

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL

**TRANSFORMER BASED INTERPRETABLE ALGORITHM FOR
DIAGNOSING MYOCARDIAL INFARCTION FROM
ELECTROCARDIOGRAPHY DATA**

BARIŞ ÖZCAN

**A THESIS OF MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING**

ADVISOR: PROF.DR. YUSUF SİNAN AKGÜL

DECEMBER 2024

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL

**TRANSFORMER BASED INTERPRETABLE ALGORITHM
FOR DIAGNOSING MYOCARDIAL INFARCTION FROM
ELECTROCARDIOGRAPHY DATA**

BARIŞ ÖZCAN

**A THESIS OF MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING**

ADVISOR: PROF.DR. YUSUF SİNAN AKGÜL

DECEMBER 2024

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ELEKTROKARDİYOĞRAFI VERİLERİNDEN MİYOKARD
ENFARKTÜSÜNÜ TEŞHİSİ İÇİN TRANSFORMER TABANLI
YORUMLANABİLİR ALGORİTMA GELİŞTİRİLMESİ

BARIŞ ÖZCAN

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMAN: PROF.DR. YUSUF SİNAN AKGÜL

ARALIK 2024

MASTER of SCIENCE JURY APPROVAL FORM

A thesis submitted by Barış ÖZCAN, defended on 10/12/2024 before the jury formed with the 02/12/2024 date and 2024/64 numbered decision of the GTU Graduate Administration Board, has been accepted as a MASTER of SCIENCE thesis in the Department of Computer Engineering Program.

JURY

MEMBER

(THESIS ADVISOR) : Prof. Dr. Yusuf Sinan AKGÜL

MEMBER

: Assoc. Prof. Yakup GENÇ

MEMBER

: Assist. Prof. Nur Banu ALBAYRAK

APPROVAL

Gebze Technical University Graduate Administration Board
...../...../..... date and/..... numbered decision.

SIGNATURE/SEAL



*To my beloved mother
and father, with all my love
and gratitude.*

ABSTRACT

Myocardial infarction (MI), a fatal cardiovascular disease, occurs when the heart muscle is damaged due to lack of oxygen. Early diagnosis is of critical importance, yet interpreting electrocardiogram(ECG) signals requires expertise and time. Therefore, the development of algorithms capable of diagnosing MI using ECG data is studied well in the fields of health and artificial intelligence. In this study, four different deep learning models were explored: Convolutional Neural Networks (CNNs),CRNNs(Convolutional Recurrent Neural Networks), CNNs with Multi-Head Attention, and Transformer-based models. These models were trained and tested on the combined dataset of the publicly available PTB, PTB-XL, and INCART datasets. The performances of the models were evaluated using metrics such as accuracy, sensitivity, specificity, and F1 score. The transformer-based model was the most successful model with an accuracy rate of 93%. On the other hand, the study focused on the transparency of the models, where Grad-CAM based methods and a perturbation algorithm inspired by adversarial attacks were used to disrupt derivations and analyze the behavior of the models.

Keywords: Electrocardiography, Deep Learning ,Attention Mechanism, Transformer, Perturbation, Adversarial Attack

ÖZET

Miyokard enfarktüsü (ME), kalp kasının oksijen eksikliği nedeniyle hasar gördüğü ölümcül kardiyovasküler hastalıklardan biridir. Erken teşhis kritik önem taşır, ancak elektrokardiyogram(EKG) ölçümlerinden yorumlanması uzmanlık ve zaman gerektiren bir işlemdir. Bu nedenle, EKG verilerini kullanarak EKG sinyallerinden ME teşhisi yapabilen algoritmaların geliştirilmesine sağlık ve yapay zeka alanlarında önem verilmektedir. Bu çalışmada, dört farklı derin öğrenme modeli denenmiştir: CNN (Evrişimsel Sinir Ağları), CRNN(Evrişimsel Yinelemeli Sinir Ağları), CNN ile Çok Başlı Dikkat ve Transformer tabanlı modeller. Bu modeller, herkese açık olarak erişilebilen PTB, PTB-XL ve INCART veri setlerinin birleştirilmiş hali üzerinde eğitildi ve test edildi. Modellerin performansı doğruluk, duyarlılık, özgüllük ve F1 skoru gibi metriklerle değerlendirilmiştir. Transformer tabanlı model %93'lük doğruluk oranıyla en başarılı model oldu. Diğer yandan, çalışmanın önemli bir kısmını modellerin şeffaflaştırılmasına ayrılmış ve Grad-CAM bazlı yöntemlerin yanında karşıt saldırıları örnek alan bir pertürbasyon algoritması sayesinde derivasyonlar sırayla bozulmuş ve modellerin davranışı analiz edilmiştir.

Anahtar Kelimeler: Elektrokardiyografi, Derin Öğrenme, Dikkat Mekanizması, Transformer, Pertürbasyon, Karşıt Saldırı

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my esteemed advisor, Prof. Dr. Yusuf Sinan AKGÜL for their invaluable guidance, wisdom, and patience throughout my master's journey. Their unwavering dedication, generosity with their time and knowledge, and readiness to offer support in every aspect of my work have been instrumental in my success. I am immensely grateful for their constant encouragement and insight.

I would also like to extend heartfelt thanks to my family, whose unwavering support and belief in me have been a constant source of strength and motivation throughout this thesis and beyond.

This thesis was supported by the TÜSEB (Türkiye Sağlık Enstitüleri Başkanlığı) 2022-EKG-01 Program (Project No: 20107). I gratefully acknowledge the financial support provided by TUSEB, which made this research possible. The funding from this program has been instrumental in facilitating the data collection, analysis, and overall execution of this study. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of TUSEB or the Health Institutes of Turkey.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	vi
ÖZET	vii
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS	ix
LIST OF SYMBOLS AND ABBREVIATIONS	xi
LIST OF FIGURES	xii
LIST OF TABLES	xiii
1. INTRODUCTION	1
1.1. Electrocardiogram	1
1.2. AI and Healthcare	4
1.3. Machine Learning and Deep Learning	7
1.3.1. Convolutional Neural Networks(CNN)	11
1.3.2. Long-Short Term Memory(LSTM)	16
1.3.3. Convolutional Neural Networks(CRNN)	22
1.3.4. Transformer and Attention	24
2. RELATED WORK	31
3. METHOD AND METHODOLOGY	39
3.1. Dataset	39
3.1.1. PTB Diagnostic ECG Database	39
3.1.2. PTB-XL Database	41
3.1.3. St Petersburg INCART 12-lead Arrhythmia Database	43
3.1.4. Merged Dataset	45
3.2. Equipment	49
3.3. CNN Model	50
3.4. CRNN Based Model	52
3.5. CNN and Attention Model	54
3.6. Our Proposed Transformer Based Model	56
3.7. Hyperparameter Optimization	60
4. RESULTS AND DISCUSSIONS	65
4.1. Result	65
4.2. Interpretability	70
4.2.1. GRAD-CAM	71
4.2.3. Lead Importance	74
5. CONCLUSIONS	80
5.1. Discussions	80
5.2. Future Work	81
5.3. Conclusion	82

REFERENCES	83
BIOGRAPHY	91
PUBLICATIONS AND PRESENTATIONS FROM THE THESIS	92



LIST OF SYMBOLS AND ABBREVIATIONS

∇	: Operator Nabla
θ	: Sigma
η	: Learning Rate
E	: Error
τ	: Tau
∂	: Partial Derivative
μ	: Mean
σ	: Variance
L	: Loss
γ	: Gamma
β	: Beta
ε	: Epsilon
α	: Alpha
Ω	: Resistance
$\odot$	: Element-Wise Multiplication
ECG	: Electrocardiogram
CVD	: Cardiovascular Diseases
BPTT	: Back Propagation Through Time
ML	: Machine Learning
NLP	: Natural Language Processing
SVM	: Support Vector Machine
DNN	: Deep Neural Network
CNN	: Convolutional Neural Network
RNN	: Recurrent Neural Network
CRNN	: Convolutional Recurrent Neural Network
LSTM	: Long Short-Term Memory
GRU	: Gated Recurrent Unit
AI	: Artificial Intelligence
OCR	: Optical Character Recognition
XAI	: Explainable Artificial Intelligence
ReLU	: Rectified Linear Unit
PTB	: PhysioNet PTB Database
INCART	: PhysioNet INCART Database
NORM	: Normal
HYP	: Hypertrophy
MI	: Myocardial Infarction
GAN	: Generative Adversarial Network
SGD	: Stochastic Gradient Descent
Adam	: Adaptive Moment Estimation
RMSProp	: Root Mean Square Propagation
AUC	: Area Under the Curve
Grad-CAM	: Gradient-weighted Class Activation Mapping
GTÜ	: Gebze Teknik Üniversitesi
TÜSEB	: Türkiye Sağlık Enstitüleri Başkanlığı

LIST OF FIGURES

	<u>Page</u>
Figure 1.1: Global Causes of Death Between 2000-2019 [2].	1
Figure 1.2: 12-Lead ECG Placement	2
Figure 1.3: ECG Signal Morphology [6].	3
Figure 1.4: Estimated AI Market Size in Healthcare from 2023 to 2034 [6].	4
Figure 1.5: AI-Enabled Devices Across Medical Disciplines [10].	5
Figure 1.6: AI Venn Diagram.	7
Figure 1.7: Example Neural Network Architecture.	8
Figure 1.8: Perceptron Diagram.	9
Figure 1.9: End to End Deep Learning Pipeline.	10
Figure 1.10: Simple CNN Network.	12
Figure 1.11: Simple Convolution Operation for 2D CNN.	13
Figure 1.12: LSTM Diagram.	19
Figure 1.13: Scaled Dot-Product Attention.	25
Figure 1.14: Multi-Head Attention Diagram [59].	26
Figure 1.15: Transformer Architecture [59].	27
Figure 2.1: Evolution of Machine Learning Methods on ECG [80].	32
Figure 3.1: Label Distribution of the PTB-XL Database.	42
Figure 3.2: Label Distribution of the Combined Dataset.	47
Figure 3.3: Age Distribution of Dataset.	48
Figure 3.4: Sex Distribution of Dataset.	49
Figure 3.5: High Level Architecture of Our Model.	57
Figure 4.1: Blackbox Diagram of Machine Learning Models [111].	70
Figure 4.2: Grad-CAM of HYP Example.	72
Figure 4.3: Grad-CAM of MI Example.	72
Figure 4.4: Grad-CAM of NORM Example.	73
Figure 4.5: Transformer Model's Heatmap of Accuracy Decrease Per Lead With Respect to Noise.	75
Figure 4.6: CNN with Attention Model's Heatmap of Accuracy Decrease Per Lead With Respect to Noise.	76
Figure 4.7: CRNN Model's Heatmap of Accuracy Decrease Per Lead With Respect to Noise.	77
Figure 4.8: CNN Model's Heatmap of Accuracy Decrease Per Lead With Respect to Noise.	78

LIST OF TABLES

	<u>Page</u>
Table 3.1: Subject Diagnoses of PTB Diagnostic ECG Database.	40
Table 3.2: St Petersburg INCART 12-lead Arrhythmia Database	44
Table 3.3: CNN Model Hyperparameter Search Ranges.	60
Table 3.4: CRNN Based Model Hyperparameter Search Ranges.	61
Table 3.5: CNN with Attention Model Hyperparameter Search Ranges.	62
Table 3.6: Transformer-Based Model Hyperparameter Search Ranges.	63
Table 4.1: Comparative Performance Metrics for All Models.	65
Table 4.2: Class-Wise Performance Metrics for All Models.	67



1. INTRODUCTION

1.1. Electrocardiogram

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, responsible for an estimated 17.9 million deaths annually, a number projected to rise to 23.6 million by 2030 [1]. The increasing burden of these diseases poses a significant threat to global health, underscoring the urgent need for effective diagnostic and therapeutic strategies. Among the various cardiovascular conditions, myocardial infarction (MI) holds a special significance due to its high mortality rate and the critical need for timely intervention.

