6.3.1 Experiments with combined lead.....	50
6.3.2 Experiments with single-lead.....	53
6.3.3 Subsets comparison for a combined and single-lead	55
6.4 Experiment using MLP for only static data classification.....	59
6.4.1 SR-SB subset	59
6.4.2 SR-ST subset.....	60
6.4.3 SR-AFIB subset	61
6.4.4 Subsets comparison.....	62
6.5 Experiment using the proposed hybrid 1D-CNN-BiGRU- BiLSTM model.....	63
6.5.1 SR-SB subset	63
6.5.2 SR-ST subset.....	64
6.5.3 SR-AFIB subset	65
6.5.4 Subsets comparison.....	66
6.6 Experiment using the Multimodal Proposed hybrid combined with MLP.....	67
6.6.1 SR-SB subset	68
6.6.2 SR-ST subset.....	68
6.6.3 SR-AFIB subset	69
6.6.4 Subsets comparison.....	70
6.7 Comparison of evaluation results	71

6.8 Analysis and discussion.....	74
6.8.1 Combined lead vs single-lead	75
6.8.2 Static data combined with time-series data.....	76
6.8.3 Proposed hybrid model performance	76
Chapter 7: Conclusion and Future Work	77
References	79

List of Tables

Table 1: Related work comparison	23
Table 2: Layers and parameters of 12-lead architecture	28
Table 3: Layers and parameters of single-lead architecture.....	29
Table 4: ECG static database attributes	37
Table 5: Distribution of rhythms based on age group and gender	38
Table 6: Combined lead SR-SB performance measurements	51
Table 7: Combined lead SR-ST performance measurements.	52
Table 8: Combined lead SR-AFIB performance measurements.....	53
Table 9: SR-SB individual leads performance measurements comparison	56
Table 10: SR-ST individual leads performance measurements comparison.....	57
Table 11: SR-AFIB individual leads performance measurements comparison.....	58
Table 12: MLP SR-SB performance measurements	60
Table 13: MLP SR-ST performance measurement.....	61
Table 14: MLP SR-AFIB performance measurement	62
Table 15: MLP model subsets comparison	63
Table 16: Proposed hybrid SR-SB performance measurements	64
Table 17: Proposed hybrid SR-ST performance measurement.....	65
Table 18: Proposed hybrid SR-AFIB performance measurement	66
Table 19: Proposed hybrid subsets comparison.....	67
Table 20: Multimodal SR-SB performance measurement	68
Table 21: Multimodal SR-ST performance measurement	69
Table 22: Multimodal SR-AFIB performance measurement.....	70
Table 23: Multimodal (proposed hybrid + MLP) subsets comparison	71
Table 24: Summary for all comparison.....	73
Table 25: Comparison with the Yildirim and his team work architecture.....	74

List of Figures

Figure 1: Visualization of the 12-lead heart electrical signal using ten electrodes	8
Figure 2: Basic ECG wave	9
Figure 3: AI terminologies	9
Figure 4: Single perceptron neural network.....	10
Figure 5: CNN architecture.....	14
Figure 6: RNN architecture.....	15
Figure 7: Overall system architecture of the proposed heart rhythm classification.....	25
Figure 8: Combination of 12-leads architecture.....	27
Figure 9: Single-lead architecture	29
Figure 10: MLP architecture for static data	30
Figure 11: Proposed hybrid architecture	32
Figure 12: Multimodal (proposed hybrid fused with MLP) architecture.....	34
Figure 13: Distribution of rhythms in the database.....	36
Figure 14: Rhythms distribution of male (left) and female (right)	39
Figure 15: Age group distribution of SB (left) and SR (right).....	40
Figure 16: Age group distribution of AFIB (left) and ST (right).....	40
Figure 17: Age group distribution of AF (left) and SI (right).....	41
Figure 18: Age group distribution of SVT (left) and AT (right).....	42
Figure 19: Age group distribution of AVNRT (left) and AVRT (right).....	42
Figure 20: Age group distribution of SAAWR.....	43
Figure 21: Organization of the dataset.....	44
Figure 22: Combined lead SR-SB accuracy (left) and loss (right)	51
Figure 23: Combined lead SR-ST accuracy (left) and loss (right).....	52
Figure 24: Combined lead SR-AFIB accuracy (left) and loss (right)	53
Figure 25: Validation of all the individual leads SR-SB.....	54
Figure 26: Validation of all the individual leads SR-ST.....	54
Figure 27: Validation of all the individual leads SR-AFIB	55
Figure 28: MLP SR-SB accuracy (left) and loss (right)	60
Figure 29: MLP SR-ST accuracy (left) and loss (right).....	61
Figure 30: MLP SR-AFIB accuracy (left) and loss (right)	62
Figure 31: Proposed hybrid SR-SB accuracy (left) and loss (right)	64
Figure 32: Proposed hybrid SR-ST accuracy (left) and loss (right).....	65
Figure 33: Proposed hybrid SR-AFIB accuracy (left) and loss (right)	66
Figure 34: Multimodal SR-SB accuracy (left) and loss (right).....	68
Figure 35: Multimodal SR-ST accuracy (left) and loss (right).....	69
Figure 36: Multimodal SR-AFIB accuracy (left) and loss (right).....	70

List of Abbreviations

AI	Artificial Intelligent
ML	Machine Learning
DL	Deep Learning
DNN	Deep Neural Network
WHO	World Health Organization
ECG	Electrocardiogram
EKG	Electrocardiogram
CVD	Cardiovascular Disease
CAD	Coronary Artery Disease
AAMI	Association for the Advancement of Medical Instrumentation
CNN	Convolutional Neural Network
ConvNet	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
INCART	St. Petersburg 12-Lead Arrhythmia Database
PPG	Photoplethysmogram
DWT	Discrete Wavelet Transform
MIT-BIH	Massachusetts Institute of Technology-Beth Israel Hospital
PVC	Premature Ventricular Contraction
MI	Myocardial Infarction
PTBDB	Physikalisch-Technische Bundesanstalt Diagnostic ECG Database
SAN	Sinoatrial Node

MLP	Multilayer Perceptron
AUC	AUC Area Under the ROC Curve
SR	Sinus Rhythm
SB	Sinus Bradycardia
ST	Sinus Tachycardia
SI	Sinus Irregularity
AFIB	Atrial Fibrillation
AF	Atrial Flutter
AT	Atrial Tachycardia
SVT	Supraventricular Tachycardia
AVNRT	Atrioventricular Node Reentrant Tachycardia
AVRT	Atrioventricular Reentrant Tachycardia
SAAWR	Sinus Atrium to Atrial Wandering Rhythm

Chapter 1: Introduction

1.1 Overview

Cardiovascular diseases (CVDs) are the most common diseases that cause death all over the world. CVDs are considered as the most prevalent disease worldwide. More than 17.9 million people yearly meet the fate of death globally from CVDs (WHO, 2017). Moreover, CVDs cause 3.9 million deaths in Europe according to European Cardiovascular Disease Statistics (European Cardiovascular Disease Statistics, 2017). Furthermore, Heart and Stroke statistics reported that more than 840,000 deaths were recorded in 2016 in the US (Nearly half of US adults have cardiovascular disease). In addition, a survey conducted on more than 1000 individuals by Cleveland Clinic Abu Dhabi reported that the UAE residents who took part in this study, have at least one heart disease (Cleveland Clinic Abu Dhabi, 2019).

However, CVDs could be prevented through analyzing the heart signals, such as Electrocardiogram (ECG). ECG is a significant tool in the medical health domain that could diagnose and identify abnormal heart conditions that can lead to CVDs. In addition, it is a valuable indicator as a health assessment for cardiologists that could detect and classify heart patterns.

ECG is a popular and cost-effective test that reflects the picture of the cardiac condition signals. Typically, ECG signals can be measured using single-lead or multiple-lead electrically on the surface of the skin by the electrodes. Those leads are distributed on different parts of the body in which they give measurement records that depend on the number of electrodes distributed on the area of the body. Electrodes

can detect the heart electrical signals and pass them through connected cables to produce the ECG graphs which are represented in the ECG machine.

There are various automated algorithms and techniques available that can help cardiologists identify and classify heart rhythms (Luz et al., 2016; Matias et al., 2021). In this work, deep learning techniques have been proposed to classify the heart rhythm using a 12-lead database and compared the performance of different Deep Learning models on several variations (single-lead vs combined lead, static vs time-series) of the 12-lead ECG data.

1.2 Motivation

Monitoring heart activity for patients with heart disease and other heart condition patterns leads to better and controlled life (Serhani et al., 2020a). Identifying heart disease from ECG needs interpretation skills and deeper understanding. Experienced cardiologists can identify heart rhythm problems by manually examining ECG data. However, a study suggests that even after years of experience in this field sometimes experienced cardiologists misinterpreted in analyzing and distinguish the irregular beats due to human error (Sampson et al., 2015).

Therefore, distinctive automated techniques to identify heart rhythm from ECG have been developed to help cardiologists in mitigating the risk of misdiagnosis and explore appropriate treatment (Singh et al., 2018)

Many of these automated techniques apply machine learning including deep learning techniques to automatically classify various heart conditions, such as normal and abnormal rhythm, and had been proven to achieve high accuracy.

This work is motivated by the possibility of enhancing these intelligent automation of heart disease classification and thus developing a model that would greatly assist caregivers in vital decision-making.

Furthermore, this thesis is also motivated by the possibility of improving prediction by combining static data with time-series data, because addition of static with time-series data has not been well investigated for ECG based heart rhythm classification problem. Finally, this work is also motivated to investigate the effectiveness of different leads of the 12-lead ECG as this initiative was also largely ignored in the literature.

1.3 Problem statement

This work aims to study and develop an efficient and effective machine learning technique to classify heart rhythms using 12-lead ECG recording data. From this point, three research questions have been identified to achieve this goal. these questions are as follows:

1. What is the relative importance of each lead in a 12-lead ECG in classifying heart rhythms? This is addressed in Section 6.3.3.
2. What is the importance of static data (e.g. demographic and clinical profile) in classifying heart rhythms? This is addressed in Section 6.4.
3. Can clinical static data be combined with ECG time-series data to improve classification accuracy? This is addressed in Section 6.6.

Extensive study, design, development, and experiments have been done to address the above research questions, and this has lead to the achievement of the

objective of developing an efficient model to classify heart rhythms from ECG data based on deep learning.

1.4 Contribution

In this thesis work, deep learning models have been developed to answer three research questions, mainly, what is the importance of each lead in a 12-lead ECG in classifying heart rhythm, what is the importance of static data in classifying heart rhythms, and whether static data can be combined with the ECG data to improve classification accuracy.

In this regard, a comprehensive study, design, development, and experiments have been terminated using large number of individual subjects. The collected database containing 12-lead ECG recording data of more than 10,000 patients with different heart rhythms. The rhythms that contain at least 1,500 subjects have been selected. This is done to ensure the reliability of the results. Then, the data is organized into three different subsets, so that each subset consists of only two rhythms: normal and abnormal rhythm, to simplify the classification problem as a binary classification. The first subset consists of normal Sinus Rhythm (SR) data and the anomaly Sinus Bradycardia (SB) data. The second subset contains normal Sinus Rhythm (SR) along with anomaly Sinus Tachycardia (ST). The third subset has normal Sinus Rhythm (SR) and anomaly Atrial Fibrillation (AFIB). Each experiment was conducted based on the three subsets.

In order to address the first question, a 2D-Convolutional Neural Network has been adopted which uses the dynamic ECG time-series data as an input. Then, the classification performance evaluated by training the model with all the 12-lead signals

and repeating the same experiment by training the model with each individual lead (single-lead) of the 12-lead ECG signal. Accordingly, the results were analyzed and proved statistically that the combined lead performs significantly better than single-lead. The second question has been addressed by using a Multi-layer Perceptron (MLP) architecture to train on the static data that are available in the same dataset. The static data of each subject contains some demographic attributes (age, gender, etc.) as well as a statistical summary of the dynamic ECG time-series data. It has been found that the static data can give some good prediction accuracy (around 80%), however, it was less than the accuracy obtained from the 2D-CNN that used the ECG time-series data. This finding could be helpful for future research.

Moreover, by understanding the previous conclusions, the study focused on identifying an efficient deep learning model for the classification task using the combined lead dynamic ECG time-series data. Thereupon, a hybrid multi-modal deep learning model have been proposed, which consists of a One-Dimensional Convolutional Neural Network with Bidirectional GRU and Bidirectional LSTM (1D-CNN-BiGRU-BiLSTM). It is shown empirically that this model significantly outperformed all other architectures in terms of classification performance.

Finally, to answer the third question, several multi-modal deep learning models that can combine static data with dynamic time-series data have been proposed. However, none of the multi-modal models achieved better accuracy than the models that used only ECG time-series data. It can be concluded that the static data does not help in improving the classification results.

Based on the knowledge obtained from related literature, this is the first work that vows to address these three important questions for a more accurate diagnosis

using the newly published database. It is understood that this study will provide valuable insights into heart rhythm classification problems and will deliver an effective and efficient tool for classifying heart rhythms. This will not only be useful for healthcare professionals, but also will pave the way for future study, research, and developments. For example, these research results will be useful in building more comprehensive heart disease diagnosis research initiatives that utilize other data including ECG such as patient clinical profile, clinical images of heart collected by Imaging devices like Echocardiogram. Also, the proposed technique can be used to develop a diagnostic tool for physicians.

Chapter 2: Background – ECG and Deep Learning

This chapter introduces a description of the main characteristics of the Electrocardiogram (ECG). Furthermore, it presents a summarization about the Neural Network, followed by Deep Learning (DL), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN).

2.1 Electrocardiogram

Electrocardiogram also called ECG or EKG, is a health assessment tool that records bioelectrical heart activities. Examining heart status can be done also by ambulatory or wearable devices which is prevalent in daily application, such as smart watches that use photoplethysmogram (PPG) as a sensor to monitor heart rate and cardiac cycle (Bashar et al., 2018; Shen et al., 2019). ECG with 12-lead is efficient and often used in healthcare which gives clear measurements of 12 different heart view dimensions, hence, decisions can be utilized by cardiologists to diagnose common heart problems, such as regular and irregular rhythms.

The standard ECG uses ten electrodes in which generates the 12-lead (Goldberger et al., 2013). The limb leads consist of six leads placed on the arms and calves are represented as lead I, II, III, aVL, aVR, and aVF. On the other hand, the rest of the leads are called precordial leads that are placed on the precordium and described as V1, V2, V3, V4, V5, and V6. Figure 1 visualizes the 12-lead ECG signals using ten electrodes distributed through a human body.

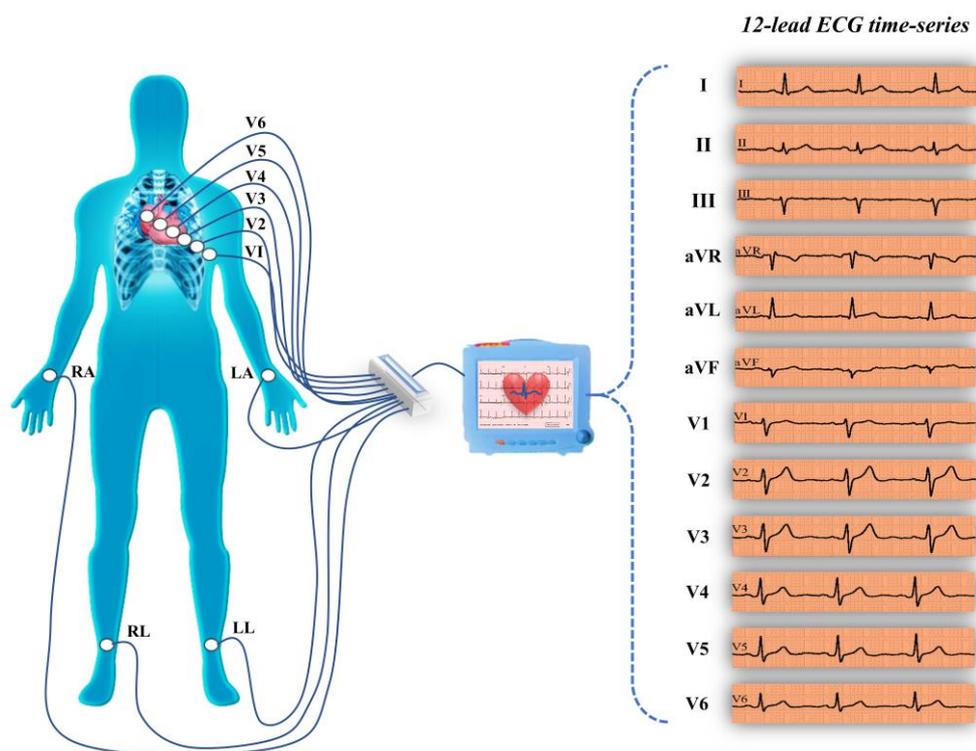


Figure 1: Visualization of the 12-lead heart electrical signal using ten electrodes

ECG typically produces heart wave pattern signals through the electrodes. At the start of the heart cycle, the heart relaxes and expands while receiving blood into the ventricles through the atrial. Therefore, the atrial chambers pump the blood into the ventricles and then relax. Those electrical signals can examine the changes in heart activity. The normal ECG cycle contains P wave, QRS complex, and T wave. The P wave indicates the atrial depolarization that spread from the sinoatrial node (SAN) to the atrial. The Q, R, and S waves are called as QRS Complex. QRS complex depicts the process of ventricular depolarization. Then, the T wave occurs after the QRS complex which illustrates the ventricular repolarization. U wave is a small deflection wave that follows the ventricular repolarization, it may not always be noticeable in the ECG. RR interval is the time elapsed between the two successive R waves of the QRS signal. Figure 2 briefly expresses the P, Q, R, S, and T waves.

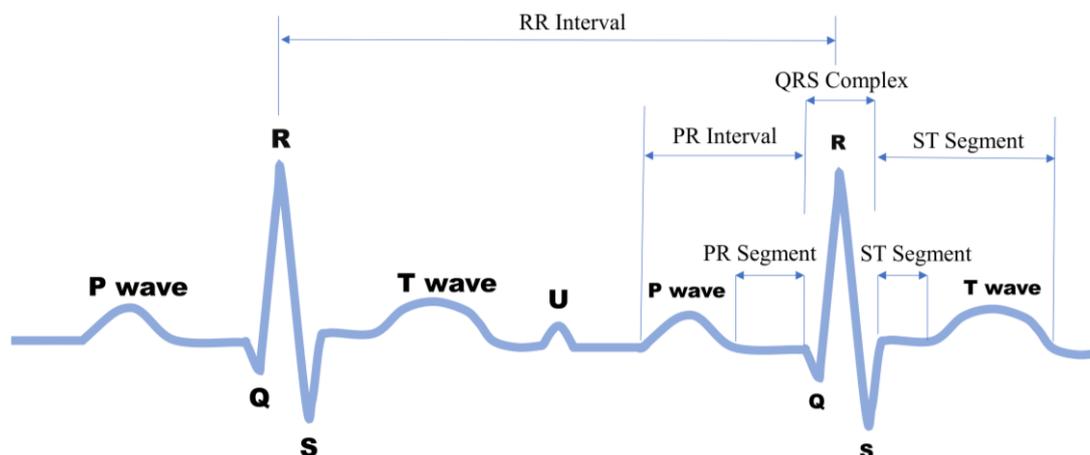


Figure 2: Basic ECG wave

2.2 Neural networks

Artificial Intelligence (AI) refers to a comprehensive term that simulates human intelligence such as Machine Learning (ML) including deep learning (DL). Neural Networks (NN) is one of the most popular and efficient computing techniques, which has its root in Artificial Intelligence (AI). It is usually used for trading systems and tasks such as classification, clustering, and prediction. Figure 3 exposes in detail the distinction between AI and other related terminologies.

