Data analytics to augment decision making in cardiac care

Faculty of Computing, Engineering and the Built Environment

Khaled Rjoob

This thesis is submitted for the degree of Doctor of Philosophy

Supervisors:

Prof. Raymond Bond

Dr. Victoria McGilligan

Prof. Aaron Peace

October 2021

I would like to dedicate this thesis to my beloved parents, my family and to my lovely fiancée.

Acknowledgement

I have to express my sincere thanks to my supervisors Prof. Raymond Bond, Dr Victoria McGilligan, Prof. Aaron Peace and Prof. Stephen Leslie for their perfect supervision, remarkable cooperation, continuous support and creative ideas during my PhD. Additionally, I would like to thank Centre for Personalised Medicine (CPM) team, especially, Prof. Tony Bjourson, Dr Donna Tedstone (CPM project manager) and Christine Stewart (administrative co-Ordinator) for their guidance and support during the PhD. I would like to thank Patricia Moran (C-TRIC facilities manager) for her help during the PhD. Also, I would like to thank my research cluster members Dr Matthew Manktelow, Aleeha Iftikhar, Charles Knoery, Anne McShane. I would like to thank Prof. Dewar Finlay for his help and support through the PhD. I thank the Doctoral College team, especially Kate McMorris for her support during my PhD. I thank all my colleagues at Ulster University. And I would like to thank my friends who I met during the PhD Ali Rababah and Motasem Bani Mustafa for their help and kindness. I acknowledge all the staff and my colleagues in the school of computing at Ulster University. Lastly, I acknowledge the INTERREG VA Programme for providing the funding needed to accomplish this work. This project is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

Abstract

Acute coronary syndrome (ACS) causes the majority of deaths in ischaemic heart disease. Hence, ACS needs to be rapidly and correctly diagnosed. The electrocardiogram (ECG) is a non-invasive method that is used to diagnose ACSs alongside cardiac biomarkers such as troponin. However, ACS diagnostics when using the ECG and blood biomarkers have several challenges such as poor ECG data quality due to electrode misplacement or noise. Moreover, there are limitations with the use and accuracy of cardiac biomarkers since they typically require a waiting time to attain the result. The aim of this PhD is to use data science to augment clinical decision-making by improving, 1) ECG data quality and 2) discovering novel biomarkers to improve the sensitivity and specificity of detecting acute myocardial infarctions, specifically ST-elevation myocardial infarction (STEMI). Datasets including blood biomarkers and ECGs have been collected from Altnagelvin hospital. In the first part of this PhD, several traditional machine learning (ML) and deep learning (DL) algorithms have been used to detect electrode misplacement in ECG data. The ML algorithms achieved high performance rates (accuracy=93% and 0.9 area under the curve). In the second part of this PhD, the impact of signal noise on ECG interpretation was investigated, and it transpired that noise has a significant impact ($P<0.001$ McNemar's) on ECG classification when using ML. In the third part of the PhD, ML has been used to discover novel blood biomarkers that could be used to discriminate between STEMI, Non-STEMI and unstable angina. In this part of the PhD, we investigated whether combining those detected blood biomarkers with ECG features can improve STEMI detection. According to the results from each part in this PhD, the use of ML demonstrated promising results when applied to these cardiology datasets. This PhD could be used to inform the development of a new medical instrument or medical laboratory kits that can provide timely and accurate diagnoses.

Abbreviations

ACS	Acute Coronary Syndrome
AHA	American heart association
AI	Artificial Intelligence
AR	Arrhythmias
AUC	Area Under the Curve
AV	Atrioventricular
B-LSTM	bidirectional long short-term memory
BSPM	Body Surface Potential Map
CHD	Coronary Heart Diseases
CNNs	Convolutional Neural Networks
CVD	Cardiovascular Disease
DL	Deep Learning
DWT	Discrete Wavelet Transform
DNNs	deep neural networks
DT	decision Tree
ECG	Electrocardiogram
H-FABP	Heart Fatty Acid Binding Protein
ICS	Intercostal Space
JMI	Joint Mutual Information
LA	Left Arm
LASSO	Least Absolute Shrinkage and Selection Operator
LIME	Local interpretable model-agnostic explanations
LL	Left Leg
LOGT	Logistic Regression
LR	Low-Risk
MI	Myocardial Infarction
MIFS	Mutual Information Feature Selection
ML	Machine Learning
MRMR	Maximum Relevance Minimum Redundancy
NSTEMI	Non-ST-Elevation Myocardial Infarction
PCA	Principal Component Analysis
PPCI	Primary Percutaneous Coronary Intervention
PVC	Premature Ventricular Contraction
RA	Right Arm
RL	Right Leg
RMSE	Root Mean Square Error
SA	sinoatrial
SFM	Select From Model
SFS	Sequential Forward Selection
STEMI	ST-Elevation Myocardial Infarction
SVM	Support Vector Machine
UA	Unstable Angina
VHR	Very High-Risk
WHO	World Health Organisation

Publications

Journal publications:

1. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Rababah, A, Iftikhar, A, Guldenring, D, Knoery, C, McShane, A & Peace, A 2021, 'Reliable Deep Learning–Based Detection of Misplaced Chest Electrodes During Electrocardiogram Recording: Algorithm Development and Validation', JMIR Medical Informatics, vol. 9, no. 4, e25347, pp. 1-10. https://doi.org/10.2196/25347

2. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Rababah, A, Guldenring, D, Iftikhar, A, Knoery, C, McShane, A & Peace, A 2020, 'Machine learning techniques for detecting electrode misplacement and interchanges when recording ECGs: A systematic review and meta-analysis ', Journal of Electrocardiology, vol. 62, pp. 116-123. https://doi.org/10.1016/j.jelectrocard.2020.08.013

3. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Iftikhar, A, Guldenring, D, Rababah, A, Knoery, C, McShane, A & Peace, A 2019, 'Data driven feature selection and machine learning to detect misplaced V1 and V2 chest electrodes when recording the 12-lead electrocardiogram', Journal of Electrocardiology, vol. 57, pp. 39-43. https://doi.org/10.1016/j.jelectrocard.2019.08.017

4. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Iftikhar, A, Guldenring, D, Knoery, C & Peace, A 2019, 'Feature Engineering and Machine Learning for the Auto-Detection of Misplaced V1 and V2 Chest Electrodes when Recording 12-lead ECGs', Journal of Electrocardiology, pp. S106. https://doi.org/10.1016/j.jelectrocard.2019.08.037

Conference publications:

5. Rjoob, K, McGilligan, VE, Bond, RR, Watterson, S, El Chemaly, M, Mc Allister, R, De Melo Malaquias, T, Leslie, SJ, Knoery, C, Iftikhar, A, McShane, A, Bjourson, AJ & Peace, A 2021, Improving the Detection of Acute Coronary Syndrome Using Machine Learning of Blood Biomarkers. in 2020 Computing in Cardiology, CinC 2020., 9344422, 2020 COMPUTING IN CARDIOLOGY, IEEE, Rimini, Italy, Computing in Cardiology 2020, Rimini, Italy, 13/09/20. https://doi.org/10.22489/CinC.2020.337

6. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Guldenring, D, Rababah, A, Iftikhar, A, Knoery, C, McShane, A & Peace, A 2021, '"What was AI thinking?": Explainable deep learning in reading of 12-lead ECGs for detecting V1 and V2 electrode misplacement', 45th International Society for Computerized Electrocardiology , 28/04/21 - 1/05/21.

7. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, SJ, Rababah, A, Iftikhar, A, Guldenring, D, Knoery, C, McShane, A & Peace, A 2020, Towards explainable artificial intelligence and explanation user interfaces to open the 'black box' of automated ECG interpretation. in T Reis, MX Bornschlegl, M Angelini & ML Hemmje (eds), Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications: AVI 2020 Workshops, AVI-BDA and ITAVIS, Ischia, Italy, June 9, 2020 and September 29, 2020, Revised Selected Papers. 1 edn, vol. 12585, Information Systems and Applications, incl. Internet/Web, and HCI, vol. 12585, Springer, Workshop on Road Mapping Infrastructures for Artificial Intelligence Supporting Advanced Visual Big Data Analysis, 9/06/21.

8. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, S, Iftikhar, A, Gueldenring, D, Rababah, A, Knoery, C & Peace, A 2019, 'Machine Learning Improves the Detection of Misplaced V1 and V2 Electrodes During 12-Lead Electrocardiogram Acquisition', Paper presented at Computing in Cardiology, Biopolis, Singapore, 8/09/19 - 11/09/19.

9. Rjoob, K, Bond, RR, Finlay, D, McGilligan, VE, Leslie, S, Iftikhar, A, Gueldenring, D, Knoery, C, Rababah, A & Peace, A 2019, 'Performance of a computer model to detect misplaced V1 and V2 electrodes on the 12-lead ECG for three dfifferent types of patients', Heart, vol. 105, no. 7, pp. A31-A32. https://doi.org/10.1136/heartjnl-2019-ICS.39

Other publications:

10. Iftikhar, A, Bond, RR, McGilligan, VE, Leslie, SJ, Knoery, C, McShane, A, Quigg, C, Campbell, R, Boyd, K, Rjoob, K & Peace, A 2021, 'Digital Forms in Healthcare: Comparing the Usability of Single-Page, Multi-Page and Conversational Forms', JMIR Human Factors, vol. 8. https://doi.org/10.2196/25787

11. Iftikhar, A, Bond, RR, McGilligan, VE, Leslie, S, Knoery, C, Shand, JA, Ramsewak, A, Sharma, D, Canning, A, McShane, A, Rjoob, K & Peace, A 2021, 'Human–Computer Agreement of Electrocardiogram Interpretation for Patients Referred to and Declined for Primary Percutaneous Coronary Intervention: Retrospective Data Analysis Study', JMIR Medical Informatics, vol. 9, no. 3, e24188, pp. 1-11. https://doi.org/10.2196/24188

12. Iftikhar, A, Bond, RR, McGilligan, VE, Leslie, SJ, McShane, A, Knoery, C, Rjoob, K & Peace, A 2020, 'Computational Time Series Analysis of Patient Referrals to a Primary Percutaneous Coronary Intervention Service', Health Informatics Journal, vol. 26, no. 3, pp. 2222-2236. https://doi.org/10.1177/1460458219899762

13. Iftikhar, A, Bond, RR, Rjoob, K, Knoery, C, Leslie, SJ, McShane, A, McGilligan, VE & Peace, A 2021, Machine Learning to Predict 30 Days and 1-Year Mortality

in STEMI and Turndown Patients. in 2020 Computing in Cardiology, CinC 2020., 9344181, 2020 COMPUTING IN CARDIOLOGY, IEEE, Rimini, Italy, Computing in Cardiology 2020, Rimini, Italy, 13/09/20. https://doi.org/10.22489/CinC.2020.389

14. Iftikhar, A, Bond, RR, McGilligan, VE, Rjoob, K, Knoery, C, Leslie, S, McShane, A & Peace, A 2020, Predicting 30 days Mortality in STEMI Patients using Patient Referral Data to a Primary Percutaneous Coronary Intervention Service. in I Yoo, J Bi & XT Hu (eds), Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019., 8983143, IEEE Xplore, San Diego, CA, USA, pp. 1315-1317, 2019 IEEE International Conference on Bioinformatics and Biomedicine, 18/11/19. https://doi.org/10.1109/BIBM47256.2019.8983143

15. Iftikhar, A, Bond, RR, McGilligan, VE, Rjoob, K, Leslie, S, Knoery, C, McShane, A & Peace, A 2020, Unsupervised Machine Learning Elicits Patient Archetypes in a Primary Percutaneous Coronary Intervention Service. in I Yoo, J Bi & XT Hu (eds), Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019., 8983318, IEEE Xplore, San Diego, CA, USA, pp. 1309-1314, 2019 IEEE International Conference on Bioinformatics and Biomedicine, 18/11/19. https://doi.org/10.1109/BIBM47256.2019.8983318

16. Iftikhar, A, Bond, RR, McGilligan, VE, McShane, A, Leslie, S, Rjoob, K, Knoery, C & Peace, A 2019, Role of dashboards in improving decision making in healthcare: Review of the literature. in ECCE 2019 Proceedings of the 31st European Conference on Cognitive Ergonomics: "Design for Cognition". ECCE 2019 - Proceedings of the 31st European Conference on Cognitive Ergonomics: "Design for Cognition", Association for Computing Machinery, pp. 215-219, 31st European Conference on Cognitive Ergonomics, Belfast, United Kingdom, 10/09/19. https://doi.org/10.1145/3335082.3335109

17. Knoery, C, McEwan, K, Manktelow, M, Watt, J, smith, J, Iftikhar, A, Rjoob, K, Bond, RR, McGilligan, VE, Peace, A, Heaton, J & Leslie, S 2021, 'Identification of the characteristics of occlusive myocardial infarction: are there any tell-tale signs?', European Heart Journal, vol. 42, no. Supplement_1, ehab724.1160. https://doi.org/10.1093/eurheartj/ehab724.1160

18. Knoery, C, Heaton, J, Polson, R, Bond, RR, Iftikhar, A, Rjoob, K, McGilligan, VE, Peace, A & Leslie, S 2020, 'Systematic review of clinical decision support systems for pre-hospital acute coronary syndrome identification', Critical Pathways in Cardiology, vol. 19, no. 3, pp. 119-125. https://doi.org/10.1097/HPC.0000000000000217

19. Knoery, C, Bond, RR, Iftikhar, A, Rjoob, K, McGilligan, VE, Peace, A, Heaton, J & Leslie, S 2019, 'SPICED-ACS: Study of the Potential Impact of a Computer-generated ECG Diagnostic Algorithmic Certainty Index in STEMI diagnosis:

towards transparent AI', Journal of Electrocardiology, vol. 57, pp. S86-S91. https://doi.org/10.1016/j.jelectrocard.2019.08.006

20. Guldenring, D, Rababah, A, Finlay, D, Bond, RR, Kennedy, A, Jennings, M, Rjoob, K & McLaughlin, J 2021, Estimating the Minimal Size of Training Datasets Required for the Development of Linear ECG-Lead Transformations. in Estimating the Minimal Size of Training Datasets Required for the Development of Linear ECG-Lead Transformations. Computing in Cardiology 2021, Brno, Czech Republic, 12/09/21.

21. Güldenring, D, Rababah, A, Finlay, D, Bond, RR, Kennedy, A, Jennings, M, Rjoob, K & McLaughlin, J 2021, Regression or Pseudo-Inverse - Which Method Should be Preferred When Developing Inverse Linear ECG-Lead Transformations? in 2020 Computing in Cardiology., 9344390, 2020 COMPUTING IN CARDIOLOGY, IEEE Xplore, Rimini, Italy, Computing in Cardiology 2020, Rimini, Italy, 13/09/20. https://doi.org/10.22489/CinC.2020.330

22. Rababah, AS, Bear, LR, Dogrusoz, YS, Good, W, Bergquist, J, Stoks, J, MacLeod, R, Rjoob, K, Jennings, M, Mclaughlin, J & Finlay, DD 2021, 'The effect of interpolating low amplitude leads on the inverse reconstruction of cardiac electrical activity', Computers in biology and medicine, vol. 136, 104666. https://doi.org/10.1016/j.compbiomed.2021.104666

23. Rababah, A, Finlay, D, Bear, L, Bond, RR, Rjoob, K & McLaughlin, J 2019, 'Interpolating Low Amplitude ECG Signals Combined with Filtering According to International Standards Improves Inverse Reconstruction of Cardiac Electrical Activity', Lecture Notes in Computer Science, vol. 11504, pp. 112-120. https://doi.org/10.1007/978-3-030-21949-9_13

24. Rababah, A, Bond, RR, Rjoob, K, Guldenring, D, McLaughlin, J & Finlay, D 2019, 'Novel hybrid method for interpolating missing information in body surface potential maps', Journal of Electrocardiology, vol. 57, pp. S51-S55. https://doi.org/10.1016/j.jelectrocard.2019.09.003

25. Al-Zaiti, S, Faramand, Z, Rjoob, K, Finlay, D & Bond, RR 2021, The role of automated 12-lead ECG interpretation in the diagnosis and risk stratification of cardiovascular disease. in AS El-Baz & J Suri (eds), Cardiovascular and Coronary Artery Imaging. vol. 1, Elsevier, pp. 45-87. https://doi.org/10.1016/B978-0-12-822706-0.00005-6

26. Bond, RR, Rjoob, K, Finlay, D, McGilligan, VE, Leslie, SJ, Knoery, C, Iftikhar, A, McShane, A, Tache, IA, Biglarbeigi, P, Manktelow, M & Peace, A 2020, 'Near future artificial intelligence in interventional cardiology: new opportunities and challenges to improve the care of STEMI patients', Journal of ESC Digital Health.

27. McGilligan, VE, Watterson, S, Rjoob, K, El Chemaly, M, Bond, RR, Iftikhar, A, Knoery, C, Leslie, S, McShane, A & Bjourson, AJ 2019, 'An exploratory analysis investigating blood protein biomarkers to augment ECG diagnosis of ACS', Journal of Electrocardiology, vol. 57, pp. S92-S97. https://doi.org/10.1016/j.jelectrocard.2019.09.002

28. Biglarbeigi, P, McLaughlin, D, Rjoob, K, Abdullah, , McCallan, N, Jasinska-Piadlo, A, Bond, RR, Finlay, D, Ng, KY, Kennedy, A & McLaughlin, J 2019, Early Prediction of Sepsis Considering Early Warning Scoring Systems. in Computing in Cardiology 2019. https://doi.org/10.22489/CinC.2019.051

29. Leslie, S, Knoery, C, Rjoob, K, Heaton, J, Bond, RR, Peace, A, Bloe, C, Iftikhar, A & McGilligan, VE 2019, 'Retrospective cross-sectional review of outcomes from a nurse delivered pre-hospital diagnostic support service for suspected acute coronary syndrome: the computer is still poor at detecting STEMI', Heart, vol. 105, no. 7. <http://dx.doi.org/10.1136/heartjnl-2019-ICS.7>

Table of Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Problems and challenges:

Cardiovascular disease (CVD) is the main cause of death globally, and most deaths in CVD occur due to ischaemic heart disease. Acute coronary syndrome (ACS) causes the most deaths in ischaemic heart disease. ACS occurs due to blockage or stenosis in the coronary arteries (blood vessels that are responsible for delivering oxygen to heart muscles) which might cause reduced blood to the heart muscles and as a consequence heart cell can die (myocardial infarction). An electrocardiogram (ECG) is the traditional method to diagnose ACS alongside cardiac biomarkers such as troponin and considering clinical risk factors including diabetes mellitus, smoking, hypertension and family history.

Primary percutaneous coronary intervention (PPCI) is a medical procedure (nonsurgical) that is used to treat ACS by inserting a catheter into the blood vessels either in the arm or in the groin using fluoroscopy to open up the stenosis or blockage in the coronary arteries by placing a stent that can improve blood flow. Clinical decision making for primary percutaneous coronary intervention (PPCI) pathways in cardiac care unit follows a sequence of procedures as shown in Figure 1.1.

The PPCI pathway begins when a clinician receives a call from a patient who needs help. Then, the clinician takes the first decision by performing the first clinical assessment and performing ECG. Based on the first clinical assessment, a physician decides whether it is a heart attack or not as shown in CD2 in Figure 1.1. If it is not a heart attack, the physician asks the ambulance service to take the patient to the nearest hospital for further diagnoses. While, if it is a heart attack, physician calls PPCI coordinator to decide whether to bring the patient for PPCI or to turn down the patient. In case that the PPCI coordinator is unable to make the decision, the PPCI coordinator calls medic to decide whether to bring the patient for PPCI or tun down the patient.

Arguably, the medical industry started using artificial intelligence (AI) in the sub-discipline of cardiology (Mayo Clinic, 2021). AI has since demonstrated promising results in different clinical areas such as radiology (Mayo Clinic, 2021). However, there are still opportunities to improve clinical decision making in cardiology such as poor ECG data quality due to lead misplacement or noise (Rjoob et al., 2019), low sensitivity and specificity of ECG (Bond et al., 2012), limited accuracy of blood biomarkers and the limited use of cardiac biomarkers since they typically require a waiting time to attain the result (Kristen et al., 2009). In addition, there is a lack of trust in AI algorithms and adopting AI in cardiac care due to issues related to transparency, explainabilty, uncertainty visualisation and accountability (Davenport et al., 2019; Gerke et al., 2020). With these factors in mind, the following section explains the aim of this PhD regarding to those challenges.

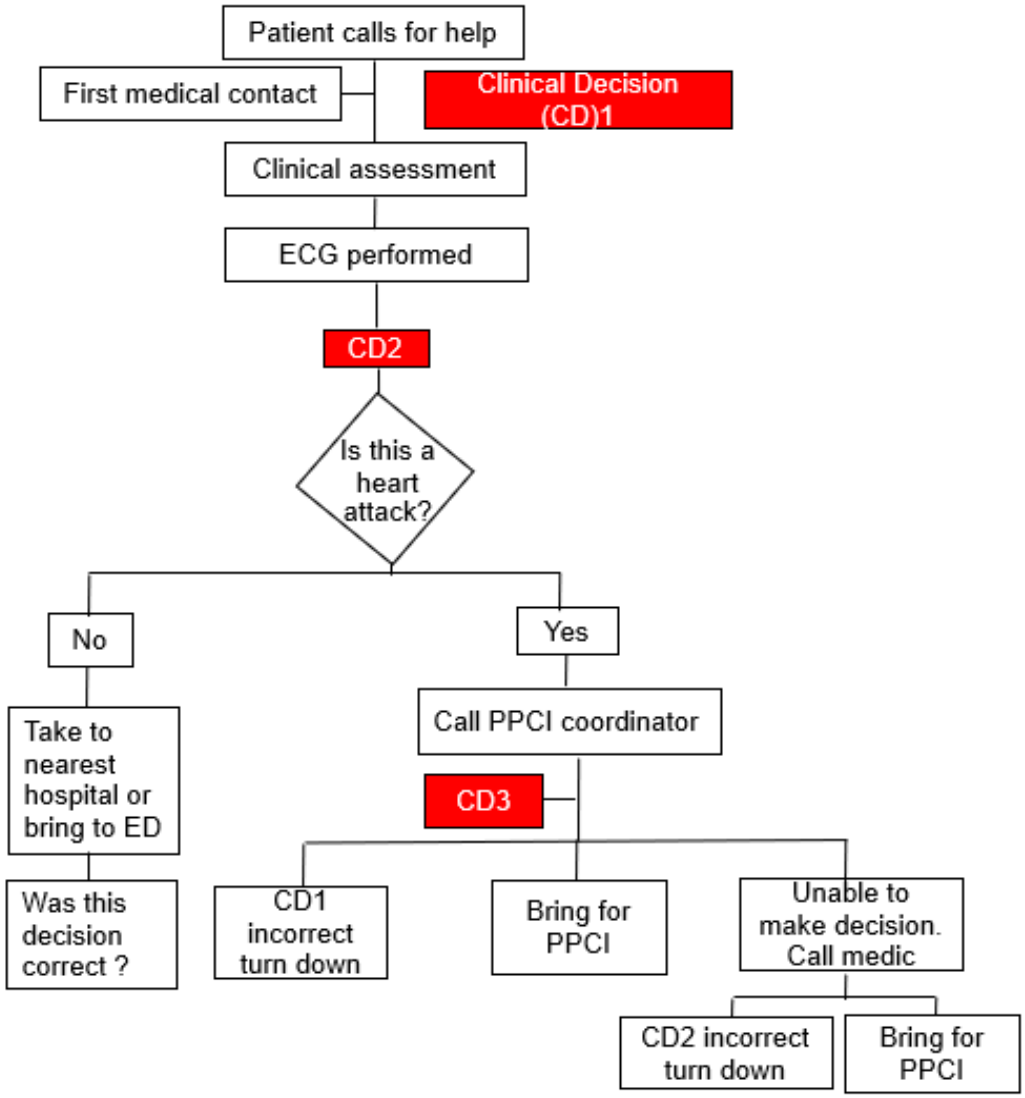


Fig. 1.1 Pathway for PPCI and points where important clinical decisions take place.

1.2 Research aim:

The aim of this PhD is to augment decision making in cardiac care using data analytics and machine learning techniques. This aim was aligned to the following objectives:

1. To improve ECG data quality by detecting lead misplacement using machine learning (ML).
2. To study the impact of noise on ECG interpretation.
3. To discover novel biomarkers that could improve clinical decision-making using ML.

Three key research questions have been investigated this PhD based on the aim and the objectives of the PhD:

4. To what extent can machine learning be used to auto-detect lead misplacement?
5. What useful and actionable insights can we find when we apply Machine Learning (ML) to combine biomarkers?

6. Can new biomarkers be combined with the ECG to improve sensitivity and specificity of ST-Elevation Myocardial Infarction (STEMI) detection?

1.3 Thesis structure:

The thesis has seven chapters to cover all the research questions as shown in Figure 1.2. The following sections summarise each chapter separately.

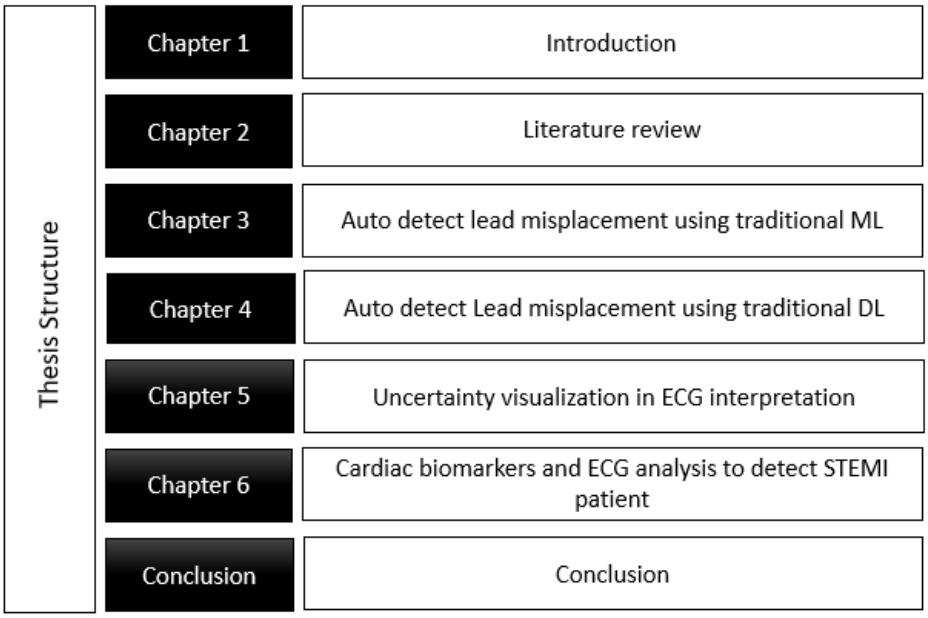


Fig. 1.2 Thesis structure.

The first chapter: Introduction

The first chapter provides an introduction to this thesis, which describes problems and challenges in specific area that need to be addressed. This chapter explains the aim, objectives, key research questions of this PhD and thesis structure.

The second chapter: Literature review

This chapter provides a literature review. In this chapter, the various topics that are related to this work are presented as well as the most up to date research in this area alongside their limitations.

In the first part of chapter 2 (cardiovascular disease or CVD), several facts and statistics are provided about CVD globally. For example, it presents the different types of CVD and how much CVD cost governments globally. Moreover, it explains how governments and the World Health Organisation (WHO) plan to reduce the cost of CVD and minimise the number of CVD patients by meeting a number of targets. Figure 1.3 shows the summary of the second chapter.

This chapter also describes acute coronary syndrome (ACS) which is a branch of CVD and is one of the main causes of death globally (British Heart Foundation, 2021; WHO, 2017; Mc Namara et al, 2019) and the guidelines of ACS diagnosis. In the second part (biomarkers) of chapter 2, several biomarkers were reviewed as they are commonly used to diagnose ACS such as:

1. Screening biomarkers.
2. Antecedent biomarkers.
3. Staging biomarkers.
4. Diagnostic biomarkers.
5. Prognostic biomarkers.

The definition and applications of those biomarkers were described in detail in the literature review chapter.

In the third part of chapter 2 (ECG interpretation), an overview of electrocardiography is provided, including how the human heart works and how the ECG signal is generated. The chapter details which parts of the heart are responsible for generating the ECG signal and details the segments of the ECG cycle. ECG electrode placement is also explained in detail as well as the ECG display format. A recommended list of diagnostic statements for ECG interpretation (which was approved by 'International Society for Computerized Electrocardiology') was also described and explained. A list of new technologies that have been involved in ECG interpretation were also discussed.

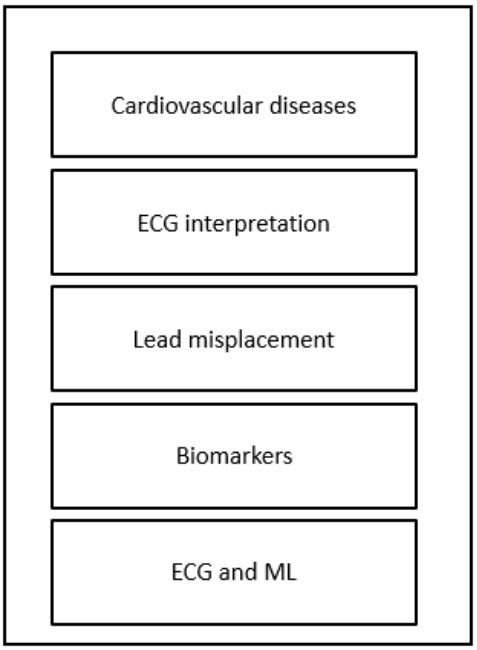


Fig. 1.3 Covered topics in the literature review.

In the fourth part of the chapter (ECG and ML), a systematic review study has been presented to summarise the history of using ML with ECG data. This section presents the limitations of current ML ECG research, the possible applications and the direction and trends for future research. This section also provides guidelines to build a trustworthy ML algorithm. A section of the literature review covered several concerns about artificial intelligence such as transparency, explainabilty and accountability since

those concerns represent barriers that prevent clinicians adopting AI. to show the history of using ECG and ML.

In the fifth part of the second chapter, the electrode misplacement issue when recording a 12-lead ECG is reviewed. This section presents the impact of lead misplacement on ECG morphology and interpretation. Furthermore, it shows what challenges and problems have been investigated in lead misplacement and what challenges remain to be addressed, including the use of ML for detecting lead misplacement. It has been found that the vertical misplacement of leads V1 and V2 have not been investigated using ML and different case studies showed that it is important to detect V1 and V2 misplacement as this can change diagnoses. Interestingly, outside of this work, ML has never been used to detect lead misplacement. It is hypothesised that deep learning (DL) could be used to detect lead misplacement as DL provides a high level of performance for different applications in cardiology such detecting arrhythmias.

The third chapter: Auto detect lead misplacement using traditional ML

In this chapter, traditional ML algorithms such as support vector machines, decision trees and logistic regression have been used to detect the vertical misplacement of leads V1 and V2 in the first, second and third intercostal space. Feature selection algorithms were applied to find the best set of features that can improve ML classifiers. All features in this chapter were manually extracted. ML algorithms obtained a high level of performance in detecting V1 and V2 misplacement in the first and second intercostal space. However, the performance declined in the third intercostal space. Hence, a new AI technology was needed to improve the performance in the third ICS. Figure 1.4 below show the structure of this chapter.

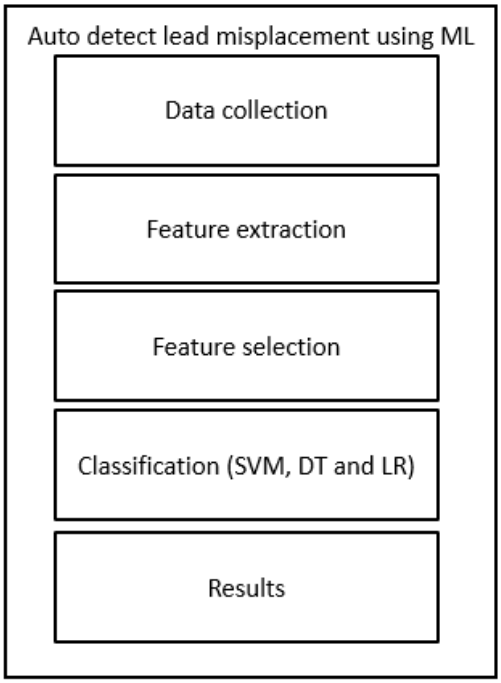


Fig. 1.4 The structure of the third chapter.

The fourth chapter: Auto detect lead misplacement using DL

In this chapter (Figure 1.5), deep learning (DL) was used for the first time in terms of lead misplacement detection. DL was used to improve lead misplacement detection in the third intercostal space. Two DL networks were used including convolutional neural networks (CNNs) and bidirectional long short-term memory network (B-LSTM).

DL obtained a high level of performance when compared to the traditional ML algorithms. However, DL is a black box and does not show the logic of the final prediction when compared to traditional ML (e.g. a decision tree). Hence, a new study was carried out to investigate the explainability and decision making of the DL algorithm. In the new study, attention maps were generated from the last layer of the DL algorithm to show high level features that were responsible for making the final classification. The performance of traditional ML and DL algorithms were benchmarked with the performance of physicians when detecting V1 and V2 misplacement.

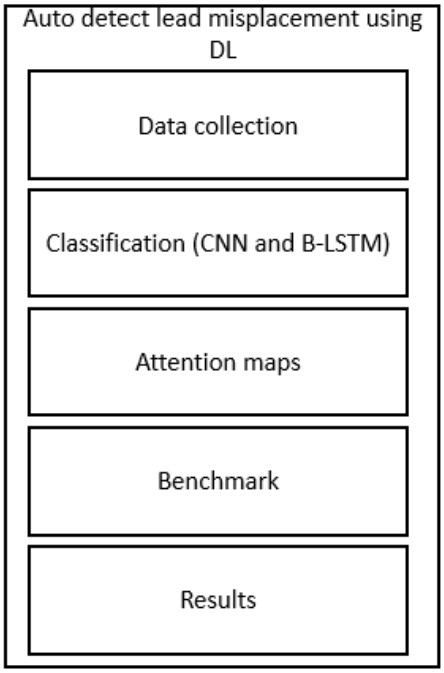


Fig. 1.5 Summary of chapter four.

However, the impact of noise on DL algorithm performance was not investigated and there is a paucity of research that have investigated this. Hence, a new study was carried out to study the impact of noisy ECG data on DL algorithm performance.

The Fifth chapter: Uncertainty visualisation in ECG interpretation

In this chapter, the impact of noise on DL performance was investigated by training and testing a DL algorithm using noisy data in one scenario and training and testing a DL algorithm using filtered data in another scenario as shown in Figure 1.6.

Furthermore, somewhat related to this work, a new uncertainty visualisation method was used for the first time. The new uncertainty visualisation method uses the poor performance of the DL algorithm on the noisy data to show the lower bound of the ‘confidence interval’, while it uses the high performance of DL using the cleaner/filtered ECG dataset to represent the upper bound of the confidence interval. It has been found that noise in the ECG can affect DL performance significantly. Improving ECG data quality by removing unwanted noise and detecting lead misplacement can improve ECG data quality and as a consequence it could improve clinical decision making in cardiac care. However, the enhanced ECG data quality needs to be interpreted alongside sensitive blood biomarkers to improve sensitivity and specificity of decision making. Hence, a new study was carried out to find novel and sensitive blood biomarkers that could be combined with ECG to make final diagnoses more accurate in cardiac care unit.

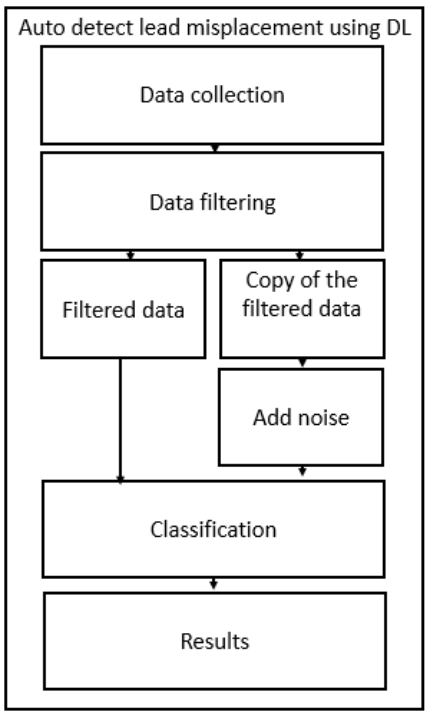


Fig. 1.6 Summary of chapter five.

The sixth chapter: Cardiac biomarkers and ECG analysis to detect STEMI patients

In this chapter, a dataset of blood biomarkers was analysed using ML to find a novel group of blood biomarkers that could improve decision making in cardiac care. In the first experiment, ML learning was used to discriminate between a very high-risk (VHR) group (patients who have ACS for example) and a low-risk group (LR). Different feature selection approaches were applied to optimise the performance of ML. A small group of blood biomarkers were found important and it can provide an accurate diagnosis to detect VHR patients. However, the VHR group has different types of patients such as ACS patients and non-ACS patients, hence, ML was used in the second experiment to discriminate between ACS and non-ACS patients.

Then, the selected blood biomarkers and ECG biomarkers were combined and used to distinguish between ST segment elevation myocardial infarction (STEMI), non-ST segment elevation myocardial infarction (NSTEMI) and unstable angina (UA) in ACS patients. Based on the results, combining blood biomarkers and ECG biomarkers improved the performance.

The seventh chapter: Conclusion

This final chapter review the thesis according to the results and discussions in each chapter and it shows limitations of this thesis and provides future work as shown in Figure 1.7.

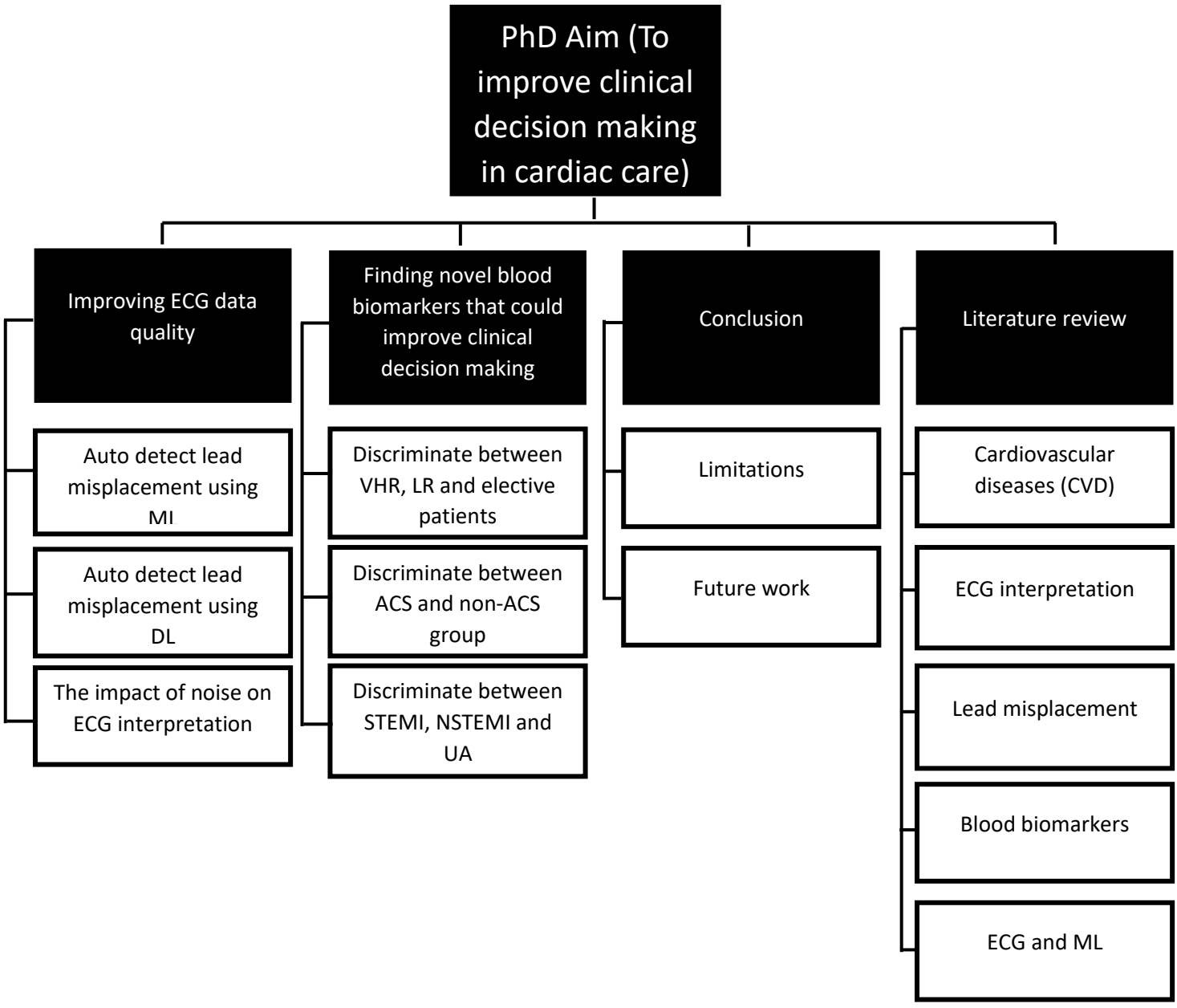


Fig. 1.7 PhD summary.

Chapter 2

Literature Review

2.1 Heart disease:

Cardiovascular disease (CVD) is the primary cause of death globally and is one of the main causes of death in the United Kingdom (UK) (British Heart Foundation, 2021; WHO, 2017; Mc Namara et al, 2019). Approximately 60 million people develop CVD annually and as of 2021, 550 million people across the world are living with the condition (Figure 2.1) (BHF, 2021). CVD costs the UK government £9 billion each year and it cost the economy £19 billion per year (BHF,2021).

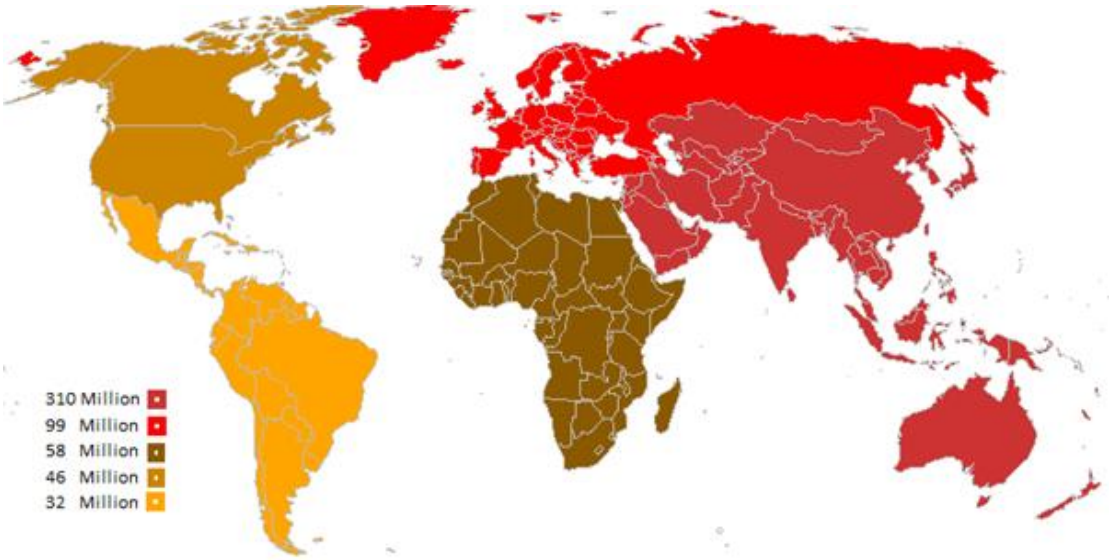


Fig. 2.1 The distribution of the CVD patients across the world.

CVD is an umbrella term encompassing different heart and blood vessels disorders (WHO, 2017; NHS, 2018) such as:

1. Coronary heart disease (CHD): a blood vessels disease that causes a blood supply problem to the heart muscle.
2. Peripheral arterial disease: a blood vessels disease that causes a blood supply problem to the arms and legs.
3. Cerebrovascular disease –a blood vessels disease that that causes a blood supply problem to the brain.
4. Rheumatic heart disease – diseases that damage the heart valves and muscle due to rheumatic fever that might occur because of a streptococcal bacteria.
5. Congenital heart disease – diseases that occur because of problems that happen during the normal development of the heart structure during pregnancy and at birth which might cause different defects in the heart structure including defects in the ventricles, atria or valves.

6. Pulmonary embolism and deep vein thrombosis– a disease that occurs due to a blood clot in the leg veins, and as a result it could move to the heart. The risk factors for CVD include 1) physical inactivity, 2) unhealthy diet, 3) tobacco use, 4) harmful use of alcohol, 5) High blood pressure, 6) high blood glucose, 7) high blood lipids and 8) obesity.

CVD can require medical devices treat some patients such as prosthetic valves, pacemakers and patches (WHO, 2017). Furthermore, CVD sometimes require costly surgical operations to treat patients including 1) balloon angioplasty, 2) coronary artery bypass, 3) heart transplantation, 4) valve repair and replacement, and 5) artificial heart operations.

The World Health Organization (WHO) outline 9 targets (WHO, 2017) to be achieved by 2025 (Figure 2.2): 1) reduction of premature mortality from noncommunicable diseases (NCDs) such as CVD, 2) reduction of physical inactivity, 3) reduction of salt consumption, 4) reduction of smoking, 5) reduction of high blood pressure, 6) 0% increase of obesity and diabetes, 7) coverage of drug therapy and counselling, 8) coverage of essential NCD medicines and technologies and 9) reduction of harmful use of alcohol.

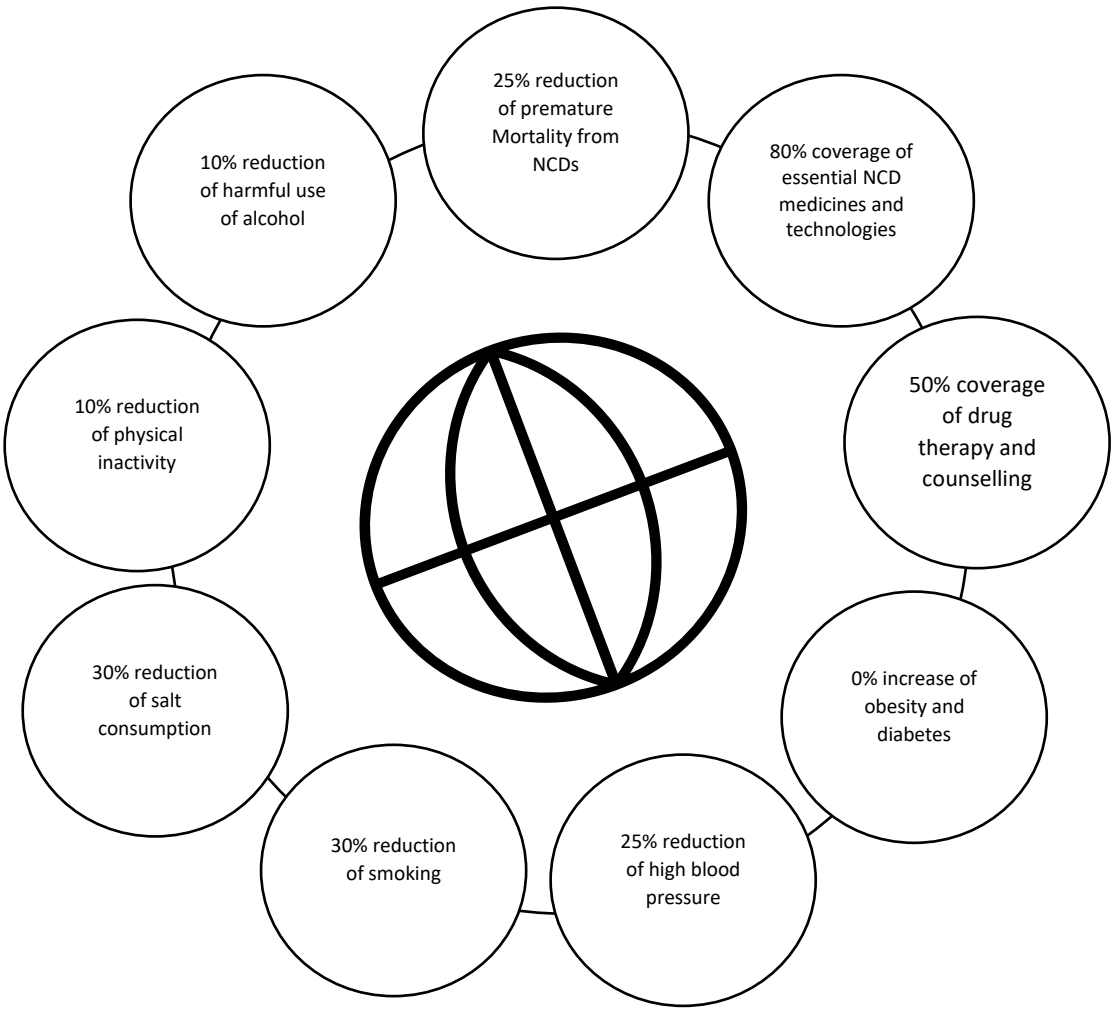


Fig. 2.2 WHO's nine targets to be achieved by 2025 (WHO, 2021).

Acute coronary syndrome (ACS) is a CVD subgroup and is one of the leading causes of death internationally (Sanchis-Gomar et al., 2016; BHF, 2021).

ACS is associated with the decrease of blood flow in the coronary artery because of a blockage in the coronary artery (Kumar et al., 2009). The blockage usually occurs due to a blood clot which can block the coronary artery partially or completely (NHS, 2020) (Figure 2.3).

Fig. 2.3 Coronary artery blockages.

ACS can lead to infarction, ischemia and injury (Akbar, 2021; Kristen, 2009) (Figure 2.4). In ischemia, oxygen level decreases in the myocardium because of the blockage for less than 20 minutes. As a result, it causes changes on T wave on the electrocardiogram (ECG). The second phase is called injury and it occurs when the oxygen reduction lasts more than 20 minutes and it causes ST abnormalities on the ECG (Kristen, 2009). However, the cell damage is still reversible in both ischemia and injury. While infarction occurs when the oxygen reduction lasts more than 2 hours and as a consequence, the cell damage is irreversible and it shows a pathological Q wave on the ECG (Kristen, 2009).

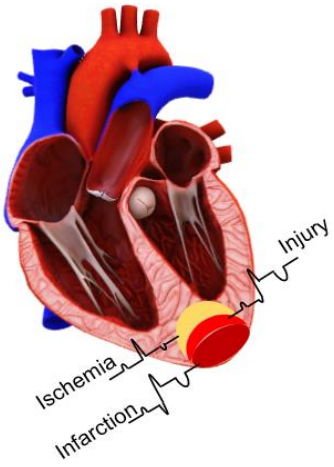


Fig. 2.4 The three phases after ACS occurrence.

ACS is the umbrella term for the symptoms of myocardial ischemia which including ST-segment elevation myocardial infarction (STEMI), non–ST elevation myocardial infarction (NSTEMI), and unstable angina (UA) (Kumar et al., 2009; Sanchis-Gomar et al., 2016). The main difference between STEMI, NSTEMI and UA is based on the

occlusion type, In STEMI patients, the coronary artery is completely blocked, while in NSTEMI and UA, there is a partial blockage in the coronary artery (Figure 2.3). The symptoms of STEMI, NSTEMI and STEMI are quite similar (Basit et al., 2021). Hence, physicians follow guidelines to discriminate between them as shown in the block diagram in Figure 2.5 (Sheridan et al., 2002, Koganti et al., 2015; Christopher et al., 2003).

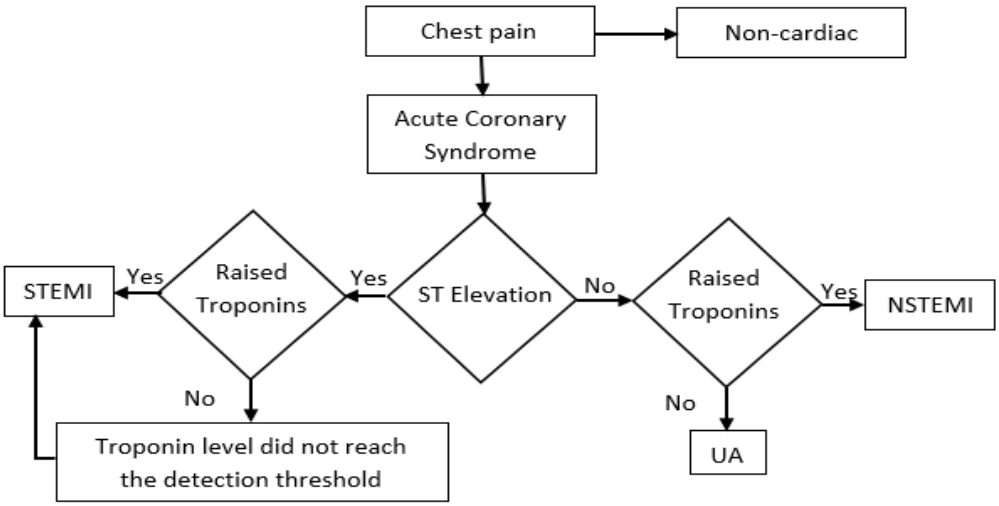


Fig.2.5 ACS diagnosis procedures. ST elevation and troponin level are two biomarkers used through ACS diagnoses.

Late or incorrect diagnosis of ACS results in a high mortality rate (Stewart et al., 2021; Simms et al., 2012; Abreu et al., 2019). Hence, Physicians use blood biomarkers and 12-lead ECG for an early and accurate assessment (Meissner et al., 2010; Loria et al., 2008; Chacko et al., 2017). However, blood biomarkers analysis might take long time (Chacko et al., 2017) and poor ECG quality can result in false diagnosis (Kania et al., 2014; Tanantong et al., 2015). Hence, new blood biomarkers need to be investigated and ECG quality needs improvement to reduce false diagnosis and to augment clinical decision making in cardiac care units.

2.2 Biomarkers:

Biomarkers are defined as features that are used as indicators to measure the development of the biological process in the human body which can be affected by different diseases (Califf, 2018; Strimbu et al., 2010; McGilligan et al., 2019). A biomarker can be a polymorphism, a gene mutation or a protein, or other clinical measurement that indicates a given disease state (Mayeux, 2004). It can be measured in different bio-samples such as a urine, blood, tissues or it could be obtained from ECG, blood pressure, Holter, echocardiogram (ECHO) or computerised tomography (CT) (Ramachandran, 2006) (Figure 2.6).

Biomarkers can be classified into five categories (Ramachandran, 2006) as follow:

1-Screening biomarkers: to screen subclinical diseases.

2-Antecedent biomarkers: to identify the risk of developing diseases.
3-Staging biomarkers: to categorise disease severity
4-Diagnostic biomarkers: to detect the presence of a disease.
5-Prognostic biomarkers: to predict future diseases including response to therapy and monitoring efficacy of treatment.

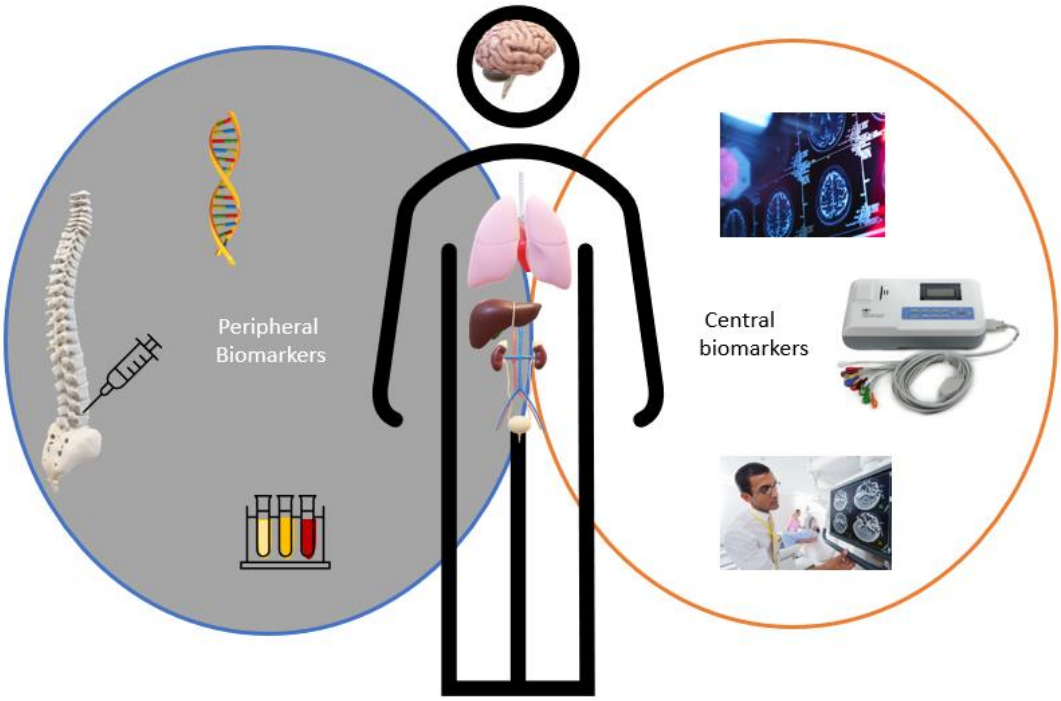


Fig. 2.6 Peripheral biomarkers and central.

Blood biomarkers play an important role to diagnose different diseases such as COVID-19, cancer, prostatitis and ACS (Ponti et al., 2020; Selleck et al., 2017, Wang et al., 2020, Fiorentino et al., 2010). In some cases, investigating blood biomarkers in an early stage might be useful to diagnose serious diseases such as cancer and cardiovascular disease to avoid any complications (Mayeux, 2004). Biomarkers can be used as predictive or prognostic biomarkers (Oldenhuis et al., 2008). The predictive biomarkers are used as an indicator to provide information about the impact of using specific drugs on patient situation, while the prognostic biomarkers are used to describe a patient situation (Oldenhuis et al., 2008).

The human body has hundreds of blood biomarkers that can describe a patient's condition. Blood biomarker however can only be useful when physicians know which biomarkers are most relevant to each case (McCarthy et al., 2017; Nicoletta et al., 2013). Hence, feature selection algorithms can be used to find the most relevant blood biomarkers in large biomarker datasets (McCarthy et al., 2017; Nicoletta et al., 2013). In this case and based on feature importance, blood biomarkers can be ranked from the most relevant to the least relevant biomarker using different feature selection approaches such the embedded method, the filter method, the wrapper method and the hybrid (which combines both the filter method and the

wrapper method) (Liu et al., 2011; Shijin et al., 2011). The highly ranked biomarkers can be selected to be the most important biomarkers based on different metrics such as F score, accuracy or area under the curve (AUC).

However, the performance of the feature selection algorithms could be different compared to other feature selection methods as each feature selection method has its own mathematical function to compute importance. Hence, different feature selection algorithms should be applied to obtain the best performance (Liu et al., 2011; Shijin et al., 2011).

To date, all aforementioned studies considered PCA as a statistical method to investigate new biomarkers and to reduce high false positive rates (Elisa et al., 2014; Gosho et al., 2012).

Several studies have been read during the literature review stage in this PhD and based on those studies, we came across a group of blood biomarkers that were suggested by previous studies to be investigated in the future to obtain fast and an accurate ACS diagnosis. Those studies have been summarised as follow (to show their methods, results and future direction based on their outcomes).

The first study used the most used blood biomarkers to detect ACS is troponin which attempts to discriminate between STEMI, NSTEMI and UA patients, however, it has been found that using troponin only provides a low specificity 17% (Kristen et al., 2009). Hence, combining heart fatty acid binding protein (H-FABP) with troponin T (cTnT) showed enhancement in sensitivity and specificity (McCann et al., 2008).