Global causes of death 2000, 2010, 2019

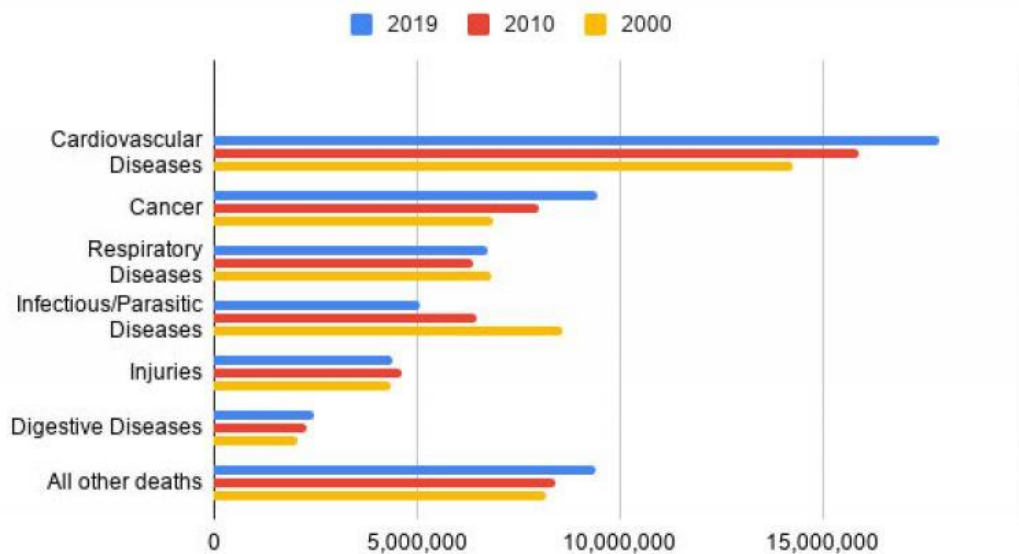


Figure 1.1 Global Causes of Death Between 2000-2019 [2].

Myocardial infarction, commonly known as a heart attack, occurs when blood flow to a part of the heart is obstructed for a long enough period to cause damage or death to heart muscle tissue. This obstruction is most often caused by the buildup of plaques in the coronary arteries, a process known as atherosclerosis. The global incidence of MI

underscores its prevalence; millions require emergency treatment annually, and many succumb to the condition.

Early diagnosis is crucial in the management of myocardial infarction and significantly improves the prognosis. Traditional diagnostic tools include the electrocardiogram (ECG) and blood tests for cardiac biomarkers. In an ECG, critical indicators such as changes in the ST segment, alterations in T-wave height and inversion, and the presence of pathological Q waves are pivotal for diagnosing MI [3]. The identification of these changes can inform timely therapeutic interventions, drastically improving patient outcomes.

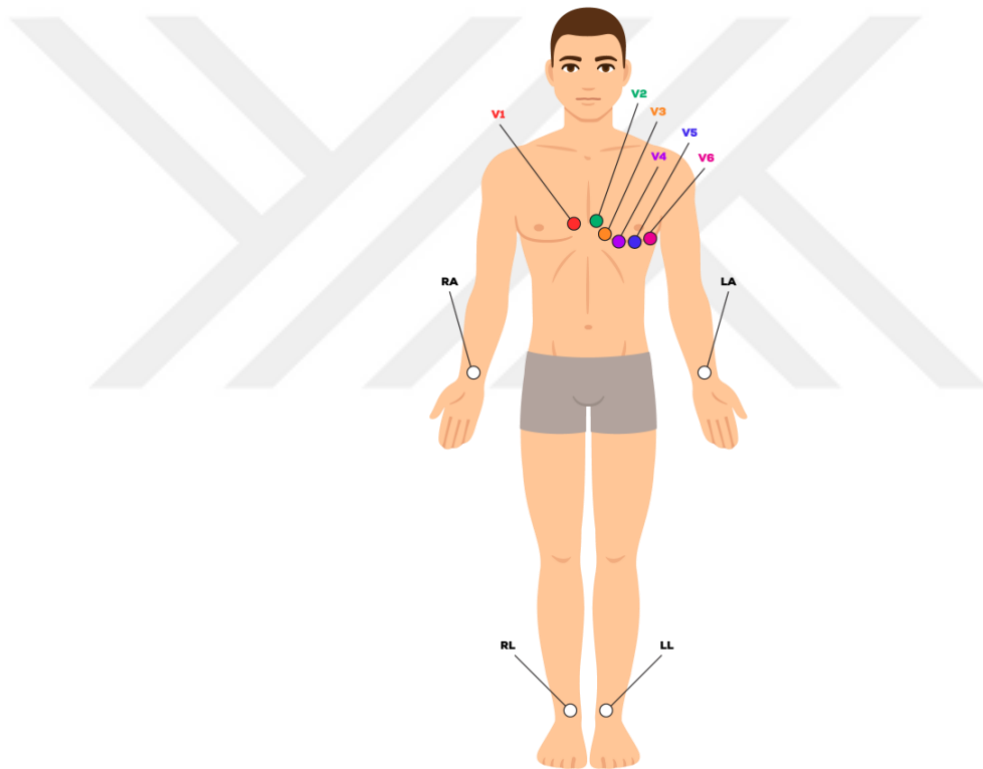


Figure 1.2 12-Lead ECG Placement [4].

Despite the effectiveness of these traditional methods, there are inherent challenges such as the need for timely interpretation by skilled clinicians, which is not always feasible, especially in under-resourced or rural areas. This gap highlights the potential impact of automatic ECG diagnostic systems. Leveraging artificial intelligence (AI), these systems can accelerate the diagnostic process, reduce human error, and extend

critical diagnostic support to regions suffering from a shortage of healthcare professionals.

The integration of AI-driven technologies in cardiovascular care not only promises to enhance the accuracy and efficiency of diagnostics but also supports a broader initiative to make healthcare more accessible and equitable across the globe. As such, the development and deployment of automated ECG diagnostic systems in clinical practice could be a pivotal advancement in combating the rising tide of cardiovascular disease[5].

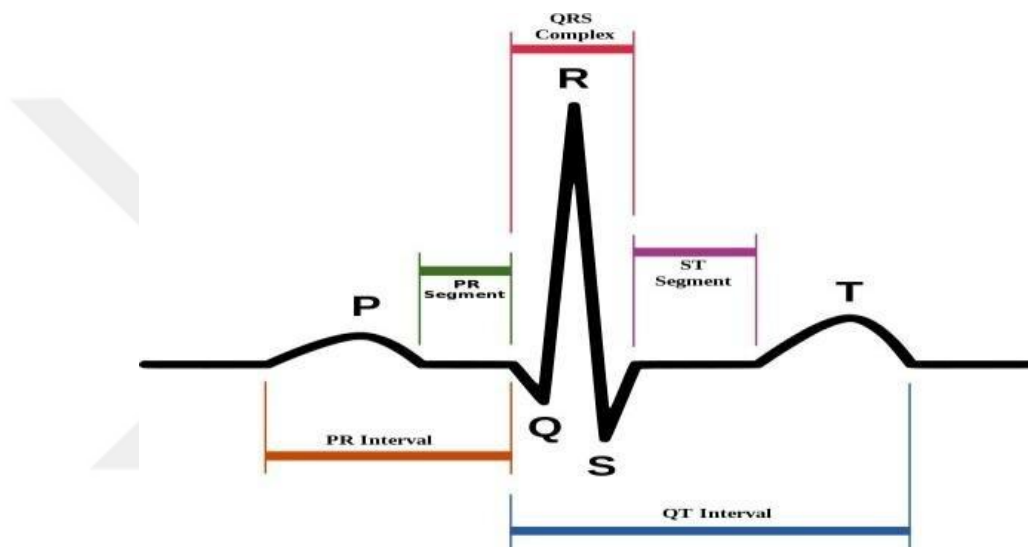


Figure 1.3 ECG Signal Morphology [6].

In conclusion, addressing the global challenge of cardiovascular diseases, with a focus on myocardial infarction, requires a multifaceted approach that includes advancements in diagnostic technologies, improved healthcare access, and ongoing research into more effective treatments. The role of technology, particularly AI in medical diagnostics, emerges as a crucial component in this endeavor, offering new avenues for saving lives and improving health outcomes worldwide.

1.2. AI and Healthcare

Artificial Intelligence stands as a beacon of innovation, significantly altering various sectors with its profound capabilities. Among these fields, healthcare witnesses an unparalleled transformation due to AI's integration. The advent of AI in medical practice holds immense promise for enhancing patient care quality, bolstering diagnostic precision, and refining therapeutic approaches. Particularly within cardiology, the utility of AI becomes paramount; the early identification and accurate diagnosis of myocardial infarction are vital for optimal patient prognoses.

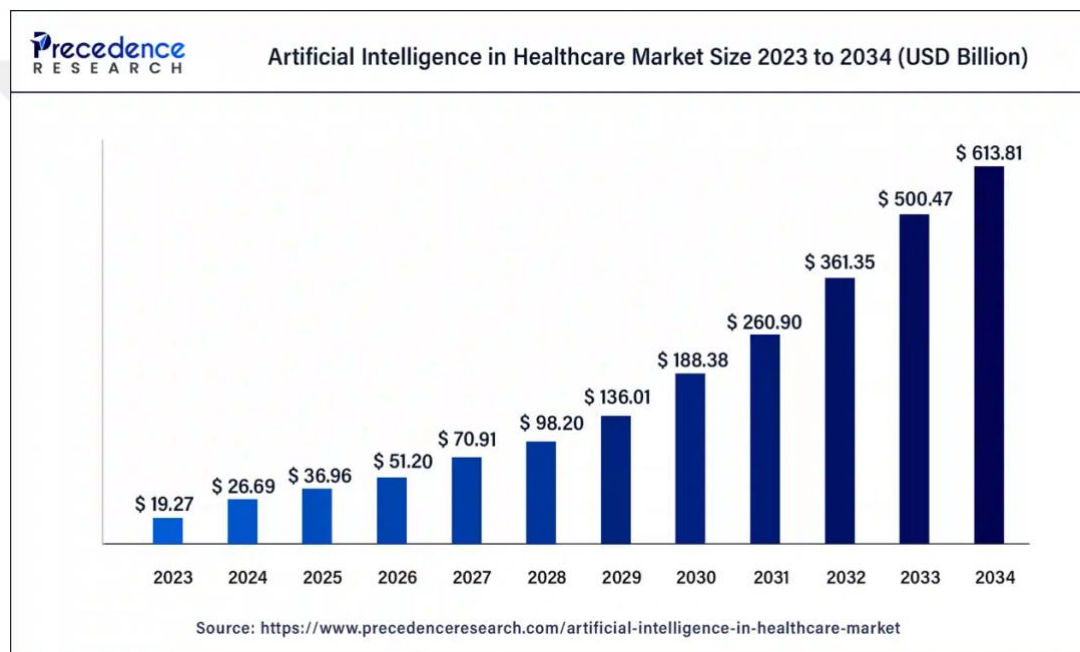


Figure 1.4 Estimated AI Market Size in Healthcare from 2023 to 2034 [7].

