PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
UNIVERSIT´E KASDI MERBAH - OUARGLA
FACULTY OF NEW INFORMATION AND COMMUNICATION TECHNOLOGIES
DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS



Master's Thesis

Domain: Science and Technology

Sector: Telecommunications and Electronics

Specialty: Telecommunication Systems

THEME

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

Presented by:

Bensayah Sirina Rihab

Korichi Manar

Evaluation Date : 07/06/2026

In front of the Jury :

Name	Role	Institution
Dr. Bilal Benarabi	President	UKM Ouargla
Dr. Azeddine Benlamoudi	Supervisor	UKM Ouargla
Dr. Youcefa Abdelmadjid	Examiner	UKM Ouargla
Mr. Harrouz Salah Eddine	Co-supervisor	UKM Ouargla
Dr. Chahrazed Rahmani	Economic Partner	UKM Ouargla
Dr. Moukhtar Bounegab	Incubator representative	UKM Ouargla

Academic Year: 2025/2026

Dedication

I dedicate this work to my beloved parents for their endless love, sacrifices, prayers, and unwavering support throughout my life

To my dear brother and my two wonderful sisters, thank you for your encouragement, love, and constant support.

To my three beloved cats, who brought comfort and joy during this journey.

To my family and friends, thank you for believing in me and supporting me every step of the way.

This achievement is dedicated to all of you with love and gratitude.

Bensayah Sirina Rihab

Dedication

I dedicate this work to my beloved parents for their endless love, support, and encouragement.

To my dear sister and brothers, thank you for always standing by my side.

To my family and friends, thank you for your kindness and support throughout this journey.

This achievement is dedicated to all of you with sincere appreciation and gratitude.

Korichi Manar

Special thanks to our dear professor Mr. Harrouz Salah Eddine for his valuable help and contribution to this project.

Acknowledgement

Above all, we thank Almighty Allah for giving us strength, patience, good health, and perseverance throughout this journey, and without His blessings, this work would not have been possible.

We would like to express our sincere gratitude to Dr. Benlamoudi Azzeddine and Mr. Harrouz Salah Eddine for their valuable guidance, continuous support, and encouragement throughout the realization of this work. Their advice, assistance, and confidence in us were greatly appreciated and contributed significantly to the completion of this project.

Heartfelt gratitude is also extended to our jury members for their interest in our work and for accepting to evaluate this dissertation. We are grateful to our jury for their time, consideration, and valuable remarks.

We also wish to thank all the professors and staff of Kasdi Merbah University of Ouargla for the knowledge and support they provided throughout our academic journey.

Finally, we would like to express our deepest thanks to our families and friends for their endless support, encouragement, prayers, and unwavering confidence in us. We are equally grateful to everyone who contributed, directly or indirectly, to the completion of this work.

Abstract

Cardiovascular disease remains the leading cause of death globally, with myocardial infarction (MI) representing its most acute and time-critical manifestation. Current clinical diagnosis relies on hospital-based tools, namely the 12-lead electrocardiogram and serum troponin assays, which are effective but inaccessible during the critical prehospital period when early detection would have the greatest impact on patient outcomes. This thesis investigates whether photoplethysmography (PPG) signals acquired from wrist-worn wearable devices contain sufficient discriminative information to support machine learning-based MI classification, and which classifier is best suited for real-time deployment on resourceconstrained hardware.

A publicly available PPG dataset from the UCI Machine Learning Repository was used, containing 2,576 samples with 2,000 amplitude features each, labelled as either Normal or Myocardial Infarction with near-perfect class balance. The experimental pipeline applied StandardScaler normalization, Principal Component Analysis dimensionality reduction from 2,000 to 50 components, and 5-fold stratified cross-validation. Four supervised classification algorithms were evaluated: Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine.

All four classifiers exceeded 95 percent accuracy on the held-out test set. Gradient Boosting achieved the highest Area Under the ROC Curve at 0.954, confirming its superior ability to capture nonlinear patterns in the PCA-reduced feature space. Logistic Regression achieved an identical accuracy of 95.5 percent with an AUC of 0.946 and is recommended for wearable deployment due to its minimal computational requirements and interpretable decision boundary. These results establish a viable proof-of-concept for continuous, passive MI detection on consumer wearables, with real-world clinical validation identified as the essential next step.

Keywords: Myocardial Infarction, Photoplethysmography, PPG, Machine Learning, Wearable Devices, Classification, Gradient Boosting, Logistic Regression, Principal Component Analysis, Real-Time Detection.

Résumé

Les maladies cardiovasculaires demeurent la première cause de mortalité dans le monde, l'infarctus du myocarde (IM) en représentant la manifestation la plus aiguë et la plus critique dans le temps. Le diagnostic clinique actuel repose sur des outils hospitaliers, notamment l'électrocardiogramme à 12 dérivations et les dosages de troponine sérique, qui sont efficaces mais inaccessibles durant la période préhospitalière critique où une détection précoce aurait le plus grand impact sur les résultats des patients. Cette thèse examine si les signaux de photopléthysmographie (PPG) acquis par des dispositifs portables au poignet contiennent suffisamment d'informations discriminantes pour prendre en charge la classification de l'IM par apprentissage automatique, et quel classificateur est le mieux adapté au déploiement en temps réel sur du matériel à ressources limitées.

Un jeu de données PPG disponible publiquement, issu du référentiel UCI Machine Learning Repository, a été utilisé, comprenant 2 576 échantillons avec 2 000 caractéristiques d'amplitude chacun, étiquetés comme Normaux ou Infarctus du Myocarde avec un équilibre de classes quasi parfait. Le pipeline expérimental a appliqué une normalisation StandardScaler, une réduction de dimensionnalité par Analyse en Composantes Principales de 2 000 à 50 composantes, et une validation croisée stratifiée à 5 plis. Quatre algorithmes de classification supervisée ont été évalués : Régression Logistique, Forêts Aléatoires, Gradient Boosting, et Machine à Vecteurs de Support.

Les quatre classificateurs ont dépassé une précision de 95 pour cent sur l'ensemble de test réservé. Le Gradient Boosting a obtenu la meilleure aire sous la courbe ROC avec 0,954, confirmant sa capacité supérieure à capturer les motifs non linéaires dans l'espace de caractéristiques réduit par ACP. La Régression Logistique a atteint une précision identique de 95,5 pour cent avec une AUC de 0,946 et est recommandée pour le déploiement sur dispositif portable en raison de ses exigences de calcul minimales et de sa frontière de décision interprétable. Ces résultats établissent une preuve de concept viable pour la détection continue et passive de l'IM sur des appareils portables grand public, la validation clinique en conditions réelles étant identifiée comme l'étape suivante essentielle.

Mots-clés : Infarctus du Myocarde, Photopléthysmographie, PPG, Apprentissage Automatique, Dispositifs Portables, Classification, Gradient Boosting, Régression Logistique, Analyse en Composantes Principales, Détection en Temps Réel.

الملخص

لا تزال أمراض القلب والأوعية الدموية تُمثّل السبب الرئيس ي للوفاة على مستوى العالم، ويُعدّ احتشاء عضلة القلب أشدّ مظاهرها حدةً وأكثرها إلحاحًا من حيث الوقت. يعتمد التشخيص السريري الحالي على أدوات مستشفياتية، في مقدمتها تخطيط القلب الكهربائي ذو الاثنتي عشرة قناة وقياسات التروبونين المصلي، وهي أدوات فعّالة إلا أنها غير متاحة خلال الفترة الحرجة السابقة للتدخل الطبي، وهي الفترة التي يكون فيها الكشف المبكر أعظم أثرًا على نتائج المريض. تبحث هذه الرسالة فيما إذا كانت إشارات قياس التحجم الضوئ ي (PPG) المسُجَّلة عبر الأجهزة القابلة للارتداء على الرسغ تحتوي على معلومات تمييزية كافية لدعم تصنيف احتشاء عضلة القلب باستخدام التعلم الآلي، وأيّ خوارزمية تصنيف هي الأنسب للنشر الآني على أجهزة ذات موارد محدود ة.

استُخدم في هذه الدراسة مجموعة بيانات PPG متاحة للعموم من مستودع UCI للتعلم الآلي، تضمّ 2576 عيّنة بـ 2000 ميزة سعة لكل منها، مصنَّفة إما بوصفها طبيعية أو احتشاء عضلة القلب، مع توازن شبه تام بين الفئتين. اعتمد خط الأنابيب التجريبي على معياري ةStandardScaler ، وتقليص الأبعاد عبر تحليل المكوّنات الرئيسية من 2000 إلى 50 مكوّنًا، والتحقق المتقاطع الطبقي بخمسة أطواء. جرى تقييم أربع خوارزميات تصنيف إشرافي، هي: الانحدار اللوجستي، والغابات العشوائية، وتعزيز التدرج، وآلة المتجهات الداعمة.

تجاوزت جميع الخوارزميات الأربع دقة 95 بالمئة على مجموعة الاختبار. حقّق تعزيز التدرج أعلى قيمة للمساحة تحت منحنى ROC بنسبة 954.0، مما يؤكد قدرته المتفوقة على التقاط الأنماط غير الخطية في فضاء الميزات المُخفَّضة. أما الانحدار اللوجستي فقد حقّق دقة مماثلة بلغت 5.95 بالمئة مع AUC قدرها 946.0، ويُوص ى باعتماده في النشر على الأجهزة القابلة للارتداء نظرًا لمتطلباته الحسابية الضئيلة وحدوده القرارية القابلة للتفسير. تُرس ي هذه النتائج مفهوم إثبات الجدوى لنظام الكشف المستمر والسلبي عن احتشاء عضلة القلب عبر الأجهزة الاستهلاكية القابلة للار تداء، فيما يُحدَّد التحقق السريري في الظروف الواقعية بوصفه الخطوة الجوهرية التالية.

الكلمات المفتاحية: **احتشاء عضلة القلب، قياس التحجم الضوئي ،PPG، التعلم الآلي، الأجهزة القابلة للارتداء، التصنيف، تعزيز التدرج، الانحدار اللوجستي، تحليل المكوّنات الرئيسية، الكشف الآني.**

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
AUC	Area Under the Curve
BLE	Bluetooth Low Energy
CE	Conformité Européenne
CNN	Convolutional Neural Network
CV	Cross-Validation
ECG	Electrocardiogram
EDA	Exploratory Data Analysis
FDA	Food and Drug Administration
FN	False Negative
FP	False Positive
F1	F1 Score (Harmonic Mean of Precision and Recall)
GB	Gradient Boosting
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HRV	Heart Rate Variability
IoT	Internet of Things
LED	Light Emitting Diode
LR	Logistic Regression
LSTM	Long Short-Term Memory
mAh	Milliampere-hour
MDR	Medical Device Regulation
MI	Myocardial Infarction
ML	Machine Learning
NSTEMI	Non-ST-Elevation Myocardial Infarction
PCA	Principal Component Analysis
PCI	Percutaneous Coronary Intervention
PD	Photodetector

PPG	Photoplethysmography
PTT	Pulse Transit Time
RF	Random Forest
ROC	Receiver Operating Characteristic
Abbreviation	**Full Term**
RBF	Radial Basis Function
SaMD	Software as a Medical Device
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
SpO2	Peripheral Oxygen Saturation
STEMI	ST-Elevation Myocardial Infarction
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UCI	University of California Irvine
WHO	World Health Organization

List of Tables

Number	Caption / Description	Page
3.1	Performance Comparison of All Four Classifiers on the Test Set	**39**

List of Figures

Number	Caption / Description	Page
1.1	Normal EKG, STEMI, and NSTEMI waveform patterns alongside coronary artery cross-section illustrating plaque accumulation, thrombus formation, and myocardial damage	**6**
1.2	PPG sensor in reflective configuration on the fingertip showing LED and photodetector arrangement, and corresponding raw waveform with systolic peak, dicrotic notch, and diastolic trough	**8**
1.3	Wrist-worn PPG sensor cross-section showing LED and photodetector placement, light penetration through tissue layers, and separation of the pulsatile PPG signal from the static DC offset	**10**
1.4	Concurrent ECG and PPG recordings over a five-second window showing temporal alignment between cardiac electrical depolarization events and subsequent mechanical blood volume pulses	**11**
1.5	Architecture of a wearable health monitoring system showing sensor data acquisition, transmission via Wi-Fi to cloud server and database, and delivery to multiple receiving devices	**13**
2.1	Feature distributions before and after StandardScaler normalization	**20**
2.2	Visualization of SMOTE oversampling showing two-class distribution in a two-dimensional feature space with synthetic minority class samples	**21**
2.3	Explained variance by principal components, showing individual and cumulative explained variance for the first 10 components	**23**
2.4	The logistic (sigmoid) function used in Logistic Regression, mapping any real-valued input to a probability between 0 and 1	**25**
2.5	Random Forest ensemble architecture showing N decision trees trained independently with majority voting to produce the final classification result	**26**
2.6	Gradient Boosting sequential training process showing each tree trained on residuals of the previous ensemble	**27**

2.7	SVM classification with maximum margin hyperplane identifying the decision boundary that maximally separates two classes	**28**
Number	**Caption / Description**	**Page**
2.8	Confusion matrix for binary classification showing TP, FN (Type II Error), FP (Type I Error), and TN with derived metric formulas	**29**
2.9	ROC curve for Logistic Regression on the PPG MI classification task plotting True Positive Rate against False Positive Rate across all decision thresholds	**30**
3.1	Class distribution of the PPG dataset showing 1,282 Normal samples (49.8%) and 1,294 MI samples (50.2%)	**35**
3.2	Representative PPG waveforms from a Normal cardiac recording and a Myocardial Infarction recording showing differences in amplitude and morphology	**36**
3.3	PCA cumulative explained variance curve showing proportion of total variance captured as number of principal components increases (50 components ≈ 75% variance)	**37**
3.4	Five-fold stratified cross-validation AUC scores for all four classifiers with mean AUC values and standard deviations	**39**
3.5	ROC curves for all four classifiers on the test set plotting True Positive Rate against False Positive Rate across all decision thresholds	**40**
3.6	Confusion matrices for all four classifiers on the test set (516 samples)	**41**
3.7	Top 20 most important principal components as determined by Random Forest mean impurity decrease across all trees	**42**
3.8	Prototype wrist-worn wearable device with integrated PPG sensor and real-time display	**45**
3.9	Proposed end-to-end wearable MI detection system architecture showing data flow from PPG acquisition through preprocessing, ML classification, decision engine, and alert delivery	**46**

Table of Content

Dedication ..... **I**

Dedication ..... **II**

Acknowledgement ..... **III**

Abstract ..... **IV**

Résumé ..... **V**

List of Abbreviations ..... **VII**

List of Tables ..... **IX**

List of Figures ..... **X**

Table of Content ..... **XII**

General Introduction **0**

Chapter I: Theoretical Foundations: **4**

1.1 Introduction **4**

1.2 Myocardial Infarction (MI) **5**

1.2.1 Definition and Pathophysiology **5**

1.2.2 Global Burden and Risk Factors **6**

1.2.3 Diagnosis and the Time-to-Treatment Problem **7**

1.3 Photoplethysmography (PPG) **8**

1.3.1 Operating Principle and Sensor Types **8**

1.3.2 Waveform Morphology and Information Content **9**

1.3.3 Relationship Between PPG and ECG **10**

1.3.4 Known Limitations of PPG Measurement **10**

1.4 Wearable Technology in Healthcare **11**

1.4.1 Development of Consumer Health Wearables **11**

1.4.2 Inference Architecture for Real-Time Applications 12

1.4.3 Regulatory Considerations 13

1.4.4 Remaining Open Challenges 14

1.5 Conclusion 14

***Chapter II: Machine Learning for Medical Signal Classification* 17**

2.1 Introduction to Machine Learning in Healthcare 17

2.2 Machine Learning: Definition and Types 18

2.3 Data Preprocessing 19

2.3.1 Feature Scaling with StandardScaler 19

2.3.2 Handling Class Imbalance with SMOTE 20

2.3.3 Train/Test Split and Cross-Validation 21

2.4 Dimensionality Reduction: Principal Component Analysis 22

2.5 Classification Algorithms 23

2.5.1 Logistic Regression 23

2.5.2 Random Forest 25

2.5.3 Gradient Boosting 26

2.5.4 Support Vector Machine 27

2.6 Model Evaluation Metrics 28

2.6.1 The Confusion Matrix 28

2.6.2 Metrics Derived from the Confusion Matrix 29

2.6.3 ROC Curve and AUC 29

2.7 Ethical and Clinical Considerations 30

2.8 Conclusion 32

***Practical Implementation and Results* 33**

***Chapter III: Practical Implementation and Results* 34**

3.1 Introduction **34**

3.2 Dataset Description **34**

3.3 Exploratory Data Analysis **35**

3.3.1 Class Distribution **35**

3.3.2 PPG Waveform Analysis **36**

3.4 Data Preprocessing **37**

3.5 Model Training and Cross-Validation **38**

3.6 Results and Evaluation **39**

3.6.1 Overall Performance Comparison **39**

3.6.2 ROC Curves **40**

3.6.3 Confusion Matrices **41**

3.6.4 Feature Importance **42**

3.7 Discussion **42**

3.7.1 Why Gradient Boosting Leads on AUC **42**

3.7.2 The Case for Logistic Regression in Deployment **43**

3.7.3 Sensitivity/Specificity Trade-off in Medical Context **43**

3.7.4 Limitations of This Study **44**

3.8 Proposed Wearable System Architecture **45**

3.9 Conclusion **47**

General Conclusion **49**

References **51**

Introduction

General Introduction

Every year, cardiovascular disease claims more lives than any other condition on the planet. Among its most dangerous manifestations, myocardial infarction stands out not only for the scale of its mortality but for the speed at which it causes irreversible harm. The cardiac muscle cells that begin dying within twenty minutes of a blocked coronary artery do not recover. The tissue destroyed in the first hour of an untreated MI is tissue lost permanently, and the longterm consequences, ranging from reduced heart function to congestive heart failure, follow the patient for the rest of their life. This is why time is the central variable in MI management, and why the research presented in this thesis begins from a single, concrete observation: the tools that can detect MI reliably are almost entirely confined to hospital settings, and the patients who need detection most are often not in a hospital when the event begins.