In the second study, A Semi-quantitative human antibody array of 174 blood biomarkers were analysed by Yi Zhang et al (Zhang et al., 2015) from 20 subjects (STEMI n=10, healthy subjects n=10). They found out that the concentration of blood serum of 21 blood biomarkers is different between healthy cases and STEMI patients. Eight of these biomarkers was taken forward for validation. A significant rise in the level of three biomarkers was observed and the study concluded that PDGF-AA, BNDF and MMP-9 could be used to distinguish between healthy subjects and STEMI patients (Zhang et al., 2015).

James et. al. used blood biomarkers to evaluate clinical treatment outcomes using blood samples that were collected from two groups of STEMI patients. In the first group, interleukin (IL)-6 levels and IL 10 were high, while in the second group, IL-6 levels and with IL10 were low (James et. al., 2012). In the first group, patients died at 6 months compared to the second group. They found out that simultaneous increases of IL6 and IL10 could differentiate STEMI patients with poor clinical treatment outcomes from other STEMI patients. (James et al., 2012). Hence, it begs a new question, “can we consider IL6 and IL10 levels to diagnose STEMI patients?”. The James et al study showed a correlation between mortality in ACS and IL6 level. They concluded that IL6 level increases in UA patients. As a result, IL6 could be used

to build a predictive model to predict myocardial infarction in future (James et al., 2012).

Yan et al recruited 60 patients to analyse the impact of stable angina (SA) and MI on blood biomarkers. Patients were divided into 3 categories (MI n=20, SA n=20 and healthy subjects n=20) (Yan et al., 2015). Based on blood biomarkers levels, they observed no considerable difference between healthy subjects and SA patients (Yan et al., 2015). However, the blood biomarkers levels in the MI group were significantly different compared to the other two categories. MI could therefore be detectable using blood biomarkers. However, blood biomarkers such as Lp-PLA2, CRP and IL-6 tend to diminish over time in patients with MI and UA (Yan et al., 2015). This highlights the importance of sampling time and should be considered in all analyses.

Haas et al studied the genotype of MI patients to find novel prognostic biomarkers. They recruited 30 subjects with acute MI and divided them into three groups with diverse Hp isoforms: the first group included α1-α1 genotype, the second group included α2- α2 genotype and the third group included α2- α1 genotype. They reported that a damaging functional outcome after MI was associated with the presence of low levels of Hp and α2 (Haas et al., 2011).

Govorukhina et al studied the impact of time clotting on blood biomarkers. They used three methods including label-free technique, stable-isotope labelling and univariate and multivariate statistical analyses to investigate the influence of time clotting on blood biomarkers (Govorukhina et al., 2009). Based on the second and the third method, clotting time had a significant influence on biomarkers level. While the first method did not show any significant impact.

Furthermore, several studies have been published about blood biomarkers validation and how to find significant blood biomarkers. Ensor et. al. investigated blood biomarkers validation as it is considered as a very important process in clinical research. Blood biomarkers validation has many concerns including correlated observations, multiplicity and bias (Ensor, 2014). Correlated observations occurs when we have multiple biomarkers for each case. Therefore, we have to select a subgroup of biomarkers that provide good results. While multiplicity might occur in case of multiple biomarkers investigation. To avoid multiplicity and correlated observations. To avoid selection bias, we need to ensure that we have homogeneous data. Hence, the age distribution of patients for example should be homogeneous in different groups to ensure that we obtain accurate results (Ensor, 2014).

Elisa et al. suggested two approaches that are used to find out significant biomarkers including the classical approach which use mono-variate analysis and the multivariate approach which is used to consider the correlation between biomarkers (Elisa et al., 2014).

Hence, based on the aforementioned studies, new blood biomarkers and approaches should be investigated to provide an accurate and fast diagnosis of ACS. Table 2.1 shows suggested blood biomarkers to be investigated in the future based on previous research as it has a correlation with ACS.

Table 2.1 Blood biomarkers associated with CVD.

Protein	Regarding to CVD
EGFR	EGFR is significantly associated with ACS and it may be used as susceptibility biomarker of the ACS (Gao et al., 2008).
OSCAR	The low level of OSCAR is associated with the presence of ACS. Hence, it could be considered to predict of ACS (Zhang et al., 2018).
Gal.9	Gal-9 is associated with coronary artery disease (CAD) and the severity of coronary arteries stenosis (Zhu et al., 2015).
NOTCH 3	Loss of Notch3 could be used to detect a severe heart failure (Hélène et al., 2016).
PSP.D	SP-D level is significantly in raised in patients with CAD. It also could be used to predict the numbers of diseased vessels (Zhen et al., 2012).
CTSL1	Cathepsin L serum level could be used as a novel and independent biomarker to predict CHD (Liu et al., 2009).
AMBP	AMBP can discriminate between four groups significantly (Májek et al., 2011): 1) patients with STEMI, 2) patients with unstable angina, 3) patients with stable angina and 4) control group of normal angiograms.
CCL15	CCL15 level is significantly associated with coronary microvascular dysfunction (Schroder et al., 2019).
OPN	Osteopontin could be used as a biomarker for incident ACS (Yu et al., 2019).
LPL	LPL D9N is associated with the risk of CAD (MA et al., 2018). Reduced level of LPL is associated with the risk for future CAD. High level of LPL may be atheroprotective through increasing HDL-C level and decreasing TG level (Rip et al., 2006). LPL could be used as a novel biomarker to evaluate the risk of patients with ACS (Tian et al., 2014).
CDH5	CHD5 may play an important role as an effective diagnostic marker to detect ACS (Cao et al., 2019).
pIgR	pIgR levels are significantly associated with ACS patients (De et al., 2013).
ADAMTS13	ADAMTS13/VWF ratio significantly reduced in the ACS patients (Pedrazzini et al., 2016).
OPG	OPG is an independent predictor of heart failure development in ACS patients (Jansson et al., 2012).
IL-6	In patients after ACS, IL-6 level is associated with an adverse cardiovascular outcome. Hence, IL-6 could be used as a potential therapeutic target with unstable ischemic heart disease (Fanola et al., 2017).
VEGF.D	VEGF is an independent predictor of ACS patients. Also, VEGF elevation is significantly correlated with the evidence of myocardial ischemia. It also indicates an adverse outcome (Heeschen et al., 2003).
MARCO	Macrophages has a pivotal role in atherosclerosis development. The current results also support the use of macrophage as a biomarker for an increased cardiovascular disease risk (Moroni et al., 2019).
IL-16	IL-16 might play an important role as an indicator in the inflammatory process in AMI patients and correlates with biochemical markers. Further studies are required (Schernthaner et al., 2017).
TIE2	Tie-2 was increased in patients with ACS. Serial changes in the Tie-2 levels between acute and chronic stages in MI (Douglas et al., 2017)
HGF	HGF is a very early biomarker of myocardial necrosis (Lee et al., 2004). In conclusion: 1- The highest HGF values were found in patients ACS. 2- HGF levels decrease to normal values quickly (within five hours from the first measurement. 3- Additional increase in HGF levels were associated with severe complications in the acute stage of STEMI. 4- Subsequent research is required to confirm the preliminary observations (Konopka et al., 2010).
TFPI	TFPI-1 levels are higher in patients with ACS (Konopka et al., 2012). TFPI is involved in the formation of thrombus in patients with ACS (Qixian et al., 2010).
MB	Myoglobin levels higher than the MI detection threshold is correlated with an increased risk of six-month mortality. Hence, myoglobin may be a useful cardiac biomarker for early risk-stratification in ACS patients (Golino et al., 2003; De et al., 2002)

2.3 ECG interpretation:

The heart is a muscular pump that pumps the blood in the circulatory system. Hence, the heart needs oxygen and energy to work. Pumping process is controlled by an electrical conduction system that adjust the contraction of the chambers of the heart including left atrium, left ventricle, right atrium, and right ventricle (Meek et al., 2020).

The electrical conduction system consists of several nodes and branches including sinoatrial (SA) node, atrioventricular (AV) node, left bundle branch and right bundle branch. The SA node initiates the electrical stimulus by generating an impulse. Then the electrical stimulus propagates through the conduction pathway to the left ventricle via the left bundle branch and to the right ventricle via AV node (Figure 2.7). The two atria are stimulated and contracted before the two ventricles and the contraction of both ventricles is called a heartbeat (Meek et al., 2020).

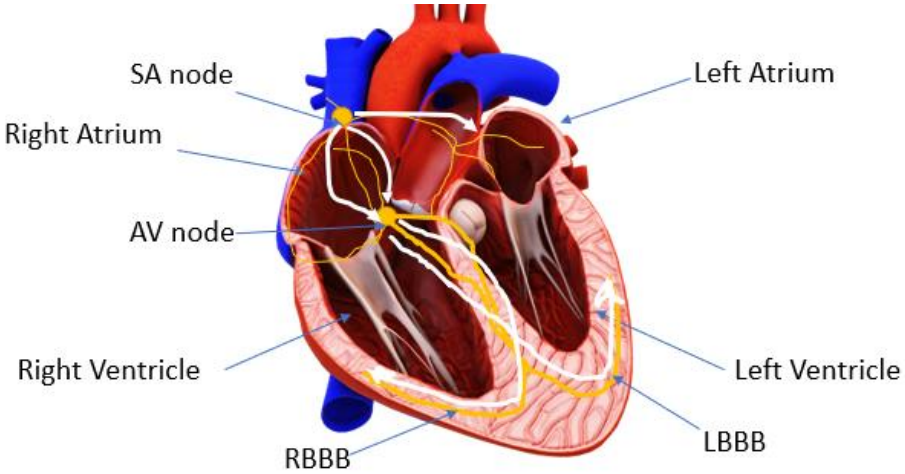


Fig. 2.7 Heart anatomy.

An ECG is the traditional diagnostic tool to detect cardiac abnormalities such myocardial infarction (MI) by recording electrical signal in the heart. Electrical signal changes reflect repolarisation and depolarisation which corresponds to heart relaxation (diastole) and contraction (systole) (Meek et al., 2020).

An ECG is a painless test that comprises 10 electrodes which are used to derive 12 leads. The twelve leads are divided between six precordial leads (V1, V2, V3, V4, V5 and V6) and six limb leads (I, II, III,aVF, aVL and aVR) (Meek et al., 2020). As shown in Figure 2.8, the aforementioned leads can be combined in a particular way to show a specific view to the heart as follow:

1- Lead I shows the left ventricle and the lateral wall.
2- Lead II, III, and aVF show the left ventricle, the inferior wall.
3- lead aVL shows the left ventricle and the high part of the lateral wall.
4- Lead aVR reciprocal of leads II, aVL, V5 and V6.
5- Lead V1 and V2 show both ventricles and the anterior wall.

6- Lead V3 and V4 show the anterior wall of the left ventricle and it also shows parts of the septum.
7- Lead V5 and V6 show the lateral wall of the left ventricle and it shows apex of the heart.

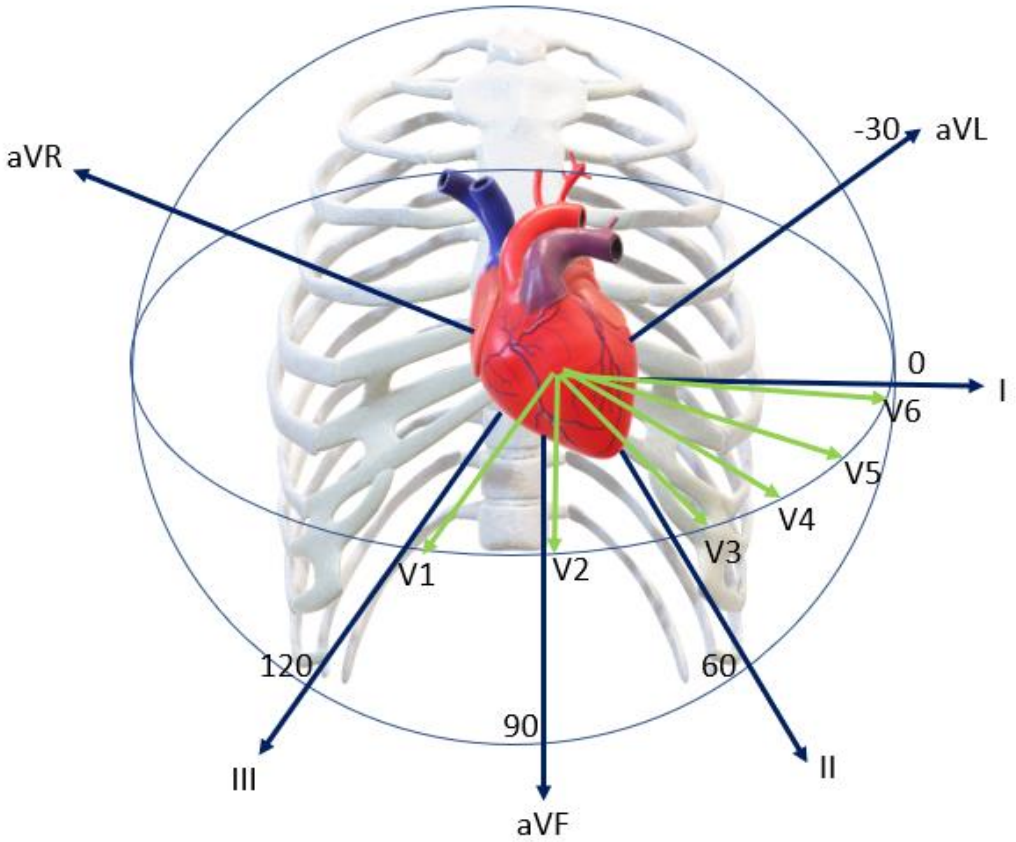


Fig. 2.8 Leads orientation.

Limb electrodes can be place in different positions to place, for example, limb electrodes can be placed at the wrists and ankles or it can be placed more proximal to the hips and shoulders (Meek et al., 2020).

As shown in Figure 2.9, chest leads are placed on six different positions as follow:

1-Lead V1 is placed on the fourth intercostal space (ICS) at the right sternal edge.
2-Lead V2 is placed in the fourth ICS at the left sternal edge.
3-Lead V3 is placed midway between V2 and V4.
4-Lead V4 is placed in the fifth ICS in the mid-clavicular line.
5-Lead V5 is placed in the left anterior axillary line at same horizontal line as V4,
6-Lead V6 is placed in the left mid-axillary level at same horizontal line as V4 and V5 (see figure 8) (Kligfield et al., 2007).

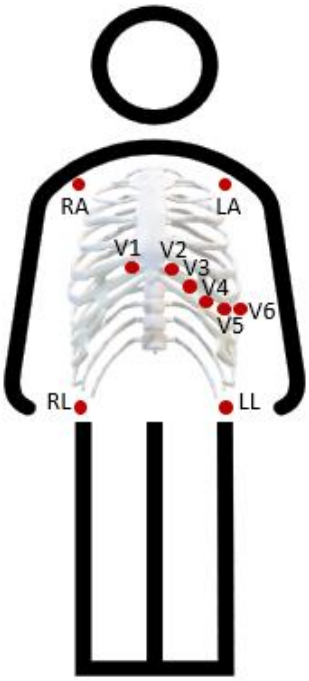


Fig. 2.9 Leads positions.

The ECG has several morphological features such as P wave which represents the depolarisation of the both atria (simultaneous activation), QRS complex represents the depolarisation of both ventricles (simultaneous activation), ST-Segment which represents the repolarisation of both ventricles, U wave which is not common in all people, PR interval which shows time interval between ventricles depolarisation and atria depolarization, QT interval which is the duration of depolarisation and repolarisation of ventricles, PP interval is the atrial cycle duration, RR interval is the ventricular cycle duration (Ashley et al., 2004; Maršánová et al., 2017; Zeraatkar et al., 2011).

The ECG signal is printed on standard paper. However, a paper speed in the US is 25 mm/s (1 mm = 0.04 s). While in other countries, a paper speed is 50 mm/s (1 mm = 0.02 s). Each ECG paper contains large squares and small squares. Each large square consists of twenty-five small squares (5 width X 5 high). Each small square measuring is 1 mm wide and equal to 0.04 s (Figure 2.10) (Meek et al., 2002). Those squares are used to compute the amplitude of P, Q, R, S, T, U wave, heart rate and to see if there is ST segment elevation. However, ECG machines may have different specifications which might cause a different evaluation of the morphological features such as P, QRS amplitudes and intervals (Kligfield et al., 2007). Furthermore, ECG signal is not stationary and varied, even for one patient due to different factors such as age, gender, and lifestyle even for one patient. Hence, American heart association (AHA) Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology is working to review all different technologies that are used in ECG recording to standardise ECG interpretation and add recommendations that could improve physician’s performance (Kligfield et al., 2007).

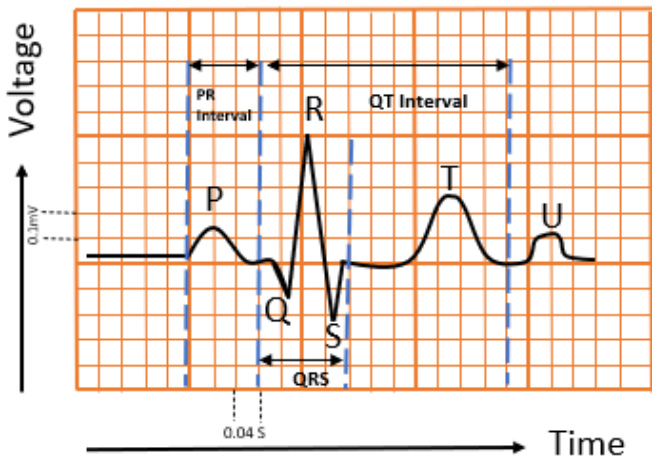


Fig. 2.10 ECG Morphology.

The AHA recommended a statement list which was approved by the 'International Society for Computerized Electrocardiology' (Mason et al., 2007). The statement list consists of diagnostic terms was used to standardise ECG interpretation. The statement list has four different statements including: 1) primary statements, 2) secondary statements, 3) modifiers statements and 4) comparison statements. Each statement has several lists of statements as shown in Table 2.2 for example.

Table 2.2 Four AHA recommended statement lists.

Primary statements	Secondary statements	Modifiers	Comparison statements
A. Overall interpretation: for example, normal ECG or abnormal ECG	Suggests: for example, Pericardial effusion, brugada abnormality, hypercalcemia, hypothermia	General: for example, borderline, possible, prominent	No significant change
B. Technical conditions: such as misplaced precordial electrodes.	Consider: for example, acute ischemia	General conjunctions such as consider, or, and, with	Significant change in rhythm
C. Sinus node rhythms and arrhythmia: for example, sinus tachycardia		MI: for example, acute, old, recent.	New or worsened ischemia or infarction
D. ST segment, T wave and U wave: ST deviation		Repolarisation abnormality such as elevation, depression	New conduction abnormality
E. Supraventricular tachyarrhythmias: for example, atrial fibrillation			Significant repolarisation change
F. Myocardial infarction: for example, anterior MI			Change in clinical status

Hence, to obtain a clear interpretation and an accurate diagnosis, the four statement should be combined together using general rules in the final diagnosis.

New technologies started appearing in ECG systems which resultled in the following trends for ECG technology (DAIC, 2015):

1. Moving from ECG paper format to full digital formats.
2. Continuous ECG recording and interpretation in the Future.
3. Using wavelet transform in ECG interpretation to offer new features for the cardiologists to make their diagnoses.
4. Toward vendor neutral interfaces by developing open platforms for ECG that can interface with other vendors' technology easily including the use of HL7 IT interfaces and DICOM format waveforms.
5. AI integration into ECG.
6. New technology to assist clinicians to place ECG electrodes properly.

2.4 ECG and machine learning:

In the 1960s, a revolution happened in 12- lead ECG interpretation when Pipberger and colleagues (Pipberger et al., 1961) developed the first computer software to interpret the ECG automatically. This revolution started after digitising the ECG which occurred 1950s by (Harold, 2018). They started converting the recorded ECG signal into a digital format (Harold, 2018) which allowed them to process the signal on a computer. As a result, different digital formats are used to store ECG (Bond et al., 2011). Hence, it enabled computers to process 12-lead ECGs. Since then, several computer algorithms were developed to improve the performance of the automated ECG interpreters such as Glasgow software which was developed by Peter Macfarlane and his team (Macfarlane et al., 2005). Glasgow software is a rule-based algorithm uses ECGs combined with a group of five clinical variables such as gender, age, medication, race and history to make the final decision. In1996, the algorithm was approved by the Food and Drug Administration (FDA) for use in the United States (US). Harold, 2019 showed that having automated ECG interpreters minimised ECG reading time between 24% to 28%, improved physician's interpretations significantly ($P<0.005$) by 25% and reduced false diagnoses in the emergency room, especially for junior doctors (Harold, 2019; Morisbak et al., 1999; Southern et al, 2010; Jakobsson et al, 1985). Hence, the presence of the automated interpreters showed a good collaboration between humans and computers in medicine and became routine in cardiac care unit (Harold, 2018).

However, automated ECG interpreters are rule-based algorithms which use a group of logical statements in a form of rules to deduce a diagnosis (Cairns et al., 2017; Geoffrey et al., 2019). Moreover, these rules are derived from known guidelines, diagnostic criteria and expert knowledge. Hence, these algorithms (also known as expert systems) are knowledge driven (deduction) as opposed data driven (induction) (Dehghan et al., 2015). Whilst these algorithms were popular in previous decades, new algorithms in machine learning (ML) form started becoming the trend towards

'data driven' algorithms. Data driven means ML algorithms can 'self-learn' with less supervision from knowledge experts to obtain the diagnosis according to diagnostic patterns that were obtained by training the ML algorithm on a dataset of labelled ECGs (Alahakoon et al., 2020). ML is a branch of artificial intelligence which enables computers to make decision by looking for a pattern in a massive amount of dataset (Feldman, 2001). ML consists of three categories including 1) supervised ML, 2) unsupervised ML and 3) semi-supervised ML. In supervised ML, the ML algorithm is trained on a labelled dataset, while in the unsupervised ML, the algorithm is trained on unlabelled dataset. In the semi-supervised ML, the algorithm is trained on a semi-labelled dataset (Feldman, 2001).

ML has traditional ML algorithms such as support vector machine (SVM) and decision tree (DT) and deep learning (DL) algorithms such as deep neural networks DNNs and convolutional neural networks (CNN) (Manal et al., 2019). In the traditional ML algorithms, data scientists start with feature engineering which includes feature extraction (manually) and feature selection and then classification (Figure 2.11a). While in DL, data are fed into the algorithm without any feature engineering as the DL extract the feature by itself automatically and then classification (Chauhan et al., 2018) (Figure 2.11b).

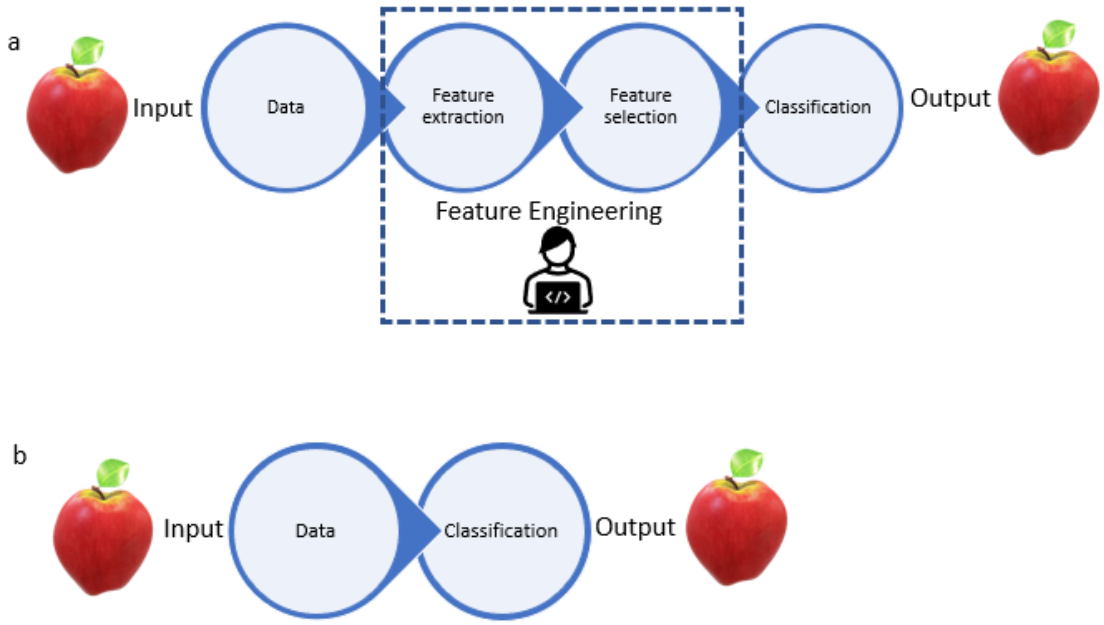


Fig. 2.11 Machine learning and deep learning structures. a represents ML structure and b represents DL structure.

However, Compared to ML, DL requires a large amount of data to ensure the performance and it requires high-performance computers (Najafabadi et al., 2015; Fang et al., 2019). Moreover, DL takes long time during the training process due to the size of the dataset (Ajay et al., 2019) (Figure 2.12).

Three different databases including 1) SCOPUS, 2) IEEE and 3) PubMed were searched and we found that ML is a rapidly evolving area in ECG interpretation, especially in the last decade (Figure 2.13).

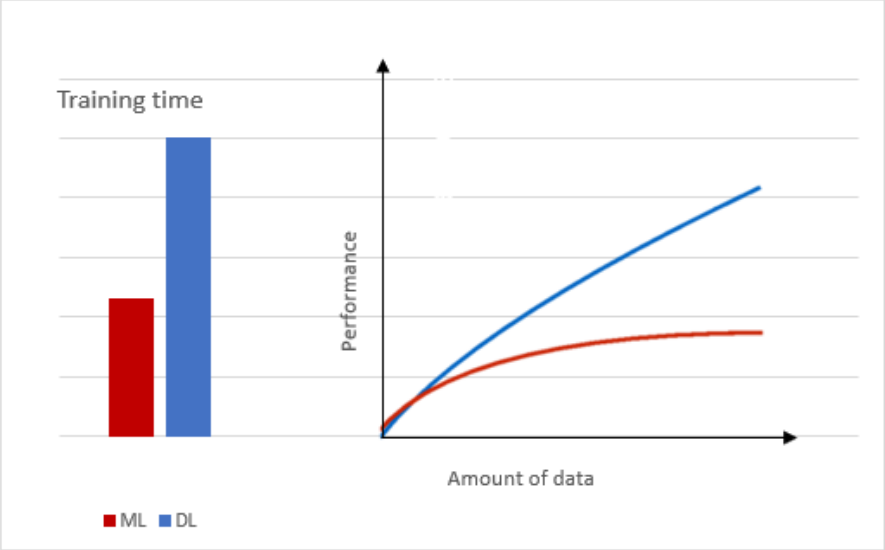


Fig. 2.12 A comparison between DL and ML in terms of training time and the performance.

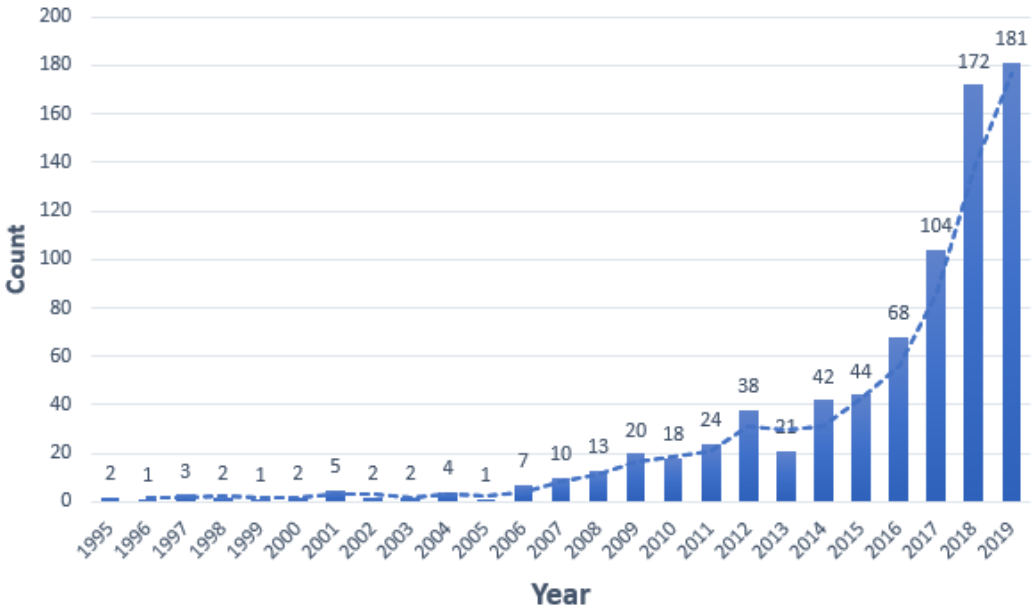


Fig. 2.13 The number of published papers in ECG and ML in the past two decades.

Keywords including "Machine Learning", "ECG", "Deep Learning" and "Artificial Intelligence" were combined in different combinations and sequences to achieve the maximum search sensitivity. The following criteria were used to include articles: 1) original articles that used ML with ECG and the articles are written in the English and 2) ECG dataset was clearly defined and a clear description of ML techniques. Articles were excluded if they did not use ML with ECG. 2116 articles were identified, of them

761 duplicate articles were removed. After removing the duplicated articles, 1355 articles were screened based on abstract and title. After this step, 757 articles were included and 598 articles were removed, and 757 articles were included. All articles which passed primary screening based on the title and abstract were considered. From each article, we extracted different results such as area under the curve (AUC), F1 score, accuracy, sensitivity and specificity and the type of ML technique. Since 2012, the number of researches that used ECG and ML increased significantly (P<0.001) (Figure 2.14).

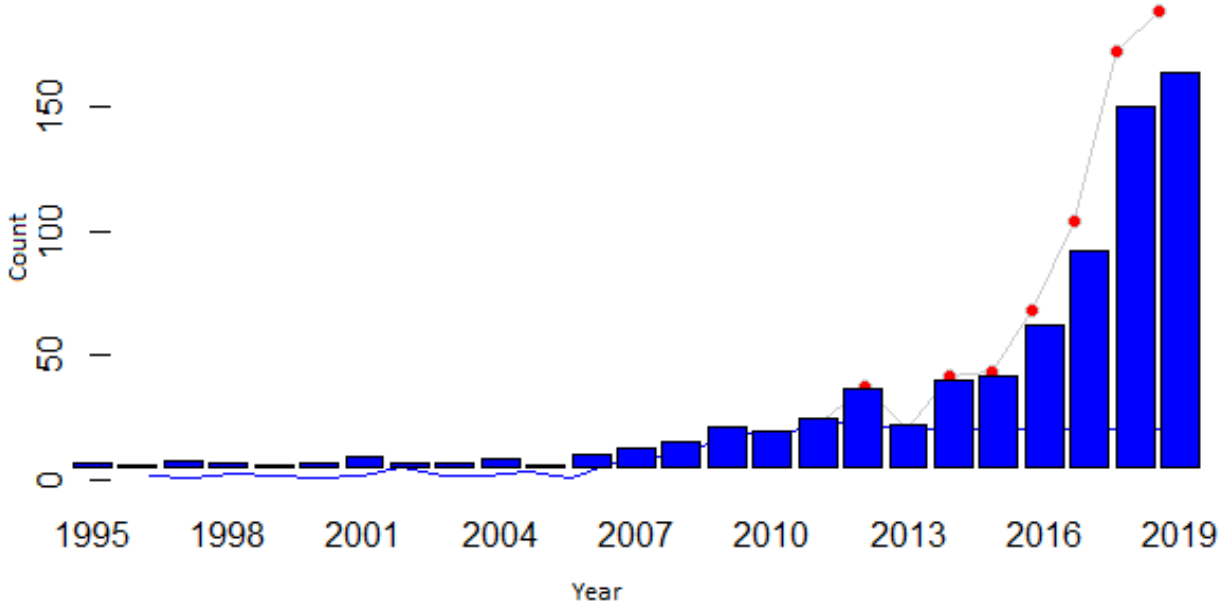


Fig. 2.14 The number of papers that address the use of ML with ECGs per year as identified in PubMed, IEEE and SCOPUS databases. Red dots represent the years where the number of published papers is different from the number of predicted published papers and the predicted number significantly increased. The blue line represents the predicted number of published papers per year, while the grey line shows the actual number of published papers per year. As this search ended at the beginning of October 2019, the last three months in 2019 were extrapolated by taking the average number of published papers per month in the previous nine months in the same year.

Based on this search, 65 different ML algorithms were used with ECG data since 1996 as shown in table 2.3. The frequency of using these techniques over the past two decades (Figure 2.15). As shown in Figure 2.15, SVM has been the most used technique with ECG data. When the ML techniques were collapsed down using five tribes taxonomy into five tribes (Domingos, 2015), the connectionist techniques which includes ANN were trending upwards from 1995 to 1996 and in 2006. And then in 2019 it re-emerged as popular. SVM and KNN techniques which are called analogizers became the trend from 2000 to 2004 and from 2007 to 2018. While Bayesian techniques were trending upwards from 1997 to 1998 (Figure 2.16).

Table 2.3 ML techniques and their frequency of use from 1995 to 2019.

ML	#	%
SVM: Support Vector Machine	332	43.9%
ANN: Artificial Neural Network	125	16.5%
KNN: K-Nearest Neighbour	73	9.6%
DT: Decision Tree	69	9.1%
RF: Random Forest	63	8.3%
CNN: Convolutional Neural Network	62	8.2%
ELM: Extreme Learning Machine	46	6.1%
NB: Naive Bayes	31	4.1%
DL: Deep Learning	29	3.8%
LOG: Logistic Regression	23	3.0%
LDA: Linear Discriminant Analysis	22	2.9%
LSTM: Long Short-Term Memory	16	2.1%
HMM: Hidden Markov Model	10	1.3%
DBN: Deep Belief Network	8	1.1%
AdaBOOST: Adaptive Boosting	8	1.1%
GA: Genetic Algorithm	7	0.9%
LR: Linear Regression	7	0.9%
K-means	6	0.8%
BPN: Back Propagation Network	5	0.7%
GMM: Gaussian Mixture Model	5	0.7%
DL-SVD: Dictionary Learning Algorithm Based on Singular Value Decomposition	4	0.5%
SVR: Support Vector Regression	3	0.4%
ESN: Echo State Networks	3	0.4%
ANFIS: Adaptive Neuro-Fuzzy Inference System	3	0.4%
SOM: Self-Organizing Map	2	0.3%
GBM: Gradient Boosting Machines	2	0.3%
MLC: Maximum-Likelihood Classifier	2	0.3%
CRF: Conditional Random Fields	2	0.3%
LVQ: Learning Vector Quantization	2	0.3%
J48	2	0.3%
SMO: Sequential Minimal Optimization	2	0.3%
BEAT: Beat-to-Beat Estimation by Adaptive Training	2	0.3%
EMD: Empirical Mode Decomposition	2	0.3%
ENL: Elastic Net Logistic	1	0.1%
SM: Statistical Model	1	0.1%
KLR: Kernel Logistic Regression	1	0.1%
SPDR: Sample Percentage in the Dynamic Range	1	0.1%
ZCR: Zero-Crossing Rate	1	0.1%
SDSM: Smart Decision Support Module	1	0.1%
TDEBOOST	1	0.1%
SL: Supper Learner	1	0.1%
TREEBOOST	1	0.1%
TASOM: Time-Adaptive Self-Organizing Map	1	0.1%
BICO: Online Clustering Algorithm	1	0.1%
RB: Rule-Based	1	0.1%
ESS: Ensemble Based Score System	1	0.1%
AMGLVQ: Adaptive Multilayer Generalized Learning Vector Quantization	1	0.1%
CFM: C-F model	1	0.1%
VF15: Voting Feature Intervals	1	0.1%
DTW: Dynamic Time Warping	1	0.1%
SKF: Switching Kalman Filter	1	0.1%
D-Logic: Decision Logic	1	0.1%
ADMM: Alternating Direction Method of Multipliers	1	0.1%
HDC-MER: HD Computing-based Multimodality Emotion Recognition	1	0.1%
ZC: Zero Crossing	1	0.1%
LTMIL: Latent Topic Multiple Instance Learning	1	0.1%
AIRS: Artificial Immune Recognition System	1	0.1%
XGBOOST	1	0.1%
FIA: Fuzzy Immune Approach	1	0.1%
FCM: Fuzzy C-Means	1	0.1%
SRC: Sparse Representation Classifier	1	0.1%
BOOSTSTRAP	1	0.1%
DFA: Discriminant Function Analysis	1	0.1%
NCA: Neighbourhood Components Analysis	1	0.1%
RVM: Relevance Vector Machine	1	0.1%

%= #/757 (where 757 is the total number of papers)

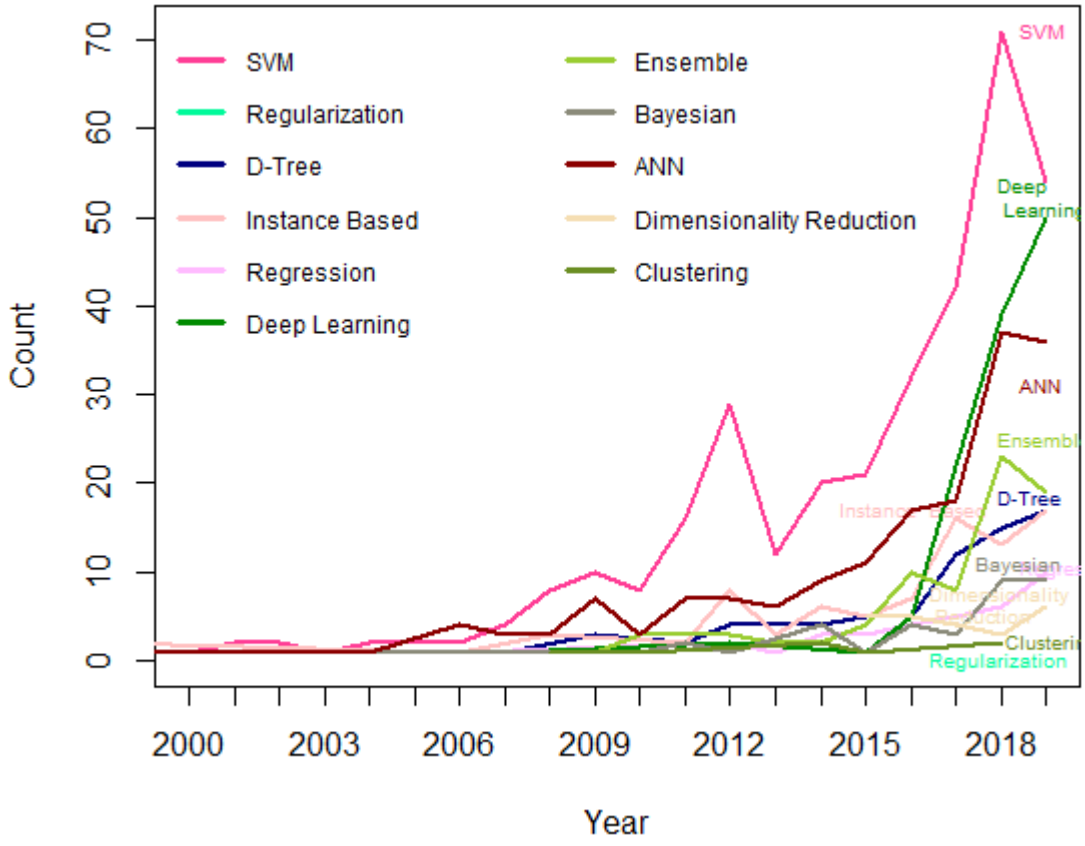


Fig. 2.15 Trending algorithms from 1995 to 2019.

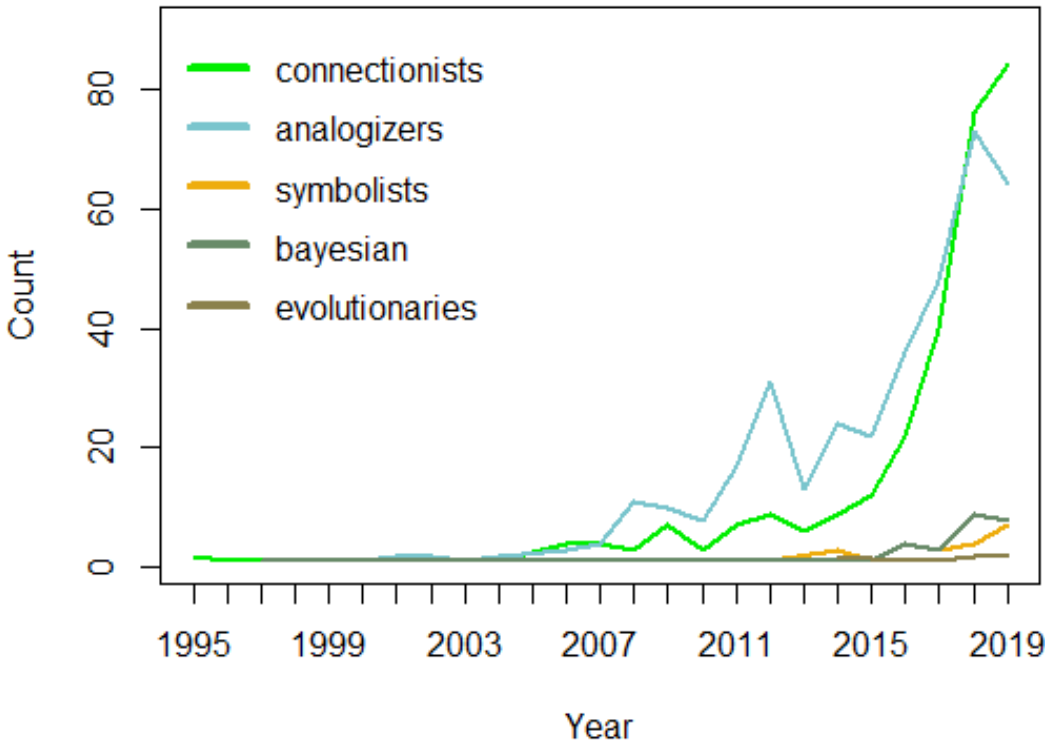


Fig .2.16 Five tribes trend.

However, DL which started increasing sharply and significantly (P<.001) in the area of ECG and ML since 2016 (Figure 2.15 and 2.17).

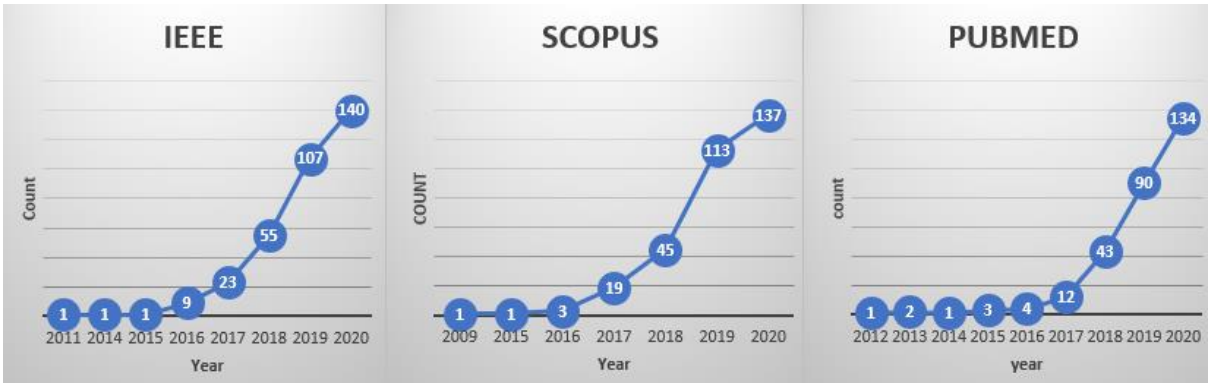


Fig. 2.17: The number of published papers in ECG and DL.

ML and ECG were used to develop several applications such as ECG classification (Ribeiro et al., 2020; Parvaneh et al., 2019), lead misplacement detection (Rjoob et al., 2019; Rjoob et al.; 2020), sleep apnoea (Mostafa et al., 2019; Álvarez et al., 2020), emotion recognition (Ekaterina et al., 2020; Domínguez-Jiménez et al.; 2020), activity classification (Luís et al., 2021; An et al., 2021) and lung disease detection (Mekov et al., 2020; Kaplan et al., 2021). And during the COVID-19, new studies started using 12-lead ECG to detect patients with COVID-19 as it has an impact on the morphology of the ECG, especially on ST segment (Angeli et al., 2020; Long et al., 2021). However, the majority of these articles (n=400/757) focused on ECG classification applications significantly (P<.001) to detect cardiac abnormalities. Cardiac abnormalities were classified into two groups: 1) arrhythmias (AR) group (n=202/400) and 2) non-arrhythmias group (non-AR) (n=62/400) (Figure 2.18a). In AR group, most of the articles (57.53%) focused on atrial fibrillation (AF) detection followed by premature ventricular contraction (PVC) (17.8%), ventricular fibrillation (VF) (9.5%), bradycardia (5.5%), tachycardia (4.1%), ventricular tachycardia (VT) (4.1%) and supraventricular tachycardia (SVT) (1.4%) (Figure 2.18b). In the non-AR group in Figure 2.18c, the majority of the articles studied ischemia and infarction detection (44.79%) followed by heart failure, structural abnormalities, coronary atherosclerosis, cardiac fibrosis, hypertrophic cardiomyopathy, abnormal myocardial relaxation, congenital heart disease, intradialytic hypotension and ventricular dysfunction. However, the performance of ML algorithms were varied between the aforementioned abnormalities in both groups AR and non-AR in terms of sensitivity and specificity. In the AR group, detecting tachycardia using ML obtained the highest sensitivity (94.5%) and specificity (97.3%) compared to the other AR categories (Figure 2.19). However, in tachycardia, all classifiers obtained a high similar sensitivity and specificity which indicates that the ECG morphological of tachycardia is prominent in most subjects which can enable most of ML algorithms to detect tachycardia easily. In terms of ML algorithms performance, KNN obtained the highest performance to detect atrial fibrillation, while SVM outperformed the other algorithms to detect premature ventricular contraction with 99% sensitivity and 95% specificity. Hence,

those results emphasise the ‘no free lunch theorem’ which tells us that there is no single ML algorithm can be the winning algorithm to solve all problems (Ho et al., 2002). Some results seem likely to be dependent on a filtered dataset, while the other results seem to have a noise or it might have artefacts such as lead misplacement.

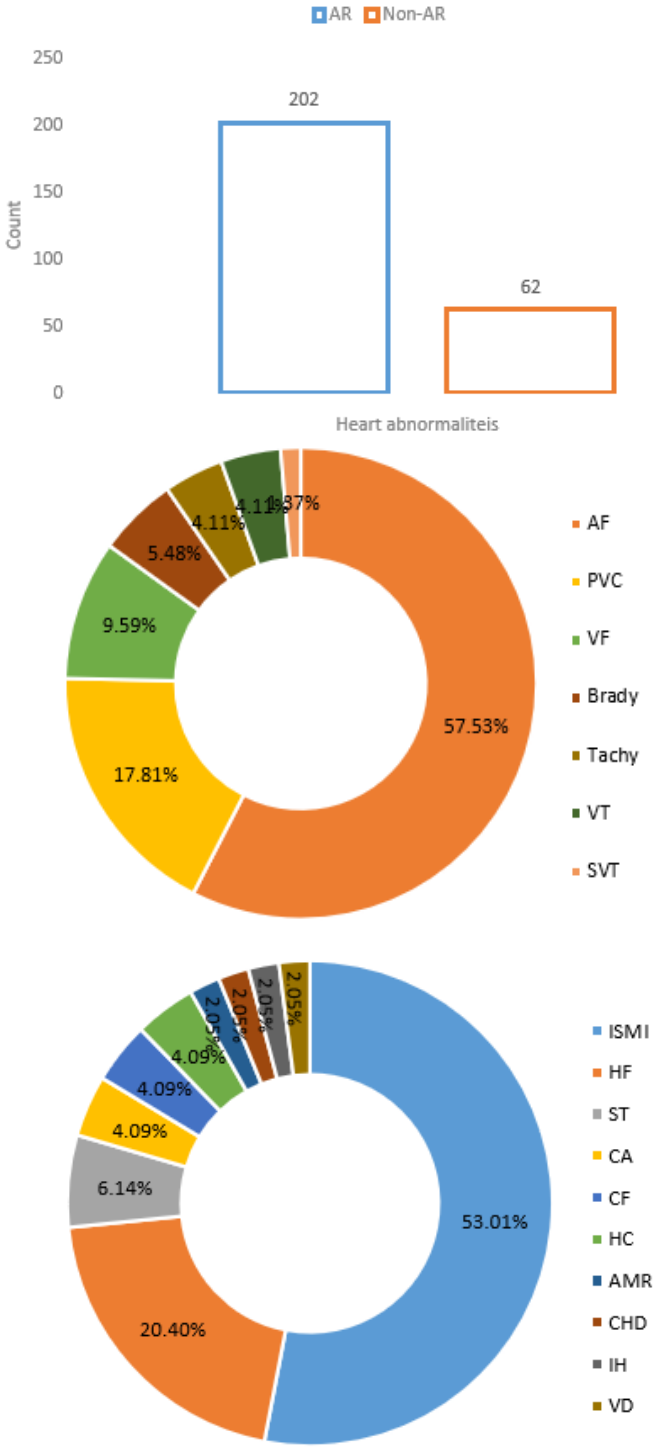


Fig. 2.18: Articles distribution. a represents distribution between arrhythmia and non- arrhythmia. b represents the arrhythmia group distribution. c represents the non-arrhythmia group distribution.

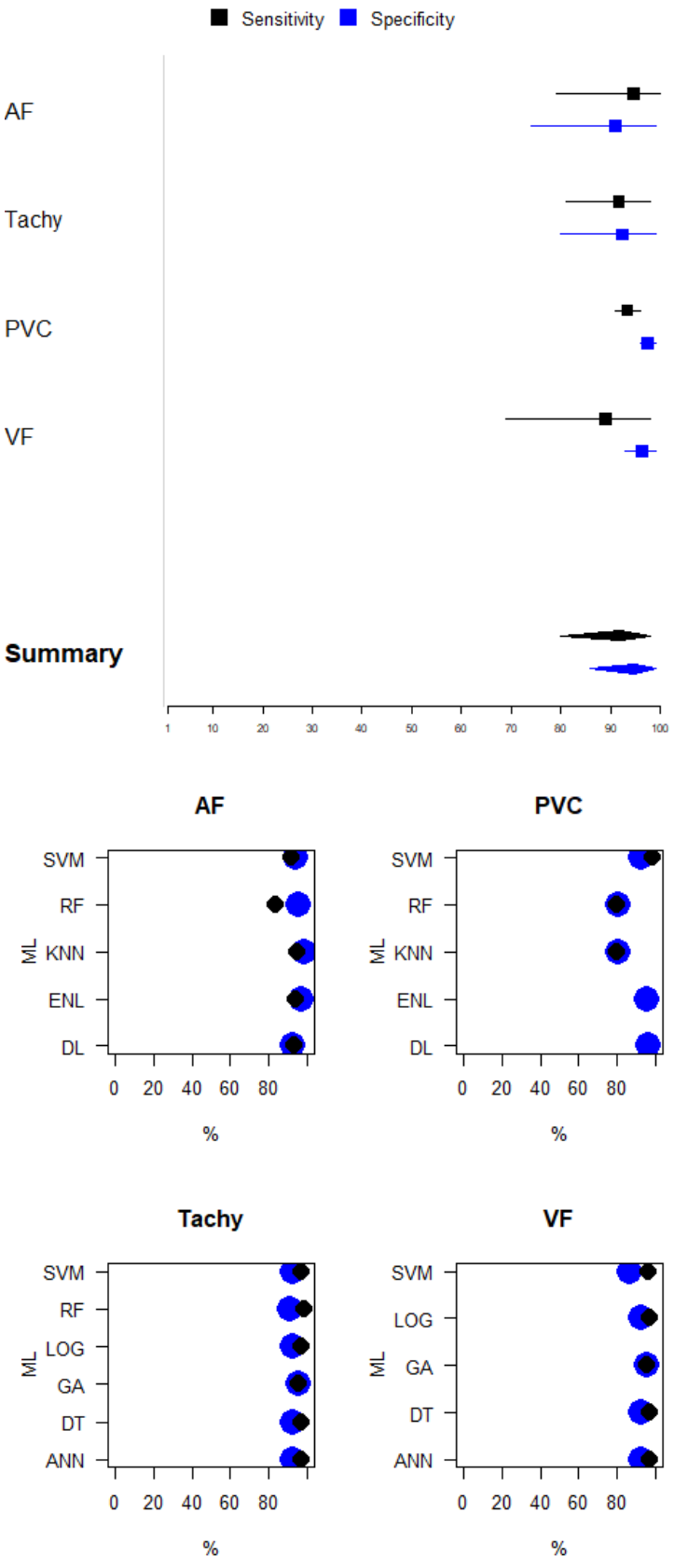


Fig. 2.19 The performance of ML techniques to detect the arrythmia group according to sensitivity and specificity.

In the non-arrhythmia group, SVM obtained the highest performance to detect coronary atherosclerosis with 92% sensitivity and 94% specificity (Figure 2.20). While in heart failure, the decision tree and KNN obtained the best performance with 99% sensitivity and 99% specificity, which might indicate that the control subjects consisted of healthy cases and the heart failure subjects were selected from the most severe patients. In ischemia and infarction detection, multiple ML achieved the best specificity (82%) and sensitivity (95%).

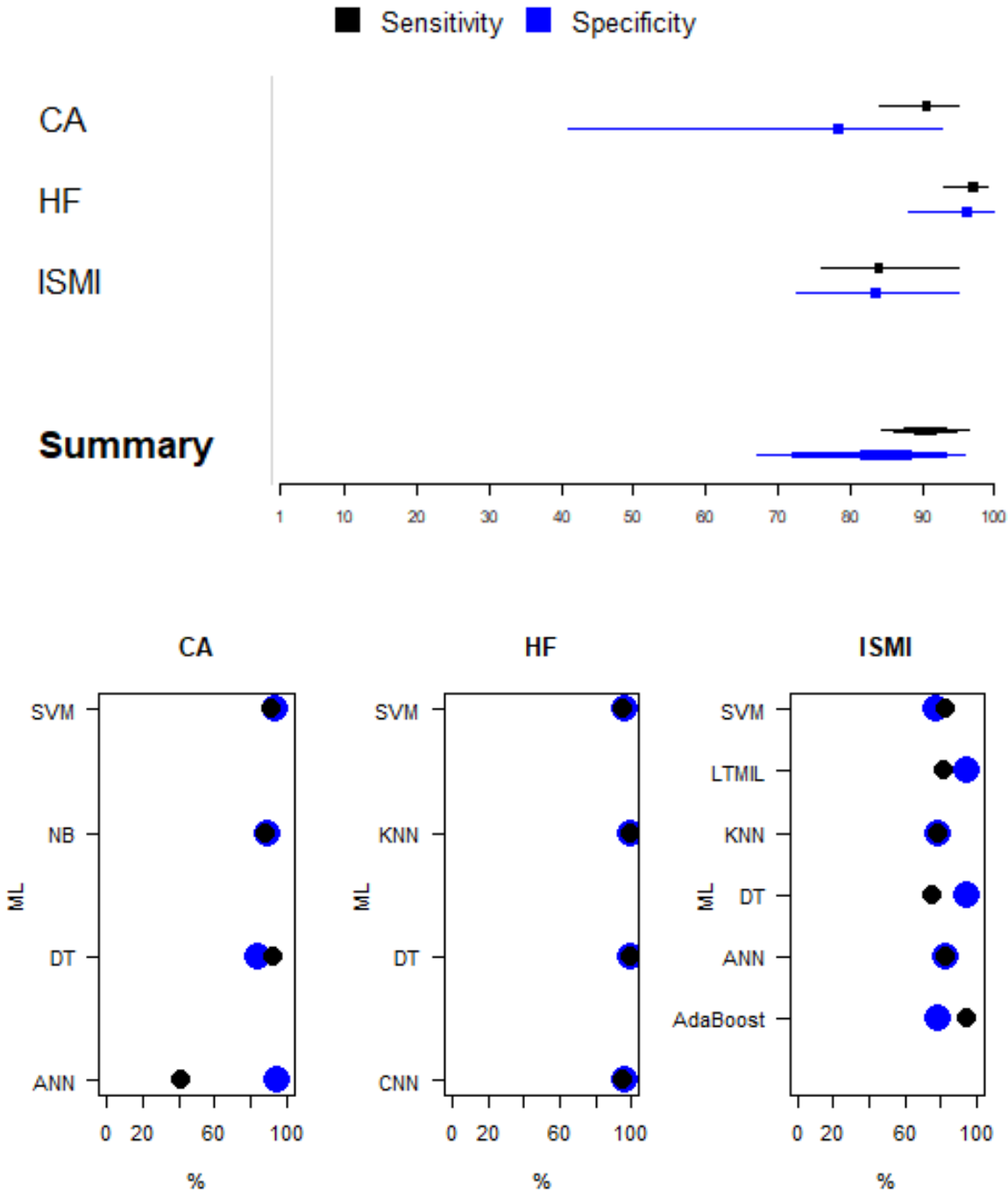


Fig. 2.20 The performance of ML techniques to detect the non-arrythmia group according to sensitivity and specificity.

As shown in Figures 2.19 and 2.20 and according to the five tribes taxonomy, analogizer algorithms such as SVM and KNN obtained the highest performance in three classification problems including: 1) cardiac arrhythmias, 2) atrial fibrillation, 3) tachycardia and 4) premature ventricular contraction. In the non-AR group, analogizer algorithms also obtained the best performance in two out of three classification problems including: 1) coronary atherosclerosis and 2) heart failure.

In terms of ML evaluation metrics, Accuracy was the most used evaluation metric in 49.2% of studies followed by sensitivity (20.3%) and specificity (17.1%) since 1995 (Figure 2.21). The frequency of using specificity and sensitivity might be expected to be the same, but this was not the case. Because other metrics such as precision and positive detection rate were used instead of specificity to complement sensitivity. Despite the obvious problem of the 'accuracy paradox', the use of accuracy increased sharply after 2013. Around 99 articles out of 757 articles (13.01%) used only accuracy to evaluate the ML performance. In the 99 articles, authors backgrounds were from computing (56/99), electrical and electronic engineering (32/99) and biomedical engineering (11/99) (Figure 2.22). Using accuracy only started increasing significantly from 2008 because accuracy was used to compare between different ML algorithms that were trained and tested on the same dataset. Since 2016, the use of other metrics including 1) precision (or positive predictive value (PPV)), 2) AUC, 3) F1 score, 4) sensitivity and 5) specificity started increasing. As a limitation, the unsupervised ML metrics were not included in this review as we focused more on the supervised ML algorithms.