A vast spectrum encompasses AI technologies such as machine learning, deep learning, and natural language processing. These methodologies empower computers to assimilate data and execute tasks conventionally reliant on human intellect. Within healthcare settings, sophisticated algorithms sift through extensive medical datasets to reveal patterns previously unseen by human eyes. This computational prowess equips clinicians with superior decision-making tools that support both the accuracy and efficacy of patient management protocols while simultaneously driving down overall healthcare expenditure.

The transformative impact of artificial intelligence within the healthcare sector is evident in its capacity to handle and analyze vast amounts of complex medical information with remarkable efficiency. Electronic health records (EHRs), sophisticated imaging modalities, and data from biosensors like electrocardiograms furnish an abundant repository for AI systems to draw upon. Through meticulous training on these datasets, machine learning models can learn patterns and associations that might escape even experienced professionals. Notably, deep learning methodologies have exhibited commendable success in parsing ECG data to detect myocardial infarctions, surpassing conventional rule-based mechanisms.

Beyond diagnostic applications, AI finds utility across multiple aspects of patient management. It contributes to risk assessment, personalized treatment strategies, and continuous monitoring from a distance. Taking wearable technology as an example: AI can analyze data from such devices to alert healthcare practitioners about potential issues or worsening conditions promptly enough for necessary interventions [8]. Moreover, the role of AI in tailoring treatments to individual patients cannot be understated. By identifying subgroups within patient populations who may react uniquely to certain therapies, it enables the crafting of customized treatment plans. Such precision medicine efforts ensure that therapeutic interventions are specifically suited to individual needs rather than applying a one-size-fits-all strategy [9].

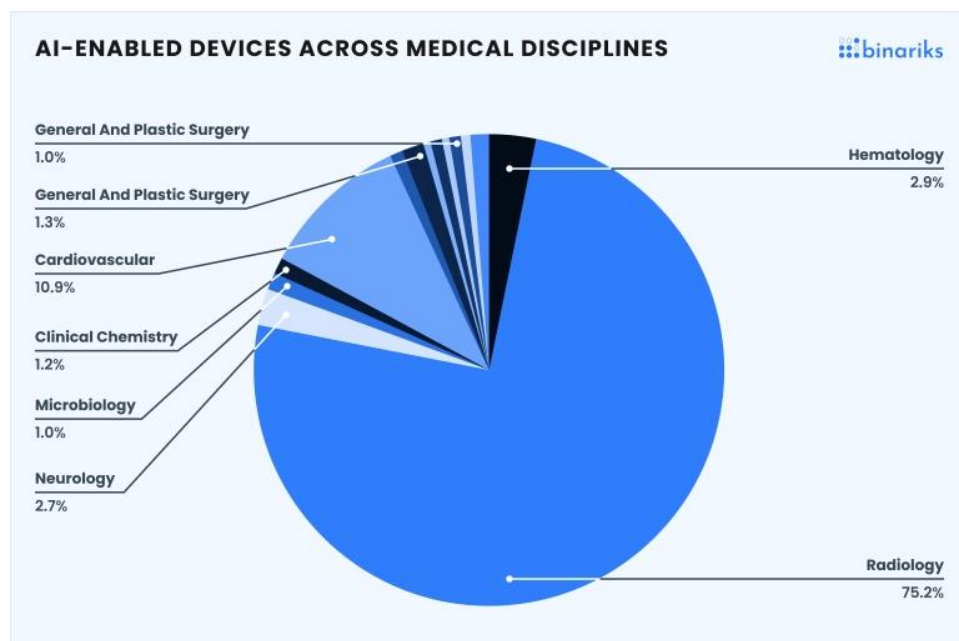


Figure 1.5 AI-Enabled Devices Across Medical Disciplines [10].

Integrating AI in healthcare brings forward challenges, notably regarding the clarity and openness of AI processes. Many sophisticated algorithms function as "black boxes," obscuring the pathways leading to their predictions or decisions. This opacity can hinder trust and limit the integration of AI tools in medical practice. To address this issue, developing explainable artificial intelligence (XAI) is imperative to ensure these systems are clear, interpretable, and reliable [11].

The significance of XAI within healthcare is crucial. Medical decisions have profound implications for patient outcomes; thus, it is essential that clinicians have confidence in AI-driven recommendations. With explainable AI, healthcare providers can grasp the variables affecting predictions made by these systems, allowing them to evaluate reliability and make well-informed choices based on their professional expertise. Furthermore, XAI promotes effective dialogue between automated systems and medical practitioners. When clinicians understand how an algorithm reaches a decision, it not only enhances their confidence but also promotes collaborative efforts toward patient care. By clarifying machine outputs through interpretability methods, clinicians are better equipped to align machine suggestions with clinical realities [12].

Techniques for making AI more transparent aim to illuminate the rationale behind algorithmic predictions. One such method is feature importance analysis, which highlights the variables that heavily influence a model's decisions. Another approach involves extracting rules that can be easily understood by humans from complex models. Visual tools like heat maps and attention mechanisms also serve this purpose, illustrating which parts of medical images are crucial for AI-generated predictions. These methods empower healthcare providers to better grasp and trust the outputs generated by AI systems. By incorporating these explainable AI techniques, healthcare professionals can better understand and validate the outputs of AI models, leading to increased confidence in their use for clinical decision-making.

Ethical and legal aspects emerge prominently when integrating AI into healthcare environments. Matters of data privacy, security, and bias must be addressed with ultimate care to ensure responsible use of these technologies. Strong regulatory frameworks are essential in overseeing the development and application of AI systems in medical settings. Moreover, continuous education for healthcare workers is crucial so they can effectively operate these tools and interpret their results accurately.

To sum up, the application of AI in healthcare holds immense potential for improving patient care and advancing medical research. The ability of AI algorithms to analyze vast amounts of medical data and provide decision support to healthcare professionals can significantly enhance the accuracy and efficiency of diagnostic processes, including the detection of myocardial infarction. However, the development of explainable AI techniques is essential to ensure the transparency and trustworthiness of AI-based decision support systems in clinical practice. As AI continues to evolve and mature, its integration into healthcare will likely become increasingly prevalent, transforming the way we approach patient care and medical decision-making. Addressing the challenges associated with AI in healthcare, such as explainability, ethics, and regulation, will be crucial to realizing its full potential and ensuring its safe and effective implementation in clinical settings.

1.3. Machine Learning and Deep Learning

The domains of Machine Learning (ML) and Deep Learning (DL), integral parts of Artificial Intelligence, have gained substantial attention in recent times, notably within healthcare. These advanced technologies provide significant capabilities for examining complex medical datasets such as electrocardiograms. The application of ML and DL has yielded encouraging outcomes in identifying and diagnosing cardiovascular conditions like myocardial infarction.

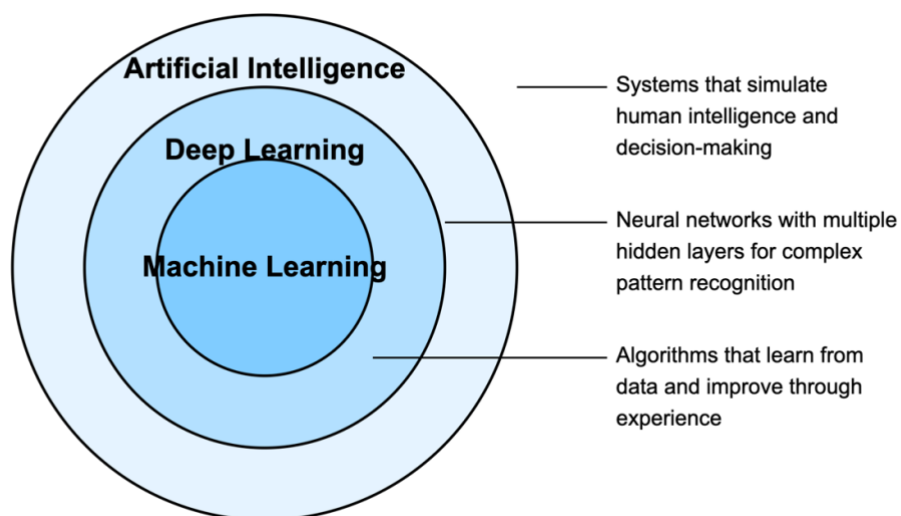


Figure 1.6 AI Venn Diagram.

Machine Learning encompasses a variety of algorithms designed to infer from data, thereby making predictions or decisions autonomously [13]. These algorithms do not rely on explicit programming to accomplish particular tasks. Within this sphere, ML can be classified into three primary categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms are instructed using labeled datasets where the target output is predetermined; conversely, unsupervised learning identifies structures within unlabeled data sets. Reinforcement learning differs by having an agent that learns optimal actions through interactions with its surroundings [14].

Deep Learning represents an advanced segment of Machine Learning, deeply rooted in artificial neural networks which draw inspiration from the human brain's architecture and functionality [15]. Deep Learning models, particularly deep neural networks (DNNs), consist of multiple layers of interconnected nodes or "neurons" that can learn hierarchical representations of data. This ability to automatically learn features from raw data makes Deep Learning particularly suitable for complex tasks such as image recognition, natural language processing, and medical diagnosis [16].

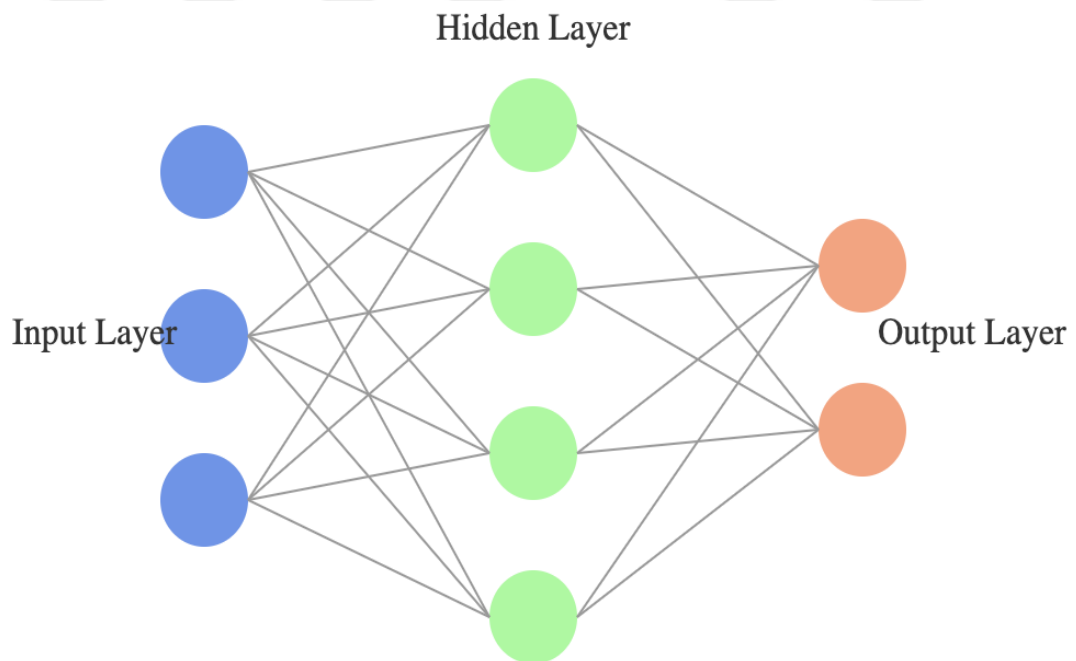


Figure 1.7 Example Neural Network Architecture.

At its core is the perceptron—an elementary yet vital component known as an artificial neuron. Each perceptron receives a different input signals; it assigns specific weights to these inputs before aggregating them into a sum that is subsequently passed through an activation function [17]. The activation function’s role is essential as it injects non-linearity into the network’s operations, thereby facilitating the learning of complex patterns embedded within the data. Widely adopted activation functions cover sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU), each contributing uniquely to pattern recognition capabilities [18].