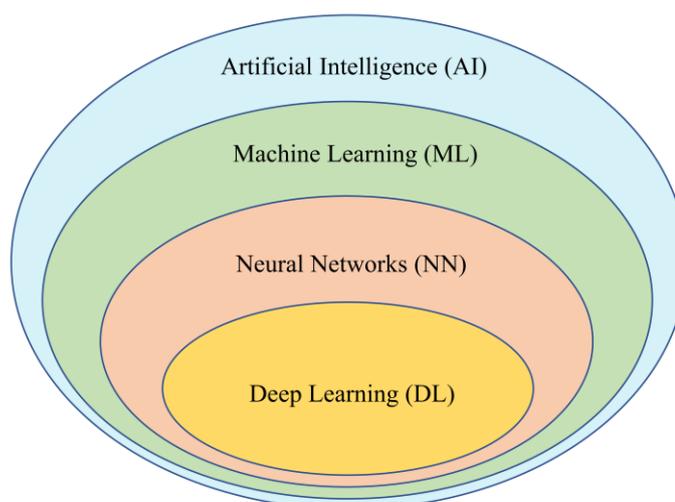


Figure 3: AI terminologies

The history of the neural network started when Walter Pitts and Warren McCulloch implemented a model based on the neural networks inspired by the human brain, which exhibit that the computable logic function could suitably learn. In 1950, Marvin Minsky and Dean Edmonds, created the first neural network, then many of the concepts were proposed such as backpropagation. Since 1943, many researchers introduced realistic models that provide computing power that allows the potential ability of the computational neuron network models to act as biological neurons that have nerve cells connected to each other. Therefore, it allows the information to flow through the connected neurons along with the associated weight to solve a particular task. However, artificial neurons are applied to various numbers of complex situations where there are large datasets in order to train static data and make predictions for multiple inputs in such there is no time constrain.

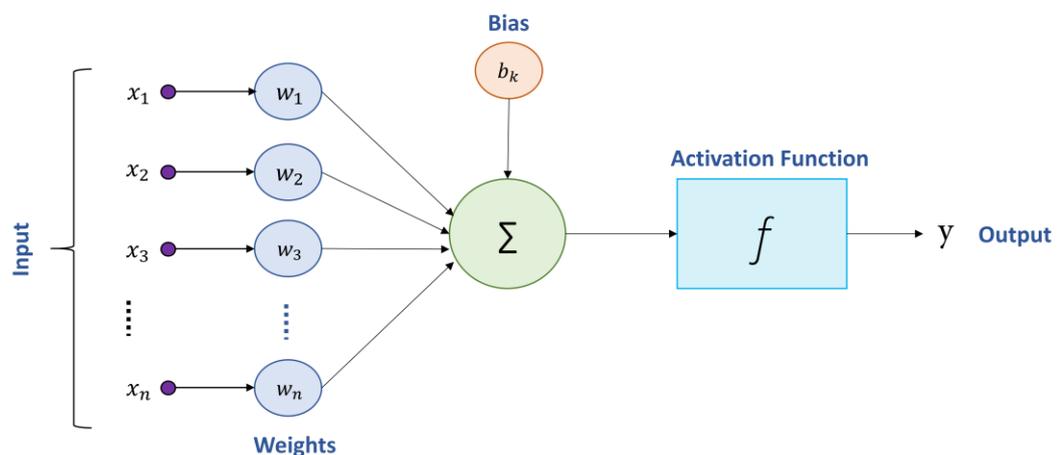


Figure 4: Single perceptron neural network

The above figure is a graphical representation of a single neural network perceptron that learns only linearly separable patterns e.g. XOR logic gate. The following elements can be seen from the figure: input layer (x_i), associated weights (w_i), net sum (Σ), Bias (b_i), activation function (g), and output layer (y). The input

layer (x_i) of the perceptron passes the input values (data) for further processing, then the input values are processed along with each connection by associated weights (w_i). Those processed values are then passed to the net sum (Σ) to calculate the total sum. The bias is an additional parameter that adds a threshold value to adjust the output along with the weighted sum of the input. Then, the activation function (g) will be applied to the net sum of the weighted inputs. Activation function (g) is mainly used to introduce nonlinearities in the network. The most common activation functions are sigmoid, hyperbolic tangent, softmax, and rectified linear unit. Finally, the output layer (y) calculates the prediction score based on the inputs, the associated weights, and the activation function as exhibited below.

$$\text{i.e., } y = g.(\sum_i^n x_i \cdot w_i)$$

Generally, the perceptrons are trained using the forward and backpropagation method to calculate the gradient descent of neural network parameters. In particular, the forward propagation sequentially calculates intermediate variables which proceed from the input to the output layer. The backpropagation sequentially calculates the gradients of intermediate variables in the reversed order of the neural network (Russell & Norvig, 2009).

In terms of multi-layer perceptron (MLP), the architecture presented and arranged as layers, each neuron of each layer will repeat the same process as clarified above, then the output result of the particular neuron is calculated. The MLP has the same structure as a single perceptron neural network with primarily two or more hidden layers.

2.3 Deep learning

In recent years, Deep Learning (DL) has been widely used in many areas, such as computer vision, speech recognition, image verification, classification, and many more. DL is a part of the Machine Learning family. Machine learning techniques are limited in terms of capability to extract features from the raw data that requires more concentration to carefully design hand-crafted high-quality features to proper representation patterns. An image, for instance, is represented by channels and arrays of pixels that require large volumes of data to train a network, hundreds of features to be extracted, and a huge amount of computational power to solve the network complexity. These limitations can be solved by deep learning. DL has the advantage of automatically extracting useful features due to the ability to deal with sophisticated and heterogeneous network structures from the substantial and high dimensional data.

The nature of deep learning able to analyze the presented patterns efficiently and effectively. However, some of the complex problems start to decrease whenever the layers increase, and this is basically because of the scalability of the DL network. Typically, the amount of training data, in addition to the size of the network help the DL approach to perform well.

Nowadays, the deep learning accuracies of different approaches energetically enhanced the state of the art application tools due to various DL architectures (Abiodun et al., 2018; Khan et al., 2020), for example, Convolutional Neural Networks (CNN) often used for learning from images, and Recurrent Neural Networks (RNN) for learning from temporal data. These variations are briefly introduced in the following paragraphs.

2.3.1 Convolutional Neural Network

Convolutional Neural network (CNN) sometimes called ConvNet, is one of the most popular variants of deep learning that is successfully used for computer vision. Conceptually, its structure performs well in detecting objects, and handwritten digits. The big data and the trend of the improvement in the technology accelerated the research in CNNs that lead to a series of improvements.

Therefore, significant improvement has been achieved to improve the representation of CNNs architectures such as using different parameters, regularization, loss, layers, and activation function. Moreover, the learning methodology helps the CNN to be performed to complex, heterogeneous, and large-scale data such as ResNet, AlexNet, VGGNet, and GoogLeNet that are used for different tasks. It is often used in many medical research fields to diagnose diseases from the images of ECG, X-ray, and RMI which supports the doctors in diagnosis and treatment.

There are three main layers in the CNN architecture: convolutional, pooling, and fully connected layer. The convolutional layer has a mathematical operation that is used to extract various features from the input data. The convolution is performed between the input and a filter of a particular size to determine the represented information of the input such as horizontal and vertical edges. The output of the convolutional layer is termed as a feature map that fed to other layers in order to learn other features. The pooling layer mostly following the convolutional layer in which decreases the size of the convolved feature map to eliminate the computational cost. Finally, the fully connected layer (FC) also known as the dense layer, in which it connects the neural network layers by a learning weight and bias vector. It is usually

placed before the last output layer of the CNN architecture, where the output layer is used to calculate the probability distribution of each class. Figure 5 below shows the three layers of CNN architecture.

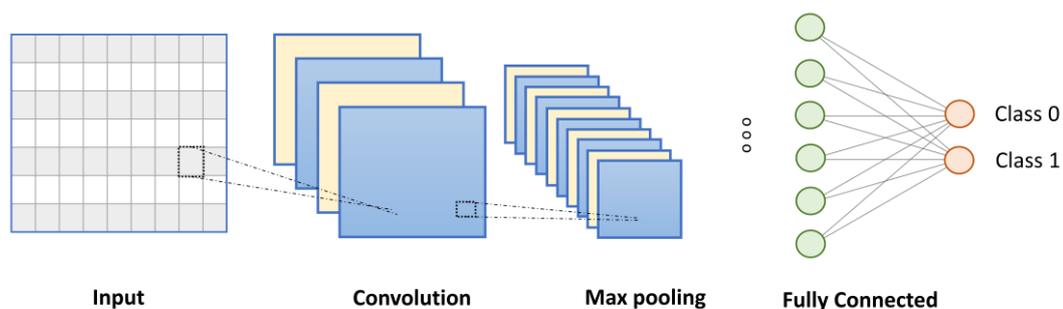


Figure 5: CNN architecture

2.3.2 Recurrent Neural Network

Recurrent Neural Networks (RNN) is a branch of Artificial Neural Networks that became popular due to the high dynamical behavior that contains cyclic connections for processing sequence of values across time, which makes it performs successfully in identical tasks for sequential data and produce an output in every time steps. This mechanism allows RNN to heavily demonstrate tasks such as time-series prediction, and language translation (LeCun et al., 2015). However, RNNs have limitations that are limited to the ability of time backpropagation which makes it unstable to capture long-term dependencies.

Hochreiter and Schmidhuber (1997) have presented long-short term memory (LSTM) which has input gates that gain the knowledge stored in the memory cell block, forget gate which learns the information to be forgotten or needed from the memory block, and output gate that can understand when to call the stored information. Figure 6 depicts the input state (X_t), current time stamp (H_t), and output state (O_t).

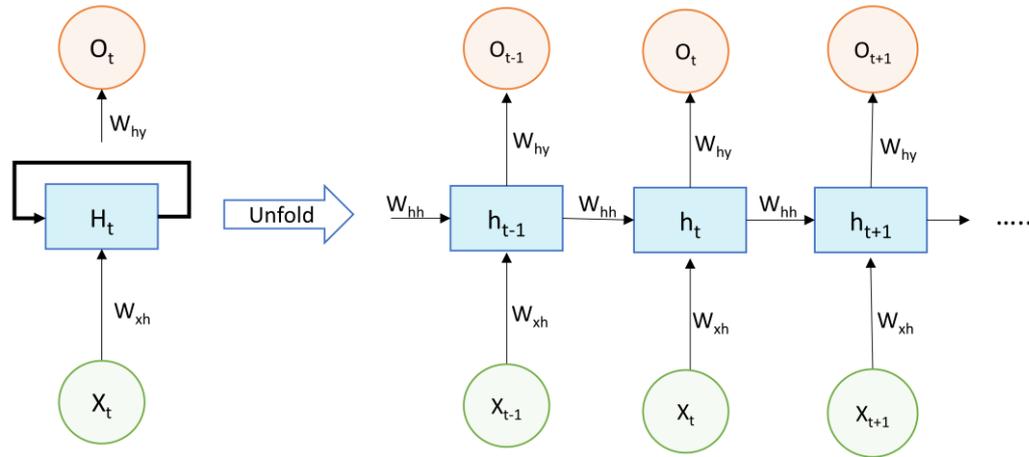


Figure 6: RNN architecture

This specific type of traditional RNN architecture is designed to use the memory blocks rather than the traditional RNN units which can control the information flow to the memory more efficiently as illustrated below.

The input gate

$$i_t = \sigma.(w_i[h_{t-1}, x_t] + b_i)$$

The forget gate

$$f_t = \sigma.(w_f[h_{t-1}, x_t] + b_f)$$

The output gate

$$o_t = \sigma.(w_o[h_{t-1}, x_t] + b_o)$$

Where i_t represents the input gate, f_t , expresses the forget gate, o_t , indicates output gate.

Chapter 3: Literature Review

Many existing research initiatives have attempted and conducted their studies on ECG and photoplethysmogram (PPG) data. Researchers frequently explored different machine learning (ML) and deep learning (DL) techniques to predict and classify various arrhythmias, such as atrial fibrillation, tachycardia, bradycardia, and ventricular arrhythmias. Furthermore, there are studies on supraventricular ectopic beat detection from ECG. Other researchers concentrated on the classes suggested by the Association for Advancement of Medical Instrumentation (AAMI) standards which distributed the heartbeat into five classes: non-ectopic (N), supraventricular ectopic (S), ventricular ectopic (V), fusion (F), and unknown beats (Q). Some researchers reviewed various sparsity based noise reduction techniques for desnoising of ECG signal (Devi et al., 2019; Keshavamurthy & Eshwarappa, 2017).

Based on the literature study of the related work from Google Scholar, IEEE Explore, ACM, Science Direct, and Elsevier databases, two main categories have been identified. The first category is the application of DL models for heart disease detection in general, and the second category is the heart rhythm classification using DL from 12-lead ECG data. The next two subsections explain those categories.

3.1 Application of DL models for heart disease detection in general

The following related works belong to the first category, i.e., heart disease detection in general, using deep learning techniques. The works could be further subcategorized into two: works that use MIT-BIH arrhythmia database and works that use other databases.

These works apply DL models on the MIT-BIH arrhythmia database. Note that the database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects.

Acharya et al. (2017c) applied a CNN model to identify five heartbeats classes according to the Association for Advancement of Medical Instrumentation (AAMI). The experiment was conducted using MIT-BIH arrhythmia database. They applied preprocessing techniques to remove noise, normalize segmentation by Z-score, and apply synthetic data to overcome the imbalance classes. The proposed CNN was trained with and without noise using augmented data and achieved accuracy of 94.03% and 93.47%, respectively.

Xu et al. (2019) presented a deep neural network for preprocessing feature extraction and beat-by-beat classification using MIT-BIH arrhythmia database. They applied raw ECG waveforms to include the extracted features such as QRS complex, P, and T waves as input for the model classifier because observations express that the P and T waves contain meaningful information to heart arrhythmias.

Jun et al. (2016) implemented an optimized deep neural network for premature ventricular contraction (PVC) beat classification performed by MIT-BIH arrhythmia database, noted that they extracted six features from the ECG signal such as R-peak, RR-interval, QRS duration, ventricular activation time, Q-peak, and S-peak that are used as an input to the deep neural network (DNN) classification model. The DNN model of six hidden layers achieved 99.41% of accuracy and sensitivity of 96.08%.

Kiranyaz et al. (2015) employed a simple 1D-CNN for ventricular ectopic beats (VEB) and supraventricular ectopic beats (SVEB) classification depends on five types

of heartbeats that are recommended by the Association for Advancement of Medical Instrumentation (AAMI) standard. MIT-BIH arrhythmia database was used in this experiment and the results indicate high classification performance accuracies of 99% and 97.6% for VEB and SVEB, respectively.

Singh et al. (2018) have applied three types of RNN algorithms for classifying normal and abnormal rhythms in an ECG. The study was conducted by MIT-BIH arrhythmia database. The binary classification of the LSTM algorithm achieved an accuracy of 88.1% without signal preprocessing.

Serhani et al. (2020b) adopted CNN model that considers various optimization of hyperparameter which can achieve higher model accuracy. The experiments conducted using MIT-BIH arrhythmia database and applied optimization schemes with batch normalization, regularization, and increasing training epochs. They also categorize clinical recommendation suggestions for five arrhythmia classes.

Gao et al. (2019) conducted a long short-term memory (LSTM) using focal loss function to eliminate the imbalanced data gained from open-source MIT-BIH arrhythmia database, discrete wavelet transform using Daubechies wavelet 6 was applied for signal noise removal, beat segmented using sliding window search method, along with, generating normalized data by Z-score technique during the preprocessing analysis. Observations proved that the Nadam gradient descent optimization and focal loss function robust solution for accurate detecting of imbalanced ECG signals which carry out 99.26% of accuracy for eight beat classes.

Oh et al. (2018a) proposed an autoencoder, derived from MIT-BIH arrhythmia database using lead II signals, further, signals were segmented into samples to allow

heterogeneous segmentation of the ECG records, Z-score normalization was used for scaling, and all the preprocessing helped to handle the U-net model, which is able to identify both arrhythmia conditions and R peaks with 97.32% of accuracy.

Romdhane et al. (2020) introduced a CNN method with an effective focal loss function using a public MIT-BIH arrhythmia and INCART databases. The cost function of the model uses focal loss to solve the imbalanced classes, the model was able to achieve 98.41% of accuracy that classifies ECG signals to five AAMI standard categories.

Li et al. (2020) deployed a deep learning model for diagnosing cardiac arrhythmia classification of five types of heartbeats classes by applying two leads from ECG signals. The preprocessing was obtained for denoising and segmentation which finally shows an accuracy of 99.38%, that obtained using MIT-BIH arrhythmia databases.

Note that, unlike the proposed work, none of the above techniques analyzed the performance of the ECG combined 12-lead with each single-lead of the 12-lead ECG or investigated the usefulness of static data. Furthermore, the MIT-BIH database consists of two-channel ECG data with a sampling rate of 360 samples per second. In contrast, this work deals with 12-channels ECG data, collected at the rate of 500 samples per second. Finally, a database consisting of more than 10,000 subjects was used since MIT-BIH database has only 47 subjects.

The following research work used other databases for heart rhythm classification.

Acharya et al. (2017b) have employed CNN trained using 10-fold cross-validation for ECG segments detection of four classes obtained from three databases available publicly and one from Creighton University Ventricular Tachyarrhythmia (CUDB). ECG signals were down sampled to 250 Hz, noise removed using Daubechies wavelet 6, and each segment normalized using Z-score normalization to solve scaling problem. Then, they used the two and five seconds durations of ECG signals without the QRS detection, which achieved accuracies of 92.50% and 94.90% for the two and five seconds, respectively. They also highlighted that the CNN is invariant to translation, hence, QRS detection is unnecessary.

Hannun et al. (2019) have implemented a deep neural network (DNN) for twelve rhythm classes by using a large dataset contains 53,549 single-lead subjects recorded by Zio monitor devices. The DNN model achieved F1-score and ROC of 0.837, and 0.97, respectively. Moreover, they further tested the proposed DNN to an external database such as 2017 PhysioNet Challenge database to capture the robust performance of the model without adjustment to the hyper-parameters and model architecture, therefore, the result demonstrated that the capability of the DNN acts well as on a different database. The author noted that this study is limited to single-lead ECG compared to the standard 12-lead ECG.

Nurmaini et al. (2020) present a low computational 1D-CNN based on 10 fold cross-validation strategy to classify ECG signals into two and three classes, where the two classes contain only normal sinus rhythm (SR) and atrial fibrillation (AFIB). The three classes consist of SR, AFIB, and non-AFIB. They conducted their experiments using three datasets that are available publicly such as MIT-BIH Malignant Ventricular Ectopy, MIT-BIH Atrial fibrillation, Physionet Atrial Fibrillation, and one external

database from an Indonesian hospital. Discrete Wavelet Transform (DWT) was used to eliminate the noise (artifact). Normalization and signal segmentation of 9 seconds produced better performance and achieved accuracies of 99.98% for the two classes and 99.17% for the three classes.

Now, will discuss the second category of the related work, which is about heart rhythm classification using 12-lead ECG with DL models.

3.2 Heart rhythm classification using DL from 12-lead ECG data

The following works apply DL models used to classify heart rhythms from 12-lead ECG signals.

Acharya et al. (2017a) have applied 11-layer convolutional neural network (CNN) for heart attack called Myocardial Infarction (MI) beats classification. This study conducted using Physikalisch-Technische Bundesanstalt diagnostic ECG database (PTBDB). PTBDB database consists of 12-lead signals of 148 MI and 52 normal data. Lead II was only used in this study in which they validate the proposed method with two datasets with and without noise. They removed the noise using Daubechies wavelet 6 mother wavelet function. Each ECG signal was segmented using normalization with Z-score normalization. Then, it was trained by 10-fold cross-validation technique which achieved accuracies of 93.53% and 95.22% with noise and without noise, respectively. This explicit that the method still can classify well by noisy ECG beat.