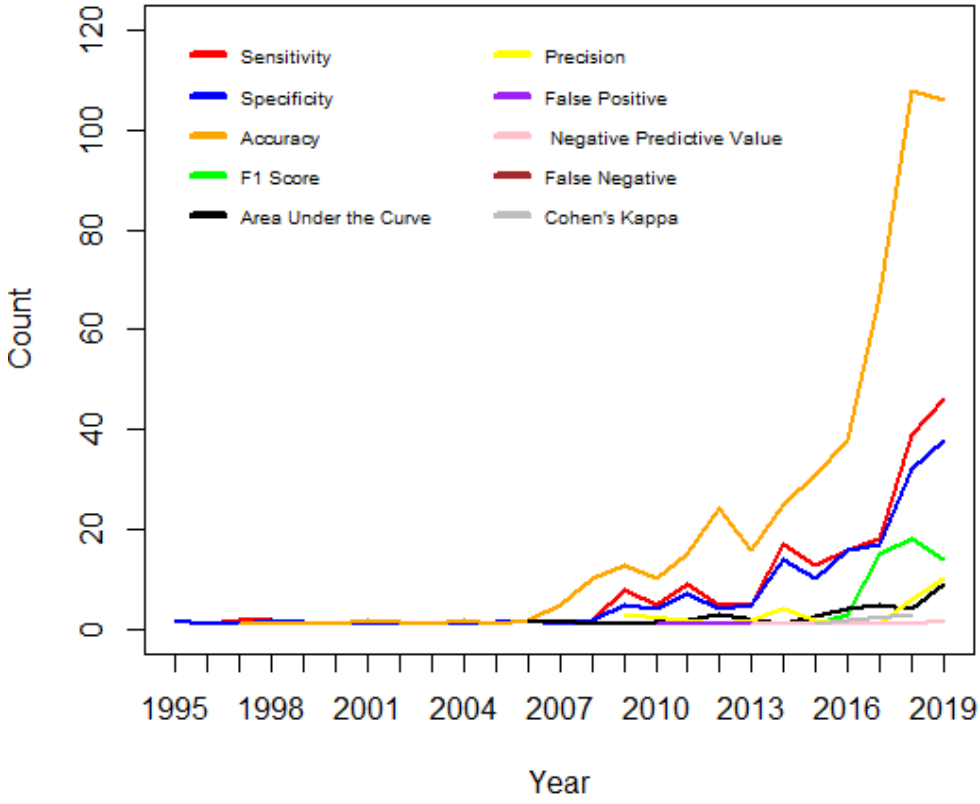


Fig. 2.21 The most frequent used metrics in ECG classification since 1995.

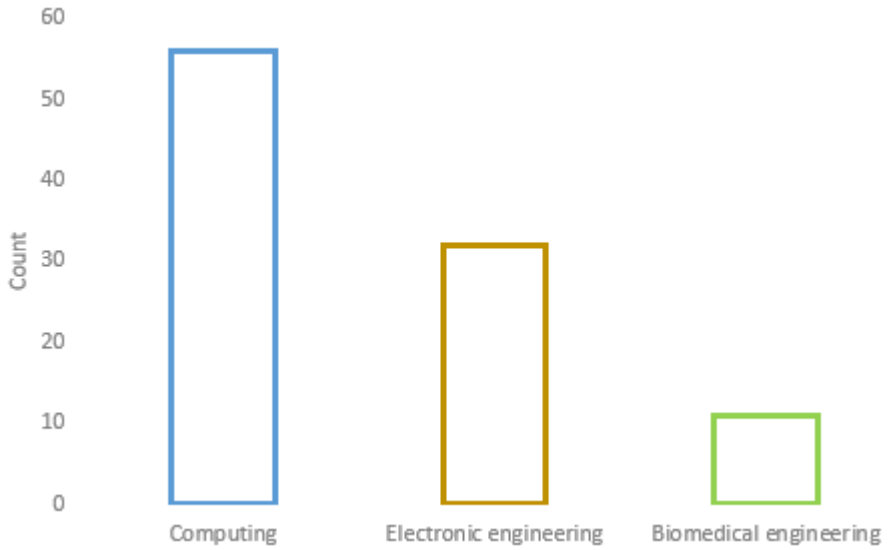


Fig. 2.22 The frequency of using accuracy as a metric in three different departments.

Around 50% of hospitals around the world plan to invest in AI as they realised that AI could improve the healthcare system performance and patient outcomes (DAIC, 2019). However, in healthcare only 50% of decision makers are familiar with AI (DAIC, 2019). Hence, there are a number of challenges before AI can be used in clinical practice. Some of those challenges depend on AI algorithms itself such as transparency and explainabilty. While other challenges depend on users of AI such as familiarity with AI, adopting AI, automation bias and reporting AI algorithms results (Davenport et al., 2019; Gerke et al., 2020). Hence, the Consolidated Standards of Reporting Trials (CONSORT) and Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) are working on an international unanimity to address those challenges (The CONSORT-AI and SPIRIT-AI Steering Group, 2019).

According to previous studies from different disciplines including cognitive science, philosophy and psychology to interpret "how people define, evaluate and present explanations?", they found that people use some certain cognitive biases and social expectations in the explanation of their answers (Miller et al., 2017). Human being brain tends to select a certain reason out of infinite number of reasons to explain the final decisions. Also, people do not ask why event "X" happened, but they ask why event "X" happened instead of event "y". Hence, these types of findings from different disciplines could be used to develop explainable AI systems. Miller et al. found that search which do not use social science to build explainable AI framework could lead to failure in AI explainabilty (Miller et al., 2017). Because AI explainabilty is also related to a human–agent interaction and not just related to AI itself. According to the human-interaction community, human–agent interaction is defined as the intersection of artificial intelligence, human–computer interaction and social science (Figure 2.23). Hence, developing a trustworthy AI system must follow eight principles (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems) including 1) preserve well-being, 2) preserve human rights, 4) transparency, 3) data agency, 4)

effectiveness, 6) accountability, 7) competence, 8) awareness of misuse as shown in table 2.4.

Table 2.4 The principles of developing a trustworthy AI system.

Principle	Explanation
human rights	AI companies, individuals and research institutions must take into account that AI shall be operated to protect and respect human rights.
well-being	AI system creators should consider prioritising and increasing human well-being as it is a primary success criterion of AI. Because AI systems (even the high-quality AI systems) might have negative consequences on society's mental health, ability to achieve goals and sense of themselves.
data agency	AI creators shall enable people to access and share their data securely, which makes them able to control over their identity.
effectiveness	AI operators and creators shall define a benchmark to measure the effectiveness of their AI systems. because AI systems will be trusted if and only if it shows effective performance. Furthermore, AI systems should not cause harm which can prevent AI adoption. Hence, associations such as ISO and IEEE should develop standard metrics to measure the effectiveness of AI systems.
transparency	transparency is a key concern in AI developing process and it must be transparent for each stakeholder in each different level; to know how and why the AI made this particular decision? especially in medical diagnosis systems as it has a real consequences to human safety. Transparency also addresses the explainabilty and interpretability as they both augment AI transparency and lack of them results less transparency, and as a consequence it increases the difficulty of accountability. However, in terms of AI, interpretability is different from explainabilty regarding to user experience level (an accident investigator, developer or data scientist). Interpretability make AI understandable for every users of all levels. Meanwhile, explainabilty makes AI understandable for data scientists to understand "why a model behaves that way?". Hence, they can explain it to all users of different experience levels.
accountability	We need clarity about the deployment of AI systems to establish accountability and responsibility to avoid potential harms. Hence, AI systems must be transparent to make it accountable.
awareness of misuse	AI might have risks coming from misuse of AI systems such as AI hacking (such as "Microsoft Tay" AI chatbot), misuse of personal data or system manipulation. Hence, a new kind of education about the risk of AI misuse is required.
competence	AI Operators should be able to understand why and how AI made the decision and the impact of the decision. Also, they should know when the AI decision should be checked and overruled. Hence, AI creators should ensure the experience of the operators of their AI system.

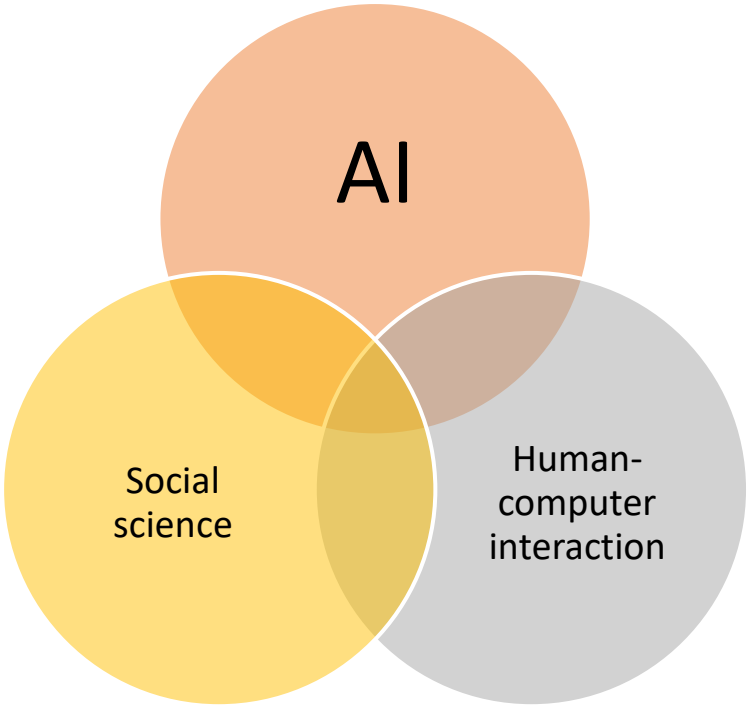


Fig. 2.23 Explainable artificial intelligence scope.

Human-computer interaction society encourages design practitioners for interdisciplinary collaboration and working on user-centered approaches to build a sufficient explainable AI (XAI) to cover the gap between XAI and use (Vera et al., 2020). Hence, a question bank as shown in table 2.5 was built by designers from different jobs based on their feedback and questions to answer the question " how to build a comprehensive XAI model?" ((Vera et al., 2020).

Table 2.5: Question bank design.

Question bank
What is the type of the input data?
What is the type of the output?
How accurate are the final predictions?
What is the overall logic in the final model?
How and why this input gives this specific output?
How and why are the other classes not predicted?
What the AI model will predicted if it is given different input?
How the given input should be changed to predict different output?
What is the range of changes that are permitted to get the same output?
How to improve the final model?

Furthermore, four major techniques were developed for making ML explainable including 1) perturbation, 2) backward propagation, 3) activation optimization and 4) proxy. In the perturbation, a noise is generated on the input features separately and then we observe the impact of that noise on the performance, and then the impact is stored in an importance score list. Local interpretable model-agnostic explanations (LIME) is an example of the perturbation technique and it uses a linear model (Marco et al., 2016).

In the backward propagation technique, importance scores are generated for the features and then the backward propagation method starts computing the contribution of each neuron in the next layer. It propagates down to the input layer. This technique has several methods (Scott et al., 2017; Marco et al., 2016; Grégoire et al., 2004; Zilke et al., 2016; Vera et al, 2020) including 1) layer-wise relevance propagation (LRP), 2) DeepLIFT, 3) smoothGrad, 4) integrated gradients (IG), 5) guided backprop (GB). This technique is used to explain DL networks.

The activation optimisation technique is used to explain how hidden layers work in deep neural networks by searching for a pattern in the input feature that maximise or minimise hidden layers contribution according to the target. Hence, it optimises input features instead of hyperparameters.

The proxy technique converts the deep neural networks to a more explainable ML algorithm such as a decision tree by creating a decision tree that mimics deep neural network structure. Hence, it maps the complex ML model to a more explainable ML model. However, this technique is only applicable with shallow ANNS.

In clinical practice, the ECG algorithms do not tell the physicians which features were used to make the diagnostic decision and how certain the algorithm is. Hence, researchers at Ulster University developed an interactive progressive-based interpretation (IPI) system which is an explainable decision support system. The IPI was developed using a differential diagnoses algorithm (DDA) which enables the IPI to suggest more than one diagnosis and shows the decision logic to improve ECG interpretation (Cairns et al., 2017). The IPI+DDA showed improvement in the interpretation and diagnosis accuracy. ECG Attention/heat map was suggested to be included in the ECG algorithms for more clarity and to solve the explainabilty issues, especially DL algorithms.

In DL, attention maps are generated from the last layer as it represents high-level features to show which input contributed more to the final prediction (Qiang et al., 2020). Those attention maps could also suggest features to the physicians that have never been used before in diagnosis. However, a combination of ML algorithm, visual explanation and a human perform better than the human and ML algorithm separately. Because this combination leads to a higher level of self-confidence and reduces automation bise in automated ECG interpretation as it has an influence on the physician's diagnosis. Based on HCI community suggestions, explanation

mechanism, question bank, AI principles and user interface guidelines, a suggested XAI model could be designed for ECG interpretation to show how the interface should look like.

Automation bias is one of the challenges in automated ECG interpretation and it occurs when physicians are biased to the diagnosis suggested by the ML algorithm (even when the ML decision is incorrect) as it influences their diagnoses, especially, for those who are not expert interpreters (Bond et al., 2018). Knoery et al found that providing probability, uncertainty index or decision logic for each prediction (Figure 21) could reduce the influence of bias on the physicians (Knoery et al., 2019).

Making AI transparent and explainable could solve the AI adoption problem as it shows how the final decision was made. However, if hospitals become more AI reliant, medicine schools will have to include some AI modules in the education system to make future physicians who can work and understand how AI work.

2.5 Lead misplacement:

Electrode misplacement affects the morphology of the ECG which can affect the ECG diagnosis. Furthermore, 12-lead ECG has a limited specificity (70–95%) and sensitivity (30–70%) to detect some heart abnormalities (Finlay et al., 2007), which can look normal in most cases for several reasons such as electrode misplacement. Also, misplaced electrodes provide incorrect ECGs which can result in significant false diagnoses (17% -34% of patients) made by computer-based algorithms or physicians, which can exhibit false-positive diagnoses of ventricular hypertrophy, anterior infarction, Brugada syndrome, ischemia (Bond et al., 2012). Moreover, it could conceal the ECG features or mimic another cardiac disease (Rudiger et al., 2003; Chanarin et al., 1990; Jowett et al., 2005), which leads to incorrect treatment which could harm the patient (O'Connor et al., 2010). Hence, placing electrodes accurately during ECG recording is important as it is linked to the integrity and reliability of the signals that are used by physicians to provide an accurate diagnosis and treatment (Walsh, 2018; Kania et al., 2014; Alioui et al., 2017). Hence, further investigation is needed to study the effect of electrode misplacement on the diagnostic specificity and sensitivity.

There are two types of errors in lead misplacement including 1) electrode interchange (e.g. the left leg electrode is placed on the right leg and vice versa) 2) chest electrode misplacement (e.g. chest electrodes are misplaced in the wrong intercostal space) (Figure 2.24).

Limb electrode interchange affects several morphological features on the ECG including P-wave, T-wave and QRS complex amplitude (Figure 2.25, Figure 2.26). However, the interchange frequency was just 0.3% in RL (Right Leg)-RA (Right Arm) (31/11,423) and the RL (Right Leg)-LA (Left Arm) interchange frequency was lower

compared to RL-RA. Those findings emphasise that ECGs are predominately free from limb electrode interchange (Finlay et al., 2010; Richard et al., 2017).

Chest electrode misplacement occurs from 40% to 60% of the time and 36% of precordial electrodes (V1, V2, V3, V4, V5, V6) are placed high or wide from the correct position (Bupp et al., 1997). V2 was the most sensitive electrode to misplacement followed by V3, V1, and V4 respectively, while there were no visible changes in V5 and V6 (Bupp et al., 1997).

The most common error in electrode misplacement is placing electrodes V1 and V2 too high and wide of their correct position (in the fourth ICS) and as a consequence it could lead to misplace the other electrodes (V3 to V6) (Figure 23) (Rajaganeshan et al.,).

Misplacement direction and distance from the correct position affect the morphology in different degrees. Noticeable changes in the ECG morphology were found for misplacement beyond 2 cm, while it has a trivial effect when it reaches up 1 cm from correct positions in any direction. 64% of chest electrodes were placed within 1.25 inches, while 27% were placed within 0.625 inches. V1 and V2 are the most common misplaced electrodes and 50% of the time they are misplaced within a radius of 0.625 inches which indicates that V1 and V2 are commonly placed wide and high of their correct position. While V4 and V6 are commonly misplaced lower and wider of their correct position in 30%-50% of the time (Wenger al., 1996). V2 misplacement has an effect on the morphology of the QRS complex more than the ST-T-U segment (Lateef et al., 2003) and vertically misplaced V1 and V2 could result in misdiagnosis of Brugada (Lateef et al., 2003).

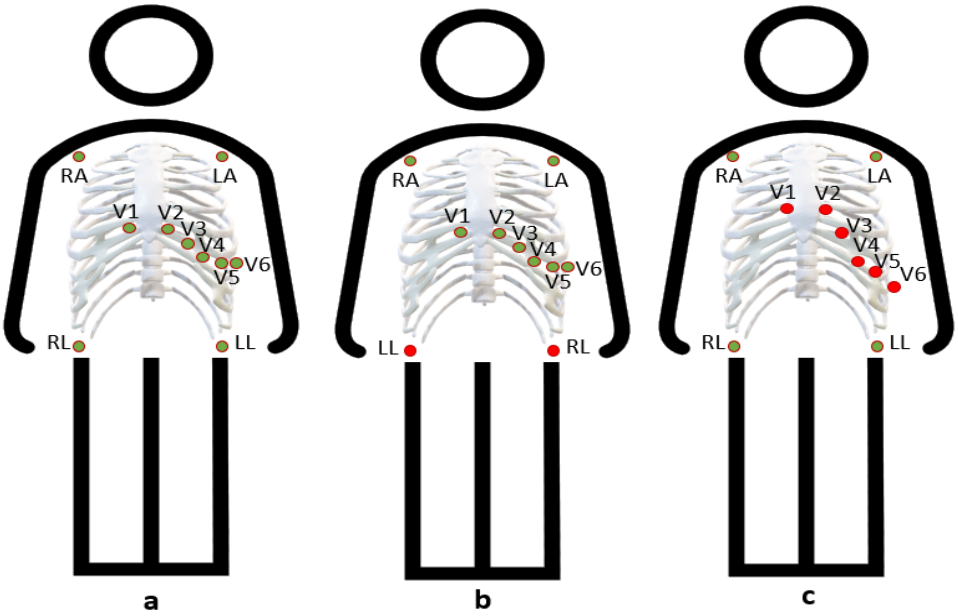


Fig. 2.24: Electrode placement. a represents correct placement. b: represents limb lead interchange. c: represents chest electrode misplacement.

In horizontal displacement, V1 and V3 misplacement has an effect on the automatic ECG interpretation in cases of LVH and MI and the worst effect happens when V1 and V2 are Horizontally shifted toward each other (Wenger al., 1996).

The most common misplacement is placing V1 and V2 in the second ICS instead of the fourth ICS. Studies found that V2 and V4 and V1 are the most affected electrodes respectively which was partially agreed by another study that found V2, V3 and V1 are the most affected electrodes respectively (Bond et al., 2012; Bond et al., 2016). Less than 20% of cardiologist and 50% of nurses can correctly place V1 and V2 in the correct position (Khunti, 2014).

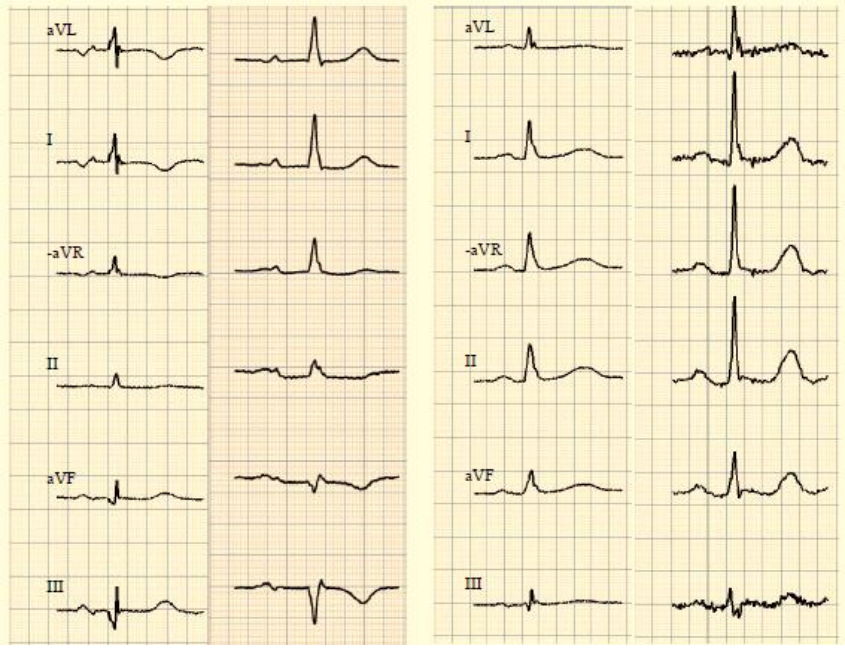


Fig. 2.25: Limb leads misplacement.

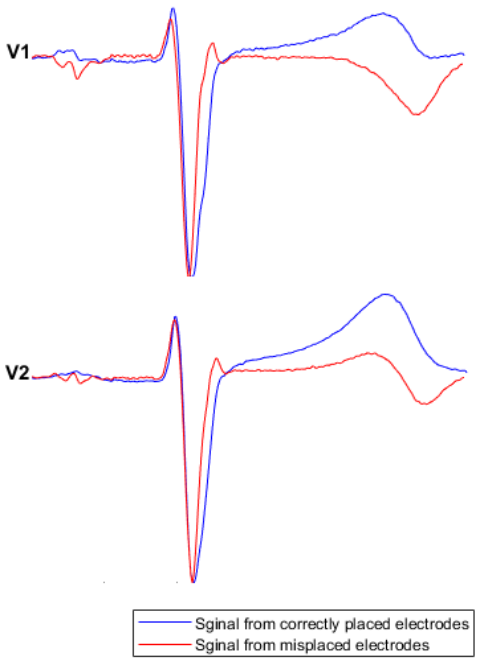


Fig. 2.26: recorded signal for vertical misplacement.

A previous study found that V1 and V2 have been shifted upward more than 1.5 cm in 50% of cases, while V4 and V6 have been shifted downward in 30% to 50% of cases (Schijvenaars et al., 2008). Furthermore, they found that misplacement of V1 can cause false QRS and ST-segment changes which might cause amisdiagnosis of LVH, BBB and myocardial infarction (Rauen et al., 2008).

30 technicians were recruited to place precordial electrodes from V1 to V6 on real patients (3 men and 5 women). 64% of precordial electrodes (for V6 and V4) were placed within a radius of 1.25 inch and 27% of them were placed within a radius of 0.625 inch (Figure 2.27). In 50% of cases, V1 and V2 were placed more than 0.625 inch of their correct position (Schijvenaars et al., 2008).

Several methods were suggested and used to control facilitate proper electrode placement such as, 1) placing stickers on the torso, 2) placing electrodes on real cases, 3) developing a human paradigm to place electrodes (Bupp et al., 1997) and 4) a sliding ruler (Bond et al., 2016). Unfortunately, those methods were not accepted by physicians, even the simplest method. Moreover, the first three methods were not sufficient for placing electrodes accurately (nurses accuracy was 18%, 25% and 9% respectively) and the electrode misplacement rate was 40%-60%. While the fourth method has not been widely adopted due to an increase in cost (Bupp et al., 1997).

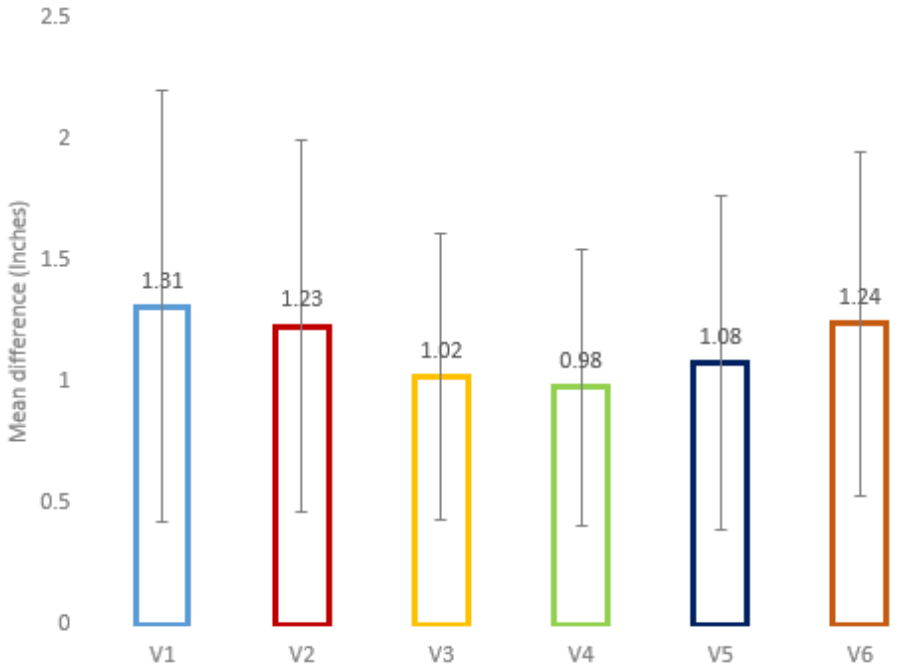


Fig. 2.27: Lead misplacement distances.

A previous study (Rjoob et al., 2020) reviewed the performance of ML algorithm to detect electrodes misplacement and interchange and which features were used to detect electrode misplacement using systematic reviews and meta-analyses (PRISMA) guidelines.

Since 1990s, only 14 articles were published in terms of using ML to detect electrode misplacement or interchanges. 12 articles studied the performance of ML to detect electrode interchanges, while 2 articles focused on electrode misplacement. The included articles are divided between five different regions (Figure 2.28).

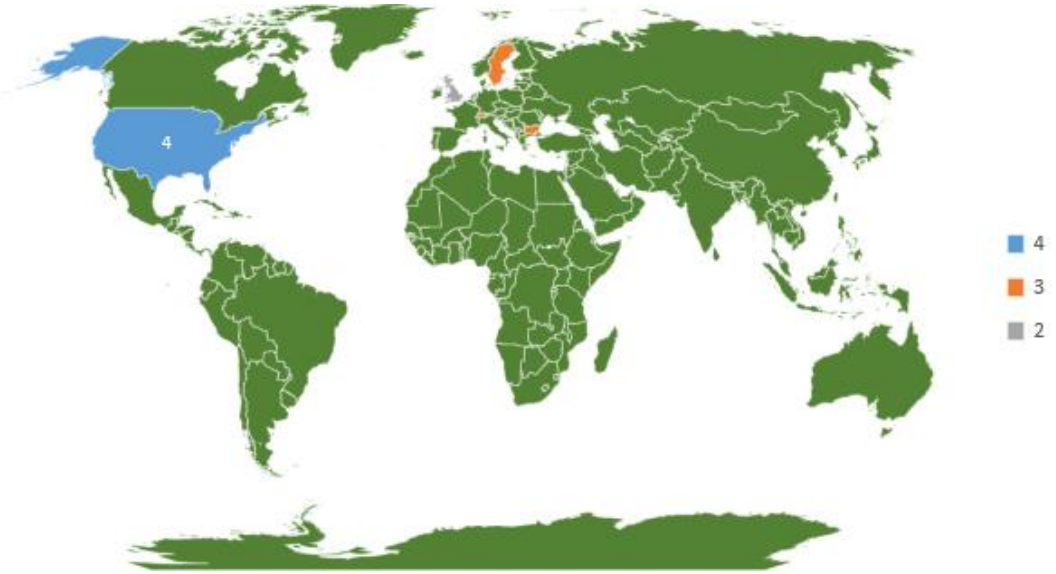


Fig. 2.28: Articles distribution. The USA (n=4), Sweden (n=3), Switzerland and Bulgaria (n=3), UK (n=2) and Netherlands (n=2).

Seven articles focused on both chest and limb electrodes interchanges, while four studies focused on limb electrode interchanges only. And two articles focused on chest electrodes vertical misplacement and one article focused on chest electrode interchange only. Four ML algorithms were applied to detect electrode misplacement/interchange including ANNs, DT, SVM and logistic regression, while correlation and amplitude threshold were also used in other studies to detect lead misplacement (Figure 2.29).

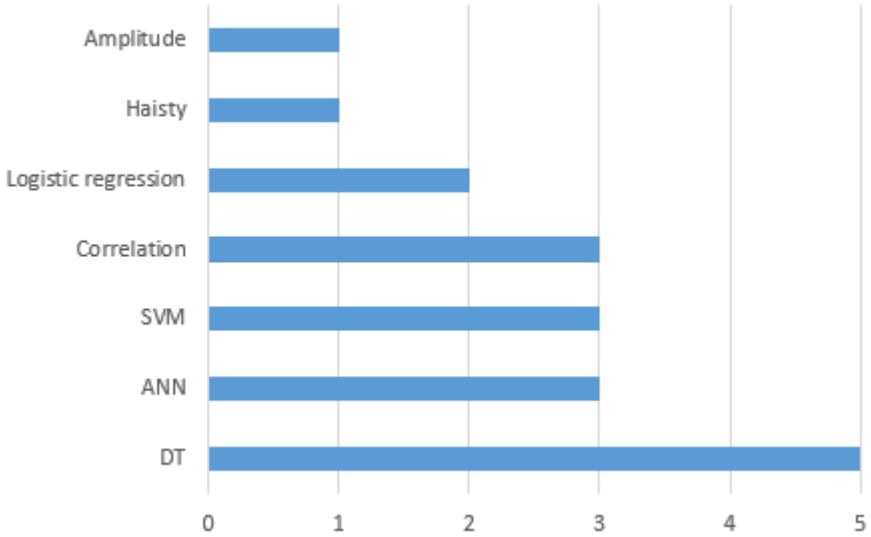


Figure 2.29: ML distribution.

All included articles were categorised into three different groups including 1) chest electrode misplacement, 2) chest electrode interchanges and 3) limb electrode interchanges (lead reversal). In chest electrode interchanges, the specificity varied from 91% to 100% and sensitivity varied from 44.5% to 99.95%. While in limb electrode interchanges, the specificity varying from 95.6% to 100% and the sensitivity varied from 20% to 99.3%. Table 2.6 shows mean specificity and mean sensitivity of each category. In terms of feature extraction, 12 articles used time domain feature, while two articles used mixed features including time domain features, time-frequency features and statistical features. Table 2.7 shows the summary of each included article including ML classifiers, features and ML performance. Specificity and sensitivity were the most used metrics to evaluate the performance of ML algorithms. Sensitivity was calculated to evaluate the performance of ML in all articles, while specificity was not mentioned in two articles.

Table 2.6: mean sensitivity and specificity.

	Study	*Mean Se*	*SD*	*Mean Sp*	*SD*	*F1*
A	(Rjoob et al, 2019)	79.6%	±8.6	84.6%	±5.9	0.816
	(Rjoob et al, 2019)	81.5%	±11.5	81.0%	±11.0	0.813
		Average =80.5%	**SD=±0.95**	**Average =82.8%**	**SD=±1.8**	
B	(Jekova et al., 2013)	97.9%	±0.31	99.1%	±0.19	0.985
	(Kors et al., 2000)	95.2%	±4.25	99.8%	±0.21	0.974
	(Kors et al., 2001)	93.2%	±6.4	99.8%	±0.11	0.964
	(de Bie et al., 2014)	88.7%	±8.80	99.8%	±0.08	0.939
	(Hedén et al., 1996)	71.6%	±14.01	99.91%	±0.03	0.938
	(Jekova et al., 2016)	95.3%	±0.35	92.2%	±1.25	0.938
	(Jekova et al., 2011)	87.0%	±0.00	97.8%	±0.00	0.920
	(Han et al., 2014)	56.1%	±27.7	99.9%	±0.04	0.718
		Average =85.6%	**SD=±13.5**	**Average =98.5%**	**SD=±2.5**	
C	(Jekova et al., 2013)	97.4%	±1.88	99.2%	±0.35	0.983
	(Hedén et al., 1995)	95.0%	±0.00	99.95%	±0.00	0.974
	(Jekova et al., 2011)	96.8%	±0.00	97.8%	±0.00	0.973
	(Han et al., 2014)	93.4%	±1.05	99.9%	±0.05	0.965
	(Richard et al., 2017)	88.1%	±3.90	99.7%	±0.20	0.935
	(Hedén et al., 1995)	83.7%	±21.31	98.5%	±2.07	0.904
	(Kors et al., 2001)	81.5%	±31.8	99.8%	±0.18	0.897
	(Han et al., 2014)	82.5%	±9.25	97.7%	±0.20	0.893
	(Heden et al., 1995)	69.0%	±11.45	99.9%	±0.01	0.816
	(Hedén et al., 1996)	57.6%	±0.00	99.97%	±0.00	0.731
	(de Bie et al., 2014)	54.6%	±28.09	99.6%	±0.15	0.705
		Average =81.7%	**SD=±14.5**	**Average =99.2%**	**SD=±0.82**	

A represents chest electrode vertical misplacement, B Represents chest electrode interchanges and C represents limb electrode interchanges.

TABLE 2.7: COMPARISON BETWEEN STUDIES.

REF	Dataset	Leads	Method	Feature	Sensitivity	Specificity
(HAN ET AL., 2014)	Body surface potential maps and Physionet PTB diagnostic ECG database	LA-RA and RA-LL using conventional and Mason Likar electrode placements	DT	P-wave frontal axis, P-wave clockwise vector loop rotation direction, QRS frontal axis, QRS clockwise vector loop rotation direction, R-wave amplitude and T-wave amplitude from lead I and lead II.	Conventional ECGs (LA-RA=91.3%, RA-LL=72.8%) and ML ECGs (LA-RA=88.9%, RA-LL=75.9%)	Conventional ECGs (LA-RA=97.9%, RA-LL=97.5%) and ML ECGs (LA-RA=96.5%, RA-LL=98.5%)
(RICHARD ET AL., 2017)	Adult 12-lead ECGs from a single teaching hospital	RA-RL and LA-RL	DT Haisty	Maximum and minimum QRS and T-wave voltages for ECG leads I, II and III.	RA-RL 84.2% LA-RL 86.6% RA-RL=93.5% LA-RL=N/A	RA-RL 99.9% LA-RL 99.9% RA-RL=99.4% LA-RL=N/A
(JEKOVA ET AL., 2013)	PTB diagnostic ECG database and the Common Standards for Electrocardiography (CSE) database	Precordial lead swaps over V1-V6	Correlation	QRS boundaries and QRS-T pattern (QRS onset to QRS offset + 3S0 ms)	95.7% (training) and 95% (testing)	93.5% (training) and 91% (testing)
(HEDEN ET AL., 1995)	ECGs recorded at the emergency department at the University hospital in Lund	LA-LL and RA-LL	ANN	ECG signal	LA-LL 57.6% RA-LL 80.5%	LA-LL 99.97% RA-LL 99.95%
(JEKOVA ET AL., 2011)	CinC Challenge 2011 dataset available from PhysioNet	Chest leads and for peripheral leads	Correlation	P-QRS-T amplitudes and polarities in I, II, III	96.8% for peripheral leads and 87% for chest leads	97.8% for peripheral and chest leads
(KORS ET AL., 2001)	The CSE database	RA-LA V1-V2 V5-V6 RA-LL V1-V3 V3-V4 LA-LL V2-V3 V4-V5 LL-RA-LA LA-LL-RA V4-V5-V6-V1-V2-V3 V6-V5-V4-V-3V2-V1	DT	Correlations between leads	Chest leads 93.2% and limb leads 81.5%	Chest leads 99.8% and limb leads 99.8%
(HAN ET AL., 2014)	Body surface potential maps and Physionet PTB diagnostic ECG database	LA–RA, RA–LL, V1–V2, V1–V3, V2–V3, V3–V4,	SVM	Including both morphology features and redundancy features	Precordial cable interchanges were 56.5% and limb cable interchange	Precordial cable interchange was 99.9%, and limb cable interchange (excluding left

		V4–V5, V4–V6, V5–V6, and no lead swap)			(excluding left arm-left leg interchange) were 93.8%	arm-left leg interchange) was 99.9%.
(JEKOVA ET AL., 2016)	Database from the Basel University Hospital, Physionet PTB diagnostic ECG database and the CSE diagnostic database	15 possible pairwise reversals in standard precordial leads	Correlation	Morphology features	Chest electrodes 97.9% and limb electrodes 97.4%	Chest electrodes 97.4% and limb electrodes 99.2%
	Table 6 continued					
(DE BIE ET AL., 2014)	Large (N18,000) hospital database for which serial ECG's were available and was based on simulated juxtaposition	RA–LA, V1–V2 LA–LL, V2–V3 RA–LL, V3–V4 LA–RL, V4–V5 RA–RL, V5–V6	Algorithm is based on QRS axis and P amplitudes for limb electrode reversals, and PQ-RS amplitude distances	P, Q, R, and S amplitudes	Chest electrodes 88.7% and limb electrodes 54.6%	Chest electrodes 99.8% and limb electrodes 99.6%
(KORS ET AL., 2000)	Database of1,220 standard 12-lead ECGs collected in the Common Standards for Quantitative ECG project	RA-LA, V1-V2, V3-V4, V5-V6 RA-LL, V1-V3, V4-V5 LA-LL, V2-V3, V4-V6	DT	Averaged representative beats	RA-LA 99.3%, V3-V4 99.3% RA-LL 98,3%, V4-V5 98.3% LA-LL 53.6%, V4-V6 87.2% V 1 -V2 98.3%, V5-V6 90.6% V1-V3 97.5%, V2-V3 95.6%	RA-LA 100%, V3-V4 100% RA-LL 100%, V4-V5 100% LA-LL 95.6%, V4-V6 99.5% V1-V2 100%, V5-V6 99.8% V1-V3 100%, V2-V3 99.5%
(HEDÉN ET AL., 1995)	The study was based on 11,432 ECGs	RA-LA	ANNs	P, QRS, and ST-T measurement	95%	99.95%
(HEDÉN ET AL., 1996)	11,432 ECGs, The emergency department at the University Hospital in Lund	LA-LL, V3-V4 V 1 -V2, V4-V5 V2-V3, V5-V6	ANNs	QRS amplitude and Area	LA-LL 57.6%, V3-V4 77.5% V1-V2 80.6%, V4-V5 83.0% V2-V3 44.5%, V5-V6 73.2%	LA-LL 99.97%, V3-V4 99.95% V1-V2 99.9%, V4-V5 99.95 % V2-V3 99.87%, V5-V6 99.88%
(RJOOB ET AL, 2019)	Body surface potential maps and Physionet PTB diagnostic ECG database	V1 and V2 vertical misplacement	SVM, DT and logistic regression	Morphological features (P, Q, R, S and T amplitudes), statistical features and WT	V1 89.8% V2 86.6%	V1 88.5% V2 86.6%
(RJOOB ET AL, 2019)	Body surface potential maps and Physionet PTB diagnostic ECG database	V1 and V2 vertical misplacement	SVM, DT, logistic regression and bagged decision tree	Morphological features (P, Q, R, S and T amplitudes), statistical features, correlation coefficients and WT	V1 and V2 93.3%	V1 and V2 92.0%

As shown in Figure 2.30, DT obtained specificity ranged from 86.6% to 100% and sensitivity ranged from 17.9% to 99.3% in five studies. While ANNs obtained specificity ranged from 99.8% to 99.9% and sensitivity ranged from 44.5% to 99.9%. In SVM, specificity ranged from 86.6% to 99.9% sensitivity ranged from 56.5% to 93.8%. Correlation obtained specificity equal 91.0% and sensitivity ranged from 87.0% to 97.8%. Using an amplitude threshold method obtained 99.8% specificity and sensitivity ranged from 20.0% to 90.0%. While Haisty obtained 99.9% specificity and 84.2% sensitivity.

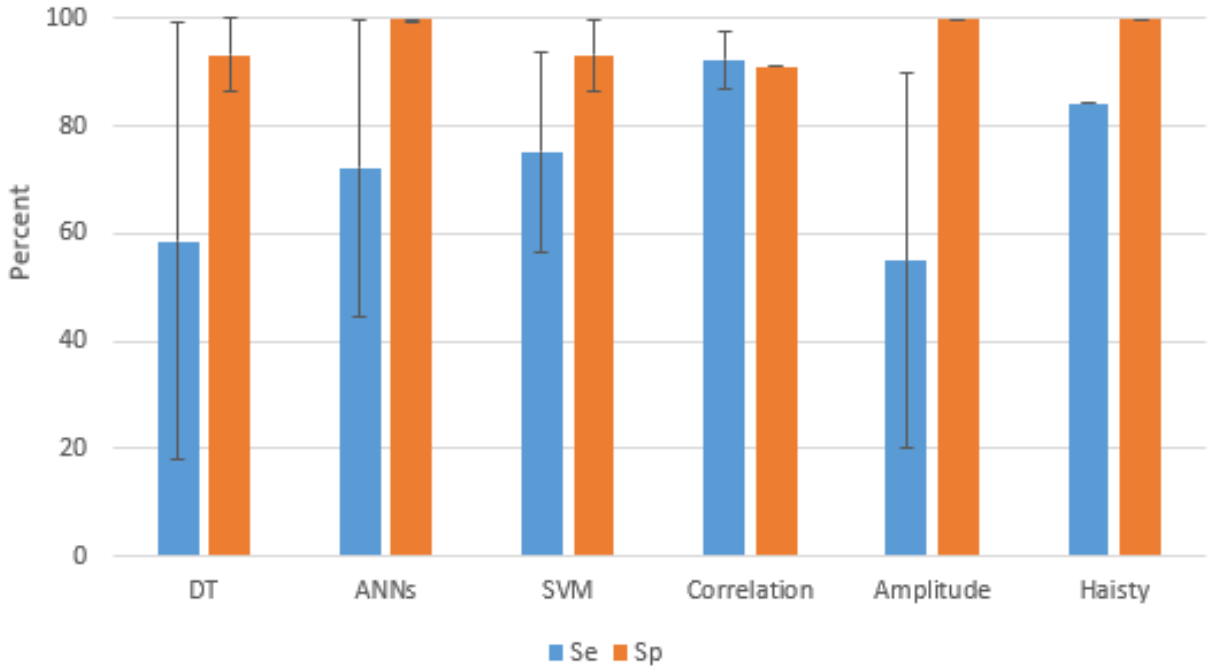


Fig. 2.30: ML sensitivity and specificity.

In each group in the three aforementioned groups (chest electrode misplacement, chest electrode interchanges and limb electrode interchanges) the ML performance was evaluated suing sensitivity, specificity and F1-score (Figures 2.31 and 2.32 respectively). In this review study we found that V1 and V2 misplacement beyond 2 cm change the morphology of P-wave and R-wave significantly which can result in misinterpretation of old anterior MI. Also, electrode V1 was found more sensitive to vertical misplacement than to horizontal misplacements which can result in false diagnoses. While there were no prominent changes when V3, V4 and V5 were misplaced up to 5 cm from their correct position (Walsh, 2018; Kania et al., 2014; Soliman, 2008).

LA-LL interchange was the most challenging interchange among the other interchanges. And ML obtained the lowest performance for detecting LA-LL interchange in terms of sensitivity, specificity and F1 score (Figures 2.31 and 2.32). While ML obtained a high performance for detecting precordial electrode misplacement and precordial electrodes misplacement.

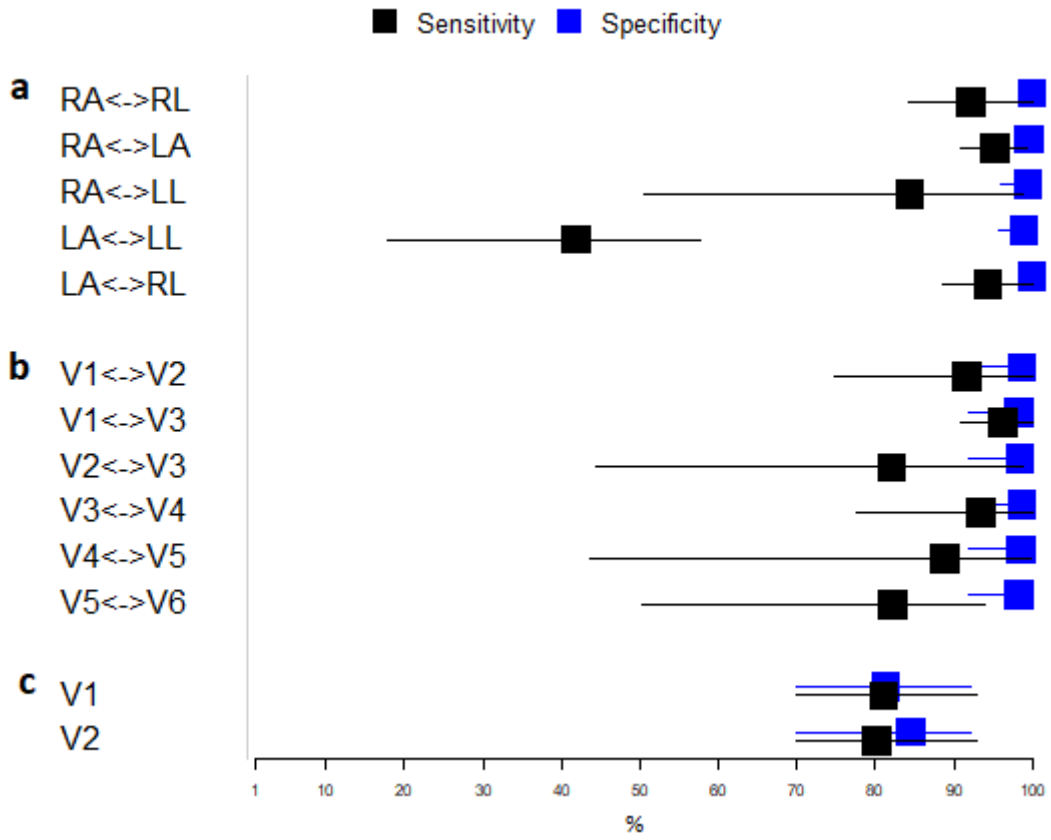


Fig. 2.31: mean sensitivity and specificity.

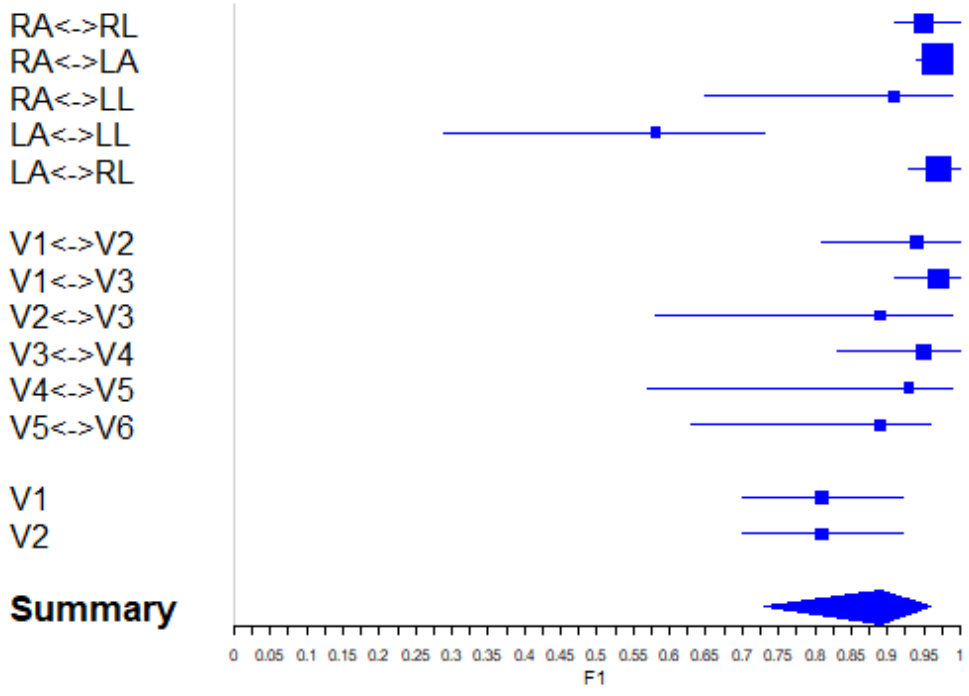


Fig. 2.32: F1-score.

The performance of each ML algorithms was analysed particularly in each misplacement/interchange scenario (Figure 2.33).

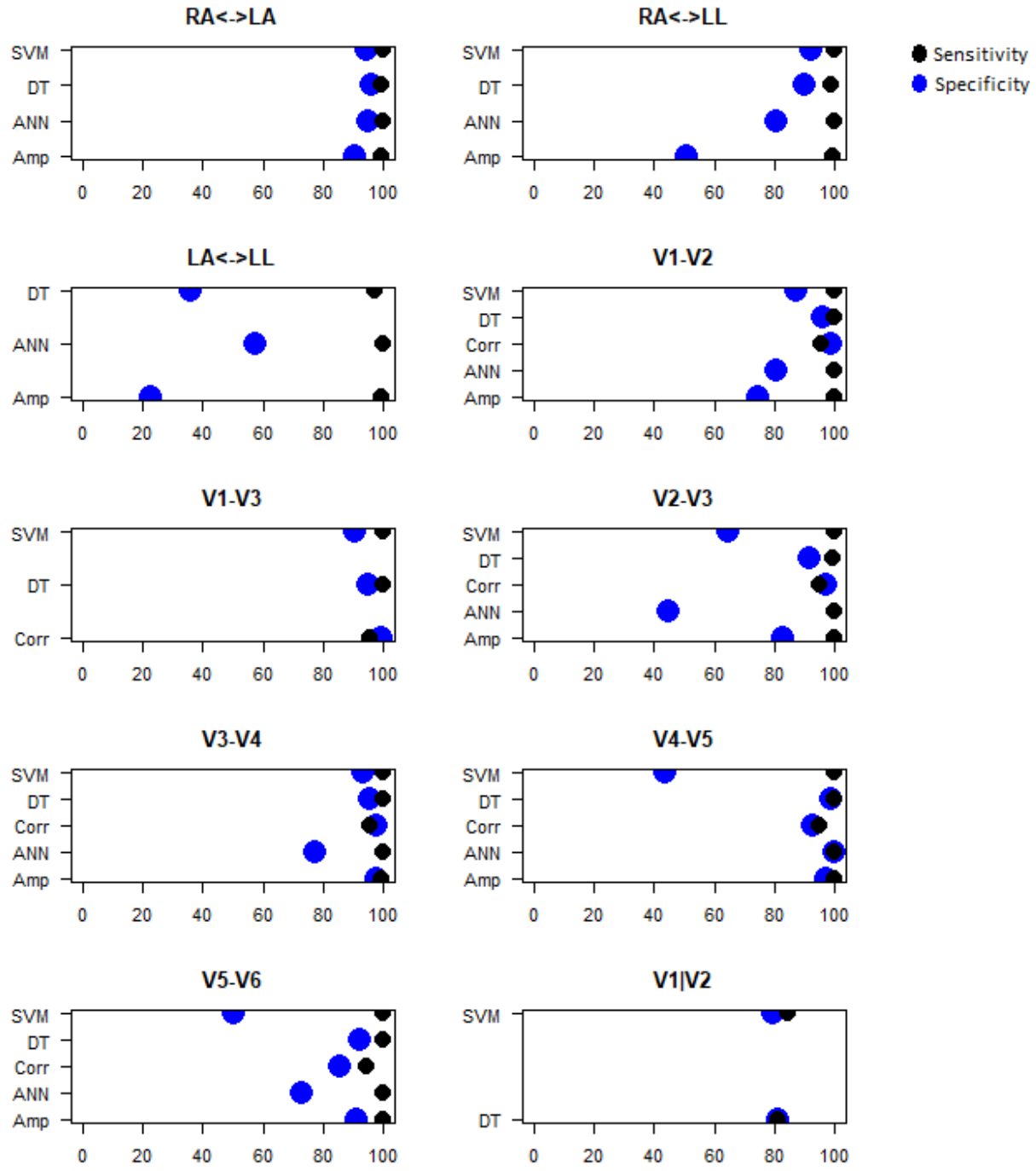


Fig. 2.33: ML performance in each interchange/misplacement scenario. Decision Tree (DT), artificial neural networks (ANNs), support vector machine (SVM), correlation (Corr), amplitude (Amp), left arm (LA), right arm (RA), left leg (LL), right leg (RL). V1IV2 represents vertical misplacement of chest electrodes, while V1-V2 represents chest electrodes interchange and RA<->LA represents limb electrode interchange (lead reversal).

As shown in figure 31, in term of ML, DT obtained the best performance in six scenarios (RA-LA, V1-V2, V1-V3, V2-V3, V3-V4 and V5-V6) out of ten scenarios of misplacement/interchange. While ANNs obtained the highest performance in three scenarios (RA-LA, LA-LL and V4-V5). And SVM reach a quite similar performance to DT and ANNs in four scenarios (RA-LA, RA-LL, V3-V4 and V1IV2). The three ML classifiers (DT, ANNs and SVM) obtained a high similar performance to detect the RA-LA scenario. However, those ML classifiers obtained a poor performance for detecting LA-LL. Hence, new ML algorithms such as DL need to be applied to improve the performance of the current technologies.

Companies started developing new ECG technologies to eliminate misplacement/interchange such as 3-D interactive Image system technology which is used to show the proper lead placement (DAIC, 2019). This technology assists clinicians to place electrodes properly and allows them to change electrodes position easily to provide a specific view to a part of the heart. However, DL has not been considered to detect electrode misplacement/interchange.

2.6 Summary:

CVD is leading the causes of premature death globally (WHO, 2017). CVD is diagnosed using traditional methods using 12-lead ECG and blood biomarkers (McGilligan et al., 2019). However, those methods are sometimes not accurate and take time to get the final diagnoses which cost healthcare systems billions of dollars around the world (British Heart foundation, 2021). For example, ECG is not accurate in some cases due to several reasons such as noise and lead misplacement (Rjoob et al, 2020).

Hence, the healthcare systems providers started looking for new technologies that could be involved in the decision-making process to improve the diagnoses accuracy and reduce the cost. One of these technologies is AI technology.

Previous studies showed that AI could be used to improve the ECG signal quality by detecting 12-lead ECG misplacement and as a consequence it could improve diagnoses (Rjoob et al, 2019). Furthermore, previous studies suggested AI to be used to find novel biomarkers that could detect CVD quickly and accurately (McGilligan et al., 2019).

Chapter 3

Auto Detect Leads Misplacement Using Traditional Machine Learning

3.1 Overview

Lead misplacement is when electrodes are incorrectly positioned when recording the 12-lead ECG. And as a consequence, misplacing electrodes can affect the morphology of the ECG and could mimic or conceal different heart diseases, which can change patient diagnoses. Electrode misplacement can take place in the limb electrodes (LA, RA, LL and RL) or in the precordial electrodes (V1, V2, V3, V4, V5 and V6). It was found that the precordial electrodes are misplaced in 40% to 60% of the time, specially, for V1 and V2. Additionally, less than 50% of nurses and 20% of cardiologists can place V1 and V2 in the right position (Khunti, 2014), which can affect diagnoses in the cardiac care unit.

In this chapter we used traditional machine learning algorithms to detect the misplacement of V1 and V2 misplacement. This chapter involves two experiments to detect V1 and V2 misplacement. In the first experiment, the misplacement of V1 and V2 was detected separately for each lead by training and testing the model on V1 and V2 features separately. While in the second experiment, the misplacement of V1 and V2 was detected by combining V1 and V2 features in the same training and testing data.

3.2 The First Experiment

3.2.1 Material and Methods

Data Collection:

The dataset was obtained from PHYSIONET and it is known as the Kornreich dataset. The dataset was recorded in the 1980s by Fred Kornreich. It involves three different types of patients including 151 myocardial infarction (MI), 151 left ventricular hypotrophy and 151 normal subjects. All MI patients had ischemic-type cardiac pain, a history of prolonged and changes in enzyme levels. The diagnosis was substantiated by ventriculography, coronary angiography, nuclear angiograms or echocardiography during which multigated, technetium99m-labeled blood pool imaging was performed.

Dalhousie torso with 117-lead body surface potential map (BSPM) electrodes and 352 nodes were used to extract the signal for each particular lead as shown in Figure 3.1. To provide a high resolution, the 117 electrodes in each BSPM were multiplied with a transformation matrix to get 352 nodes.

Precordial electrodes V1 and V2 were extracted from the 352 nodes (refer to Figure 3.1) as follow 1) node169 represents the correct position of V1 and 2) node171 represents the correct position of V2. The green color represent the correct position of the precordial leads (V1 to V6). While the blue color represents the wrong position. As

leads V1 and V2 are commonly misplaced vertically (high and wide of their correct position), nodes 126, 83 and 43 were considered to represent V1 misplacement in the third intercostal space (ICS), the second ICS and the first ICS respectively. While nodes 128, 85 and 43 were considered to represent V2 misplacement in the third intercostal space (ICS), the second ICS and the first ICS respectively.

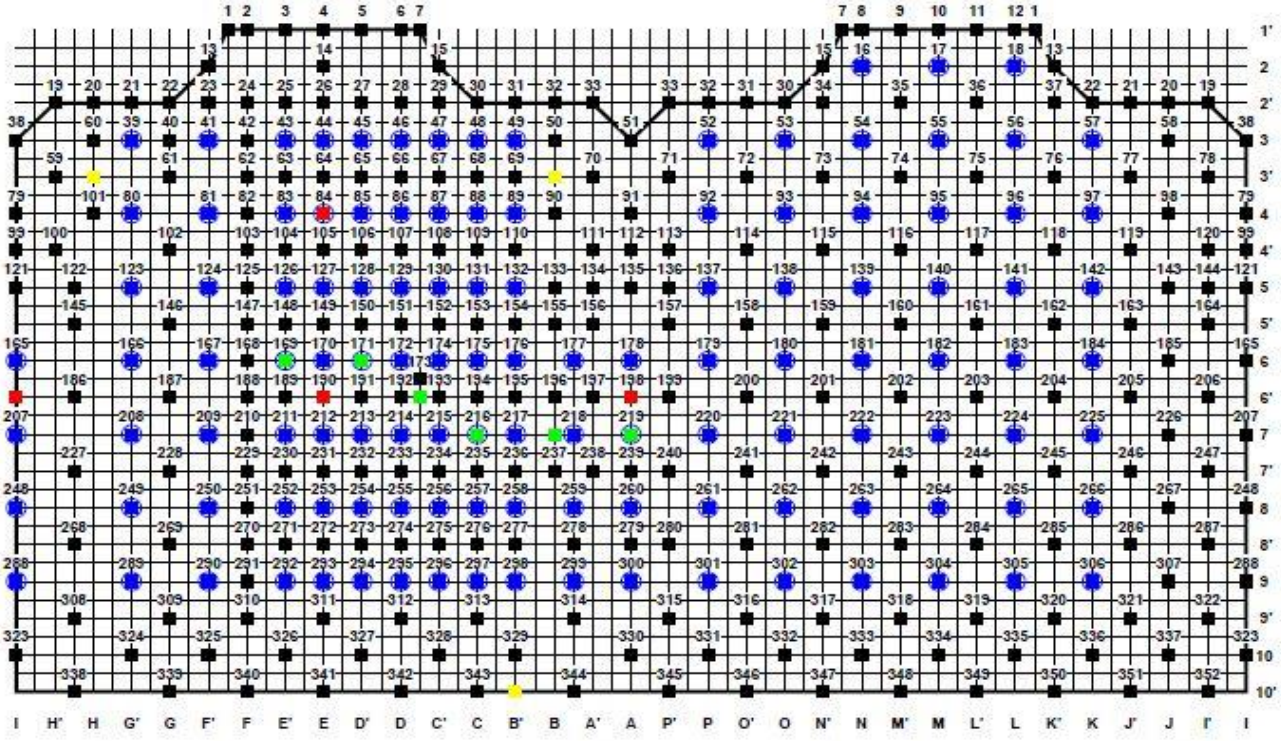


Fig. 3.1 Dalhousie torso with 352 nodes.

Feature Extraction:

ECG morphological features of V1 and V2 were extracted from three different domains as shown in Table 3.1 including 1) time domain, 2) time-frequency domain and 3) statistical domain. In the time domain, the P-wave amplitude, R-wave amplitude, S-wave amplitude, PR interval, QRS onset value and offset of the QRS were considered as features. In the time-frequency domain, discrete wavelet transform (DWT) using symlets wavelet mother function and 4 levels was applied. Then, four detailed coefficients (D1, D2, D3, D4) were retrieved. The first maximum three values of D3 were considered as features. In the statistical domain, skewness, kurtosis, mean, and standard deviation of the ECG signal were considered as features.