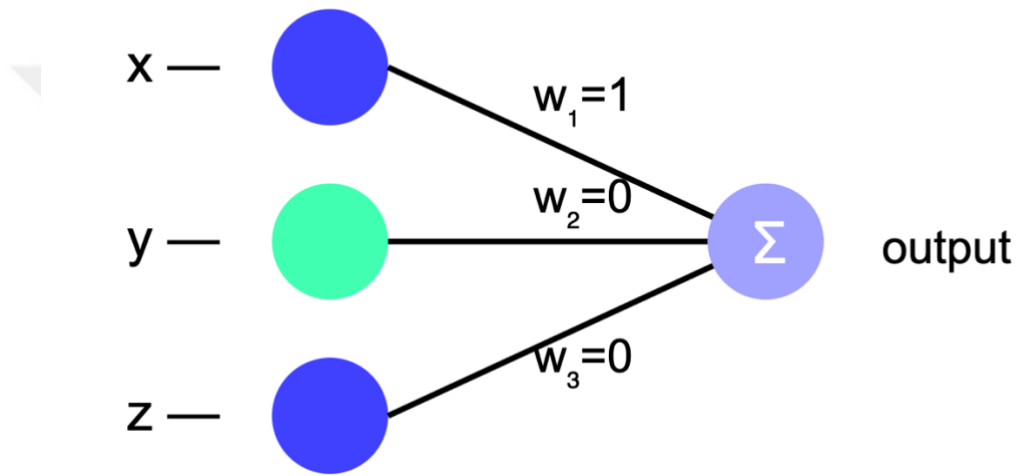


Figure 1.8 Perceptron Diagram.

$$y = f(\sum_{i=1}^n w_i x_i + b) \quad (1.1)$$

Where y is the output, f is the activation function, w_i are the weights, x_i are the inputs and b is the bias.

Neural networks, structured in distinct layers, start with an input layer to gather initial data, proceed through one or more hidden layers for advanced processing, and conclude with an output layer that delivers the ultimate results. The term "deep" within Deep Learning means the inclusion of numerous hidden layers within this architecture [19]. These intermediate stages are crucial as they progressively extract sophisticated features from the raw input, allowing the network to recognize and learn rich representations.

The learning process in neural networks involves adjusting the weights and biases of the connections between neurons to minimize the difference between the network's predictions and the actual outcomes. This is typically achieved through an optimization algorithm called backpropagation, which calculates the gradient of the error with respect to the network's parameters and updates them accordingly [20].

$$w_{i,j}^{(t+1)} = w_{i,j}^{(t)} - \eta \frac{\partial E}{\partial w_{i,j}} \quad (1.2)$$

Where $w_{i,j}^{(t+1)}$ is the weight at time t , η is the learning rate, and $\frac{\partial E}{\partial w_{i,j}}$ is the gradient of the error with respect to weight.

Various optimization techniques, such as stochastic gradient descent (SGD) and its variants (e.g., Adam, RMSprop), are used to efficiently train neural networks [21].

$$\theta := \theta - \eta \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)}) \quad (1.3)$$

Where θ represents the parameters, η is the learning rate, and $J(\theta; x^{(i)}; y^{(i)})$ is the cost function for i -th training example.

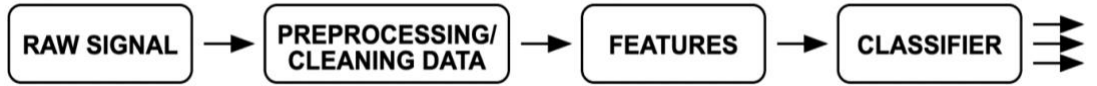


Figure 1.9 End to End Deep Learning Pipeline.

A prominent advantage of Deep Learning is its capability to autonomously extract significant features from unprocessed data, bypassing the necessity for manual feature engineering. This proves particularly advantageous in medical fields, where data complexity often complicates the identification of pertinent features manually [22]. In analyzing ECGs, these models can distinguish patterns and morphological alterations indicative of conditions like myocardial infarction or other cardiac issues. However, their success relies significantly on access to extensive, high-quality datasets for training purposes. Collecting such datasets in medical contexts presents difficulties due to privacy constraints, diverse data sources, and the need for expert labeling.

Transfer learning techniques, which involve using pre-trained models on large datasets and fine-tuning them for specific tasks, can help mitigate this issue to some extent [23].

Another important consideration in the application of Deep Learning to medical diagnosis is the interpretability of the models. While Deep Learning models can achieve high accuracy, their decision-making process is often opaque, leading to the "black box" problem [24]. This lack of transparency can be a significant barrier to the adoption of these models in clinical practice, where understanding the reasoning behind a diagnosis is crucial. Various techniques for explainable AI, such as attention mechanisms, saliency maps, and layer-wise relevance propagation, are being developed to address this challenge [25].

1.3.1. Convolutional Neural Networks(CNN)

Convolutional Neural Networks (CNNs) represent an advanced class of neural networks tailored to process data with grid-like structures, such as two-dimensional images or one-dimensional time series [26]. Introduced by LeCun and colleagues in 1989 [27], these networks have transformed the field of computer vision, becoming the preferred architecture for numerous spatial data tasks. Their effectiveness in image recognition was evidently demonstrated through achievements like AlexNet in the ImageNet competition [28], which subsequently encouraged their adoption across various fields including medical imaging and signal processing [29]. The core innovation of CNNs is their use of convolutional layers. These layers employ a set of learnable filters that slide over input data to perform element-wise multiplication followed by summation, resulting in feature maps [30]. This technique allows CNNs to efficiently capture local patterns and hierarchical spatial features within the data—a critical capability when dealing with images or temporal sequences where spatial relationships are essential. One significant advantage of CNNs lies in their ability to learn relevant features directly from raw data, thereby removing the necessity for manual feature engineering often required by traditional machine learning methods. The shared weights used in convolutional layers further enhance efficiency by reducing parameter count compared to fully connected networks [31]. This not only improves computational performance but also mitigates overfitting issues when handling high-dimensional datasets. Furthermore, this architectural approach enables CNNs to exploit inherent spatial or temporal structures present within diverse types of

data effectively. As a result, they achieve exceptional performance across a broad spectrum of applications ranging from object detection and facial recognition to natural language processing and time series analysis [32].

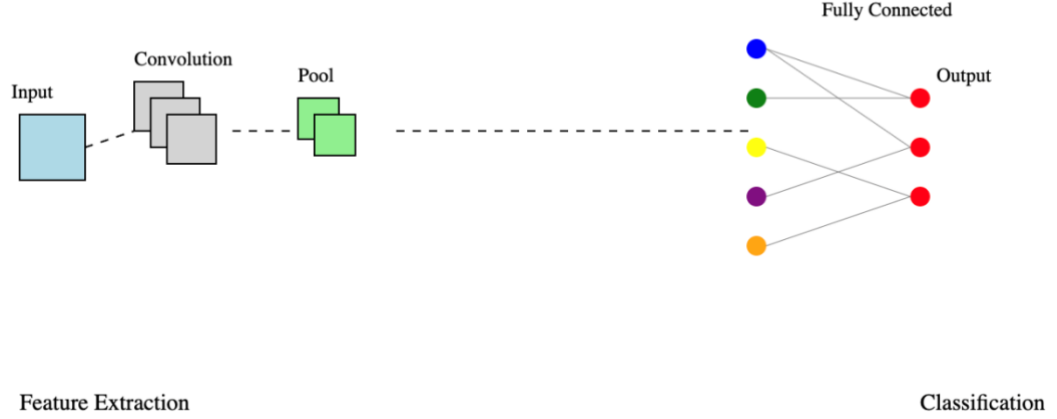


Figure 1.10 Simple CNN Network.

The architecture of CNNs typically consists of several key components, each playing an important role in the network's ability to learn and generalize from data. Convolutional layers form the core of a CNN and are responsible for feature extraction. Each convolutional layer applies a set of filters to the input, creating feature maps that highlight different aspects of the data. The convolution operation can be mathematically expressed as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (1.4)$$

where f is the input function, g is the filter, and $(f * g)$ is the convolved output [33]. In practice, for discrete data like images or time series, this operation is implemented as a sum:

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n - m] \quad (1.5)$$

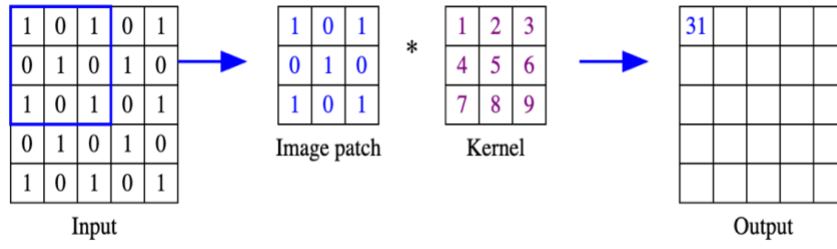


Figure 1.11 Simple Convolution Operation for 2D CNN.

Following the convolutional layers, non-linear activation functions are applied to introduce non-linearity into the network. The most commonly used activation function in CNNs is the Rectified Linear Unit (ReLU), defined as:

$$f(x) = \max(0, x) \quad (1.6)$$

In conjunction with convolutional layers, activation functions such as ReLU play an integral part in CNN architecture. Known for their straightforward implementation and efficacy in addressing vanishing gradient issues, ReLU functions have been widely adopted [34]. Alternatives like Leaky ReLU, Parametric ReLU, and Swish have emerged over time to serve specific needs across different models. Pooling layers are another essential component, reducing the spatial dimensions of the feature maps and providing a form of translation invariance while reducing computational complexity. The most common pooling operations are max pooling and average pooling. For a 2x2 max pooling operation:

$$y_{ij} = \max(x_{2i,2j}, x_{2i,2j+1}, x_{2i+1,2j}, x_{2i+1,2j+1}) \quad (1.7)$$

Pooling functions play a significant role in diminishing the sensitivity of outputs to minor translations and distortions within inputs [35].

This process, typically involving operations such as max pooling or average pooling, condenses information while maintaining key features. Meanwhile, fully connected layers generally reside towards the terminal stages of a neural network. These layers are integral for classification tasks as they connect every neuron from one layer to each

neuron in another, enabling the model to detect overall patterns across the feature space [36].

The training process for CNNs typically involves backpropagation and gradient-based optimization methods, which are crucial for the network to learn effective representations from the data. The backpropagation algorithm calculates the gradient of the loss function with respect to each weight by the chain rule, iteratively from the last layer to the first. This can be expressed as:

$$\partial L / \partial w_i = \partial L / \partial y \cdot \partial y / \partial x_i \cdot \partial x_i / \partial w_i \quad (1.8)$$

where L is the loss function, y is the output, x_i is the input to the i -th layer, and w_i are the weights of the i -th layer [37]. Various optimization algorithms are used to update the weights based on these gradients, with Stochastic Gradient Descent (SGD) and its variants like Adam and RMSprop being popular choices [38]. The general update rule for SGD can be written as:

$$w_{t+1} = w_t - \eta \nabla L(w_t) \quad (1.9)$$

where η is the learning rate and $\nabla L(w_t)$ is the gradient of the loss function with respect to the weights at time t . Techniques like batch normalization are often employed to normalize the inputs of each layer, which can speed up training and improve generalization [39]. The batch normalization operation can be expressed as:

$$y = \gamma(x - \mu_B) / \sqrt{(\sigma_B^2 + \varepsilon)} + \beta \quad (1.10)$$

where μ_B and σ_B^2 are the mean and variance of the mini-batch, γ and β are learnable parameters, and ε is a small constant for numerical stability. These training techniques, combined with the architectural innovations of CNNs, enable the networks to learn complex, hierarchical representations of the input data, making them particularly effective for tasks involving spatial or temporal patterns [40].