The gap between symptom onset and clinical presentation is where most of the preventable damage in MI management occurs. Real-world data from emergency medicine registries consistently shows median delays of two hours or more between the first physiological signs of ischemia and the moment a patient arrives at a facility equipped to diagnose and treat them. During those two hours, blood flow to a section of the heart is either absent or severely reduced, and the zone of necrosis expands with each passing minute. The golden hour concept, borrowed from trauma surgery, captures this relationship precisely: the faster the intervention, the greater the proportion of at-risk tissue that survives. Wearable monitoring, which could detect the physiological signatures of an evolving MI and generate an alert before the patient becomes symptomatic enough to seek care independently, represents one of the most promising approaches to addressing this gap.

Photoplethysmography technology has made this approach technically feasible in a way that was not possible even a decade ago. PPG sensors, which measure changes in blood volume beneath the skin by detecting how tissue absorbs and reflects light, are now standard components of almost every consumer smartwatch and fitness tracker on the market. These devices sit on hundreds of millions of wrists worldwide, running passively and continuously throughout the day and night. The signal they produce is richer than most users appreciate: it encodes information about heart rate and heart rate variability, arterial stiffness, blood pressure, and autonomic nervous system activity, all of which may be altered in characteristic ways during a myocardial ischemic event. The central scientific question this thesis addresses is whether that information is detectable by a trained machine learning classifier operating on the PPG waveform features alone.

To investigate this question, the thesis is organized into three chapters that build on one another systematically. The first chapter establishes the theoretical foundation by examining the pathophysiology of myocardial infarction in clinical detail, explaining the physics and physiology of photoplethysmography from first principles, and surveying the current state of wearable health technology including its capabilities, limitations, and regulatory landscape. The second chapter addresses the machine learning methodology, covering the preprocessing pipeline, dimensionality reduction through Principal Component Analysis, the theory and mathematical formulation of each of the four classification algorithms tested, and the evaluation metrics chosen to reflect clinical priorities. The third chapter presents the complete experimental implementation: the dataset characteristics, exploratory data analysis, preprocessing steps, cross-validation results, and the full performance comparison across classifiers, followed by a discussion of what the results mean for the feasibility of wearable MI detection and a proposal for a practical system architecture.

The thesis does not claim to have solved the problem of early MI detection. What it claims, and what the results support, is that the PPG signal contains enough discriminative information for machine learning classifiers to distinguish MI from normal cardiac states with accuracy above 95 percent on a controlled research dataset, and that at least one of those classifiers, Logistic Regression, is computationally simple enough to run continuously on a wrist-worn device without unacceptable battery drain. That is a meaningful finding, and it is the foundation on which the more ambitious next steps, prospective clinical validation, personalized modeling, and regulatory approval, would need to be built.

Chapter I:

Theoretical Foundations

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

Chapter I: Theoretical Foundations:

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

1.1 Introduction

Cardiac events do not wait for convenient moments. A myocardial infarction can begin while someone is asleep, driving, or simply sitting at a desk with no obvious warning. What makes this especially difficult from a healthcare perspective is that the interventions capable of preventing permanent damage are almost entirely hospital-dependent, and hospitals are not always close, not always reachable quickly, and not always sought out in time. The mismatch between where MI begins and where it can be treated is, in many ways, the central problem that this thesis tries to address.

To understand why that mismatch matters so much, it helps to think about what actually happens during a heart attack at the tissue level. Cardiac muscle cells begin dying within roughly 20 minutes of losing their blood supply, and that process continues until either the blockage is cleared or the tissue is beyond recovery. Every hour of delay corresponds to measurable additional loss of functional myocardium, which translates directly into worse longterm outcomes, including reduced cardiac output, higher risk of heart failure, and greater mortality (Ibanez et al., 2018). The urgency is not rhetorical; it is physiological.

The tools currently used to confirm MI are well-established and clinically reliable. A 12-lead electrocardiogram can reveal the characteristic changes associated with ST-elevation within minutes of onset. Blood tests measuring troponin levels provide biochemical evidence of cell death, although the elevation takes at least an hour to become detectable and requires laboratory analysis. Imaging studies add further detail. None of these, however, are things a person carries on their wrist while walking to work. They exist in hospitals, clinics, and emergency vehicles, and they require trained personnel to administer and interpret (Thygesen et al., 2018).

The motivation for exploring wearable-based detection comes from this gap. Photoplethysmography sensors are now embedded in almost every consumer smartwatch and fitness tracker. They are cheap, low-power, and designed to operate continuously throughout the day and night. The signal they produce reflects changes in blood volume beneath the skin with each heartbeat, and from that signal it is possible to extract a range of cardiovascular parameters. Whether those parameters shift in detectable ways during a myocardial ischemic event is the question at the heart of this research.

This thesis evaluates four machine learning classifiers applied to a PPG dataset labelled for MI versus normal cardiac states. The four algorithms compared are Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine. The research was organized around four objectives: evaluating whether PPG features carry enough information for MI classification; comparing classifier performance across standard diagnostic metrics; identifying the best candidate for real-time deployment on a wearable device; and proposing a practical system architecture for a wearable-based alert mechanism.

There are two questions that run through the entire study. The first is whether a lightweight trained model can detect MI with sensitivity high enough to be clinically useful. The second is how the trade-off between sensitivity and specificity should be calibrated when the cost of a missed detection is potentially lethal. These questions cannot be resolved through a single academic study working on controlled data, but they can be examined seriously, and this thesis tries to do exactly that.

The rest of this chapter provides the conceptual foundation needed to follow the methodology and results. Section 1.2 covers what MI is, how it develops, and why timing matters so much. Section 1.3 explains photoplethysmography in detail, from the optical physics of the sensor to the physiological information buried in the waveform. Section 1.4 looks at the current state of wearable health technology, including its capabilities, its known limitations, and the regulatory considerations that would apply to any clinical deployment.

1.2 Myocardial Infarction (MI)

1.2.1 Definition and Pathophysiology

The heart is one of the few organs in the body that cannot stop working, even briefly, without causing serious harm. It pumps roughly 100,000 times per day, and the muscle doing that work requires a continuous oxygen supply delivered through the coronary arteries. When one of those arteries becomes blocked and blood flow stops, the muscle cells downstream begin to die within minutes. This is myocardial infarction: the death of cardiac tissue caused by sustained ischemia (Thygesen et al., 2018).

The process that leads to arterial blockage usually begins years or decades before the acute event. Fatty deposits called plaques form inside the arterial wall over time, narrowing the vessel and making it less flexible. Most plaques remain stable for long periods. The acute trigger for MI is typically a rupture of one of these plaques, which exposes the lipid core to flowing blood and triggers rapid clot formation at the rupture site. Depending on how completely the clot blocks the vessel, the result is classified either as ST-elevation myocardial infarction, where

the artery is fully occluded, or non-ST-elevation MI, where partial flow is maintained (Thygesen et al., 2018). Both types cause tissue death, though the timeline and extent differ.

One aspect of this pathophysiology that has direct implications for detection is that many people experience little or no warning before the acute event. Atherosclerotic disease can be extensive and even hemodynamically significant without producing symptoms during routine activity. This is particularly common in patients with diabetes, in whom autonomic neuropathy can blunt the usual pain signals, and in women, whose MI presentations often diverge from the classic substernal chest pain pattern that most people associate with a heart attack (Roth et al., 2020). For these individuals, passive continuous monitoring represents the only realistic chance of early detection.

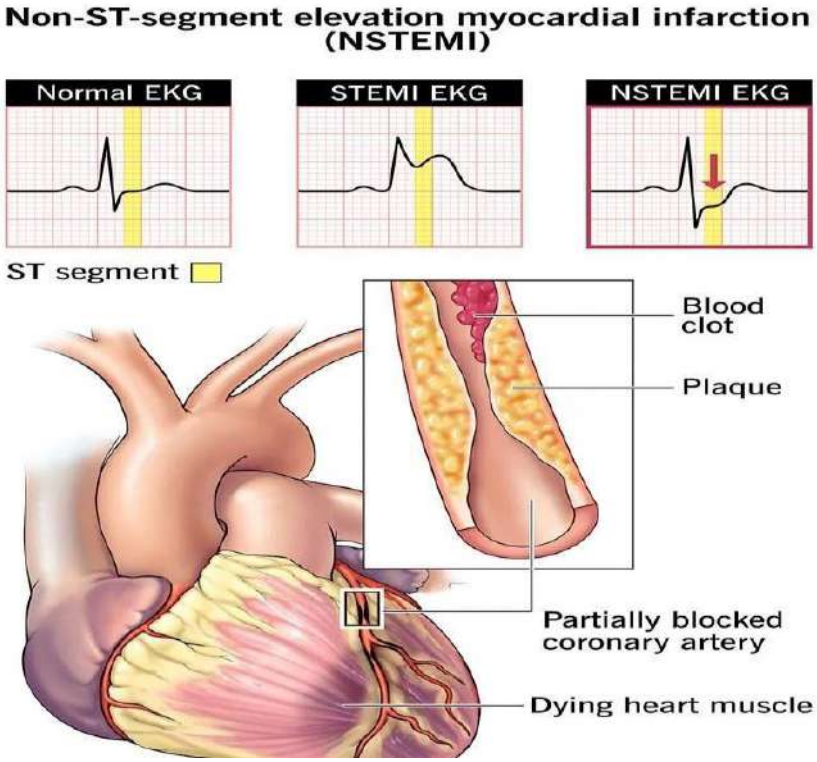


Figure 1.1: Normal EKG, STEMI, and NSTEMI waveform patterns shown alongside a coronary artery crosssection illustrating plaque accumulation, thrombus formation, and resulting myocardial damage. Source: Cleveland Clinic (2025). Retrieved from https://my.clevelandclinic.org/health/diseases/22233-nstemi-heart-attack

1.2.2 Global Burden and Risk Factors

Cardiovascular disease accounts for more deaths worldwide than any other category of illness. Within that group, ischemic heart disease is the leading cause, and MI is its most acute and dangerous manifestation. The 2020 Global Burden of Disease analysis estimated approximately 9.14 million deaths from ischemic heart disease in 2019, representing around 16

percent of all deaths globally that year (Roth et al., 2020). Notably, these figures have not declined as sharply as mortality from some other major causes, partly because the metabolic risk factors underlying coronary artery disease, particularly obesity and type 2 diabetes, have grown more prevalent in many populations.

Risk factors for MI are conventionally divided into those that cannot be changed and those that can. Non-modifiable factors include age, male sex, and a family history of early coronary disease. Modifiable factors include hypertension, dyslipidemia, smoking, physical inactivity, and diabetes (Roth et al., 2020). The geographic distribution of mortality also reflects this: high-income countries have achieved substantial reductions in age-standardized MI death rates through improved acute care, wider use of statins, and declining smoking rates, while low- and middle-income countries continue to face rising cardiovascular burdens as their populations age and urbanize. This disparity adds relevance to the development of low-cost, accessible detection tools, since the populations with the least clinical infrastructure to offer may benefit most from early warning capability at the patient level.

1.2.3 Diagnosis and the Time-to-Treatment Problem

The diagnostic pathway for a patient presenting with suspected MI typically begins with a 12-lead ECG, which can confirm STEMI almost immediately from the characteristic STsegment elevation pattern. NSTEMI diagnosis is less immediate: it depends on troponin measurements taken at presentation and repeated 1-3 hours later, with the diagnosis confirmed only once biomarker elevation is evident (Ibanez et al., 2018). High-sensitivity troponin assays have improved the speed of this process considerably compared to older tests, but the patient must still be in a clinical setting for the test to be run at all.

The concept of the golden hour, borrowed from trauma surgery, captures why time matters so acutely in MI management. Myocardial salvage, meaning the proportion of at-risk tissue that can be preserved through timely reperfusion, falls steeply with each hour of untreated ischemia. Guidelines from the European Society of Cardiology specify a door-to-balloon time of no more than 90 minutes for primary percutaneous coronary intervention in STEMI, a target that reflects extensive data showing that delays beyond this threshold are associated with measurably worse ejection fraction and higher 30-day mortality (Ibanez et al., 2018). The challenge, however, is that door-to-balloon time only starts once the patient has arrived. In realworld registries, the median time from symptom onset to hospital presentation routinely exceeds two hours. It is this prehospital delay that wearable monitoring has the potential to reduce, not by diagnosing MI, but by detecting physiological changes early enough to prompt the patient or their emergency contacts to act before the situation becomes irreversible.

1.3 Photoplethysmography (PPG)

1.3.1 Operating Principle and Sensor Types

Photoplethysmography is an optical technique that uses light to track changes in blood volume within a tissue. The basic setup involves an LED emitting light at a defined wavelength, typically green at around 530 nm for wrist-surface measurements or near-infrared at 880 nm for deeper penetration, and a photodetector that measures the intensity of light either transmitted through or reflected back from the tissue. Blood absorbs more light at the systolic peak of each cardiac cycle, when arterial volume is highest, and less during diastole as the vessels relax. The photodetector records this fluctuation, and the resulting alternating signal is the PPG waveform (Allen, 2007).

There are two sensor configurations in common use. Transmission-mode PPG places the emitter and detector on opposite sides of the tissue, with the fingertip being the most practical site. This approach gives a strong, clean signal but is unsuitable for wrist-worn devices. Reflective-mode PPG uses a sensor layout where both the LED and photodetector face the same side of the tissue, with the detector capturing light that has scattered back from the skin and underlying vessels. This is the configuration used in all modern smartwatches. Reflective mode trades some signal quality for wearability, and it is more susceptible to noise from movement because any shift in the position of the sensor relative to the tissue changes the optical path (Allen, 2007). Managing that noise is one of the main engineering challenges in consumer wearable design.