Feature Selection:

A feature selection method was applied to find an optimal set of features that provide good classification results. The hybrid feature selection approach (which combines the filter method followed by the wrapper method) was applied. Firstly, five filter methods were applied including 1) maximum relevance minimum redundancy (MRMR) using Equation (1), 2) mutual information feature selection (MIFS) using Equation (2), 3) Entropy using equation (3), 4) joint mutual information (JMI) using Equation (1) and 5) Relief using equation (4) have been applied to rank the features from the most important feature to the least important. Secondly, the wrapper method

has been applied by applying the backword elimination algorithm to remove the least important features that do not improve the performance. Figure 3.2 shows the feature selection process.

$$f_t = \arg\max \mathrm{I}(x_i; y) - \left[\alpha \sum_{k=1}^{t-1} I(x_{fk}; x_i) - \beta \sum_{k=1}^{t-1} I(x_{fk}; x_i|y)\right] \quad (1)$$

Where (I) represents information gain, (x) represents features, (y) represents labels and (t) represents the number of features. While α and β are weights calculated as follow:

In JMI: $\alpha = \frac{1}{t-1}$ and $\beta = \frac{1}{t-1}$

In MIFS: $\alpha = 0$ and $\beta = 0$

In MRMR: $\alpha = \frac{1}{t-1}$ and $\beta = 0$

$$I(X; Y) = H(X) - H(X|Y) \quad (2)$$

Where (I) represents information gain, (H) represents entropy, (X) represents Features and (Y) represents labels.

$$H(X) = -\sum p(x) \log_2 p(x) \quad (3)$$

Where (x) represents features and (H) represents entropy.

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad (4)$$

(W) is the weighted vector initialised with zeros, (x) represents feature, nearHit represents the same feature from the closest instance in the same class and nearMiss represents the same feature from the closest instance in the different class.

MIFS maximises the mutual information between labels and features, while MRMR minimises the redundancy between features and maximises relevance between features and labels. JMI performs an extra process beyond MRMR to remove relevance between features given the labels.

Table 3.1 Feature extraction

Feature ID	Domain	Feature Description
1	Time domain	P wave amplitude
2	Time domain	PR interval
3	Time domain	QRS beginning value/ P wave amplitude
4	Time domain	P wave amplitude/ QRS beginning
5	Time domain	R wave/ end of QRS
6	Time domain	S wave amplitude
7	Time domain	End of QRS/ R wave
8	Statistical domain	Mean of ECG signal
9	Statistical domain	Variance of ECG signal
10	Statistical domain	Standard deviation of ECG signal
11	Statistical domain	Skewness of ECG signal
12	Statistical domain	Kurtosis of ECG signal
13	Time-Frequency domain	First maximum value of D3/Third maximum value of D3
14	Time-Frequency domain	First maximum value of D3/Second maximum value of D3
15	Time-Frequency domain	Second maximum value of D3/Third maximum value of D3
16	Time-Frequency domain	First maximum value of D3/ First maximum value of D3

Tables 3.2, 3.3 and 3.4 show the importance of each feature according to each feature selection algorithm in the filter method.

Table 3.2 Features importance in each feature selection method in the first ICS.

V1 1st Feature Selection Algorithms with Selected Features					V2 1st Feature Selection Algorithms with Selected Features				
JMI	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**
3	3	3	8	6	2	2	2	10	6
2	6	6	1	12	16	8	8	6	12
7	9	8	10	13	7	10	10	14	13
4	10	9	14	16	3	14	1	11	16
16	14	1	9	15	12	6	14	12	15
12	8	4	11	2	4	9	6	15	2
15	13	10	6	7	8	1	9	2	7
11	1	14	12	11	15	13	13	16	11
8	5	13	13	5	11	11	4	1	5
1	15	15	7	8	5	15	15	13	3
5	11	11	3	3	1	5	11	3	8
13	4	5	2	4	13	4	5	7	4
6	12	2	5	9	6	12	3	4	1
9	16	12	4	1	9	16	12	5	9
10	7	16	16	10	10	3	16	8	10
14	2	7	15	14	14	7	7	9	14

Each number in the table represents feature ID

Table 3.3 Features importance in each feature selection method in the second ICS.

V1 2nd Feature Selection Algorithms with Selected Features					V2 2nd Feature Selection Algorithms with Selected Features				
JMI	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**
4	4	4	6	6	2	2	2	10	6
2	6	6	10	12	16	6	6	9	12
7	9	1	1	13	12	9	8	8	13
3	10	8	9	16	3	10	1	6	16
16	14	9	8	15	7	14	9	14	15
12	1	10	11	2	4	8	10	15	2
15	8	3	16	7	15	1	14	12	7
11	13	14	7	11	11	13	13	2	5
1	5	13	13	5	8	11	4	11	11
8	15	15	5	8	5	15	11	3	3
5	11	5	4	3	1	5	15	16	4
13	12	11	3	4	13	4	5	1	8
6	16	2	14	9	6	12	12	5	1
9	3	12	2	1	9	16	3	7	9
10	7	16	15	10	10	3	16	4	10
14	2	17	12	14	14	7	7	13	14

Each number in the table represents feature ID

Table 3.4 Features importance in each feature selection method in the third ICS.

V1 3rd Feature Selection Algorithms with Selected Features					V2 3rd Feature Selection Algorithms with Selected Features				
JMI	MIFS	MRMR	RELIEF	ENTROPY	JMI	MIFS	MRMR	RELIEF	ENTROPY
4	4	4	1	6	7	7	7	15	6
2	6	6	12	12	2	6	6	8	12
16	9	9	8	16	16	9	9	2	13
7	10	1	9	15	4	10	10	10	16
12	14	10	11	13	12	14	14	14	15
3	1	14	10	2	3	1	1	9	2
15	13	13	7	7	15	13	13	12	7
11	8	8	16	11	11	8	8	5	5
5	5	5	4	5	5	11	4	6	11
8	15	15	3	8	13	15	11	4	4
1	11	11	2	4	8	5	15	7	3
13	12	3	13	3	1	4	5	3	1
6	16	12	5	9	6	12	12	13	9
9	3	16	15	1	9	3	3	16	8
10	7	7	14	10	10	16	16	1	10
14	2	2	6	14	14	2	2	11	14

Classification:

V1 and V2 have been classified separately using five ML classifiers to detect misplacement including 1) fine decision tree (F-Tree), 2) Coarse decision tree (C-Tree), 3) Linear support vector machine (L-SVM) ,4) Quadratic support vector machine (Q-SVM) and 5) Logistic regression (LOGT). The hold-out validation method has been used to evaluate the models. Hence, 70% of the dataset were randomly selected for training and 30% for testing. The performance of the five ML classifiers was evaluated based on accuracy (Acc), sensitivity (Se) and specificity (Sp) using equations (4), (5) and (6) respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

Gini impurity has been used as a splitting criterion to split the decision tree into branches using equation 7.

$$fI_G(p) = 1 - \sum_{i=1}^{n} p_i^2 \quad (7)$$

(P) represents the probability of the class and (n) represents the number of classes

In L-SVM, the algorithm incorporates two parallel hyperplanes to separate the data into two different classes with an optimal margin between hyperplanes as shown in equations 8 and 9. The difference between L-SVM and Q-SVM is based on the kernel function that has been used to split the dataset.

$$w.x - b = 1 \quad (8)$$

$$w.x - b = -1 \quad (9)$$

where (w) represent weight corresponding to each feature, (x) features and (b) represent bias term. Cases above this hyperplane or on the hyperplane should be in class 1 and cases below this hyperplane or on the hyperplane should be in class 0.

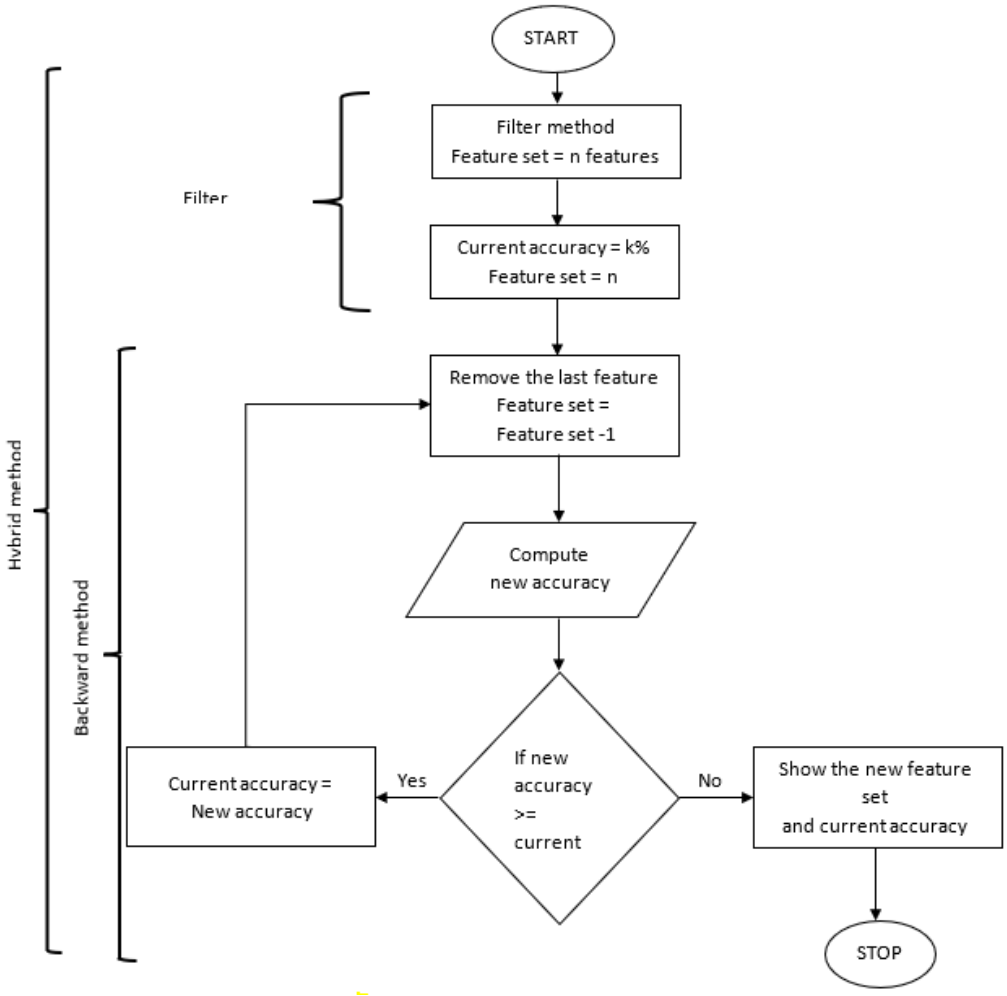


Fig.3.2 The hybrid feature selection process.

Logistic regression (logit) was used as it is a common statistical technique for binary classification. It uses log-odds (which represents a linear combination of model parameters and features) as computed in equation 10.

$$L=\alpha_0 + \alpha_1.x_1 + \cdots + \alpha_n.x_n \quad (10)$$

where (α_0) coefficients are the model parameters and (x) are the features.

Odds (o) computation represents the exponent that is used to compute odds using equation 11 and the corresponding probability has been computed using equation 12.

$$o = e^{\alpha_0+\alpha_1.x_1+\cdots+\alpha_n x_n} \quad (11)$$

$$p = \frac{1}{1+e^{\alpha_0+\alpha_1.x_1+\cdots+\alpha_n x_n}} \quad (12)$$

McNemar's test has been used to check if the performance is statistically significant. For the null hypothesis, we assume that the probabilities p(model1 incorrect and model2 correct) and p(model1correct and model 2 incorrect) are the same (none of the two models outperforms the other). While in the alternative hypothesis, we assume that the performances of the models are not equal. A contingency matrix was created from the predicted and actual labels for the models. Then the contingency matrix has been used to find the p-value using McNemar's test as follow:

p-value = mcnemar(ary=contingency matrix, corrected=True).

While in the case of small sample size, the exact p-value was calculated from the binomial distribution as follows:

p-value = mcnemar(ary=contingency matrix, exact=True)

3.2.2 Results

As shown in Table 3.5. 3.6 and 3.7, the best performance for V2 misplacement detection was 93.9%, 89.3%, 72.8% accuracy, 96.6%, 89.9%, 71.1% sensitivity and 91.2%, 88.5%, 74.5% specificity in the first, second and third ICS respectively. In V1, the best performance for V2 misplacement detection was 93.6%, 86.6% and 68.1% accuracy, 95.3%, 86.6%, 67.1% sensitivity and 91.9%, 86.6%, 69.1% specificity in the first, second and third ICS respectively. Tables 3.8, 3.9 and 3.10 and Figure 33. show the agreement between the feature selection methods using cohen's kappa statistics in each ICS. As shown in Figure 3.4, among the five ML classifiers, QSVM obtained the best performance for detecting V1 lead misplacement in the first and second ICS and outperformed the other classifier significantly (P<0.001), while LOGT obtained the best performance in the third ICS. In V2, QSVM obtained the best performance for detecting misplacement in the first and second and third ICS.

In terms of feature selection, MIFS, JMI and entropy provided the most relevant and important features that provided good classification results. And the minimum number of required features for detecting lead V1 and V2 misplacement in the first ICS was fifteen and thirteen features respectively. While in the second ICS, the minimum number of features was 16 features. And in the third ICS, the minimum number of features was 15 features.

Table 3.5 ML accuracy for V1 and V2 lead misplacement detection in the 1st ICS.

	V1 1st Intercostal Space					V2 1st Intercostal Space				
	JMI^{15f}	MIFS^{16f}	MRMR^{13f}	RELIEF^{16f}	ENTR^{15f}	JMI^{15f}	MIFS^{13f}	MRMR^{13f}	RELIEF^{16f}	ENTR^{15f}
FTREE	89.9%	89.9%	91.6%	87.2%	89.3%	89.3%	87.2%	87.9%	91.3%	90.9%
CTREE	88.3%	84.2%	85.9%	86.6%	89.3%	86.2%	87.2%	86.2%	88.3%	86.9%
LOGT	87.9%	84.9%	85.2%	83.9%	83.2%	91.3%	89.3%	87.6%	85.6%	85.9%
LSVM	87.6%	82.9%	82.9%	82.6%	83.6%	88.6%	88.6%	84.9%	86.9%	84.9%
QSVM	**93.9%**	90.3%	87.9%	88.9%	91.6%	**93.6%**	**93.6%**	89.3%	87.2%	92.3%

FTREE: fine tree, CTREE: coarse tree, LOGT: logistic regression, LSVM: linear support vector machine, QSVM: quadratic support vector machine, JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as MIFS^{13f} represent the best number of features to provide a good accuracy in feature selection algorithm.

Table 3.6 ML accuracy for V1 and V2 lead misplacement detection in the 2nd ICS.

	V1 2nd Intercostal Space					V2 2nd Intercostal space				
	JMI^{15f}	MIFS^{16f}	MRMR^{16f}	RELIEF^{15f}	ENTR^{16f}	JMI^{15f}	MIFS^{16f}	MRMR^{12f}	RELIEF^{16f}	ENTR^{15f}
FTREE	85.2%	88.6%	82.6%	84.9%	84.2%	78.2%	**86.6%**	84.6%	79.5%	83.6%
CTREE	83.6%	82.2%	81.9%	85.9%	86.2%	75.8%	79.9%	79.2%	79.5%	79.5%
LOGT	82.9%	82.9%	81.9%	82.6%	81.5%	76.2%	79.5%	79.5%	75.5%	81.9%
LSVM	84.9%	84.2%	83.6%	84.6%	82.6%	77.5%	80.9%	83.2%	76.2%	80.2%
QSVM	85.6%	86.6%	86.2%	86.9%	**89.3%**	80.9%	84.2%	81.9%	77.5%	84.2%

FTREE: fine tree, CTREE: coarse tree, LOGT: logistic regression, LSVM: linear support vector machine, QSVM: quadratic support vector machine, JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as MIFS^{13f} represent the best number of features to provide a good accuracy in feature selection algorithm.

Table 3.7 ML accuracy for V1 and V2 lead misplacement detection in the 3rd ICS.

	V1 3rd Intercostal Space					V2 3rd **I**ntercostal space				
	JMI^{15f}	MIFS^{15f}	MRMR^{16f}	RELIEF^{16f}	ENTR^{15f}	JMI^{15f}	MIFS$^{\mathbf{15f}}$	MRMR^{12f}	RELIEF^{15f}	ENTR^{15f}
FTREE	63.4%	60.1%	69.1%	65.8%	61.7%	58.7%	60.1%	59.1%	60.4%	57.7%
CTREE	63.4%	69.5%	68.5%	71.1%	68.5%	58.7%	59.1%	61.7%	63.1%	57.0%
LOGT	70.1%	**72.8%**	67.4%	67.8%	70.8%	67.1%	66.4%	64.8%	68.5%	62.1%
LSVM	70.8%	71.1%	66.8%	70.8%	70.1%	64.8%	64.4%	61.4%	63.1%	62.4%
QSVM	65.4%	68.1%	69.5%	68.5%	65.1%	61.1%	**68.1%**	65.1%	65.4%	61.1%

FTREE: fine tree, CTREE: coarse tree, LOGT: logistic regression, LSVM: linear support vector machine, QSVM: quadratic support vector machine, JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as MIFS^{13f} represent the best number of features to provide a good accuracy in feature selection algorithm.

Table 3.8 Agreement between feature selection methods in the first ICS.

	V1 1st Intercostal Space					V2 1st Intercostal Space				
	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.93	0.75	0.93	1	1	0.75	0.75	0.93	1
MIFS	0.93	1	0.81	1	0.93	0.75	1	0.87	0.81	0.75
MRMR	0.75	0.81	1	0.82	0.75	0.75	0.87	1	0.81	0.75
RELIEF	0.93	1	0.82	1	0.93	0.93	0.81	0.81	1	0.93
ENTR	1	0.93	0.75	0.93	1	1	0.75	0.75	0.93	1

JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as MIFS^{13f} represent the best number of features to provide a good accuracy in feature selection algorithm.

Table 3.9 Agreement between feature selection methods in the second ICS.

	V1 2nd Intercostal Space					V2 2nd Intercostal Space				
	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.93	0.93	0.87	0.93	1	0.93	0.68	0.93	1
MIFS	0.93	1	1	0.93	1	0.93	1	0.75	1	0.93
MRMR	0.93	1	1	0.93	1	0.68	0.75	1	0.75	0.68
RELIEF	0.87	0.93	0.93	1	0.93	0.93	1	0.75	1	0.93
ENTR	0.93	1	1	0.93	1	1	0.93	0.68	0.93	1

JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as $MIFS^{13f}$ represent the best number of features to provide a good accuracy in feature selection algorithm.

Table 3.10 Agreement between feature election methods in the third ICS.

	V1 3rd Intercostal Space					V2 3rd Intercostal Space				
	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.87	0.93	0.93	1	1	0.87	0.68	0.87	1
MIFS	0.87	1	0.93	0.93	0.87	0.87	1	0.81	0.87	0.87
MRMR	0.93	0.93	1	1	0.93	0.68	0.81	1	0.68	0.68
RELIEF	0.93	0.93	1	1	0.93	0.87	0.87	0.68	1	0.87
ENTR	1	0.87	0.93	0.93	1	1	0.87	0.68	0.87	1

JMI: joint mutual information, MIFS: mutual information feature selection, MRMR: maximum relevance minimum redundancy, ENTR: entropy. Superscripts such as $MIFS^{13f}$ represent the best number of features to provide a good accuracy in feature selection algorithm.

	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.93	0.75	0.93	1	1	0.75	0.75	0.93	1
MIFS	0.93	1	0.81	1	0.93	0.75	1	0.87	0.81	0.75
MRMR	0.75	0.81	1	0.82	0.75	0.75	0.87	1	0.81	0.75
RELIEF	0.93	1	0.82	1	0.93	0.93	0.81	0.81	1	0.93
ENTR	1	0.93	0.75	0.93	1	1	0.75	0.75	0.93	1

	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.93	0.93	0.87	0.93	1	0.93	0.68	0.93	1
MIFS	0.93	1	1	0.93	1	0.93	1	0.75	1	0.93
MRMR	0.93	1	1	0.93	1	0.68	0.75	1	0.75	0.68
RELIEF	0.87	0.93	0.93	1	0.93	0.93	1	0.75	1	0.93
ENTR	0.93	1	1	0.93	1	1	0.93	0.68	0.93	1

	JMI	MIFS	MRMR	RELIEF	ENTR	JMI	MIFS	MRMR	RELIEF	ENTR
JMI	1	0.87	0.93	0.93	1	1	0.87	0.68	0.87	1
MIFS	0.87	1	0.93	0.93	0.87	0.87	1	0.81	0.87	0.87
MRMR	0.93	0.93	1	1	0.93	0.68	0.81	1	0.68	0.68
RELIEF	0.93	0.93	1	1	0.93	0.87	0.87	0.68	1	0.87
ENTR	1	0.87	0.93	0.93	1	1	0.87	0.68	0.87	1

perfect agreement
substantial agreement
moderate agreement
fair agreement

Fig. 3.3 Agreement between feature selection methods.

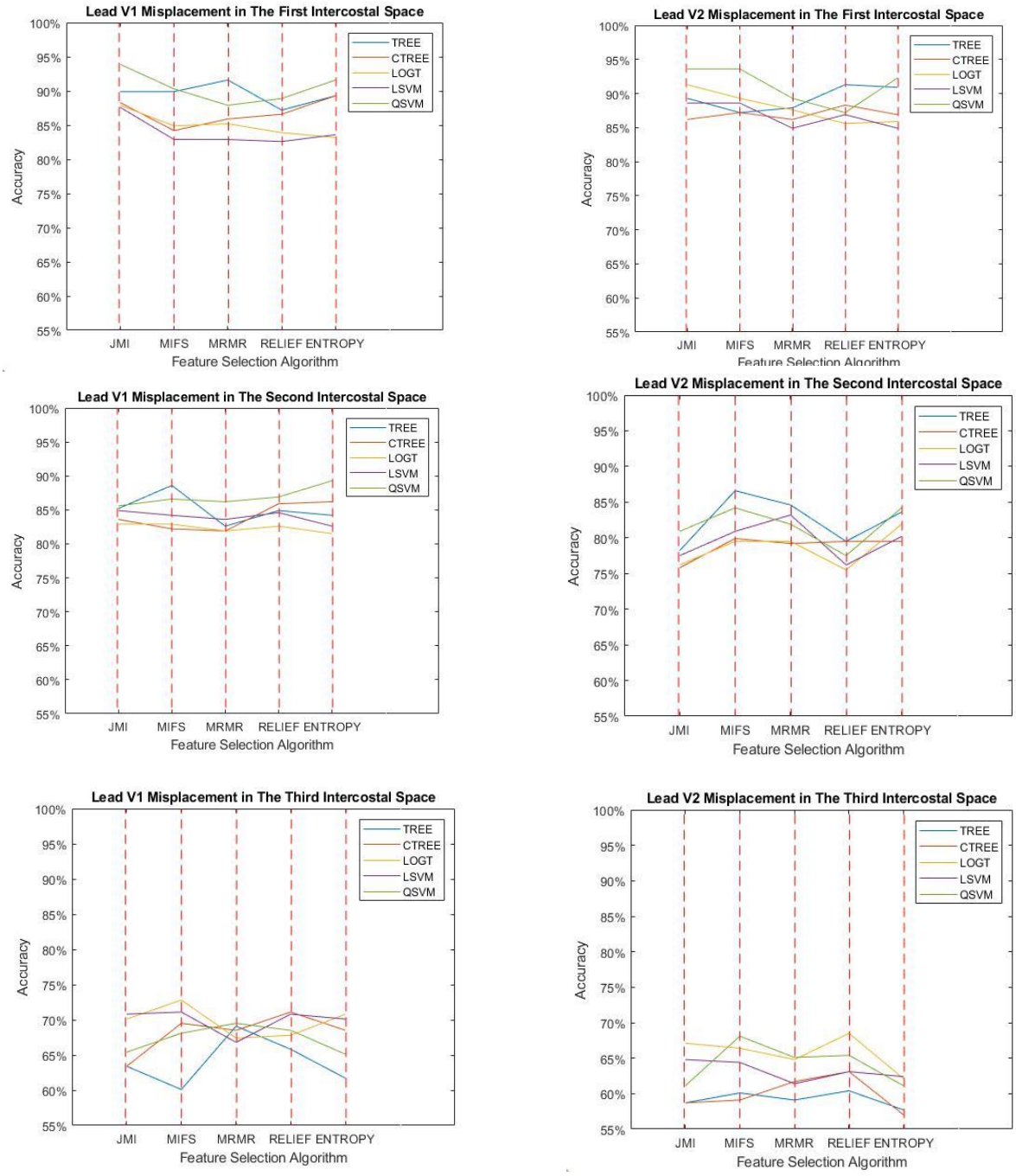


Fig.3.4 ML classifiers performance in each ICS for detecting V1 and V2 lead misplacement. Each subplot represents the performance in each combination between feature selection and ML classifiers in 1st, 2nd and 3rd ICS.

3.2.3 Discussion

In this study, chest lead misplacement was detected using a new method that achieved high performance, especially in the first and second ICS. However, there was a significant decline in the performance ($P<0.001$, McNemar's) when the misplacement occurs in the third ICS, which is theoretically expected as the

morphological features of the ECG looks similar to features obtained from the correct position (4th ICS). F-TREE and Q-SVM obtained the best performance in the first and second ICS using MIFS, JMI and entropy as a feature selection algorithm. In this study, the same morphological features in previous research were used but with new different classifiers such as QSVM and by including new features such as time-frequency features. According to the performance, the algorithm could be used to reduce the number of incorrectly recorded ECGs which might reduce the cost and time of performing a new ECG. Hence, physicians could get an accurate using correct recorded ECG. Further research in the future could be carried out to detect lead misplacement in the third ICS using new ML methods such as deep learning and new features could be investigated.

And as a limitation of this work, the hyperparameters were not tuned in this work as we used the default values of the hyperparameters. Furthermore, in some experiments, the McNemar's test was used with small discordant pairs, which is considered as a limitation of using the test, and as a consequence it might cause issues of interpretation due to low statistical power. Also, the Deep learning should be considered to detect lead misplacement as it showed promising performance in several applications. Alternatively, based on the ML results, medical device companies could deploy a ML algorithm to detect leads misplaced. Hence, the quality of recorded ECG could be iteratively improved. For further insights, the performance of the aforementioned ML could be compared with healthcare professionals.

3.3 The Second Experiment

3.3.1 Material and Methods

Data Collection:

The same dataset in the first experiment was used in the second experiment. Dalhousie torso in Figure 3.1 with 117-lead body surface potential map (BSPM) electrodes and 352 nodes were used to extract the signal for leads V1 and V2. Node169 represents V1 and node 171 represents V2. The green color represent the correct position of the precordial leads (V1 to V6). While the blue color represents the incorrect position of V1 and V2. Nodes 126, 83 and 43 were considered to represent V1 misplacement in the third intercostal space (ICS), the second ICS and the first ICS respectively. While nodes 128, 85 and 43 were considered to represent V2 misplacement in the third intercostal space (ICS), the second ICS and the first ICS respectively. To provide a class balance and to avoid bias, 50% of cases were incorrect ECGs and 50% were correct.

Feature Extraction:

A total of sixteen ECG features have been extracted in three different domains as shown in Table 3.11 including 1) time domain features (including P wave amplitude, QRS onset value, PR interval, R-wave peak amplitude, S wave amplitude, and offset of the QRS), 2) statistical domain features (including the skewness, kurtosis, mean, standard deviation of the ECG signal, the root mean square error (RMSE) between V1 and V2 electrodes and Pearson correlation coefficient (as given that V1 and V2 electrodes are commonly misplaced together), and 3) time-frequency features by

applying a discrete wavelet transform (DWT) using symlets wavelet mother function using four levels to obtain four detailed coefficients (D1, D2, D3, D4) and four approximation coefficients (A1, A2, A3, A4). Furthermore, the maximum, minimum and mean values of D4 were considered as features.

Table 3.11 Features domains and ID.

Feature ID	Domain	Feature Description
1	Time	P wave amplitude
2	Time	PR interval
3	Time	QRS beginning value
4	Time	R amplitude
5	Time	End of QRS value
6	Time	S wave amplitude
7	Statistical	Mean of ECG signal
8	Statistical	Variance of ECG
9	Statistical	Standard deviation of ECG signal
10	Statistical	Skewness of ECG
11	Statistical	Kurtosis of ECG signal
12	Time-Frequency domain	Maximum value of D4 (decomposition coefficient 4)
13	Time-Frequency domain	Minimum value of D4
14	Time-Frequency domain	Mean value of D4
15	Statistical	Correlation coefficient between V1 and V2 ECGs
16	Statistical	Root mean square error (RMSE) between V1 and V2 ECGs

Feature Selection:

Firstly, the hybrid approach feature selection algorithm was applied to rank the sixteen using different filter methods as shown in table 3.12 including, mutual information feature selection (MIFS), maximum relevance minimum redundancy (MRMR), joint mutual information JMI, Entropy and Relief. Secondly, the backwards elimination algorithm has been applied to find the best group of features as inputs to the ML classifier. The backwards algorithm started removing feature by feature until it the ML classifiers obtained the best performance.

Classification:

Six ML classifiers were used including 1) F-Tree, 2) C-Tree, 3) Bagged tree, 4) L-SVM, 5) Q-SVM and 5) LOGT. The difference between the three different types of the decision tree is based on the number of leaves and the splitting method. For example, C-TREE makes few branches to enable the tree to make coarse distinctions between classes. F-TREE enables the tree to make fine distinctions between classes. While bagged tree uses bootstrapping method with replacement to produce multiple training data and then it takes the majority outcome from multiple trees. The hold-out validation method has been used to validate the models (70% of the data were used for training and 30% for testing).

3.3.2 Results:

In feature selection process, feature 2 has the highest priority in JMI, MIFS and MRMR, while features 5 and 11 has the highest priority when using RELIEF and Entropy respectively in the second ICS. In the third ICS, feature 2 has the highest priority when using JMI, MIFS and MRMR, meanwhile features 6 and 11 has the highest priority when using RELIEF and Entropy respectively. Table 3.13 and figure 3.5 show the agreement between feature selection algorithms in each ICS.

Table 3.12 Feature ranking per feature selection technique.

Feature Selection Algorithms with Selected Features in 2nd ICS						**Feature Selection Algorithms with Selected Features in 3rd ICS**					
Feature Rank	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**	**Feature Rank**	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**
1	2	2	2	5	11	**1**	2	2	2	6	11
2	11	6	6	13	2	**2**	11	6	6	5	2
3	10	8	8	6	1	**3**	10	8	8	13	3
4	7	9	9	4	3	**4**	3	9	9	4	1
5	1	12	12	9	8	**5**	7	12	12	16	10
6	3	16	16	12	10	**6**	1	16	16	12	8
7	4	15	1	8	7	**7**	4	14	15	14	7
8	14	14	14	14	12	**8**	15	15	14	9	12
9	5	5	15	16	9	**9**	14	5	1	8	9
10	13	13	5	10	4	**10**	5	13	5	7	4
11	15	1	13	3	16	**11**	13	1	13	3	16
12	6	4	4	7	15	**12**	6	4	4	10	14
13	8	7	7	11	14	**13**	8	7	7	11	15
14	9	3	3	1	6	**14**	9	3	3	15	6
15	12	10	10	2	5	**15**	12	10	10	1	5
16	16	11	11	15	13	**16**	16	11	11	2	13

Table 3.13 Agreement between feature selection methods.

	Agreement in the 2nd ICS					**Agreement in the 3rd ICS**				
	ENTROPY	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**	**JMI**	**MIFS**	**MRMR**	**RELIEF**
Entropy	1	0.87	0.81	0.75	0.93	1	0.87	0.75	0.81	0.86
JMI	0.87	1	0.81	0.75	0.93	0.87	1	0.86	0.93	1
MIFS	0.81	0.81	1	0.93	0.87	0.75	0.86	1	0.93	0.86
MRMR	0.75	0.75	0.93	1	0.81	0.81	0.93	0.93	1	0.93
RELIEF	0.93	0.93	0.87	0.81	1	0.86	1	0.86	0.93	1

	1st ICS					2nd ICS				
	ENTROPY	**JMI**	**MIFS**	**MRMR**	**RELIEF**	**ENTROPY**	**JMI**	**MIFS**	**MRMR**	**RELIEF**
Entropy	1	0.87	0.81	0.75	0.93	1	0.87	0.75	0.81	0.86
JMI	0.87	1	0.81	0.75	0.93	0.87	1	0.86	0.93	1
MIFS	0.81	0.81	1	0.93	0.87	0.75	0.86	1	0.93	0.86
MRMR	0.75	0.75	0.93	1	0.81	0.81	0.93	0.93	1	0.93
RELIEF	0.93	0.93	0.87	0.81	1	0.86	1	0.86	0.93	1

perfect agreement
substantial agreement
fair agreement

Fig.3.5 Agreement between feature selection methods.

In terms of ML classifiers performance, the best ML classifier performance in the second ICS was bagged decision tree followed by C-Tree, F-Tree, LOGT, Q-SVM and L-SVM respectively. In the third ICS, the best ML classifier performance was also a bagged decision tree followed by C-Tree, LOGT, Q-SVM, L-SVM and F-Tree respectively. Table 3.14, Figure 3.6 and Figure 3.7 show the accuracy of the six ML classifier corresponding to each feature selection algorithm. The numeric appended to the label of each feature selection method represents the optimal number of features that was selected to achieve the best accuracy. Based on ML classifiers performance, MRMR and JMI were the best feature selection algorithms for detecting misplacement in the second ICS and JMI and RELIEF for the third ICS. Figure 3.8 shows area under the curve (AUC) for the bagged tree algorithm in the second and third ICS.

Table 3.14 ML classifiers performance for detecting V1 and V2 misplacement.

Classification accuracy using ML										
	Accuracy in 2nd ICS					Accuracy in 3rd ICS				
	ENTROPY15	**JMI15**	**MIFS14**	**MRMR13**	**RELIEF16**	**ENTROPY14**	**JMI16**	**MIFS14**	**MRMR15**	**RELIEF16**
FTREE	85.0%	85.0%	85.0%	82.7%	84.0%	60.0%	59.0%	60.3%	58.3%	59.0%
CTREE	87.7%	87.7%	87.7%	85.7%	87.7%	69.7%	69.7%	69.7%	69.7%	69.7%
LOG	82.7%	82.7%	83.7%	81.3%	82.7%	64.3%	63.7%	65.7%	63.7%	63.7%
SVM	78.7%	78.7%	75.7%	76.0%	78.7%	59.0%	60.3%	61.0%	61.3%	60.3%
QSVM	79.0%	79.0%	78.3%	79.0%	79.0%	58.7%	60.0%	62.7%	60.3%	60.0%
BAGGED	88.3%	92.3%	90.3%	**92.7%**	90.7%	69.3%	**70.0%**	66.3%	69.0%	**70.0%**

Superscripts such as MIFS^{13f} represent the best number of features to provide a good accuracy in feature selection algorithm

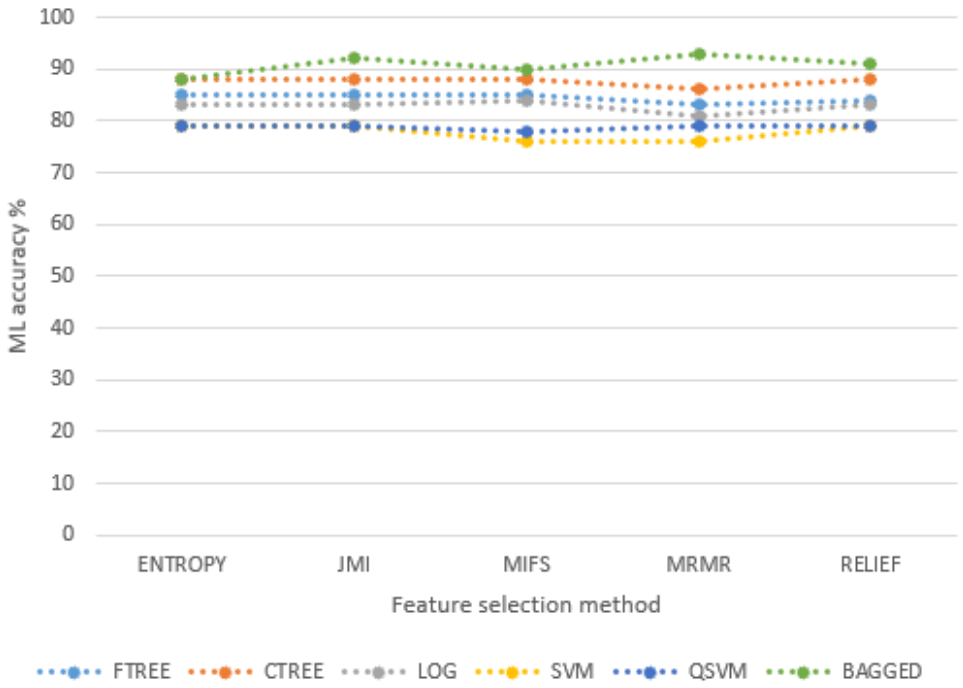


Fig.3.6 ML classifiers performance corresponding to each feature selection algorithm in the 2nd ICS

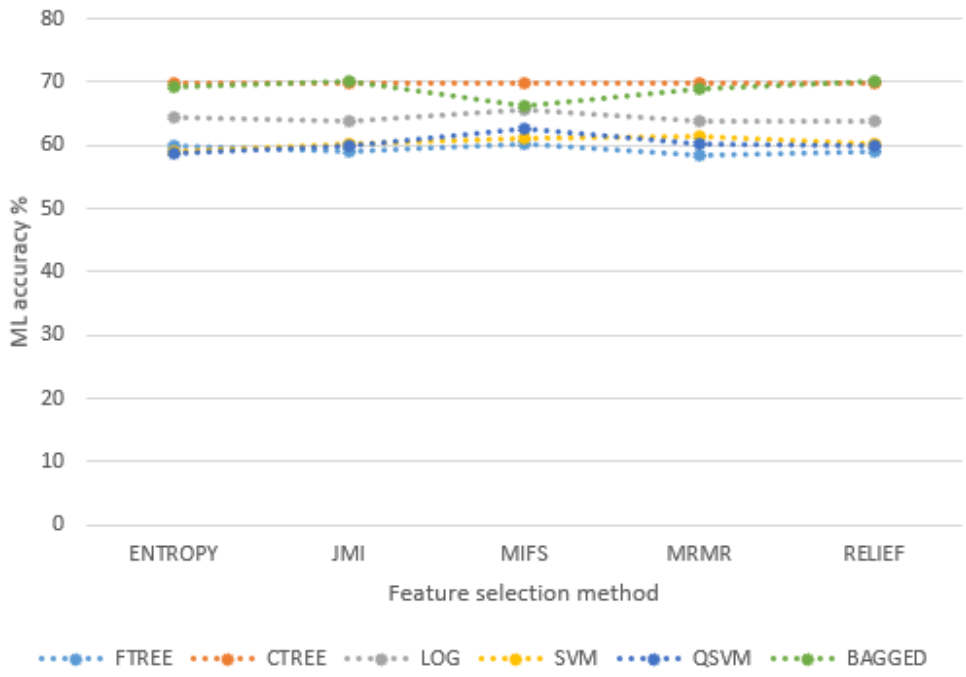
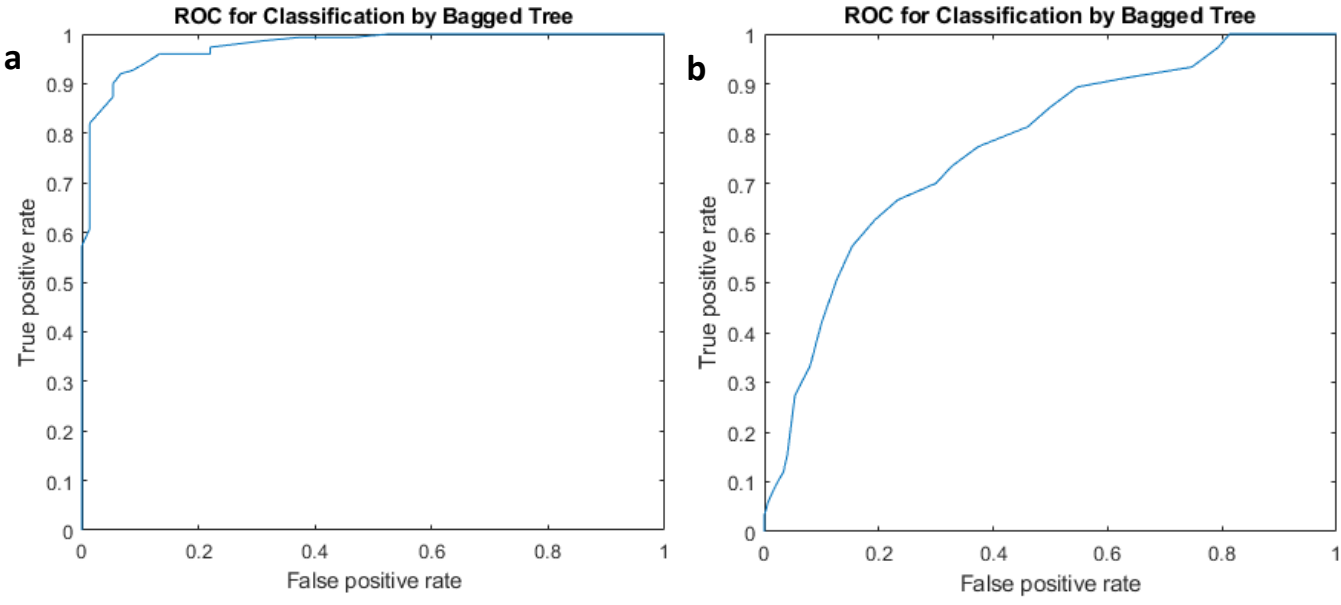


Fig.3.7 ML classifiers performance corresponding to each feature selection algorithm in the 2nd ICS.



Prevalence: 50.0%
Sensitivity: 93.3%
Specificity:92.0%, Accuracy:92.7%
Predictivity of positive test: 92.1%
Predictivity of negative test 93.2% F1-score: 0.9

Prevalence: 50.0%
Sensitivity: 70.0%
Specificity:70.0%, Accuracy: 70.0%
Predictivity of positive test: 70.0%
Predictivity of negative test:70.0% F1-score: 0.7

Fig.3.8. ROC curve for bagged tree in each ICS. a represents performance in the second ICS and b represents the performance in the third ICS.

3.3.3 Discussion:

In this study, lead V1 and V2 misplacement was detected using a new method that achieved high performance, especially in the second ICS. However, there was a significant decline in the performance ($P<0.001$, McNemar's) when the misplacement occurs in the third ICS as the features of the ECG looks similar to features obtained from the correct position. Six ML were used to detect chest lead misplacement. Bagged tree ML classifiers outperformed the other five ML classifiers significantly in the second ICS ($P<0.05$, McNemar's) and third ICS. MRMR, JMI and relief were the best feature selection algorithm in the second and third ICS respectively. However, as a limitation of this work, the hyperparameters were not tuned in this work as we used the default values of the hyperparameters. Based on the results, the developed algorithms could be used to reduce the number of incorrectly recorded ECGs which reduce the cost and time of performing a new ECG.

However, lead misplacement detection in the third ICS is challenging. Hence, further research could be carried out to detect lead misplacement in the third ICS using new ML technology such as deep learning. Medical device companies could deploy one of the aforementioned algorithms and included it in their products to improve the quality of recorded ECG and as consequence to improve physician's performance.

3.4 Conclusion:

In this chapter two different scenarios have been applied to detect lead V1 and V2 misplacement. In the first scenario, V1 and V2 were detected separately, while in the second experiment, V1 and V2 were detected together. Based on the results, ML classifiers obtained a similar level of performance in both experiments. Bagged tree ML algorithm obtained the highest performance do detect lead misplacement, especially, in the second ICS ($P<0.05$) compared to the other ML classifiers. However, the performance of ML classifiers declined significantly in both experiments in the third ICS. Hence, DL algorithms should be applied to detect lead misplacement, especially in the third ICS. Additionally, all features were extracted manually in both experiments. Hence, an automatic feature extraction method such as deep learning feature extraction method should be applied to see if both humans and computers consider the same features or not to detect chest lead misplacement. And a new study should be carried out to compare the accuracy of lead misplacement detection between physicians and ML algorithms.

Chapter 4

Auto Detect Leads Misplacement Using Deep learning

4.1 Overview

In the previous chapter, several traditional ML algorithms such as SVM, DT have been used to detect lead misplacement with promising results. However, these traditional ML algorithms obtained a low accuracy to detect lead misplacement in the third ICS. Furthermore, misplacement in the third ICS was challenging as features look similar to features in the correct position (fourth ICS). Hence, deep learning (DL) should be investigated to see if it improves the performance for detecting misplacement as it has demonstrated a promising performance in different ECG applications. Also, DL has never been used before to detect lead misplacement. This chapter describes the first study that used DL to detect lead misplacement.

In this chapter, DL techniques are used to detect V1 and V2 lead misplacement in two different scenarios. In the first scenario, DL used ECG signals to detect lead misplacement. While in the second scenario, DL used ECG images to detect lead misplacement. Also, in the second scenario, heat maps were generated from each prediction from the last layer to show which features were responsible for making the final classification and which feature contributed the most.

Objectives

The objectives of this chapter are:

- Discover the accuracy of DL algorithms in classifying V1 and V1 lead misplacement and to compare this to classical ML algorithms
- To discover if attention maps can provide explainable insights into how DL algorithms made their classifications

4.2 The first experiment

4.2.1 Material and Methods

Data collection:

Kornreich dataset which is available on the PHYSIONET database was used in this experiment. The dataset includes 151 normal subjects, 151 left ventricular hypertrophy patients and 151 myocardial infarction patients. Dalhousie torso with 117-lead body surface potential map (BSPM) electrodes and 352 nodes were used to extract the signal for each particular lead as shown in Figure 4.1. The 352 nodes were obtained from the BSPMs by multiplying the 117-lead BSPM electrodes with a transformation matrix to provide a high resolution. Figure 4.2 shows a simple flow diagram showing data science pipeline.

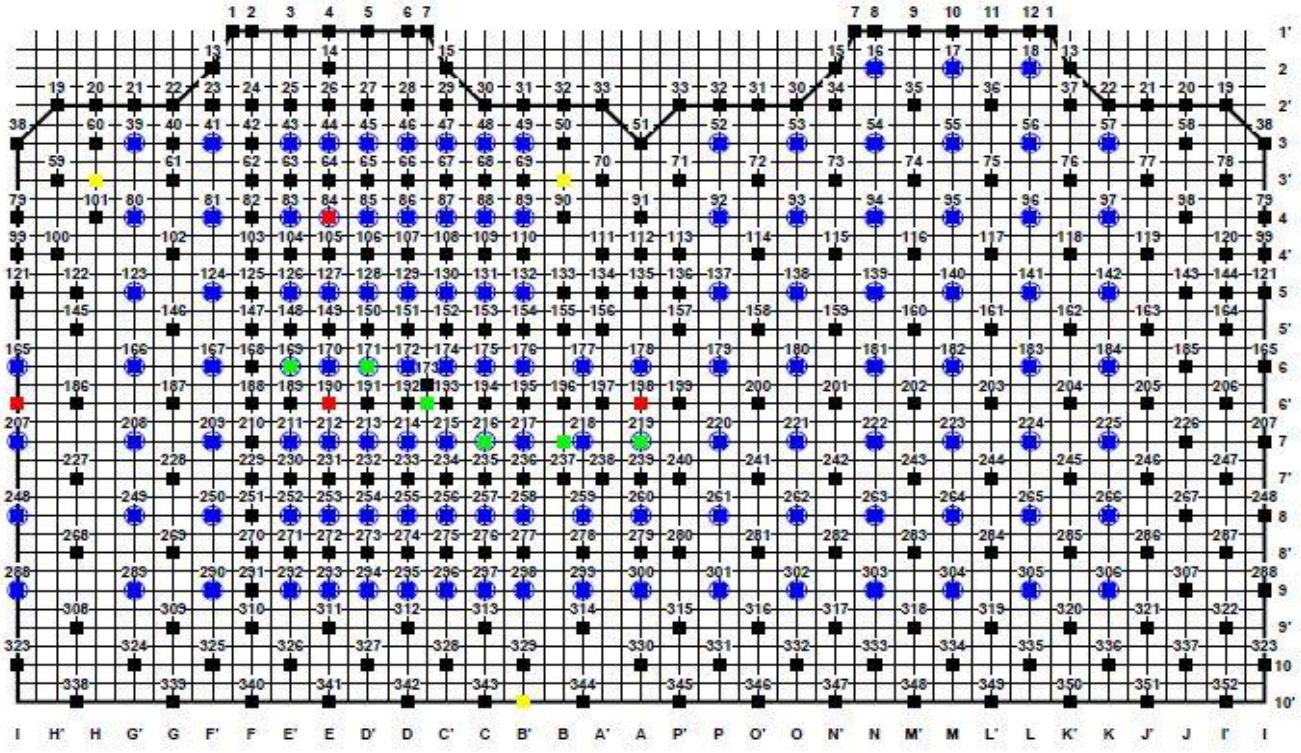


Fig. 4.1 Dalhousie torso with 352 nodes.

Chest leads V1 and V2 were extracted from Dalhousie torso as shown in Figure 4.1. Node 169 in green color represents the correct position for V1 (4th ICS). And node 171 in green color represents the correct position for V2 (4th ICS). While nodes in blue color represent the incorrect position for V1 and V2. As leads V1 and V2 are commonly misplaced vertically, nodes 126 and 83 were considered to represent V1 misplacement in the third intercostal space (ICS) and the second ICS respectively. While nodes 128 and 85 were considered to represent V2 misplacement in the third intercostal space (ICS) and the second ICS respectively. There is no feature engineering in DL (which includes feature extraction and feature selection). Hence, the extracted signals are fed into a DL network directly.

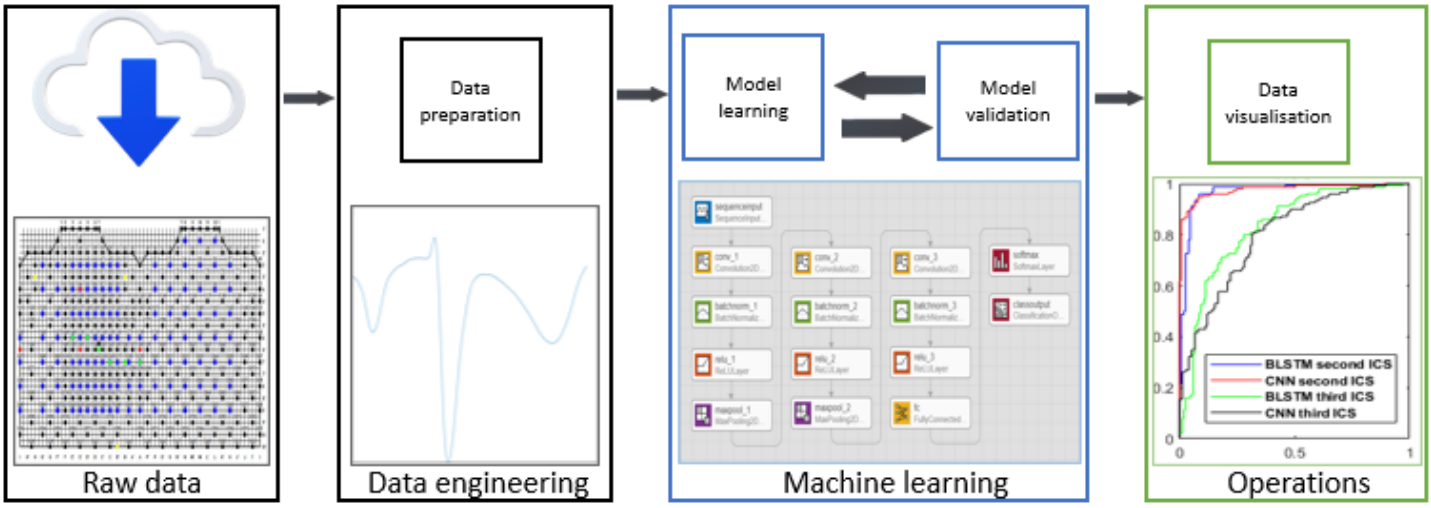


Fig. 4.2 Data science process.

The extracted ECG signals for V1 and V2 were divided into 70% for training and 30% for testing. In both testing and training sets, 50% of the signals were extracted from the correct position for V1 and V2. And the other 50% of the signals were extracted from the wrong position for V1 and V2 to make the dataset balanced and to avoid bias.

Data analysis software:
MATLAB toolbox was used to build the deep learning models.

Classification:

Two DL algorithms were developed including: 1) convolutional neural network as shown in Figure 4.3 and 2) Bidirectional long short-term memory (B-LSTM) network.

In the CNN, the network has been built using the following structure: fifteen layers were divided as follows: 1 input layer (which is used to feed in the input signals), 3 hidden convolution layers (which uses a filter to transform the input ECG into a convolution layer as shown in Figure 4.4), 3 batch normalization layers (to normalise the output of a previous layer), 3 rectified linear unit (ReLu) layers (which is used to remove negative values as shown in Figure 4.5) as it achieves good performance and is easier to train, 2 max-pool layers (which reduces the output of the previous layer into one single value to reduce the amount of computation in the network as shown in figure 4.6), 1 fully connected layer (which is used to connect each neuron in one layer to each neuron in next layer as shown in Figure 4.7), 1 soft max layer (which uses logistic regression to generate probability for each label using Equation (1) and one classification output layer to show the final prediction.

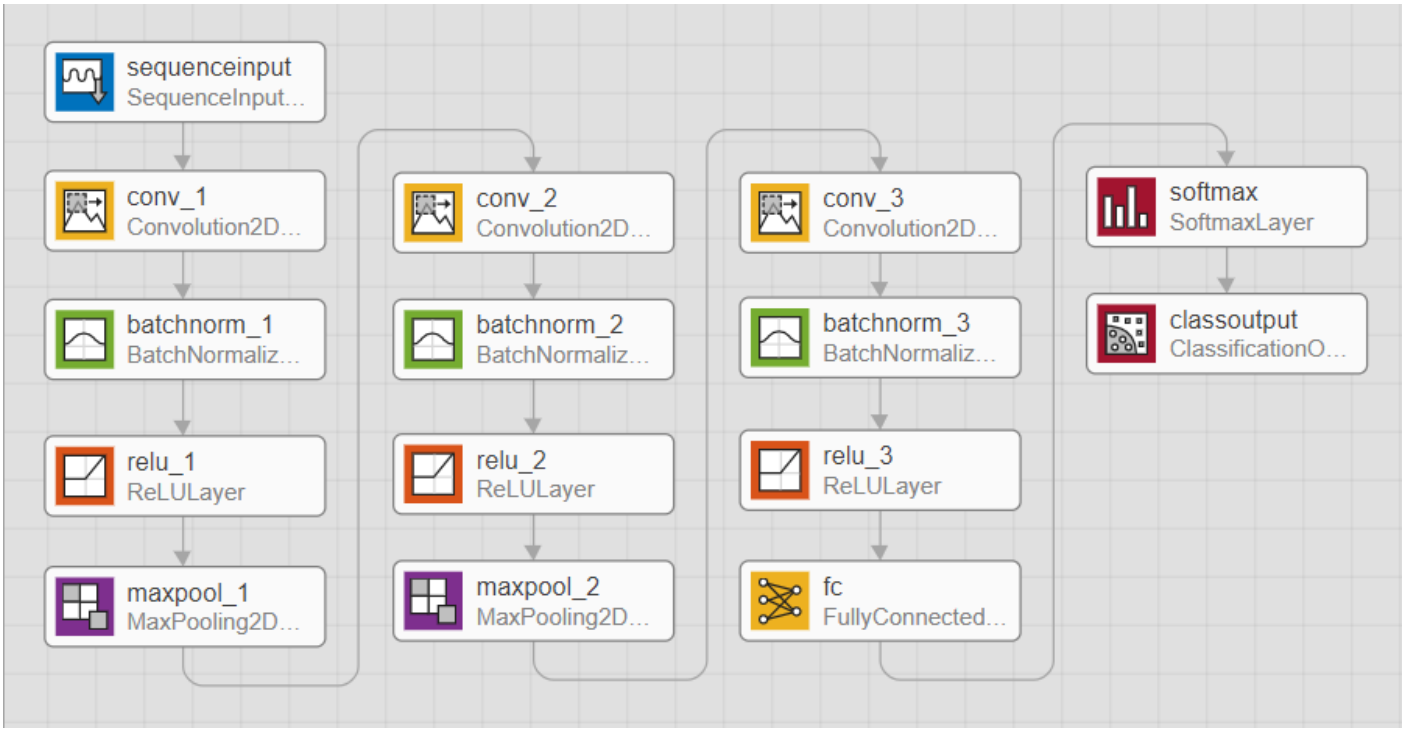


Fig. 4.3 Convolutional neural network structure.

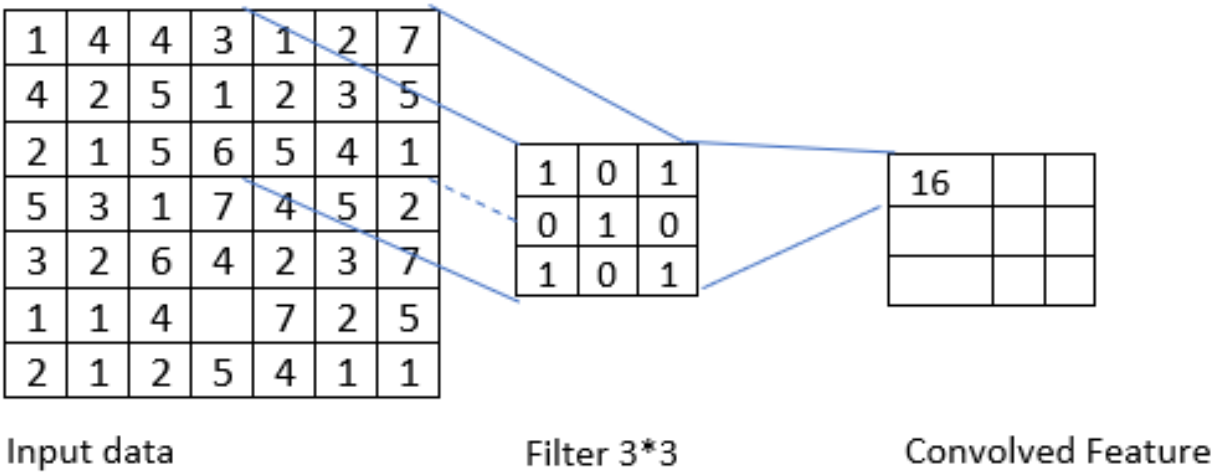


Fig. 4.4 Converting the input data into a convolutional layer.

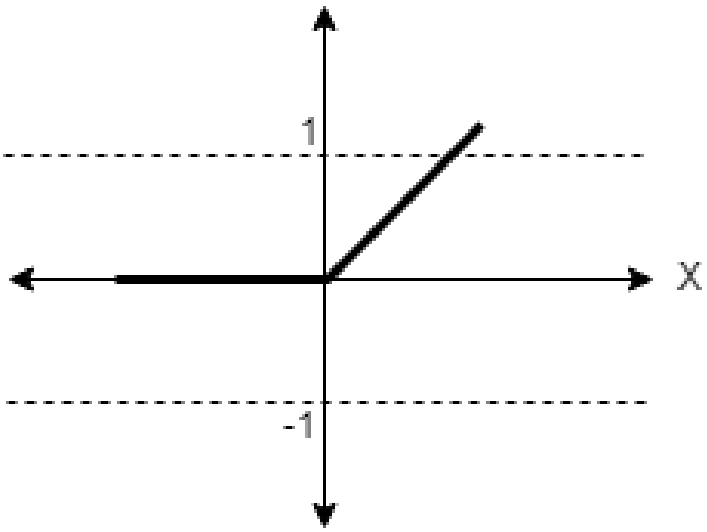


Fig. 4.5 Rectified linear unit (RLU) Activation function.