While 2D CNNs have been widely successful in image processing tasks, 1D CNNs have gained significant consideration for analyzing time series data, including ECG signals. 1D CNNs operate on one-dimensional data, making them particularly suitable for tasks involving temporal or sequential data [41]. The 1D convolution operation can be mathematically expressed as:

$$y[n] = (x * w)[n] = \sum_{m=0}^{M-1} x[n + m] \cdot w[m] \quad (1.11)$$

where x is the input signal, w is the filter (kernel) of length M , and y is the output feature map [42]. 1D convolutional neural networks present notable advantages for the analysis of electrocardiogram data. These networks excel in extracting temporal features efficiently, while requiring less computational power than their 2D counterparts. This makes them ideal for handling inputs of varying lengths and enables them to discern local patterns and attributes within the time domain. Such capability is essential in ECG interpretation as convolutional layers can learn to identify specific waveforms like P waves and QRS complexes, along with their temporal interrelations [43]. The reduced computational requirements of 1D CNNs make them more suitable for real-time applications and deployment on resource-constrained devices, an important consideration in many medical applications [44]. Furthermore, the integration of pooling layers enhances the system's durability to minor shifts or variations in input signals. Given that ECG signals can exhibit slight differences in feature timing across different heartbeats or patients, this robustness proves highly beneficial [45]. Recent investigations have highlighted the proficiency of 1D CNNs across a spectrum of ECG analysis tasks such as detecting arrhythmias, diagnosing myocardial infarctions, classifying heartbeats, and identifying sleep apnea events. Often outperforming traditional machine learning methods, these studies underscore the significant promise that 1D CNNs hold for improving diagnostic accuracy and efficiency within clinical settings [46], [47], [48], [49].

The application of Convolutional Neural Networks, particularly one-dimensional CNNs, in electrocardiogram analysis has achieved considerable success. Nevertheless, numerous obstacles persist that researchers are actively attempting to tackle. One major issue is the scarcity of data, which is particularly problematic for rare medical conditions. Deep learning models typically demand large labeled datasets for effective training; thus, strategies such as data augmentation and transfer learning are being examined to address this limitation [50]. Another significant challenge lies in the interpretability of CNN models. This concern is especially pronounced in medical contexts where comprehending the rationale behind a model's predictions is critical for clinical validation. Researchers are focused on developing explainable AI techniques tailored specifically for healthcare applications involving CNNs. These efforts aim to illuminate the features and patterns that neural networks leverage when

making decisions. The robustness of CNNs among various types of noise inherent in ECG signals remains an area under examination. Methods like adaptive filtering and noise-aware training are being explored to enhance model performance under real-world conditions replete with inconsistencies and disturbances in signal quality [51]. Furthermore, integrating CNNs with other forms of data such as patient histories or laboratory results promises a more holistic approach to cardiovascular risk assessment. Such integration could significantly enrich predictive capabilities by considering a broader array of factors contributing to cardiovascular health [52]. As these challenges are addressed and the technology continues to evolve, CNNs, especially 1D CNNs, are likely to play an increasingly important role in automated ECG interpretation and cardiovascular disease diagnosis, potentially revolutionizing clinical practice and improving patient outcomes [53].

1.3.2. Long-Short Term Memory(LSTM)

Long Short-Term Memory (LSTM) networks have innovated the field of sequence modeling and have become a cornerstone in various applications of deep learning to sequential data problems. Introduced by Hochreiter and Schmidhuber in 1997 [54], LSTMs were designed to address the vanishing gradient problem inherent in traditional Recurrent Neural Networks (RNNs). This issue, which restricts the ability of RNNs to learn long-term dependencies, occurs when gradients are propagated backwards through time, often resulting in their exponential decay. LSTMs overcome this limitation through a sophisticated architecture that allows the network to selectively remember or forget information over extended sequences.

At the heart of an LSTM unit lies a complex structure comprising gates and a cell state. This architecture enables LSTMs to regulate the flow of information through the network with remarkable efficiency. The primary components of an LSTM include an input gate, a forget gate, an output gate, and a cell state. These elements work in concert to manage the storage, modification, and output of information as it traverses the network.

The mathematical formulation of an LSTM can be expressed through a series of equations that describe the operations performed at each time step. Let us denote the

input at time t as x_t , the hidden state as h_t , and the cell state as C_t . The LSTM updates for a given time step t are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1.12)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1.13)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (1.14)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (1.15)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (1.16)$$

$$h_t = o_t \odot \tanh(C_t) \quad (1.17)$$

In these equations, σ represents the sigmoid function, $\tanh$ is the hyperbolic tangent function, $\odot$ denotes element-wise multiplication, W are weight matrices, and b are bias vectors. The forget gate f_t (Equation 1.12) determines what information should be discarded from the previous cell state. The input gate i_t (Equation 1.13) decides which values from the input should be used to modify the cell state. Equation 1.14 creates a vector of new candidate values $\tilde{C}_t$ that could be added to the cell state. The cell state C_t is then updated in Equation 1.15 by forgetting old information and adding new information. The output gate o_t (Equation 1.16) controls what parts of the cell state should be output. Finally, the hidden state h_t is computed in Equation 1.6 as a filtered version of the cell state.

This complex structure allows LSTMs to maintain and propagate relevant information over long sequences, effectively mitigating the vanishing gradient problem. The cell state serves as a memory channel that runs through the entire chain, with only minor linear interactions. This design enables LSTMs to capture long-term dependencies by allowing relevant information to flow unchanged through the cell state while selectively updating and outputting information through its gating mechanism.

The effectiveness of LSTMs in handling long-term dependencies has made them particularly valuable in a wide range of applications. In natural language processing, LSTMs have been successfully employed for tasks such as language modeling, machine translation, and sentiment analysis. Graves et al. demonstrated the power of LSTMs in speech recognition, achieving state-of-the-art results on the TIMIT

phoneme recognition benchmark [55]. In the domain of time series prediction, LSTMs have shown remarkable performance in forecasting financial data, weather patterns, and energy consumption.

The training process for LSTMs typically involves backpropagation through time (BPTT), a technique that unfolds the recurrent network over multiple time steps and applies the standard backpropagation algorithm. However, the computational complexity of BPTT for long sequences has led to the development of various optimizations. Truncated BPTT, for instance, limits the number of time steps over which gradients are propagated, trading off some long-term dependency learning for computational efficiency.

One of the key advantages of LSTMs is their ability to handle the exploding and vanishing gradient problems that plague standard RNNs. The gating mechanism allows gradients to flow more easily through the network, even over many time steps. This is particularly important for tasks that require the model to remember information for long periods, such as language translation or video captioning.

To further understand the dynamics of LSTMs, we can examine the role of each gate in more detail:

1. **Forget Gate:** The forget gate determines what information from the previous cell state should be discarded. It takes the previous hidden state h_{t-1} and the current input x_t , passes them through a sigmoid function, and produces a value between 0 and 1 for each number in the cell state. A value closer to 1 means "keep this information," while a value closer to 0 means "forget this information."
2. **Input Gate:** The input gate decides what new information should be stored in the cell state. It consists of two parts: a sigmoid layer that determines which values we'll update, and a tanh layer that creates a vector of new candidate values that could be added to the state.
3. **Cell State Update:** The cell state is updated by first multiplying the old state by the forget vector (forgetting things we decided to forget earlier), then adding the product of the input gate and the new candidate values. This step allows the LSTM to selectively update its internal memory.

4. **Output Gate:** The output gate decides what parts of the cell state we're going to output. It starts with a sigmoid layer that determines what parts of the cell state we're going to output, then multiplies this by a tanh of the cell state to produce the final output.

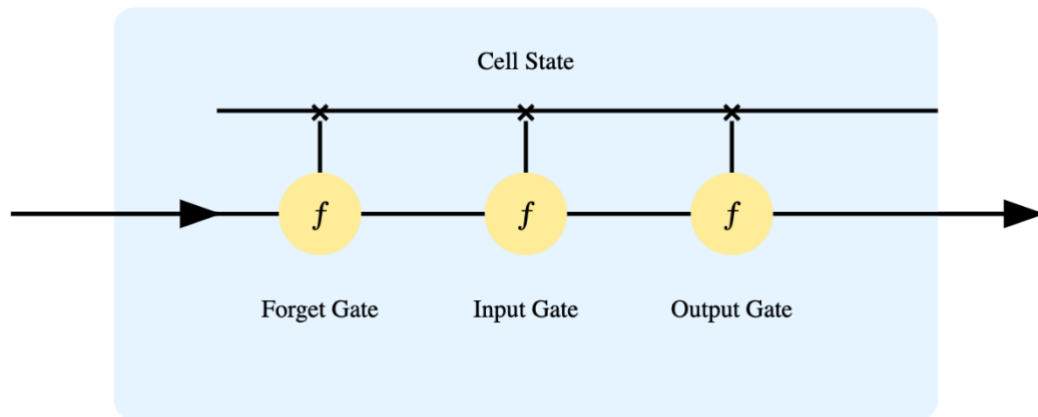


Figure 1.12 LSTM Diagram.

The combination of these gates allows LSTMs to perform complex operations on the input sequence. For example, they can learn to recognize important features in the input and store them for later use, or to forget irrelevant information that might otherwise interfere with the task at hand.

LSTMs have been particularly successful in tasks that require modeling long-term dependencies. In natural language processing, for instance, LSTMs can maintain context over long sentences or even paragraphs, allowing them to perform well in tasks like machine translation, where the meaning of a word often depends on words that appeared much earlier in the sentence.

In time series analysis, LSTMs have shown remarkable ability to capture complex temporal patterns. They can learn to recognize trends, seasonality, and even rare events in sequential data. This has made them valuable tools in fields such as financial forecasting, where they can analyze historical price data to predict future market movements, or in anomaly detection systems, where they can learn to recognize unusual patterns in streaming data.

Despite their power, LSTMs are not without limitations. The increased complexity of LSTM units compared to simple RNN cells results in higher computational

requirements and potentially longer training times. Additionally, the sequential nature of LSTMs can pose challenges for parallelization, which is increasingly important in modern deep learning frameworks.

To address some of these limitations, researchers have proposed numerous variations of the basic LSTM architecture. One notable variation is the introduction of peephole connections by Gers and Schmidhuber [56]. Peephole connections allow the gates to look at the cell state, providing them with more contextual information for their decisions. The mathematical formulation for an LSTM with peephole connections modifies Equations 1.1, 1.2, and 1.5 as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + W_{f_c} \odot C_{t-1} + b_f) \quad (1.18)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + W_{i_c} \odot C_{t-1} + b_i) \quad (1.19)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + W_{o_c} \odot C_t + b_o) \quad (1.20)$$

Here, W_{f_c} , W_{i_c} , and W_{o_c} are diagonal weight matrices for the peephole connections. These connections allow the gates to have more direct access to the cell state, potentially improving the network's ability to learn precise timing and counting tasks.

Another significant variation is the Gated Recurrent Unit (GRU), introduced by Cho et al [57]. GRUs simplify the LSTM architecture by combining the forget and input gates into a single "update gate" and merging the cell state and hidden state. The GRU update equations are:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (1.21)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (1.22)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \quad (1.23)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (1.24)$$

Here, z_t is the update gate, r_t is the reset gate, and $\tilde{h}_t$ is the candidate activation. GRUs have fewer parameters than LSTMs, which can make them faster to train and less prone to overfitting on smaller datasets. In practice, GRUs have been found to perform comparably to LSTMs on many tasks, while being computationally more efficient.