Tan et al. (2018) applied an algorithm of CNN with LSTM for normal and coronary artery ECG classification. Lead II ECG data was preprocessed and resampled for consistency (Hz), next applied (discrete wavelet transform) Daubechies wavelet 6

mother wavelet function to ensure noise removal, consequent, data segmented to 5 seconds without detecting the R-peak, and finally generate augmentation procedures such as re-normalize ECG segment before forwarding it to the training stage. PhysioNet and St Petersburg INCART 12-lead Arrhythmia Database were used in this study which accomplished overall performance accuracy of 99.85%.

Yildirim et al. (2020) designed efficient Deep Neural Network (DNN) with high performance for heart rhythm classification using each single-lead signal of the 12-lead. They show a promising result on all the ECG 12-lead using the newly published database contains more than 10000 records. Mainly, two experiments were conducted and each lead was classified separately. The first experiment involves seven rhythms, and the second experiment contains rhythms that are merged due to the insufficient number of subjects in the database. Lead 2 obtained the highest lead among all other leads in the 12-lead ECG in both experiments, which achieved 92.24% and 96.13% for the first and the second experiment, respectively.

The proposed work is more relevant to the second category, i.e., classification of heart rhythms from 12-lead ECG data. However, this work is distinguished from the above work in that this work evaluates the effectiveness of every single-lead in a 12-lead ECG signal and a combination of all leads (12-lead). Moreover, the possibility of utilizing static data in addition to the ECG data to improve the classification performance is investigated. Table 1 displays the other related work summarizes along with the features.

Chapter 4: Proposed Approach

This work performs a comprehensive evaluation of the 12-lead ECG to classify normal and abnormal heart rhythms. A deep learning technique based on 2D-CNN have been adapted to adequate for combined lead and single-lead. This allows to evaluate the relative importance of combined lead and single-lead from the dynamic ECG time-series data. Also, an MLP architecture have been developed to classify only the static data from the ECG data for investigatigating the effect of using static data classification. Moreover, a novel hybrid 1D-CNN-BiGRU-BiLSTM architecture has been proposed, which evaluates the combination of all leads (12-lead) from the dynamic ECG time-series data. Finally, the possibility of combining static data with the dynamic ECG time-series data has been investigated by designing multi-modal deep learning that uses the proposed hybrid DL fused with the MLP model.

The overall architecture consists of ECG database, data preprocessing, sampling, model deployment, and model evaluation. Two types of ECG databases were used, namely, the dynamic time-series, and the basic ECG measurements static data. The ECG database is passed to a data pre-processing stage to perform the data cleaning, transformation, and reduction. Then, it is sampled to training, validation, and testing data in order to be forwarded to the model deployment phase. The model deployment phase provides all the models that are developed to answer the thesis questions. Each of the models has been evaluated based on the accuracy and F1-score. Finally, the average evaluation result is reported by cross-validation. However, the above-mentioned detection and classification model architectures are evaluated on real-world patient data and their performances are compared in Chapter 6. Figure 7 illustrates the overall system architecture of the heart rhythm classification.

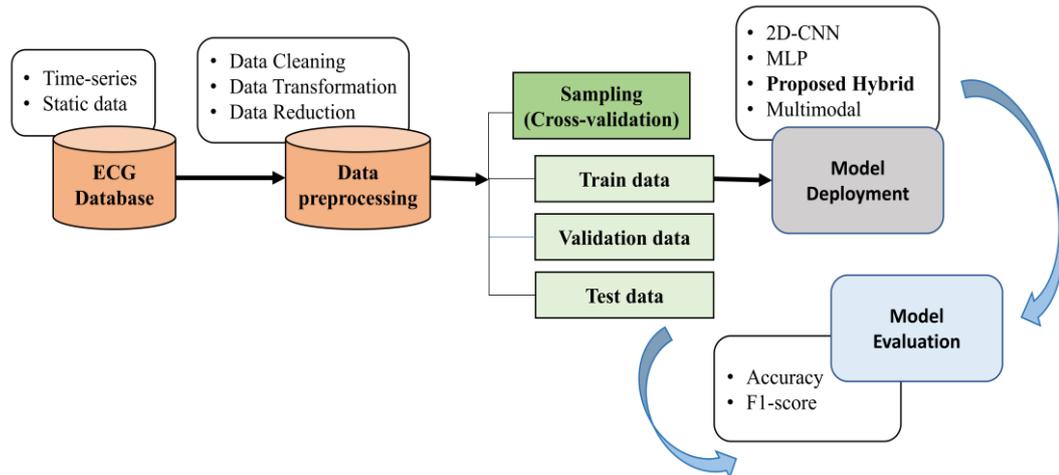


Figure 7: Overall system architecture of the proposed heart rhythm classification

The next sections demonstrated the arrangements of 2-Dimensional Convolutional Neural Network (2D-CNN) for combined and single-lead, proposed hybrid (1D-CNN-BiGRU-BiLSTM) for the combined lead as well as the multimodal deep learning (Proposed hybrid + combined with MLP) architectures.

4.1 2D-Convolutional Neural Network architecture

Justification of the architecture: the CNN model is considered as fast and most commonly used for time-series data in related literature (Ebrahimi et al., 2020; Khan et al., 2021). CNN with a fewer number of layers has the advantages of lower hardware specifications and ensures shorter time during training compared to their deeper counterparts (Gu et al., 2015). It also accommodates to optimize more hyperparameters and facilitates the training process.

Thus, the ECG time-series architectures have been developed for the combined and single-lead using the 2D-CNN. The first architecture is convenient for the combination of the 12-lead and the second architecture is adequate for the individual lead (single-lead).

Justification of combined and single-lead architectures: before selecting the architectures, many experiments were investigated, single layer to deeper network structure along with the number of parameters and hyper-parameters were observed. The structure of 2D-CNN in this work consists mainly of two convolutional layers. Primarily the two convolutional layers are suitable in both combined and single-lead data, but they slow down the training process of the combined lead architecture. Adding max-pooling into the combined lead and single-lead architecture will boost the processing power of the combined lead but at the same time, it will not fit for the single-lead architecture due to the different dimensions. Therefore, the max-pooling was added only for the combined lead architecture after each convolutional layer to increase the processing power. In addition, the kernel size of the first convolutional layer for combined lead has been increased to ensure better performance. All the other parameters and hyperparameters are not adjusted. That is, the architectures are convenient for both combined and single lead, and they are analyzed based on their performance and classification.

Subsection 4.1.1 shows the 2D-CNN architectures for combined lead and subsection 4.1.2 illustrates the 2D-CNN for single-lead.

4.1.1 2D-CNN architecture for combined lead data

The architecture of the 2D-CNN model for the combined lead consists of two conv2D layers along with two max-pooling, the first convolution layer has 8 filters with a kernel size of 5 and a max-pooling size of 2. Subsequently, the second layer of conv2D uses a kernel size of 3 along with 16 filters. Next, a stride, max-pooling of the size of 2 applied to the max-pooling layer to produce down sample operation and extract essential features from the previous feature map. Additionally, it controls the

speed in training duration. The flatten layer transforms the two-dimensional matrix to be fed into the fully connected layers. Overfitting also plays an important role during training which can't be neglected. One drop layer with a rate of 0.5 was added which is proved to be a very effective technique for reducing the overfitting. Finally, all the neurons were connected as a fully connected layer to form a single output that computes the distribution of binary classification. Figure 8 expresses the combination of 12-lead structure and Table 2 briefly summarizes the following structure along with layer parameters, output shape, and the number of parameters in detail.

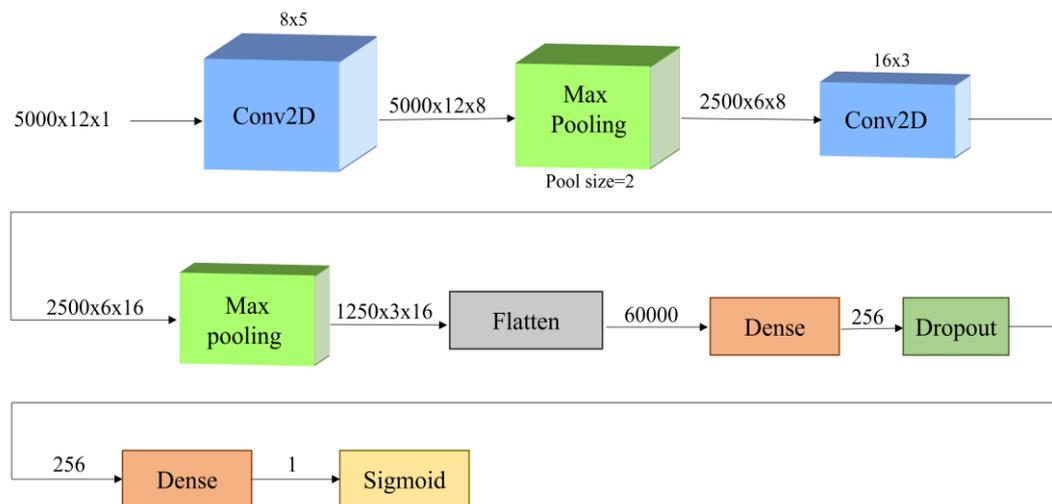


Figure 8: Combination of 12-leads architecture

Table 2: Layers and parameters of 12-lead architecture

Layer (Type)	Layer Parameters	Output Shape	Number of parameters
Conv2D	Filters=8, Kernel Size= 5 Padding = "same"	5000x12x8	208
Max Pooling	Pool size=2	2500x6x8	0
Conv2D	Filters=16, Kernel Size= 3 Padding = "same"	2500x6x16	1168
Max Pooling	Pool size=2 Stride=2	1250x3x16	0
Flatten		60000	0
Dense		256	15360256
Dropout	Rate=0.5	256	0
Dense		1	257

4.1.2 2D-CNN architecture for single lead data

The CNN model structure for classifying single-lead ECG contains two conv2D layers. The first conv2D layer had 8 filters, 5 kernel sizes, and padding of zero (Layer 1), the second layer had conv2D with 8 filters, kernel size of 1, and padding of zero. Finally, all the neurons were connected by flatten layer then fed into a 256 dense layer (Fully connected layer). Moreover, one dropout layer regularization with a rate of 0.5 was added, to eliminate the overfitting. However, other hyperparameters of the CNN model were not altered, which was set as the rates of the first experiment. Figure 9 displays the single-lead architectures and Table 3 summarizes the structure corresponding to layer parameters.

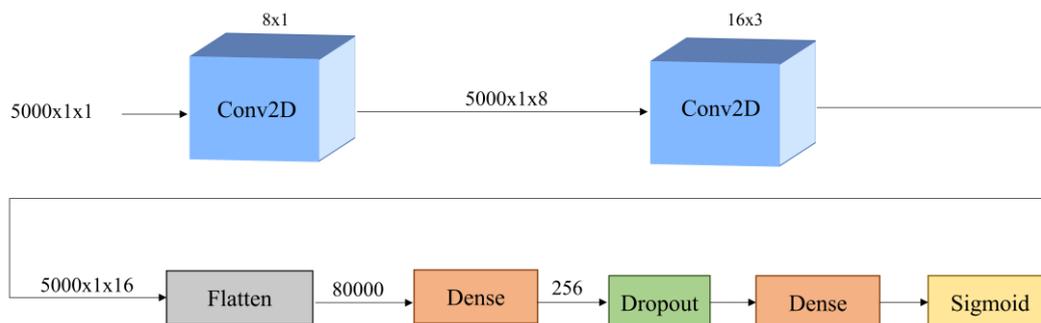


Figure 9: Single-lead architecture

Table 3: Layers and parameters of single-lead architecture

Layer (Type)	Layer Parameters	Output Shape	Number of parameters
Conv2D	Filters=8, Kernel Size= 1, Padding = "same"	5000x1x8	16
Conv2D	Filters=16, Kernel Size= 3, Padding = "same"	5000x1x16	1168
Flatten		80000	0
Dense		256	20480256
Dropout	Rate=0.5	256	0
Dense		1	257

4.2 Multi-layer perceptron architecture for only static data

Justification of the architecture: multi-layer perceptron (MLP) has the processing elements that learn from the relation input values of the static data. It has been widely used due to its capability of solving problems related to classification that are not linearly separated.

MLP was developed after studying possible cases that concern the model's selection such as the number of neurons and hidden layers. The number of neurons in the intermediate layer was tested between different numbers of neurons. Finally, the MLP model has two intermediate hidden layers of 6 neurons each, which were constituted by 11 input parameters, and a single output neuron. The single output

neuron has a sigmoid activation function which will further classify the heart rhythm based on the binary classification.

The ECG static values are forwarded as an input to the input layer of the MLP. The multiplication will be performed to each input data with the corresponding weight to forward it to the hidden layers, then it is passed to the output layer for binary classification. Figure 10 illustrates the multilayer perceptron (MLP) that is designed for investigating only the static data.

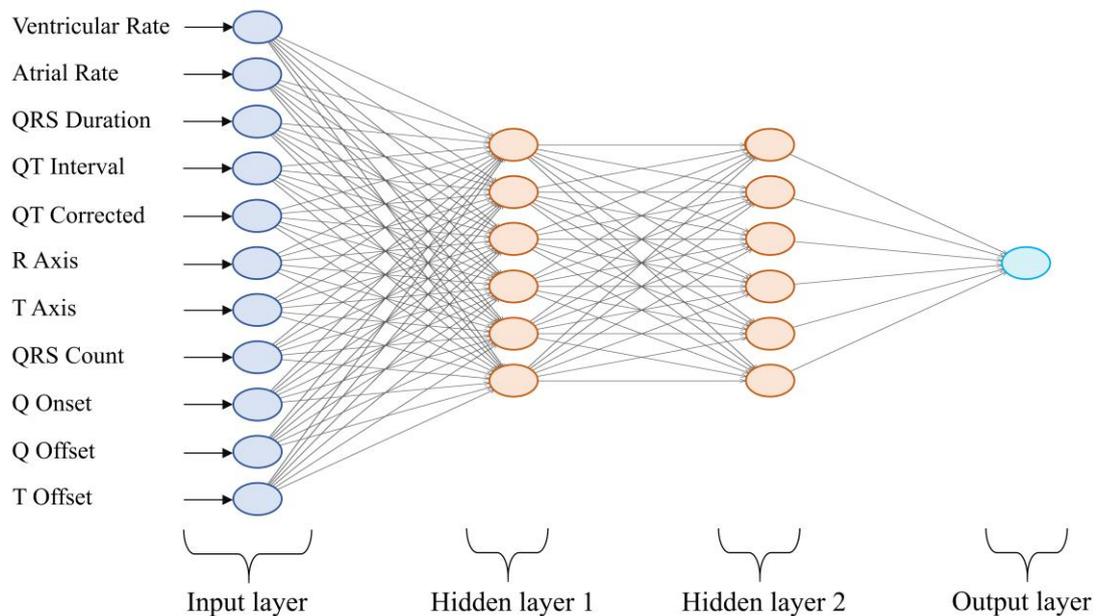


Figure 10: MLP architecture for static data

4.3 Proposed Hybrid 1D-CNN with-Bidirectional-GRU-and Bidirectional LSTM architecture

The hybrid architecture of a one-dimensional convolutional neural network (1D-CNN) with the bidirectional gated recurrent unit (BiGRU) and bidirectional long short-term memory (BiLSTM) represented as 1D-CNN-BiGRU-BiLSTM. This architecture has been termed as a proposed hybrid.

Justification of the architecture: hybrid deep learning structure is able to reveal promising results by considering a hybrid DL model that couples CNN with RNN to ensure robust classification (Hong et al., 2020). The proposed hybrid architecture uses the first layer of a 1-Dimensional Convolutional Neural Network that can process the time-series data through its image processing capabilities. This allows spatial representation learning from the time-series data. The BiGRU and BiLSTM layers are capable to learn from the temporal dependencies in the ECG data. Thus, the proposed hybrid method can capture the temporal dependencies from different lead data, as well as the inter-relationships among the lead data.

The proposed hybrid uses the combined 12-lead of the dynamic ECG time-series data for heart rhythm classification. However, the component of this model consists of different layers. Mainly, employing three consecutive one-dimensional convolutional neural networks (1D-CNN) with different filter sizes and a max-pooling layer following each 1D-CNN layer. The first Conv1D layer has 8 filters and a kernel size of 5 with padding. Then, a 1D-Max-pooling size of 3 is applied to extract essential features. Dropout with the dropout rate of 0.3 value, and Batch Normalization layer to accelerate the training and reduce the generalization error. Subsequently, the second layer of Conv1D convolved with 16 filters, kernel size of 11 using padding, and stride set to 2. Next, a 1D-Max-pooling size of 3 along with a 0.3 dropout rate. The third Conv1D layer has 32 number of filters and a kernel that has a size of 3, with padding and 2 stride length. 1D-Max-pooling size of 3 as well, coupled with 0.3 dropout rate, and then batch Normalization layer. Additionally, two bidirectional layers consist of one GRU layer and one LSTM layer with 128 number of units that can capture the dynamic information in serialized data, along with the Batch Normalization layer. Moreover, a flatten layer was added, then fed into the fully connected layers. Finally,

all the neurons were fully connected within the layer to form a single output that computes the distribution of binary classification, where the sigmoid was chosen as the activation function. Figure 11 illustrates this architecture in detail.

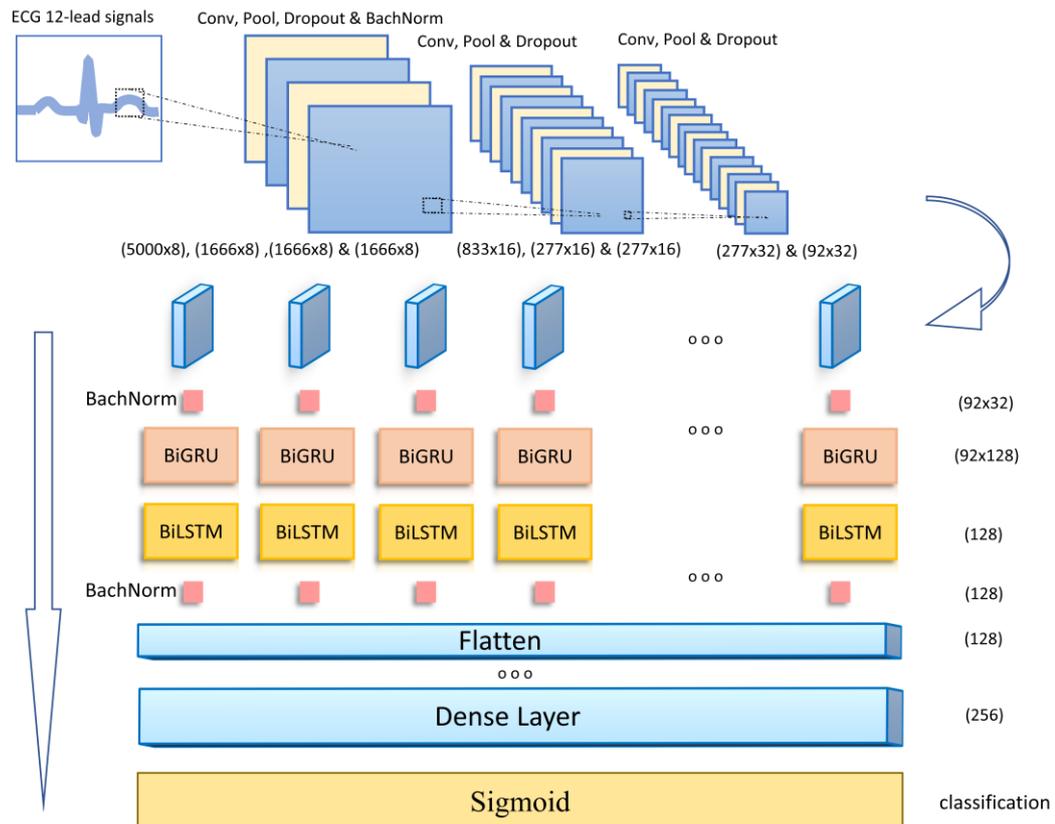


Figure 11: Proposed hybrid architecture

4.4 Multimodal deep learning (Proposed Hybrid combined with MLP)

Justification of the architecture: the multimodal architecture allows to deal with different modalities of information to improve the performance. This can be done by integrating multi-dimensional data. The most common method of combining those modalities sources by concatenating them through the concatenation layer. The multimodal allows capturing of the represented features of static data along with the

temporal dependencies of the dynamic ECG time-series data for classifying the heart rhythms.