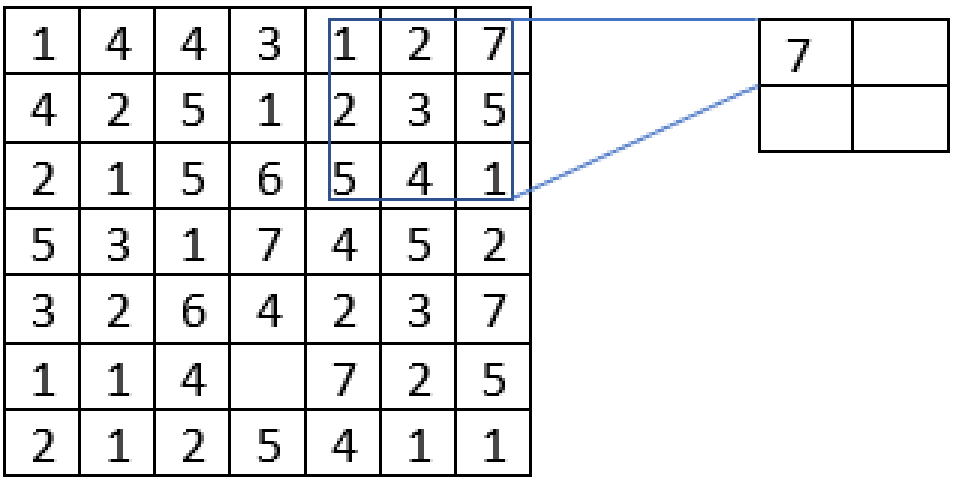


Fig. 4.6 Max pooling layer.

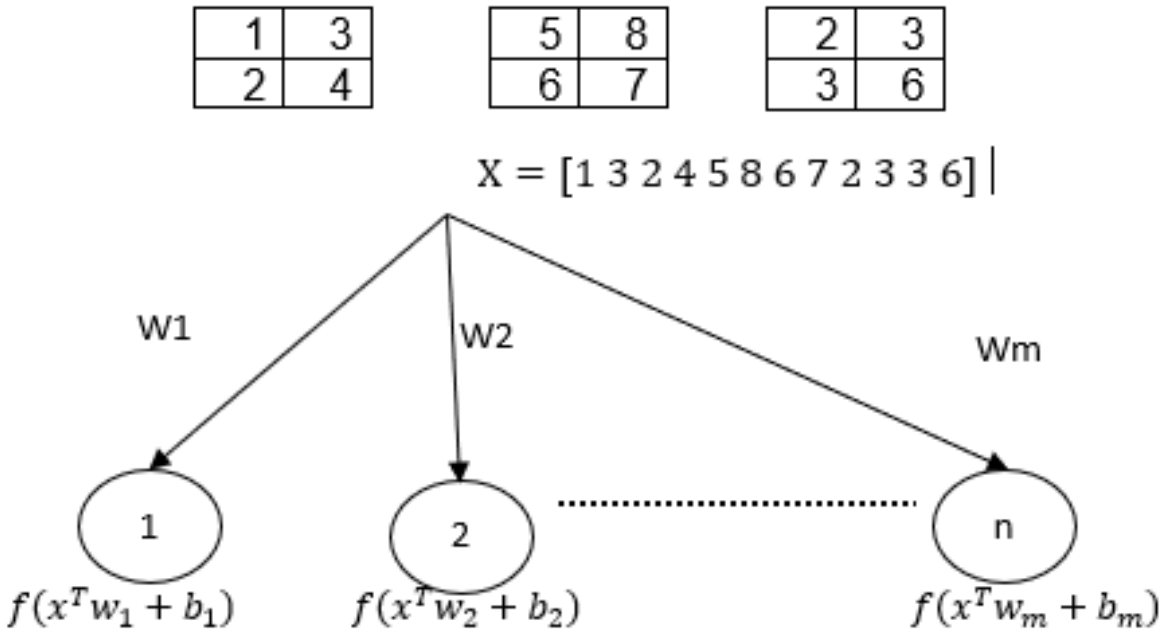


Fig. 4.7 The transition from a convolutional layer to a fully connected layer. In this figure for example, the fully connected layer reshapes the tensors (2X2X3) to be a vector (1x12). Then, each neuron in the fully connected layer will be associated with 12 different weights (based on the vector length) to produce a linear combination of a new vector. Hence, each neuron will produce an output using function $f(x^T w_m + b_m)$, where x represents the feature vector, w represents the weight and b is the bias term.

$$P(y = c|w, x, b) = softmax_c(x^T w + b) = \frac{e^{x^T w_c + b_c}}{\sum_j e^{x^T w_j + b_j}} \qquad (1)$$

Where (y) represents the predicted label, (c) represents class, (w) is the weights associated with each neuron and (b) is the bias value associated with each neuron of the output layer.

In the BLSTM as shown in Figure 4.8, the DL model were built using the following structure:

- One sequence input layer to read the ECG signal.
- Two B-LSTM layers (which are used to learn the DL network through each complete time series at each iteration).
- One fully connected layer.
- One soft max layer.
- One classification output layer.

In each subject, leads V1 and V2 signals were combined in one matrix of size 2x352 with one label (whether it is 0 for incorrect placement or 1 for correct placement). The first row in the matrix represents V1 and the second row represents V2. Each matrix has one label whether it is 0 or one.

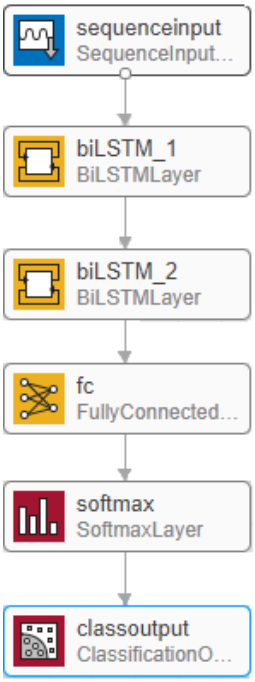


Fig. 4.8 Long short-term memory (LSTM) network structure.

4.2.2 Results

Two deep learning networks including 1) B-LSTM and 2) CNN were developed to detect leads V1 and V2 misplacement. ECGs were fed into the DL networks directly without feature engineering (which includes feature extraction and selection). B-LSTM network outperformed the CNN in the second and third ICS as shown in Figures 4.9, 4.10 and 4.11.

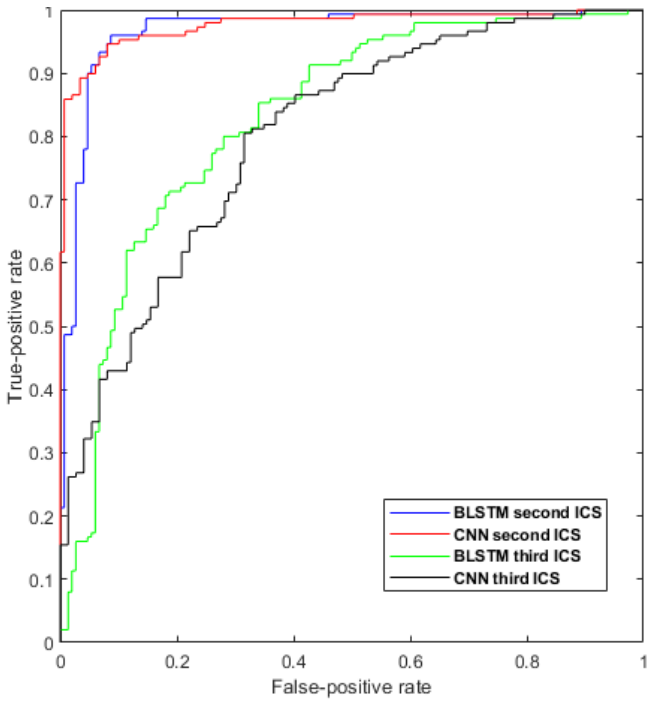


Fig 4.9 Roc curve of BLSTM and CNN performance in the second and third ICS.

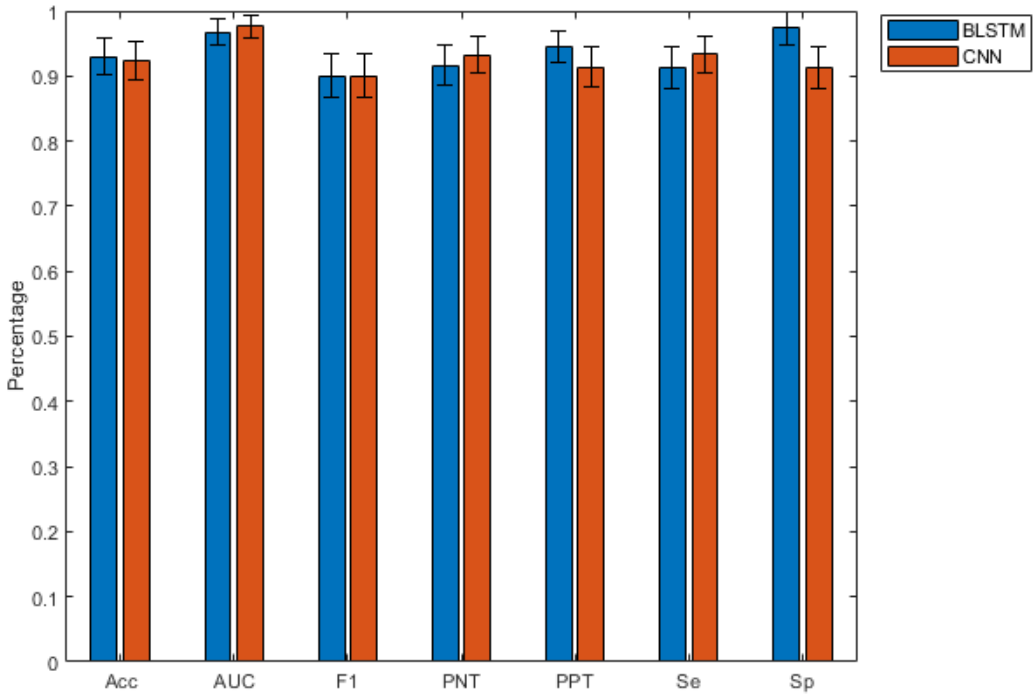


Fig. 4.10 Accuracy (Acc), area under the curve (AUC), F1 score, predictivity of negative test (PNT), predictivity of positive test (PPT), sensitivity (Se) and specificity of B-LSTM and CNN performance in the second ICS.

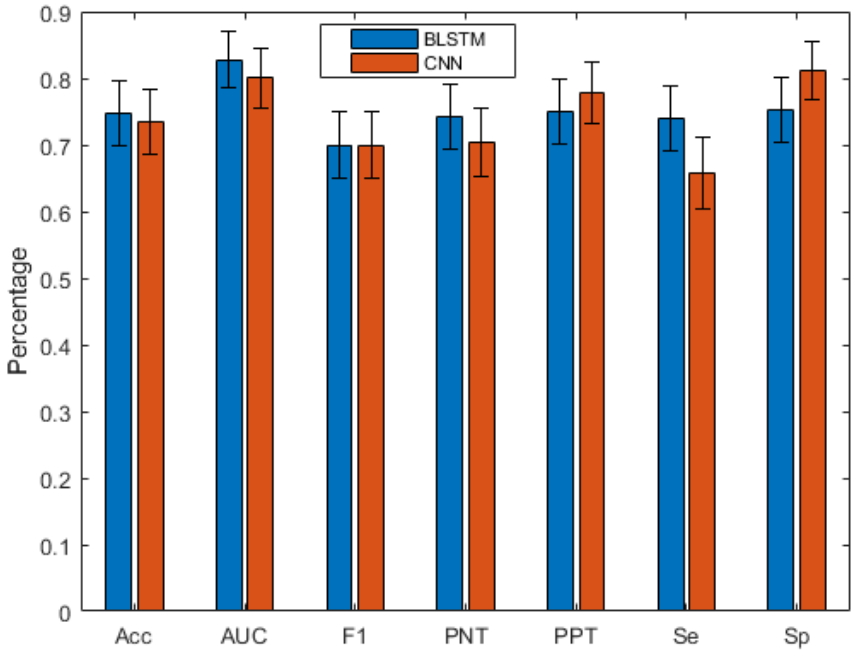


Fig. 4.11 Accuracy (Acc), area under the curve (AUC), F1 score, predictivity of negative test (PNT), predictivity of positive test (PPT), sensitivity (Se) and specificity of B-LSTM and CNN performance in the third ICS.

B-LSTM obtained a high accuracy compared to CNN, especially in the third ICS (74.7%) as shown in table 4.1. In terms of sensitivity and specificity, BLSTM obtained higher specificity (97.4%, P<0.01, McNemar's) and lower sensitivity (91.3%)

compared to CNN which obtained 91.3% specificity and 93.3% sensitivity. There was no significant difference in terms of F1 score and AUC between the two models.

Table 4.1 Classification accuracy using B-LSTM and CNN.

Classification accuracy using DL			
Accuracy in the 2nd ICS		**Accuracy in the 3rd ICS**	
B-LSTM	**93.0%**	**B-LSTM**	**74.7%**
CNN	92.3%	**CNN**	73.5%

4.2.3 Discussion

Based on the literature review, 12-lead ECG lead misplacement is one of the most critical problems that affect ECG interpretation. Leads V1 and V2 are one of the most commonly misplaced electrodes. Traditional ML learning algorithms have been used to detect misplacement in the second and third ICS. In this study, a new approach was applied to detect leads V1 and V2 misplacement using two DL networks including 1) B-LSTM network and 2) CNN.

In those DL networks, there is no feature engineering which consists of feature extraction and selection. Hence, the features were extracted automatically using DL networks. This experiment is the first experiment that uses DL to detect chest electrode misplacement whereas all the previous works used traditional ML algorithms to detect limb lead interchanges. The limitation of using CNN in this experiment was the size of the dataset which was small. However, the dataset was used to show the ability of CNN to detect lead misplacement for future research.

B-LSTM obtained 93.0% accuracy when leads V1 and V2 are misplaced in the second ICS and 74.7% accuracy when they are misplaced in the third ICS. The performance in the 3rd ICS declined as the morphology of the recorded signal in the third ICS is close to the recorded signal form the 4th ICS.

While CNN obtained 92.3% accuracy in the second ICS and 73.5% accuracy when they are misplaced in the third ICS. Hence, DL improved the performance of detecting V1 and V2 misplacement in the third ICS compared to the performance of traditional ML algorithms such as bagged tree and SVM. Using DL algorithms in applications to detect medical errors like lead misplacement detection has a low risk as it is not used for diagnosis. Hence, medical companies could consider using DL in areas that does not have a high risk.

4.3 The second experiment

4.3.1 Material and Methods

Data collection:

The same dataset in the first experiment was used in the second experiment. However, in this experiment, all the extracted signals for V1 and V2 were converted to images as shown in Figure 4.12. Dalhousie torso in Figure 4.1 with 117-lead body

surface potential map (BSPM) electrodes and 352 nodes were used to extract the signal for leads V1 and V2, then the signals were converted to images. Node169 represents V1 and node171 represents V2. The green color represent the correct position of the precordial leads (V1 to V6).

While the blue color represents the incorrect position of V1 and V2. Node 83 was considered to represent V1 misplacement in the second ICS. While node 85 was considered to represent V2 misplacement in the second ICS. To provide a class balance and to avoid bias, 50% of cases were incorrect ECGs and 50% were correct.

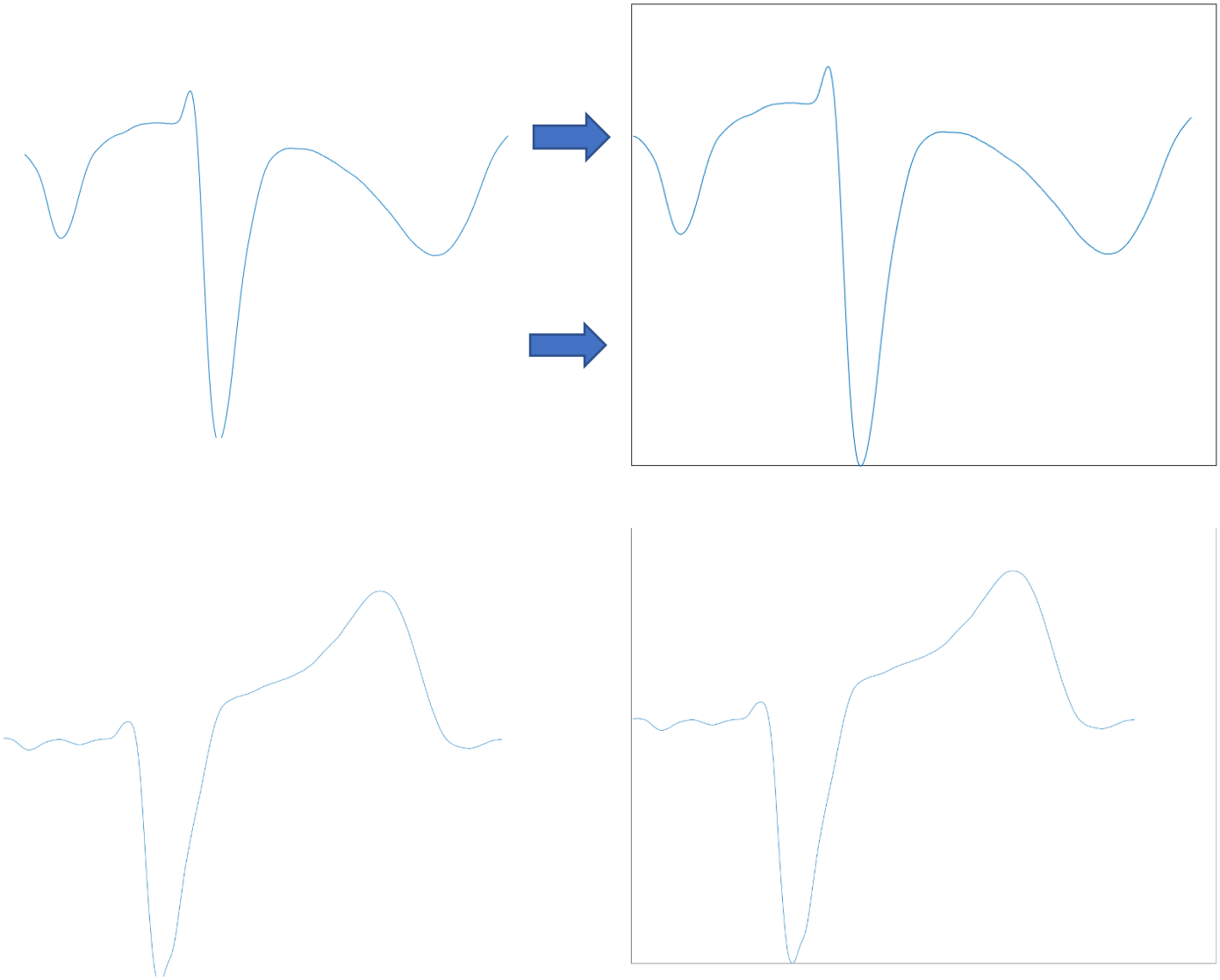
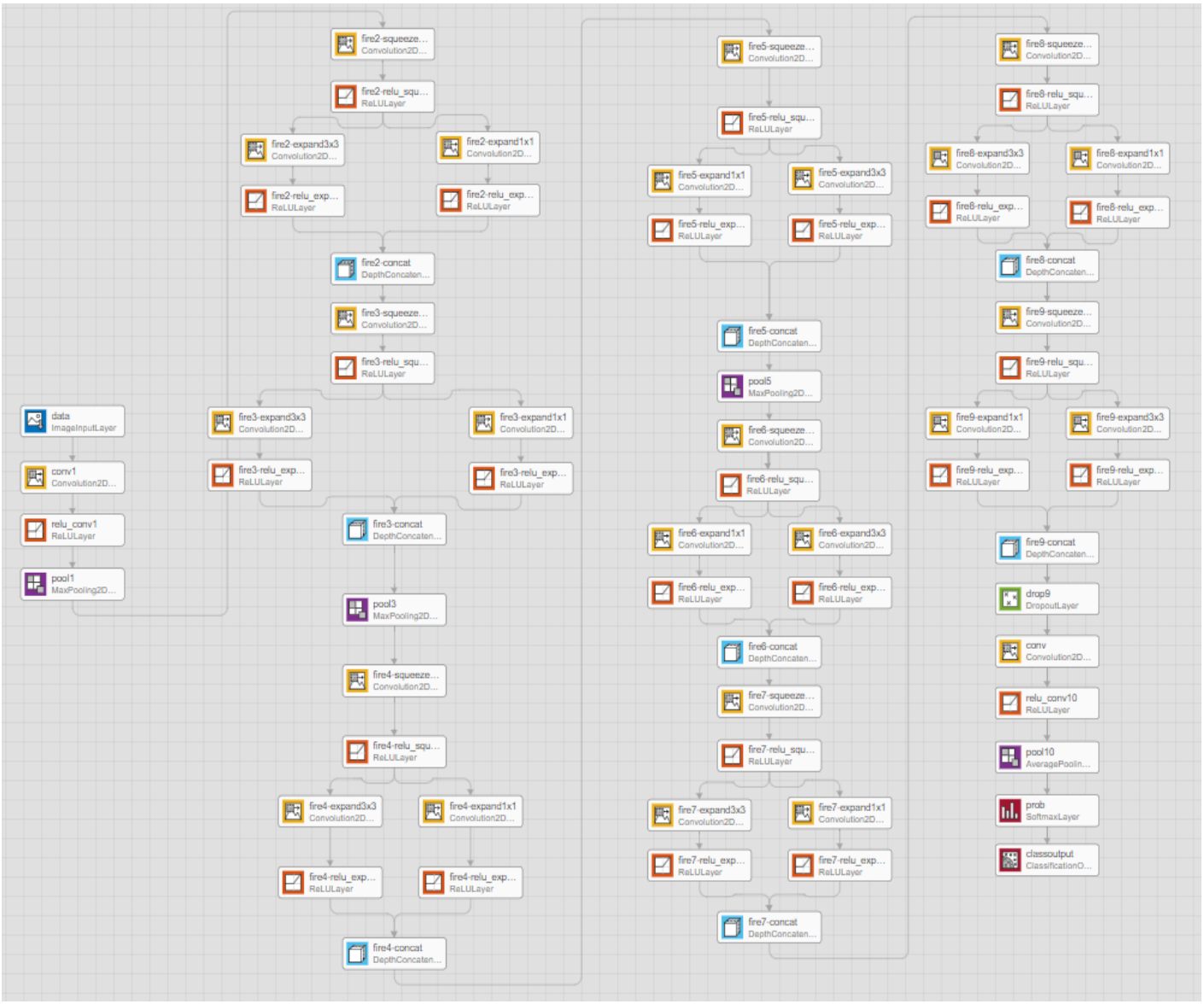


Fig. 4.12 Converting ECG signals to images for V1 and V2. **a** represents converting V1 signal to image and **b** represents converting V2 signal to image.

Classification:

A CNN was developed using 68 hidden layers as shown in Figure 4.13. then the CNN was trained using 70% of the images and tested using 30% of the generated images. Then attention maps were generated from the last convolution layer to show

the contribution of each feature in the final prediction as shown in Figure 4.14. The attention maps were inspected visually and the features were highlighted manually to analyse the frequency of feature use



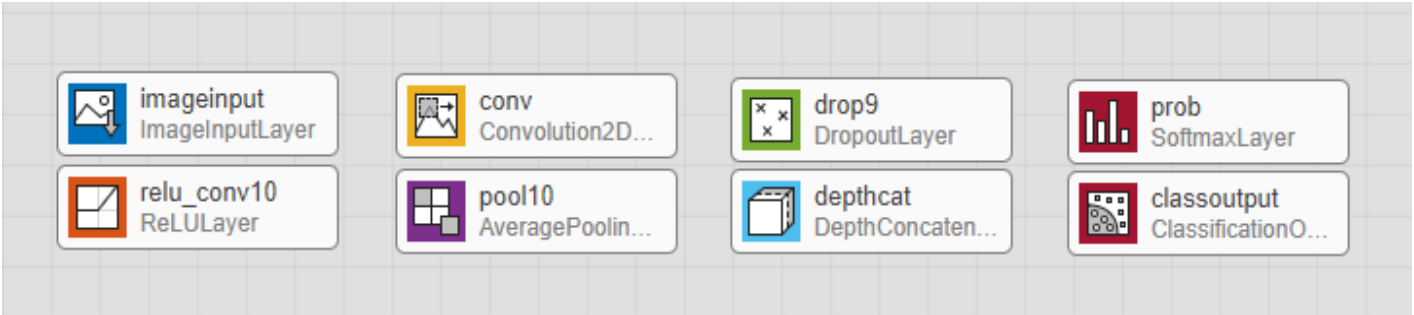


Fig. 4.13 CNN structure.

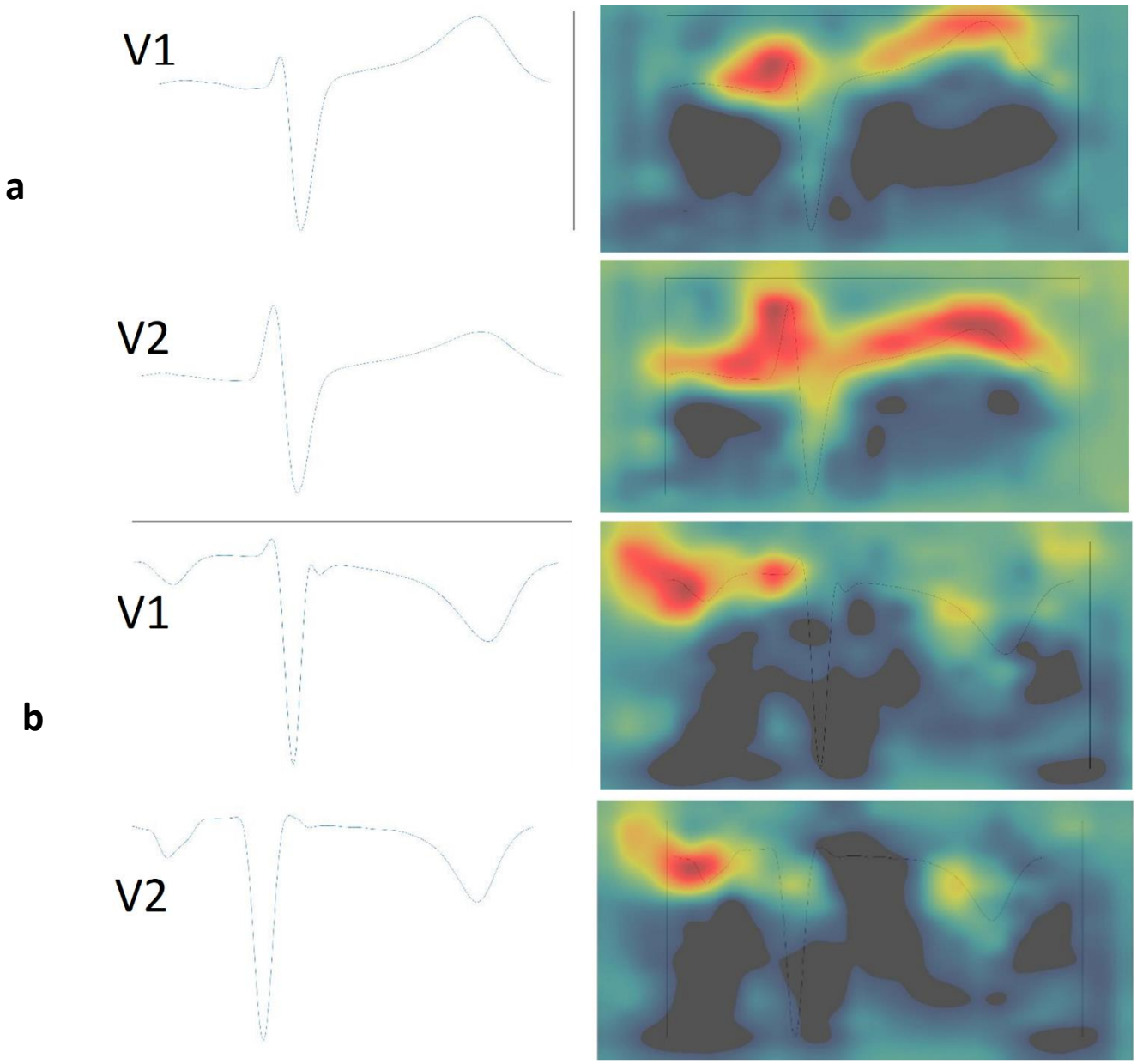


Fig 4.14 Attention maps for V1 and V2. **a** represents the attention map of V1 and V2 when they were in the correct position. While **b** represents the attention map of V1 and V2 when they were in the incorrect position.

4.3.2 Results:

A developed CNN were used in this experiment to detect lead misplacement in the second ICS using images instead of signals. The CNN obtained a high performance in terms of accuracy (Acc), true negative (TN), true positive (TP), false negative (FN) and false positive (FP) as shown in table 4.2 and Figure 4.15. P waves, T waves and R waves contributed the most in the final prediction respectively (R waves (56%), T waves (55%) and R waves (48%)) as shown in Figures 4.16 and 4.17 based on the generated attention maps in Figure 4.18. While S waves were not considered related in most cases in detecting lead misplacement of V1 and V2. And the rest of the features such as the PR interval, J point and Q waves contributed 29%, 27% and17% respectively to the final predicted classes.

Table 4.2 CNN performance in the second ICS for detecting lead V1 and V2 misplacement.

	Acc	**TN**	**TP**	**FN**	**FP**
CNN	92.6%	291/300 (97%)	265/300 (88.3%)	35/300 (11.6%)	9/300 (3%)
Acc: Accuracy, TN: True Negative, TP: True Positive, FN: False Negative, FP: False Positive.					

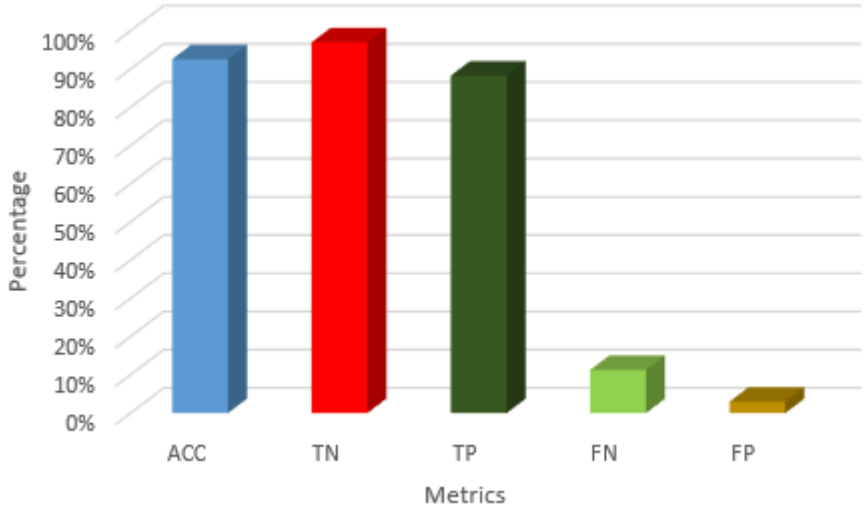


Fig.4.15 CNN performance to detect lead misplacement.

	Total P wave	Total PR wave	Total Q wave	Total R wave	Total S wave	J point	Total T wave
V1+V2 correct	187	79	44	181	0	61	225
V1+V2 incorrect	148	96	59	107	5	103	107
total(correct+incorrect)	335	175	103	288	5	164	332
(correct+incorrect) %	56%	29%	17%	48%	1%	27%	55%

Fig.4.16 Feature contribution in the final prediction based on attention maps from CNN.

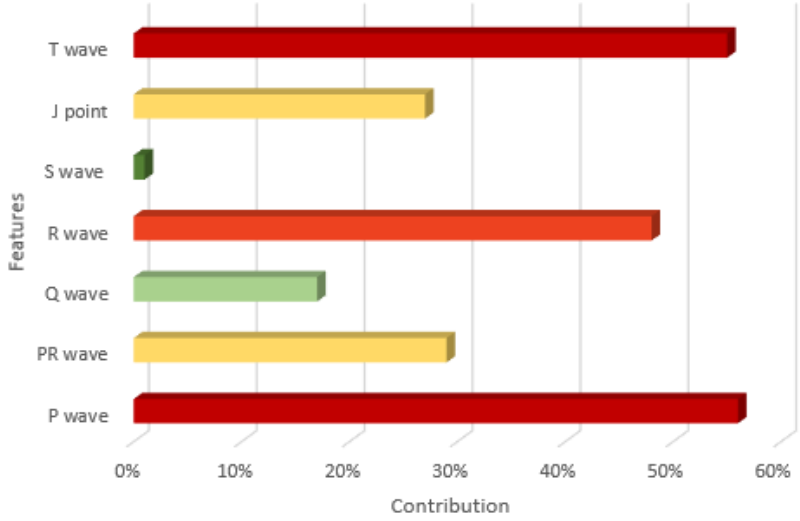


Fig.4.17 Feature contribution in the final prediction (correct and incorrect classification) based on attention maps.

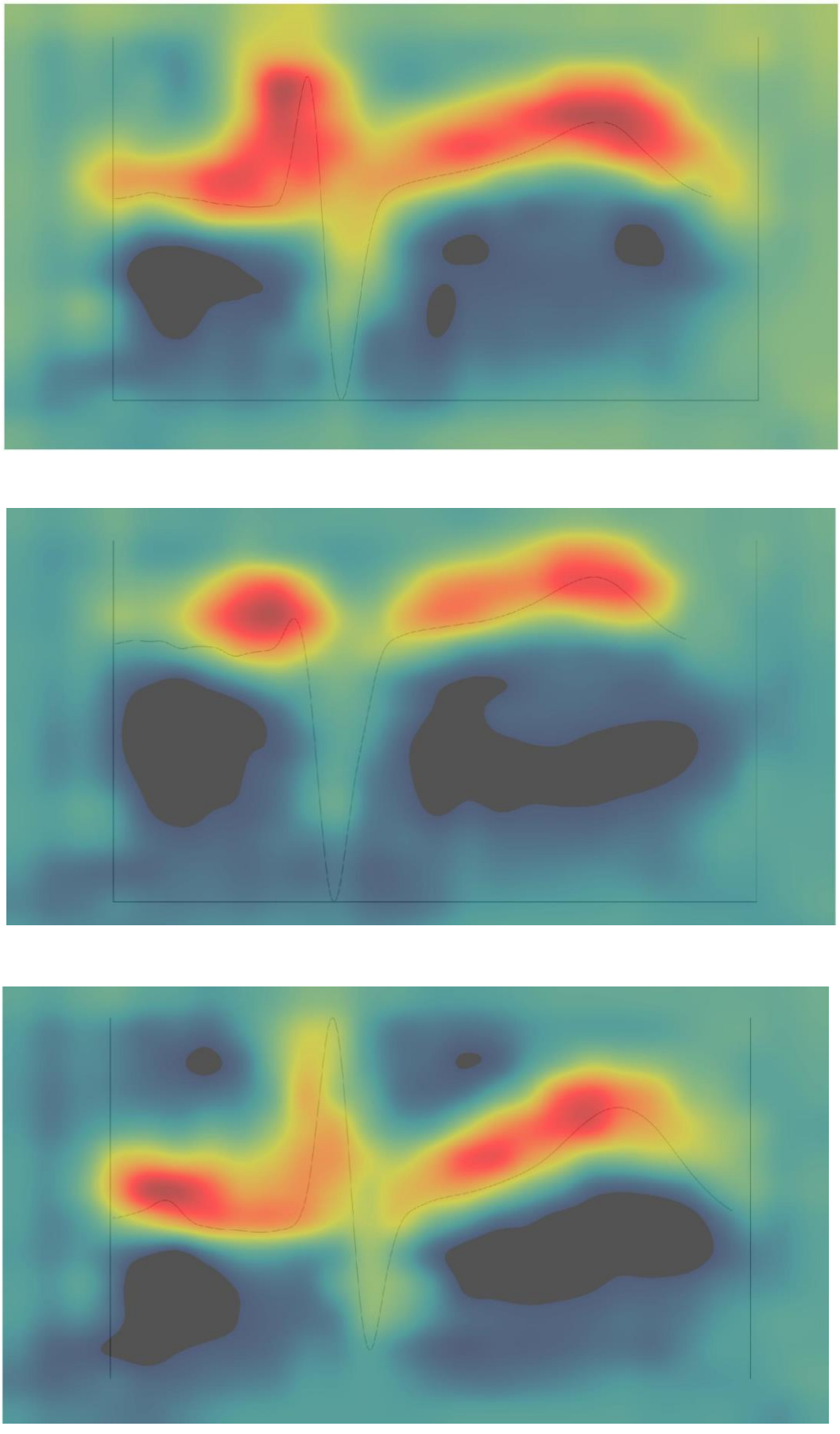


Fig.4.18 Generated heat maps from the last layer in the CNN.

4.3.3 Discussion:

CNN obtained a high accuracy (92.6%) to detect lead misplacement. Attention maps were extracted from the last layer in the CNN to show the importance of each feature and the contribution of each feature in the final prediction. Hence, attention maps increased the transparency and the explainabilty of the algorithm during decision-making by showing what the ML algorithm looked at before making the decision. According to the results, P wave, R wave and T wave contributed more than the other features and they should be considered as important features by physicians. Hence, attention maps could be used by physicians to calibrate their trust with ML. The remaining morphological features including S wave, PR interval, Q wave and J point were considered as mid-level features.

According to the results from chapter 3 and chapter 4, the DL learning and traditional ML algorithms obtained a similar performance in the second ICS, while the DL outperformed the traditional ML algorithms significantly in the 3rd ICS ($P< 0.05$, McNemar's).

4.4 Conclusion:

DL algorithms including CNN and B-LSTM improved the accuracy of detecting lead V1 and V2 misplacement significantly ($P<0.05$) especially in the third ICS compared to the performance of traditional ML algorithms such as SVM and DT as shown in chapter 3. However, traditional ML algorithms were more transparent and explainable especially DT compared to the DL algorithms. Hence, to overcome the explainabilty and transparency issues, attention maps were used in the deep CNN to make the algorithm transparent. This chapter also showed that using ECG as a signal or as an image dose not effect DL algorithms performance as the DL algorithms performance was quite similar in both datasets (signals dataset and images dataset). The developed algorithms could be used by medical industry companies to make their products perform better and to make it transparent and accurate. However, those experiments have a limitation including dataset size as DL require a large dataset to obtain a high performance.

Chapter 5

Uncertainty Visualisation in Automated ECG Interpretation

5.1 Overview

The Electrocardiogram (ECG) is the traditional method to diagnose heart abnormalities by measuring heart's electrical activity and rhythm. ECG interpretation is the most important step after ECG recording as it is used to inform treatments. However, the ECG has a limited specificity (70–95%) and sensitivity (30–70%) to detect some heart abnormalities (Finlay et al., 2007). Hence, new methods such as automated ECG interpretation was developed to improve ECG interpretation by using AI technologies such as ML. In automated ECG interpretation, ML obtained a high accuracy to detect several heart abnormalities such as atrial fibrillation. However, human ECG interpretation and automated ECG interpretation could be affected by different factors such as lead misplacement as discussed in the third and fourth chapter and noises such as the muscular noise, the respiratory noise and the power-line interference which can make human and ML interpretation less reliable. Automated ECG interpretation algorithms should convey the level of certainty to show "how confident they are?" in its interpretation. ECG algorithms today often do convey some level of confidence by using words such as 'possibly', 'probably', or 'consider diagnosis X' but there is a lack of quantified confidence/certainty indices.

Generally speaking, a confidence interval (CI) is a method that is used to show the confidence level of a human or ML decision or recommendation (Pang, A. et al, 1997). In statistics, CI is calculated using several methods such as standard error and standard deviation as shown in Figure 5.1.

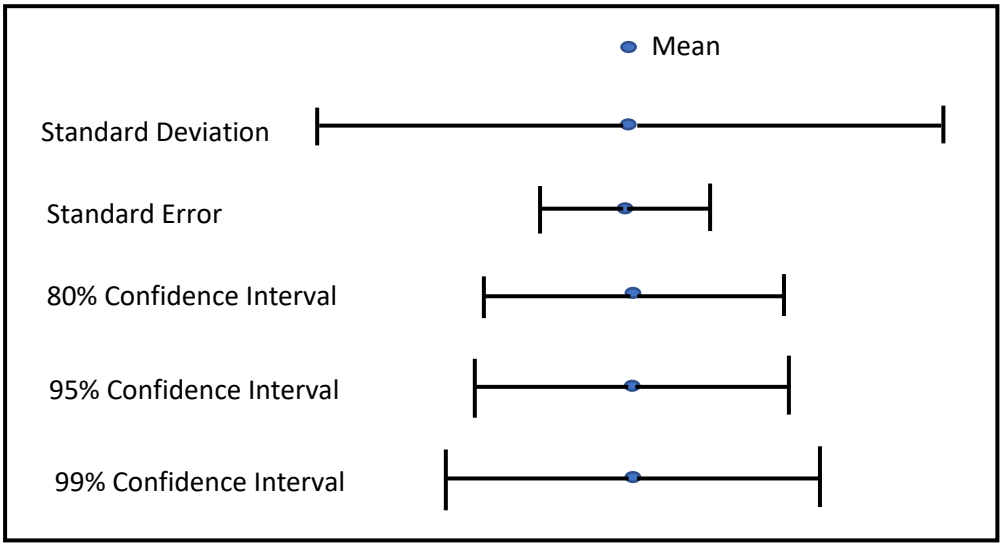


Fig. 5.1 Confidence interval computation methods.

However, a new uncertainty visualisation could be developed using a new CI computation method. Hence, in this chapter, a new uncertainty visualisation was

developed to better calibrate the confidence a person should have in classifications suggested by an ML algorithm. In this new method, the confidence interval was computed based on the ML performance on a noisy and clean (filtered) data. The ML performance on the noisy data represents the lower bound of the CI and the ML performance on the filtered data represents the upper bound of the CI. Conceptually, this is very different when compared to using standard error. It also avoids showcasing ML performance on 'ideal' cases only (clean/filtered data) and provides more of a 'realist' performance.

5.2 Material and Methods

Three bidirectional long short-term memory networks (B-LSTM) DL algorithms were trained and tested using three slightly different datasets. These three different DL algorithms reflect three different scenarios, including 1) training on filtered/clean data and tested on filtered data and noisy data, 2) training on noisy data and tested on noisy and filtered data and 3) training on noisy and filtered data together and tested on noisy and filtered data as shown in Figure 5.2. In the testing step, the bootstrapping method has been used to ensure the performance measured is representative.

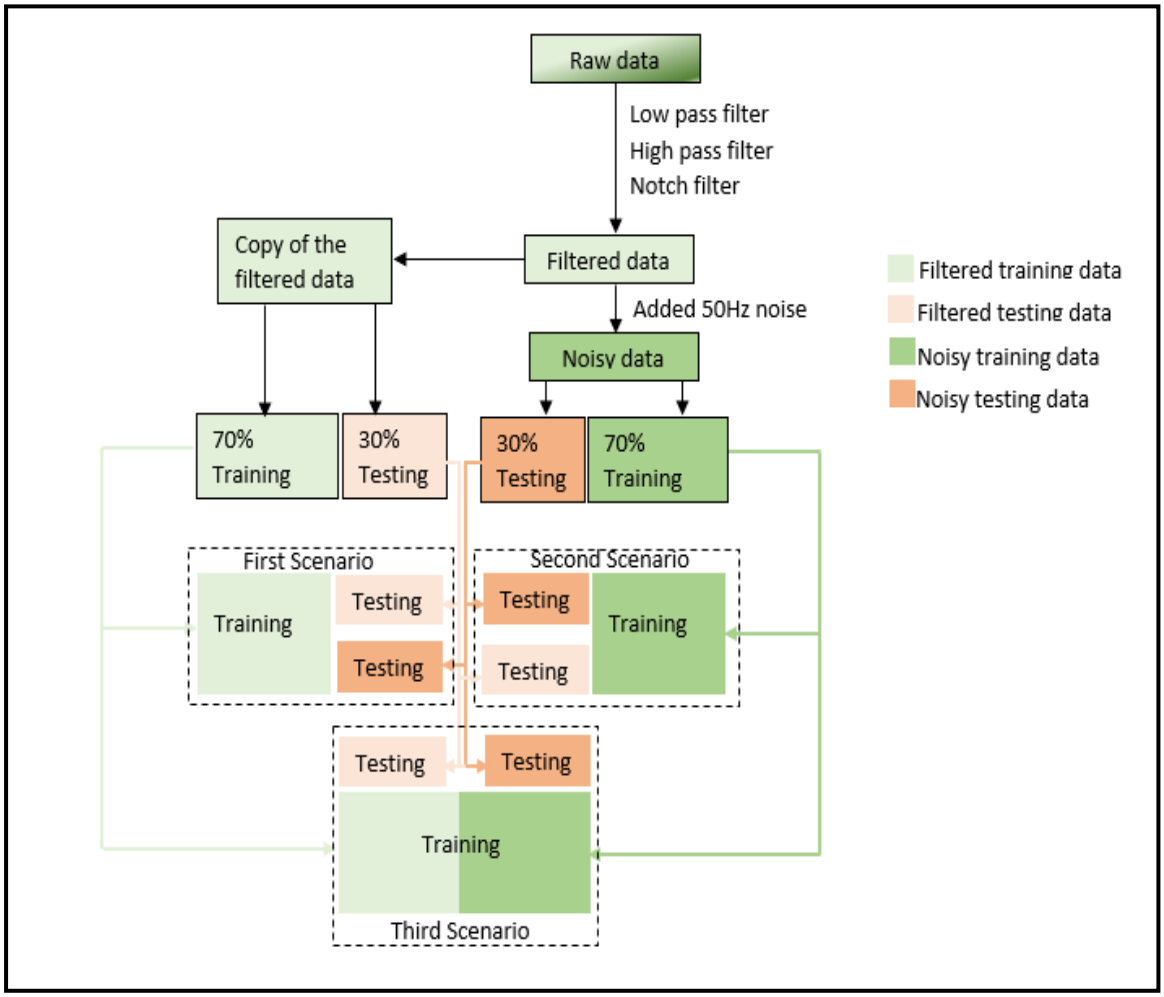


Fig. 5.2 Training the DL based on three different scenarios/training datasets.

Data collection:

The dataset was extracted from the “MIT-BIH Arrhythmia Database” which was collected by the laboratories at Boston’s Beth Israel Hospital. The dataset includes ECGs obtained from 47 patients. QRS complexes were extracted from these ECGs and included complexes that are considered premature ventricular contractions (PVCs) and non-PVCs. Each instance of a PVC/non-PVC included 360 samples per second. The final dataset included a class balance of 7,130 PVCs records and 7,130 non-PVCs. The dataset was balanced to avoid classification bias. Each row in the dataset has 60 samples which includes 30 samples before the R peak and 30 samples after the R peak. Hence, each row in the dataset represents ~167 milliseconds (number of samples (60) / (360) sampling frequency). Figure 5.3 below show a sample of a PVC and a non-PVC record.

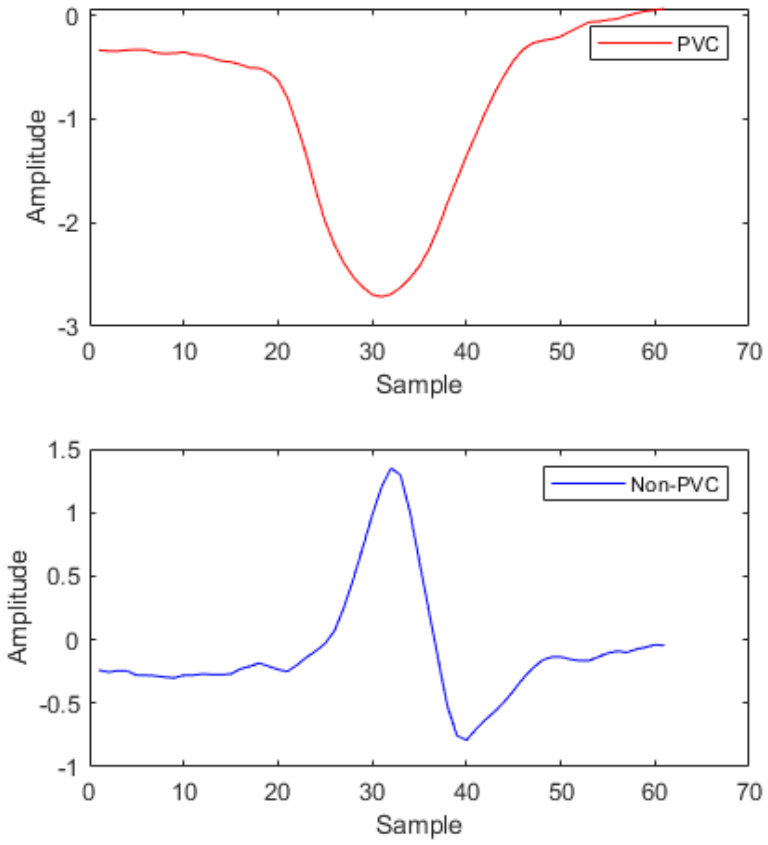


Fig. 5.3 A QRS complex for PVC patient and a QRS complex for a non-PVC subject.

Data analysis software:

MATLAB toolbox was used to build the deep learning models.

Digital signal processing:

From the raw dataset, we generated a filtered dataset by applying three digital filters including low pass filter (<80Hz), high pass filter (>0.5Hz) and notch filter (at 60 Hz) to remove the muscular noise, the respiratory noise and the power-line interference respectively as shown in Figure 5.4. The noisy dataset was generated by adding 50Hz noise to the filtered dataset as shown in Figure 5.5. Figure 5.6 below

shows a QRS example of the raw, filtered and noisy signal for normal and PVCs subjects.

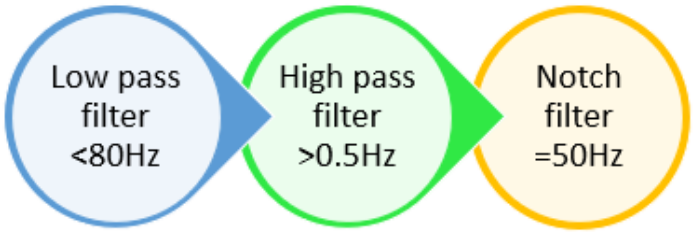


Fig.5.4 Signal filtering using a sequence of three digital filters.

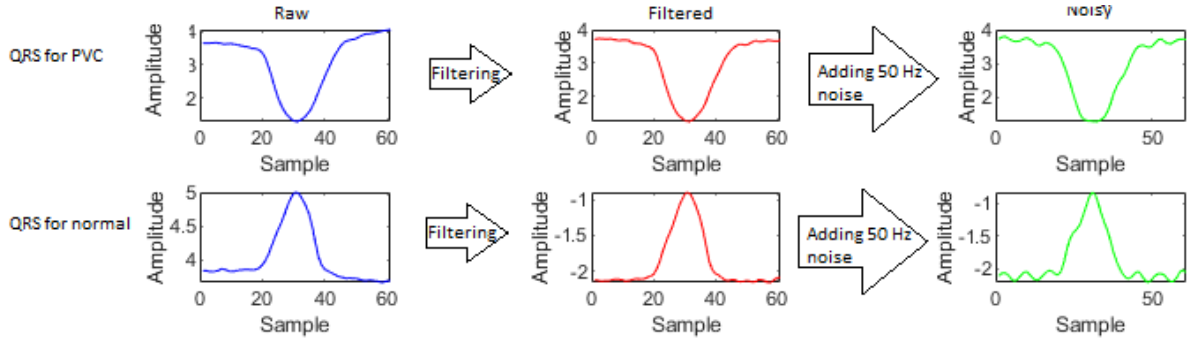


Fig.5.5 Signal processing process.

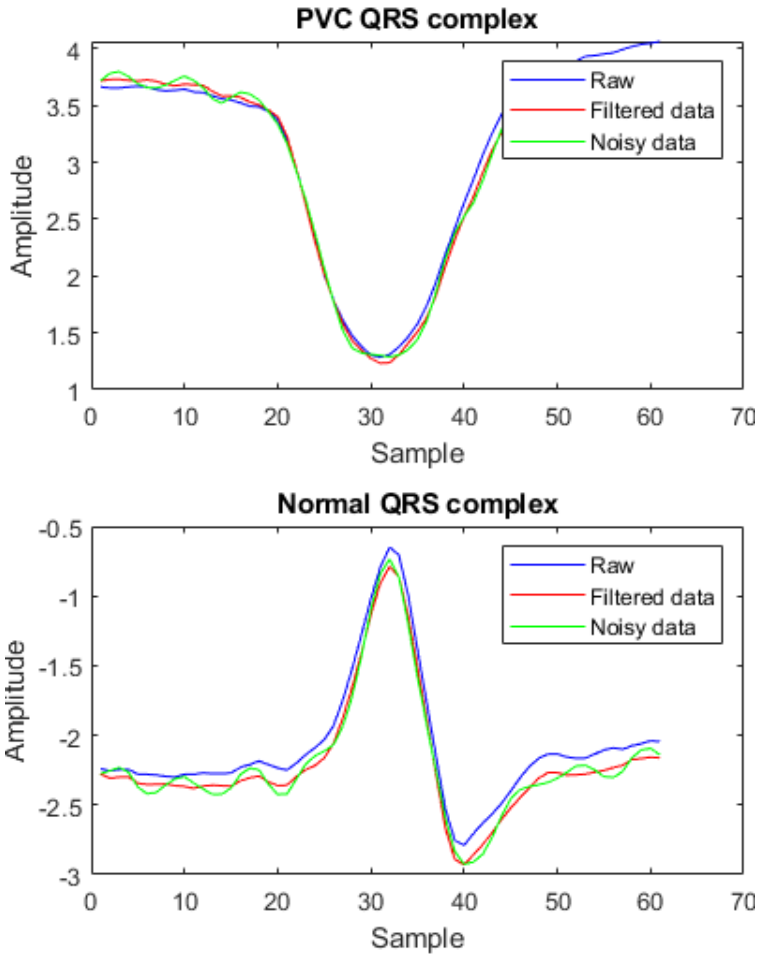


Fig. 5.6 QRS complex for a PVCs patient and QRS complex for a normal subject.

Classification:

The DL model was built using 7 layers as shown in figure 5.7 using Deep Learning Toolbox in MATLAB. In the training step, the DL model was trained using 70% of the data in three different scenarios as shown in Figure 5.5. In the testing step, the trained model was tested on both noisy and filtered data using 30% of the data in each scenario.

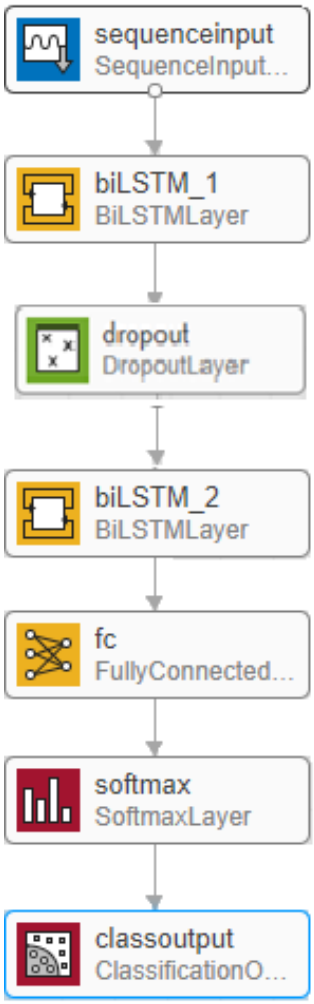


Fig. 5.7 B-LSTM DL network structure.

The performance of the DL algorithms was evaluated using four metrics including accuracy (Acc), sensitivity (Se), specificity (Sp) and area under the curve (AUC) using bootstrapping method. McNemar's test was used to compare performance between the algorithms.

5.3 Results

The performance of the three different trained DL algorithms was significantly different ($P<0.02$, McNemar's) in terms of accuracy in each scenario as shown in

table 5.1. Figure 5.8 below shows the performance of DL in each scenario in details. Based on the results, DL in the first scenario obtained the best performance compared to the other two scenarios.

Table 5.1 The performance of DL in three different scenarios (first scenario: training on filtered/clean data and tested on filtered data and noisy data, second scenario: training on noisy data and tested on noisy and filtered data, third scenario: training on noisy data and tested on noisy and filtered data together and tested on noisy and filtered data).

Scenarios	**Tested on Filtered Data**				**Tested on Noisy Data**				**Absolute accuracy difference between filtered and noisy data**	**P Value**
	Acc	**Se**	**Sp**	**AUC**	**Acc**	**Se**	**Sp**	**AUC**	**|Acc|**	
1st scenario: DL trained on clean data	92.0 %	91.4%	92.7%	0.976	90.0%	90.3%	90.8%	0.971	2%	**P<0.02**
2nd scenario: DL trained on noisy data	89.6 %	93.9%	85.0%	0.966	88.8%	95.2%	81.8%	0.964	0.8%	**P<0.05**
3rd scenario: DL trained on clean & noisy data	87.6 %	94.1%	80.5%	0.969	85.8%	94.6%	76.5%	0.964	1.8%	**P<0.02**
Acc: Accuracy, Se: Sensitivity, Sp: Specificity, AUC: Area Under the Curve.										

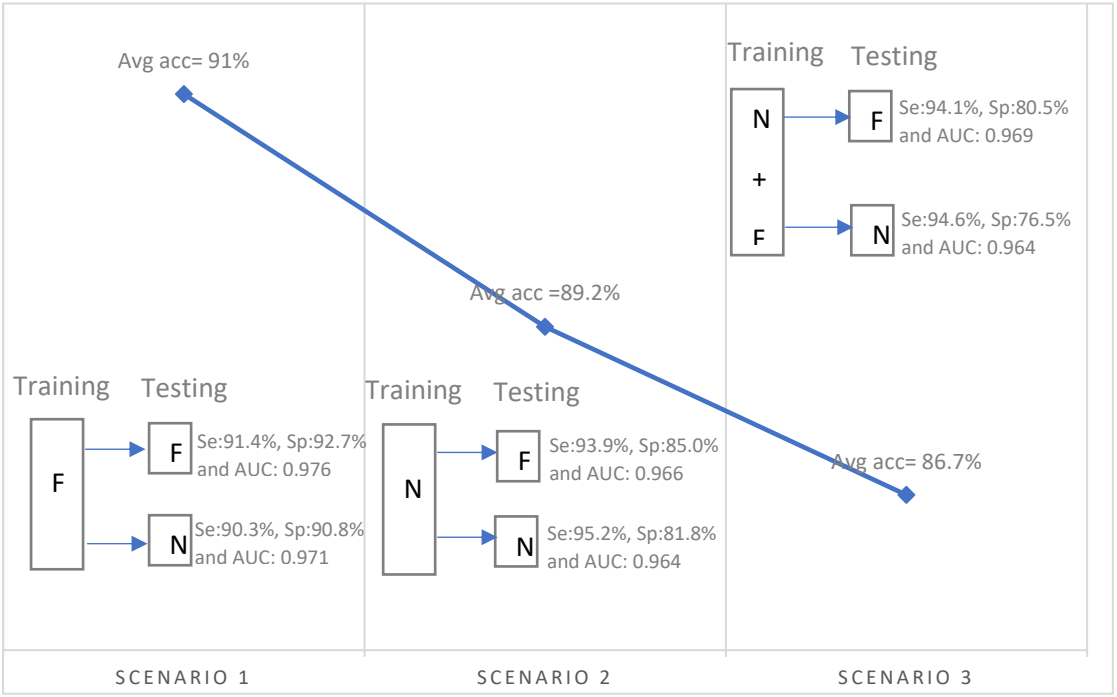


Fig. 5.8 DL algorithms performance in the three different scenarios.

The first scenario:

In the testing step, the DL obtained 92.0% Acc, 91.4% Se, 92.7% Sp with AUC = 0.975 when tested on filtered data. While it obtained lower Acc (90.0%), Se (90.3%), Sp (90.8%) and AUC (0.971) when tested on noisy data as shown in table 5.1 with a significant difference (P<0.02).