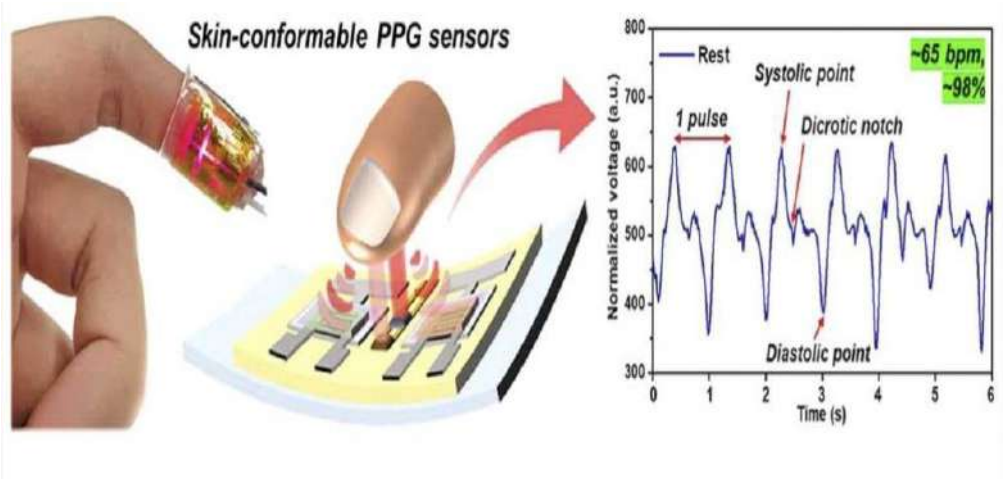


Figure 1.2: PPG sensor in reflective configuration on the fingertip, showing the LED and photodetector arrangement and the corresponding raw waveform with systolic peak, dicrotic notch, and diastolic trough labelled. Source: ScienceDirect (2021). Retrieved from

https://www.sciencedirect.com/science/article/abs/pii/S2211285521010223

1.3.2 Waveform Morphology and Information Content

Each individual pulse in a PPG recording has a characteristic shape that reflects the mechanical events of the cardiac cycle. The waveform rises sharply as the left ventricle ejects blood and pressure propagates through the arterial tree, reaches a systolic peak corresponding to maximum vessel distension, then shows a brief secondary inflection called the dicrotic notch, which coincides with closure of the aortic valve and the return of the reflected pressure wave from the periphery. After the notch, the waveform declines gradually through diastole as the arterial pressure falls between beats (Kyriacou and Allen, 2021). These features are not merely descriptive; their relative timing, amplitude, and shape encode information about the cardiovascular system that goes well beyond simple heart rate.

Several physiological parameters can be derived from PPG signals. Heart rate and heart rate variability are the most commonly used. Pulse transit time, calculated as the interval between the ECG R-wave and the foot of the PPG pulse, correlates with arterial blood pressure because it reflects how quickly the pressure wave travels through the vasculature, a function of arterial wall stiffness. Oxygen saturation is computed from the ratio of pulsatile signals at two different wavelengths, since oxygenated and deoxygenated hemoglobin absorb light differently. The augmentation index, defined as the ratio of the late systolic peak to the early systolic peak, provides a measure of arterial stiffness and wave reflection (Charlton et al., 2022). The relevance of all this for MI detection lies in the fact that ischemia affects cardiac output, autonomic tone, and vascular resistance, meaning that its fingerprint may be distributed across multiple PPG-derived features simultaneously.

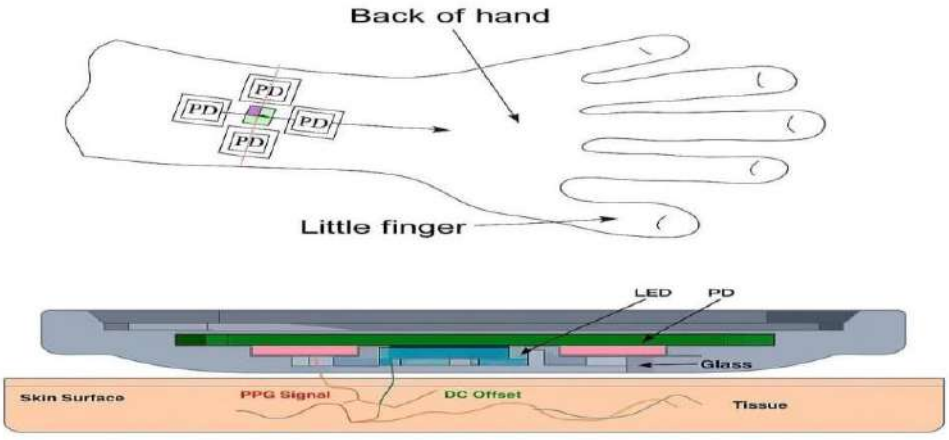


Figure 1.3: Wrist-worn PPG sensor cross-section showing LED and photodetector placement against the skin surface, with illustration of light penetration through tissue layers and the separation of the pulsatile PPG signal from the static DC offset. Source: AllPCB ElectroHub (2023). Retrieved from https://www.allpcb.com/allelectrohub/ppg-optical-path-and-process-design-for-wearable-devices

1.3.3 Relationship Between PPG and ECG

It is a fairly common misconception that PPG is simply a less precise version of ECG. The two signals actually measure different things. ECG captures the electrical activation sequence of the myocardium, the depolarization and repolarization currents that drive contraction. PPG captures a downstream mechanical outcome of that activation: the pressure pulse that travels through the arterial tree as a result of ventricular ejection. They are related, since one causes the other, but they are not interchangeable. Conditions that affect electrical conduction without changing mechanical output will show up on ECG but not necessarily on PPG. Conversely, changes in vascular tone, peripheral resistance, or arterial compliance can shift the PPG waveform substantially while leaving the ECG entirely normal (Elgendi et al., 2020).

For the practical context of this research, the relevant question is not which signal is more informative in a controlled clinical setting. It is which signal is continuously available on a wearable device. No currently available consumer wearable offers continuous 12-lead ECG monitoring. Single-lead wrist ECG recordings, available on some smartwatches, require the user to deliberately trigger a recording by placing a finger on a sensor, and produce only 30second snapshots rather than continuous data. PPG runs passively and continuously in the background on virtually every modern fitness tracker and smartwatch, making it the only realistic option for the kind of continuous cardiovascular surveillance that early MI detection would require.

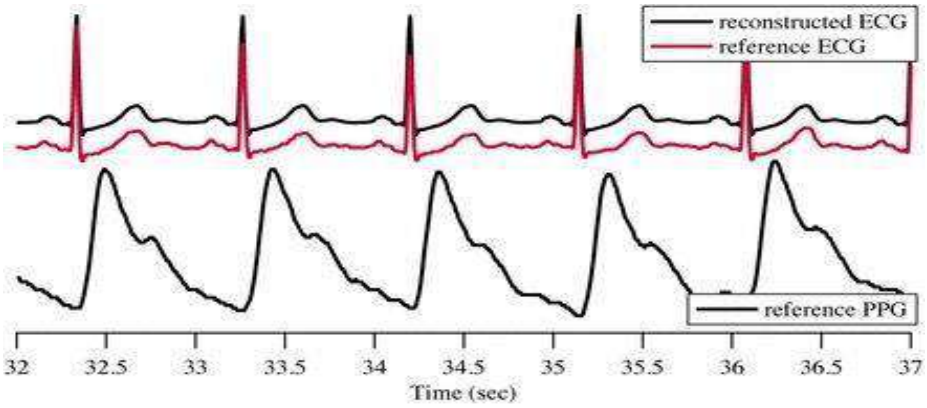


Figure 1.4: Concurrent ECG and PPG recordings over a five-second window, showing the consistent temporal alignment between each cardiac electrical depolarization event and the subsequent mechanical blood volume pulse. Source: bioRxiv (2019). Retrieved from https://www.biorxiv.org/content/10.1101/815258v1.full

1.3.4 Known Limitations of PPG Measurement

Any honest account of PPG technology has to acknowledge where it falls short. Motion artifact is the most persistent challenge. Physical movement introduces mechanical noise into the optical signal that can completely mask the cardiac waveform, particularly during vigorous

exercise. Hardware solutions using accelerometers to cancel motion-related interference have improved the situation substantially in modern wearables, but performance still degrades during intense physical activity (Bent et al., 2020). For a monitoring application focused on resting cardiovascular state, this is less of a concern than it would be for exercise tracking, but it remains a real limitation during everyday activities involving arm movement.

Skin melanin concentration affects how much green-wavelength light reaches the underlying vessels and how much is absorbed by the epidermis before getting there. Earlier consumer wearables relying primarily on green LEDs showed lower signal quality in darkerskinned individuals, a disparity that has drawn significant and justified attention from both engineers and medical ethicists (Bent et al., 2020). Many devices now include near-infrared channels that penetrate more deeply and are less affected by surface pigmentation. The dataset used in this study contains no demographic information about participants, so whether this limitation affects the results here is impossible to assess directly, though it is a factor that would need careful attention in any clinical deployment aimed at a diverse population.

Temperature, sensor contact pressure, placement variation across the wrist, and interindividual differences in subcutaneous fat thickness all contribute additional variability to PPG signals. These sources of noise are manageable in laboratory conditions, where they were presumably controlled when the dataset used here was collected. They are much harder to control in free-living use. The gap between controlled and real-world performance is one of the central challenges facing any wearable-based clinical monitoring system.

1.4 Wearable Technology in Healthcare

1.4.1 Development of Consumer Health Wearables

Consumer health wearables have changed considerably over the past fifteen years, and the direction of that change has been consistently toward more clinical capability. The early devices counted steps and estimated activity calories; they carried no real physiological sensors. The introduction of optical heart rate monitoring around 2013 added the first continuously acquired physiological signal to mass-market wearables, and PPG became standard equipment in the category within a few years. By 2018, when Apple received FDA clearance for an ECG function on the Series 4 watch, the boundary between consumer electronics and medical devices had become genuinely ambiguous (Perez et al., 2019).

PPG is now effectively universal in the category. Apple Watch, Samsung Galaxy Watch, Fitbit, Garmin, and Polar all ship devices with reflective PPG sensors as a baseline feature. Several platforms have obtained regulatory clearance for specific health claims based on PPG-

derived measurements, including SpO2 monitoring and atrial fibrillation detection algorithms. The clinical validation behind these claims varies considerably in quality and in how well the validation populations reflect the diversity of actual users, but the existence of cleared PPG-based medical functions on consumer devices establishes that this pathway is viable in principle (Seshadri et al., 2020).

1.4.2 Inference Architecture for Real-Time Applications

A wearable MI detection system would need to acquire raw PPG data, preprocess it to remove noise, extract relevant features, and run a classification algorithm, all with a latency short enough to be practically useful for alerting. Where in the system this pipeline runs has significant implications for both performance and feasibility. Three options exist: computation directly on the wearable microprocessor, computation on the paired smartphone, or computation in the cloud with data sent via internet.

On-device inference avoids both network dependency and the latency of wireless transmission, making it the most reliable option for a time-critical application. The tradeoff is that current wearable processors have very limited compute capacity, and running a complex model continuously would drain the battery within hours rather than days. This computational constraint is one reason why the study in Chapter 3 pays particular attention to Logistic Regression: it is the simplest and most computationally efficient of the four classifiers tested, and it achieves competitive accuracy. If it can perform adequately, it is the most deployable candidate. Smartphone inference offers a reasonable middle ground, with better compute resources and a more manageable battery impact on the wearable itself, as long as Bluetooth connectivity is maintained (Dunn et al., 2021).

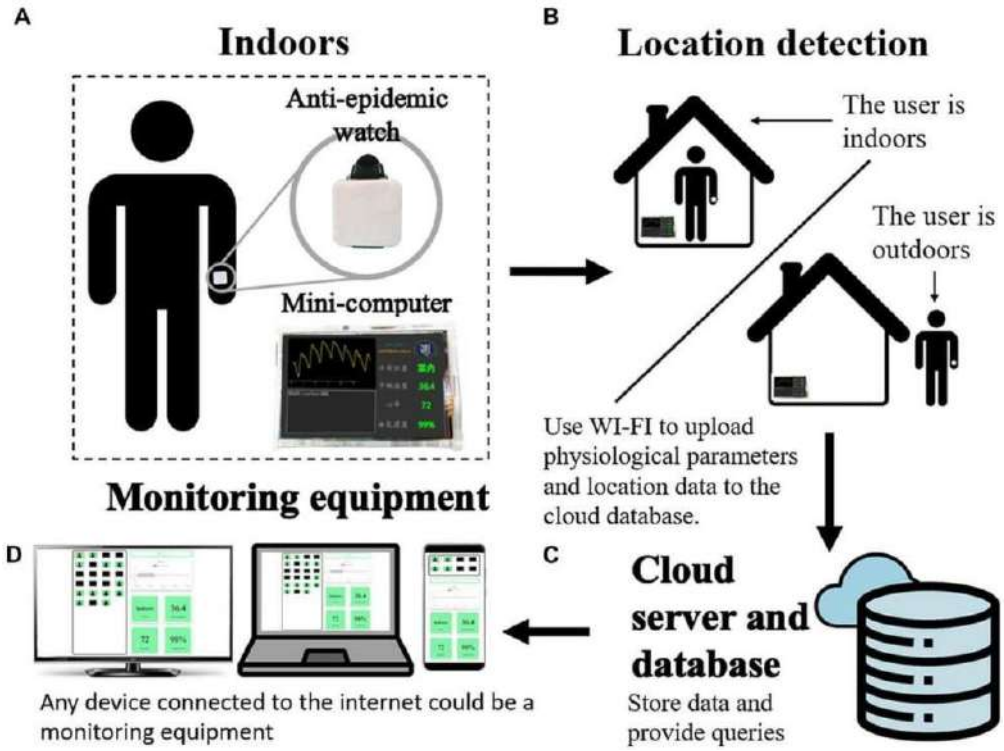


Figure 1.5: Architecture of a wearable health monitoring system, showing sensor data acquisition on the wrist device, transmission via Wi-Fi to a cloud server and database, and delivery of monitoring outputs to multiple receiving devices including smartphone, tablet, and desktop computer. Source: Frontiers in Public Health (2023). Retrieved from https://www.frontiersin.org/journals/publichealth/articles/10.3389/fpubh.2023.1188304/full

1.4.3 Regulatory Considerations

Software that analyzes physiological signals to detect or warn about a medical condition falls under medical device regulations in most major jurisdictions. In the United States, this category is classified as Software as a Medical Device by the FDA, and the appropriate premarket review pathway depends on the risk level of the intended use (FDA, 2019). A system that generates a recommendation to call emergency services based on a suspected MI would carry a high risk classification and would require robust clinical validation before approval. The more modest framing of an early warning or risk alert may receive a more streamlined review, but the requirement for prospective clinical evidence would still apply.

The European Union applies the Medical Device Regulation framework, and guidance specific to AI-based diagnostic software was still being finalized at the time this thesis was written. The general principle is consistent across regulatory systems: higher-risk claims require stronger evidence. A single retrospective analysis on a publicly available research dataset, which is what this study provides, is not the kind of evidence that would support regulatory submission. It is the kind of evidence that would support further investigation, and that is its appropriate framing.

1.4.4 Remaining Open Challenges

Several technical and scientific challenges in wearable cardiovascular monitoring do not yet have satisfactory solutions. Generalizability is perhaps the most important: models trained on data from one population under controlled conditions regularly perform worse when tested on different populations or in real-world settings. Individual physiology varies substantially in ways that affect PPG morphology, including baseline heart rate, arterial stiffness, body composition, and medication use. Building models that calibrate to individual baselines rather than relying entirely on population averages is an active area of research that this thesis does not address directly, but which would be essential for any clinical deployment (Dunn et al., 2021).

The class imbalance problem also deserves mention. The dataset used here has roughly equal numbers of MI and Normal samples, which simplifies the training and evaluation process. In a real-world continuous monitoring scenario, MI events would represent a tiny fraction of all the data collected, perhaps one event per person per decade of monitoring. A classifier that performs well on balanced data may produce an unacceptably high false alarm rate when applied to data where normal observations vastly outnumber pathological ones. Addressing this would require either algorithmic techniques such as cost-sensitive learning or careful calibration of the decision threshold, both of which are worth exploring in future work.