A multimodal deep learning architecture has been conducted for heart rhythm classification using static and combined 12-lead ECG time-series data. As illustrated in Figure 12, the overall framework structure consists of two main inputs. That is, the first input contains statistical ECG time-series data, where the second input has the dynamic heart rhythm time-series data. The static data forwarded to the Multi-layer Perceptron (MLP) model and the dynamic ECG-time series data fed to the proposed hybrid model. Thus, the formation of the multimodal deep learning architecture is explained in detail in the upcoming paragraphs.

The structures of both the hybrid model for the combined 12-lead and the MLP model that uses the static data are explained previously in the proposed approach chapter.

In this current work, the sigmoid function was removed from the last layer of the models (MLP and proposed hybrid DL), then added to the last layer of the multimodal deep learning architecture. However, the static data are initially passed through the input layer of the MLP model and the corresponding dynamic time-series signals are forwarded through the proposed hybrid (1D-CNN-BiGRU-BiLSTM) model. Then, the two models are merged with a concatenate layer using the functional APIs. Functional APIs is more flexible to handle non-linear topology structure, therefore, functional APIs has the advantage over the sequential APIs to share multiple inputs or outputs of deep learning models.

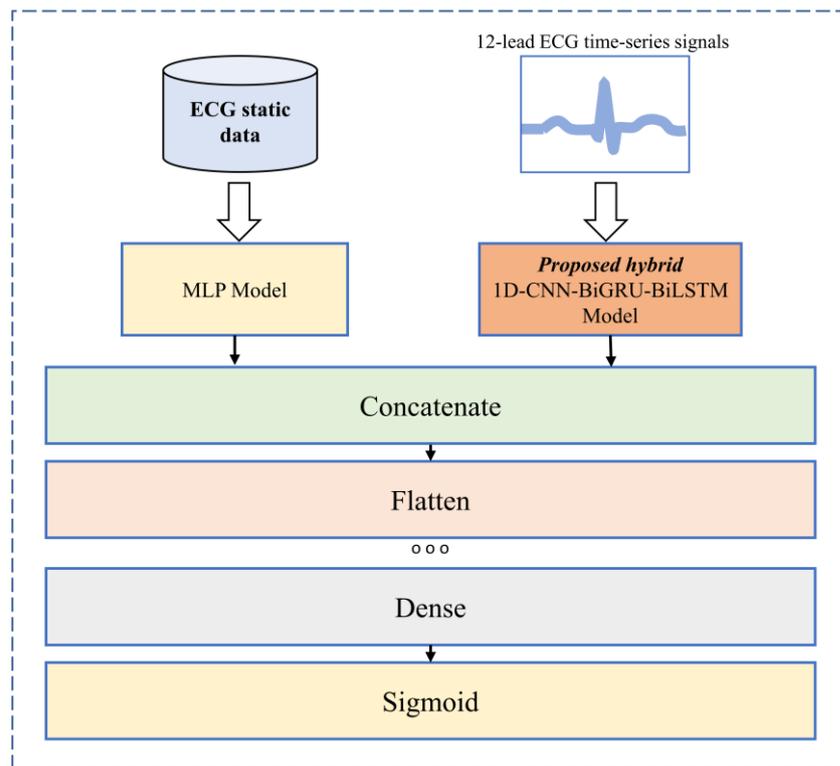


Figure 12: Multimodal (proposed hybrid fused with MLP) architecture.

After the concatenation layer, the flatten layer has been added and then the dropout of rate 0.1 was applied to prevent overfitting. Five dense layers were added, along with each dropout. The first dense layer consists of 256 neurons, where the second and the third dense layer contains 128 neurons, and the fourth layer is expressed by 64 neurons. Finally, all the neurons were connected as a fully connected layer in order to form a single output that can generate the distribution of binary classification using the sigmoid activation function. All the baseline of the model optimizers was discussed in the parameter setting Section 6.1.2.

Chapter 5: Dataset

This chapter discusses the dataset used in this study and demonstrates the analysis of data distribution that were analyzed from the database, followed by data preparation and preprocessing used in this regard.

There are different ECG databases that contain various subjects grouped as normal and abnormal, most of these databases are open source and available online. On the contrary, there are some of the ECG classification experiments in relevant literature have been employed using private databases which are not available to the public. However, the databases which are available to the public contain different subjects, sample rate (frequency), ages, rhythms, ECG time length, and either single-lead or more.

For the present study, the published database that accommodates a large number of subjects (10646) including male and female, the highest sample rate (500 Hz), and a large number of leads (12-lead) have been obtained. The database was collected by Chapman University and Shaoxing People's Hospital (Shaoxing Hospital Zhejiang University School of Medicine).

This database is collected with respect to the ECG time-series of the 12-lead from the electrical cardiac muscle activities (Zheng et al., 2020). Each lead in the 12-lead ECG has 500 Hz samples per second for 10 seconds, which is equal to 5000 values for each lead. Basically, leads contain low and high-frequency noise, consequently, due to the reflection of the interpreted signals with artifacts such as motion noise, electrode connection, and baseline wandering, which affects the readings of the raw data are processed by signal processing approaches: Butterworth low pass filter to clear

out frequency above 50 Hz, Locally Weighted Scatterplot Smoothing (LOSS) applied to free the effects of baseline wandering, and Non-Local Means (NLM) to eliminate the remaining noise. Finally, those processed data (Denoised) got to perform the preprocessing and prepare the data for the heart rhythm classification.

This database contains 11 rhythms such as Sinus Bradycardia (SB), Sinus Rhythm (SR), Atrial Fibrillation (AFIB), Sinus Tachycardia (AF), Sinus Irregularity (SI), Supraventricular Tachycardia (SVT), Atrial Tachycardia (AT), Atrioventricular Node Reentrant Tachycardia (AVNRT), Atrioventricular Reentrant Tachycardia (AVRT), and Sinus Atrium to Atrial Wandering Rhythm (SAAWR). In addition to 56 cardiovascular conditions. Figure 13 introduces the distribution of the 11 rhythms in the database.

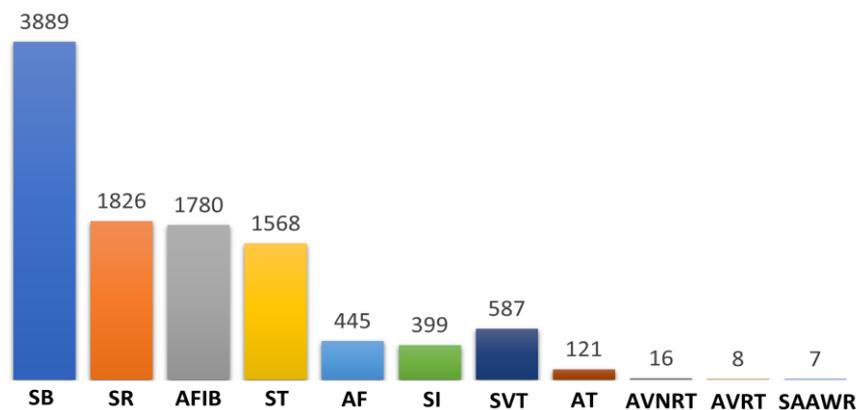


Figure 13: Distribution of rhythms in the database

Moreover, it includes ECG 12-lead signal and ECG basic measurements static data for each individual subject such as Gender, Patient age, Date of Birth, Atrial Rate, QRS Counts, QT Interval, Atrial Beat Rate, Ventricle Beat Rate, Q offset, and T offset, et. Table 4 illustrates all the static data attributes in the dataset.

Table 4: ECG static database attributes

Attributes	Types	Value Range	Description
File Name	String		ECG data file name (unique ID)
Rhythm	String		Rhythm Label
Beat	String		Other conditions Label
Patient Age	Numeric	0-999	Age
Date Of Birth	Date		Date of Birth
Gender	String	Male/Female	Gender
Ventricular Rate	Numeric	0-999	Ventricular rate in BPM
Atrial Rate	Numeric	0-999	Atrial rate in BPM
QRS Duration	Numeric	0-999	QRS duration in msec
QT Interval	Numeric	0-999	QT interval in msec
QT Corrected	Numeric	0-999	Corrected QT interval in msec
R Axis	Numeric	-179~180	R axis
T Axis	Numeric	-179~181	T axis
QRS Count	Numeric	0-254	QRS count
Q Onset	Numeric	16 Bit unsigned	Q onset (In samples)
Q Offset	Numeric	17 Bit Unsigned	Q offset (In samples)
T Offset	Numeric	18 Bit Unsigned	T offset (In samples)

5.1 Analysis of data distribution

The distribution of different rhythms based on age groups and gender has been analyzed from the database. Table 5 reports this distribution of all rhythms and age categories group. For instance, the rhythm SB contains 3889 patients, those patients engaged 2481 males and 1408 females, and 11 individuals from the overall aged between 91 to 98. The second column clearly demonstrates the total values of all presented rhythms, where SB rhythm has the greatest number among all rhythms.

This analysis will help to identify the prevalence of different heart rhythm problems at different ages which will give a good understanding of developing a more effective classification model (Khan et al., 2011). In particular, the highest frequency age group for each rhythm can be chosen to train a classification model that is believed to be more effective than developing a model without considering the age of the patient. However, this requires further investigation and more data which could be explored in the future.

Table 5: Distribution of rhythms based on age group and gender

Rhythm	Total	Male	Female	Age Group e.g, 5 (years) – 10 (years)									
				5 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71 - 80	81 - 90	91 - 98
SB	3889	2481	1408	0	28	112	208	498	909	1074	479	169	11
SR	1826	802	1024	2	26	111	175	285	389	373	201	84	3
AFIB	1780	1041	739	0	0	0	6	50	135	367	509	458	56
ST	1568	799	769	41	48	119	120	177	239	286	191	149	13
AF	445	257	188	0	0	3	4	20	39	92	120	80	26
SI	399	223	176	56	66	57	43	38	27	31	28	6	1
SVT	587	279	308	1	9	40	53	102	99	107	57	48	12
AT	121	64	57	3	2	5	2	4	10	23	39	16	7
AVNRT	16	4	12	0	0	2	1	0	5	4	0	3	0
AVRT	8	5	3	0	0	1	0	2	0	3	2	0	0
SAAWR	7	1	6	0	2	1	0	1	0	0	1	2	0
Total	10646	5956	4690	103	181	451	612	1177	1852	2360	1627	1015	129

The section is further divided into two subsections to analyze the distribution of rhythms based on gender and age group. Subsection 5.1.1 analyzes the distribution of rhythms based on gender, while Subsection 5.1.2 analyzed based on age group.

5.1.1 Based on gender

In this section, the rhythms are distributed based on gender and displayed into a pie chart that determines information about the proportion of the patients based on each rhythm.

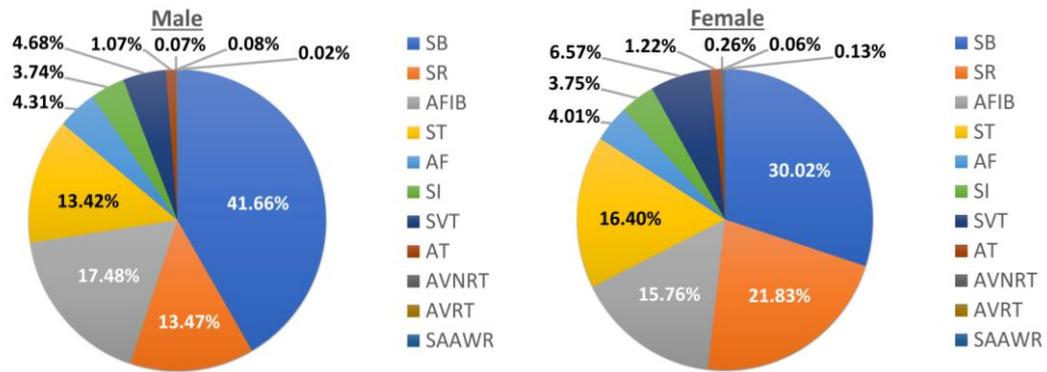


Figure 14: Rhythms distribution of male (left) and female (right)

These pie charts in Figure 14, represent the male at left and female at right, which indicate the distribution of total males and total females in the database. It is clear from the charts that the SB has the majority in both males and females, where it occupies 41.66% of the total males and 30.02% of the total females. The SR consumption represents 13.47% of the total males, and 21.83% of the total females. AFIB shares consumption category of 17.48% of total males, and 15.76% of total females. Followed by ST, which has 13.42% of total males and 16.40% of the total females as well. Then, the SVT at 4.68% of the male's total, and 6.57% of the female's total. AF and SI share around 4% of the total distribution in both males and females. Approximately, one percent of AT in both females and males. Finally, AVNRT, AVRT, and SAAWR have less than 15 patients for males and females which represent less than 0.1 percent.

5.1.2 Based on age group

In this section, all the rhythms are distributed based on age group segment correlation and represented them into a pie chart that provides information about the proportion of the patients based on each rhythm. Figures 15-20 are showing the distributions of ages group graphically.

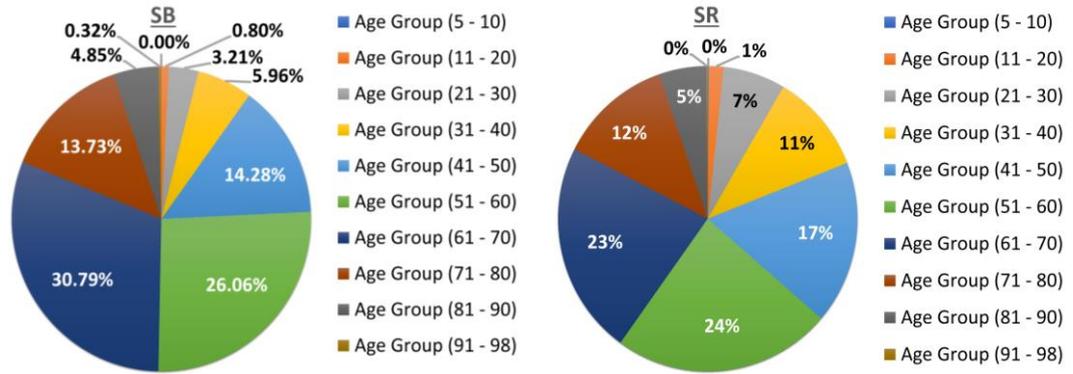


Figure 15: Age group distribution of SB (left) and SR (right)

In terms of the most significant feature, the largest group category of SB was the group aged 61 to 70, which is represented as 30.79%. On the other hand, the majority group of the SR is accounted as 23.59% between ages 51 to 60. The next segment of the SB majority was announced at 26.06% and aged between 51 to 70. In contrast, the SR group age between 61-70 achieved a consumption of 22.62%. In addition, those aged between 41 to 50 in the SB shares 26.06 percent. Moreover, the other age groups contributed lower rates which can be clearly comprised in the above pie charts.

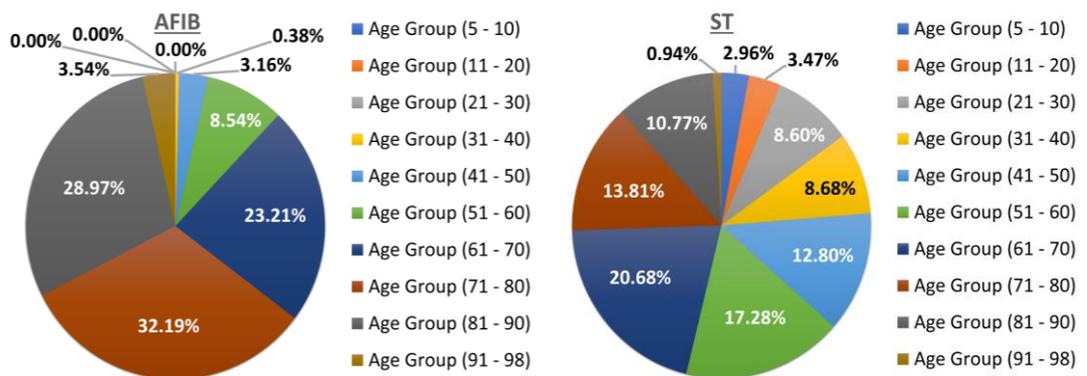


Figure 16: Age group distribution of AFIB (left) and ST (right).

In the charts above, the age group from 71 to 80 consumed the most for AFIB at a rate of 32.19%. On the flip side, the most significant group age found at ST is between ages from 61 to 70. The subsequent important segment of AFIB for those aged between 81 to 90 at the percentage of 28.97. In contrast, the ST participates at a rate of 17.28% for ages between 51 to 60. The third best results are indicated for those aged between 61 to 70 for AFIB and ages category of 71 to 80 for ST. However, the other distribution of age categories is concluded in the above charts.

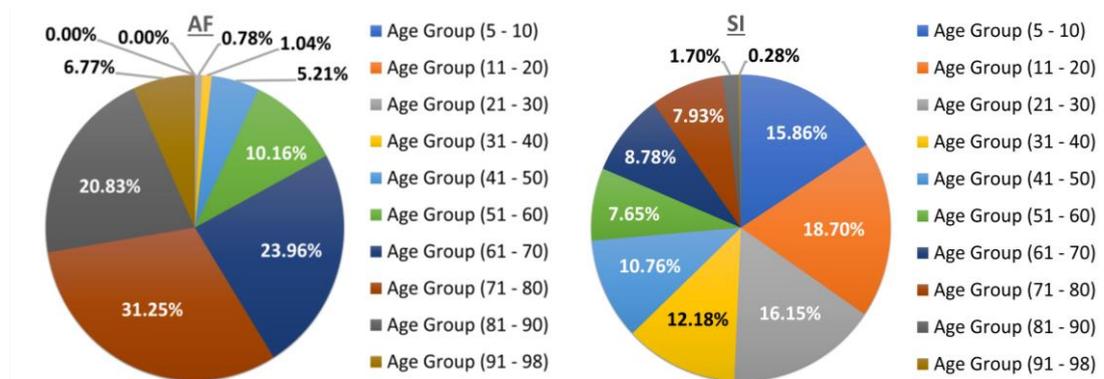


Figure 17: Age group distribution of AF (left) and SI (right)

The side-by-side pie charts above illustrate the age groups of AF and SI, where the most significant age group reported at 31.25% between 71 to 80 in AFIB. The corresponding majority age group of SI indicated between 11 to 20 at 18.70%. The second majorities of AF and SI are indicated at 23.96% for ages 61 to 70 and 16.15% for ages 21 to 30, respectively. The third majority of the age group of AF shares a percentage of 20.83% between 81 to 90. On the other side, SI has 15.86% aged between 5 to 10. The other age distribution depicts the rest of the age categories as shown above.