The second scenario:

In this scenario, the DL model obtained quite similar results in both filtered and noisy data with no significant difference as shown in table 5.1.

The third scenario:

In this scenario, the DL performance was significantly different when tested on noisy and filtered data (P<0.02) as shown in table 5.1.

Uncertainty visualisation:

In terms of uncertainty visualisation, a new uncertainty visualisation method was used to show the uncertainty as shown in Figures 5.9, 5.10, 5.11 and 5.12. In the new method as shown in the purple bar in Figure 5.9, the lower bound of the error bar represents the DL performance when tested on the noisy data, which is akin to the worst-case scenario, while the upper bound represents the DL performance when tested on the filtered/clean data which is akin to the best-case scenario.

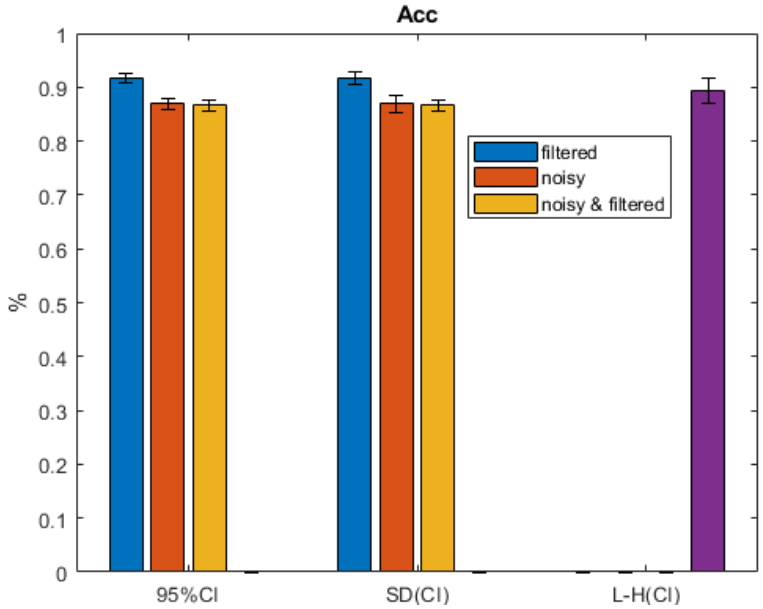


Fig. 5.9 The confidence intervals (CI) of the DL learning performance based on accuracy using three different methods including 1) 95% CI, 2) standard deviation CI and 3) the new CI method which is based on the DL performance on the noisy and filtered dataset. The blue bar represents the first scenario, the orange bar represents the second scenario and the yellow bar represents the third scenario, while the purple bar represents the new CI method.

Finally, the midpoint on the purple bar is the average between the lower and upper bound. It can be seen that this method of presenting uncertainty might feel more intuitive to end-users as it represents how well the algorithm is expected to work when assessing a case with clean data and how well it works when assessing a case with

noisy data. In this specific study, Figures 5.9, 5.10, 5.10 and 5.11 clearly show that this method can be more modest, for example, it exhibits greater uncertainty when presenting specificity.

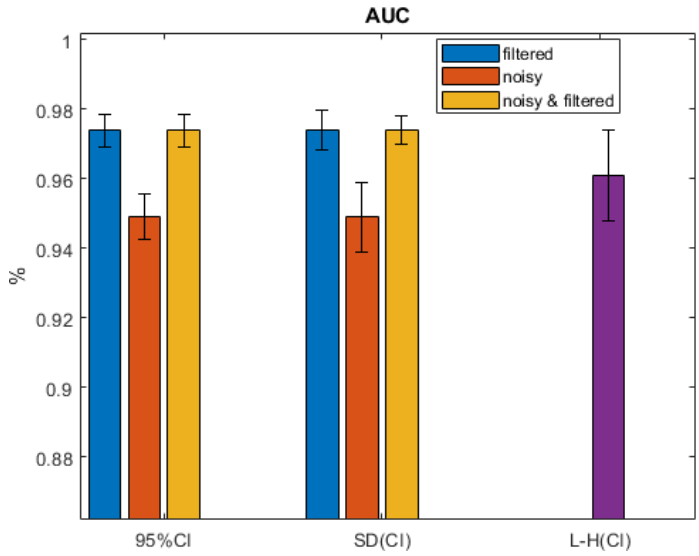


Fig. 5.10 The confidence intervals (CI) of the DL learning performance based on AUC using three different methods including 1) 95% CI, 2) standard deviation CI and 3) the new CI method which is based on the DL performance on the noisy and filtered dataset. The blue bar represents the first scenario, the orange bar represents the second scenario and the yellow bar represents the third scenario, while the purple bar represents the new CI method.

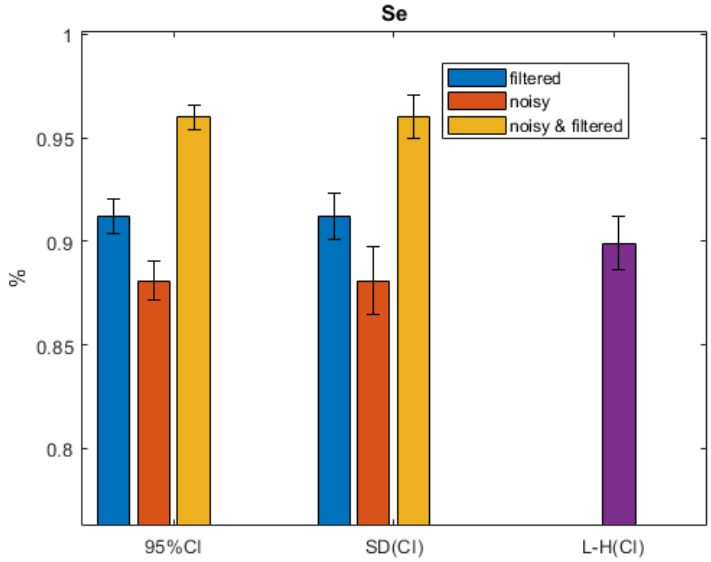


Fig. 5.11 The confidence intervals (CI) of the DL learning performance based on sensitivity using three different methods including 1) 95% CI, 2) standard deviation CI and 3) the new CI method which is based on the DL performance on the noisy and filtered dataset. The blue bar represents the first scenario, the orange bar represents the second scenario and the yellow bar represents the third scenario, while the purple bar represents the new CI method.

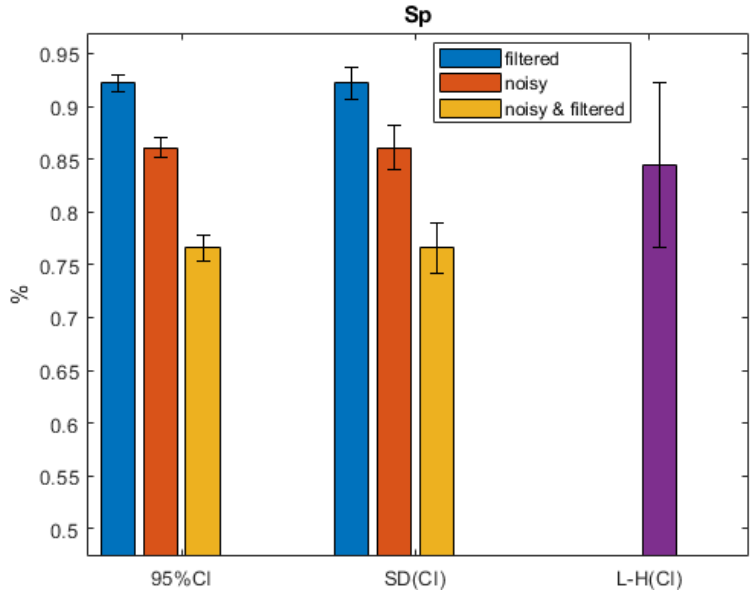


Fig. 5.12 The confidence intervals (CI) of the DL learning performance based on sensitivity and specificity using three different methods including 1) 95% CI, 2) standard deviation CI and 3) the new CI method which is based on the DL performance on the noisy and filtered dataset. The blue bar represents the first scenario, the orange bar represents the second scenario and the yellow bar represents the third scenario, while the purple bar represents the new CI method.

The new CI type was combined with the other CI types in Figure 5.13 to show the tightness of each CI based on accuracy.

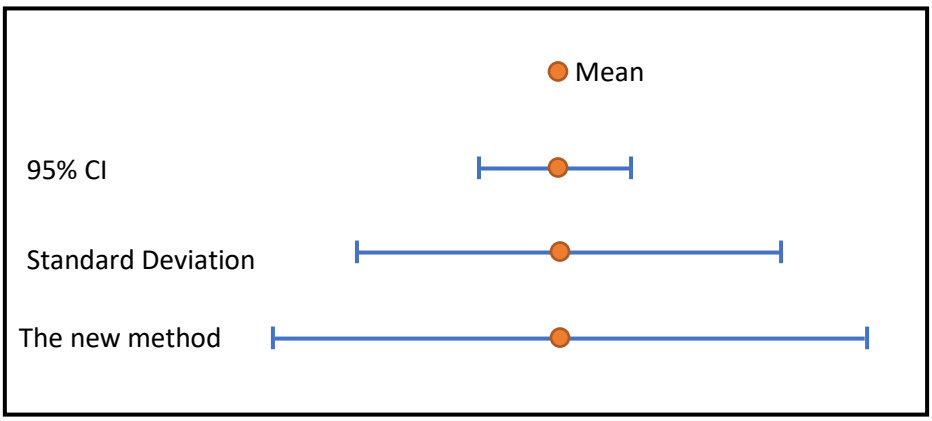


Fig. 5.13 The new CI combined with the traditional CI.

5.4 Discussion

DL obtained a high accuracy (92.0%) when discriminating between PVC and non-PVC complexes. According to the results section, using a filtered dataset for training a model as shown in the first scenario improved the DL performance significantly ($P<0.001$) when compared to the second and the third scenario. There was no significant difference between the second and the third scenario in terms of

DL performance. Hence, including noisy data in the training process (whether its partially or completely included) led to a decrease in DL performance ($P<0.001$).

The impact of noise was investigated by previous studies that showed that noise can lead to a decrease in ML performance. Uncertainty in ML could happen in different stages during building the algorithm due to data collection process, derive and visualisation as shown in figure 5.14. The new uncertainty visualisation method (which uses a new CI method to show a kind of confidence level) looks reliable as the lower bound of the new CI represents the lowest performance attainable when using the algorithm to make a classification when given noisy data and the upper bound of the CI represents the best performance of the algorithm when it is used to make a classification when given filtered data. This kind of CI portrays a level of accuracy that can be expected in the real world where data can be noisy and less than ideal.

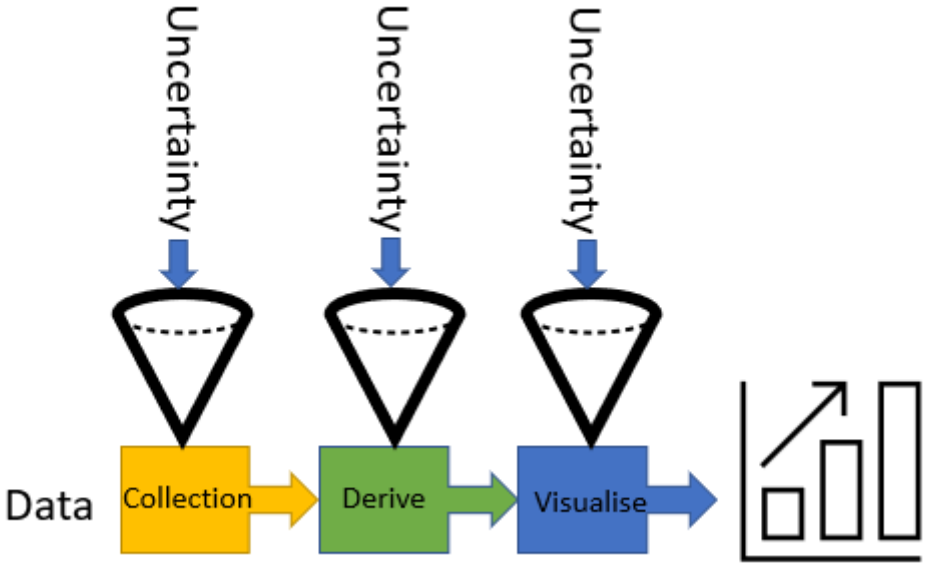


Fig.5.14 Sources of uncertainty in machine learning.

5.5 Conclusion

In this study we investigated the influence of noise on DL performance in ECG classification by evaluating the DL performance in the presence and absence of noisy ECG cases. Hence, three different scenarios were investigated and based on the results in those scenarios, the DL algorithm that was only trained using the clean/filtered data achieved the best performance. This emphasises that the presence of noise significantly influences DL performance and that adding noisey cases to a training set doesn't help the algorithm to generalise. However, an ML algorithm should be at least tested on noisey data.

We also recommend that data scientists present different types of confidence intervals, including the performance of the algorithm when reading clean data and

when reading noisy data. This is important since many real-world applications involve real world data that are often noisy and less than ideal.

Acknowledgement

We thank Professor Dewar Finlay for preparing the dataset that we used in this chapter and for his advises.

Chapter 6

Cardiac Biomarkers and ECG Analysis to Detect STEMI Patients

6.1 Overview

Cardiovascular disease (CVD) is the leading cause of death worldwide killing 18 million people (WHO, 2011). Coronary Artery Disease (CAD) the largest contributor of CVD (Eur Cardiovas Disease Statistics, 2012), accounts for over a quarter of all deaths in N. Ireland (NICHS, 2016). CAD can result in an acute coronary syndrome (ACS) or can present as stable angina. There is an urgent clinical need to develop better ways to diagnose acute coronary syndromes (ACS) that result from CAD. ACS includes patients with ST-elevated myocardial infarction (STEMI), Non-ST elevated myocardial infarction (NSTEMI) or patients with Unstable angina (UA). Currently ACS is diagnosed based on the 12-lead electrocardiogram (ECG). Diagnosis using Troponin blood testing which confirms damage to the heart muscle is then used to indicate a diagnosis of acute myocardial infarction (AMI) which included both STEMI and NSTEMI. Although the ECG is the first and only diagnostic option, recent work has shown that interpretation of the 12-lead ECG is complex and that clinicians are sub-optimal in their interpretation of the ECG and which can result in errors of significant importance. ECG diagnosis of ACS has variable sensitivity (3.1%-92%) and specificity (35%-99%) (Barstow et al., 2017; Boersma et al., 2000; Sakamoto et al., 2016; Yoon et al.,2017; Califf, 2018; Strimbu et al., 2010). Additionally, troponin blood tests are timely and need to be repeated over time. Diagnosis is confirmed and treated with an invasive strategy such as cardiac catheterization (Kumar et al., 2009). Hence, there is a clear clinical need therefore to improve ACS diagnosis. Machine learning (ML) approaches have previously identified protein and ECG biomarkers with potential to augment traditional risk factors in cardiovascular disease (CVD) diagnosis (Weng et al., 2017; Alaa et al., 2019. In the current study, we aimed to identify biomarkers that could differentiate between patients with ACS, stable coronary artery disease undergoing elective (ELEC) coronary angiography and a low risk (LR) group with no risk factors of CVD as defined by HeartSCORE (European Guidelines, 2016).

In the previous chapters, ML was used to improve ECG data quality by detecting lead misplacement in 12-lead ECG before diagnosis in cardiac care unit. While in this chapter, ML was used to investigate the utility of novel blood protein biomarkers alone and in combination with ECG biomarkers to improve the diagnosis of ACS. In the first experiment, ML and blood biomarker dataset were used to discriminate between two public groups including a very high-risk (VHR) patients group (which includes ACS and elective (ELE) patients) and a low risk (LR) patients group. In the second experiment, ML and the blood biomarkers dataset were used to discriminate between ACS patients and non-ACS (ELE and LR). In the third experiment, ML, the blood biomarkers dataset and the ECG dataset were used to discriminate between STEMI and NSTEMI patients. Figure 6.1 below shows the general structure of this chapter six.

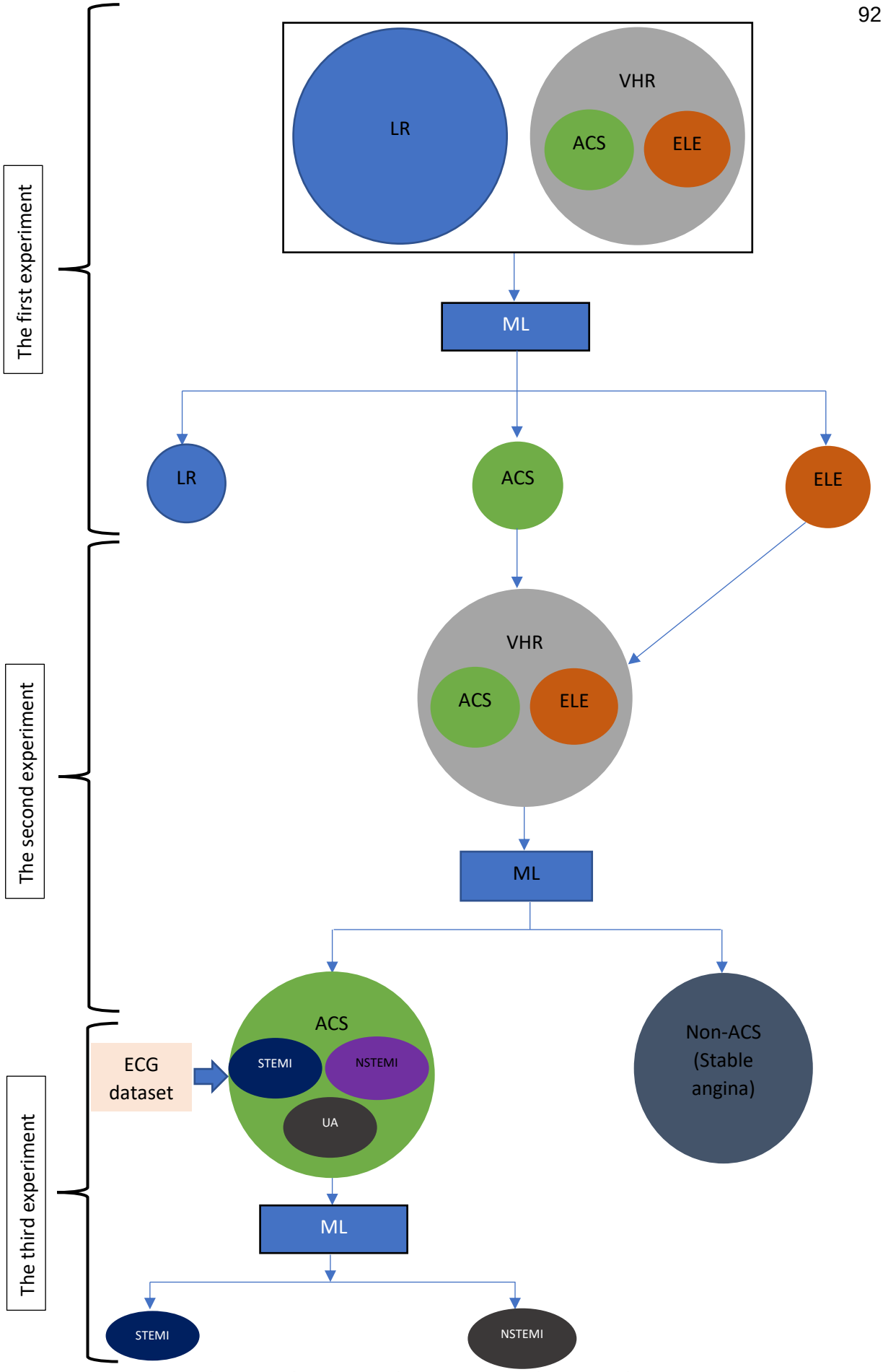


Fig. 6.1 The general structure of the experiments.

6.2 The first experiment

6.2.1 Material and Methods

Data collection:

A data set from a CVD patient cohort previously recruited as part of the CVE Risk study was used in this investigation. The previous study collected peripheral blood samples from consecutive patients attending for cardiac catheterization l at Altnagelvin hospital, Northern Ireland from 2014 to 2017. Extracted plasma was then analysed using multiplex proteomic analysis (Olink).Protein biomarkers (n=327) were acquired from two different groups: group 1) very high risk (VHR) group (mean age=65.3±10.6, 150 males,38 females) which includes acute coronary syndrome (ACS) patients (n=91) including those with STEMI, NSTEMI or UA, 2) stable angina elective (ELEC) patients (n=97), and 3) low risk (LR) group (n=61(12 males, 49 females), mean age=41.5±9.3) who had ten years fatal CVD risk < 10% according to the European systematic coronary risk evaluation (SCORE) (Table 6.1).

Table 6.1 Patients groups distribution

	Gender		Mean age
	Males	Females	
VHR	N=150	N=38	65.3±10.6
LR	N=12	N=49	41.5±9.3
VHR: very high risk, **LR**: low risk			

Classification:

The classification was subdivided as follows: 1) classifying ACS vs. ELEC, 2) classifying ACS vs. LR and 3) classifying ELEC vs. LR, see Figure 6.2.

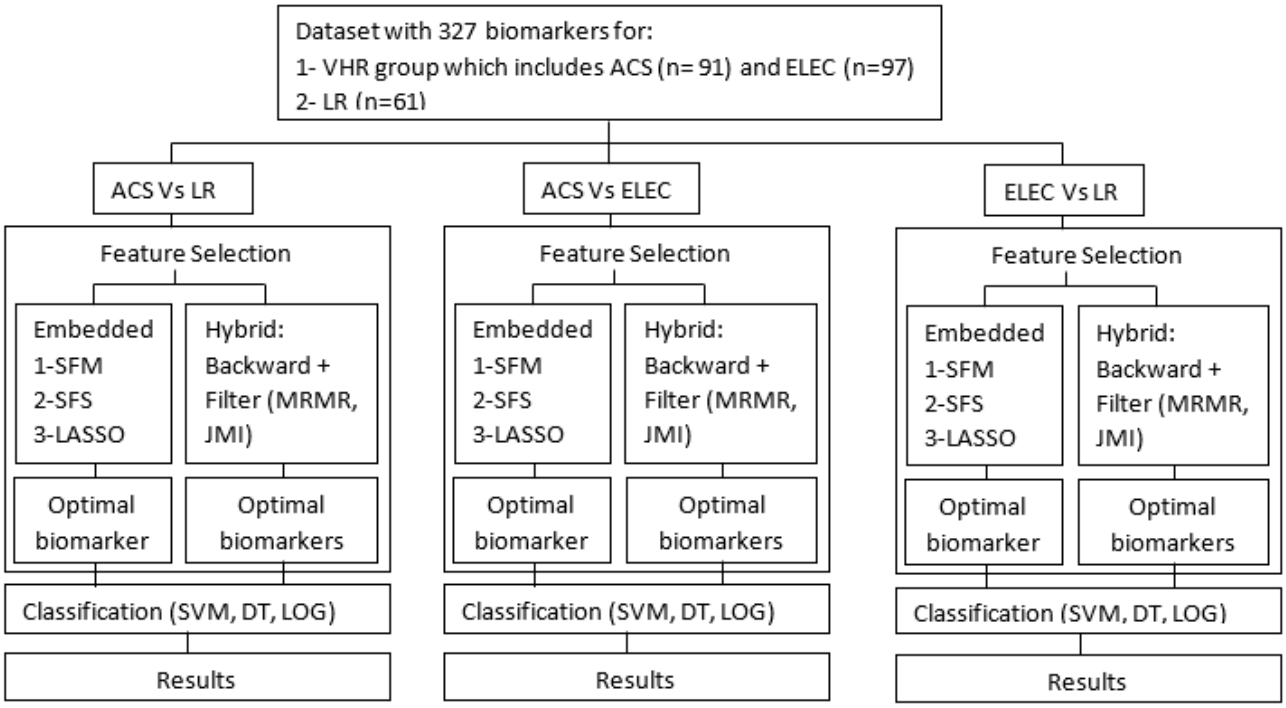


Fig. 6.2 Block diagram to recap the whole work in this study.

As shown in Figure 6.2, in each sub-study, to select the most important and relevant biomarkers out of 327 biomarkers; two different feature selection approaches were applied: an embedded method that selects the most relevant features during the training process and a hybrid method which selects the features before the training process. After feature selection and finding the most important features, three machine learning (ML) classifiers were applied: support vector machine (SVM) with linear kernel function, decision tree (DT) and logistic regression (LOG).

ACS vs ELE study:

A total of 188 patients (ACS n=91, ELEC n=97) were included with 327blood protein biomarkers. A hybrid feature selection approach was applied which included filter method and backward method. Hybrid approach starts with filter method firstly to rank features from the most important to the least important feature. Two algorithms were applied in filter method including joint mutual information (JMI) and maximum relevance minimum redundancy (MRMR) to rank features as shown in Equation (1) in chapter 3. After the filter method, backward elimination was applied by removing feature by feature from the least to the most important ranked features in features list until feature removal stopped improving the performance. Then three ML classifiers (SVM, DT and LOG) were applied with 5-fold cross validation using the selected biomarkers to ensure performance.

In the embedded method which applies feature selection during the training process, the dataset was split into 5-folds. In each fold, three embedded feature selection algorithms were applied including 1) select from model (SFM) algorithm which considers feature importance (if the corresponding feature importance value is greater than or equal to a threshold), 2) sequential forward selection (SFS) which is a greedy search algorithm to reduce the feature space from S to K features where K < S as shown in Equation (1) and 3) least absolute shrinkage and selection operator (LASSO) algorithm. Each feature selection algorithm from the three mentioned algorithms was combined with three ML algorithms (SVM, DT and LOG) during training process in each fold to select the most important features. Then, SVM, DT and LOG were applied for classification using 5-fold cross validation.

$$x^{new} = \arg\max \mathrm{J}(x_k + x) \qquad (1)$$
$$x_{k+1} = x_k + x^{new},$$

where x represents the new added feature, x_k represents the selected subset of features and x^{new} represents the new selected feature that maximize the classifier performance, k is the size of the subset.

ACS vs LR study:

In this study, 152 patients (ACS n=91, LR n=61) were included with 327 biomarkers. The hybrid feature selection approach and embedded method were applied as in (ACS vs ELEC study). Then three ML classifiers (SVM, DT and LOG) were applied with 5-fold cross validation using the selected biomarkers.

ELE vs LR study:

For this aim, 158 patients (ELEC n=97, LR=61) were included with 327 biomarkers. The hybrid feature selection approach and embedded method were

applied as in the previous two studies. Then three ML classifiers (SVM, DT and LOG) were applied with 5-fold cross validation using the selected biomarkers.

6.2.2 Results

Three ML classifiers were applied after each feature selection method with 5-fold cross validation to identify the best ML algorithm and the best feature selection method. Each scenario's results were analysed separately as follows:

Differentiating between patients with ACS and elective patient:

Using the hybrid method, 20 biomarkers were considered as important biomarkers out of 327 to discriminate between ACS and ELEC patients including ('OPN','IL_6','OPG','MB','CCL15','TFPI','HGF','TIE2','PRSS8','IL16','LPL','AMBP','ADAM.TS13','CTRC','CDH5','VEGF.D','PIgR','PD.L2','MARCO','Protein.BOC'). Subsequently, three ML classifiers SVM, DT and LOG were applied with 5-fold cross validation to classify patients. SVM achieved the best performance according to AUC (0.8) with 95% CI [0.74,0.85] and accuracy (81.0%) with 95% CI[75.3%,86.6%] as shown in table 6.2.

Table 6.2 ML classifiers and feature selection methods performance in each scenario.

	Feature Selection	ML	ACS vs ELEC		ACS vs LR		ELEC vs LR	
			AUC	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Embedded Method	**SFM**	**SVM***	**0.82**	**83.0%**	**0.96**	**96.7%**	**0.96**	**96.8%**
		DT*	0.66	66.0%	**0.99**	**99.3%**	**0.98**	**96.8%**
		LOG	0.50	52.0%	**0.96**	**96.7%**	**0.96**	**96.8%**
	LASSO	**SVM**	**0.81**	**82.0%**	**0.96**	**96.7%**	**0.96**	**96.8%**
		DT	0.60	61.0%	**0.96**	**96.0%**	**0.98**	**97.4%**
		LOG	0.57	57.0%	**0.96**	**96.7%**	**0.94**	**94.8%**
	SFS	**SVM**	0.62	62.0%	**0.96**	**96.7%**	**0.96**	**96.8%**
		DT	0.60	60.0%	**0.99**	**99.3%**	**0.90**	**91.1%**
		LOG	0.55	54.0%	0.50	59.8%	0.66	70.8%
Hybrid Method	**Filter method + backward algorithm**	**SVM**	**0.80**	**81.0%**	**0.96**	**96.7%**	**0.96**	**96.8%**
		DT	0.60	61.0%	**0.98**	**98.0%**	**0.95**	**94.9%**
		LOG	0.52	53.0%	0.68	61.3%	**0.92**	**91.7%**

Using embedded method, SFM considered 20 biomarkers as important including ('CDH5', 'PSP.D', 'OPN', 'CCL15', 'Protein.BOC', 'EGFR', 'PRSS8', 'hOSCAR', 'IL.6', 'Gal.9', 'PIgR', 'CTRC', 'LPL', 'MARCO', 'PD.L2', 'VEGF.D', 'PRELP', 'CTSL1', 'AMBP', 'Notch.3') and SFS considered 4 biomarkers ('EIF5A','IL.1.alpha','Protein.BOC','TNF.R2'), while LASSO selected 15 biomarkers as

an important ('ADAM.TS13', 'Ep.CAM', 'PIgR', 'PD.L2', 'CDH5', 'AMBP', 'PSP.D', 'CTSL1', 'LPL', 'KLRD1', 'MGMT', 'CCL15', 'OPG', 'TR.AP', 'OPN'). SFM with SVM classifier achieved the best AUC (0.82) with 95% CI [0.76,0.87] and accuracy (83.0%) with 95% CI [77.6%,88.3%] as shown in table 6.2.

Table 6.3 shows similarity between the different feature selection methods according to the selected biomarkers and it was calculated using Cohen's kappa as shown in Equation (2). The agreement between the feature selection algorithms was predominant in the first scenario (ACS vs ELEC), while in the other scenarios, the agreement was poor. Which means that the selected biomarkers are robust in terms of discriminating between ACS and elective as they have been selected several times by different feature selection algorithms.

According to table 6.2, SFM was the best feature selection method to discriminate between ACS and ELEC using SVM, while LASSO and Hybrid methods were slightly different from SFM and quite similar to each other according to accuracy and AUC. SFS obtained the least performance ($p<0.001$) with (AUC = 0.55) with 95% CI [0.47,0.62] and accuracy = 54.0% with 95% CI [46.8%,61.1%].

$k = 1 - \frac{1-p_0}{1-p_e}$, where p_0 is the agreement among algorithms and p_e represents the probability of chance agreement. (2)

Table 6.3. Agreement between feature selection algorithms in each comparison according to Cohen's kappa.

Feature Selection	ACS vs ELEC				ACS vs LR				ELEC vs LR			
	SFM	**LASSO**	**SFS**	**Hybrid**	**SFM**	**LASSO**	**SFS**	**Hybrid**	**SFM**	**LASSO**	**SFS**	**Hybrid**
SFM	1.00	0.50	0.06	0.63	1.00	0.32	0.07	0.08	1.00	0.43	0.12	0.42
LASSO	0.50	1.00	0.00	0.49	0.32	1.00	0.07	0.08	0.43	1.00	0.13	0.19
SFS	0.06	0.00	1.00	0.06	0.07	0.07	1.00	0.00	0.12	0.13	1.00	0.21
Hybrid	0.63	0.49	0.06	1.00	0.08	0.08	0.00	1.00	0.42	0.19	0.21	1.00

Cohen's kappa value (k) interpretation: 0.01 – 0.20 slight agreement, 0.21 – 0.40 fair agreement, 0.41 –0.60 moderate agreement, 0.61 – 0.80 substantial agreement, 0.81 – 1.00 almost perfect or perfect agreement

Differentiating between patients with ACS and LR patients:

Using hybrid method, 17 biomarkers were considered as important biomarkers to discriminate between ACS and LR patients including ('IL.1.alpha', 'LIF.R', 'IL.15RA', 'DNER', 'IL.12B', 'IL.18R1', 'IL18', 'CXCL10', 'MASP1', 'IL.10RB', 'CX3CL1', 'ITGA11', 'CCL19', 'VEGFA', 'TNFB', 'CASP.8', 'CCL4') and then three ML classifiers SVM, DT and LOG were applied with 5-fold cross validation to classify patients. DT achieved the best performance according to AUC (0.98) with 95% CI [0.95,1.0] and accuracy (98.0%) with 95% CI [95.7%,100.0%] as shown in table 6.2. In embedded method, SFM considered 3 biomarkers as important including ('IL.1.alpha', 'CCL4', 'THBS2') and SFS considered 19 biomarkers ('TNFRSF14', 'MEPE', 'IL.1.alpha', 'LDL.receptor', 'ITGB2', 'IL.17RA', 'TNF.R2', 'MMP.9', 'EPHB4', 'IL2.RA', 'OPG', 'ALCAM', 'TFF3', 'SELP', 'CSTB', 'MCP.1', 'CD163', 'Gal.3', 'GRN'), while LASSO selected 3 biomarkers as an important ('PSP.D', 'IL.1.alpha', 'LPL'). SFM and SFS with DT classifier achieved the best AUC (0.99) with 95% CI [0.97,1] and accuracy (99.3%) with 95% CI [97.9%,100.0%] as shown in table 6.2. Similarity between different feature selection

methods according to the selected biomarkers was calculated as shown in table 6.3. According to table 6.2, SFM and SFS were the best feature selection methods to discriminate between ACS and LR using DT, while LASSO and Hybrid methods were slightly different from SFM and SFS according to accuracy and AUC.

Differentiating between elective patients and LR patients:

In hybrid method, 3 biomarkers were considered as important biomarkers to discriminate between ELEC and LR patients including ('IL.1.alpha' , 'LIF.R' , 'IL.15RA') and then three ML classifiers SVM, DT and LOG were applied with 5-fold cross validation to classify patients. SVM and DT achieved the best performance according to AUC (0.95 with 95% CI [0.91,0.98] and 0.96 with 95% CI [0.92,0.99] respectively) and accuracy (96.8% with 95% CI [94.0%,99.5%] and 94.9% with 95% CI [91.4%,98.3%] respectively) as shown in table 6.2. In embedded method, SFM selected 9 biomarkers as important ('IL.1.alpha', 'IL.15RA', 'TR.AP', 'PSP.D', 'Ep.CAM', 'REN', 'PRSS8', 'MMP.12', 'LIF.R') and SFS considered 6 biomarkers ('TNFRSF14', 'LDL.receptor', 'ITGB2', 'TFF3', 'IL.1.alpha', 'BLM.hydrolase'), while LASSO selected 6 biomarkers as an important ('IL.1.alpha', 'REN', 'CCL11', 'PSP.D', 'PRSS8', 'CCL4'). SFM, LASSO and hybrid method with SVM, DT and LOG classifiers obtained the best AUC and accuracy while SFS obtained the least performance as shown in table 6.2. SFM, LASSO and hybrid method were the best feature selection methods to discriminate between ELEC and LR using SVM, DT and LOG. In each scenario, there is an overlapping between feature selection methods according to the selected biomarkers as shown in figure 6.3.

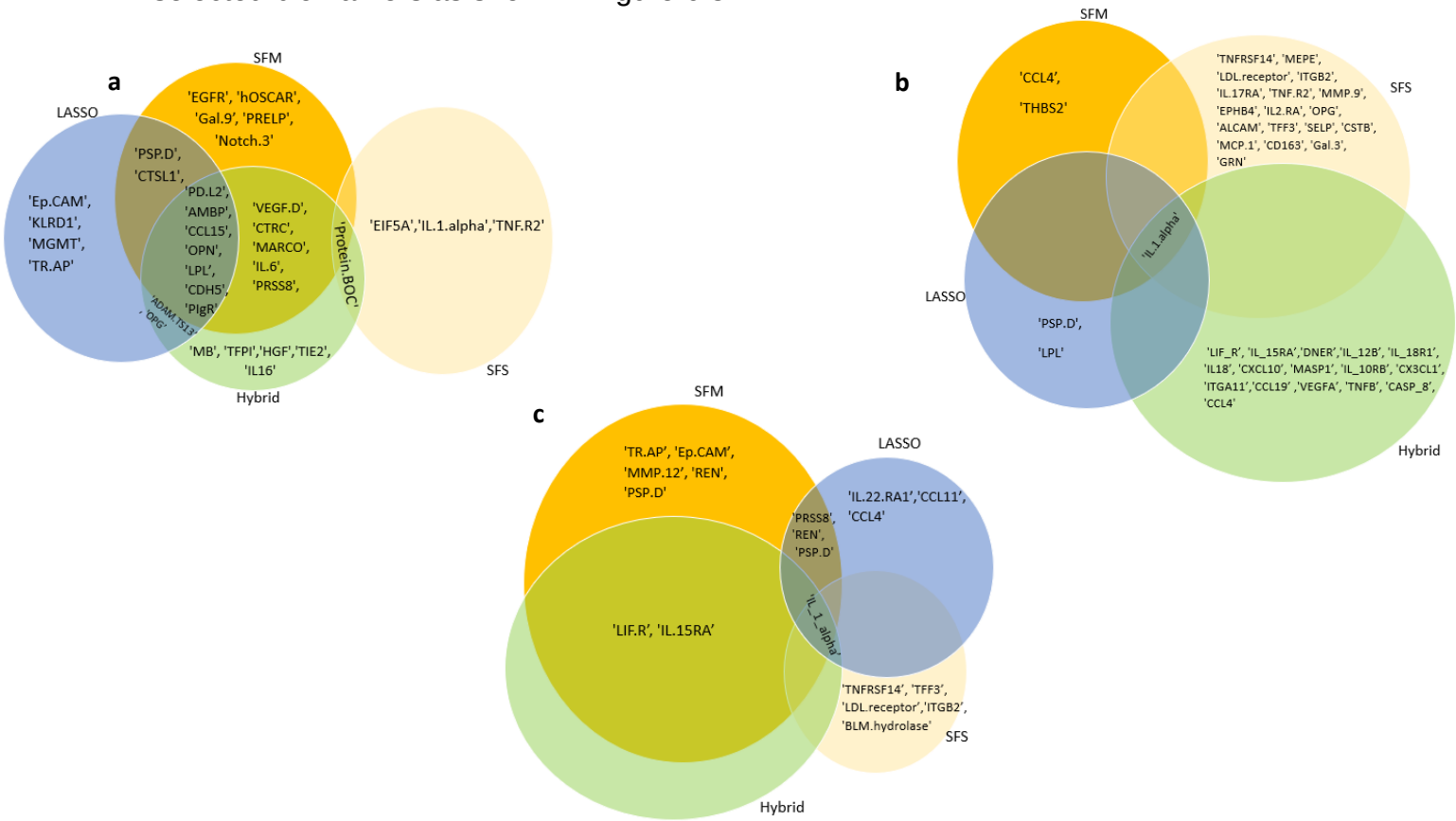


Fig. 6.3 Overlapping between feature selection method in each scenario, a: represents ACS vs ELEC, b: ACS vs LR and c: ELEC vs LR.

Also, there was an overlapping between aforementioned scenarios regarding to the best feature selection methods performance and selected biomarkers ('PRSS8', 'LPL', 'PSP.D', 'Ep.CAM', 'TR.AP', 'IL.1.alpha', 'CCL4', 'IL.24.RA1') as shown in Figure 6.4.

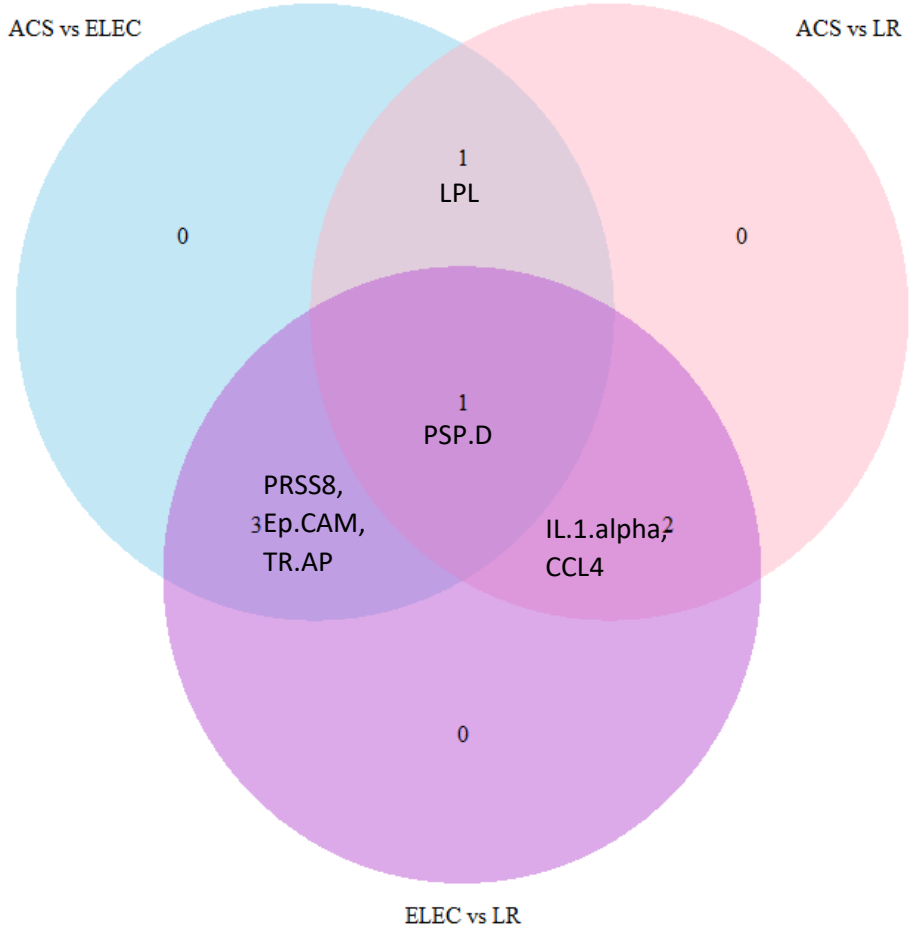


Fig. 6.4 Overlapping biomarkers between aforementioned scenarios.

6.2.3 Discussion

In this study, different ML algorithms and feature selection methods were used to differentiate between three different types of patients ACS, ELEC and LR in three different combinations: 1) ACS vs ELEC, 2) ACS vs LR and 3) ELEC vs LR. This study focused on ACS, ELEC and a LR group, to allow us to identify biomarkers that can differentiate between these patient groups to augment ACS diagnosis.

Three ML classifiers were applied SVM, DT and LOG with two different feature selection methods: embedded (SFM, SFS and LASSO) and hybrid which includes a filter step and backwards step. SFM was the best feature selection method for ACS vs ELEC and ACS vs LR while LASSO and SFM were the best methods in the third for ELEC vs LR. SVM was the best classifier in the first scenario (ACS vs ELEC) while DT was the best in the second scenario (ACS vs LR). In the third scenario (ELEC vs LR), three ML classifiers (SVM, DT and LOG) performed similarly. However, these ML classifiers have strengths and weaknesses, for example DT and LOG are easy to interpret and explain while they may overfit in case of a large number of features.

Meanwhile, SVM is fairly robust against overfitting, while it's not as interpretable as DT and LOG. Hence, different feature selection methods were applied to reduce overfitting, improve accuracy and to reduce training time. The ACS vs LR comparison obtained the highest AUC and accuracy followed by ELEC vs LR and then ACS vs ELEC which means LR can be easily detected among the others using the selected biomarkers and ML classifiers, because in LR group, the selected biomarkers have a quite significant different mean level ($p<0.001$) compared to ACS and ELEC.

Here we have identified a number of well characterised CAD biomarkers, as well as more novel candidates that are drivers of atherosclerotic plaque progression and cardiac remodelling post-acute myocardial infarction. Osteopontin (OPN) is an extracellular matrix regulatory protein, possessing documented correlation with atherosclerosis and prediction of acute events in elective patients [10]. The epidermal growth factor receptor (EGFR) signalling pathway has pleiotropic roles in atherosclerosis progression and in cardiac remodelling [20]. Lipid level changes are associated with the post-myocardial infarction acute phase response [21] and lipoprotein lipase (LPL) is a key lipid metabolism enzyme that may have value in ACS risk stratification [22]. Cadherins play important roles in cardiac remodelling and Cadherin 5 (CDH5) upregulation has been identified at STEMI occlusion sites [23]. PDL2 signalling negatively regulates Th2 cell responses and has a protective role in CVD, with its inhibition in mice being proatherogenic [24]. Proinflammatory cytokines such as IL6 and the downstream functional integrator CRP have been associated with ACS. However, the cytokine repertoire driving atherosclerosis and post-AMI inflammation is recognised to be much more complex [25]. We also identified IL6 and CCL15 as being predictive for ACS over elective patients, likely reflecting the greater inflammatory burden in that cohort. AMBP has been associated with renal tubule damage that is an independent CVD risk factor the elderly [26]. We have identified it as an ACS (over elective) associated biomarker and follow-up work could assess chronic kidney disease (CKD) comorbidity and AMBP protein expression correlation in this cohort. In this study, IL-1α was considered as a very important biomarkers to discriminate ACS from the others (LR an ELEC), which emphasizes the fact that IL-1α is a crucial biomarkers in myocardial infarction detecting [27][28]. In addition, pulmonary surfactant-associated protein-D (PSP.D) is one of the selected proteins and it was used in previous research to assess severity of coronary lesions [29].

This promising pilot study has identified potential ACS biomarkers with biological and clinical support, that could augment ECG in diagnosis and improve its sensitivity and specificity. Clinical uptake of such biomarkers however requires careful validation in multiple large independent cohorts. They must display robust analytical detection characteristics and must be cost effective. Larger cohorts may also be amenable to deep learning and artificial neural networks could capture spatiotemporal complexity within the signalling networks and biomarker interactions with clinical parameters. Furthermore, the ML classifiers used the default values of the hyperparameters, hence, hyperparameters could be tuned for a more robust performance.

6.3 The second experiment

In this experiment, ML was used to discriminate between ACS and non-ACS using 20 selected blood biomarkers.

6.3.1 Material and Methods

Data collection:

The same dataset in the first experiment was used in the second experiment. In this experiment ML were used distinguish between ACS patients and non-ACS patients. The first group includes ACS patients (n=91) and the second group includes non-ACS (n=97).

Feature selection:

The hybrid feature selection method (which combines the filter method and the wrapper method as shown in Figure 6.5) was applied to find a novel set of blood biomarkers that could be used to discriminate between ACS and non-ACS with high performance. The 327 blood biomarkers were ranked using the filter method by applying the joint mutual information formula in Equation 4 to rank the blood biomarkers from the most important to the least important as shown in Figure 6.6. Then, the wrapper method was applied to the ranked blood biomarkers by applying the backward elimination algorithm. The backward elimination algorithm works by removing feature by feature from the ranked dataset until removing a feature that does not improve the performance.

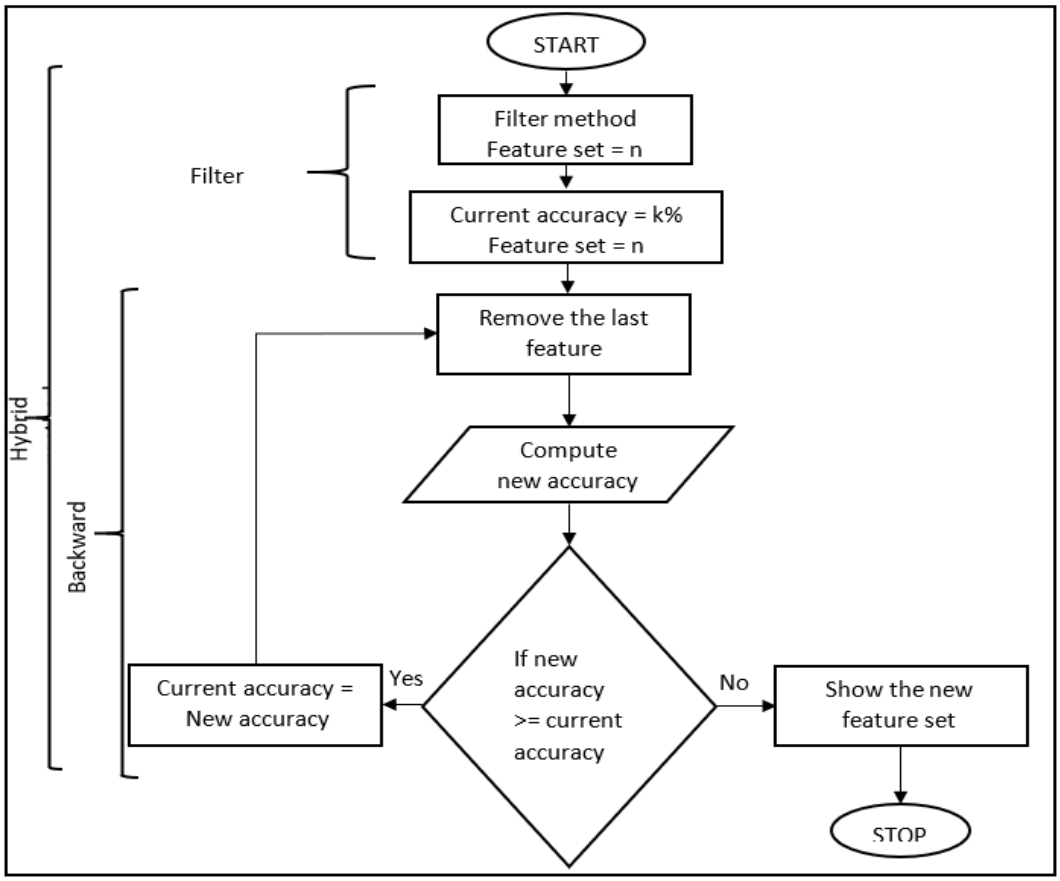


Fig. 6.5 The structure of the hybrid feature selection approaches.

$$f_t = \arg\max \mathrm{I}(x_i; y) - \left[\alpha \sum_{k=1}^{t-1} I(x_{fk}; x_i) - \beta \sum_{k=1}^{t-1} I(x_{fk}; x_i|y)\right] \quad (1)$$

where x represents features and y represents labels

In JMI: $\alpha = \frac{1}{t-1}$ and $\beta = \frac{1}{t-1}$

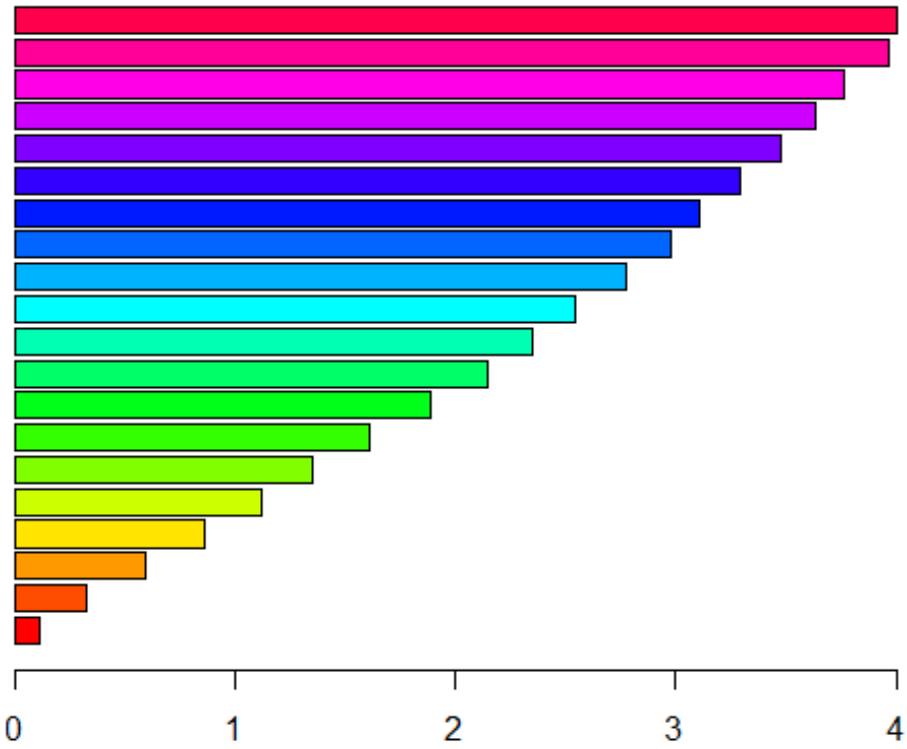


Fig. 6.6 Ranking using joint mutual information feature selection.

Classification:

Two ML algorithms were applied including decision tree (DT) and logistic regression (LR) to discriminate between ACS and non-ACS. The two ML algorithms were trained and tested using 5-fold cross validation as shown in figure 6.7. ML performance was evaluated using accuracy, sensitivity (Se) and specificity (SP) using Equations 2, 3 and 4.

Run 1	Run 2	Run 3	Run 4	Run 5
Training	Training	Training	Training	Testing
Training	Training	Training	Testing	Training
Training	Training	Testing	Training	Training
Training	Testing	Training	Training	Training
Testing	Training	Training	Training	Training
Accuracy 1	Accuracy 2	Accuracy 3	Accuracy 4	Accuracy 5

Average accuracy = (Accuracy 1+ Accuracy 2+ Accuracy 3+ Accuracy 4 + Accuracy 5) / 5

Fig. 6.7 Five-fold cross validation.

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$ Where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

$$Se = \frac{TP}{TP+FN} \quad (3)$$

$$Sp = \frac{TN}{TN+FP} \qquad (4)$$

6.3.2 Results:

Using the hybrid feature selection approach and ML, 20 blood biomarkers were selected out of 327 blood biomarkers as an optimal set of proteins that could discriminate between ACS and non-ACS as shown in Figure 6.8. DT and logistic regression were used to check the efficiency of the selected blood biomarkers to detect ACS patients.

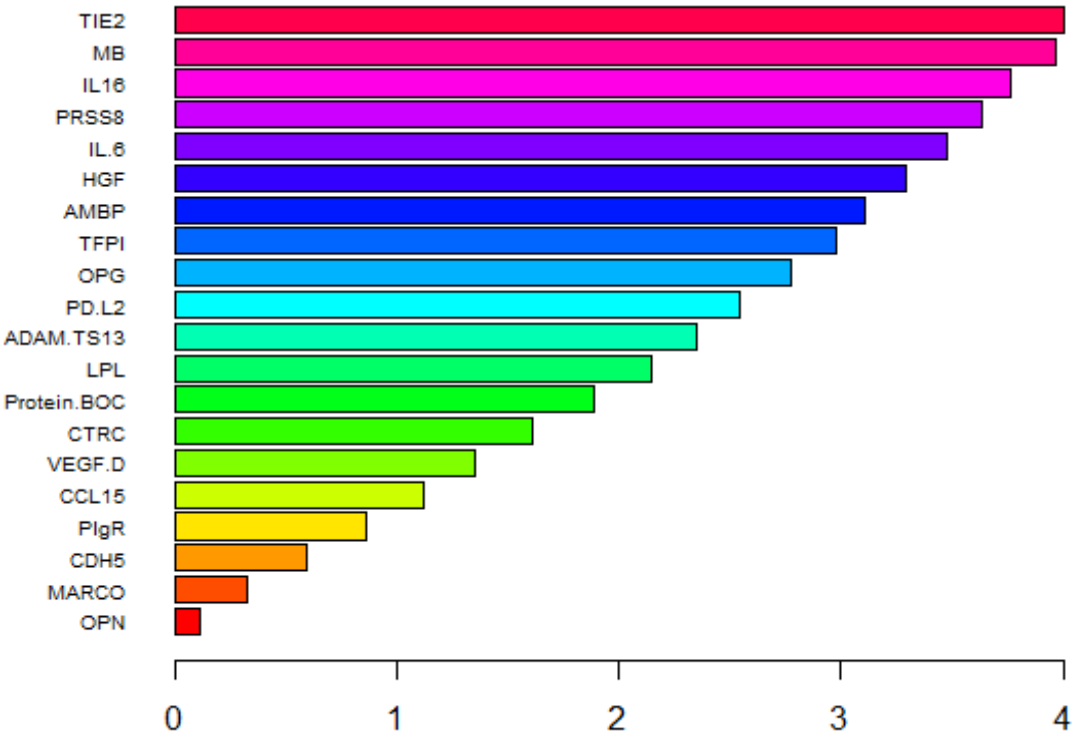


Fig. 6.8 The selected blood biomarkers that obtained a high performance to detect ACS patients.

LR obtained the best performance in terms of accuracy (82.5% [95% CI=0.72, 0.92]) and it outperformed DT significantly (P<0.03). While DT was less accurate (64.9% [95%CI=0.52,0.77]) compared to LR. LR also obtained high area under the curve (AUC) (0.89, [95%CI=0.69,0.9]), sensitivity (85.2%) and specificity (80.0%) compared to DT which achieved 0.59 [95%CI=0.47,0.72] AUC, 60.0% specificity and 70.4% sensitivity. Table 6.4 below shows the performance of the two ML classifiers (DT, LOG). Figure 6.9 represents the confusion matrix of each ML classifier. And Figure 6.10 represents receiver operating characteristic (ROC) curves for LR and DT.

Table 6.4 ML algorithms performance.

ML	Accuracy	AUC	Se	Sp
DT	64.90%	0.590	70.40%	60.00%
LR	82.50%	0.890	85.20%	80.00%

DT: Decision tree, **LR**: Logistic regression, **AUC**: Area under the curve, **SE**: Sensitivity, **SP**: Specificity.

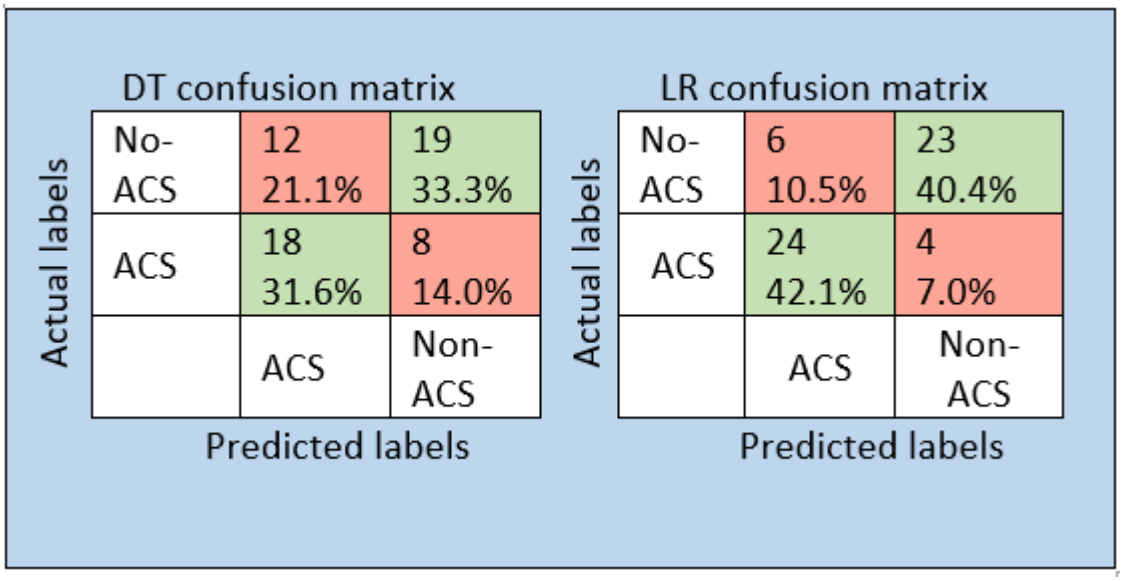


Fig. 6.9 The confusion matrix for LR and DT.