One area where LSTMs continue to excel is in processing variable-length sequences. Unlike transformer models, which typically require fixed-length inputs (often achieved through padding), LSTMs can naturally handle sequences of different lengths. This makes them particularly useful in scenarios such as speech recognition, where the input sequences can vary significantly in length.

LSTMs have also found applications in multimodal learning tasks, where they can be used to process and integrate information from different modalities. For example, in video captioning tasks, LSTMs can be used to process both the visual features extracted from video frames and the textual features of the generated captions, allowing the model to generate coherent and contextually relevant descriptions of video content.

In the field of music generation, LSTMs have been used to create models that can compose music in various styles. By training on large datasets of musical scores, these models can learn to capture complex musical structures and generate new compositions that adhere to the learned style.

Another interesting application of LSTMs is in predictive maintenance for industrial equipment. By analyzing time series data from sensors, LSTM models can learn to predict when a piece of equipment is likely to fail, allowing for proactive maintenance and reducing costly downtime.

As research in deep learning continues to evolve, new architectures and techniques are constantly being developed. However, the fundamental principles behind LSTMs – the ability to selectively remember or forget information over long sequences – remain highly relevant. Many modern architectures, including transformers with their self-attention mechanisms, can be seen as building upon and extending these principles.

In conclusion, Long Short-Term Memory networks represent a powerful and flexible approach to sequence modeling, offering a solution to the long-standing challenge of capturing long-term dependencies in sequential data. Their sophisticated gating mechanism and cell state design enable them to selectively retain and propagate information over extended time steps, making them an essential tool in various applications of deep learning to sequential data problems. While newer architectures have emerged that outperform LSTMs in certain tasks, the fundamental insights provided by LSTMs continue to influence the development of sequential models. As

the field of deep learning continues to advance, it is likely that the principles underlying LSTMs will remain relevant, continuing to inspire new approaches to handling complex sequential data.

1.3.3. Convolutional Neural Networks(CRNN)

Convolutional Recurrent Neural Networks, commonly known as CRNNs, represent an advanced category of neural models that elegantly combine the benefits found in Convolutional Neural Networks with those in Recurrent Neural Networks. They were popularized by Baidu and collaborators back in 2015 [58]. This blend allows CRNNs to skillfully handle and interpret information that presents both spatial characteristics and temporal features. These qualities make them particularly suitable for numerous applications where this kind of data is widespread. One significant advantage tied to CRNNs is their ability to perform all-encompassing learning from beginning to end, which simplifies the training procedure and facilitates efficient model application for complicated assignments. By integrating CNNs' capacity for identifying elaborate patterns within space through convolutional mechanisms with RNNs' expertise in recognizing sequential relationships via looping formations, CRNNs are capable of executing complex tasks. These combined abilities are crucial when dealing with understanding spoken language, analyzing written content, or examining video sequences—fields where spatial and temporal aspects play key roles in ensuring accurate outcomes.

The architecture of a CRNN typically consists of three main components, each serving a distinct purpose:

1. Convolutional Layers:

These layers are designed to extract spatial features from the input data. Through a series of convolutional operations, the network identifies local patterns and spatial hierarchies. Convolutional layers are often followed by activation functions such as ReLU (Rectified Linear Unit) to introduce non-linearity, enhancing the network's ability to model complex patterns. Pooling layers, such as max pooling or average pooling, are sometimes used to downsample the feature maps, reducing the computational load and focusing on the most salient features.

2. Recurrent Layers:

Positioned after the convolutional layers, recurrent layers are responsible for modeling temporal dependencies within the data. Long Short-Term Memory (LSTM) units or Gated Recurrent Units (GRUs) are commonly employed due to their ability to capture long-range dependencies and mitigate issues like the vanishing gradient problem. These layers enable the network to maintain contextual information over sequences, which is crucial for understanding temporal patterns in data such as audio signals or video frames.

3. Fully Connected Layers:

The final stage of the CRNN architecture involves fully connected layers that perform high-level reasoning and decision-making based on the features extracted by the previous layers. The output from the recurrent layers is typically flattened and passed through these layers, which culminate in the network's prediction or classification output. These layers are crucial for translating the learned features into actionable insights or decisions.

CRNNs have stepped into diverse fields, illustrating their adaptability. Within the sphere of recognizing spoken language, these systems excel with audio inputs, leveraging convolutional elements to extract frequency-based characteristics and employing recurrent parts to capture temporal features. This method improves the accuracy of detecting spoken words. Similarly, in tasks involving text identification like Optical Character Recognition(OCR), CRNNs manage various character sizes and shapes effectively. Their distinctive combination of spatial understanding and sequential processing is advantageous in analyzing video content as well, enabling them to examine series of frames and identify patterns or activities over time with ease.

Convolutional Recurrent Neural Networks present an abundance of advantages, yet they bring their own challenges. A significant challenge involves the large requirement for computational capacity, which might limit implementation on gadgets with constrained capabilities. Additionally, training deep layers that depend on feedback loops can become complicated due to obstacles such as the vanishing gradient problem, potentially inhibiting advancement in training. In response to these difficulties, scholars are earnestly involved in enhancing CRNN architectures by

developing new techniques for learning and exploring groundbreaking applications in emerging fields like autonomous vehicles and real-time surveillance systems.

In conclusion, Convolutional Recurrent Neural Networks represent a significant advancement in the field of deep learning, providing a versatile and powerful framework for handling complex data with both spatial and temporal dimensions. Their ability to integrate feature extraction and sequence modeling into a single architecture makes them an invaluable tool in numerous applications, pushing the boundaries of what is achievable in machine learning and artificial intelligence.

1.3.4. Transformer and Attention

Transformers, introduced by Vaswani et al. in 2017 [59], signify a significant change in deep learning structures, especially for natural language processing (NLP). The primary innovation of Transformers is their use of the attention mechanism. This technique allows the model to focus on various segments of the input dynamically while generating each part of the output. Previous models like recurrent neural networks and long short-term memory networks faced challenges with handling extended dependencies and parallel computation [60].

The concept of attention was first brought into prominence by Bahdanau et al. [61] for neural machine translation. It has since become foundational in contemporary deep learning models. This mechanism enables models to prioritize relevant parts of an input, mirroring human cognitive functions. In Transformer architecture, self-attention expands this idea further; it permits every element within a sequence to consider all other elements, thereby capturing both detailed and broad patterns.

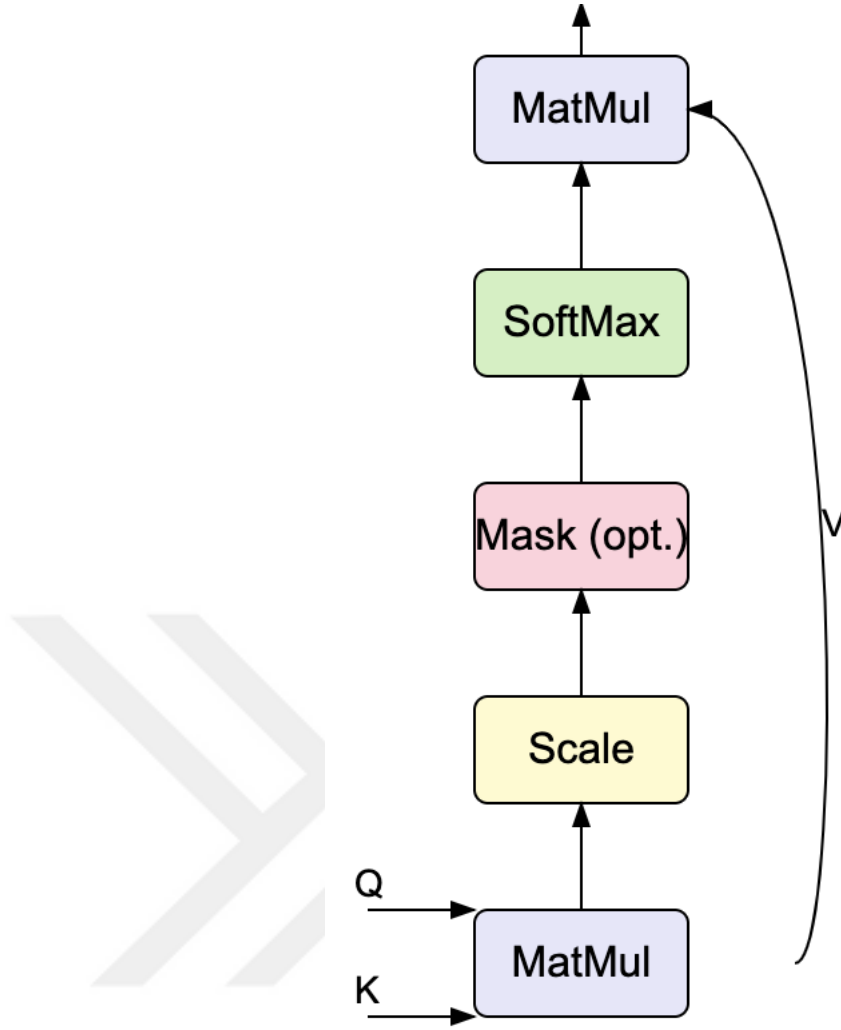


Figure 1.13 Scaled Dot-Product Attention.

A pivotal element within Transformer architectures is the scaled dot-product attention mechanism. This approach determines attention weights through a process involving the dot product of query and key matrices, followed by scaling to avoid overly large values. Subsequently, a softmax function is applied. These computed weights facilitate a weighted aggregation of value matrices.

Expressed mathematically:

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{d_k})V \quad (1.25)$$

where Q, K , and V represent the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors [59]. The term QK^T computes the similarity between each pair of elements in the sequence, while the scaling factor $\sqrt{d_k}$ prevents

the dot products from growing too large in magnitude. The softmax function normalizes the attention weights, and the final multiplication with V produces the output.

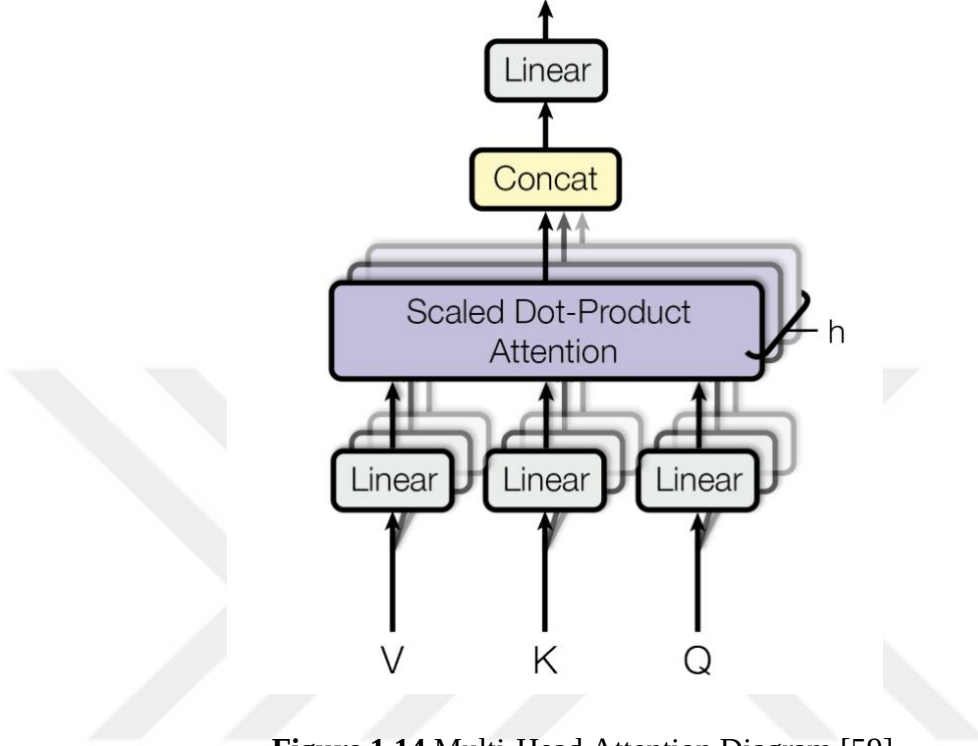


Figure 1.14 Multi-Head Attention Diagram [59].