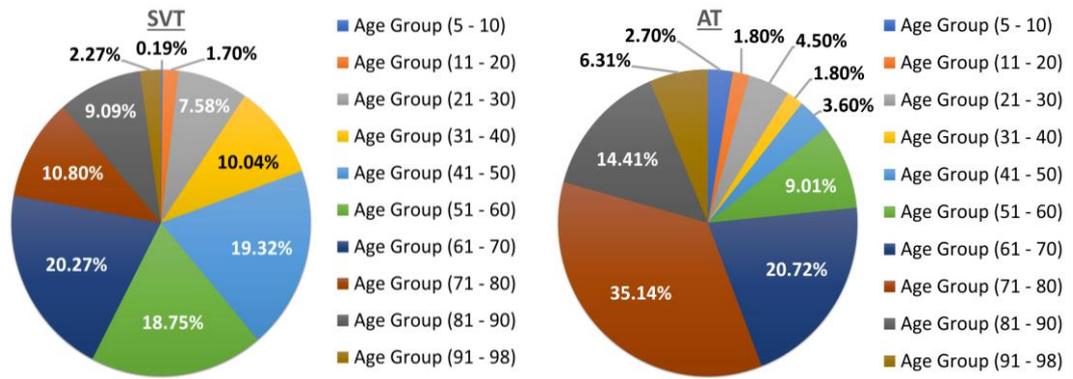


Figure 18: Age group distribution of SVT (left) and AT (right)

The above pie charts show the age group distribution of SVT and AT. The largest group category of SVT is grouped at ages between 61 to 70, which represented a proportion of 20.27%. On the other hand, AT accounted for 35.14% between ages 71 to 80. The second majority of the SVT segment is indicated at 19.32% for ages between 41 to 50. In contrast, the AT group age between 61 to 70 results in consumption of 20.72%. Furthermore, those aged from 51 to 60 in SVT announced 18.75% and 14.41% of AT categorized between ages 81 to 90. Referring to the above charts, the other age groups contributed at lower overall rates.

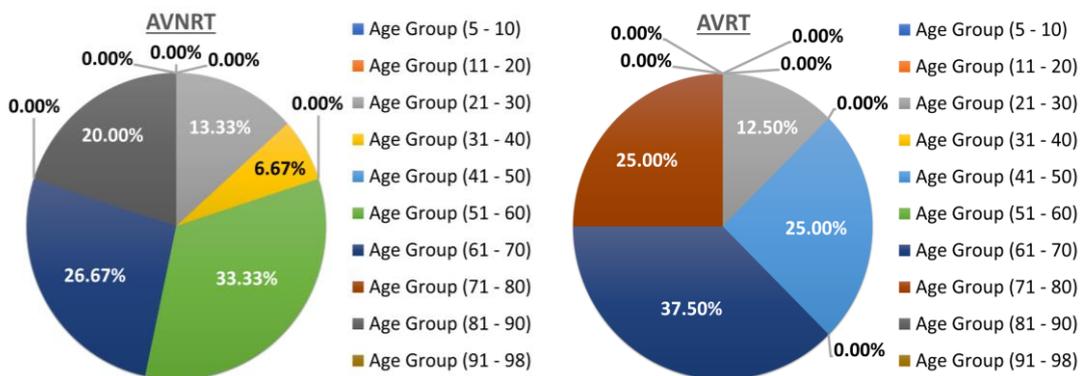


Figure 19: Age group distribution of AVNRT (left) and AVRT (right)

The pie charts side-by-side express the AVNRT and AVRT group age that contains a total of 16 patients in AVNRT and 8 patients in AVRT. However, the representation of 33.33% has a maximum of 5 patients categorized between ages 51 to 60. On the other hand, the AVRT majority defined as 37.50% which consists of 3 patients from ages 61 to 70. The other age categories are displayed clearly in the above charts.

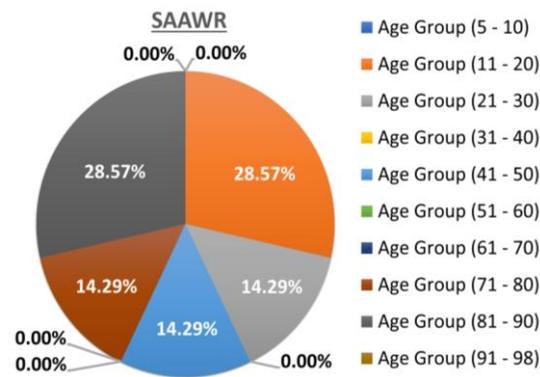


Figure 20: Age group distribution of SAAWR

The SAAWR consumed the lowest consumption held among all age group distribution. It contains seven patients; two segments are represented as 29% contain only 2 patients. The other 3 segments indicated as 14% has one patient each.

5.2 Preprocessing and data preparation

In the present work, various experiments were performed on the new ECG dataset that contains more than 10,000 individual subjects. The rhythms that contain more than 1,500 subjects were selected such as SB, SR, AFIB, and ST. Then, the data was organized into three different subsets based on these classes as shown in Figure 21. The first subset consists of normal sinus rhythm (SR) data and the anomaly sinus bradycardia (SB) data. This subset is referred to as SR-SB. The second subset consists of normal sinus rhythm (SR) along with the sinus tachycardia (ST) which is indicated

as SR-ST. The third group consists of normal sinus rhythm (SR) and atrial fibrillation (AFIB). This can be represented as (SR-AFIB). There are in total 1825 sinus rhythm (SR), 3888 sinus bradycardia (SB), 1564 sinus tachycardia (ST), and 1780 of AFIB. This method allows testing abnormal types of cardio conditions along with the normal rhythm to observe the distinguishing behavior between normal and anomaly rhythms.

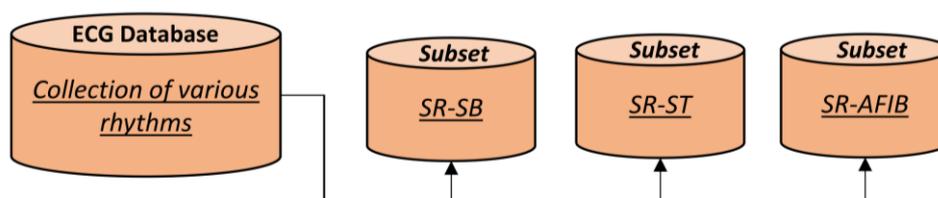


Figure 21: Organization of the dataset

During the preprocessing stage, vital recommendations for the data preparation have been suggested in the relevant work to use normalization techniques in order to improve the attainment of the model, where all the selected rhythms in this work contain large scale ECG subjects, therefore, those subjects were normalized to a min-max normalization to rescale features variables and eliminate the scaling problem before forwarding the trained dataset to the model. Normalization function ensures amplitude scaling of the signal range constrained between 0 and 1 without affecting the morphology of the signal. Moreover, during the data cleaning stage, some of the individual records contain null values which were carried out from the dataset as follows: two SB and five ST records were excluded.

Chapter 6: Experiments and Results

This chapter explains the experimental setup, followed by the evaluation setup. It also comes up with different experiments to answer the three research questions, mainly, what is the importance of each lead in a 12-lead ECG in classifying heart rhythms, what is the importance of static data in classifying heart rhythms, and whether static data can be combined with ECG data to improve classification accuracy.

In this regard, the classification experiments are essentially attempted by three different subsets, so that each classification experiment is repeated for each subset such as SR-SB, SR-ST, and SR-AFIB. Four different experiments are conducted as follows. First, experiment with the dynamic ECG time-series data using the combination of 12-lead, in addition to each individual leads (single-lead) of the 12-lead with the help of the 2D-CNN. This study aims to show that the combination of 12-lead ECG achieves better results than single-lead and it has been proved by statistical inference using t-test. Therefore, the rest of the experiments were further evaluated with the combined 12-lead due to both its high accuracy and the fact that it has been proved statistically. Second, classify only static data that are available in the same subsets using MLP. The study is aimed to investigate the performance of only static data classification. Third, experiment with an effective hybrid DL (proposed hybrid) model for the classification of the heart rhythm, which achieved superior performance among other architectures of this study. Finally, the static data are fused with the dynamic ECG time-series data of the 12-lead combined by using the multimodal framework. This allows examining the effectiveness of the static data with dynamic ECG time-series data. The finding emphasizes that the static data decreases the performance ability compared to the proposed hybrid DL model. This evidence can be helpful for the future research study.

6.1 Experimental setup

The structure of this section is described as follows: The hardware and software subsections presents the hardware and software specifications, parameter setting subsection summarizes the optimization algorithm and the activation functions that were used in the experiments. Finally, the overfitting avoidance subsection that highlights the technique used to avoid overfitting during the experiments.

6.1.1 Hardware and software

All the experiments were executed on CIT DGX-1 server, which consists of 8 Tesla NVIDIA GPUs with 32 GB RAM each. The models were implemented using Python 3.7.6, Keras 2.3.1, Scikit-learn, and other deep learning related dependencies.

6.1.2 Parameter setting

The Rectified Linear Unit (ReLU) was mainly used as an activation function in all the experiment's convolutional layers, in addition to the MLP layers of the static data experiment as well. The ReLU activation function has been used with default arguments such as maximum activation value set to 'none', and zero values for both negative slope and threshold. Batch normalization was added in some of the layers of the proposed hybrid (1D-CNN-BiGRU-BiLSTM) architecture and set to -1 and 0.99 for axis and momentum, respectively. Furthermore, the sigmoid activation function is applied to the last layer in each architecture to ensure binary classification. Differently, all the sigmoid activation functions were removed from the 2D-CNN and proposed hybrid before merging them using concatenate layer to deploy the multimodal architecture, where the sigmoid function was only added in the last layer of the multimodal architecture. Researchers use gradient descent optimization algorithms

such as Nadam, Momentum, AdaMax, and Adam to find the local minimum values of a given function. In the conducted experiments, adam optimizer achieves good results which were coupled with, learning rate, beta 1, beta 2 of 0.001, 0.9, 0.999 respectively.

Various empirical studies suggested different cost functions for different arrhythmia classification such as focal loss function (Romdhane et al., 2020) and batch-weighted loss function (Sellami & Hwang, 2019), hence, in this approach, the Mean Square Error (MSE) loss function accomplished the best-presented results among all other cost functions. All those parameter adjustments ensure the optimal outcome results. The training was derived of 50 epochs in all the conducted experiments, except for the MLP experiment where it has been set to 100 epochs due to the continuous error dropping, to guarantee the minimum error rate, consequent, the validation was computed subsequently after each round.

Bengio (2012) presented a practical recommendation on choosing the hyperparameters of a model. A grid search has been performed to gather the best parameters and hyperparameters to improve the optimal performance of the networks. The accuracy values have been determined by evaluating the classification accuracy with respect to the loss function. Also, various network architectures, gradient descent optimization algorithms, dropouts, and loss functions have been evaluated. Accordingly, by comparing the experimental results of multiple cases that were tested, it is confirmed that the proposed parameters and hyperparameters obtain the best classification accuracy.

Two metrics of the performance measurements are estimated to evaluate the model performance such as Accuracy and F1-measure which were considered in each iteration.

6.1.3 Overfitting avoidance

In this thesis work, many techniques were used to avoid overfittings like the cross-validation strategy, which is a powerful preventative measure against overfitting. It allows tuning hyperparameters with the original training set. This is done in all the conducted experiments. In addition, regularization techniques are considered as an effective technique to prevent overfitting situations, it was performed in all the case studies architecture that have been described briefly in the proposed approach Chapter (Chapter 4). Moreover, the grid search method was performed to optimize various parameters and hyperparameters.

6.2 Evaluation setup

There are two main criteria for data splitting such as normal split and cross-validation. The normal split is mainly considered as a classic approach in which the data split randomly into training and testing sets. On the other hand, the cross-validation criteria are often preferred, the dataset splits into a number of given folds in order to divide the dataset to the corresponding to the selected folds. In the thesis experiments, 10 fold cross-validation were done. In each fold, there is 90% training data and 10% test data. The training data is further split into training and validation sets. The training set is used to train the network and the validation set to validate the model. Then the test data is used to evaluate the model. After completing the 10-fold, the average evaluation result is reported. No early stop strategy was used.

Evaluating machine learning algorithms is essential for any use case. There are many evaluation metrics used to evaluate the quality of the machine learning performance such as accuracy, F1-score, precision, recall, specificity, and ROC Curve (AUC). The accuracy is the most intuitive measure, it's a ratio of correctly predicted

to the total observations. F1-score is used when the false negatives and false positives are crucial. Precision is a measure of the true predicted positive to the total predicted positive. Recall (sensitivity) is the ratio of the true predicted positive to the total of the actual positive. The ROC area under the curve is a measure of the culmination of the model, which gives an idea about the true-positive rate for a given false-positive rate and provides a summary indicator of the classifier attainment.

However, during the approach iteration, the accuracy and F1-score metrics were estimated to evaluate the model performance. Accuracy works best when the false positives and false negatives have similar costs and it is mostly used when all the classes are equally balanced. The binary classification subsets are a disproportionate ratio of observations in each class, therefore, F1-score is the harmonic mean between precision and recall and it is used to ensure reliability when dealing with imbalanced data. The following performance metrics can be evaluated as follow:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Where:

TP: refers to the number of correctly predicted positive samples.

TN: refers to the number of correctly predicted negative samples.

FP: refers to the number of negative samples incorrectly predicted as positive.

FN: refers to the number of positive samples incorrectly predicted as negative.

6.3 Experiments using 2D-CNN of single-lead and combined lead classification

In this section, an experimental classification of the normal and abnormal rhythms for leads combined (12-lead) and single-lead are provided. Mainly, a 2D-Convolutional Neural network was used in this experiment, which is most commonly used for time-series data in related work. The purpose of this study, primarily to answer the first research question of the effective leads in the 12-lead ECG. The adapted 2D-CNN for both combined and single lead approach is introduced in more detail in the proposed approach chapter (Chapter 4).

Two experiments were attended, the first scenario is to classify the abnormalities based on the combination of all leads (12-lead combined), the second scenario is to identify which lead of the 12-lead is the most effective. Then comparing the performance of the 12-lead combined with the single-lead. Each subset is divided into training and validation in which each subset being compared separately. Ultimately, to confirm the results and analysis, the results of the subsets are compared and come with a hypothesis test (one-tail t-test) to prove the results.

6.3.1 Experiments with combined lead

The purpose of this section is to attend an experiment of 2D-CNN using all leads combined (the 12-lead). Figure 8 in Section 4.1.1 shows the architecture of the combination of all leads used in this regard. The section is divided based on the subsets that were described previously.

6.3.1.1 SR-SB subset

The performance results of the experiment that obtained for SR-SB subset classification achieved Accuracy and F1 Score as 98.43% and 96.73%, respectively.

Figure 22 briefly clarifies the accuracy and corresponding loss function during training and testing. Table 6 determines the training accuracies and validation accuracy concerning the epochs.

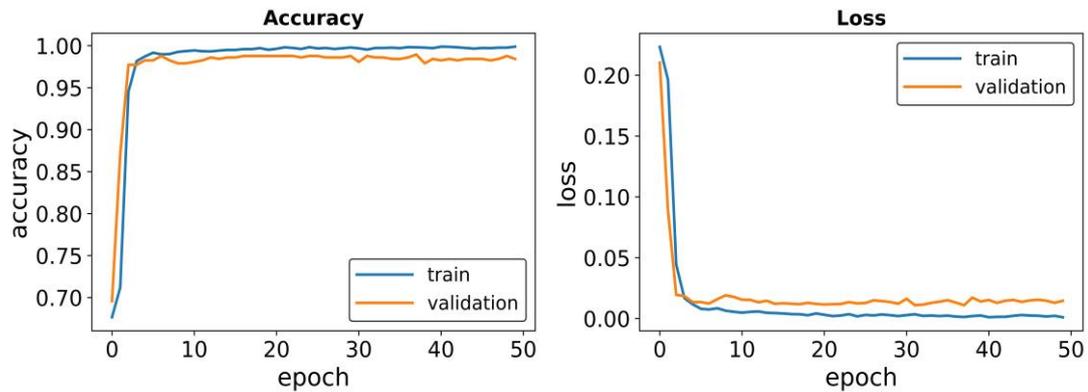


Figure 22: Combined lead SR-SB accuracy (left) and loss (right)

Table 6: Combined lead SR-SB performance measurements

Epochs	Validation	
	Accuracy	F1-score
10	0.9790	0.9590
20	0.9878	0.9776
30	0.9878	0.9738
40	0.9843	0.9672
50	0.9843	0.9673

6.3.1.2 SR-ST subset

The experiment that was conducted for the classification of the SR-ST subset, defines that the Accuracy and F1 Score obtained as 97.94% and 97.71%, respectively. Figure 23 displays the accuracy and loss function performance. Table 7 exhibits the performance ranking.

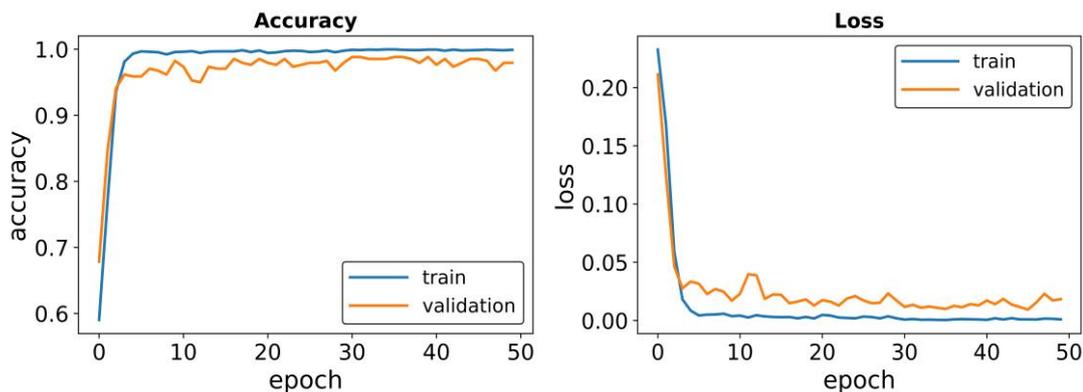


Figure 23: Combined lead SR-ST accuracy (left) and loss (right)

Table 7: Combined lead SR-ST performance measurements.

Epochs	Validation	
	Accuracy	F1-score
10	0.9823	0.9816
20	0.9853	0.9867
30	0.9794	0.9786
40	0.9882	0.9874
50	0.9794	0.9771

6.3.1.3 SR-AFIB subset

The classification experiment that uses the SR-AFIB subset achieved Accuracy and F1-score of 83.10% and 83.46%, respectively. SR-AFIB classification reveals that it achieves high accuracy during the training, but low accuracy during the validation. This overfitting could be shown from the training and validation sets as depicts in Figure 24. Table 8 refers to the experiment performance outcomes.

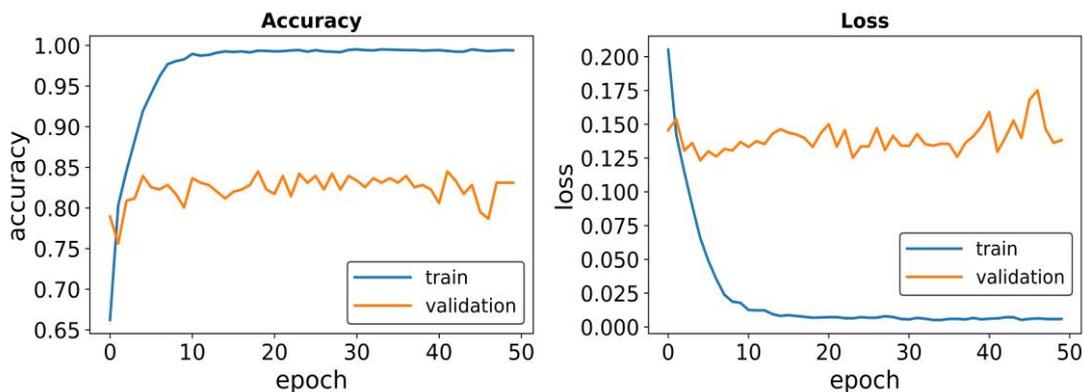


Figure 24: Combined lead SR-AFIB accuracy (left) and loss (right)

Table 8: Combined lead SR-AFIB performance measurements.