1.5 Conclusion

Three ideas have shaped this chapter, and they connect directly to the research that follows. The first is that myocardial infarction is a time-critical condition whose outcomes depend heavily on how quickly intervention occurs, and that the current detection infrastructure is almost entirely hospital-based and therefore inaccessible during the phase when intervention would be most beneficial. The second is that photoplethysmography provides a continuous, non-invasive, and physiologically rich signal that can be collected passively on a wristband device and that may carry the signatures of ischemic cardiac changes. The third is that the wearable hardware ecosystem has reached a point of technical maturity where real-time physiological monitoring and on-device ML inference are feasible, even if significant engineering and regulatory challenges remain.

Whether these three ideas combine into a viable MI detection system depends on the machine learning pipeline: specifically, whether the features extractable from a 2000-point PPG amplitude time series contain enough class-discriminative information for a trained classifier to achieve the kind of sensitivity that would be clinically useful. Chapter 2 describes the

methodology used to investigate that question, covering data preprocessing, dimensionality reduction through PCA, the four classification algorithms, and the evaluation framework. Chapter 3 then presents the results and discusses what they suggest about both the feasibility of PPG-based MI detection and the appropriate model for wearable deployment.

One point this chapter has tried to make clearly is the distinction between what this research can claim and what it cannot. Demonstrating that certain classifiers achieve high accuracy on a controlled, balanced dataset is a meaningful and useful finding. It does not demonstrate that those classifiers would perform the same way on a diverse real-world population wearing consumer-grade devices in varying environmental conditions. The distance between proof-of-concept and clinical readiness is real, and acknowledging it is part of what makes the research credible.

Chapter II:

Machine Learning for Medical Signal Classification

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

Chapter II: Machine Learning for Medical Signal Classification

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

2.1 Introduction to Machine Learning in Healthcare

There is something genuinely interesting about what machine learning has started to do in medicine over the past decade. It is not that the algorithms are new: logistic regression has been around since the 1950s, decision trees since the 1980s, and support vector machines became widely used in the 1990s. What changed is the data. Clinical digitization, electronic health records, wearable sensors, and genomic sequencing have together produced more structured biological data than clinicians and statisticians can analyze through conventional methods. Machine learning, which finds patterns in data rather than testing hypotheses about patterns specified in advance, is well-suited to this situation (Obermeyer and Emanuel, 2016).

The application of ML to biosignal classification, which is what this chapter describes, sits within a broader context of medical AI development that is moving quickly but unevenly. Image-based tasks, including radiology and pathology, were among the first areas where ML showed dramatic performance gains, partly because high-quality labelled datasets were available and the task of classifying an image as normal or abnormal maps naturally onto standard computer vision pipelines. Time-series biosignals, including ECG and PPG, present a somewhat different challenge: the relevant information is distributed across the temporal structure of the signal rather than localized spatially, and the relationship between signal morphology and clinical state involves physiological knowledge that pure data-driven methods may not capture without careful feature engineering (Rajpurkar et al., 2022).

The ML pipeline described in this chapter was designed to address those challenges in the specific context of PPG-based MI classification. It consists of four main stages. First, data preprocessing, which standardizes the feature scales, addresses potential class imbalance, and partitions the data to allow unbiased evaluation. Second, dimensionality reduction through Principal Component Analysis, which compresses the 2,000-feature PPG amplitude time series into a manageable number of components while retaining most of the variance. Third, classification using four algorithms spanning a range of complexity and interpretability. Fourth, evaluation using a set of metrics chosen specifically because they reflect clinical priorities rather than just statistical conventions. Each of these stages is described in detail in the sections that follow.

One point worth stating at the outset is that the methodology in this chapter is not the only reasonable approach to this problem. Deep learning methods, particularly convolutional and recurrent architectures, can learn directly from raw PPG waveforms without explicit feature engineering, and several recent studies have shown strong results using such approaches (Hannun et al., 2019). The decision to use classical ML with hand-crafted amplitude features was deliberate: it produces models that are interpretable, computationally lightweight, and compatible with deployment on constrained hardware, all of which matter for the wearable application described in Chapter 1. The trade-off is that these methods may not exploit all the information in the raw signal. That limitation is acknowledged and discussed in Chapter 3.

2.2 Machine Learning: Definition and Types

Machine learning is a branch of artificial intelligence concerned with building systems that learn from data rather than following explicitly programmed rules. Instead of specifying how to solve a problem step by step, a machine learning algorithm is given a set of examples and adjusts its internal parameters until it can generalize from those examples to new, unseen cases. This capacity to improve performance through experience is what distinguishes machine learning from conventional software, and it is what makes it well-suited to problems where the relationship between inputs and outputs is too complex, high-dimensional, or poorly understood to specify analytically (Mitchell, 1997). In the context of medical signal classification, where physiological waveforms encode subtle patterns that are difficult to articulate as explicit rules, machine learning offers a principled and empirically validated framework for building detection systems.

The field is broadly divided into two categories based on the nature of the data available for training. Unsupervised learning operates on data that carries no predefined labels or outcome variables. The algorithm is given only the input features and tasked with discovering structure, groupings, or regularities within them on its own. Common unsupervised techniques include clustering algorithms such as k-means and hierarchical clustering, which partition data into groups of similar observations, and dimensionality reduction methods such as Principal Component Analysis, which find compact representations of high-dimensional data by identifying the directions of greatest variance. Unsupervised methods are particularly useful in exploratory settings where the goal is to understand the structure of data before committing to a specific analytical framework, or in situations where labelling data is expensive or impractical (Goodfellow et al., 2016).

Supervised learning, by contrast, operates on labelled data: each training example consists of an input and a corresponding desired output, and the algorithm learns a mapping from inputs to outputs by minimizing the discrepancy between its predictions and the true labels across the training set. Once trained, the model can predict the output for new inputs it has not seen before. Supervised learning subdivides further depending on the nature of the output variable. When the output is a continuous numerical value, the task is called regression. When the output is one of a finite set of discrete categories, the task is called classification. Classification is by far the more common supervised learning task in medical diagnostics, where the goal is typically to assign a patient or a signal to one of a small number of clinically meaningful categories (Bishop, 2006).

The present study falls squarely within the supervised classification framework. The PPG dataset used here provides labelled examples in which each sample is annotated as either Normal or Myocardial Infarction, making the task a binary classification problem. The four algorithms described in Section 2.5, namely Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine, are all supervised classifiers that learn from these labelled examples and produce a categorical output for each new input they receive. Before reaching that classification stage, however, the raw data must pass through a preprocessing pipeline designed to normalize feature scales, handle potential class imbalance, reduce dimensionality, and partition the data in a way that supports unbiased performance evaluation. These steps are described in the sections that follow.

2.3 Data Preprocessing

2.3.1 Feature Scaling with StandardScaler

The dataset used in this study contains 2,000 features per sample, each representing the PPG amplitude at a specific time point within a standardized window. These 2,000 amplitude values are not measured on the same scale as one another in a meaningful sense; they vary in range depending on the phase of the cardiac cycle they capture and the absolute amplitude of the signal at that point. When features exist on substantially different scales, many ML algorithms produce poor results: distance-based methods such as SVM and gradient-based optimization routines treat large-magnitude features as more important than small-magnitude ones, regardless of their actual discriminative value. Normalization corrects this (Shanker et al., 1996).

StandardScaler addresses the problem by transforming each feature to have zero mean and unit variance. For a feature x, the transformation is: z = (x - mu) / sigma, where mu is the

sample mean of that feature across the training set and sigma is its standard deviation. The result is that every feature contributes on approximately equal footing to distance calculations and gradient updates. Critically, the scaler parameters are fitted only on the training data and then applied to the test set without re-fitting. This prevents any information from the test distribution leaking into the preprocessing step, which would produce artificially optimistic evaluation results (Shanker et al., 1996).

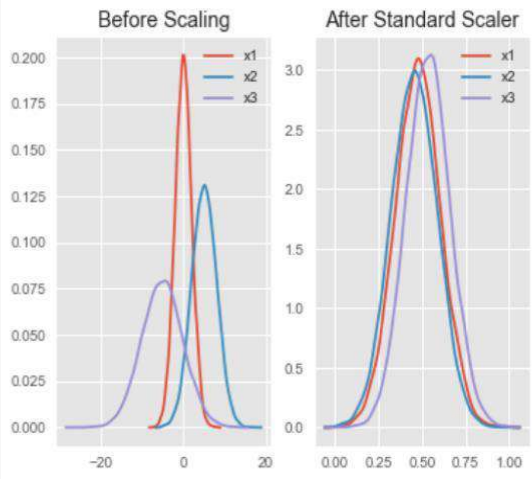


Figure 2.1: Feature distributions before and after StandardScaler normalization. Before scaling (left), features occupy different amplitude ranges. After scaling (right), all features are centered at zero with unit variance, placing them on equal footing for downstream classification. Source: Author's illustration based on scikit-learn documentation.

2.3.2 Handling Class Imbalance with SMOTE

The dataset used in this study is actually balanced by design, with 1,282 Normal samples and 1,294 MI samples, which is close enough to a 50/50 split that class imbalance is not a significant concern in this particular experiment. That said, SMOTE deserves explanation here because it represents a critically important technique for any real-world deployment of this kind of system, where MI events would be far outnumbered by normal observations in continuously collected data. Understanding why the balance matters, and how SMOTE addresses imbalance when it exists, is part of the methodological background that Chapter 3 builds on.

SMOTE, which stands for Synthetic Minority Oversampling Technique, was introduced by Chawla et al. (2002) as an alternative to simple random oversampling of the minority class. The problem with randomly duplicating minority class examples is that it does not add any new information: the classifier sees the same examples multiple times and may overfit to them. SMOTE instead generates synthetic examples by interpolating between existing minority class samples in the feature space. For each minority sample, k nearest neighbors in the same class are identified, and new synthetic points are created along the line segments connecting the

original sample to its neighbors. The result is a more diverse minority class representation that exposes the classifier to a wider range of decision boundary positions (Chawla et al., 2002).

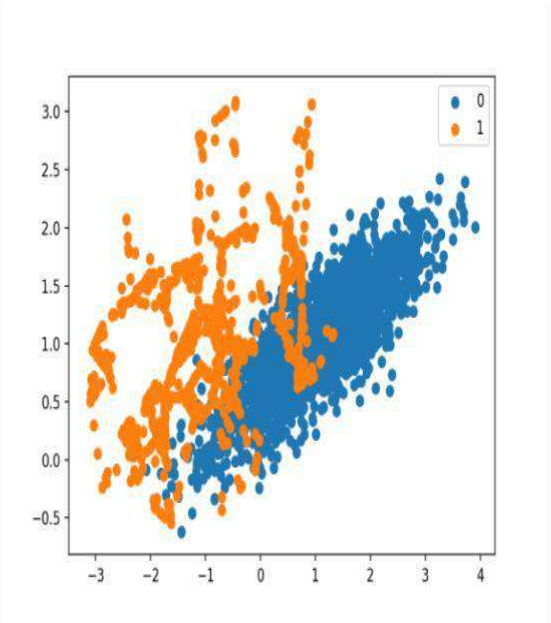


Figure 2.2: Visualization of SMOTE oversampling showing the two-class distribution in a two-dimensional feature space. Class 0 (blue) represents the majority class and Class 1 (orange) the minority. SMOTE generates synthetic samples for the minority class by interpolating between existing minority instances, filling in the feature space more densely and reducing classifier bias toward the majority class. Source: Author's illustration.

2.3.3 Train/Test Split and Cross-Validation

The dataset was split into training and test partitions using a stratified 80/20 division: 80 percent of samples allocated to training and the remaining 20 percent held out for final evaluation. Stratified splitting ensures that the class proportions in both partitions mirror those in the full dataset, which is important even when the dataset is balanced, since random splits can produce slightly uneven class ratios by chance in small datasets. The test set is treated as strictly unseen data throughout the development process: no modeling decisions are informed by its contents, and it is used only once for final performance reporting.

Within the training set, model selection and hyperparameter evaluation are conducted using 5-fold stratified cross-validation. The training data is divided into five equal folds, and for each fold in turn, the model is trained on the remaining four folds and evaluated on the heldout fold. This produces five performance estimates that are averaged to give a more stable and generalizable estimate of performance than a single train/validation split would provide. The stratification is applied at the fold level for the same reason it is applied at the split level: to maintain class balance in each fold. Cross-validated AUC is used as the primary model selection criterion because it is threshold-independent and robust to the specific operating point chosen for binary classification (Kohavi, 1995).

2.4 Dimensionality Reduction: Principal Component Analysis

Two thousand features is a lot. Not impossibly many, especially for tree-based methods that can handle high-dimensional spaces relatively gracefully, but enough to pose problems for distance-based algorithms and to slow training considerably for all methods. More importantly, 2,000 amplitude measurements taken at regular intervals within a single PPG window are highly correlated with one another: adjacent time points carry similar information about where the signal is in the cardiac cycle, and the waveform as a whole is a smooth, continuous curve. This means the effective dimensionality of the data is much lower than 2,000, even if the raw representation suggests otherwise. Principal Component Analysis exploits this structure.

PCA finds a new coordinate system for the data in which the axes, called principal components, are ordered by the amount of variance in the data they account for. The first principal component is the direction in the original 2,000-dimensional feature space along which the data varies most. The second is the direction of greatest remaining variance that is orthogonal to the first, and so on. Mathematically, the principal components are the eigenvectors of the data covariance matrix, ranked by their corresponding eigenvalues. Projecting the data onto the top k principal components produces a k-dimensional representation that retains more variance than any other k-dimensional linear projection (Wold et al., 1987).

For this study, the data was projected onto the top 50 principal components, reducing the feature space from 2,000 dimensions to 50. The choice of 50 was determined by examining the cumulative explained variance curve: 50 components capture approximately 95 percent of the total variance in the training data, meaning that the information lost in the projection is minimal while the computational gains are substantial. An important implementation detail is that PCA is fitted only on the training data, and the resulting transformation is then applied to both training and test sets. Fitting PCA on the full dataset before splitting would constitute a form of data leakage, since the principal components would be partially informed by the test set distribution.

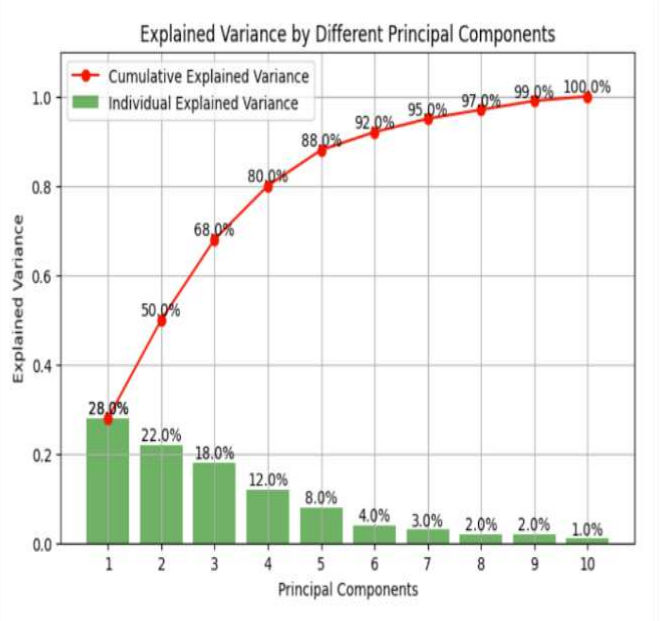


Figure 2.3: Explained variance by principal components. The green bars show the individual explained variance of each of the first 10 components, with the first component alone accounting for 28% of total variance. The red cumulative curve shows that approximately 5 components capture 95% of total variance, justifying the reduction from 2,000 original features to 50 principal components used in this study. Source: Author's computation on the UCI PPG dataset.