Transformers typically employ multi-head attention, which allows the model to jointly attend to information from different representation subspaces:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (1.26)$$

$$\text{where } head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (1.27)$$

Here, W_i^Q, W_i^K, W_i^V , and W^O are learnable parameter matrices [59]. Multi-head attention enables the model to capture different types of relationships within the data simultaneously, enhancing its representational power. This mechanism can be thought of as allowing the model to attend to different aspects of the input in parallel, similar to how humans can focus on various features of a scene simultaneously.

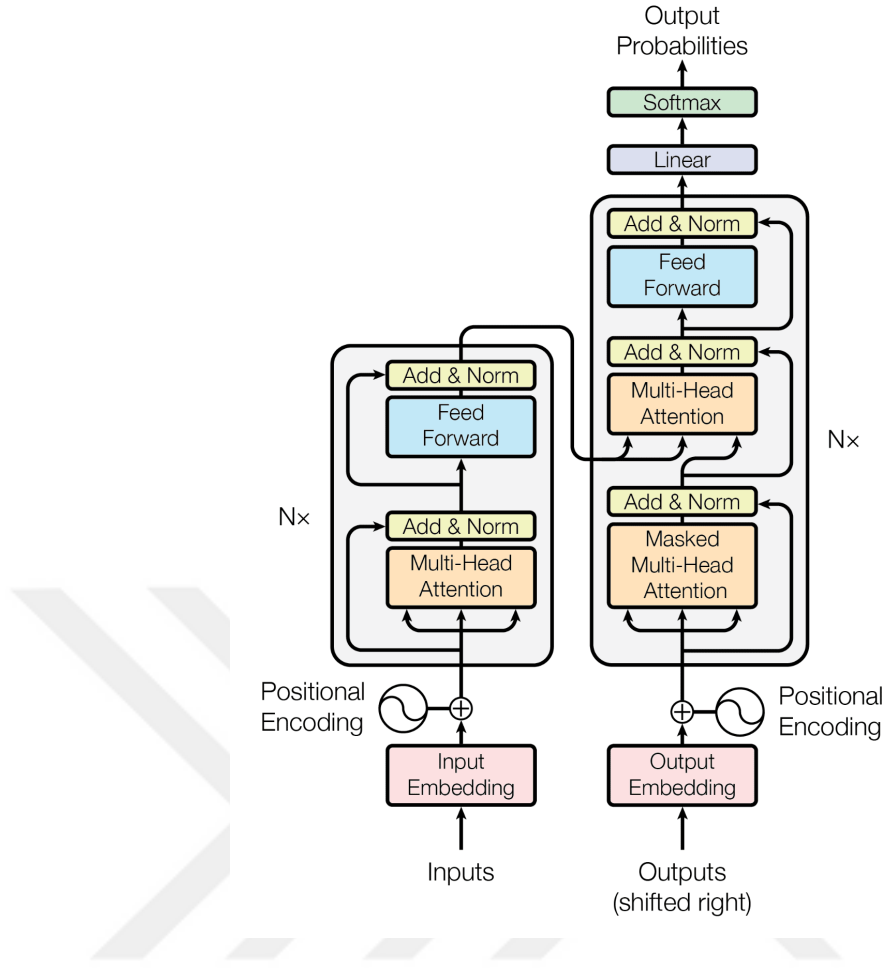


Figure 1.15 Transformer Architecture [59].

In addition to self-attention, Transformers incorporate several other key components such as Position-wise Feed-Forward Networks, Layer Normalization, Residual Connections, and Positional Encoding. Position-wise feed-forward networks apply two linear transformations with a ReLU activation in between:

$$FFN(x) = \max(0, xW_1 + b_1) W_2 + b_2 \quad (1.28)$$

where W_1, W_2, b_1 , and b_2 are learnable parameters [59]. This component introduces non-linearity and allows the model to capture more complex patterns. Layer normalization applied after each sub-layer to stabilize the activations and improve training dynamics [62]. Residual connections are another pivotal feature incorporated into these models. By facilitating gradient flow throughout the network, they support effective training even in deeper architectures [63]. This mechanism helps overcome common issues related to vanishing or exploding gradients during backpropagation.

Since Transformers process input elements in parallel, they lack an inherent notion of sequence order. Positional encodings are added to the input embeddings to inject information about the relative or absolute position of the tokens in the sequence. The original Transformer paper proposed sinusoidal positional encodings:

$$PE(pos, 2i) = \sin(pos/10000^{(2i/d_{model})}) \quad (1.29)$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{(2i/d_{model})}) \quad (1.30)$$

Where pos is the position and i is the dimension [59].

These encodings have the advantage of allowing the model to easily attend to relative positions. The training process for Transformers, similar to other deep learning models, typically involves backpropagation and gradient-based optimization. The loss function used depends on the specific task, with cross-entropy being common for classification tasks and mean squared error for regression tasks. The gradients are computed using the chain rule:

$$\partial L / \partial \theta = \partial L / \partial y \cdot \partial y / \partial a \cdot \partial a / \partial \theta \quad (1.31)$$

where L is the loss function, y is the output, a represents the attention weights, and θ are the model parameters. Optimization algorithms like Adam or AdamW are often used to update the parameters:

$$\theta_t = \theta_{t-1} - \eta \cdot m_t / (\sqrt{v_t} + \varepsilon) \quad (1.32)$$

where η is the learning rate, m_t and v_t are the first and second moments of the gradients, and ε is a small constant for numerical stability [64].

While Transformers were initially designed for NLP tasks, their ability to capture long-range dependencies and process sequences in parallel has made them attractive for other domains, including time series analysis and processing of 1D signals. In these contexts, each time step or segment of the signal can be treated as a token, analogous to words in NLP tasks. The self-attention mechanism allows the model to focus on relevant parts of the signal when making predictions, potentially capturing important relationships between different segments of the input.

Transformers present several compelling advantages for time series and 1D signal processing tasks. Through their attention mechanism, these models excel at modeling relationships between distant parts of the signal, making them particularly effective for data with long-term patterns or periodicities. Unlike Recurrent Neural Networks' sequential processing approach, Transformers can process entire input sequences simultaneously, significantly improving training efficiency and inference speeds—a crucial benefit for extensive datasets or real-time applications. Their inherent flexibility in handling variable-length sequences without requiring padding or truncation methods proves especially valuable when working with signals of varying durations. The combination of multi-head attention and hierarchical structures enables Transformers to simultaneously detect patterns across different temporal scales, making them well-suited for data containing both short-term fluctuations and long-term trends. Additionally, the transparency offered by attention weights provides valuable insights into the model's focus areas during prediction stages, facilitating understanding of the decision-making process and highlighting critical signal segments. Furthermore, the ability to leverage pre-trained Transformer models through fine-tuning for specific 1D signal processing tasks offers a powerful approach to improve performance, particularly in scenarios with limited data availability.

Despite these advantages, applying Transformers to time series or 1D signal processing tasks presents several significant challenges. The quadratic computational complexity of self-attention with respect to sequence length poses a particular concern for long signals, though researchers have proposed various solutions including sparse attention patterns, low-rank approximations, and hierarchical architectures [65] [66]. Unlike RNNs, Transformers lack inherent biases toward temporal modeling, making positional encodings crucial; however, designing effective encoding schemes remains an active research area, especially given the diverse nature of 1D signals encountered, from audio and physiological readings to financial series. The challenge of data scarcity further compounds these issues, as training large-scale transformers typically requires substantial datasets, necessitating the implementation of regularization techniques and robust augmentation strategies. Moreover, the distinctive attributes of different types of 1D signals often require customized modifications to standard transformer architectures to better align with their specific characteristics, ensuring

optimal results across diverse applications in areas such as audio analytics, physiological monitoring, and financial forecasting scenarios.

To address these challenges and adapt Transformers for 1D signal processing, several variants and modifications have been proposed. By combining convolutional layers with self-attention mechanisms [67], the model can effectively capture both local nuances and broader dependencies within signals, enhancing the accuracy of feature extraction from input data. Further developments include mechanisms allowing the model to learn the appropriate attention span for different parts of the input signal [68], potentially improving efficiency and performance. The integration of Fourier Transform layers into Transformer architecture [69] offers an innovative approach to capturing frequency-domain details within signals, enriching the model's ability to understand and manipulate intricate patterns that manifest across diverse frequencies. Additionally, employing multiple Transformer layers each attuned to distinct temporal resolutions [70] facilitates recognition of patterns over varying time spans, ensuring that both transient and long-term trends are adequately captured and analyzed. The adaptation of Transformer architectures for continuous-time signal management [71] introduces significant flexibility in modeling irregularly sampled time series, further expanding their applicability in various domains.

As research advances in this field, we are prepared to witness further refinements and enhancements of Transformer architectures across various sequential data processing applications. The adaptability and efficacy of the attention mechanism, paired with ongoing efforts to overcome its constraints, position Transformers as a powerful solution for addressing complex challenges in time series analysis and one-dimensional signal processing.

In conclusion, transformers and their attention mechanisms illustrate a dynamic approach to sequence modeling that has already demonstrated substantial promise in natural language processing. Their application is expanding into other areas such as time series analysis and one-dimensional signal interpretation. As these models progress, they offer transformative potential in how we tackle sequential data analysis tasks, enabling new insights into complex temporal patterns and interrelations within an array of 1D signals.

2. RELATED WORK

The development of automatic ECG classification has witnessed notable progress in recent years, as researchers have pursued diverse strategies to enhance accuracy and clinical relevance. This review seeks to outline the progression of ECG classification methodologies, tracing their journey from traditional machine learning paradigms to contemporary deep learning techniques. Historically, this field has faced substantial challenges due to the complex nature of cardiac signals and individual variability. Early automated classification efforts relied significantly on handcrafted feature extraction methods combined with conventional machine learning models. While these foundational approaches provided a starting point, they often grappled with issues related to high-dimensionality and the temporal characteristics inherent in ECG data.

A landmark contribution came from Laguna et al. [72], who developed an algorithm for QRS complex detection by differentiating ECG signals. This pioneering work established a foundation for subsequent research focused on extracting meaningful features from cardiac signals. Building on this, Thakor et al.[73] introduced adaptive filtering methods aimed at reducing noise within ECG recordings an essential step that significantly improved the quality of input data fed into classification algorithms.

As machine learning methodologies evolved, researchers delved into more refined techniques. For instance, S. Celin and associates [74] worked with a 2-channel MIT-BIH Arrhythmia dataset, training a support vector machine (SVM) that achieved an accuracy rate of 87.5%. This study showcased the efficacy of SVMs in managing the complex feature space inherent in ECG signals. Similarly, Dias et al. [75] utilized the same dataset but chose a linear discriminant classifier for their analysis, resulting in an accuracy of 90.22%. These investigations underscored the significance of selecting appropriate features and classifiers to attain superior performance.

Moving beyond traditional methods, Acharya and team [76] combined nonlinear features with support vector machines for classifying cardiac health states. Their work achieved a notable accuracy of 95.85% when distinguishing between normal, pre-event, and event conditions. The research stood out due to its innovative use of higher-

order spectra and cumulants as features elements that effectively differentiated subtle variances within ECG classes.