Epochs	Validation	
	Accuracy	F1-score
10	0.8006	0.8027
20	0.8227	0.8278
30	0.8393	0.8464
40	0.8227	0.8144
50	0.8310	0.8346

6.3.2 Experiments with single-lead

The purpose of this experiment is to see the effect of each lead by using 2D-CNN adapted for a single-lead. This section provides classification analysis for the leads available in the subsets of SR-SB, SR-ST, and SR-AFIB. The best behavior of the single-lead CNN network is clarified in Figure 9 in Section 4.1.2.

6.3.2.1 SR-SB subset

The experiment emphasizes that the best three highest leads out of the 12-lead for SR-SB classification in terms of Accuracy and F1-score are lead 4, lead 1, and lead 2, respectively. These highest leads achieved more than 98% and 96% of Accuracy and F1-score, respectively. Other leads obtained less than 98% of Accuracy and F1-

score. Figure 25 symbolizes the validation accuracy along with its corresponding loss function through epoch 50.

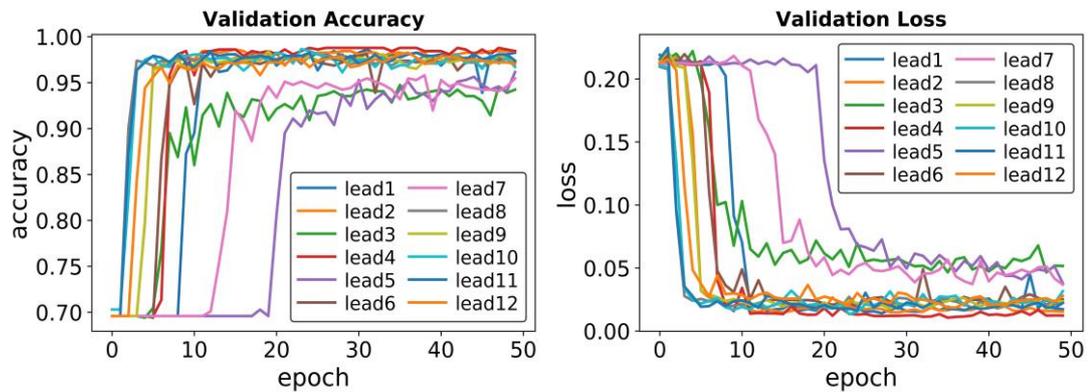


Figure 25: Validation of all the individual leads SR-SB

6.3.2.2 SR-ST subset

The investigation of the SR-ST subset classification examined that the highest three single-leads in terms of Accuracy and F1-score achieved as follows: lead 9, lead 8 and then lead 10, respectively. It attained more than 93% of Accuracy and 92% of F1-score. The other single-lead obtained more than 85% and 82% of Accuracy and F1-score, respectively. Figure 26 displays the accuracy and cost function values.

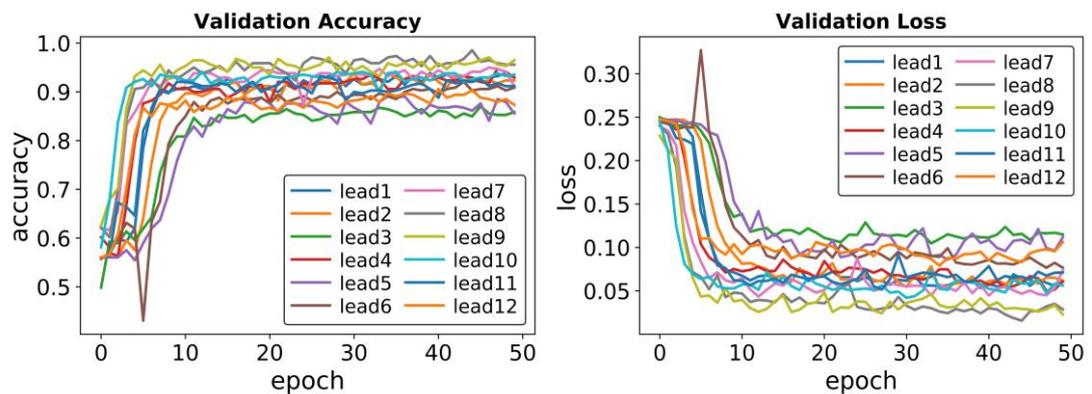


Figure 26: Validation of all the individual leads SR-ST

6.3.2.3 SR-AFIB subset

An experiment that has been employed for the SR-AFIB subset classification captured the highest three leads in terms of Accuracy and F1-score as lead 8, lead 2, and lead 9, respectively. These highest leads achieved Accuracy and F1-score above 70%, while the others achieved between 70%-60% of accuracy. Figure 27 captures the accuracy of all leads and cost function.

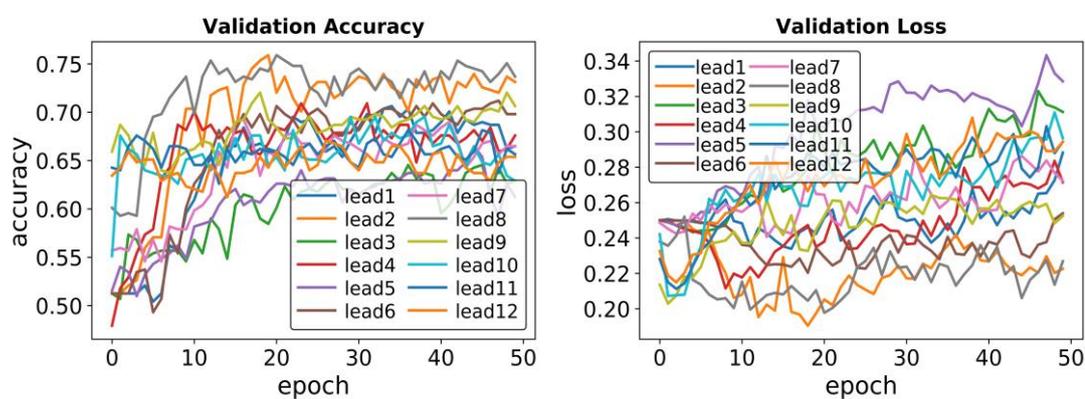


Figure 27: Validation of all the individual leads SR-AFIB

6.3.3 Subsets comparison for a combined and single-lead

It can be noticed from the experiments of combined and single-lead that the combined lead achieved better results compared to single-lead. It is interesting to observe that lead 4 of SR-SB achieves the same performance as the combined lead in terms of Accuracy, and less performance in terms of F1-score. Table 9, 10, and 11 summarize the results obtained for combined lead and single-lead.

As it can be observed from the Tables that the SR-SB and SR-ST subsets classification obtained acceptable performance compared to SR-AFIB in both combined and single-lead. In general, SR-AFIB performed well in training, but it didn't generalize well during validation. This overfitting can be avoided if the characteristic

of the AFIB is mainly understood. Unfortunately, AFIB contains characteristics of rhythm pattern conditions such as atrial flutter (AF) or sinus tachycardia (ST), etc. (Nurmaini et al, 2020). In addition, AFIB signals have irregular time elapse between RR intervals which means that the inconsistencies in the RR interval and other characteristics in AFIB classification will affect the outperform of the model measurements. Some reserchers have proposed methods for detecting arrhythmias using RR interval from ECG data (Kim et al., 2011; Lian et al., 2011). However, studies reveal promising results that considering the capability of CNN coupled with the RNN model to ensure robust detection of AFIB (Oh et al., 2018b; Murat et al., 2020). Therefore, the CNN structure can be integrated with the RNN to form a hybrid deep learning network for more accurate diagnosis. This possibility have been investigated and proposed an effective hybrid deep learning architecture, as explained in the coming Section (6.5).

As a result, several important observations can be made. First, the SR-SB and SR-ST achieved higher accuracy compared to the SR-AFIB. Second, lead 4 achieved the same accuracy as the combined lead but still with a lower F1-score in the SR-SB subset. For all other subsets, combined lead performed better than all the single-lead. Therefore, it proves the hypothesis that “combined lead performance is better than single-lead”. The result of the test is shown after the tables.

Table 9: SR-SB individual leads performance measurements comparison

Leads	Accuracy	F1 Score
All Leads	0.984266	0.974359
Lead 1 (I)	0.982517	0.971751
Lead 2 (II)	0.980769	0.968116

Table 9: SR-SB individual lead performance measurements comparison (continued)

Leads	Accuracy	F1 Score
Lead 3 (III)	0.942308	0.900901
Lead 4 (aVR)	0.984266	0.974063
Lead 5 (aVL)	0.961538	0.935673
Lead 6 (aVF)	0.972028	0.953488
Lead 7 (V ₁)	0.954545	0.922619
Lead 8 (V ₂)	0.973776	0.956268
Lead 9 (V ₃)	0.973776	0.956522
Lead 10 (V ₄)	0.966990	0.944625
Lead 11 (V ₅)	0.973776	0.9566522
Lead 12 (V ₆)	0.970280	0.951009

Table 10: SR-ST individual leads performance measurements comparison

Leads	Accuracy	F1 Score
All Leads	0.979351	0.976271
Lead 1 (I)	0.929204	0.918367
Lead 2 (II)	0.923304	0.911565
Lead 3 (III)	0.858407	0.840000
Lead 4 (aVR)	0.929204	0.918367
Lead 5 (aVL)	0.855457	0.829268
Lead 6 (aVF)	0.908555	0.896321
Lead 7 (V ₁)	0.911504	0.895833
Lead 8 (V ₂)	0.955752	0.94915
Lead 9 (V ₃)	0.964602	0.959732
Lead 10 (V ₄)	0.935103	0.926667
Lead 11 (V ₅)	0.911504	0.903226
Lead 12 (V ₆)	0.873156	0.856187

Table 11: SR-AFIB individual leads performance measurements comparison

Leads	Accuracy	F1 Score
All Leads	0.831025	0.840731
Lead 1 (I)	0.664820	0.709832
Lead 2 (II)	0.731302	0.739946
Lead 3 (III)	0.628809	0.625698
Lead 4 (aVR)	0.67590	0.69940
Lead 5 (aVL)	0.612188	0.621622
Lead 6 (aVF)	0.698061	0.728180
Lead 7 (V ₁)	0.664820	0.675603
Lead 8 (V ₂)	0.736842	0.755784
Lead 9 (V ₃)	0.706371	0.710383
Lead 10 (V ₄)	0.628809	0.676329
Lead 11 (V ₅)	0.656510	0.653631
Lead 12 (V ₆)	0.653740	0.670185

Statistical test to confirm the results: the hypothesis that “combined lead performs better than single lead” is further investigated using hypothesis testing. In particular, as it recognizes that the combined lead accuracies of SR-SB, SR-ST, and SR-AFIB are 98.42%, 97.93%, and 83.10%, respectively. The t-test experiment has been performed and it is found that the combined lead (mean = 0.93, standard deviation = 0.07, sample size = 36) is in fact statistically better than the single-lead (mean = 0.85, standard deviation = 0.13, sample size = 36). This difference was significant in terms of P-value of 0.0012, which is less than the α of 0.05, and test statistic (t) of 3.14 greater than the critical value of 1.66 (one-tail).

6.4 Experiment using MLP for only static data classification

This experiment has involved only the static data for heart rhythms classification to answer the second question (importance of static data in heart rhythm classification) by using the Multi-layer Perceptron (MLP) architecture. This empowers to investigate the effectiveness of the static data in the classification of heart rhythms.

The static data are available in the same subsets for each individual. It contains some demographic attributes (patient age, gender, date of birth, etc.), in addition to the ECG time-series statistical summary of the ECG time-series data. Static data are already described in Table 4 of Chapter 5. However, the statistical summary of the ECG time-series was only included as an input to the MLP model, while other feature characteristics e.g., age, gender, etc, were not involved. These statistical summaries have 11 features such as Ventricular Rate, Atrial Rate, QRS Duration, QT Interval, QT Corrected, R Axis, T Axis, QRS Count, Q Onset, Q Offset, and T Offset. During the experiment observation, it was noticed that the validation and training error continues dropping when the epoch was set to 50. Therefore, it is adjusted to epoch 100 in order to increase the possibility of terminating training based on the minimum error rate. Each experiment was repeated and compared based on each subset such as (SR-SB, SR-ST, and SR-AFIB).

6.4.1 SR-SB subset

The observation of the experiment using SR-SB subset classification obtained 98.25% and 96.85% in terms of Accuracy and F1-score, respectively. Figure 28 below shows the accuracy and loss across epoch 100. Table 12 reports the performance measurements of the SR-SB classification.

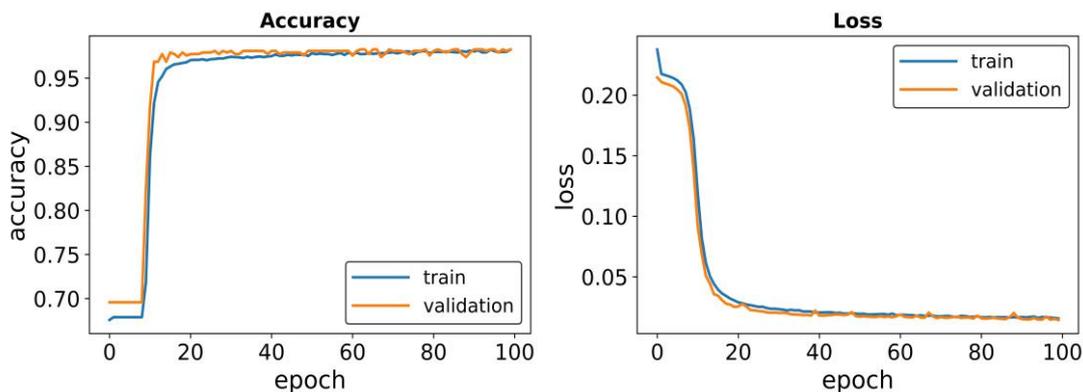


Figure 28: MLP SR-SB accuracy (left) and loss (right)

Table 12: MLP SR-SB performance measurements

Epochs	Validation	
	Accuracy	F1-score
20	0.9755	0.9553
40	0.9773	0.9577
60	0.9808	0.9643
80	0.9755	0.9546
100	0.9825	0.9685

6.4.2 SR-ST subset

The experiment of SR-ST subset classification attained 99.02% and 99.07% in terms of Accuracy and F1-score, respectively. Figure 29 illustrates the loss iteration and the corresponding accuracy to 100 epochs. Moreover, Table 13 reveals the performance rating of each epoch along with the Accuracy and F1-score.

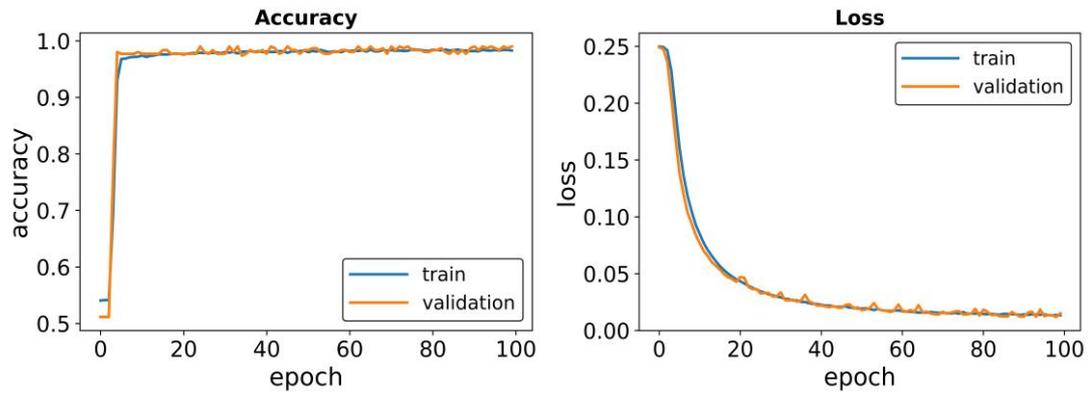


Figure 29: MLP SR-ST accuracy (left) and loss (right)

Table 13: MLP SR-ST performance measurement

Epochs	Validation	
	Accuracy	F1-score
20	0.9770	0.9782
40	0.9836	0.9835
60	0.9836	0.9835
80	0.9803	0.9808
100	0.9902	0.9907

6.4.3 SR-AFIB subset

The examination employed for the SR-AFIB subset classification captures the Accuracy and F1-score as 84.0% and 83.78%, respectively. Figure 30 depicts the accuracy and the loss cost plots during the 100 epochs iteration. Table 14 below shows the performance metrics of the SR-AFIB subset classification during the validation.

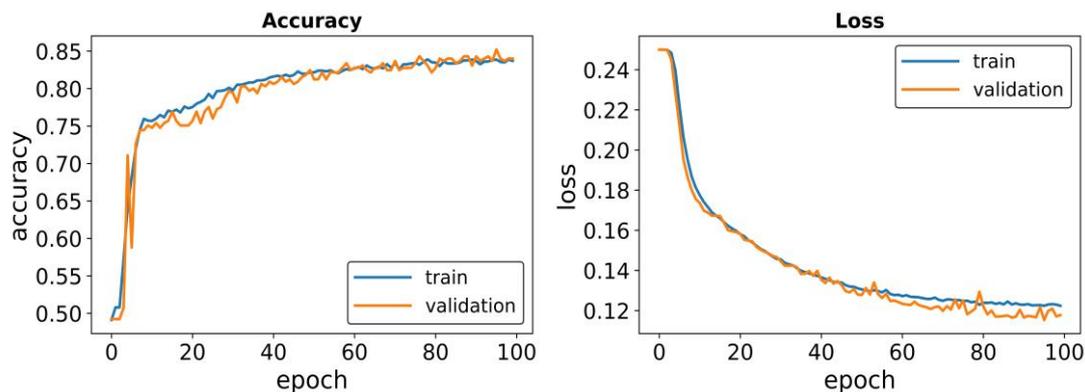


Figure 30: MLP SR-AFIB accuracy (left) and loss (right)

Table 14: MLP SR-AFIB performance measurement

Epochs	Validation	
	Accuracy	F1-score
20	0.7508	0.7673
40	0.8092	0.8183
60	0.8246	0.8224
80	0.8215	0.236
100	0.8400	0.8378

6.4.4 Subsets comparison

It can be seen from Table 15 that the SR-SB and SR-ST carry out good performance results compared to the SR-AFIB. For example, the average accuracy conducted using 10 fold cross-validation of SR-SB, SR-ST, and SR-AFIB are 97.93%, 98.22%, and 80.59%, respectively, which indicates SR-AFIB has depressed accuracy in contrary to SR-SB and SR-ST. This indication emphasizes that using only static data does not help in improving the classification. Hence, a combination of the static data with the dynamic time-series ECG data was proposed to form a multimodal network, which examines whether adding static data able to provide better results or not as shown in subsequent sections.

Table 15: MLP model subsets comparison

Subset	Accuracy (%)	F1 Score (%)
SR-SB	97.93	96.72
SR-ST	98.22	98.08
SR-AFIB	80.59	82.27

6.5 Experiment using the proposed hybrid 1D-CNN-BiGRU-BiLSTM model

In this section, the proposed hybrid architecture was analyzed using a 1-Dimensional Convolutional Neural Network with Bidirectional Gated Recurrent Unit and Bidirectional Long-short term memory (1D-CNN-Bi-GRU-LSTM) model (shown in Figure 9, Section 4.2) using all the leads combined from the dynamic time-series data. This architecture attained effective results in the classification of heart rhythms. The upcoming subsections discuss the performances of each subset e.g. SR-SB, SR-ST, and SR-AFIB.

6.5.1 SR-SB subset

The experiment for SR-SB subset classification accomplished higher attainment in terms of Accuracy and F1-score as 99.61% and 99.60%, respectively, during the validation stage. Figure 31 shows briefly the validation accuracy and the corresponding iteration loss along the 50 epochs. Moreover, the performance measurements are shown briefly in Table 16 below.