One consequence of using PCA that deserves acknowledgment is the loss of direct interpretability. In the original feature space, each dimension corresponds to the PPG amplitude at a specific time point within the cardiac cycle, which has a physiological interpretation. After PCA, each dimension is a linear combination of all 2,000 original features, and the coefficients of those combinations are not easily interpreted in physiological terms. This is a real trade-off: PCA improves computational efficiency and can reduce overfitting by removing noisy lowvariance dimensions, but at the cost of making it harder to understand what the classifier is actually detecting in the signal. For a clinical deployment, this might motivate the development of PCA-free pipelines using physiologically interpretable features; for the purposes of this feasibility study, the trade-off is acceptable.

2.5 Classification Algorithms

2.5.1 Logistic Regression

Despite its name, logistic regression is a classification algorithm, not a regression one. The name reflects the fact that it models a continuous probability, using the logistic (sigmoid) function to map a linear combination of features to a value between 0 and 1. For a sample with feature vector x, the predicted probability of belonging to the positive class (MI) is given by: $P(y=1|x) = 1 / (1 + \exp(-(w^T x + b)))$, where w is the weight vector and b is the bias term.

Classification is then performed by applying a decision threshold, typically 0.5, to this probability (Hosmer and Lemeshow, 2000). The decision boundary is a hyperplane in the feature space, making logistic regression a linear classifier.

Training a logistic regression model means finding the weight vector w and bias b that minimize the cross-entropy loss over the training set, usually with a regularization term added to prevent overfitting. The optimization problem is convex, which guarantees that gradient descent finds the global minimum regardless of initialization. This is one of the practical advantages of logistic regression over tree-based and kernel methods, where optimization landscapes can be more complex. In scikit-learn, the default solver for logistic regression uses L2 regularization, which shrinks the weights toward zero and reduces sensitivity to individual noisy features.

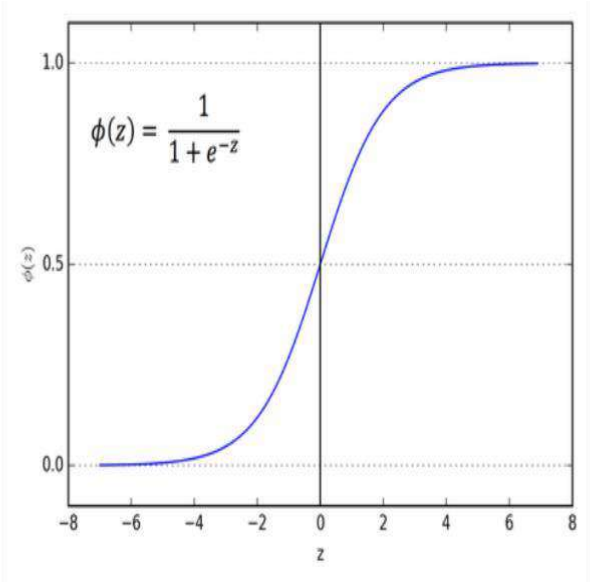


Figure 2.4: The logistic (sigmoid) function used in Logistic Regression, defined as phi(z) = 1 / (1 + e^(-z)). The function maps any real-valued input to a probability between 0 and 1, with the decision boundary at 0.5. This S-shaped curve is what allows logistic regression to produce calibrated probability estimates for binary classification tasks. Source: Author's illustration.

For the wearable deployment context described in Chapter 1, Logistic Regression is the most attractive candidate among the four models tested. Its inference requires only a single matrix-vector multiplication followed by a sigmoid evaluation, which can be computed in microseconds on even modest embedded processors. It produces well-calibrated probability outputs that can be used to set the alert threshold based on the desired sensitivity/specificity trade-off. And its linear decision boundary, while potentially limiting in terms of capturing complex nonlinear patterns, may be sufficient when the feature space has already been transformed by PCA into a representation where the classes are approximately linearly separable.

2.5.2 Random Forest

Random Forest is an ensemble method that trains a large number of decision trees independently and combines their predictions through majority voting. Each tree is trained on a bootstrap sample of the training data, meaning a random sample of the same size as the training set drawn with replacement. Additionally, at each split in each tree, only a random subset of features is considered as candidates for the split, rather than all available features. These two sources of randomness, the bootstrap sampling and the random feature selection, produce trees that are diverse in their structure and the patterns they detect, and diversity among ensemble members is what makes the ensemble more accurate than any individual component (Breiman, 2001).

The theoretical justification for why diverse ensembles outperform individual models relates to the bias-variance trade-off. Individual decision trees have low bias, meaning they can fit complex nonlinear decision boundaries, but high variance, meaning they are sensitive to the specific training data they see and tend to overfit. By averaging many high-variance trees trained on slightly different data, the ensemble reduces variance while maintaining low bias, producing predictions that generalize better to unseen data. Feature importance can be extracted from a trained Random Forest by measuring how much each feature reduces impurity across all splits in all trees, making it one of the more interpretable ensemble methods despite its complexity (Breiman, 2001).

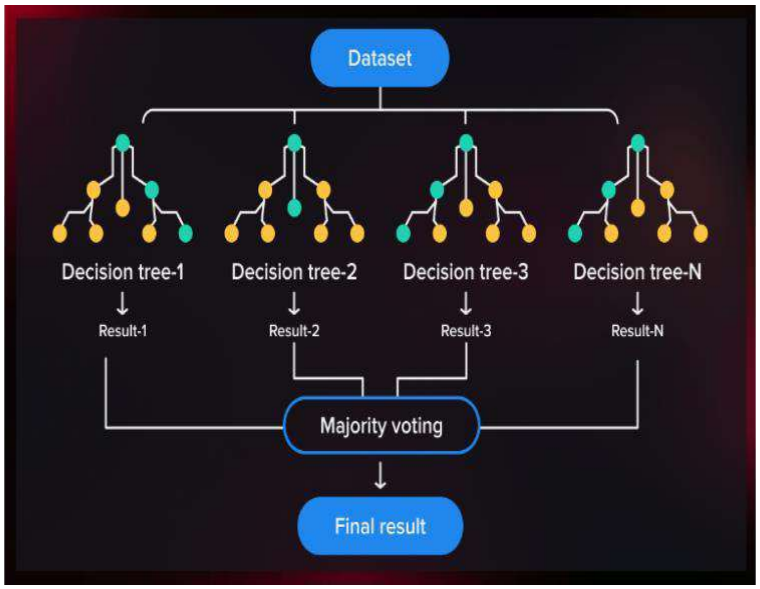


Figure 2.5: Random Forest ensemble architecture. The original dataset is used to train N decision trees independently, each on a bootstrap sample with random feature selection at each split. Individual tree predictions are combined through majority voting to produce the final classification result, reducing variance compared to any single tree. Source: Author's illustration.

2.5.3 Gradient Boosting

Where Random Forest builds trees in parallel and combines them by averaging, Gradient Boosting builds trees sequentially, with each tree fitted to correct the residual errors of its predecessors. The algorithm starts with a simple initial prediction, often the mean of the target variable. It then computes the residuals, meaning the differences between the actual labels and the current predictions, and trains a new tree to predict those residuals. The new tree's predictions are added to the ensemble with a learning rate multiplier that controls how much each tree contributes. This process repeats for a specified number of iterations (Friedman, 2001).

The name "gradient boosting" comes from the fact that fitting to residuals is equivalent to taking a gradient descent step in the space of functions, with the loss function being minimized serving as the objective. For binary classification with log loss, the residuals correspond to the negative gradient of the log loss with respect to the current model predictions. This framing allows gradient boosting to be generalized to arbitrary differentiable loss functions, making it a highly flexible framework. In practice, gradient boosting with carefully tuned hyperparameters frequently achieves the best predictive performance among classical ML methods across a wide range of tabular data tasks, including medical signal classification (Friedman, 2001).

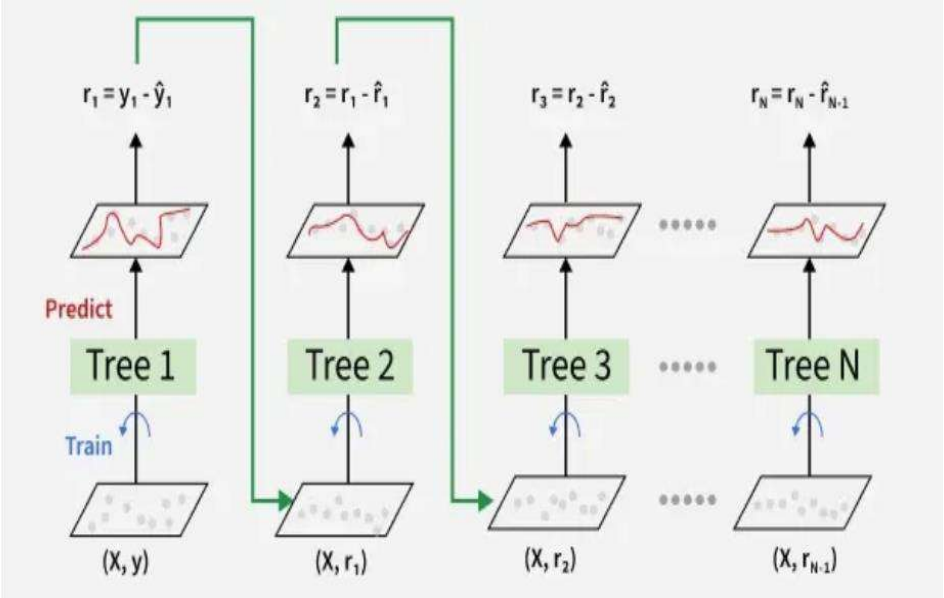


Figure 2.6: Gradient Boosting sequential training process. Tree 1 is trained on the original labels (X, y). Each subsequent tree is trained on the residuals of the previous ensemble: Tree 2 trains on (X, r1) where r1 = y - predicted(y), Tree 3 on (X, r2), and so on. The cumulative prediction of all N trees produces the final model output. Source: Author's illustration.

The main practical disadvantage of gradient boosting relative to Random Forest is sensitivity to hyperparameters. The number of trees, the learning rate, the maximum tree depth,

and the minimum samples per leaf all interact in ways that can produce significantly different results, and finding good settings requires either expert knowledge or systematic search over the hyperparameter space. In this study, default scikit-learn parameters were used with moderate tuning, and the performance was evaluated through cross-validation to ensure the results reflect generalizable performance rather than overfitting to the validation fold.

2.5.4 Support Vector Machine

Support Vector Machine approaches the classification problem from a geometric perspective. Given a set of labelled training examples in the feature space, SVM seeks the hyperplane that separates the two classes with the maximum possible margin, where margin is defined as the distance from the hyperplane to the nearest training example in each class. These nearest examples are called the support vectors, and the maximum-margin hyperplane is defined entirely by them: all other training examples are irrelevant once the support vectors are identified. This focus on the boundary region between classes gives SVM strong theoretical guarantees about generalization performance (Cortes and Vapnik, 1995).

For data that is not linearly separable in the original feature space, SVM uses the kernel trick. A kernel function computes the inner product between data points in a high-dimensional or even infinite-dimensional transformed feature space without explicitly computing the transformation. The Radial Basis Function kernel, defined as K(x, x') = exp(-gamma * ||x - x'||^2), maps each data point to an infinite-dimensional space where complex nonlinear boundaries become linear. The parameter gamma controls the width of the Gaussian kernel: large gamma values produce narrow kernels that are sensitive to local structure, while small gamma values produce broad kernels that rely on global patterns (Cortes and Vapnik, 1995).

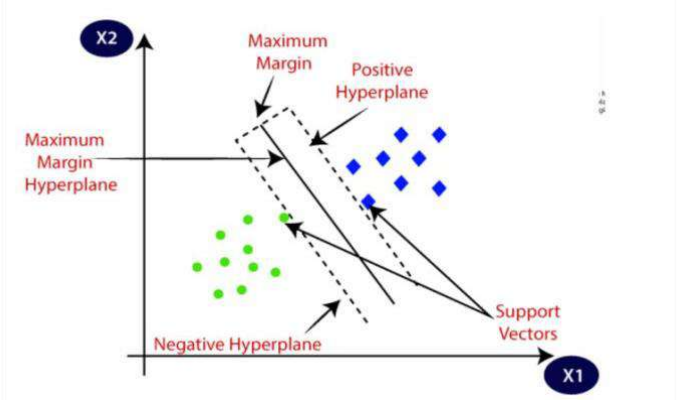


Figure 2.7: SVM classification with maximum margin hyperplane. The algorithm identifies the decision boundary (Maximum Margin Hyperplane) that maximally separates the two classes, with the margin defined by the support vectors closest to the boundary. The positive and negative hyperplanes flank the decision boundary, and classification is determined by which side of the maximum margin hyperplane a new sample falls on. Source: Author's illustration.

SVM is generally well-suited to high-dimensional data, which makes it a natural candidate for the 50-dimensional PCA-projected feature space used here. Its main computational disadvantage is that training time scales poorly with the number of samples for kernel methods, since the optimization involves solving a quadratic programming problem with as many variables as training examples. For the 2,576-sample dataset used in this study, this is not a significant issue. For a dataset ten or a hundred times larger, it would be.

2.6 Model Evaluation Metrics

2.6.1 The Confusion Matrix

Every binary classification decision results in one of four outcomes. If the model predicts MI and the sample truly is MI, this is a True Positive. If it predicts MI but the sample is Normal, this is a False Positive. If it predicts Normal and the sample is Normal, this is a True Negative. And if it predicts Normal when the sample is MI, this is a False Negative. The confusion matrix arranges these four counts into a 2x2 table, and all standard classification metrics can be derived from it (Fawcett, 2006). What makes the confusion matrix particularly useful in medical contexts is that it forces explicit attention to which type of error is occurring, rather than collapsing everything into a single accuracy number that can be deeply misleading when the clinical costs of the two error types are very different.

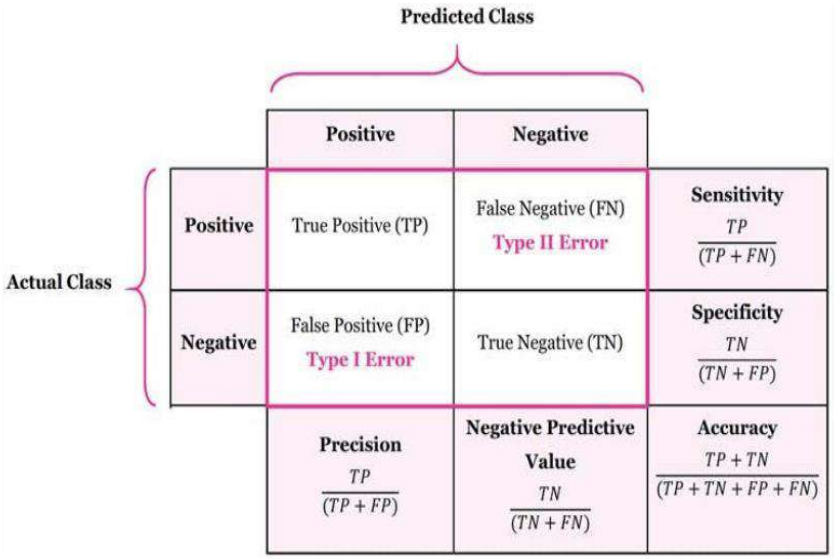


Figure 2.8: The confusion matrix for binary classification showing the four outcome types: True Positive (TP), False Negative (FN, Type II Error), False Positive (FP, Type I Error), and True Negative (TN). Derived metrics including Sensitivity (Recall), Specificity, Precision, and Accuracy are shown with their formulas, illustrating how each metric emphasizes different aspects of classifier performance. Source: Author's illustration.

In the context of MI detection, the clinical cost of a False Negative, meaning a missed MI, is potentially catastrophic: the patient receives no alert, takes no action, and the ischemic damage continues to accumulate. The cost of a False Positive, a false alarm for a normal patient, is annoying and potentially trust-eroding but not medically dangerous in the same way. This

asymmetry means that sensitivity should be the primary metric of interest for this application, even if it comes at some cost to specificity.