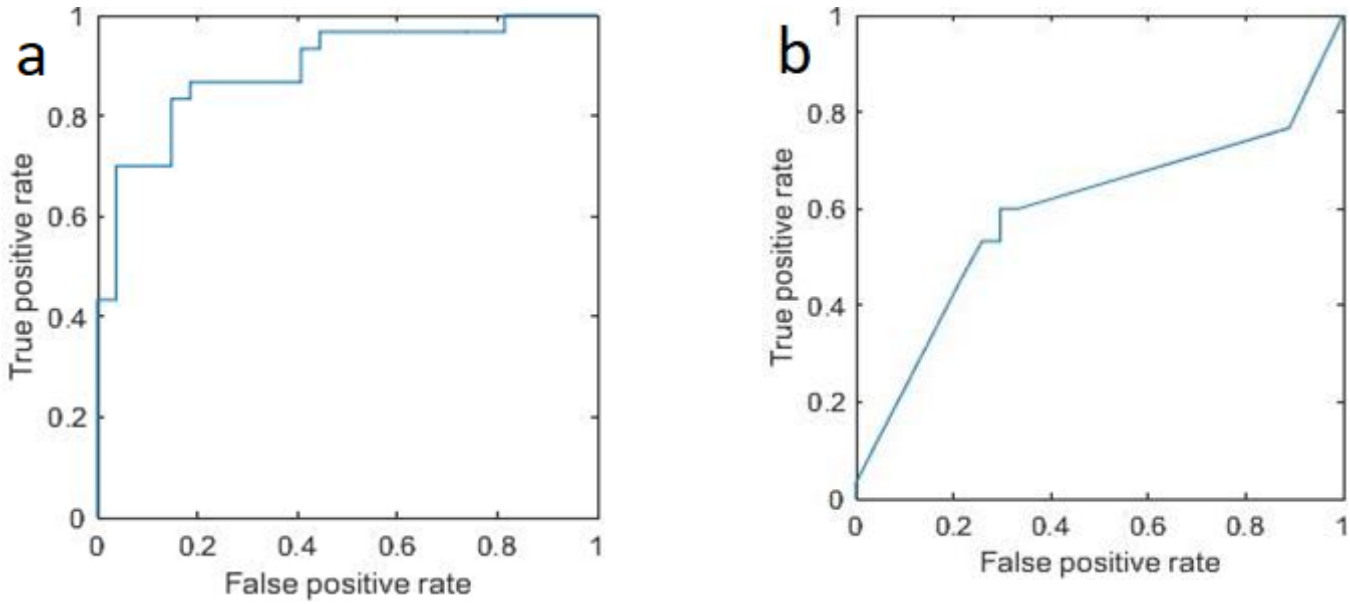


Fig. 6.10 ML ROC curves. **a** represents ROC curve for LR and **b** represents ROC curve for DT. LR AUC=0.89 and DT AUC=0.59.

6.3.3 Discussion:

In this experiment, we found out that ML is able to distinguish between ACS patients and non-ACS effectively using 20 blood biomarkers. Hence, results could be used to improve diagnoses in the cardiac care unit, especially for patients who have ACS. The 20 blood biomarkers were prioritized as the most important features by the joint mutual information feature selection algorithm. The 20 proteins have established roles in myocardial injury, progressive atherosclerotic plaques and inflammation. LR classifier obtained the best performance (accuracy= 82.50%, AUC= 0.89, sensitivity= 85.20% and specificity= 80%) to distinguish between ACS and non-ACS using 20 proteins as

these two types of patients have an overlap according to symptoms. While previous researches obtained less than 0.82 AUC using LR to detect ACS, which concludes that using feature selection methods could improve the performance of ML classifiers. Furthermore, in this study, a large proteomic dataset was analysed which provided an opportunity to identity additional novel blood biomarkers and to assess a wider dataspace. However, new ML algorithms could be involved in future to improve the performance and clinical decision making. Again, more critical discussion- how does this compare to other studies in the literature?

6.4 The third experiment

In this experiment, ML was used to detect STEMI and NSTEMI using the same dataset of blood biomarkers with ECG biomarkers including P-wave, T-wave, S-wave and ST-segment. The aim of this experiment is to find out if combining blood biomarkers with ECG biomarkers could improve the detection of STEMI and NSTEMI.

6.4.1 Material and Methods

Data collection:

The same dataset was used in this experiment, but in this experiment, ECG biomarkers were added alongside the blood biomarkers. Among those patients in the blood biomarkers dataset, 33 patients (NSTEMI n=28 and STEMI n=5) were selected as they have their blood biomarkers alongside the ECGs. The morphological features of the ECG dataset including P-wave, T-wave, S-wave and ST-segment were extracted from the ECG images and added in the final dataset as features as shown in figure 6.11, 6.12 and 6.13.

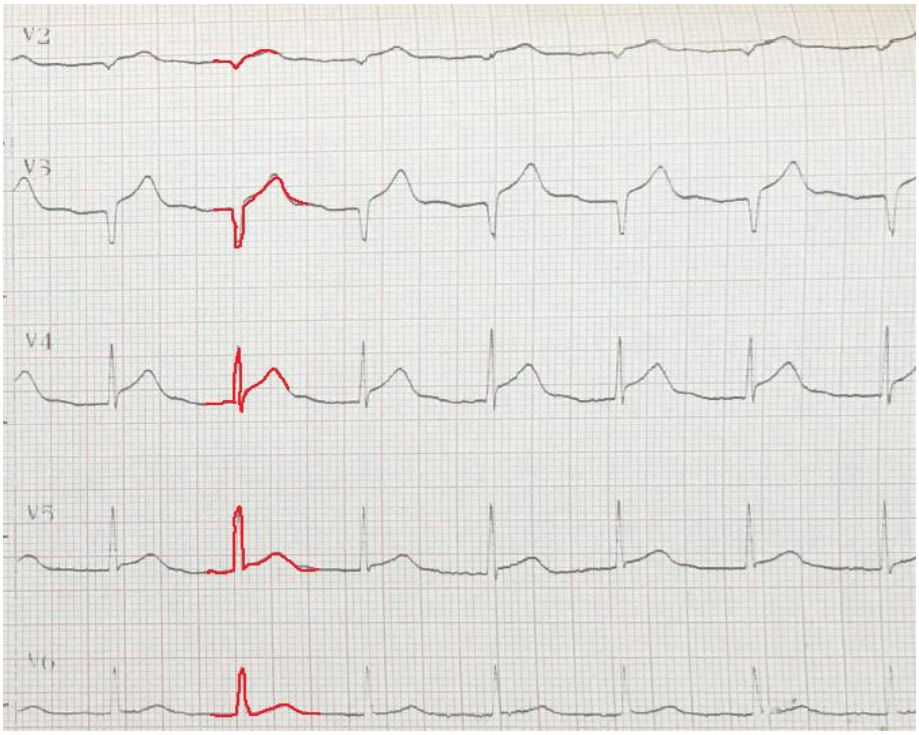


Fig. 6.11 Extracted features from an ECG image of STEMI patient.

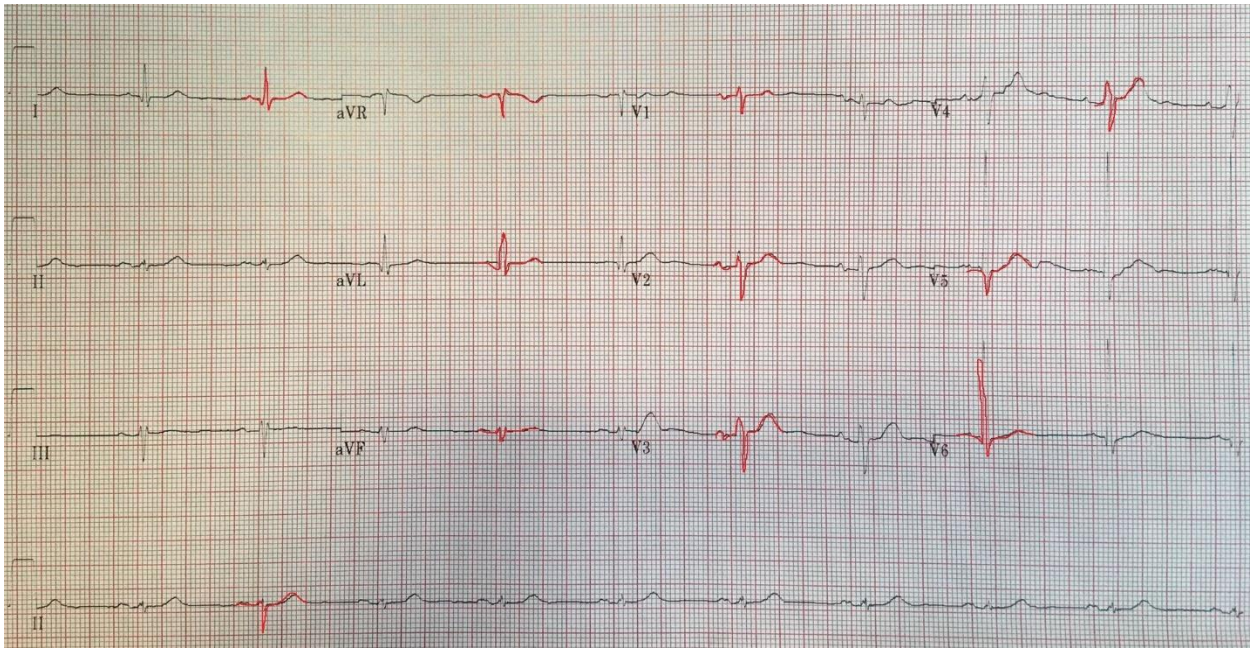


Fig. 6.12 Extracted features from an ECG image of NSTEMI patient.

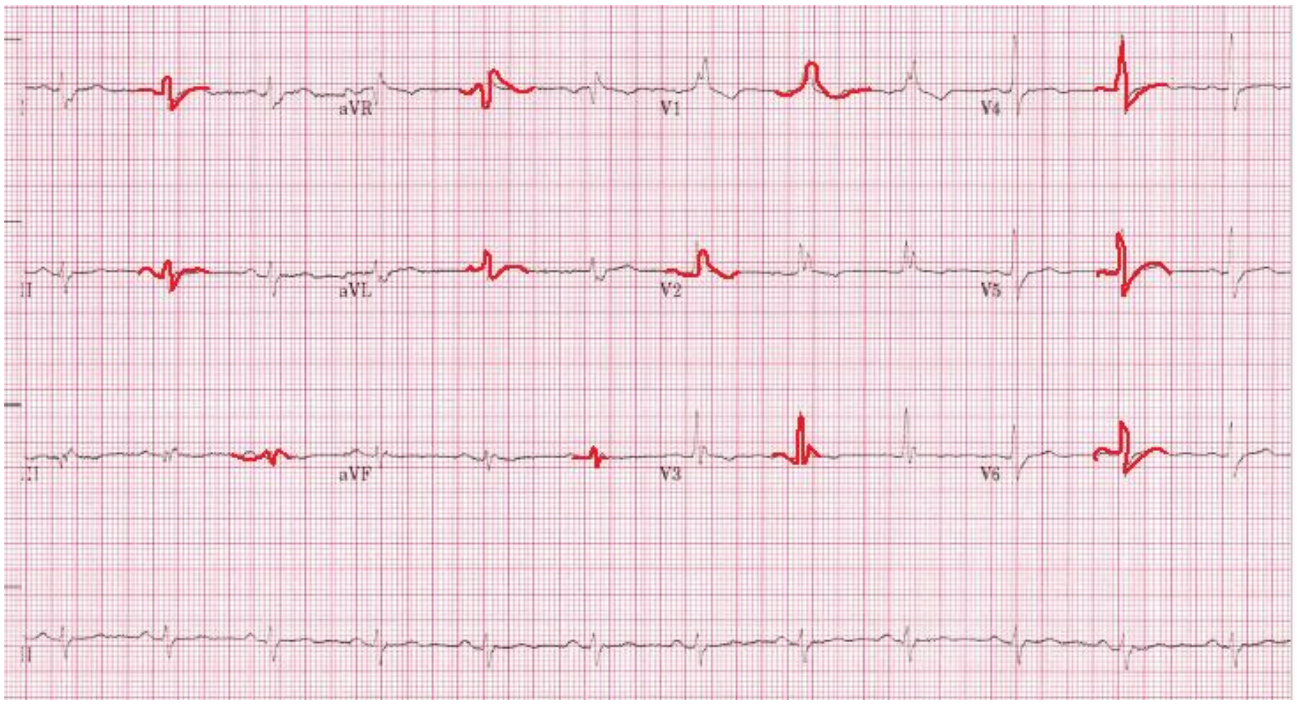


Fig. 6.13 Extracted features from an ECG image of UA patient.

Feature selection:

Two feature selection methods were applied (including sequential forward selection (SFS) and select from model (SFM)) to select to obtain an optimal set of blood biomarkers that can be combined with ECG biomarkers to improve detection of STEMI and NSTEMI. Then, the selected blood biomarkers were combined with the ECG biomarkers in one dataset.

Classification:

Two ML classifiers were applied including SVM and DT. Because the dataset is small, leave one out cross validation (LOOCV) method was applied. In the LOOCV, the model was trained n iteration (n represents the number of samples) and in each iteration, one data point was left out of the training data to test the model using that data point. The performance of ML was tested using three scenarios to discriminate between STEMI and NSTEMI. In the first scenario, ML classifiers were tested using blood biomarkers only. And in the second scenario, ML classifiers were tested using ECG biomarkers only. While, in the third scenario, ML classifiers were tested using blood biomarkers and ECG biomarkers together.

6.4.2 Results:

In terms of feature selection, FSM model selected the following blood biomarkers ('152_Ep-CAM','159_XCL1','CNTNAP2', '132_CCL24', '144_IGFBP-1') as the most important proteins that can discriminate between STEMI and NSTEMI. While, SFS model selected those the following blood biomarkers as the most important blood biomarkers (152_Ep-CAM', '103_ITGB2', '111_ALCAM', '117_Gal-3', '101_TNFRSF14', '105_IL-17RA', '108_EPHB4', '110_OPG', '124_Notch 3','120_MEPE', '106_TNF-R2', '116_CD163', '115_MCP-1', '112_TFF3', '128_TLT-2', '142_IL-6RA', 'CLEC4G'). Figure 6.14 show the overlap between the FSM and the SFS model. 'Ep-CAM' blood biomarker was selected as an important by both algorithms. There is a slight agreement between SFM and SFS (Cohen's k= 0.07)

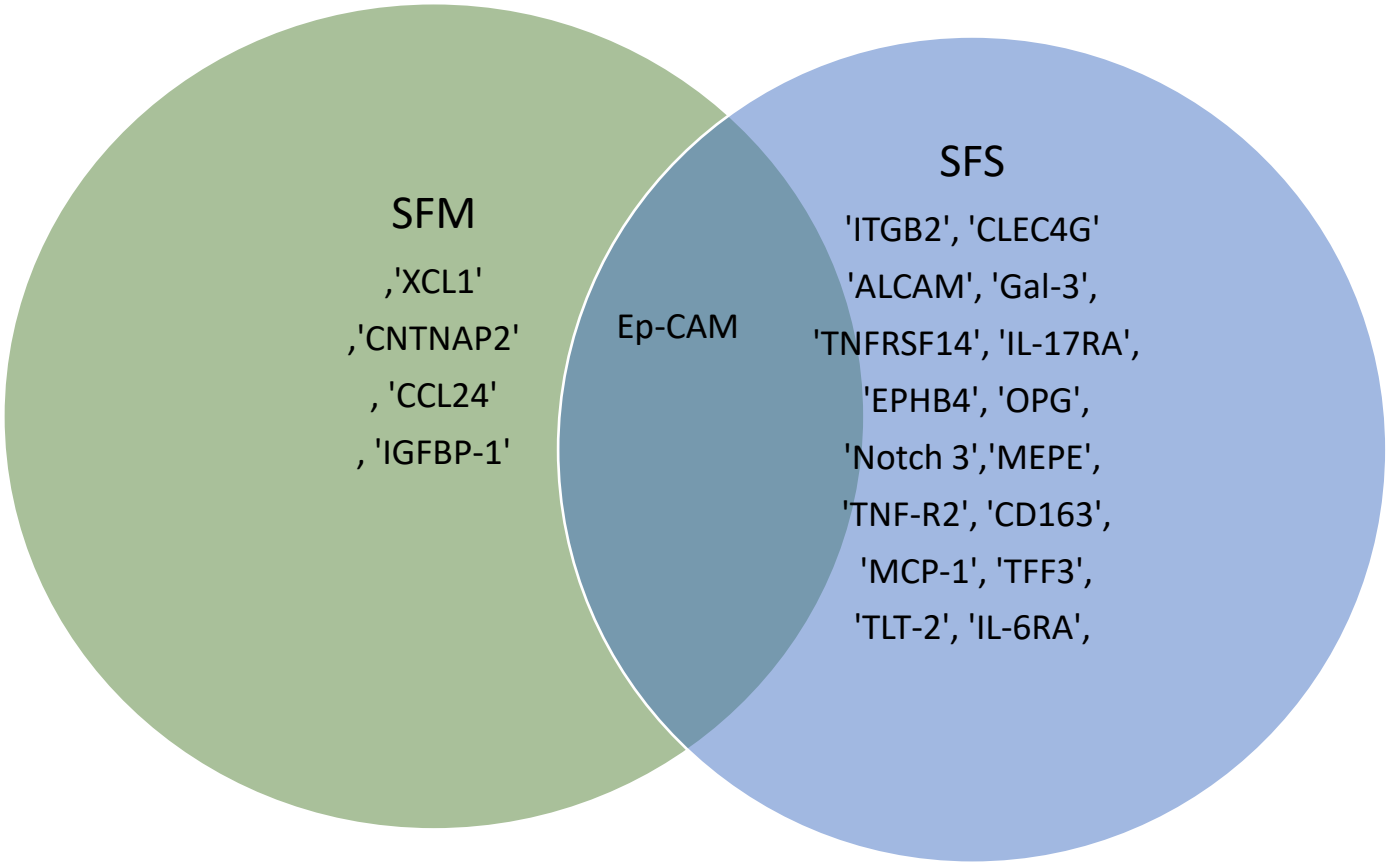


Fig. 6.14 Overlap between the two feature selection methods.

In the SFM model, five blood biomarkers were selected as the most important blood biomarkers. DT and SVM were applied to discriminate between STEMI and NSTEMI using three different scenarios as follow:

The first scenario:

In this scenario, ML classifiers (SVM and DT) were trained and tested using blood biomarkers only. SVM obtained the best performance compared to DT with 100% sensitivity, 40% specificity and 0.7 AUC.

while DT obtained 96.4% sensitivity, 40% specificity and 0.68 AUC as shown in Table 6.5. Tables 6.6, 6.7 below show the confusion matrix of DT and SVM respectively.

Table 6.5 ML classifiers performance for detection STEMI, NSTEMI and UA using blood biomarkers only.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	100%	40%	0.7
DT	96.4%	40%	0.68

Table 6.6 The confusion matrix of DT.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	27	1
	STEMI	3	2

Table 6.7 The confusion matrix of SVM.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	28	0
	STEMI	3	2

The second scenario:

In this scenario, ML classifiers (SVM and DT) were trained and tested using ECG biomarkers only. DT obtained the best performance (82% sensitivity, 60% specificity and 0.71 AUC) compared to SVM (100% sensitivity, 0% specificity and 0.5 AUC).

Hence, DT is more accurate based on AUC as shown in table 6.8. Tables 6.9, 6.10 below show the confusion matrix of DT and SVM respectively.

Table 6.8 ML classifiers performance for detection STEMI, NSTEMI and UA using ECG biomarkers only.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	100%	0%	0.5
DT	82%	60%	0.71

Table 6.9 The confusion matrix of DT.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	23	5
	STEMI	2	3

Table 6.10 The confusion matrix of SVM.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	28	0
	STEMI	5	0

The third scenario:

In this scenario, ML classifiers (SVM and DT) were trained and tested using blood biomarkers and ECG biomarkers. SVM obtained the best performance compared to DT with 100% sensitivity, 60% specificity and 0.8 AUC.

While DT obtained 96% sensitivity, 40% specificity and 0.68 AUC as shown in Table 6.11. Tables 6.12, 6.13 below show the confusion matrix of DT and SVM respectively.

Table 6.11 ML classifiers performance for detection STEMI, NSTEMI and UA using blood biomarkers and ECG biomarkers.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	100%	60%	0.80
DT	96%	40%	0.68

Table 6.12 The confusion matrix of DT.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	27	1
STEMI	3	2

Table 6.13 The confusion matrix of SVM.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	28	0
STEMI	2	3

In the SFS feature selection model, seventeen blood biomarkers were selected as the most important blood biomarkers. SVM and DT were trained and tested using the same three scenarios in the SFM model.

The first scenario:

In this scenario, SVM and DT were trained and tested using blood biomarkers only. SVM outperformed DT according to sensitivity and AUC (96% sensitivity and 0.48 AUC) as shown in table 6.14. Tables 6.15 and 6.16 show the confusion matrix of each ML classifier.

Table 6.14 ML classifiers performance for detection STEMI, NSTEMI and UA using blood biomarkers only.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	96%	0%	0.48
DT	0.82	0%	0.41

Table 6.15 The confusion matrix of DT.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	23	5
STEMI	5	0

Table 6.16 The confusion matrix of SVM.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	27	1
STEMI	5	0

The second scenario:

In this scenario, ML classifiers (SVM and DT) were trained and tested using ECG biomarkers only. DT obtained the best performance (82% sensitivity, 60% specificity and 0.71 AUC) compared to SVM (100% sensitivity, 0% specificity and 0.5 AUC).

Hence, DT is more accurate based on AUC as shown in table 6.17. Tables 6.18, 6.19 below show the confusion matrix of DT and SVM respectively.

Table 6.17 ML classifiers performance for detection STEMI, NSTEMI and UA using ECG biomarkers only.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	100%	0%	0.5
DT	82%	60%	0.71

Table 6.18 The confusion matrix of DT.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	23	5
STEMI	2	3

Table 6.19 The confusion matrix of SVM.

Predicted labels	*Actual labels* NSTEMI	STEMI
NSTEMI	28	0
STEMI	5	0

The third scenario:

In this scenario, SVM and DT were trained and tested using blood biomarkers and ECG biomarkers. DT outperformed SVM according to AUC (DT 0.51 AUC vs SVM 0.5 AUC) as shown in table 6.20. Tables 6.21 and 6.22 show the confusion matrix of each ML classifier.

Table 6.20 ML classifiers performance for detection STEMI, NSTEMI and UA using blood biomarkers and ECG biomarkers.

ML	*Sensitivity*	*Specificity*	*AUC*
SVM	100%	0%	0.5
DT	82%	20%	0.51

Table 6.21 The confusion matrix of DT.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	23	5
	STEMI	4	1

Table 6.22 The confusion matrix of SVM.

		Actual labels	
		NSTEMI	STEMI
Predicted labels	NSTEMI	28	0
	STEMI	5	0

6.4.3 Discussion:

In this experiment, we found that ML achieved a high performance for detection of STEMI and NSTEMI patients. According to results, we found that the third scenario (which combines blood biomarkers and ECG biomarkers) improves the ML performance compared to the second and first scenario according to AUC and sensitivity ((100% sensitivity and 0.8 AUC) vs ((82% sensitivity and 0.71 AUC) and (100% sensitivity and 0.7 AUC)). Hence, blood biomarkers combined with ECG biomarker made ML classifiers more accurate. In terms of feature selection, SFM obtained the best performance compared to SFS method. In terms of ML classifiers, SVM obtained the best performance compared to DT. Hence, suing SFM feature

selection method and SVM classifier with five blood biomarkers and four ECG biomarkers provide the best performance to distinguish between STEMI and NSTEMI.

6.5 Conclusion:

In the first experiment, ML classifiers such as SVM and DT showed promising results for biomarkers analysis to discriminate between three different patients including ACS, ELEC and LR. ML was able to identify novel and relevant biomarkers related to ACS, ELEC and LR patients. Regarding to feature selection methods, SFM, LASSO and hybrid method showed a high performance in cardiac biomarkers selection based on ML classifiers results. ECGs or clinical risk factors were used in previous studies to predict CHD but they were lack sufficient to provide a high AUC to predict CHD [4-8], which means new methods such as laboratory biomarkers could be used to improve AUC of predicting CHD.

In the second experiment, logistic regression showed high performance according to the results for distinguishing between non-ACS and ACS. And in terms of feature selection, JMI improved logistic regression performance by diminishing the number of blood biomarkers to the optimum number and as a consequence, it improves performance, reduces testing and training time and revokes overfitting.

In the third experiment, combining blood biomarkers with ECG biomarkers provides a high performance to discriminate between STEMI and NSTEMI compared to experiments that used blood biomarkers only or ECG biomarkers only. Hence, SFM feature selection method and SVM classifier with five blood biomarkers and four ECG biomarkers provide the best performance to distinguish between STEMI and NSTEMI.

Deep learning was not suitable for those experiments as the dataset is imbalanced and the size is small and this was considered as a limitation. Based on these results, new panels could be developed using these biomarkers such as Olink's panels for more efficient, accurate and better decisions in biomarkers analysis which might help physicians in their decision making in cardiac care unit. Implementing these results in clinical environment could make decision making easier, faster and with low coast as it does not require high-cost procedures such as angiogram. However, future analysis should be applied to improve ML performance by adding more examples or investigating new biomarkers. Furthermore, new deep learning algorithms could be applied in future work such as artificial neural networks (ANNs).

Chapter 7

Discussions, Conclusions and Future Work

7.1 Overview

Clinical decision-making in the cardiac care unit determines the patient pathway and decision aids can help improve the quality of the decisions. However, clinical decision-making in cardiac care could be affected by different factors such as poor-quality recorded ECGs (which might occur due to several reasons such as noise and lead misplacement). Furthermore, the traditional cardiac biomarkers that are used to help diagnose cardiac diseases take a long time to release in the blood such as troponin (> 2 hours) which might result in late diagnoses. Hence, there is a need to design and evaluate other tools that can be used to augment these decisions.

This PhD was carried out to improve clinical decision-making in cardiac care units by attempting to meet the following objectives:

1. Improving ECG data quality by detecting lead misplacement using ML.
2. Showing the impact of noise on automatic ECG interpretation.
3. Finding novel blood biomarkers using ML that could be used in the cardiac care units to make fast and accurate diagnoses.

This PhD was carried out to answer three research question:

7. To what extent can machine learning be used to auto-detect lead misplacement?
8. What useful and actionable insights can we find when we apply Machine Learning (ML) to combine biomarkers?
9. Can new biomarkers be combined with the ECG to improve sensitivity and specificity of ST-Elevation Myocardial Infarction (STEMI) detection?

This chapter will recapitulate how the previous chapters addressed those research questions with contributions achieved and outcomes.

7.2 Improving ECG data quality by detecting lead misplacement using ML.

12-lead ECG lead misplacement is one of different factors that can affect clinical decision making in cardiac care. Lead misplacement could occur in limb leads which is called limb lead reversal such as left arm-left leg (LA-LL) reversal, right arm-left leg (RA-LL) and right arm-right leg (RA-RL) reversal. Furthermore, lead misplacement could occur in chest leads (V1, V2, V3, V4, V5 and V6 precordial leads).

Lead misplacement in chest leads could be in different scenarios such as V1-V6 switching, V1-V6 horizontal displacement and V1-V6 vertical misplacement. Figure 7.1 below shows the different lead misplacement scenarios including chest leads misplacement and limb leads reversal.

Chest electrode misplacement occurs from 40% to 60% of the time and 36% of precordial electrodes (V1, V2, V3, V4, V5, V6) are placed high or wide from the correct position (Bupp et al., 1997).

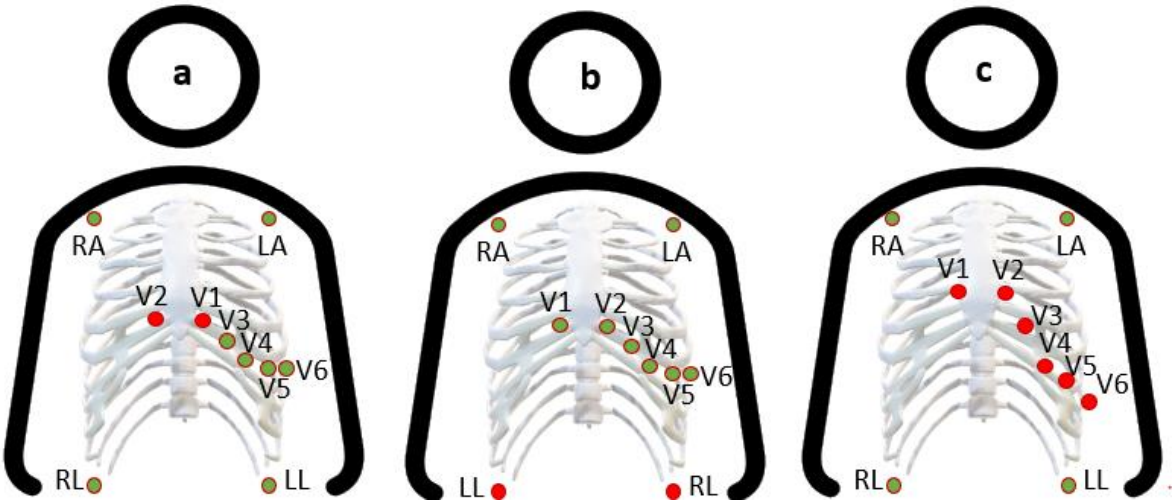


Fig. 7.1 Lead misplacement scenarios. a represents chest leads switching, b represents limb leads reversal and c represents chest lead vertical misplacement.

Lead misplacement has an in impact on ECG morphology and as a result it can affect final diagnoses made by physicians or automated ECG interpreters as shown in Figure 7.2 and Figure 7.3. It could cause false diagnoses of ischemia, ventricular hypertrophy, anterior infarction and Brugada syndrome (Bond et al., 2012) as shown in. Moreover, it could mimic other cardiac abnormalities (see Figure 7.4) or conceal important ECG features (Rudiger et al., 2003; Chanarin et al., 1990; Jowett et al., 2005), which can significantly result in incorrect treatment and as a consequence, it could harm the patient (O'Connor et al., 2010).

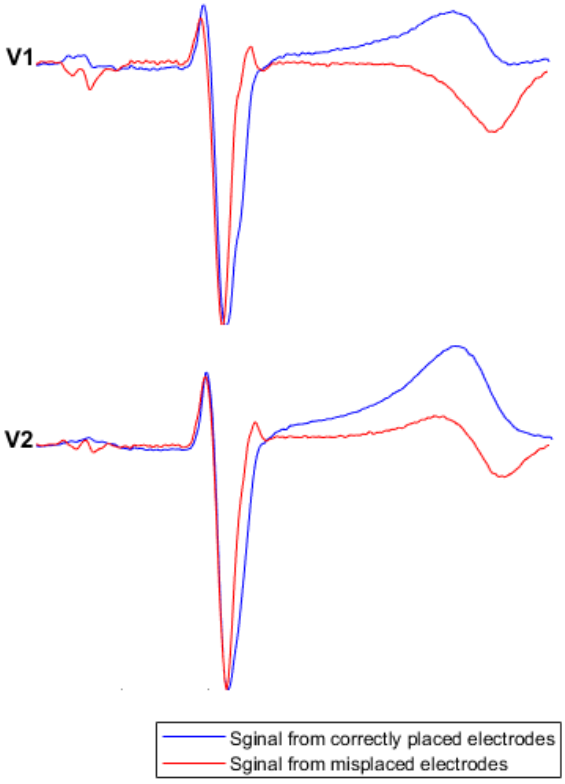
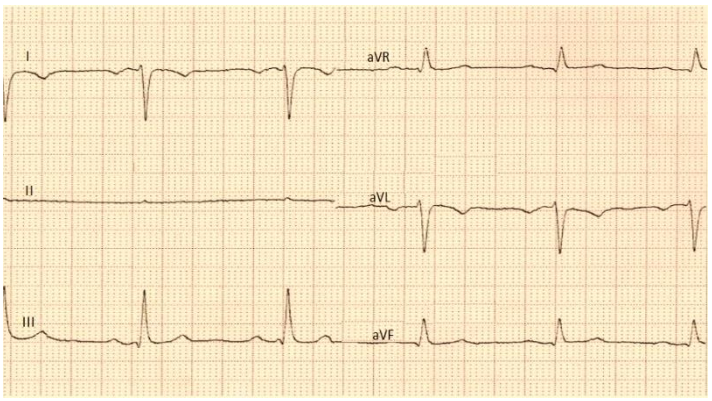
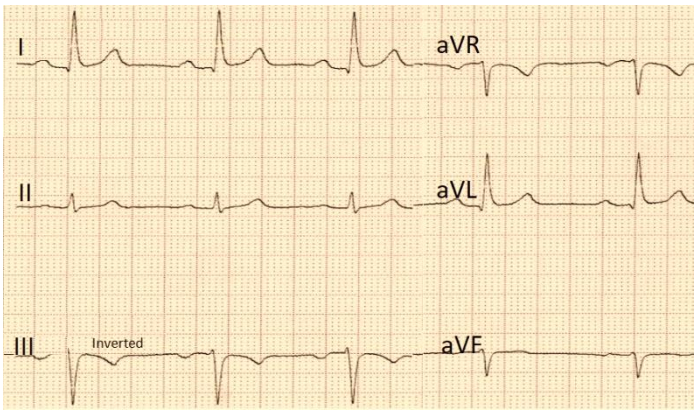
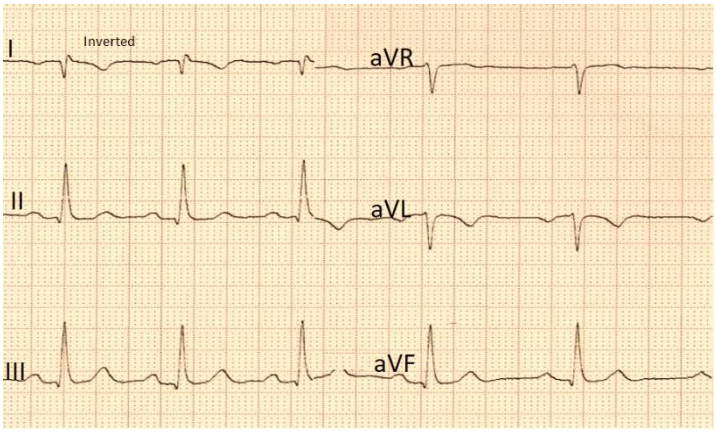
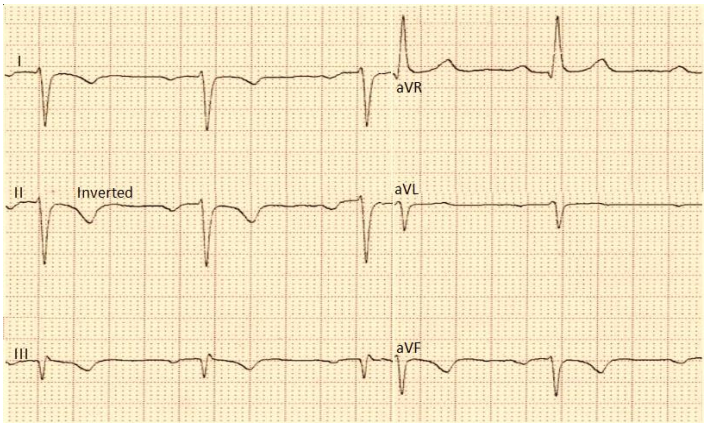


Fig. 7.2 Lead misplacement impact on ECG morphology.









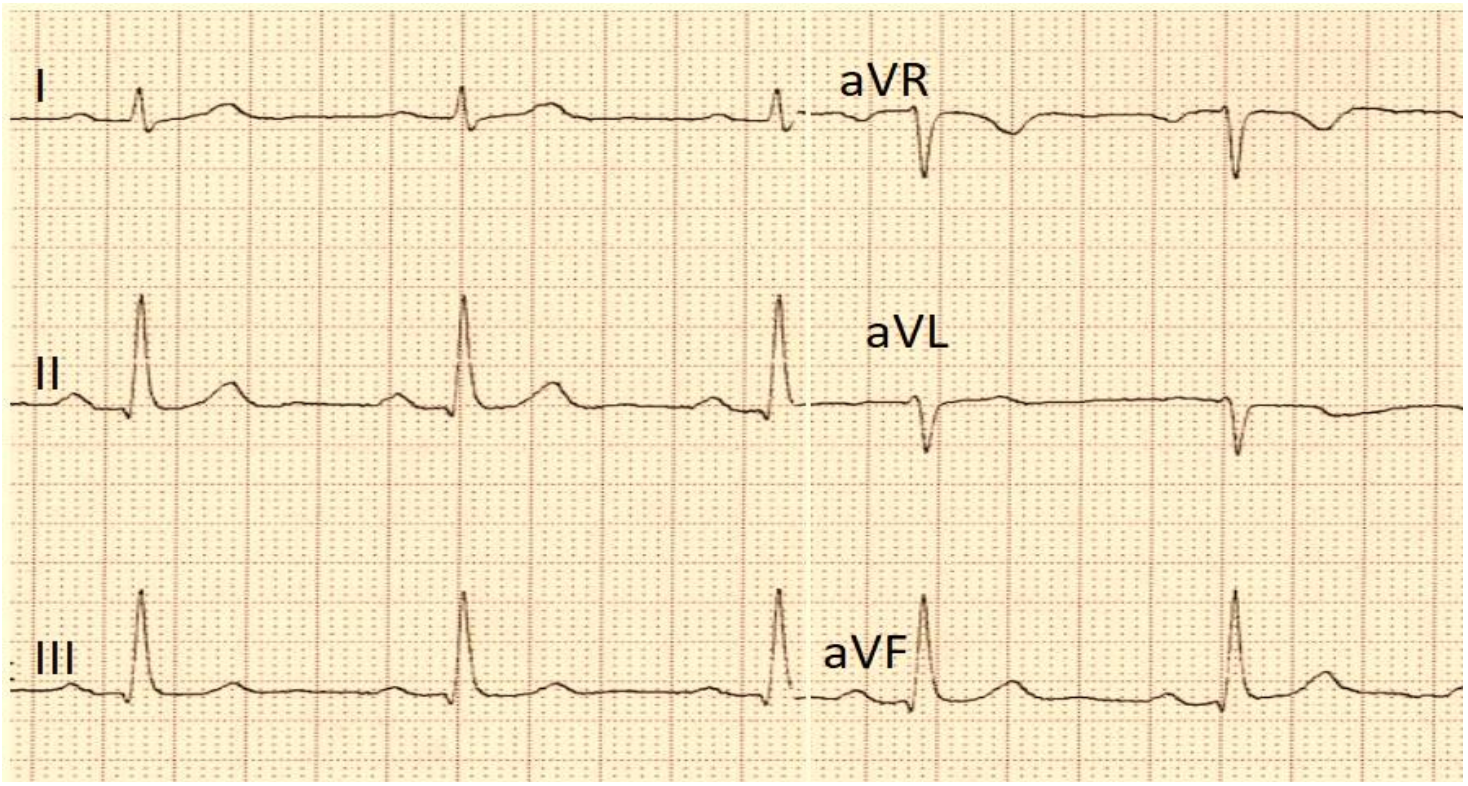


Fig. 7.3 Limb leads reversal impact on ECG morphology. a represents RA-RL reversal, b represents LA-LL reversal, c represents LA-RA reversal, d represents RA-LL reversal and e represents normal baseline ECG.

RA-RL has a significant impact on lead I (lead I becomes inverted) compared to the baseline ECG as shown in Figure 7.3.a. In LA-LL, P wave, QRS complex and T wave are inverted compared to the baseline in leads III, aVR and aVF. According to previous studies that studied limb leads reversal impact on ECG interpretation, they found out that LA-LL is the most challenging limb lead reversal, especially in the absence of a bassline ECG (see Figure 7.3.b). In LA-RA, P wave, QRS complex and T wave are inverted (see Figure 7.3.c). While in RA-LL, all leads are inverted except lead aVL.

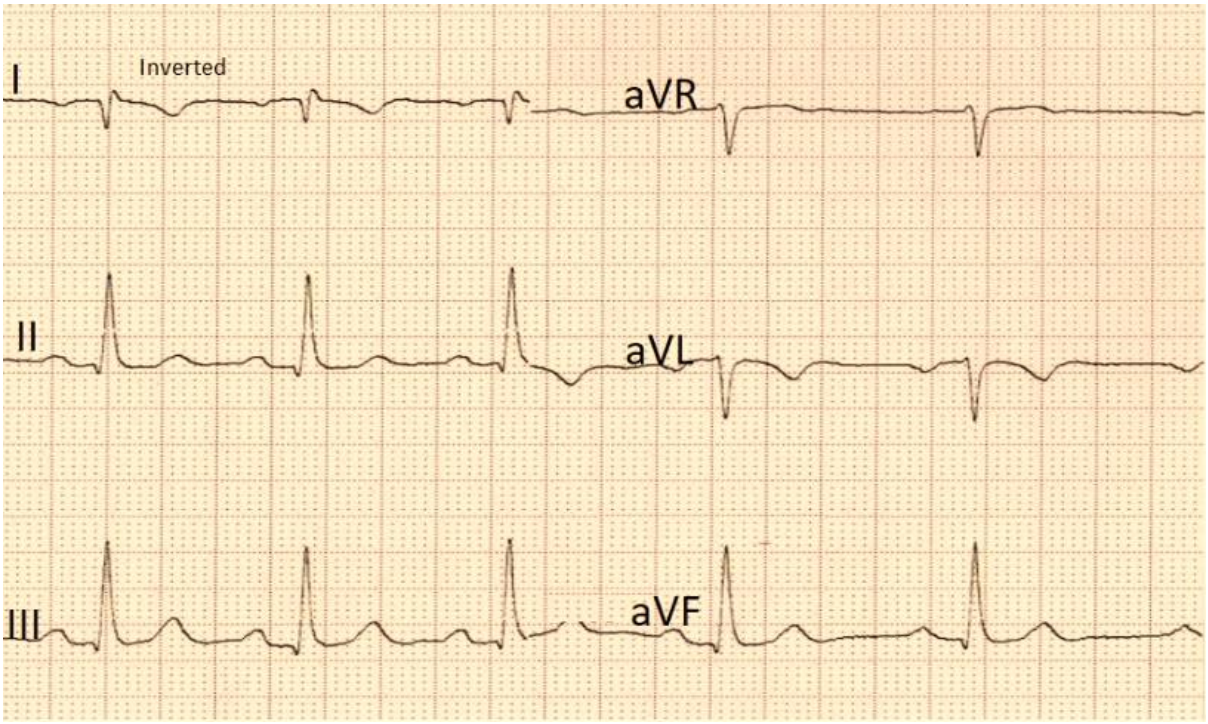


Fig. 7.4 LA-RA limb leads reversal which simulates dextrocardia.

In chest lead misplacement, the most common error misplacement occurs in leads V1 and V2 and they commonly misplaced high and wide of their correct position. Which could lead to misplace the other leads (V3 to V6) (Rajaganeshan et al.,). Lead V2 and V1 misplacement has an impact on the ECG morphology including QRS complex, ST-T-U segment and it could result in misdiagnosis of Brugada (Lateef et al., 2003).

Traditional ML algorithms including support vector machine, logistic regression and decision tree were used to detect limb leads reversal and chest lead misplacement. Traditional ML obtained a good performance to detect limb lead reversal expect LL-LA reversal. Furthermore, lead V1 and V2 misplacement were not investigated using whether traditional ML or deep learning (DL). Hence, this study is the first study that used traditional ML and DL to detect V1 and V2 misplacement. Furthermore, this study used attention maps for the first time to show which morphological features were responsible for making the final prediction. In V1 and V2 misplacement, traditional ML obtained a high performance to detect V1 and V2 misplacement in the first and second intercostal space (ICS), while the performance

declined in the third ICS as those leads were very close to the correct position (4th ICS).

Hence, to improve the performance in the third ICS, DL algorithms including convolutional neural networks (CNNs) and bidirectional long short-term memory network were used for the first time to detect V1 and V2 misplacement. According to the results, DL achieved a high performance in the second ICS and improved the performance in the third ICS significantly (P<0.001) compared to the traditional ML algorithms. DL also showed which features contributed more to the final results as shown in Figure 7.5.

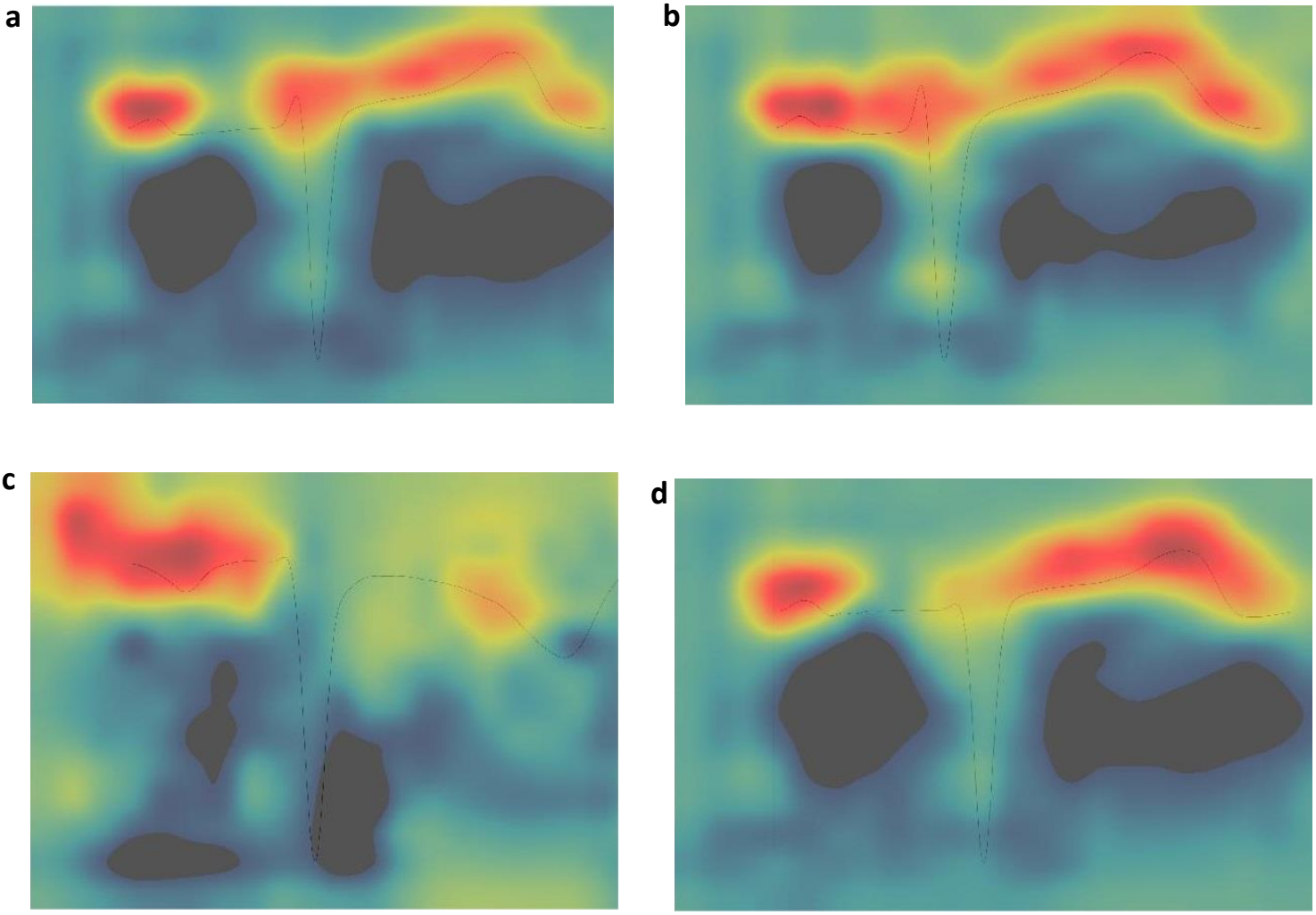


Fig. 7.5 attention maps generated form DL to show the contribution of each feature to the final decision. Those attention maps were generated from the last layer in the DL network as the last layer shows high-level features. a and b represent attention maps of V1 and V2 in respectively in their correct position. While c and d represent attention maps of V1 and V2 in the wrong position respectively. The red colour represents features that significantly contributed to final prediction. While the other colours represent features that did not contribute to the final classification significantly.

Based on those maps, P wave, R wave and T wave have a significant contribution in the final classification, while S wave has no significant contribution. Meanwhile, PR interval, Q wave and J point have a moderate contribution. Hence, physicians and automated ECG interpreters should consider those features to detect V1 and V2 misplacement as they could cause misdiagnoses. Figure 7.6 blow show the summery of using the traditional ML and DL algorithms.

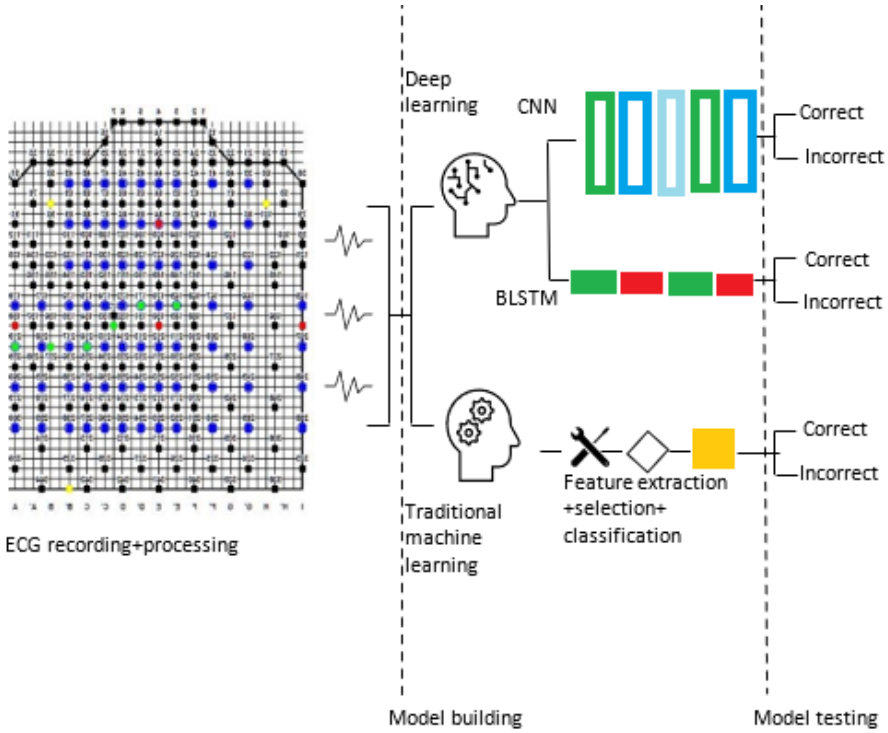


Fig. 7.6 The data pipeline in this study (data engineering, analytics and delivery) to detect leads V1 and V2 misplacement.

Then, the DL algorithms were benchmarked with eleven physicians who are experienced in reading ECG. Based on the results, ML and DL outperformed the physicians significantly (P<0.001) as shown in Figure 7.7.

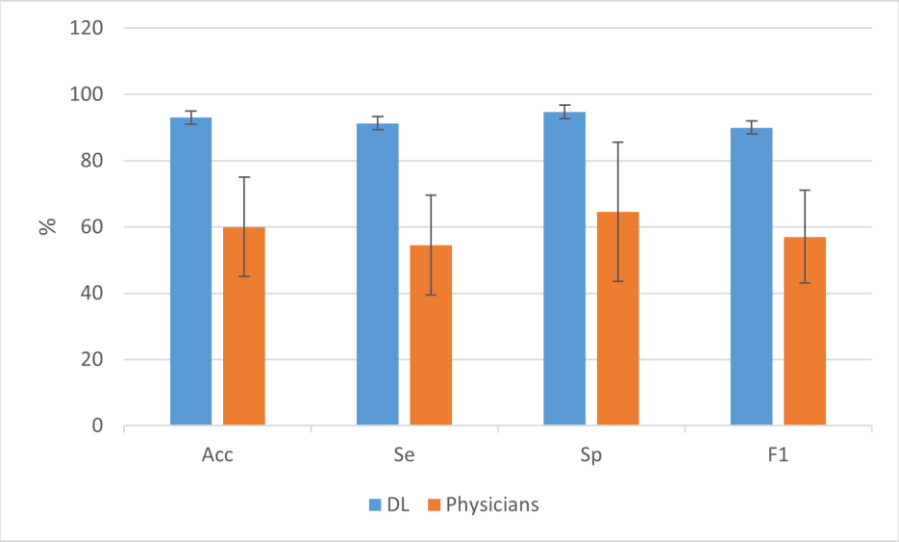


Fig. 7.7 DL algorithm performance compared to physicians.

Hence, physicians in cardiac care units could consider using DL technology to detect lead misplacement before making their final diagnoses.

7.3 The impact of noise on automatic ECG interpretation.

ECG interpretation could be affected by noise (whether it is internal noise such as noise during respiration or from muscles or external noise such as noise comes from medical devices in the same room). Hence, the aim of this experiment is to find the impact of noise on automatic ECG interpretation using DL. Furthermore, a new uncertainty visualisation method was developed. The dataset was filtered using three different digital filters to remove different types of noises as follow:

1-Low pass filter (<80Hz) to remove the muscular noise.

2-High pass filter (>0.5Hz) to remove the respiratory noise.

3-Notch filter (at 60 Hz) to remove the power-line interference

Then, a copy of the filtered dataset was used to inject 50Hz noise to generate a noisy dataset. A Bidirectional long short-term memory (B-LSTM) DL was built to study the impact of noise on ECG interpretation using DL. The DL model were trained suing three different scenarios. In the first scenario, DL model were trained on filtered dataset, in the second scenario, the DL model was trained using noisy dataset, while in the third scenario, the DL model was trained using noisy dataset. In each scenario, the DL model were tested using filtered dataset and noisy dataset.

Based on results, we found out that noise has a high impact on automatic ECG interpretation using DL. Applying digital filters improved DL performance significantly ($P<0.001$, McNemar's). Hence, medical industry could include different types of filters on the software of the ECG devices which improves the quality of ECG images and as a result it could improve automatic ECG interpretation and physicians diagnoses.

A new uncertainty visualisation method was used to show the confidence interval using a non-traditional method. In this new method, the confidence interval is computed based on the DL model performance using the filtered/clean and noisy test sets. The lower bound of the confidence interval represents DL performance using noisy data, while the upper bound of the confidence interval represents DL performance on the filtered data. Hence, in the worst-case scenario, DL performance will not be lower than the lower bound of the confidence interval and it will be within the interval. Figure 7.8 below shows the new method combined with the traditional confidence interval.

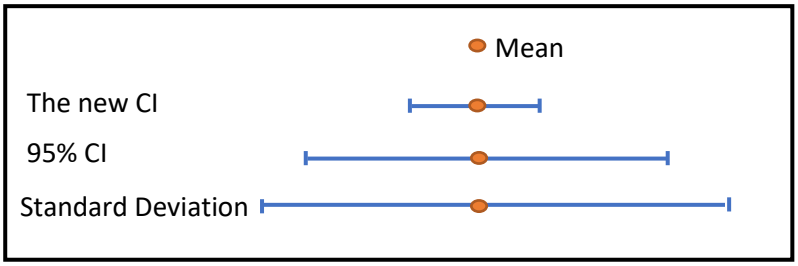


Fig. 7.8 The new uncertainty visualisation method combined with traditional confidence interval method.

Hence, a new application was proposed and conceptually designed by this study for future work to be developed and used in automatic ECG interpretation by physician in cardiac care units as shown in Figure 7.9.

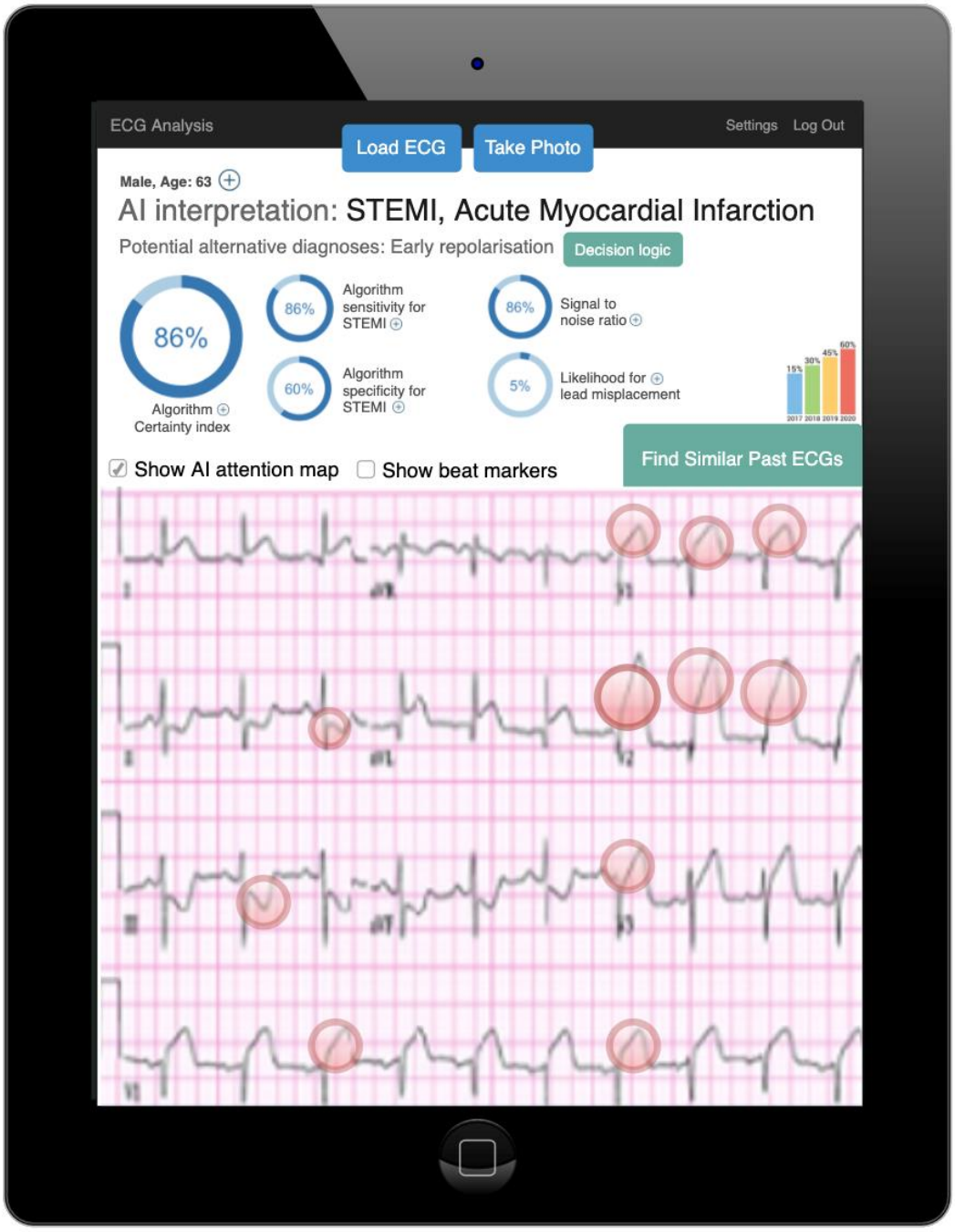


Fig. 7.9 Suggested application interface according to the findings in this study. Physicians can insert general metadata such as patient gender and age. Also, physicians can use prior similar ECGs as a base line to compare it with the current ECG from a local database of ECGs. Attention maps could be involved in the interface to show the most important morphological features that should be considered by physicians in decision making. The interface should include facts about the developed application such as confounding decisions, signal to noise ratio, potential artifacts, rules and an algorithm certainty index.