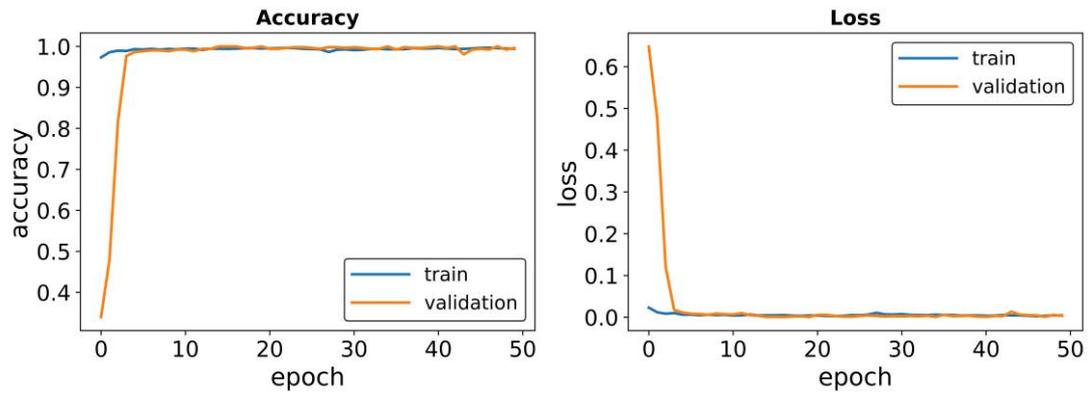


Figure 31: Proposed hybrid SR-SB accuracy (left) and loss (right)

Table 16: Proposed hybrid SR-SB performance measurements

Epochs	Validation	
	Accuracy	F1-score
10	0.9922	0.9862
20	0.9942	0.9806
30	0.9961	0.9893
40	0.9981	0.9960
50	0.9961	0.9960

6.5.2 SR-ST subset

This investigation of SR-ST subset classification achieved higher performance as 99.42% and 99.07% in terms of Accuracy and F1-measure, respectively. The illustration in Figure 32 shows the accuracy and loss function during the model iteration. Table 17 displays the performance evaluation of the SR-ST subset during validation.

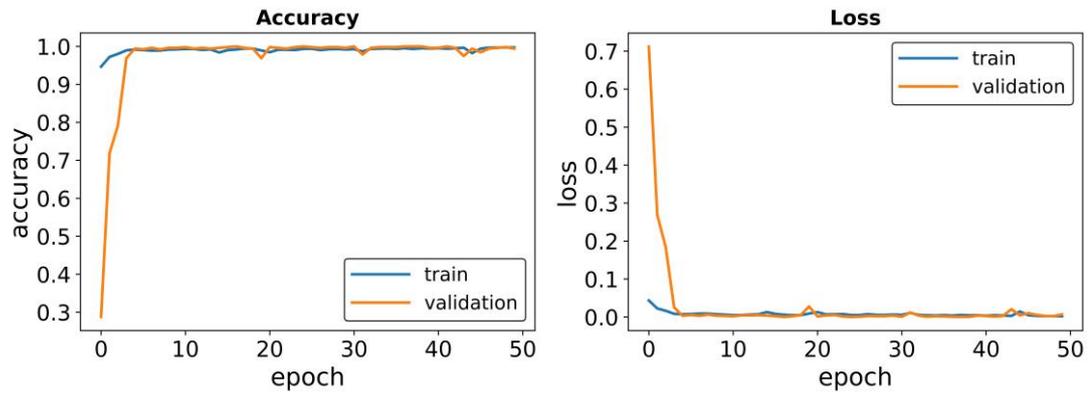


Figure 32: Proposed hybrid SR-ST accuracy (left) and loss (right)

Table 17: Proposed hybrid SR-ST performance measurement

Epochs	Validation	
	Accuracy	F1-score
10	0.9981	0.9941
20	0.9689	0.9531
30	0.9981	0.9946
40	0.9961	0.9946
50	0.9942	0.9907

6.5.3 SR-AFIB subset

The trial of SR-AFIB subset classification obtained superior results in terms of Accuracy and F1-score as 99.02% and 99.14%, respectively. Figure 33 indicates the accuracy with the corresponding loss during validation. Table 18 presents the measurement outcomes.

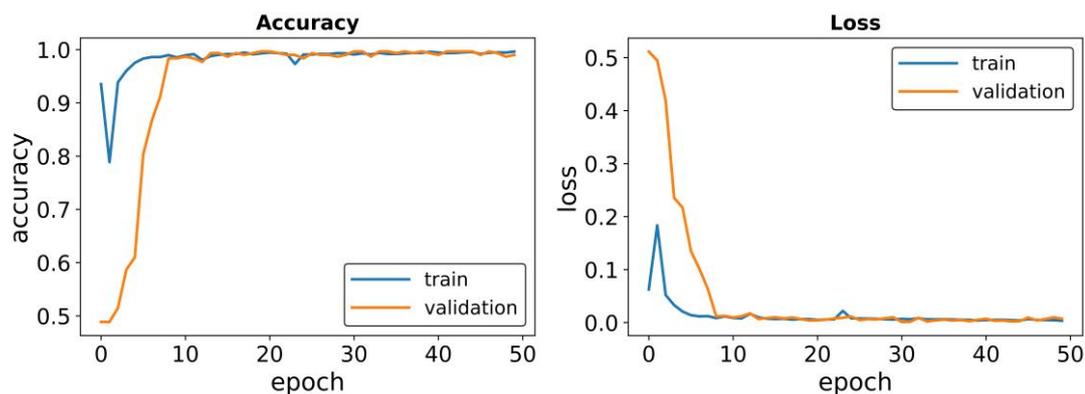


Figure 33: Proposed hybrid SR-AFIB accuracy (left) and loss (right)

Table 18: Proposed hybrid SR-AFIB performance measurement

Epochs	Validation	
	Accuracy	F1-score
10	0.9836	0.9849
20	0.9967	0.9970
30	0.9902	0.9894
40	0.9934	0.9941
50	0.9902	0.9914

6.5.4 Subsets comparison

The observations from Table 19 show that the SR-SB, SR-ST, and SR-AFIB achieved satisfactory outcomes result. For instance, as it is recognized from the Table below that the 10-fold cross-validation average accuracy of SR-SB, SR-ST, and SR-AFIB are 98.18%, 98.30%, and 98.73%, respectively, which shows that all current subsets including SR-AFIB achieved higher accuracy in comparison to previous experiments.

Table 19: Proposed hybrid subsets comparison

Subset	Accuracy (%)	F1 Score (%)
SR-SB	98.18	97.06
SR-ST	98.30	98.15
SR-AFIB	98.73	98.73

Accordingly, in the upcoming section, an experiment to combine the static data with the time-series ECG data was attempted by using the proposed hybrid fused with the MLP to form multimodal DL in order to accomplish an enhanced result. Regardless, adding static data does not enhance the performance compared to the proposed hybrid.

6.6 Experiment using the Multimodal Proposed hybrid combined with MLP

In the previous experiments, it was found that the proposed hybrid produces higher performance in all the subsets, where it achieves more than 98% in terms of Accuracy and F1-measure. This study tries to analyze the impact of the static data on improving the outcomes to answer the third research question of the thesis. Therefore, an experiment was implemented to involve the static data fused with the dynamic time-series data to form a multimodal of (MLP + proposed hybrid DL). Multimodal deep learning has the ability to deal with multi-dimensional datasets of different modalities that have channels of information sources. This illustration of the multimodal architecture was introduced briefly in section 4.4 in the proposed approach chapter. The experiment was repeated for each subset e.g. SR-SB, SR-ST, and SR-AFIB. Each subset was compared separately. Finally, through a comparative table (in the subset comparison subsection), the results have been analyzed for each subset obtained by cross-validation.

6.6.1 SR-SB subset

The observation of the experiment that was conducted for SR-SB subset classification carries out Accuracy and F1-score as 99.13% and 98.07%, respectively. Figure 34 shows the accuracy and loss cost during validation deployment. Table 20 reports the performance evaluation results of the subset of SR-SB.

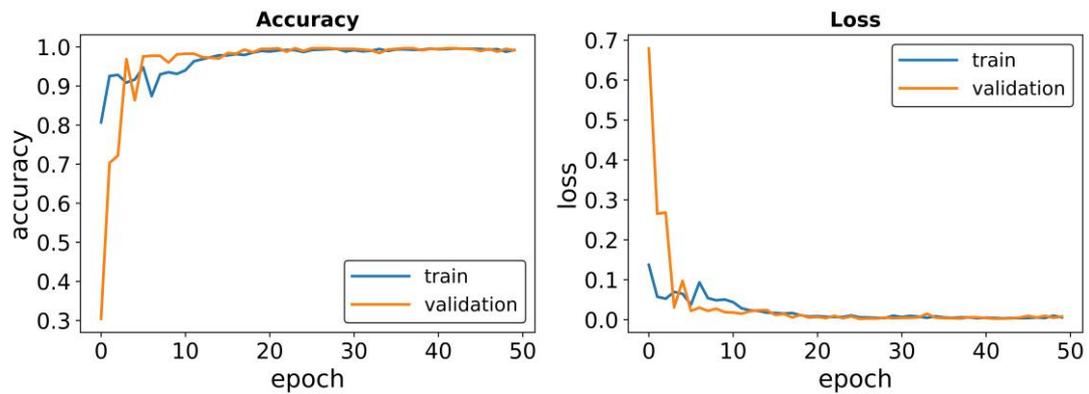


Figure 34: Multimodal SR-SB accuracy (left) and loss (right)

Table 20: Multimodal SR-SB performance measurement

Epochs	Validation	
	Accuracy	F1-score
10	0.9808	0.9656
20	0.9948	0.9922
30	0.9948	0.9891
40	0.9948	0.9859
50	0.9913	0.9807

6.6.2 SR-ST subset

The empirical study for SR-ST subset classification earns 99.12% for Accuracy and 99.26% for F1-score. Figure 35 demonstrates the accuracy and loss across epoch 50. Table 21 determines the validation evaluation results of the subset of SR-SB for varying number of epochs upto 50.

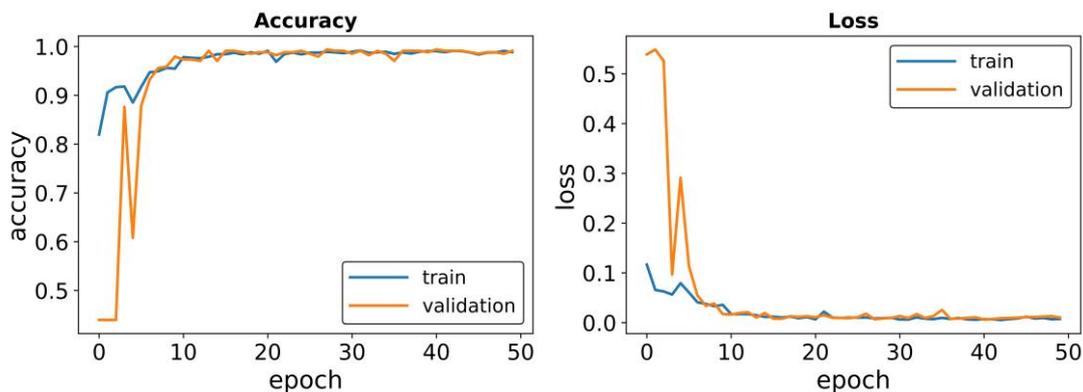


Figure 35: Multimodal SR-ST accuracy (left) and loss (right)

Table 21: Multimodal SR-ST performance measurement

Epochs	Validation	
	Accuracy	F1-score
10	0.9794	0.9803
20	0.9882	0.9890
30	0.9912	0.9926
40	0.9882	0.9890
50	0.9912	0.9926

6.6.3 SR-AFIB subset

This trial study for SR-AFIB classification achieved 99.45% and 99.60% as regards Accuracy and F1-score, respectively. Figure 36 below briefly clarifies the accuracy and loss during training and validation. Table 22 displays the validation evaluation outcomes of the subset of SR-AFIB.

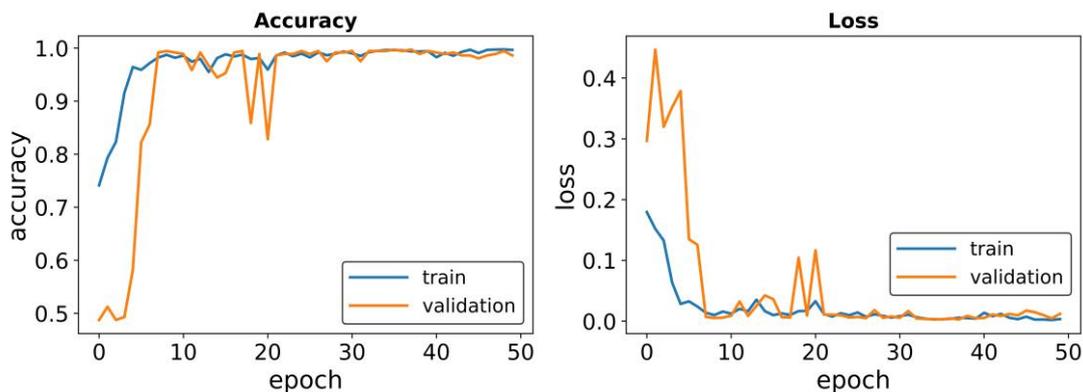


Figure 36: Multimodal SR-AFIB accuracy (left) and loss (right)

Table 22: Multimodal SR-AFIB performance measurement

Epochs	Validation	
	Accuracy	F1-score
10	0.9861	0.9875
20	0.9861	0.9886
30	0.9861	0.9881
40	0.9917	0.9929
50	0.9945	0.9960

6.6.4 Subsets comparison

It could be observed from Table 23 that the SR-SB, SR-ST, and SR-AFIB classification carry out a satisfying performance. For example, the average accuracy conducted using 10 fold cross-validation of SR-SB, SR-ST, and SR-AFIB are 97.69%, 97.73%, and 97.87%, respectively. Nevertheless, this indication confirms that the combination of the static data with the ECG time-series data does not enhance the model improvement of the classification. Since the proposed hybrid DL model performs better in terms of all the performance measurements, where it achieves an average accuracy of more than 98% during the cross-validation iteration in all the subsets.

Table 23: Multimodal (proposed hybrid + MLP) subsets comparison

Subset	Accuracy (%)	F1 Score (%)
SR-SB	97.69	96.47
SR-ST	97.73	97.59
SR-AFIB	97.87	97.88

6.7 Comparison of evaluation results

This section summarizes the performance of all the experiments that were conducted of different models and provides the answer to the three research questions stated in the problem statement of the introduction.

To address the first question (the importance of each lead in a 12-lead ECG in classifying heart rhythm). A comprehensive experiment was executed using the 2D-CNN for the single and combined lead, where the combined lead (12-lead) performed significantly better than single-lead. This fact was emphasized statistically using a T-test. Therefore, all the other experiments were conducted based on the combination of all leads (12-lead) in which it was inserted as an input to the models to investigate the remaining questions.

The second question was addressed (the importance of static data in classifying heart rhythms) by using a Multi-layer Perceptron (MLP) architecture to train on the only static data that are available in the same subset, which contains statistical ECG basic measurements of the time-series data. This provides evidence of the performance of the static data classification. Since the MLP network has the advantage to extract the important features from the static raw data. It has been found that the static data can give low prediction accuracy, however, it was less than the accuracy obtained from

the 2D-CNN that used the combined leads ECG time-series data. For instance, the MLP model subsets for the SR-SB and SR-ST achieved accuracy around 97%, while the SR-AFIB attained approximately around 80%.

In contrast, the proposed hybrid is extremely better than other methods. The classification of the SR-AFIB, SR-SB and SR-ST subsets shows a performance of more than 98% in terms of Accuracy and F1-measure.

The fact of fusing the static data with the proposed hybrid DL model is further investigated. This examination can identify whether the static data enhance the attainment of the model to answer the third thesis question (can clinical static data be combined with ECG time-series data to improve classification). It could be found that it achieved less performance (about 97.87%) compared to the proposed hybrid DL.

Furthermore, different possibilities were also investigated to utilize the static data in order to improve the classification performance by including it into the model using MLP. This was done by fusing primarily the MLP model with 2D-CNN and the proposed hybrid to form a multimodal fusion (MLP + 2D CNN + Proposed hybrid). In addition to including the MLP with 2D-CNN to come up with a multimodal of (2D-CNN+MLP), as shown in the summary for all comparison table. However, all the above-mentioned models were producing lower accuracy when fused with the MLP model. This concludes that adding static data does not help improving classification performance. Table 24 displays the comparison of the other cross-validation accuracy of the mentioned models.

Accordingly, it is emphasized finally that the proposed hybrid deep learning model exhibits superior performance to all the subsets which carried out robust outcomes results compared to all the other methods.

Table 24: Summary for all comparison

Measurements	Model	SR-SB (%)	SR-ST (%)	SR-AFIB (%)
Accuracy	2D-CNN	98.12	97.07	80.33
	MLP	97.93	98.22	80.5
	Multimodal (2D-CNN+MLP)	97.56	97.52	79.44
	Multimodal (Proposed hybrid +MLP)	97.69	97.73	97.87
	Multimodal fusion (proposed hybrid + 2D-CNN+MLP)	97.69	97.57	97.87
	Proposed hybrid	98.18	98.30	98.73
F1-score	2D-CNN	97.05	96.80	81.11
	MLP	96.72	98.08	82.27
	Multimodal (2D-CNN+MLP)	96.19	96.16	80.02
	Multimodal (Proposed hybrid +MLP)	96.47	97.59	97.88
	Multimodal fusion (proposed hybrid + 2D-CNN+MLP)	96.55	97.35	97.83
	Proposed hybrid	97.06	98.15	98.73

Moreover, the subsets of SR-SB, SR-ST, and SR-AFIB were tested with a very recent related work (Yildirim et al., 2020) that used the same 12-lead database and a hybrid deep learning architecture (which is different from the hybrid architecture). Yildirim et al. model code was not available, so it was implemented by two different models. The first model was implemented using the same architecture along with the

parameters and hyperparameters. The second model was adapted by considering hyperparameter optimization and tuning to choose a set of optimal hyperparameters. However, the comparison proves the superiority of the proposed hybrid model because Yildirim et al architecture was mainly customized for a single-lead, therefore, it does not perform well in these experiments. These results are presented in Table 25.

Table 25: Comparison with the Yildirim and his team work architecture

Measurements	Model	SR-SB (%)	SR-ST (%)	SR-AFIB (%)
Accuracy	Yildirim et al., 2020 (original architecture)	96.86	65.73	63.7
	Yildirim et al., 2020 (adapted architecture)	97.72	71.58	63.7
	Proposed hybrid	98.18	98.30	98.73
F1-score	Yildirim et al., 2020 (original architecture)	94.86	57.64	66.23
	Yildirim et al., 2020 (adapted architecture)	96.4	64.89%	61.08
	Proposed hybrid	97.06	98.15	98.73

6.8 Analysis and discussion

Here the results were analyzed and the outcomes are explained based on the research findings: subsection (6.8.1) explains why combined leads are better than single-lead, Subsection (6.8.2) discusses why time-series data are better than static data in classification. In addition to the effect of adding static data to ECG time-series data, and the last Subsection (6.8.3) summarizes why the proposed hybrid model achieves the best result.

6.8.1 Combined lead vs single-lead

Many researchers have attempted heart rhythm classification based on a single-lead. This is because most of those studies have widely used publicly available databases such as MIT-BIH arrhythmia database (Moody & Mark, 2001; Sahoo et al., 2019). This directory contains 47 subjects with 2 leads, studied by the BIH Arrhythmia Laboratory between 1975 and 1979 and were sampled at 360 samples per second. Other researchers conducted their experiments based on the INCART database, which contains records extracted from the Holter monitor. This database records 48 subjects sampled at 257 Hz using 12-lead. Other ECG databases contain different distributions of anomaly which is available to the public as well. Most of those databases contain limitations such as imbalanced classes, either single-lead or more, and low sampling frequencies, thus, the new large database plays a vital role in this study. In this thesis work, a newly published database was used containing more than 10,000 subjects with various rhythms that are higher than the usual sample rate. It was sampled at a rate of 500 Hz and containing a large age group between 4 to 98.