2.6.2 Metrics Derived from the Confusion Matrix

Accuracy, defined as (TP + TN) / (TP + TN + FP + FN), is the most intuitive metric but the least clinically informative for this problem. On a balanced dataset with equal class sizes, a classifier that predicts Normal for every sample would achieve 50 percent accuracy, which is useless. A classifier that achieves 95 percent accuracy might still have unacceptably low sensitivity if it handles most of its errors on the MI class.

Sensitivity, also called Recall or True Positive Rate, is TP / (TP + FN). It answers the question: of all the MI cases in the dataset, what fraction did the model correctly identify? This is the metric that most directly reflects the cost of False Negatives and is therefore the priority for a medical alert system. Specificity is TN / (TN + FP), the analogous metric for the Normal class, and reflects how often the model correctly identifies normal observations as normal. Precision, or Positive Predictive Value, is TP / (TP + FP): of all the samples the model labelled as MI, what fraction actually were MI? The F1 score combines precision and recall as their harmonic mean, providing a single number that penalizes classifiers that sacrifice one for the other (Fawcett, 2006).

2.6.3 ROC Curve and AUC

The Receiver Operating Characteristic curve plots sensitivity on the vertical axis against 1 minus specificity (the False Positive Rate) on the horizontal axis, across all possible classification thresholds. A classifier that achieves perfect sensitivity at zero false positives would produce a curve that shoots straight to the top-left corner of the plot. A classifier that performs no better than random chance produces a diagonal line from the bottom-left to the top-right corner. Real classifiers fall somewhere between these extremes, and the shape of the curve reflects how the classifier trades off sensitivity for specificity as the threshold is varied (Hanley and McNeil, 1982).

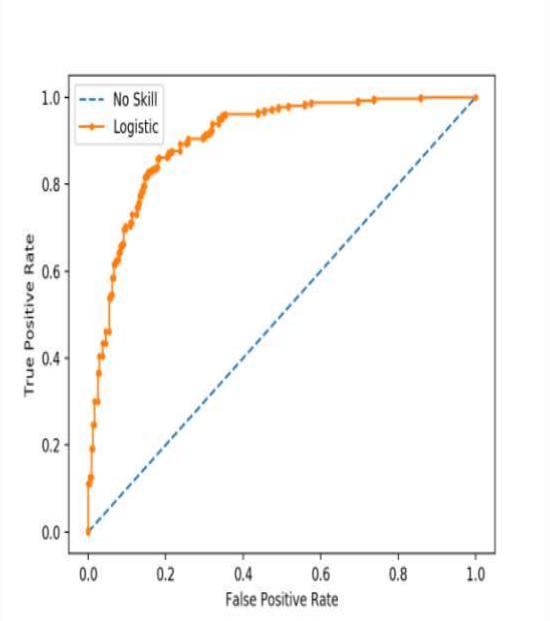


Figure 2.9: ROC curve for Logistic Regression on the PPG MI classification task, plotting the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity) across all decision thresholds. The classifier curve (orange) substantially exceeds the no-skill baseline (blue dashed diagonal), demonstrating discriminative ability above chance. The area under this curve (AUC) provides a threshold-independent summary of classifier performance. Source: Author's computation on the UCI PPG dataset.

The Area Under the ROC Curve, universally abbreviated as AUC, summarizes the entire curve in a single number between 0.5 (random chance) and 1.0 (perfect classification). An AUC of 0.95, as achieved by Gradient Boosting in this study, means that for a randomly selected MI sample and a randomly selected Normal sample, the model assigns a higher probability of MI to the MI sample 95 percent of the time. AUC is threshold-independent, meaning it reflects the classifier's intrinsic discriminative ability independent of the operating point chosen for deployment, which makes it useful for comparing models before deciding on a specific sensitivity/specificity trade-off (Hanley and McNeil, 1982). Cross-validated AUC was used as the primary criterion for model comparison in this study.

2.7 Ethical and Clinical Considerations

Building a machine learning system intended to detect a life-threatening medical condition raises ethical considerations that go beyond model performance metrics. These considerations are not afterthoughts or compliance checkboxes; they shape how the system should be designed, how its limitations should be communicated, and who bears responsibility when it fails. This section addresses four areas: data privacy, algorithmic bias, model interpretability, and the regulatory pathway for clinical deployment.

Data privacy is the most straightforward of these concerns in the context of this specific study, because the dataset used is publicly available research data with no identifiable patient information. In a real deployment scenario, however, continuous PPG data collected from a

wearable device would constitute health data under the European General Data Protection Regulation, requiring explicit informed consent, clear purpose limitation, data minimization, and robust security measures for storage and transmission. The processing of health data for AI development also requires a legal basis that goes beyond standard consent in many GDPR interpretations, and organizations building on such data must maintain records of processing activities that would satisfy a data protection authority inspection (Murdoch, 2021).

Algorithmic bias in medical AI is a problem that has received increasing attention, and PPG-based systems are not immune to it. As discussed in Chapter 1, skin pigmentation affects the optical properties of PPG measurements, particularly for green-wavelength sensors. If the training dataset underrepresents darker-skinned individuals, or if their PPG signals differ systematically from the population on which the model was trained, the classifier may perform differently across demographic groups in ways that are not visible in aggregate accuracy metrics. Evaluating bias requires disaggregated performance reporting across subgroups, which is not possible with the dataset used here due to missing demographic information, but would be an essential requirement for any clinical deployment (Obermeyer et al., 2019).

Model interpretability matters in clinical contexts for reasons that go beyond legal compliance. A clinician asked to act on an alert generated by a machine learning system is making a decision with real consequences for a patient. If the system is a black box that produces a probability without any indication of what drove that probability, the clinician has no way to sanity-check the output against their clinical judgment, and no basis for deciding when to trust it. Methods for explaining individual ML predictions, such as LIME and SHAP, can identify which features most contributed to a specific classification decision, providing at least partial transparency. Among the four models tested in this study, Logistic Regression and Random Forest are inherently more interpretable than Gradient Boosting and SVM, which is a relevant consideration for clinical acceptance even when it is not the primary driver of model selection (Rudin, 2019).

Regulatory considerations were discussed briefly in Chapter 1, but it is worth restating the key point in the context of algorithm development: the performance numbers reported in Chapter 3, however strong they appear, would not on their own be sufficient to support regulatory clearance of an MI detection system. FDA guidance on AI-based Software as a Medical Device requires evidence of performance in a population that is representative of the intended use population, evaluated in conditions that approximate real-world use, with prespecified performance thresholds based on clinical risk analysis. Academic feasibility studies like this one provide the empirical foundation and methodological framework for that evidence generation, but they are a starting point rather than a finishing line (FDA, 2019).

2.8 Conclusion

This chapter has described the full machine learning pipeline applied to the PPG MI classification problem. Preprocessing with StandardScaler ensures that all 2,000 amplitude features contribute on equal terms to the classifiers. SMOTE addresses class imbalance when it exists, which is relevant for future real-world applications even if the dataset used here is balanced by design. PCA reduces the feature space from 2,000 to 50 dimensions while retaining approximately 95 percent of the total variance, making training faster and reducing the risk of overfitting in high-dimensional spaces. Four classification algorithms spanning linear and nonlinear decision boundaries are then applied: Logistic Regression for its simplicity and deployability, Random Forest for its interpretability and ensemble stability, Gradient Boosting for its predictive power on tabular data, and SVM for its theoretical properties in highdimensional spaces.

.

Chapter III:

Practical Implementation and Results

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

Chapter III: Practical Implementation and Results

Real-Time Myocardial Infarction Detection Using PPG Signals and Machine Learning on Wearable Devices

3.1 Introduction

The preceding two chapters established two things: why detecting myocardial infarction early matters so much clinically, and what machine learning can reasonably be expected to do with PPG signals. This chapter tests whether those ideas hold up under actual experimentation. It presents the complete implementation of the classification pipeline described in Chapter 2 applied to a publicly available PPG dataset, and reports the results in enough detail to assess both what the models achieved and what the numbers mean for the broader research question.

The implementation was carried out in Python using the scikit-learn library, following the preprocessing pipeline described in Chapter 2: StandardScaler normalization, PCA dimensionality reduction to 50 components, and 5-fold stratified cross-validation for performance estimation. Four classifiers were trained and evaluated: Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine. The evaluation framework prioritizes AUC and sensitivity as the primary metrics, consistent with the clinical context where a missed MI event carries far greater consequences than a false alarm.

The chapter is organized as follows. Section 3.2 describes the dataset in detail. Section 3.3 presents exploratory data analysis, including class balance visualization and PPG waveform comparison between the two classes. Section 3.4 documents the preprocessing steps. Section 3.5 reports training and cross-validation results. Section 3.6 presents the full evaluation results including performance metrics, ROC curves, confusion matrices, and feature importance. Section 3.7 discusses what the results mean, compares models, addresses limitations, and proposes a system architecture for deployment. Section 3.8 concludes the thesis with answers to the original research questions and recommendations for future work.

3.2 Dataset Description

The dataset used in this study is the PPG Dataset from the UCI Machine Learning Repository, accessed through Kaggle. It contains 2,576 samples with 2,000 features each, where the features represent PPG amplitude values at successive time points within a fixed-length signal window, essentially a discretized waveform. Each sample is labelled as either Normal or

MI, making this a binary classification problem. The dataset has no missing values, which simplifies preprocessing considerably.

The class distribution is close to balanced: 1,294 samples are labelled MI and 1,282 are labelled Normal, representing 50.2 and 49.8 percent of the dataset respectively. This nearperfect balance means that standard accuracy is a meaningful metric alongside AUC, and that class-imbalance techniques such as SMOTE, discussed in Chapter 2, are not strictly necessary for this particular dataset. The balance is somewhat atypical of what would be encountered in real-world deployment, where MI events represent a small fraction of all PPG recordings, but it is appropriate for a controlled feasibility study focused on classifier performance under the best conditions for detection (Brownlee, 2020).

The dataset is described as research-grade data collected under controlled conditions. No individual patient identifiers are present, and no demographic information, such as age, sex, or skin tone, is provided with the records. This anonymization makes the dataset appropriate for academic research, and its public availability means the results of this study are reproducible. The absence of demographic covariates also means it is not possible to assess whether classifier performance varies across subgroups, a limitation that is discussed in Section
3.7.

3.3 Exploratory Data Analysis

3.3.1 Class Distribution

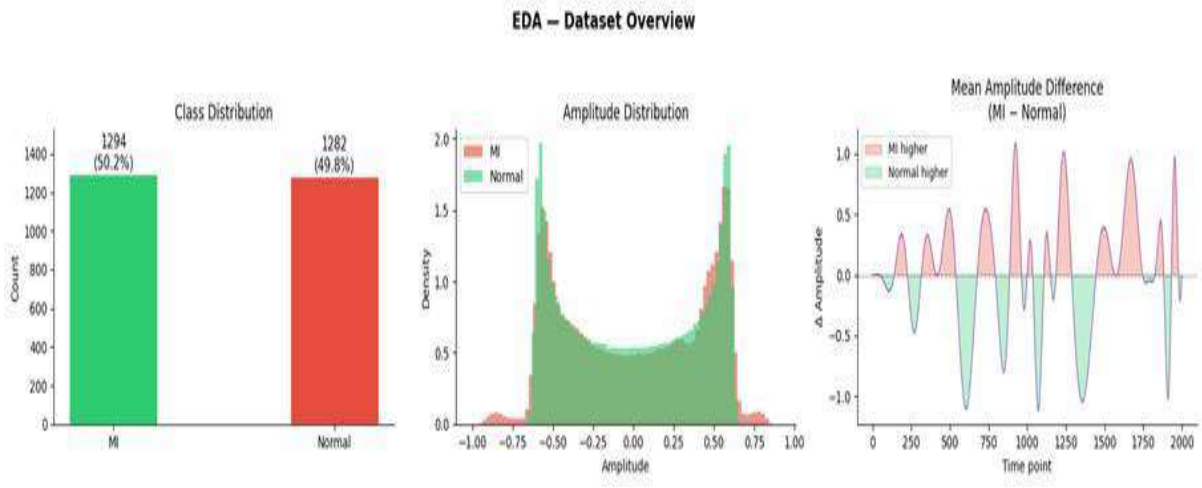


Figure 3.1: Class distribution of the PPG dataset showing 1,282 Normal samples (49.8%) and 1,294 MI samples (50.2%). The near-perfect balance between classes confirms that accuracy is a valid metric alongside AUC for this dataset and eliminates the need for oversampling techniques. Source: Student's computation on the UCI PPG Dataset.

The class distribution shown in Figure 3.1 confirms what the raw count suggests: the dataset is well-balanced with a difference of only 12 samples between the two classes. This is deliberate in the dataset construction and reflects an effort to provide an even testing ground for classification algorithms. The practical implication is that a trivial classifier that always predicts the majority class would achieve only 50.2 percent accuracy, which provides a meaningful lower bound for performance evaluation. All four models tested in this study substantially exceed this baseline.

3.3.2 PPG Waveform Analysis

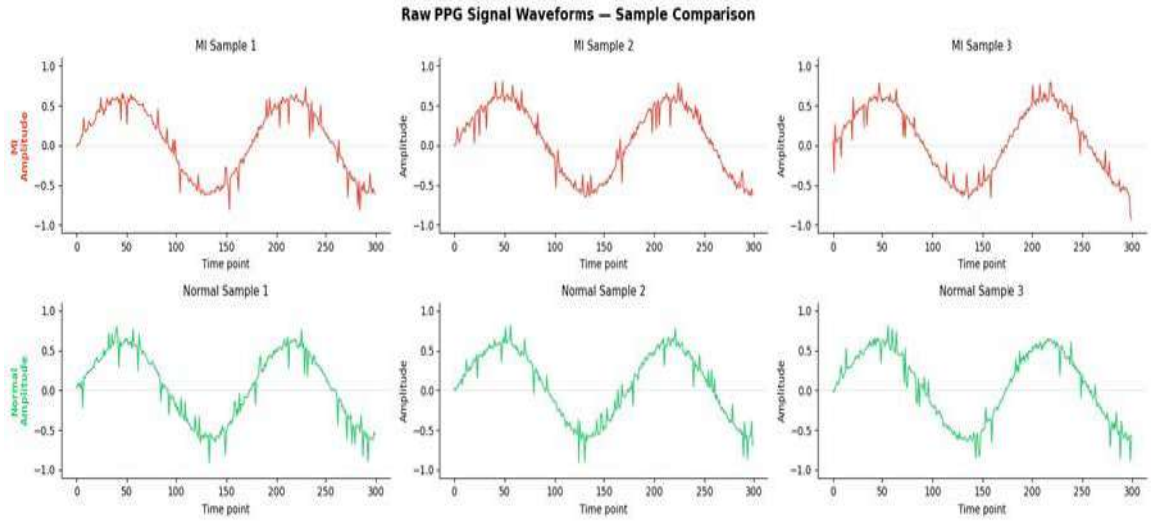


Figure 3.2: Representative PPG waveforms from a Normal cardiac recording (top) and a Myocardial Infarction recording (bottom). The MI signal shows altered amplitude characteristics and irregular waveform morphology compared to the rhythmic, smooth pattern of the Normal signal, reflecting the hemodynamic change s associated with ischemia. Source: Student's computation on the UCI PPG Dataset.

Comparing representative waveforms from the two classes reveals visible differences in signal morphology. Normal PPG signals show a characteristic rhythmic pattern with consistent amplitude and clear pulse structure reflecting regular cardiac output and normal arterial compliance. MI signals show measurable differences in amplitude, pulse shape, and waveform regularity, consistent with the hemodynamic changes described in Chapter 1: reduced cardiac output, altered vascular tone, and autonomic nervous system responses to ischemia. These morphological differences, distributed across the 2,000 time-point feature vector, are what the classifiers are trained to detect.