7.4 Novel blood biomarkers to improve clinical decision-making in cardiac care.

Cardiovascular diseases (CVD) are one of the main causes of death in the world. clinical risk factors including smoking, hypertension, diabetes mellitus and family history are currently used in risk prediction algorithms to derive risk scores for CVD. However, these risk factors have low accuracy to detect acute coronary syndrome (ACS) which is a part of CVD. Troponin is used as blood biomarkers to predict and diagnose ACS. However, troponin takes long time (> 2 hours) to be released in the blood. Previous studies have shown that ML technology can identify blood biomarkers that have the potential to augment these traditional risk factors in cardiovascular disease (CVD) diagnosis. A dataset of blood biomarkers (n=327) and ECGs (n=39) were used to find a group of blood biomarkers to augment decision-making in cardiac care. In the first part of this study, ML and feature selection were applied to find novel biomarkers that could be used to distinguish between very high-risk groups (VHR) and low-risk group (LR). According to each scenario in this part, a small group of blood biomarkers was able to differentiate between ACS, LR and elective patients using ML. Regarding to feature selection methods, SFM, LASSO and hybrid method showed a high performance in cardiac biomarkers selection based on ML classifiers results.

In the second part, ML was used to distinguish between ACS and non-ACS patients in the VHR group. Twenty blood biomarkers were selected as the most important to discriminate between elective and ACS patients. In terms of ML, logistic regression obtained the best performance to differentiate between ACS and non-ACS subjects. The combination of ML and a high-dimensional proteomic data shows promising results to differentiate ACS, ELEC and LR patients with high AUC and accuracy.s

In the third part, ML and blood biomarkers with ECG images were used to discriminate between ST-segment elevation myocardial infarction (STEMI), NSTEMI and unstable angina (UA) in the ACS group. Using biomarkers combined the ML performance compared to using biomarkers only to ECGs only. Hence, five blood biomarkers and four ECG biomarkers were able to distinguish between STEMI, NSTEMI and UA. In terms of ML, SVM and select from the model (SM) features selection were the best method to find novel biomarkers. Hence, new panels could be developed using these biomarkers such as Olink's panels for more efficient, accurate and better decisions in biomarkers analysis which might help physicians in their decision-making. However, future work could focus on distinguishing ACS from ELEC patients, using a bigger cohort of subjects (including blood biomarkers and ECGs), or using new machine learning techniques such as deep learning as it showed promising results in different heart diseases diagnoses such as atrial fibrillation.

7.5 The new decision making for PPCI pathway.

As the aim of this PhD is to improve clinical decision making in cardiac care, a new improved PPCI pathway was proposed according to the main findings and objectives

of this PhD as shown in Figure 7.10. In this pathway, ML could be introduced in critical decisions making steps in the PPCI pathway to make the final diagnosis accurate and to avoid misdiagnoses or false alarm as shown in Figure 7.11. According to the proposed design, the first intervention of ML will be before clinical decision (CD) two. In this step ML will be used to improve recorded ECG data quality by detecting lead misplacement before making any decision related to diagnosis. The second intervention of ML will be before clinical decision three, when the PPCI coordinator try to make the decision including (bring the patient for PPCI, Unable to make a decision, hence, call medics or clinical decision one incorrect then turn down the patient). In this case ML learning using ECGs and a group of novel blood biomarkers will be available to make the decision more accurate. Figure 7.11 below shows clinical decision-making process for PPCI pathway and where the result of this study could be used to improve the pathway.

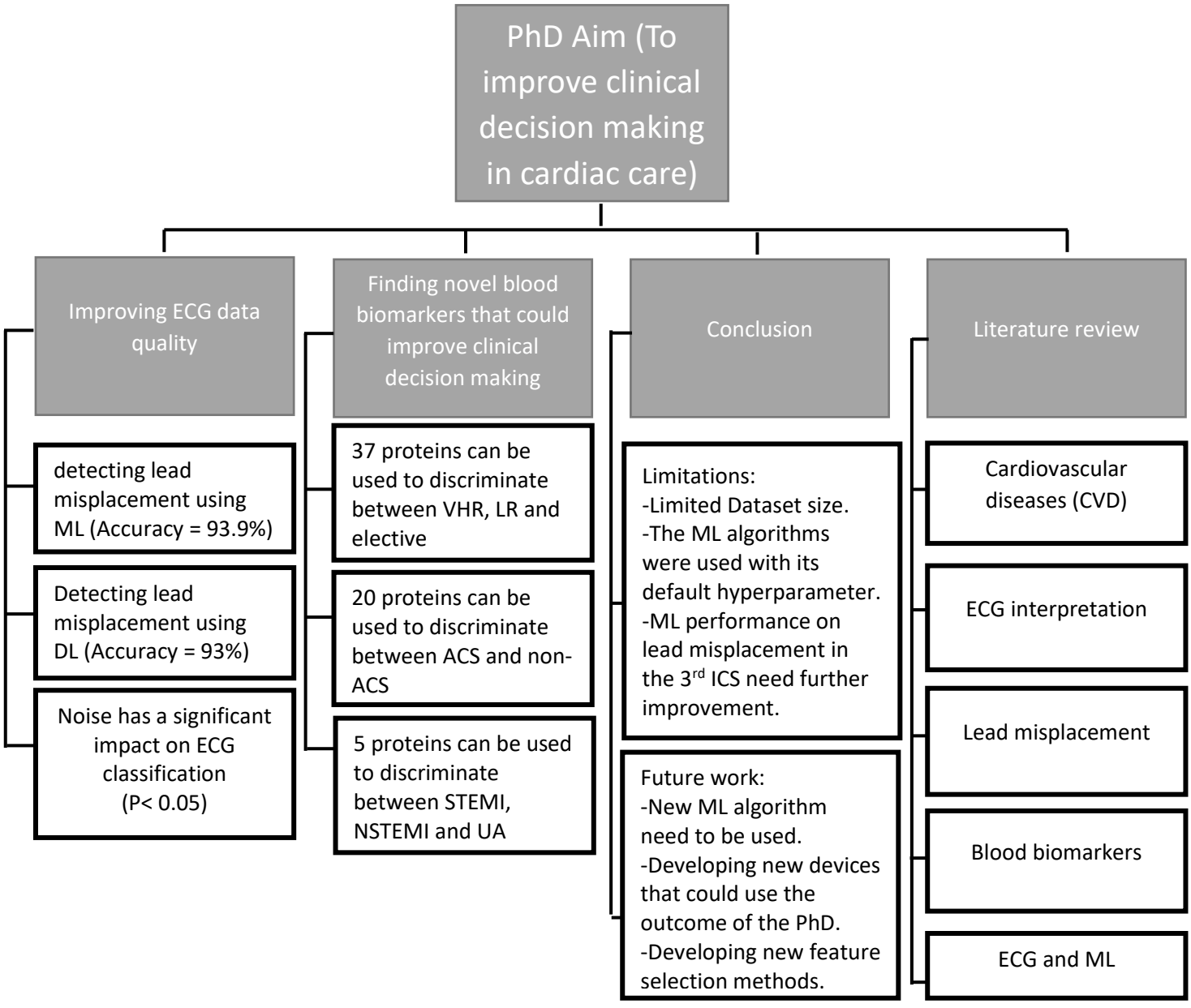


Fig. 7.10: Summary of the outcomes of the thesis.

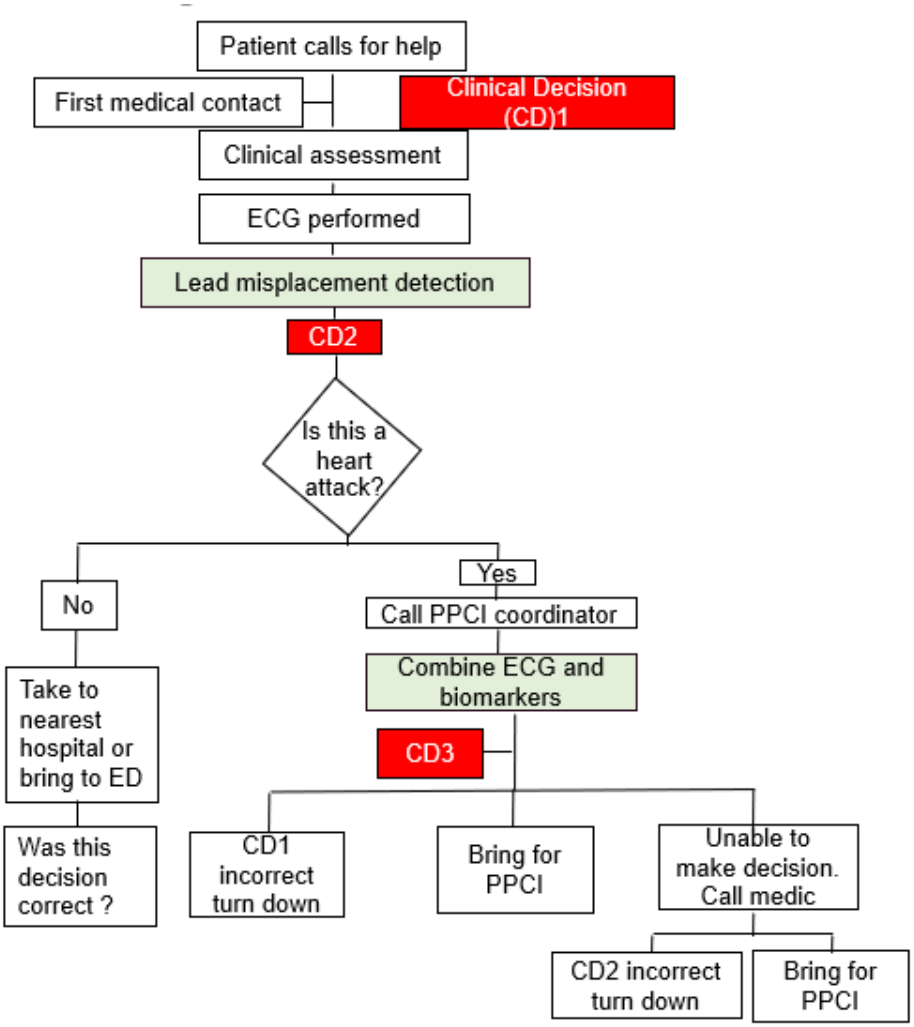


Fig. 7.11 A proposed PPCI pathway according to the results of the PhD.

7.6 Policy and practice implications.

Policy makers in healthcare should include artificial intelligence in their system as it showed promising results in healthcare and reduces time and cost. For example, several healthcare organisations such as Mayo Clinic started involving artificial intelligence technology in their healthcare system, especially in cardiovascular medicine to improve ECGs interpretation and to provide fast and an accurate diagnosis to improve care. Medical device company should consider including AI technologies such as DL algorithm in their new products such as including DL algorithms in 12-lead ECG devices to auto detect lead misplacement before making diagnoses as lead misplacement can significantly change diagnoses. Furthermore, new blood biomarkers panel combined with ML algorithms could be developed to analyse blood biomarkers accurately and quickly. Physicians and scientists should collaborate and work together in multidisciplinary teams to share their knowledge that could be used to combine clinical practice with AI for better diagnoses as they complement each other in terms of intelligent diagnosis.

7.7 Limitations of the PhD work.

There are a few limitations that could be addressed in future work such as the use of small datasets in this PhD, especially in chapters four and six as the data collection was during the pandemic and the nurses were not able to collect the data. Hence, the dataset size needs to be increased for future investigation. Furthermore, in terms of lead misplacement, V1 and V2 were only considered as they are commonly misplaced leads, however, the other chest leads (V3-V6) need to be explored. Hence, the performance of ML for detecting lead misplacement in the third intercostal space (ICS) could be improved by including more leads in the process of data extraction.

The benchmarking in chapter four includes comparing the algorithm with a small group of cardiologists and physicians; hence, the algorithm needs to be benchmarked against a larger group of cardiologists and physicians to generalise the model and strengthen the findings.

The recorded ECG in chapter six was paper-based and the quality was poor. Hence, the low quality of the recorded ECG might have affected the performance of the ML algorithms. Moreover, the ML algorithms in chapter six need to be benchmarked against a large group of physicians (as we did in the fourth chapter) because the algorithms demonstrated promising results. The hyperparameters of the used ML algorithms were set using their default values; hence, these parameters could be optimised to enhance their performance.

7.8 Future work.

In terms of detecting lead misplacement, the performance of the ML algorithms require further improvement, especially in the third ICS. Hence, new features such as rSr' prime and new feature selection algorithms could be used to improve the performance of the ML algorithms. Furthermore, the impact of V1 and V2 lead misplacement on the other leads (V3-V6) requires further investigation. Moreover, leads V3-V6 could be considered in future work with leads V1 and V2 to improve the performance of the ML algorithms to detect misplacement. Improving the transparency and the explainabilty of the ML algorithms to justify the results of the ML algorithms and to improve the human decisions could be carried out. The impact of noise on ECG classification should also be further investigated using different types and ranges of noise artefacts. This would help in developing new applications and medical devices using the benchmarked ML algorithms to improve ECG data quality and diagnoses.

In terms of STEMI and NSTEMI diagnosis using blood biomarkers, a new group of biomarkers such as neurological biomarkers could be considered to improve the performance of ML for detecting STEMI and NSTEMI as they associated with cardiovascular diseases (Daniela Carnevale et al., 2021). The selected group of blood biomarkers could be benchmarked in the future with physicians to check the ability of those biomarkers to discriminate between STEMI and NSTEMI. Furthermore, new test panels could be developed using the selected biomarkers after benchmarking them to improve diagnoses. And those new panels should have the ability to be linked with healthcare systems to enable ML algorithms to provide prompt diagnoses.

7.9 Reflection

During this PhD I have been working with interdisciplinary supervisory team including data scientists, biologist, cardiologists, biomedical engineers, physicians and nurses. As a computer scientist I spent a lot of effort to understand some medical terms and procedures. Furthermore, working with people from different backgrounds required more effort to make the results understood and explainable. A lot of time has been spent to apply for the ethical approval to collect the dataset. Furthermore, collecting the dataset during the pandemic took a lot of time as the nurses were busy and could not be able to collect the data, specially at the beginning of the pandemic. Working from home at the beginning during the pandemic put a lot of pressure and caused some health issues. However, my supervisors helped me a lot and were very supportive to solve my health issue. There are many things I like about this PhD, but the most thing I like about this PhD is my supervisory team and the way that my supervisors directed me during my PhD and I hope that I can do that in my future career as a supervisor. Also, I like the events that have been organised by Ulster University, doctoral college and school of computing as it improved my skills. And I like people who I met during this PhD journey, and I will never forget them for ever. What I dislike about this PhD is it kept me away from my family in the last three years as traveling requires me a lot of time to travel to my home country. If I have been given the chance to do this PhD again, I would spend more time on learning medical terminologies and procedures as healthcare area is my favourite area to apply ML. I would suggest a few adevices for other PhDs from my experience during my PhD including 1) they should balance between their PhD work and their social life, 2) they should have good time management during the PhD, 3) networking with other PhD students to gain their experience, 3) don't feel shy if you don't understand something and try to ask other PhD students who can answer your question, 4) keep learning new skills in this field and keep yourself updated as this field is developing rapidly, 5) publish journal paper as much as you can during the PhD journey to secure a job quickly after finishing the PhD.

References

Abreu, D., Pinto, F. J., Matias-Dias, C., & Sousa, P. (2019). Trends of case-fatality rate by acute coronary syndrome in Portugal: Impact of a fast track to the coronary unit. *JRSM cardiovascular disease*, 8, 2048004019851952. https://doi.org/10.1177/2048004019851952

Ajay, S. and Ausif M. (2019). Review of Deep Learning Algorithms and Architectures. *IEEE Access*, 7(2019), 53040-53065. https://doi.org/10.1109/ACCESS.2019.2912200

Akbar H, Foth C, Kahloon RA, et al (2021). Acute ST Elevation Myocardial Infarction. [Updated 2021 Aug 9]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK532281/

Alahakoon, D., Nawaratne, R., Xu, Y., Daswin, D. S., Uthayasankar, S. and Bhumika, G. (2020). Self-Building Artificial Intelligence and Machine Learning to Empower Big Data Analytics in Smart Cities. *Inf Syst Front* (2020). https://doi.org/10.1007/s10796-020-10056-x

Alioui, S., Kastelein M., Van Dam E. M. and Van Dam P. M. (2017). Automatic registration of 3D camera recording to model for leads localization. *2017 Computing in Cardiology (CinC)*, Rennes, France, 2017, pp. 1-4, https://doi.org/10.22489/CinC.2017.298-171.

Álvarez, D., Cerezo-Hernández, A., Crespo, A. et al (2020). A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow. *Sci Rep*, 10, 5332. https://doi.org/10.1038/s41598-020-62223-4

An X., Zhou L. Liu Z., Wang C., Li P., Li Z. (2021). Dataset and benchmark for detecting moving objects in construction sites. *Automation in Construction*, 122,103482, https://doi.org/10.1016/j.autcon.2020.103482.

Angeli, F., Spanevello, A., De Ponti, R., Visca, D., Marazzato, J., Palmiotto, G., Feci, D., Reboldi, G., Fabbri, L. M., & Verdecchia, P. (2020). Electrocardiographic features of patients with COVID-19 pneumonia. *European journal of internal medicine*, 78, 101–106. https://doi.org/10.1016/j.ejim.2020.06.015

Ashley, E. A., Niebauer, J. Cardiology Explained. London: Remedica; 2004. Chapter 3, Conquering the ECG. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2214/

Basit, H., Malik, A., Huecker, M.R. (2021) Non ST Segment Elevation Myocardial Infarction. In: StatPearls [Internet]. *Treasure Island (FL): StatPearls Publishing*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK513228/

Bond, R. R., Finlay, D. D., Nugent, C. D., & Moore, G. (2011). A review of ECG storage formats. *International journal of medical informatics*, 80(10), 681–697. https://doi.org/10.1016/j.ijmedinf.2011.06.008

Bond, R. R., Finlay, D. D., Nugent, C. D., Breen, C., Guldenring, D., & Daly, M. J. (2012). The effects of electrode misplacement on clinicians' interpretation of the standard 12-lead electrocardiogram. *European journal of internal medicine*, 23(7), 610–615. https://doi.org/10.1016/j.ejim.2012.03.011

Bond, R. R., Finlay, D. D., Nugent, C. D., Breen, C., Guldenring, D., & Daly, M. J. (2012). The effects of electrode misplacement on clinicians' interpretation of the standard 12-lead electrocardiogram. *European journal of internal medicine*, 23(7), 610–615. https://doi.org/10.1016/j.ejim.2012.03.011

Bond, R. R., Finlay, D. D., McLaughlin, J., Guldenring, D., Cairns, A., Kennedy, A., Deans, R., Waldo, A. L., & Peace, A. (2016). Human factors analysis of the CardioQuick Patch®: A novel engineering solution to the problem of electrode misplacement during 12-lead electrocardiogram acquisition. *Journal of electrocardiology*, 49(6), 911–918. https://doi.org/10.1016/j.jelectrocard.2016.08.009

British Heart Foundation (2021) Heart statistics [online]. Available at: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics (Accessed: 26 March 2021)

Bupp, J. E., Dinger, M., Lawrence, C., & Wingate, S. (1997). Placement of cardiac electrodes: written, simulated, and actual accuracy. *American journal of critical care : an official publication, American Association of Critical-Care Nurses*, 6(6), 457–462.

Cairns, A. W., Bond, R. R., Finlay, D. D., Guldenring, D., Badilini, F., Libretti, G., Peace, A. J., & Leslie, S. J. (2017). A decision support system and rule-based algorithm to augment the human interpretation of the 12-lead electrocardiogram. *Journal of electrocardiology*, 50(6), 781–786. https://doi.org/10.1016/j.jelectrocard.2017.08.007

Califf R. M. (2018). Biomarker definitions and their applications. *Experimental biology and medicine (Maywood, N.J.)*, 243(3), 213–221. https://doi.org/10.1177/1535370217750088

Cao, S., Li, L., Geng, X., Ma, Y., Huang, X., & Kang, X. (2019). The upregulation of miR-101 promotes vascular endothelial cell apoptosis and suppresses cell migration in acute coronary syndrome by targeting CDH5. International journal of clinical and experimental pathology, 12(9), 3320–3328.

Carnevale D., L. Giuseppe (2021). Neuroimmune interactions in cardiovascular diseases. Cardiovascular Research, Volume 117, Issue 2, Pages 402–410, https://doi.org/10.1093/cvr/cvaa151

Cau, R., Bassareo, P. P., Mannelli, L., Suri, J. S., & Saba, L. (2020). Imaging in COVID-19-related myocardial injury. *The international journal of cardiovascular imaging*, 1–12. Advance online publication. https://doi.org/10.1007/s10554-020-02089-9

Chacko, S., Haseeb, S., Glover, B. M., Wallbridge, D., & Harper, A. (2017). The role of biomarkers in the diagnosis and risk stratification of acute coronary syndrome. *Future science OA*, 4(1), FSO251. https://doi.org/10.4155/fsoa-2017-0036

Chanarin, N., Caplin, J., & Peacock, A. (1990). "Pseudo reinfarction": a consequence of electrocardiogram lead transposition following myocardial infarction. *Clinical cardiology*, 13(9), 668–669. https://doi.org/10.1002/clc.4960130916

Chauhan N. K. and Singh K. (2018). A Review on Conventional Machine Learning vs Deep Learning. *International Conference on Computing, Power and Communication Technologies (GUCON)*, Greater Noida, India, 347-352. https://doi.org/10.1109/GUCON.2018.8675097.

Christopher P. Cannon and Alexander G.G. Turpie (2003) Unstable Angina and Non–ST-Elevation Myocardial Infarction. *Circulation*. 107(21), 1640-2645. https://doi.org/10.1161/01.CIR.0000072246.69344.2D

DAIC (2019) Half of Hospital Decision Makers Plan to Invest in AI by 2021. Available at: https://www.dicardiology.com/content/half-hospital-decision-makers-plan-invest-ai-2021 (Accessed 27 March 2021)

DAIC. (2019). ECG System Uses 3-D Interactive Image to Show Proper Lead Placements [online]. Available at: https://www.dicardiology.com/videos/video-ecg-system-uses-3-d-interactive-image-show-proper-lead-placements (Accessed on 28 March 2021)

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

De Bie, J., Mortara, D. W., & Clark, T. F. (2014). The development and validation of an early warning system to prevent the acquisition of 12-lead resting ECGs with interchanged electrode positions. *Journal of electrocardiology*, 47(6), 794–797. https://doi.org/10.1016/j.jelectrocard.2014.08.015

De Hoog, V. C., Timmers, L., Schoneveld, A. H., Wang, J. W., van de Weg, S. M., Sze, S. K., van Keulen, J. K., Hoes, A. W., den Ruijter, H. M., de Kleijn, D. P., & Mosterd, A. (2013). Serum extracellular vesicle protein levels are associated with acute coronary syndrome. European heart journal. Acute cardiovascular care, 2(1), 53–60. https://doi.org/10.1177/2048872612471212

De Lemos, J. A., Morrow, D. A., Gibson, C. M., Murphy, S. A., Sabatine, M. S., Rifai, N., McCabe, C. H., Antman, E. M., Cannon, C. P., & Braunwald, E. (2002). The prognostic value of serum myoglobin in patients with non-ST-segment elevation acute coronary syndromes. Results from the TIMI 11B and TACTICS-TIMI 18 studies. Journal of the American College of Cardiology, 40(2), 238–244. https://doi.org/10.1016/s0735-1097(02)01948-4

Dehghan, A., Kovacevic, A., Karystianis, G., Keane, J. A., & Nenadic, G. (2015). Combining knowledge- and data-driven methods for de-identification of clinical narratives. Journal of biomedical informatics, 58 Suppl(Suppl), S53–S59. https://doi.org/10.1016/j.jbi.2015.06.029

Diagnostic and Interventional Cardiology (DAIC) (2015), Advances in ECG Technology [online]. Available at: https://www.dicardiology.com/article/advances-ecg-technology (Accessed: 26 March 2021)

Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

Domínguez-Jiménez, J. A., Campo-Landines, K. C., Martínez-Santos, J. C., Delahoz, E. J. and Contreras-Ortiz, S. H. (2020). A machine learning model for emotion recognition from physiological signals. *Biomedical Signal Processing and Control*, 55, 101646. https://doi.org/10.1016/j.bspc.2019.101646.

Douglas, S. H. and Americo, S. (2017). Immune Response Biomarkers Characterize Endothelial Injury and Predict Acute Coronary Syndrome. Arteriosclerosis, Thrombosis, and Vascular Biology, 37(1). https://www.ahajournals.org/doi/10.1161/atvb.37.suppl_1.214

Ekaterina, I. and Georgii B. (2020) .Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions. *Procedia Computer Science*, 169, 244-248. https://doi.org/10.1016/j.procs.2020.02.143

Elisa, R., Marcello, M. and Emilio, M. (2014). Biomarkers Discovery through Multivariate Statistical Methods: A Review of Recently Developed Methods and Applications in Proteomics. *Journal of Proteomics & Bioinformatics*, S3: 003. doi:10.4172/jpb.S3-003

Ensor J. E. (2014). Biomarker validation: common data analysis concerns. *The oncologist*, 19(8), 886–891. https://doi.org/10.1634/theoncologist.2014-0061

Fang, Y., Wang, J., Chen, S., Shen, S., Zhang, Z. and Hu W. (2019). Impact of Dataset Size on Deep Learning-Based Auto Segmentation for Head and Neck Cancer. *International Journal of Radiation Oncology, Biology, Physics*, 105(1), E129 - E130. https://doi.org/10.1016/j.ijrobp.2019.06.2258

Fanola, C. L., Morrow, D. A., Cannon, C. P., Jarolim, P., Lukas, M. A., Bode, C., Hochman, J. S., Goodrich, E. L., Braunwald, E., & O'Donoghue, M. L. (2017). Interleukin-6 and the Risk of Adverse Outcomes in Patients After an Acute Coronary Syndrome: Observations From the SOLID-TIMI 52 (Stabilization of Plaque Using Darapladib-Thrombolysis in Myocardial Infarction 52) Trial. Journal of the American Heart Association, 6(10), e005637. https://doi.org/10.1161/JAHA.117.005637

Feldman, J. (2001). Artificial Intelligence. *Encyclopedia of Neuroscience*, 561-564. https://doi.org/10.1016/B978-008045046-9.00434-4

Finlay, D. D., Nugent, C. D., Kellett, J. G., Donnelly, M. P., McCullagh, P. J., & Black, N. D. (2007). Synthesising the 12-lead electrocardiogram: Trends and challenges. *European journal of internal medicine*, 18(8), 566–570. https://doi.org/10.1016/j.ejim.2007.04.011

Finlay, D. D., Nugent, C. D., Nelwan, S. P., Bond, R. R., Donnelly, M. P., & Guldenring, D. (2010). Effects of electrode placement errors in the EASI-derived 12-lead electrocardiogram. *Journal of electrocardiology*, 43(6), 606–611. https://doi.org/10.1016/j.jelectrocard.2010.07.004

Fiorentino, M., Capizzi, E., & Loda, M. (2010). Blood and tissue biomarkers in prostate cancer: state of the art. *The Urologic clinics of North America*, 37(1), 131–Contents. https://doi.org/10.1016/j.ucl.2009.11.006

Gao, L. B., Zhou, B., Zhang, L., Wei, Y. S., Wang, Y. Y., Liang, W. B., Lv, M. L., Pan, X. M., Chen, Y. C., & Rao, L. (2008). R497K polymorphism in epidermal

growth factor receptor gene is associated with the risk of acute coronary syndrome. *BMC medical genetics*, 9, 74. https://doi.org/10.1186/1471-2350-9-74

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial Intelligence in Healthcare*, 295–336. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Geoffrey, H. T., Jeffrey, Z., Francesca, N. D., Rahul, C. D. (2019). Automated and Interpretable Patient ECG Profiles for Disease Detection, Tracking, and Discovery. Circulation: *Cardiovascular Quality and Outcomes*, 12(9), e005289. https://doi.org/10.1161/CIRCOUTCOMES.118.005289

Giustino, G., Pinney, S. P., Lala, A., Reddy, V. Y., Johnston-Cox, H. A., Mechanick, J. I., Halperin, J. L., & Fuster, V. (2020). Coronavirus and Cardiovascular Disease, Myocardial Injury, and Arrhythmia: JACC Focus Seminar. *Journal of the American College of Cardiology*, 76(17), 2011–2023. https://doi.org/10.1016/j.jacc.2020.08.059

Gosho, M., Nagashima, K., & Sato, Y. (2012). Study designs and statistical analyses for biomarker research. Sensors (Basel, Switzerland), 12(7), 8966–8986. https://doi.org/10.3390/s120708966

Golino, P., Ravera, A., Ragni, M., Cirillo, P., Piro, O., & Chiariello, M. (2003). Involvement of tissue factor pathway inhibitor in the coronary circulation of patients with acute coronary syndromes. Circulation, 108(23), 2864–2869.

Govorukhina, N. I., de Vries, M., Reijmers, T. H., Horvatovich, P., van der Zee, A. G., & Bischoff, R. (2009). Influence of clotting time on the protein composition of serum samples based on LC-MS data. *Journal of chromatography*. B, Analytical technologies in the biomedical and life sciences, 877(13), 1281–1291. https://doi.org/10.1016/j.jchromb.2008.10.029

Grégoire M., Wojciech S., Klaus-Robert M. (2004). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, 1-15. https://doi.org/10.1016/j.dsp.2017.10.011.

Haas, B., Serchi, T., Wagner, D. R., Gilson, G., Planchon, S., Renaut, J., Hoffmann, L., Bohn, T., & Devaux, Y. (2011). Proteomic analysis of plasma samples from patients with acute myocardial infarction identifies haptoglobin as a potential prognostic biomarker. *Journal of proteomics*, 75(1), 229–236.

Han, C., Gregg, R. E., Feild, D. Q., & Babaeizadeh, S. (2014). Automatic detection of ECG cable interchange by analyzing both morphology and interlead relations.

Journal of electrocardiology, 47(6), 781–787. https://doi.org/10.1016/j.jelectrocard.2014.08.006

Han, C., Gregg R. E. and Babaeizadeh S. (2014). Automatic detection of ECG lead-wire interchange for conventional and Mason-Likar lead systems. *Computing in Cardiology 2014*, Cambridge, MA, USA, 2014, 145-148.

Heden, B., Ohlsson, M., Rittner, R., Pahlm, O., Edenbrandt L. and Peterson, C. (1995). Misplacement of the left foot ECG electrode detected by artificial neural networks. *Computers in Cardiology 1995*, Vienna, Austria, 1995, 225-228, doi: 10.1109/CIC.1995.482613.

Hedén, B., Ohlsson, M., Edenbrandt, L., Rittner, R., Pahlm, O., & Peterson, C. (1995). Artificial neural networks for recognition of electrocardiographic lead reversal. *The American journal of cardiology*, 75(14), 929–933. https://doi.org/10.1016/s0002-9149(99)80689-4

Hedén, B., Ohlsson, M., Holst, H., Mjöman, M., Rittner, R., Pahlm, O., Peterson, C., & Edenbrandt, L. (1996). Detection of frequently overlooked electrocardiographic lead reversals using artificial neural networks. *The American journal of cardiology*, 78(5), 600–604. https://doi.org/10.1016/s0002-9149(96)00377-3

Heeschen, C., Dimmeler, S., Hamm, C. W., Boersma, E., Zeiher, A. M., Simoons, M. L., & CAPTURE (c7E3 Anti-Platelet Therapy in Unstable REfractory angina) Investigators (2003). Prognostic significance of angiogenic growth factor serum levels in patients with acute coronary syndromes. Circulation, 107(4), 524–530. https://doi.org/10.1161/01.cir.0000048183.37648.1a

Hélène, R., Astrid, M., Mathilde, B., Fériel, A., Loubina, F., Régine, M., Evelyne, P., Alain, C., Claude, D., Nicolas, V., Christos, C., and Jane-Lise, S. (2016). *Hypertension*, 68(2), 392–400. https://doi.org/10.1161/HYPERTENSIONAHA.116.07694

Ho, Y., Pepyne, D. (2002). Simple Explanation of the No-Free-Lunch Theorem and Its Implications. *Journal of Optimization Theory and Applications*, 115, 549–570. https://doi.org/10.1023/A:1021251113462

Jakobsson, A., Ohlin, P., & Pahlm, O. (1985). Does a computer-based ECG-recorder interpret electrocardiograms more efficiently than physicians?. *Clinical physiology* (Oxford, England), 5(5), 417–423. https://doi.org/10.1111/j.1475-097x.1985.tb00772.x

James, T., Edward, T. H. and Emerson, C. P. (2012). Cytokine Profile and ST-Elevation Myocardial Infarction. *Circulation Research*, 111(10), 1256–1257. https://doi.org/10.1161/CIRCRESAHA.112.279380Circulation

Jansson, A. M., Hartford, M., Omland, T., Karlsson, T., Lindmarker, P., Herlitz, J., Ueland, T., Aukrust, P., & Caidahl, K. (2012). Multimarker risk assessment including osteoprotegerin and CXCL16 in acute coronary syndromes. Arteriosclerosis, thrombosis, and vascular biology, 32(12), 3041–3049. https://doi.org/10.1161/ATVBAHA.112.300326

Jekova, I., Krasteva, V., Dotsinsky, I., Christov, I. and Abächerli, R. (2011). Recognition of diagnostically useful ECG recordings: Alert for corrupted or interchanged leads. *2011 Computing in Cardiology*, Hangzhou, China, 2011, 429-432.

Jekova, I. , Krasteva V. and Abächerli R. (2013). Detection of electrode interchange in precordial and orthogonal ECG leads. *Computing in Cardiology 2013*, Zaragoza, Spain, 2013, 519-522.

Jekova, I., Krasteva, V., Leber, R., Schmid, R., Twerenbold, R., Müller, C., Reichlin, T., & Abächerli, R. (2016). Inter-lead correlation analysis for automated detection of cable reversals in 12/16-lead ECG. *Computer methods and programs in biomedicine*, 134, 31–41. https://doi.org/10.1016/j.cmpb.2016.06.003

Jowett, N. I., Turner, A. M., Cole, A., & Jones, P. A. (2005). Modified electrode placement must be recorded when performing 12-lead electrocardiograms. *Postgraduate medical journal*, 81(952), 122–125. https://doi.org/10.1136/pgmj.2004.021204

Kania, M., Rix, H., Fereniec, M., Zavala-Fernandez, H., Janusek, D., Mroczka, T., Stix, G., & Maniewski, R. (2014). The effect of precordial lead displacement on ECG morphology. *Medical & biological engineering & computing*, 52(2), 109–119. https://doi.org/10.1007/s11517-013-1115-9

Kaplan, A., Cao, H., FitzGerald, J. M., Iannotti, N., Yang, E., Kocks, J., Kostikas, K., Price, D., Reddel, H. K., Tsiligianni, I., Vogelmeier, C. F., Pfister, P., & Mastoridis, P. (2021). Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. *The journal of allergy and clinical immunology*. In practice, S2213-2198(21)00194-X. Advance online publication. https://doi.org/10.1016/j.jaip.2021.02.014

Khunti, K. (2014). Accurate interpretation of the 12-lead ECG electrode placement: A systematic review. *Health Education Journal*, 73(5), 610–623. https://doi.org/10.1177/0017896912472328

Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., van Herpen, G., Kors, J. A., Macfarlane, P., Mirvis, D. M., Pahlm, O., Rautaharju, P., Wagner, G. S., American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology, American College of Cardiology Foundation, Heart Rhythm Society, Josephson, M., Mason, J. W., Okin, P., Surawicz, B., … Wellens, H. (2007). Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation*, 115(10), 1306–1324. https://doi.org/10.1161/CIRCULATIONAHA.106.180200

Kochi, A. N., Tagliari, A. P., Forleo, G. B., Fassini, G. M., & Tondo, C. (2020). Cardiac and arrhythmic complications in patients with COVID-19. *Journal of cardiovascular electrophysiology*, 31(5), 1003–1008. https://doi.org/10.1111/jce.14479

Koganti, S., & Rakhit, R. D. (2015). Management of high-risk non-ST elevation myocardial infarction in the UK: need for alternative models of care to reduce length of stay and admission to angiography times. *Clinical medicine* (London, England), 15(6), 522–525. https://doi.org/10.7861/clinmedicine.15-6-522

Konopka, A., Janas, J., Piotrowski, W., & Stepińska, J. (2010). Hepatocyte growth factor--a new marker for prognosis in acute coronary syndrome. Growth factors (Chur, Switzerland), 28(2), 75–81. https://doi.org/10.3109/08977190903403984

Konopka, A., Janas, J., & Stępińska, J. (2012). Hepatocyte growth factor concentration during the first day of acute coronary syndrome. Archives of medical science : AMS, 8(2), 389–391. https://doi.org/10.5114/aoms.2012.28571

Kors, J. A., & van Herpen, G. (2000). A novel method to detect electrocardiographic electrode interchanges. *Journal of electrocardiology*, 33 Suppl, 209–210. https://doi.org/10.1054/jelc.2000.20352

Kors, J. A., & van Herpen, G. (2001). Accurate automatic detection of electrode interchange in the electrocardiogram. *The American journal of cardiology*, 88(4), 396–399. https://doi.org/10.1016/s0002-9149(01)01686-1

Kristen, J. (2009). Acute Coronary Syndrome. *AJN*, Vol.109, pp. 42-52.

Kumar, A., & Cannon, C. P. (2009). Acute coronary syndromes: diagnosis and management, part I. *Mayo Clinic proceedings*, 84(10), 917–938. https://doi.org/10.1016/S0025-6196(11)60509-0

Lateef, F., Nimbkar, N., Da, Z. K., & Min, F. R. (2003). Vertical displacement of the precordial leads alters electrocardiographic morphology. *Indian heart journal*, 55(4), 339–343.

Lee, K. W., Lip, G. Y., & Blann, A. D. (2004). Plasma angiopoietin-1, angiopoietin-2, angiopoietin receptor tie-2, and vascular endothelial growth factor levels in acute coronary syndromes. Circulation, 110(16), 2355–2360.

Liu, H., Li, M., Zhao, J., Mo, Y. (2011). An Effective Feature Selection Method Using Dynamic Information Criterion. In: Deng H., Miao D., Lei J., Wang F.L. (eds) Artificial Intelligence and Computational Intelligence. AICI 2011. *Lecture Notes in Computer Science, 7002. Springer*, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23881-9_59

Liu, Y., Li, X., Peng, D., Tan, Z., Liu, H., Qing, Y., Xue, Y., & Shi, G. P. (2009). Usefulness of serum cathepsin L as an independent biomarker in patients with coronary heart disease. *The American journal of cardiology*, 103(4), 476–481.

Long, B., Brady, W. J., Bridwell, R. E., Ramzy, M., Montrief, T., Singh, M., & Gottlieb, M. (2021). Electrocardiographic manifestations of COVID-19. *The American journal of emergency medicine*, 41, 96–103. https://doi.org/10.1016/j.ajem.2020.12.060

Loria, V., Leo, M., Biasillo, G., Dato, I., & Biasucci, L. M. (2008). Biomarkers in Acute Coronary Syndrome. *Biomarker insights*, 3, 453–468. https://doi.org/10.4137/bmi.s588

Ludwig, T. W., Adrian, C., Stephanie, B., Angelina, K., Johannes, B., Johannes, C. H., Maximilian. M., Clemens, S., Ines, S., Michael, I., Stefan, M., Jens, R., Karin, K., Stefan, K., Mathias, O., Steffen, M., Jörg, H., Ulrich, G. (2021). Myocardial Inflammation and Dysfunction in COVID-19–Associated Myocardial Injury. *Circulation: Cardiovascular Imaging*, 14(1). https://doi.org/10.1161/CIRCIMAGING.120.011713

Luís, S., Diego, C., João, P. M, Nuno, M.M. R., Pedro, M., Miguel, C. G. and Hipólito, S. (2021). Activity classification using accelerometers and machine learning for complex construction worker activities. *Journal of Building Engineering*, 35, 102001. https://doi.org/10.1016/j.jobe.2020.102001.

Ma, W. Q., Wang, Y., Han, X. Q., Zhu, Y., & Liu, N. F. (2018). Associations between LPL gene polymorphisms and coronary artery disease: evidence based on an updated and cumulative meta-analysis. Bioscience reports, 38(2), BSR20171642. Advance online publication. https://doi.org/10.1042/BSR20171642

Macfarlane, P., Devine, B. and Clark E. (2005). The university of glasgow (Uni-G) ECG analysis program. *Computers in Cardiology*, 2005, Lyon, France, pp. 451-454. https://doi.org/10.1109/CIC.2005.1588134

Májek, P., Reicheltová, Z., Suttnar, J., Malý, M., Oravec, M., Pečánková, K., & Dyr, J. E. (2011). Plasma proteome changes in cardiovascular disease patients: novel isoforms of apolipoprotein A1. *Journal of translational medicine*, 9, 84. https://doi.org/10.1186/1479-5876-9-84

Manal B., Liping Z. (2019). A review of machine learning algorithms for identification and classification of non-functional requirements. *Expert Systems with Applications: X*, 1(2019). https://doi.org/10.1016/j.eswax.2019.100001.

Marco T. R., Sameer S., Carlos G. (2016). Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* August, 1135–1144. https://doi.org/10.1145/2939672.2939778

Maršánová, L., Ronzhina, M., Smíšek, R., Vítek, M., Němcová, A., Smital, L., & Nováková, M. (2017). ECG features and methods for automatic classification of ventricular premature and ischemic heartbeats: A comprehensive experimental study. Scientific reports, 7(1), 11239. https://doi.org/10.1038/s41598-017-10942-6

Mason, J. W., Hancock, E. W., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Josephson, M., Kligfield, P., Kors, J. A., Macfarlane, P., Pahlm, O., Mirvis, D. M., Okin, P., Rautaharju, P., Surawicz, B., van Herpen, G., Wagner, G. S., Wellens, H., American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology, American College of Cardiology Foundation, … Heart Rhythm Society (2007). Recommendations for the standardization and interpretation of the electrocardiogram: part II: electrocardiography diagnostic statement list a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. Journal of the American College of Cardiology, 49(10), 1128–1135. https://doi.org/10.1016/j.jacc.2007.01.025

Mayeux R. (2004). Biomarkers: potential uses and limitations. NeuroRx *: the journal of the American Society for Experimental NeuroTherapeutics*, 1(2), 182–188. https://doi.org/10.1602/neurorx.1.2.182

Mayo Clinic, 2021, Artificial Intelligence (AI) in Cardiovascular Medicine, 28/09/2021, https://www.mayoclinic.org/departments-centers/ai-cardiology/overview/ovc-20486648.

McCann, C. J., Glover, B. M., Menown, I. B., Moore, M. J., McEneny, J., Owens, C. G., Smith, B., Sharpe, P. C., Young, I. S., & Adgey, J. A. (2008). Novel biomarkers in early diagnosis of acute myocardial infarction compared with cardiac troponin T. *European heart journal*, 29(23), 2843–2850.

McCarthy, C. P., van Kimmenade, R., Gaggin, H. K., Simon, M. L., Ibrahim, N. E., Gandhi, P., Kelly, N., Motiwala, S. R., Belcher, A. M., Harisiades, J., Magaret, C. A., Rhyne, R. F., & Januzzi, J. L., Jr (2017). Usefulness of Multiple Biomarkers for Predicting Incident Major Adverse Cardiac Events in Patients Who Underwent Diagnostic Coronary Angiography (from the Catheter Sampled Blood Archive in Cardiovascular Diseases [CASABLANCA] Study). *The American journal of cardiology*, 120(1), 25–32. https://doi.org/10.1016/j.amjcard.2017.03.265

McGilligan, V., Watterson, S., Rjoob, K., Chemaly, M., Bond, R., Iftikhar, A., Knoery, C., Leslie, S. J., McShane, A., Bjourson, A., & Peace, A. (2019). An exploratory analysis investigating blood protein biomarkers to augment ECG diagnosis of ACS. *Journal of electrocardiology*, 57S, S92–S97. https://doi.org/10.1016/j.jelectrocard.2019.09.002

Mc Namara, K., Alzubaidi, H., & Jackson, J. K. (2019). Cardiovascular disease as a leading cause of death: how are pharmacists getting involved?. *Integrated pharmacy research & practice*, 8, 1–11. https://doi.org/10.2147/IPRP.S133088

Meek, S., & Morris, F. (2002). ABC of clinical electrocardiography.Introduction. I-Leads, rate, rhythm, and cardiac axis. *BMJ* (Clinical research ed.), 324(7334), 415–418. https://doi.org/10.1136/bmj.324.7334.415

Meissner, A., Trappe, H. J., de Boer, M. J., Gorgels, A. P., & Wellens, H. J. (2010). The value of the ECG for decision-making at first medical contact in the patient with acute chest pain. *Netherlands heart journal: monthly journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation*, 18(6), 301–306. https://doi.org/10.1007/BF03091780

Morisbak, B., & Gjesdal, K. (1999). Datamaskintolking av EKG--veiledende eller villedende? [Computer-based interpretation of ECG--guiding or misleading?]. Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke, 119(23), 3441–3444.

Mekov, E., Miravitlles, M., & Petkov, R. (2020). Artificial intelligence and machine learning in respiratory medicine. *Expert review of respiratory medicine*, 14(6), 559–564. https://doi.org/10.1080/17476348.2020.1743181

Miller, Tim et al (2017). Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. ArXiv, Vol.abs/1712.00547, 2017, available at: https://people.eng.unimelb.edu.au/tmiller/pubs/explanation-inmates.pdf

Moroni, F., Ammirati, E., Norata, G. D., Magnoni, M., & Camici, P. G. (2019). The Role of Monocytes and Macrophages in Human Atherosclerosis, Plaque Neoangiogenesis, and Atherothrombosis. Mediators of inflammation, 2019, 7434376. https://doi.org/10.1155/2019/7434376

Mostafa, S. S., Mendonça, F., Ravelo-García, A. G., & Morgado-Dias, F. (2019). A Systematic Review of Detecting Sleep Apnea Using Deep Learning. *Sensors* (Basel, Switzerland), 19(22), 4934. https://doi.org/10.3390/s19224934

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M. et al. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1). https://doi.org/10.1186/s40537-014-0007-7

NHS (2020) Arterial thrombosis [online]. Available at: https://www.nhs.uk/conditions/arterial-thrombosis/ (Accessed: 26 March 2021)

NHS (2018) Cardiovascular disease [online]. Available at: https://www.nhs.uk/conditions/cardiovascular-disease/ (Accessed: 26 March 2021)

Nicoletta, D., Emanuele, P., and Barbara, P. (2013). A Comparative Analysis of Biomarker Selection Techniques. *BioMed Research International*, 2013. https://doi.org/10.1155/2013/387673

O'Connor, R. E., Brady, W., Brooks, S. C., Diercks, D., Egan, J., Ghaemmaghami, C., Menon, V., O'Neil, B. J., Travers, A. H., & Yannopoulos, D. (2010). Part 10: acute coronary syndromes: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, 122(18 Suppl 3), S787–S817. https://doi.org/10.1161/CIRCULATIONAHA.110.971028

Oldenhuis, C. N., Oosting, S. F., Gietema, J. A., de Vries, E. G. (2008). Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer, 44(7):946-53. doi: 10.1016/j.ejca.2008.03.006.

Parvaneh, S., Rubin, J., Babaeizadeh, S., & Xu-Wilson, M. (2019). Cardiac arrhythmia detection using deep learning: A review. *Journal of electrocardiology*, 57S, S70–S74. https://doi.org/10.1016/j.jelectrocard.2019.08.004

Pedrazzini, G., Biasco, L., Sulzer, I., Anesini, A., Moccetti, T., Kremer Hovinga, J. A., & Alberio, L. (2016). Acquired intracoronary ADAMTS13 deficiency and VWF retention at sites of critical coronary stenosis in patients with STEMI. Blood, 127(23), 2934–2936. https://doi.org/10.1182/blood-2015-12-688010

PIPBERGER, H. V., ARMS, R. J., & STALLMANN, F. W. (1961). Automatic screening of normal and abnormal electrocardiograms by means of digital electronic computer. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* (New York, N.Y.), 106, 130–132. https://doi.org/10.3181/00379727-106-26260

Ponti, G., Maccaferri, M., Ruini, C., Tomasi, A., & Ozben, T. (2020). Biomarkers associated with COVID-19 disease progression. *Critical reviews in clinical laboratory sciences*, 57(6), 389–399. https://doi.org/10.1080/10408363.2020.1770685

Qiang, Z., Evan, H., Konrad, W., Cody, W., Iulia, P., Elena, L., Ahmet, B., Vanessa, M. F., Stefan, K. P. (2020). Deep learning with attention supervision for automated motion artefact detection in quality control of cardiac T1-mapping. *Artificial Intelligence in Medicine*, 110. https://doi.org/10.1016/j.artmed.2020.101955.

Qixian, W., Zhanxin, W. 0448 (2010). A study of relation between tissue factor pathway and acute coronary syndrome. Heart,96, A139-A140. http://dx.doi.org/10.1136/hrt.2010.208967.448

Rajaganeshan, R., Ludlam, C. L., Francis, D. P., Parasramka, S. V., & Sutton, R. (2008). Accuracy in ECG lead placement among technicians, nurses, general physicians and cardiologists. *International journal of clinical practice*, 62(1), 65–70. https://doi.org/10.1111/j.1742-1241.2007.01390..x

Ramachandran, S. Vasan (2006). Biomarkers of Cardiovascular Disease. *Circulation*, 113(9),2335–2362. https://doi.org/10.1161/CIRCULATIONAHA.104.482570

Rauen, C. A., Chulay, M., Bridges, E., Vollman, K. M., & Arbour, R. (2008). Seven evidence-based practice habits: putting some sacred cows out to pasture. *Critical care nurse*, 28(2), 98–124.

Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M. et al (2002). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*, 11(1760). https://doi.org/10.1038/s41467-020-15432-4

Richard E. G., William H., Saeed B. (2017). Detecting ECG Limb Lead-wire Interchanges Involving the Right Leg Lead-wire. *Computing in Cardiology*, 44. https://doi.org/10.22489/CinC.2017.014-061

Rip, J., Nierman, M. C., Wareham, N. J., Luben, R., Bingham, S. A., Day, N. E., van Miert, J. N., Hutten, B. A., Kastelein, J. J., Kuivenhoven, J. A., Khaw, K. T., & Boekholdt, S. M. (2006). Serum lipoprotein lipase concentration and risk for future coronary artery disease: the EPIC-Norfolk prospective population study. Arteriosclerosis, thrombosis, and vascular biology, 26(3), 637–642. https://doi.org/10.1161/01.ATV.0000201038.47949.56

Rjoob K, Bond R. et al. (2019). Machine Learning Improves the Detection of Misplaced v1 and v2 Electrodes During 12-Lead Electrocardiogram Acquisition. *CinC 2019*. https://doi.org/10.22489/CinC.2019.035.

Rjoob, K., Bond, R., Finlay, D., McGilligan, V., Leslie, S. J., Iftikhar, A., Guldenring, D., Rababah, A., Knoery, C., McShane, A., & Peace, A. (2019). Data driven feature selection and machine learning to detect misplaced V1 and V2 chest electrodes when recording the 12-lead electrocardiogram. *Journal of electrocardiology*, 57, 39–43. https://doi.org/10.1016/j.jelectrocard.2019.08.017

Rjoob, K., Bond, R., Finlay, D., McGilligan, V., Leslie, S. J., Iftikhar, A., Guldenring, D., Rababah, A., Knoery, C., McShane, A., & Peace, A. (2020). Machine learning techniques for detecting electrode misplacement and interchanges when recording ECGs: A systematic review and meta-analysis. *Journal of electrocardiology*, 62, 116–123. https://doi.org/10.1016/j.jelectrocard.2020.08.013

Robert, B. (1969) Early phase of myocardial ischemic injury and infarction. *American Journal of Cardiology*, 6(24), 753. https://doi.org/10.1016/0002-9149(69)90464-0

Rudiger, A., Schöb, L., & Follath, F. (2003). Influence of electrode misplacement on the electrocardiographic signs of inferior myocardial ischemia. *The American journal of emergency medicine*, 21(7), 574–577. https://doi.org/10.1016/j.ajem.2003.08.007

Sanchis-Gomar, F., Perez-Quilis, C., Leischik, R., & Lucia, A. (2016). Epidemiology of coronary heart disease and acute coronary syndrome. *Annals of translational medicine*, 4(13), 256. https://doi.org/10.21037/atm.2016.06.33

Schernthaner, C., Paar, V., Wernly, B., Pistulli, R., Rohm, I., Jung, C., Figulla, H. R., Yilmaz, A., Cadamuro, J., Haschke-Becher, E., Schulze, P. C., Hoppe, U. C., Lichtenauer, M., & Kretzschmar, D. (2017). Elevated plasma levels of interleukin-16 in patients with acute myocardial infarction. Medicine, 96(44), e8396. https://doi.org/10.1097/MD.0000000000008396Copy

Schijvenaars, B. J., van Herpen, G., & Kors, J. A. (2008). Intraindividual variability in electrocardiograms. *Journal of electrocardiology*, 41(3), 190–196. https://doi.org/10.1016/j.jelectrocard.2008.01.012

Schroder, J., Mygind, N. D., Frestad, D., Michelsen, M., Suhrs, H. E., Bove, K. B., Gustafsson, I., Kastrup, J., & Prescott, E. (2019). Pro-inflammatory biomarkers in women with non-obstructive angina pectoris and coronary microvascular dysfunction. *International journal of cardiology*. Heart & vasculature, 24, 100370. https://doi.org/10.1016/j.ijcha.2019.100370

Scott L., Su-In L. (2017). A Unified Approach to Interpreting Model Predictions, arXiv:1705.07874, 2017. Availabl at: https://arxiv.org/abs/1705.07874

Selleck, M. J., Senthil, M., & Wall, N. R. (2017). Making Meaningful Clinical Use of Biomarkers. *Biomarker insights*, 12, 1177271917715236. https://doi.org/10.1177/1177271917715236

Sheridan, P. J., & Crossman, D. C. (2002). Critical review of unstable angina and non-ST elevation myocardial infarction. *Postgraduate medical journal*, 78(926), 717–726. https://doi.org/10.1136/pmj.78.926.717

Shijin, L., Hao, W., Dingsheng, W., Jiali, Z. (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, 24 (1), 40-48, https://doi.org/10.1016/j.knosys.2010.07.003.

Simms, A. D., Batin, P. D., Kurian, J., Durham, N., & Gale, C. P. (2012). Acute coronary syndromes: an old age problem. *Journal of geriatric cardiology: JGC*, 9(2), 192–196. https://doi.org/10.3724/SP.J.1263.2012.01312

Smulyan H. (2019). The Computerized ECG: Friend and Foe. *The American journal of medicine*, 132(2), 153–160. https://doi.org/10.1016/j.amjmed.2018.08.025

Soliman E. Z. (2008). A simple measure to control for variations in chest electrodes placement in serial electrocardiogram recordings. *Journal of electrocardiology*, 41(5), 378–379. https://doi.org/10.1016/j.jelectrocard.2008.05.008

Southern, W. N., & Arnsten, J. H. (2009). The effect of erroneous computer interpretation of ECGs on resident decision making. Medical decision making : an *international journal of the Society for Medical Decision Making*, 29(3), 372–376. https://doi.org/10.1177/0272989X09333125

Stewart, R. A. H., Jones, P., Dicker, B., Jiang, Y., Smith, T., Swain, A. et al (2021). High flow oxygen and risk of mortality in patients with a suspected acute coronary syndrome: pragmatic, cluster randomised, crossover trial. *BMJ*, 372 (355). https://doi:10.1136/bmj.n355

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers?. *Current opinion in HIV and AIDS*, 5(6), 463–466. https://doi.org/10.1097/COH.0b013e32833ed177v

Tanantong, T., Nantajeewarawat, E., & Thiemjarus, S. (2015). False alarm reduction in BSN-based cardiac monitoring using signal quality and activity type information. *Sensors* (Basel, Switzerland), 15(2), 3952–3974. https://doi.org/10.3390/s150203952

The CONSORT-AI and SPIRIT-AI Steering Group., Liu, X., Rivera, S.C. et al. (2019). Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med*, 25, 1467–1468. https://doi.org/10.1038/s41591-019-0603-3

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (The IEEE Global Initiative). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e), Creative Commons At-tribution Non-Commercial 4.0 United States License. https://ethicsinaction.ieee.org/#read.

Tian, L., Li, C., Liu, Y., Chen, Y., & Fu, M. (2014). The value and distribution of high-density lipoprotein subclass in patients with acute coronary syndrome. *PloS one*, 9(1), e85114. https://doi.org/10.1371/journal.pone.0085114

Vera, L., Daniel, G., and Sarah, M. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery*, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

Walsh B. (2018). Misplacing V1 and V2 can have clinical consequences. *The American journal of emergency medicine*, 36(5), 865–870. https://doi.org/10.1016/j.ajem.2018.02.006

Wang, X. Y., Zhang, F., Zhang, C., Zheng, L. R., & Yang, J. (2020). The Biomarkers for Acute Myocardial Infarction and Heart Failure. *BioMed research international*, 2020, 2018035. https://doi.org/10.1155/2020/2018035

Wenger, W., & Kligfield, P. (1996). Variability of precordial electrode placement during routine electrocardiography. *Journal of electrocardiology*, 29(3), 179–184. https://doi.org/10.1016/s0022-0736(96)80080-x

World Health Organization (2017) Cardiovascular diseases facts sheet [online]. Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (Accessed: 26 March 2021)
World Health Organization (2021) Noncommunicable diseases [online]. Available at: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (Accessed: 26 March 2022)

Yan, W., Wen, S., Wang, L., Duan, Q., & Ding, L. (2015). Comparison of cytokine expressions in acute myocardial infarction and stable angina stages of coronary artery disease. International journal of clinical and experimental medicine, 8(10), 18082–18089.

Yu, K., Yang, B., Jiang, H., Li, J., Yan, K., Liu, X., Zhou, L., Yang, H., Li, X., Min, X., Zhang, C., Luo, X., Mei, W., Sun, S., Zhang, L., Cheng, X., He, M., Zhang, X., Pan, A., Hu, F. B., Wu, T. (2019). A multi-stage association study of plasma cytokines identifies osteopontin as a biomarker for acute coronary syndrome risk and severity. *Scientific reports*, 9(1), 5121. https://doi.org/10.1038/s41598-019-41577-4

Zeraatkar, E., Kermani, S., Mehridehnavi, A., Aminzadeh, A., Zeraatkar, E., & Sanei, H. (2011). Arrhythmia Detection based on Morphological and Time-frequency Features of T-wave in Electrocardiogram. Journal of medical signals and sensors, 1(2), 99–106.

Zhang, Q., Qiao, W., Zhou, L., Jin, H., Zheng, K. L., Zhao, D. S., & Lu, H. H. (2018). Correlations of soluble osteoclast-associated receptor (sOSCAR) with acute coronary syndrome. *Annals of translational medicine*, 6(20), 408. https://doi.org/10.21037/atm.2018.10.05

Zhang, Y., Lin, P., Jiang, H., Xu, J., Luo, S., Mo, J., Li, Y., & Chen, X. (2015). Extensive serum biomarker analysis in patients with ST segment elevation myocardial infarction (STEMI). *Cytokine*, 76(2), 356–362. https://doi.org/10.1016/j.cyto.2015.06.015

Zhen, D., Peiying, F. and Zhong, C. (2012). AS-239 Elevated Serum Surfactant Protein D is Associated with the Severity of Coronary Lesions and Poor Prognosis in Patients with Coronary Artery Disease and Chronic Kidney Disease. *American Journal of Cardiology*, 109 (7), S116. https://doi.org/10.1016/j.amjcard.2012.01.272

Zhu R, Liu C, Tang H, Zeng Q, Wang X, Zhu Z, Liu Y, Mao X, Zhong Y. Serum Galectin-9 Levels Are Associated with Coronary Artery Disease in Chinese Individuals. Mediators Inflamm. 2015;2015:457167. doi: 10.1155/2015/457167. Epub 2015 Nov 18. PMID: 26663989; PMCID: PMC4667018.

Zilke J. R., Loza M. E., Janssen F. (2016). DeepRED – Rule Extraction from Deep Neural Networks. In: Calders T., Ceci M., Malerba D. (eds) Discovery Science. DS 2016. *Lecture Notes in Computer Science*, 9956. Springer, Cham. https://doi.org/10.1007/978-3-319-46307-0_29