However, the signals that are extracted from ECG devices may contain noises such as line interference, electrode connection noise, motion artifact, and other random noises. Combining all lead data gives a better view of the state of the heart rhythm and reduces the effect of noise rather than using a single-lead only. This is why combined lead performs better than individual lead. It is believed that the outcome of this work about the superiority of combined lead will have an impact on future studies.

6.8.2 Static data combined with time-series data

The static data consists of different statistical measures that are extracted from the ECG time-series data. Therefore, combining static data does not enhance the performance of the Deep Learning model, because it just adds redundancy in the data.

In general, the structure of the DL models that contains many hidden layers able to learn from the time-series data in a way that is superior to hand-crafted statistical features generated from the time-series data. This special structure allows to learn powerful representations and extract features automatically from the ECG time-series training data.

6.8.3 Proposed hybrid model performance

The hybrid approach performs well because of the advantage of combining the 1D-Convolutional Neural Network with the Bidirectional Gated Recurrent Unit (BiGRU) and Bidirectional Long-short-term memory (BiLSTM). The 1D-CNN is capable of extracting out significant items and learning complex features from the data in a way similar to image learning. Bidirectional Long short-term memory (BiLSTM) is an extension of traditional LSTM which consists of memory blocks that have proven to be very useful in learning from temporal data. Therefore, a combination of the 1D-CNN with BiGRU and BiLSTM ensures better learning compared to the other approaches such as 2D-CNN.

Chapter 7: Conclusion and Future Work

A DL model is proposed to diagnose heart rhythm anomalies from 12-lead ECG data. This study is motivated by the fact that manual diagnosis of heart rhythms from ECG signals by experts may be erroneous, and therefore, an automated heart rhythm diagnosis approach will help caregivers to make more informed decisions with less chance of misdiagnosis. A Large database of 12-lead ECG data consisting of more than 10 thousand subjects has been collected and three research challenges were identified to accomplish the goal of developing the automated diagnosis model. The first challenge was to understand the role of each lead of the 12-lead ECG in classifying heart rhythms. This challenge was addressed by proposing a CNN model to evaluate the efficacy of individual lead and the combined lead, concluded with the help of statistical tests that the combined lead data are much more effective than the single-lead. The second challenge was to understand the effectiveness of static data that was part of the database. To address this, an MLP model was proposed to evaluate the performance of the static data. Finally, the third challenge of understanding the effectiveness of fusing the static data with the ECG time series data for heart rhythm classification was addressed by proposing different multimodal DL models and evaluating the combined data. Thereby, it can be concluded that the static data does not help in improving the classification performance.

Furthermore, an effective hybrid DL model (1D-CNN BiGRU-BiLSTM) have been proposed to classify heart rhythm and showed its effectiveness over other models.

In the future, utilizing other databases would be considered as well as utilizing demographic data (e.g. age, gender) and clinical background in improving the classification performance using multi-class classification and attention-based models.

Existing deep learning architectures such as transfer learning can be investigated, which may have great impact by applying the fine-tuning to the model after retraining the network weights using the new database. Also, considering the diagnosis of other types of heart disease such as Myocardial Infarction (MI) and use other types of data like image. Additionally, classifying the heartbeats based on the recommendations of the Association for the Advancement of Medical Instrumentation (AAMI) standard that categorized heartbeats into five classes using the heartbeats features available in the same database. Finally, collaborating with a local medical facility to collect patient data and develop a full-fledged application that would be used in the field.

References

- Abiodun, O., Jantan, A., Omolara, A., Dada, K., Mohamed, N., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938. <https://doi.org/10.1016/j.heliyon.2018.e00938>.
- Acharya, U., Fujita, H., Oh, S., Hagiwara, Y., Tan, J., & Adam, M. (2017a). Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. *Information Sciences*, 415. <https://doi.org/10.1016/j.ins.2017.06.027>.
- Acharya, U., Fujita, H., Oh, S., Hagiwara, Y., Tan, J., & Adam, M. (2017b). Automated Detection of Arrhythmias Using Different Intervals of Tachycardia ECG Segments with Convolutional Neural Network. *Information Sciences*, 405. <https://doi.org/10.1016/j.ins.2017.04.012>.
- Acharya, U., Oh, S., Hagiwara, Y., Tan, J., Adam, M., Gertych, A., & Tan, San. (2017c). A Deep Convolutional Neural Network Model to Classify Heartbeats. *Computers in Biology and Medicine*, 89. <https://doi.org/10.1016/j.combiomed.2017.08.022>.
- Al Rahhal, M., Bazi, Y., Almubarak, H., Alajlan, N., & Al Zuair, M. (2019). Dense Convolutional Networks With Focal Loss and Image Generation for Electrocardiogram Classification. *IEEE Access*, 7, 182225–182237. <https://doi.org/10.1109/ACCESS.2019.2960116>.
- Anwar, S. M., Gul, M., Majid, M., & Alnowami, M. (2018). Arrhythmia Classification of ECG Signals Using Hybrid Features. *Computational and Mathematical Methods in Medicine*, 2018, 1–8. <https://doi.org/10.1155/2018/1380348>.
- Bashar, S., Han, D., Soni, A., Mcmanus, D., & Chon, K. (2018). Developing a novel noise artifact detection algorithm for smartphone PPG signals: Preliminary results, 79-82. <https://doi.org/10.1109/BHI.2018.8333374>.
- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. <http://arxiv.org/abs/1206.5533>.
- Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. (2018). Retrieved April 25, 2021, from <https://www.ahajournals.org/doi/10.1161/CIR.0000000000000558>
- Cleveland Clinic Abu Dhabi (2019). Focuses on Prevention- Urgent need to Tackle Root Causes of Heart Disease: Obesity, Lack of Exercise, Smoking and Stress. Retrieved 10 April 2021, from <https://www.clevelandclinicabudhabi.ae/en/media-center/news/pages/survey-reveals-high-number-of-heart-risk-factors-across-uae-population.aspx>
- Devi, R., Tyagi, H., & Kumar, D. (2019). Performance Comparison and Applications of Sparsity Based Techniques for Denoising of ECG Signal. 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 346–351. <https://doi.org/10.1109/SPIN.2019.8711632>.

- Ebrahimi, Z., & Loni, M., Daneshtalab, M., & Gharehbaghi, A. (2020). A Review on Deep Learning Methods for ECG Arrhythmia Classification. *Expert Systems with Applications: X*, 7, 100033. <https://doi.org/10.1016/j.eswax.2020.100033>.
- European Cardiovascular Disease Statistics 2017. (2017). Retrieved April 25, 2021, from <https://ehnheart.org/cvd-statistics/cvd-statistics-2017.html>
- Ganguly, B., Ghosal, A., Das, A., Das, D., Chatterjee, D., & Rakshit, D. (2020). Automated Detection and Classification of Arrhythmia From ECG Signals Using Feature-Induced Long Short-Term Memory Network. *IEEE Sensors Letters*, 4(8), 1–4. <https://doi.org/10.1109/LSSENS.2020.3006756>.
- Gao, J., Zhang, H., Lu, P., & Wang, Z. (2019). An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset. *Journal of Healthcare Engineering*, 1–10. <https://doi.org/10.1155/2019/6320651>.
- Gawande, N., & Barhatte, A. (2017). Heart diseases classification using convolutional neural network. 2017 2nd International Conference on Communication and Electronics Systems, 17–20. <https://doi.org/10.1109/CESYS.2017.8321264>.
- Goldberger, A., Goldberger, Z., & Shvilkin, A. (2013). *Goldberger's Clinical Electrocardiography* (Eighth Edition). 233. <https://doi.org/10.1016/B978-0-323-08786-5.16001-4>.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., & Wang, G. (2015). Recent Advances in Convolutional Neural Networks. *Pattern Recognition*, 77. <https://doi.org/10.1016/j.patcog.2017.10.013>.
- Hammad, M., Zhang, S., & Wang, K. (2019). A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication. *Future Generation Computer Systems*, 101, 180–196. <https://doi.org/10.1016/j.future.2019.06.008>.
- Hannun, A., Rajpurkar, P., Haghpanahi, M., Tison, G., Bourn, C., Turakhia, Mi., & Ng, A. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25. <https://doi.org/10.1038/s41591-018-0268-3>.
- He, R., Wang, K., Zhao, N., Liu, Y., Yuan, Y., Li Q., & Zhang, H. (2018). Automatic Detection of Atrial Fibrillation Based on Continuous Wavelet Transform and 2D Convolutional Neural Networks. *Frontiers in physiology*, 9, 1206. <https://doi.org/10.3389/fphys.2018.01206>.
- Hochreiter, S. & Schmidhuber, J. (1997). Long Short-term Memory. *Neural computation*, 9, 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hong, S., Zhou, Y., Shang, J., Xiao, C., & Sun, J. (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine*, 122, 103801. <https://doi.org/10.1016/j.combiomed.2020.103801>.

- Hsieh, C., Li, Y., Hwang, B., & Hsiao, C. (2020). Detection of Atrial Fibrillation Using 1D Convolutional Neural Network. *Sensors*, 20(7), 2136. <https://doi.org/10.3390/s20072136>.
- Izci, E., Ozdemir, M., Degirmenci, M., & Akan, A. (2019). Cardiac Arrhythmia Detection from 2D ECG Images by Using Deep Learning Technique. 2019 Medical Technologies Congress (TIPTEKNO), 1–4. <https://doi.org/10.1109/TIPTEKNO.2019.8895011>.
- Jun, T., Nguyen, H., Kang, D., Kim, D., Kim, Y., & Kim, D. (2018). ECG arrhythmia classification using a 2-D convolutional neural network. <https://arxiv.org/abs/1804.06812>.
- Jun, T., Park, H., Minh, N., Kim, D., & Kim, Y. (2016). Premature Ventricular Contraction Beat Detection with Deep Neural Networks. 2016 15th IEEE International Conference on Machine Learning and Applications, 859–864. <https://doi.org/10.1109/ICMLA.2016.0154>.
- Kanani, P., & Padole, M. (2020). ECG Heartbeat Arrhythmia Classification Using Time-Series Augmented Signals and Deep Learning Approach. *Procedia Computer Science*, 171. <https://doi.org/10.1016/j.procs.2020.04.056>.
- Keshavamurthy, T. G., & Eshwarappa, M. (2017). Review paper on denoising of ECG signal. 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1–4. <https://doi.org/10.1109/ICECCT.2017.8117941>.
- Khan, A. H., Hussain, M., & Malik, M. K. (2021). Arrhythmia Classification Techniques Using Deep Neural Network. *Complexity*, 2021, 1–10. <https://doi.org/10.1155/2021/9919588>.
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. (2020). A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artificial Intelligence Review*, 53(8), 5455–5516. <https://doi.org/10.1007/s10462-020-09825-6>.
- Khane, R., Surdi, A., & Bhatkar, R. (2011). Changes in ECG pattern with advancing age. *Journal of Basic and Clinical Physiology and Pharmacology*, 22(4). <https://doi.org/10.1515/JBCPP.2011.017>.
- Kim, J., Kang, M., & Hwang, B. (2011). A method for detecting arrhythmia using a RR interval from ECG data in U-Health system. Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, 1–5. <https://doi.org/10.1145/1968613.1968632>.
- Kiranyaz, S., Ince, T., & Gabbouj, M. (2015). Real-Time Patient-Specific ECG Classification by 1D Convolutional Neural Networks. *IEEE transactions on biomedical engineering*, 63. <https://doi.org/10.1109/TBME.2015.2468589>.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep Learning. *Nature*, 521, 436–44. <https://doi.org/10.1038/nature14539>.

- Li, Z., Zhou, D., Wan, L., Li, J., & Mou, W. (2020). Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram. *Journal of Electrocardiology*, 58. <https://doi.org/10.1016/j.jelectrocard.2019.11.046>.
- Lian, J., Wang, L., & Muessig, D. (2011). A Simple Method to Detect Atrial Fibrillation Using RR Intervals. *The American Journal of Cardiology*, 107(10), 1494–1497. <https://doi.org/10.1016/j.amjcard.2011.01.028>.
- Lih, O., Jahmunah, V., San, T., Ciaccio, E., Yamakawa, T., Tanabe, M., Kobayashi, M., Faust, O., & Acharya. (2020). Comprehensive electrocardiographic diagnosis based on deep learning. *Artificial intelligence in medicine*. 103. <https://doi.org/10.1016/j.artmed.2019.101789>.
- Liu, M., & Kim, Y. (2018). Classification of Heart Diseases Based On ECG Signals Using Long Short-Term Memory. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2707–2710. <https://doi.org/10.1109/EMBC.2018.8512761>.
- Luo, J., Fu, C., Bai, M., & Zhao, Y. (2018). Atrial Fibrillation Detection with Convolutional Neural Network. *CSAI '18: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 94–98. <https://doi.org/10.1145/3297156.3297169>.
- Luz, E., Schwartz, W., Chávez, G., & Menotti, D. (2016). ECG-based Heartbeat Classification for Arrhythmia Detection: A Survey. *Computer Methods and Programs in Biomedicine*, 127. <https://doi.org/10.1016/j.cmpb.2015.12.008>.
- Ma, F., Zhang, J., Chen, W., Liang, W., & Yang, W. (2020). An Automatic System for Atrial Fibrillation by Using a CNN-LSTM Model. *Discrete Dynamics in Nature and Society*, 1–9. <https://doi.org/10.1155/2020/3198783>.
- Mathews, S., & Kambhamettu, C., & Barner, K. (2018). A novel application of deep learning for single-lead ECG classification. *Computers in Biology and Medicine*. 99. <https://doi.org/10.1016/j.combiomed.2018.05.013>.
- Matias, I., Garcia, N., Pirbhulal, S., Felizarco, V., Pombo, N., Zacarias, H., Sousa, M., & Zdravevski, E. (2021). Prediction of Atrial Fibrillation using artificial intelligence on Electrocardiograms: A systematic review. *Computer Science Review*, 39. <https://doi.org/10.1016/j.cosrev.2020.100334>.
- Moody, G., & Mark, R. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society*, 20, 45-50. <https://doi.org/10.1109/51.932724>.
- Murat, F., Yildirim, O., Talo, M., Baloglu, U., Demir, Y., & Acharya, U. (2020). Application of deep learning techniques for heartbeats detection using ECG signals- Analysis and Review. *Computers in Biology and Medicine*, 120. <https://doi.org/10.1016/j.combiomed.2020.103726>.

- Nurmaini, S., Tondas, A., Darmawahyuni, A., Rachmatullah, M. & Partan, R., Firdaus, F., Tutuko, B. Pratiwi, F., Juliano, A., & Khoirani, R. (2020). Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. *Future Generation Computer Systems*, 113. <https://doi.org/10.1016/j.future.2020.07.021>.
- Oh, S., Ng, E., Tan, R., & Acharya, U. (2018a). Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Computers in Biology and Medicine*, 105. <https://doi.org/10.1016/j.combiomed.2018.12.012>.
- Oh, S., Ng, E., Tan, R., & Acharya, U. (2018b). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Computers in Biology and Medicine*, 102. <https://doi.org/10.1016/j.combiomed.2018.06.002>.
- Rafi, S. M., & Akthar, S. (2021). ECG Classification using a Hybrid Deep learning Approach. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 302–305. <https://doi.org/10.1109/ICAIS50930.2021.9395897>.
- Romdhane, T., Alhichri, H., Ouni, R. & Atri, M. (2020). Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. *Computers in Biology and Medicine*, 123, 103866. <https://doi.org/10.1016/j.combiomed.2020.103866>.
- Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd.ed.).
- Sahoo, S., Dash, M., & Sabut, S. (2019). Machine Learning Approach to Detect Cardiac Arrhythmias in ECG Signals: A Survey. *IRBM*, 41(4), 185–194 <https://doi.org/10.1016/j.irbm.2019.12.001>.
- Sampson, M. & Mcgrath, A. (2015). Understanding the ECG. Part 1: Anatomy and physiology. *British Journal of Cardiac Nursing*, 10, 548-554. <https://doi.org/10.12968/bjca.2015.10.11.548>.
- Sannino, G., & De Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Generation Computer Systems*. 86. <https://doi.org/10.1016/j.future.2018.03.057>.
- Schwob, M. R., Dempsey, A., Zhan, F., Zhan, J., & Mehmood, A. (2020). Robust Multimodal Heartbeat Detection Using Hybrid Neural Networks. *IEEE Access*, 8, 82201–82214. <https://doi.org/10.1109/ACCESS.2020.2990607>.
- Nearly half of US adults have cardiovascular disease, study says. Retrieved April 20, 2021, from <https://edition.cnn.com/2019/01/31/health/heart-disease-statistics-report/index.html>
- Sellami, A. & Hwang, H. (2019). A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Systems with Applications*. 122, 75–84. <https://doi.org/10.1016/j.eswa.2018.12.037>.

- Serhani, M., El Kassabi, H., Ismail, H., & Nujum, R. (2020a). ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges. *Sensors*, 20(6), 1796. <https://doi.org/10.3390/s20061796>.
- Serhani, M., Nujum, R., Ashwal, H., & Qirim, N. (2020b). ECG-based Arrhythmia Classification & Clinical Suggestions: An Incremental Approach of Hyperparameter Tuning. *Proceedings of the 13th International Conference on Intelligent Systems*, 1–7. <https://doi.org/10.1145/3419604.3419787>.
- Shen, Y., Voisin, M., Aliamiri, A., Avati, A., Hannun, A., Ng, A. (2019). Ambulatory Atrial Fibrillation Monitoring Using Wearable Photoplethysmography with Deep Learning. 1909–1916. <https://doi.org/10.1145/3292500.3330657>.
- Singh, S., Pandey, S., Pawar, U., & Janghel, R. (2018). Classification of ECG Arrhythmia using Recurrent Neural Networks. *Procedia Computer Science*, 132, 1290–1297. <https://doi.org/10.1016/j.procs.2018.05.045>.
- Tan, J., Hagiwara, Y., Pang, W., Lim, I., Oh, S., Adam, M., Tan, R., Chen, M., & Acharya, U. (2018). Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Computers in Biology and Medicine*, 94. <https://doi.org/10.1016/j.combiomed.2017.12.023>.
- Wang, J., & Li, W. (2020). Atrial Fibrillation Detection and ECG Classification based on CNN-BiLSTM. <http://arxiv.org/abs/2011.06187>.
- Institute for Quality and Efficiency in Health Care. (2019). What is an electrocardiogram (ECG)?. (2019). Retrieved April 21, 2021, from <https://www.ncbi.nlm.nih.gov/books/NBK536878/>
- World Health Organization. Cardiovascular diseases (CVDs). (2017). Retrieved April 25, 2021, from [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Xiong, Z., Nash, M., Cheng, E., Fedorov, V., Stiles, M., & Zhao, J. (2018). ECG Signal Classification for the Detection of Cardiac Arrhythmias Using a Convolutional Recurrent Neural Network. *Physiological Measurement*, 39(9). <https://doi.org/10.1088/1361-6579/aad9ed>.
- Xu, S., Mak, M., & Cheung, C. (2019). Towards End-to-End ECG Classification With Raw Signal Extraction and Deep Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, PP, 1–1. <https://doi.org/10.1109/JBHI.2018.2871510>.
- Yildirim, O., Talo, M., Ciaccio, E., Ru San Tan, & Acharya, U. (2020). Accurate Deep Neural Network Model to Detect Cardiac Arrhythmia on More Than 10,000 Individual Subject ECG Records. *Computer Methods and Programs in Biomedicine*. 197. <https://doi.org/10.1016/j.cmpb.2020.105740>.
- Zheng, J., Zhang, J., Danioko, S., Hai, Y., Hangyuan, G., & Rakovski, C. (2020). A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Sci Data*, 7, 48. <https://doi.org/10.1038/s41597-020-0386-x>.