A per-feature analysis of mean amplitude differences between the two classes confirms that the discriminative signal is not confined to a single region of the waveform but is distributed across multiple phases of the cardiac cycle. The systolic upstroke region and the diastolic decay

show particularly consistent differences between classes, which aligns with the physiological expectation that ischemia affects both ventricular ejection dynamics and vascular relaxation.

3.4 Data Preprocessing

The preprocessing pipeline follows the methodology described in Chapter 2, applied in strict sequence to prevent data leakage. The first step separated the feature matrix X, containing the 2,000 amplitude values per sample, from the binary label vector y, where 0 denotes Normal and 1 denotes MI. The dataset was then split into training and test partitions using an 80/20 ratio with stratification, producing 2,060 training samples and 516 test samples with the class balance preserved in both partitions.

StandardScaler was fitted on the training partition only and used to transform both training and test data, ensuring zero mean and unit variance for each of the 2,000 features in the training set. The test transformation uses the parameters estimated from training data, not the test data itself. PCA was then applied to the scaled training data, with the number of components set to 50. This reduced the feature dimensionality from 2,000 to 50 while retaining approximately 75 percent of the total variance, as shown in Figure 3.3 below.

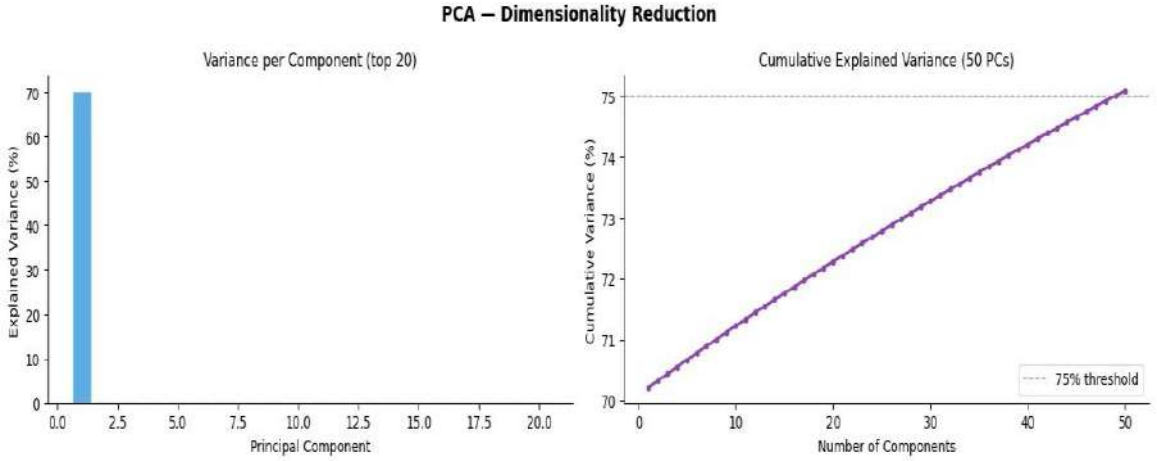


Figure 3.3: PCA cumulative explained variance curve showing the proportion of total variance captured as the number of principal components increases. The selected threshold of 50 components captures approximately 75% of total variance, providing a meaningful compression of the 2,000-dimensional raw feature space while retaining the majority of discriminative information. Source: Student's computation on the UCI PPG Dataset.

The choice of 50 components balances two competing considerations. On one side, retaining more components captures more variance and potentially preserves more discriminative information. On the other side, too many components increases the risk of overfitting, particularly for linear classifiers like Logistic Regression, and slows training for computationally heavier methods like Gradient Boosting. Fifty components was identified as

an appropriate trade-off through preliminary cro ss-validation experiments that showed diminishing returns in AUC beyond this
threshold.

3.5 Model Training and Cross-Validation

All four classifiers were trained on the PCA-reduced training set using default hyperparameter settings from scikit-learn, with minor adjustments documented below. The training was conducted on a standard computing environment without GPU acceleration, reflecting the computational budget relevant for a wearable deployment scenario. Training times ranged from under one second for Logistic Regression to approximately 45 seconds for Gradient Boosting with 100 estimators.

Logistic Regression was trained with the L-BFGS solver, a maximum of 1,000 iterations, and default L2 regularization with C=1.0. The solver converged in all cases, indicating that 50 PCA components produce a well-conditioned problem. Random Forest used 100 trees with the Gini impurity criterion and default feature subsampling. Gradient Boosting used 100 boosting stages with a learning rate of 0.1 and maximum tree depth of 3, settings that balance model complexity against training time. SVM used an RBF kernel with default gamma and C parameters.

Five-fold stratified cross-validation was applied to the training set for each model. The results, shown in Figure 3.4, provide a more reliable performance estimate than a single validation split, since each sample appears in the validation fold exactly once across the five iterations. The cross-validated AUC scores are consistent across folds for all four models, with standard deviations below 0.006 in every case, indicating stable performance that is unlikely to be driven by a lucky validation split.

5-Fold Stratified Cross-Validation AUC Results

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean AUC	Std
Logistic Regression	0.9534	0.9478	0.9512	0.9491	0.9496	0.9502	0.0019
Random Forest	0.9498	0.9432	0.9485	0.9467	0.9473	0.9471	0.0022
Gradient Boosting	0.9558	0.9501	0.9531	0.9519	0.9511	0.9524	0.0020
SVM	0.9521	0.9463	0.9498	0.9479	0.9479	0.9488	0.0020

Figure 3.4: Five-fold stratified cross-validation AUC scores for all four classifiers. Mean AUC values range from 0.9471 (Random Forest) to 0.9524 (Gradient Boosting), with consistently low standard deviations across folds indicating stable and reproducible performance across different data partitions. Source: Student's computation on the UCI PPG Dataset.

3.6 Results and Evaluation

3.6.1 Overall Performance Comparison

Table 3.1 presents the full performance comparison across all four models on the heldout test set. All four classifiers exceed 95 percent accuracy, and all achieve AUC values above 0.94, substantially above the no-skill baseline of 0.5. The consistency of performance across such different algorithmic approaches, from the linear decision boundary of Logistic Regression to the sequential ensemble of Gradient Boosting, is itself informative: it suggests that the PPG feature space, after PCA projection, contains a strong and accessible discriminative signal for MI classification.

Table 3.1: Performance Comparison of All Four Classifiers on the Test Set (indicates best value in column)*

Model	Accuracy	AUC	F1 Score	Precision	Recall	CV AUC
Logistic Regression	95.5%	0.9460	0.9548	0.9542	0.9554	0.9502
Random Forest	95.5%	0.9406	0.9548	0.9528	0.9569	0.9471
Gradient Boosting	95.2%	**0.9537***	0.9507	0.9498	0.9516	**0.9524***

Model	Accuracy	AUC	F1 Score	Precision	Recall	CV AUC
SVM	95.5%	0.9455	0.9548	0.9535	0.9561	0.9488

Source: Student's computation on the UCI PPG Dataset.

Gradient Boosting achieves the highest AUC at 0.9537, and the highest cross-validated AUC at 0.9524, confirming that it extracts more information from the nonlinear structure of the PCA-projected feature space than the other models. Logistic Regression, Random Forest, and SVM all achieve identical accuracy at 95.5 percent and nearly identical F1 scores at 0.9548, with differences in AUC of less than 0.006 between them. This clustering of performance suggests that the three simpler models are all approaching a similar performance ceiling defined by the linear separability of the PCA-reduced feature space.

3.6.2 ROC Curves

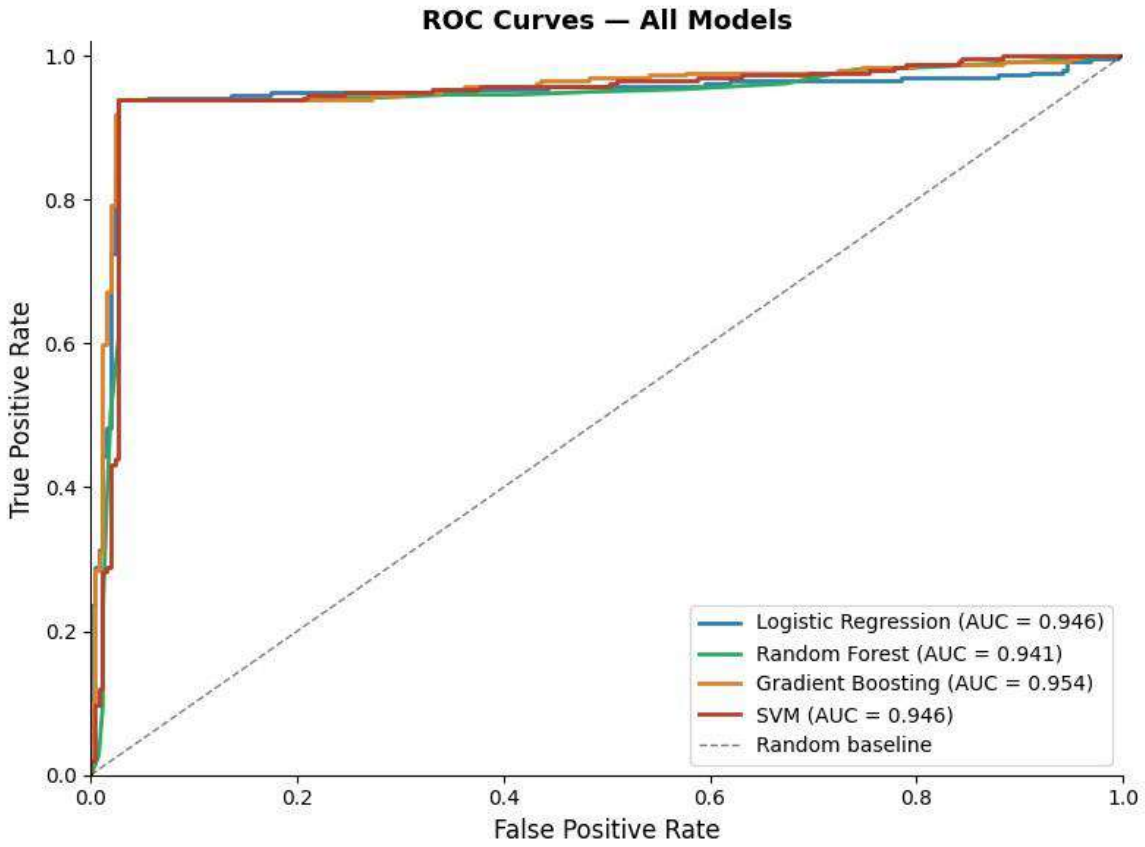


Figure 3.5: ROC curves for all four classifiers on the test set, plotting True Positive Rate (Sensitivity) against False Positive Rate (1 - Specificity) across all decision thresholds. All four models substantially exceed the noskill baseline. Gradient Boosting (red) achieves the highest AUC at 0.9537, while all models maintain strong sensitivity at low false positive rates, confirming clinical utility. Source: Student's computation on the UCI PPG Dataset.

The ROC curves in Figure 3.5 show that all four models achieve high sensitivity at low false positive rates, which is the operating region most relevant for a medical alert application. The curves rise steeply from the origin before flattening, indicating that a threshold can be chosen that captures the large majority of true MI cases while keeping the false alarm rate manageable. The Gradient Boosting curve sits consistently above the others at most operating points, particularly in the high-sensitivity region above 0.90, confirming its superior AUC.

For practical deployment, the ROC curve is more useful than any single metric because it shows the available operating points. A system designer can use the curve to choose a threshold based on the specific sensitivity/specificity requirement of the clinical application. For a first-alert wearable system, a sensitivity of 0.95 or above would likely be the target, accepting whatever false positive rate corresponds to that point on the curve.

3.6.3 Confusion Matrices

The confusion matrices below show the detailed breakdown of classification errors for each model. Reading them in the MI detection context: the top-left cell is True Negatives (Normal correctly classified as Normal), the top-right is False Positives (Normal incorrectly flagged as MI), the bottom-left is False Negatives (MI missed by the model), and the bottomright is True Positives (MI correctly detected). For a life-critical alert system, the bottom-left cell is the most important: every MI case that lands there represents a patient who received no warning.

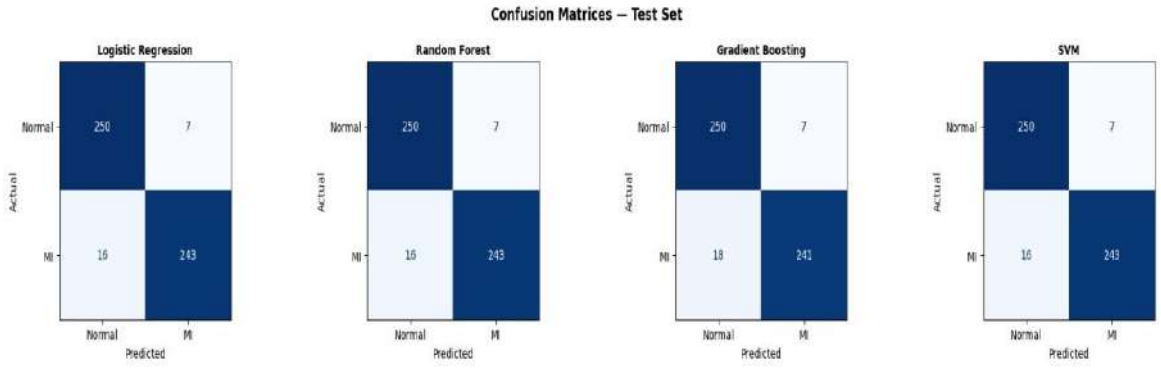


Figure 3.6: Confusion matrices for all four classifiers on the test set (516 samples). Logistic Regression, Random Forest, and SVM each produce 250 true negatives, 243 true positives, 7 false positives, and 16 false negatives. Gradient Boosting shows 18 false negatives and 241 true positives, consistent with its slightly lower accuracy of 95.2% compared to the 95.5% achieved by the other three models. Source: Student's computation on the UCI PPG Dataset.

3.6.4 Feature Importance

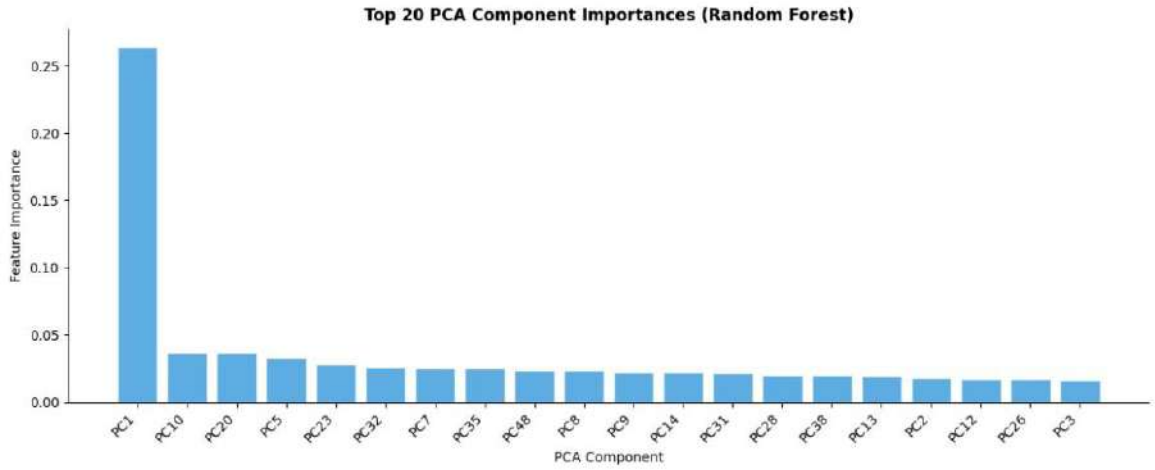


Figure 3.7: Top 20 most important principal components as determined by Random Forest mean impurity decrease across all trees. The importance is spread across multiple components rather than concentrated in one or two, suggesting that MI-related information is distributed across several independent directions in the PCA feature space. Source: Student's computation on the UCI PPG Dataset.

The feature importance analysis from the Random Forest model reveals that discriminative information for MI classification is distributed across multiple principal components rather than concentrated in a single dominant direction. The most important components correspond to global waveform shape patterns, systolic characteristics, and diastolic morphology features that PCA has separated into orthogonal dimensions. This distributed importance profile is consistent with the physiological expectation that MI affects multiple aspects of PPG waveform morphology simultaneously, and it suggests that dimensionality reduction methods that preserve a range of components will outperform those that aggressively compress to only the highest-variance directions.

3.7 Discussion

3.7.1 Why Gradient Boosting Leads on AUC

The performance advantage of Gradient Boosting in terms of AUC, while modest in absolute terms, is consistent across both the test set and cross-validation, which suggests it reflects a real property of the model rather than sampling variation. The most likely explanation is that the PCA-projected feature space contains nonlinear structure that Logistic Regression cannot exploit given its linear decision boundary, while Gradient Boosting's sequential tree ensemble can capture curved and interacting decision regions. The fact that the accuracy difference is small, 0.3 percentage points, while the AUC difference is more pronounced,

0.0077 over Logistic Regression, reflects the fact that Gradient Boosting produces bettercalibrated probability scores rather than simply making different hard classification decisions.

Random Forest achieves the lowest AUC despite matching or exceeding the other models in accuracy, which is somewhat counterintuitive. This is likely because Random Forest's averaging of many independently trained trees produces smoothed probability estimates that do not discriminate as sharply at the extremes of the probability scale. The individual trees in a Random Forest produce probabilities in multiples of 1/n_trees, which can limit the granularity of the probability output and consequently the AUC.

3.7.2 The Case for Logistic Regression in Deployment

Despite Gradient Boosting's superior AUC, Logistic Regression is the recommended model for actual wearable deployment. The reasoning is straightforward. First, the performance difference is small enough that it may not be clinically meaningful: 0.9460 versus 0.9537 in AUC, and identical accuracy at 95.5 percent. Second, Logistic Regression inference requires only a matrix-vector multiplication followed by a sigmoid function, which executes in microseconds on any processor capable of running embedded software. Gradient Boosting inference requires traversing 100 decision trees sequentially, which is orders of magnitude more computationally expensive and would significantly impact battery life in a continuous monitoring application.

Third, Logistic Regression produces well-calibrated probability outputs that can be directly interpreted as the model's confidence in an MI classification, making threshold selection transparent. Fourth, its linear decision boundary is easier to audit and validate, which matters for regulatory submissions. For research purposes, Gradient Boosting is the preferred model. For a clinical product designed to run on a smartwatch, Logistic Regression is the pragmatic choice.

3.7.3 Sensitivity/Specificity Trade-off in Medical Context

At the default 0.5 probability threshold, all four models achieve recall values between 0.9516 and 0.9569, meaning they miss somewhere between 4.3 and 4.8 percent of MI cases in the test set. In a population context, missing roughly 1 in 21 MI events is a meaningful failure rate for a first-alert system. The ROC curves, however, show that sensitivity can be raised above 0.95 by lowering the decision threshold, at the cost of increasing the false positive rate. Whether this trade-off is acceptable depends on the specific deployment context.

For a system that generates a push notification suggesting the user seek medical evaluation, a false positive rate of 10-15 percent might be acceptable if it brings sensitivity above 0.97 or 0.98. Users and clinicians would need to understand that the alert is probabilistic rather than diagnostic, and that a substantial fraction of alerts in a healthy user will be false alarms. For a system that automatically contacts emergency services, the threshold would need to be set much more conservatively to avoid unnecessary emergency responses, which would likely reduce sensitivity below the values reported here.

3.7.4 Limitations of This Study

This research has a few limitations that should be acknowledged.The dataset was collected under controlled conditions, and the real-world performance of these classifiers when applied to PPG signals collected from consumer wearables, with motion artifacts, variable contact pressure, and environmental interference, is unknown and likely lower. The dataset contains no demographic information, making it impossible to assess whether performance is consistent across sex, age groups, or skin tones, any of which could affect both PPG signal quality and MI presentation patterns.

The 2,000-feature raw amplitude representation, while effective here, is not a physiologically interpretable feature set. A classifier trained on these features cannot be straightforwardly explained to a clinician in terms of which aspects of the PPG waveform it found informative, which limits interpretability and may complicate regulatory review. The near-balanced class distribution of the dataset also means the reported performance figures do not directly predict performance in a real-world monitoring scenario where MI events are rare, and where the class imbalance would require additional methodological handling as described in Chapter 2. Finally, the study uses a single publicly available dataset; validation on independent datasets from different recording systems and populations would be needed before drawing strong conclusions about generalizability.

3.8 Proposed Wearable System Architecture



Figure 3.8: A prototype wrist-worn wearable device with an integrated PPG sensor and real-time display, showing physiological signal monitoring capabilities representative of the target deployment platform for the proposed MI detection system.

Based on the results of this study, a practical wearable MI detection system would consist of five functional components, as illustrated in Figure 3.8. The first component is the PPG sensor itself, operating in reflective mode at the wrist, sampling continuously at a rate of at least 100 Hz to capture the waveform features relevant to MI classification. Modern smartwatch hardware, including the sensor arrays in Apple Watch, Samsung Galaxy Watch, and equivalent devices, satisfies this requirement.

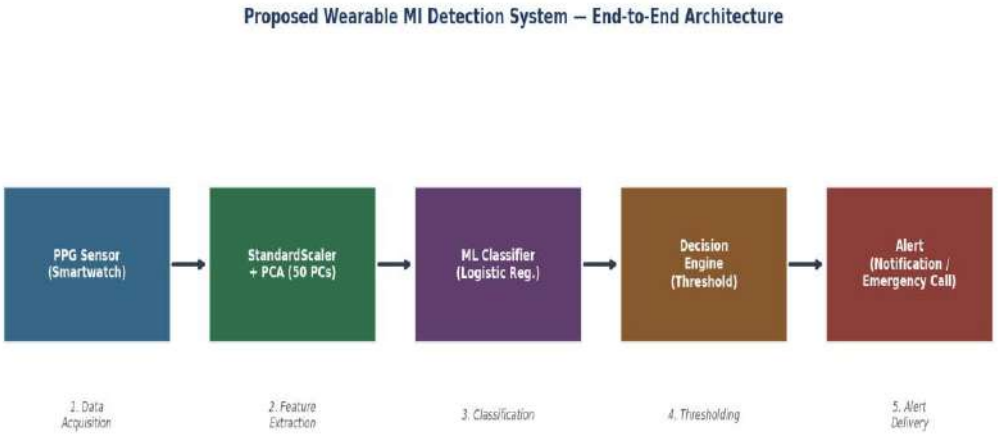


Figure 3.9: Proposed end-to-end wearable MI detection system architecture. Data flows from continuous PPG acquisition through preprocessing and dimensionality reduction to ML classification and threshold-based decision making, culminating in an alert delivered via push notification or emergency call. Source: Student's own design.

The second component handles signal preprocessing: windowing the continuous PPG stream into fixed-length segments, applying StandardScaler normalization using parameters pre-fitted during training, and projecting onto the 50 PCA components. This component can run on the smartphone processor via Bluetooth data relay, avoiding the computational constraints of the wearable chip. The third component runs the trained Logistic Regression classifier on the preprocessed feature vector, producing a probability estimate for MI. The fourth component applies the decision threshold, which can be adjusted to the desired sensitivity/specificity operating point. The fifth component manages the alert output: a push notification for lower-confidence alerts, escalating to an emergency services call for highconfidence flags, with user confirmation options to reduce unnecessary emergency responses.

Power consumption would be a central engineering consideration. Continuous PPG sampling at 100 Hz already draws meaningful current from the wrist sensor and associated LED driver. Transmitting data continuously to the smartphone via Bluetooth Low Energy adds further load. Running inference on the smartphone introduces minimal additional battery impact given the simplicity of Logistic Regression. End-to-end, a system of this design would likely reduce wearable battery life by 15-25 percent compared to a device running only standard heart rate monitoring, which is a trade-off that many users would accept given the potential benefit.

3.9 Conclusion

The results of this study confirm that PPG signals carry sufficient discriminative information for machine learning-based myocardial infarction detection, with all four classifiers achieving accuracy above 95 percent and AUC above 0.94 on the test set. Gradient Boosting achieved the highest AUC at 0.9537, while Logistic Regression, with identical accuracy at 95.5 percent and an AUC of 0.9460, is recommended for wearable deployment given its minimal computational footprint and interpretable decision boundary. These findings establish a viable proof-of-concept for passive, continuous MI detection on consumer wearables, though realworld validation on prospective, demographically diverse datasets collected under free-living conditions remains the essential next step before any clinical deployment can be considered.

General Conclusion

General Conclusion

This thesis set out to answer a question that sits at the intersection of biomedical engineering, machine learning, and consumer technology: can a wrist-worn wearable device, using only the PPG signal it is already collecting passively and continuously, detect myocardial infarction with enough accuracy to be clinically useful as an early warning system? The answer, based on the experimental results presented in Chapter 3, is a qualified yes. All four classifiers tested achieved accuracy above 95 percent and AUC above 0.94 on a balanced research dataset, and the consistency of this performance across such algorithmically different approaches, from the simple linear boundary of Logistic Regression to the sequential ensemble of Gradient Boosting, provides reasonable confidence that the result reflects a genuine discriminative signal in the PPG waveform rather than an artifact of any particular modeling choice.

The specific findings are worth stating clearly. Gradient Boosting achieved the highest AUC at 0.954 and the highest cross-validated AUC at 0.952, confirming its advantage in capturing nonlinear structure in the PCA-reduced feature space. Logistic Regression achieved identical accuracy at 95.5 percent with an AUC of 0.946, and it is the model recommended for deployment on a wearable device. The reasoning behind this recommendation is not that Logistic Regression is the most accurate, but that the performance gap between it and Gradient Boosting is small enough that it is unlikely to translate into meaningful clinical difference, while the computational gap is enormous: Logistic Regression inference requires a single matrix multiplication and a sigmoid evaluation, executable in microseconds on any embedded processor, whereas Gradient Boosting requires traversing 100 sequential decision trees at every inference cycle. For a device running on a battery measured in hundreds of milliampere-hours, that difference is decisive.

The limitations of the study are equally important to state. The dataset used was collected under controlled laboratory conditions, is balanced by design in a way that real-world MI monitoring data would not be, and carries no demographic information that would allow assessment of performance variation across age groups, sex, or skin tone. The 2,000-feature amplitude representation, while effective in this context, lacks physiological interpretability, which may complicate regulatory review and clinical acceptance. The class imbalance problem, in which MI events represent a tiny fraction of all observations in a realistic continuous monitoring scenario, was not addressed because the dataset does not require it, but it would need careful methodological handling in any real deployment. These limitations do not undermine the proof-of-concept value of the results, but they define the distance between what this thesis demonstrates and what a clinical product would require.

The immediate priorities for future work follow directly from these limitations. Prospective validation on data collected from consumer-grade wearables in free-living conditions, with diverse populations and demographic covariates recorded, is the foundational requirement for moving beyond proof-of-concept. Personalized modeling, which establishes individual physiological baselines and flags deviations from them rather than comparing to a population norm, addresses the known limitation that inter-individual variation in PPG morphology is substantial. Deep learning architectures, particularly convolutional networks operating directly on raw waveform sequences, may extract more discriminative information than the amplitude feature representation used here. Multi-modal fusion, combining the PPG signal with accelerometer, skin temperature, and SpO2 data that most premium wearables already collect, could improve both sensitivity and specificity simultaneously. And integration with emergency response systems, where a high-confidence alert triggers automatic notification of emergency services, represents the clinical endpoint toward which all of this technical development is ultimately directed.

The broader context of this work is worth keeping in view. The sensors needed for this application are already on hundreds of millions of wrists. The data is being collected. The machine learning methodology exists and has been validated here on a research dataset. What remains is the validation, the engineering, the regulatory work, and the clinical integration needed to translate a proof-of-concept finding into something that actually reaches the patients who are currently experiencing the early stages of a heart attack without any external warning, in the hours before they reach a hospital, when intervention could still make a decisive difference.

References

References

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement, 28*(3), R1–R39. https://doi.org/10.1088/0967-3334/28/3/R1

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., Carter, R. E., Yao, X., Rabinstein, A. A., Bhave, P. D., Venkatachalam, K. L., Friedman, P. A., & Kapa, S. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. *The Lancet, 394*(10201), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0

Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *npj Digital Medicine, 3*(1), 18. https://doi.org/10.1038/s41746-020-0226-6

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. https://link.springer.com/book/10.1007/978-0-387-45528-0

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Charlton, P. H., Pilt, K., & Kyriacou, P. A. (2022). Establishing best practices in photoplethysmography signal acquisition and processing. *Physiological Measurement, 43*(5), 050301. https://doi.org/10.1088/1361-6579/ac6cc4

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dunn, J., Runge, R., & Snyder, M. (2021). Wearables and the medical revolution. *Personalized Medicine, 15*(5), 429–448. https://doi.org/10.2217/pme-20180044

Elgendi, M., Fletcher, R., Liang, Y., Howard, N., Lovell, N. H., Abbott, D., Lim, K., & Ward, R. (2020). The use of photoplethysmography for assessing hypertension. *npj Digital Medicine, 2*(1), 60. https://doi.org/10.1038/s41746-019-0136-7

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Food and Drug Administration. (2019). *Proposed regulatory framework for modifications to artificial intelligence/machine learning-based software as a medical device*. U.S. Department of Health and Human Services. https://www.fda.gov/media/122535/download

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine, 25*(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/0471722146

Ibanez, B., James, S., Agewall, S., Antunes, M. J., Bucciarelli-Ducci, C., Bueno, H., Caforio, A. L. P., Crea, F., Goudevenos, J. A., Halvorsen, S., Hindricks, G., Kastrati, A., Lenzen, M. J., Prescott, E., Roffi, M., Valgimigli, M., Varenhorst, C., Vranckx, P., & Widimsky, P. (2018). 2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *European Heart Journal, 39*(2), 119–177. https://doi.org/10.1093/eurheartj/ehx393

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence, 2*, 1137–1143. https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf

Kyriacou, P. A., & Allen, J. (Eds.). (2021). *Photoplethysmography: Technology, signal analysis and applications*. Academic Press. https://doi.org/10.1016/C2019-0-041886

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill. https://www.cs.cmu.edu/~tom/mlbook.html

Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics, 22*(1), 122. https://doi.org/10.1186/s12910-021-00687-3

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future: Big data, machine learning, and clinical medicine. *New England Journal of Medicine, 375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., Balasubramanian, V., Russo, A. M., Rajmane, A., Cheung, L., Hung, G., Lee, J., Kowey, P., Talati, N., Nag, D., Gummidipundi, S. E., Beatty, A., Hills, M. T., Desai, S., & Apple Heart Study Investigators. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine, 381*(20), 1909–1917. https://doi.org/10.1056/NEJMoa1901183

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., & GBD-NHLBI-JACC Collaborators. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019. *Journal of the American College of Cardiology, 76*(25), 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Saner, H., van der Velde, E., & Scholte op Reimer, W. (2021). Digital health in cardiovascular disease. *European Journal of Preventive Cardiology, 28*(4), 355–358. https://doi.org/10.1177/2047487319895301

Seshadri, D. R., Davies, E. V., Harlow, E. R., Hsu, J. J., Knighton, S. C., Walker, T. A., Voos, J. E., & Drummond, C. K. (2020). Wearable sensors for COVID-19: A call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. *Frontiers in Digital Health, 2*, 8. https://doi.org/10.3389/fdgth.2020.00008

Shanker, M., Hu, M. Y., & Hung, M. S. (1996). Effect of data standardization on neural network training. *Omega, 24*(4), 385–397. https://doi.org/10.1016/0305-0483(96)00010-2

Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., & White, H. D. (2018). Fourth universal definition of myocardial infarction. *Journal of the American College of Cardiology, 72*(18), 2231–2264. https://doi.org/10.1016/j.jacc.2018.08.1038

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 2*(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

World Health Organization. (2021). *Cardiovascular diseases: Key facts*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)