Ye et al. [77], on the other hand, progressed further by integrating morphological features with independent component analysis alongside an SVM classifier. This strategy reached an exceptional accuracy rate of 99.3% in detecting five arrhythmia types—a feat not just impressive due to its high precision but also for demonstrating how advanced signal processing can be seamlessly combined with machine learning techniques for outstanding results.

The research conducted by Martis and colleagues [78] marked a key advancement in the application of principal component analysis for reducing data dimensions, subsequently employing diverse classifiers like artificial neural networks and support vector machines. Their innovative approach achieved an impressive 99.52% classification accuracy across five distinct ECG beat types. This investigation was particularly significant due to its extensive evaluation of various feature extraction methods coupled with different classification strategies, offering profound insights into the comparative efficacy of these techniques, strengths of various approaches.

As advancements unfolded, scholars delved into sophisticated machine learning methods. Oster et al. [79] proposed an innovative technique employing semi-supervised learning for ECG classification. Their approach cleverly utilized both labeled and unlabeled data, offering a promising solution in scenarios with limited labeled datasets. This development was particularly significant given the difficulty of amassing extensive, labeled ECG datasets in medical environments.

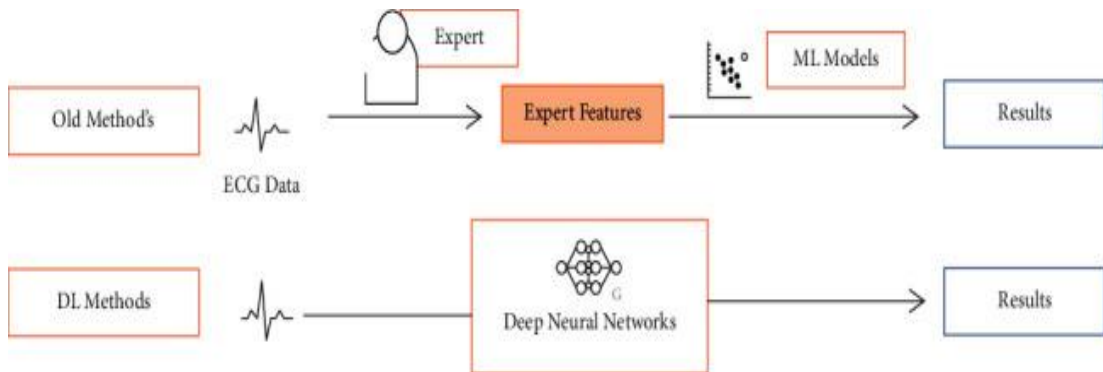


Figure 2.1 Evolution of Machine Learning Methods on ECG [80].

The emergence of deep learning brought a transformative change to ECG classification. Convolutional Neural Networks, initially crafted for image processing tasks, were successfully repurposed for analyzing ECG signals. Hannun et al. [81] made a notable contribution with their 34-layer CNN model trained on a unique single-channel dataset. The model, leveraging residual networks to sustain high depth, distinguished between 12 classes with accuracy surpassing that of seasoned cardiologists. This pioneering work showcased not just impressive precision but also the potential for deep learning models to excel beyond human expertise in specific diagnostic tasks.

Following the achievements with CNNs, Acharya et al.[82] introduced a 9-layer CNN designed for automated myocardial infarction detection, achieving an impressive accuracy rate of 93.53% using raw ECG signals directly without manual feature extraction. This study underscored the capacity of deep learning models to autonomously identify critical features from raw data effectively. Similarly, Kachuee et al. [83] presented a deep residual network for ECG classification that demonstrated exceptional performance across various datasets and set new benchmarks on the MIT-BIH Arrhythmia dataset through its use of residual connections that enabled deeper network structures while addressing common challenges like vanishing gradients in neural networks.

Researchers have focused on understanding the time-sensitive nature of ECG signals, leading to the creation of architectures that can better capture these temporal patterns. One such development is the integration of Convolutional Neural Networks with Recurrent Neural Networks. A significant contribution in this area came from Oh et al. [84], who combined CNNs and long short-term memory networks for arrhythmia diagnosis automation. Their model reached an F1 score of 0.982 across 17 rhythm categories, showcasing how hybrid architectures excel in identifying intricate patterns within ECG data.

Expanding upon this hybrid approach, Warrick and Homsy [85] developed a Convolutional Recurrent Neural Network (CRNN) for ECG classification. This architecture demonstrated significant improvement over conventional machine learning techniques by effectively blending convolutional layers for spatial feature extraction with recurrent layers dedicated to modeling temporal dependencies. Such a

strategy proved exceptionally proficient in handling the dynamic characteristics of ECG signals while preserving their spatial consistency.

Another notable advancement in this field has been the application of the Attention Mechanism, initially devised for natural language processing as detailed by Vaswani et al. [86] in their pivotal work "Attention is All You Need." The mechanism's ability to prioritize relevant parts of input sequences makes it particularly advantageous for lengthy ECG recordings analysis. By focusing on crucial segments, attention-based models enhance precision and reliability when examining extended ECG datasets.

Yan et al. [87] pioneered the integration of transformer models with ECG classification by merging these architectures with temporal features, yielding impressive results compared to traditional CNN and RNN frameworks. Their approach excelled in recognizing long-range dependencies within ECG sequences, opening ways for more comprehensive analysis, particularly in detecting arrhythmias over extended periods. Building on this foundation, Li et al. [88] advanced a hierarchical transformer network customized for multi-lead ECG classification, achieving top-tier performance on the PTB-XL dataset. This hierarchical strategy allowed effective handling of multi-lead data, showcasing how adaptable transformers can be for complex medical signal analyses. Notably, their work underscored efficiency while dealing with high-dimensional 12-lead ECG inputs. Further demonstrating versatility, Che et al. [89] developed a model capable of managing both single and 12-lead ECG data formats—a crucial feature in diverse clinical environments—and demonstrated the benefits of transfer learning where pre-trained models could be refined across various tasks or datasets.

With increasing advancement in deep learning applications for ECG classification came a growing concern regarding interpretability. The unclear nature of deep neural networks posed significant barriers to their clinical acceptance since unraveling model predictions is essential for healthcare providers' trust and understanding. Addressing this challenge head-on, Raghunath et al. [90] utilized interpretation techniques within deep learning frameworks to detect specific ECG characteristics predictive of left ventricular dysfunction—offering not only high accuracy but also valuable insights into critical aspects of the signals themselves. Similarly focused on transparency and clarity, Strodthoff et al. [91] delved into identifying which segments of the ECG were significant to neural network decisions—blending predictive prowess with an

interpretative angle vital for earning clinicians' confidence and potentially uncovering novel patterns that hold clinical significance.

The application of deep learning to ECG analysis has also led to novel approaches in data augmentation and handling class imbalance, common challenges in medical datasets. Jun et al. [92] proposed a novel data augmentation technique using generative adversarial networks (GANs) to synthesize realistic ECG signals. Their approach not only improved classification performance but also demonstrated the potential of generative models in addressing data scarcity issues in medical signal analysis.

Recent years have seen an increased focus on developing models that can handle multi-modal data, incorporating not just ECG signals but also patient metadata and other clinical information. Mousavi et al. [93] proposed a multi-modal deep learning approach that combines ECG signals with patient demographic data for improved arrhythmia detection. Their model demonstrated superior performance compared to single-modality approaches, highlighting the potential of integrating diverse data sources in ECG analysis.

The challenge of limited labeled data in medical domains has led to increased interest in self-supervised learning techniques. Sarkar and Etemad [94] introduced a self-supervised pre-training method for ECG classification, which significantly improved performance on downstream tasks, especially in scenarios with limited labeled data. This approach could be particularly valuable in overcoming the challenge of obtaining large, labeled ECG datasets in clinical settings. Privacy concerns in healthcare have driven research into federated learning approaches for ECG analysis. Liu et al. [95] demonstrated the feasibility of federated learning for ECG classification, achieving comparable performance to centralized learning while maintaining data privacy. This approach allows for collaborative model training across multiple institutions without sharing raw patient data, addressing a critical concern in medical data analysis.

The integration of domain-specific knowledge into deep learning frameworks for ECG analysis represents a transformative advancement, enhancing both the predictive accuracy and the interpretability of these models. As evidenced in recent research published in BMC Medical Informatics and Decision Making [96], the incorporation of pre-existing medical insights into machine learning algorithms significantly augments their diagnostic precision and aligns computational outputs with clinical

reasoning. This methodological synergy not only fortifies model robustness but also ensures that the analytical processes resonate with established clinical paradigms, thereby amplifying their applicability and reliability in medical diagnostics.

Recent research has ventured into employing ECG signals for applications beyond traditional arrhythmia detection. A notable study by Attia et al. [97] developed a deep learning model to identify asymptomatic left ventricular dysfunction from standard ECGs. This work highlights the potential of artificial intelligence in uncovering subtle ECG patterns that could signal cardiac conditions, often missed by conventional analysis.

Deep learning's application in ECG analysis has also found relevance in wearable devices and continuous monitoring scenarios. Hannun et al. [98] created a deep neural network capable of detecting arrhythmias using single-lead ECG data gathered from wearable technology. Their model matched cardiologist-level performance across various arrhythmias, showcasing AI's capability to facilitate widespread and uninterrupted cardiac surveillance. As advancements continue, there's increasing interest in adaptive models that cater to individual patients while managing concept drift in ECG signals. For instance, Kwon et al.'s [99] proposed adaptive deep learning model adjusts predictions based on specific patient data, addressing inter-patient variability challenges and enhancing long-term monitoring accuracy.

Moreover, the fusion of ECG analysis with other elements of artificial intelligence is yielding promising outcomes. Porumb et al., [100] for example, investigated using deep learning for biometric identification through ECG signals, demonstrating its feasibility as a unique identifier within healthcare settings.

In another innovative approach focusing on explainable AI, Zhang et al.'s [101] recent work revolves around attention-based models that offer both high classification accuracy and visual explanations for their predictions. By employing attention maps to highlight crucial parts of the ECG signal relevant to each prediction, this methodology provides clinicians with deeper insights into the model's reasoning process.

Integrating reinforcement learning into electrocardiogram analysis shows potential for enhancing long-term patient care strategies. Nemati and colleagues [102] introduced a framework that personalizes the management of atrial fibrillation, utilizing ECG

alongside other clinical data to inform treatment choices. This advancement signals a shift towards more adaptive and individualized methods in cardiac healthcare.

Looking ahead, numerous challenges and prospects in ECG analysis continue to surface. One significant area involves developing models capable of efficiently processing multi-lead ECG data while integrating it with other physiological metrics and medical records. Another research avenue focuses on creating algorithms that can recognize and adjust to evolving ECG patterns over time, factoring in aging or disease progression. In parallel, there's growing interest in crafting lightweight models designed for use on edge devices. Such innovations would enable real-time ECG monitoring via wearable technology or mobile health applications. Research exemplified by Yao et al. [103] highlights efforts aimed at making advanced ECG analytics accessible even in settings with limited resources, potentially transforming how personal health is monitored.

Yan et al. [104] showcased a fresh perspective on the use of the Transformer setup for classifying ECG heartbeats, traditionally used in language-related tasks, they cleverly applied it to analyzing ECG signals. The researchers selected just the encoder portion of this architecture and swapped dropout elements with batch normalization, aligning better with their classification goals. They skillfully combined crafted temporal features, such as RR intervals, with insights gained by the Transformer itself to effectively grasp both shape and rhythm aspects in ECG data. This blending of time-based details and self-focus methods showed exciting outcomes. Their model performed exceptionally well on the MIT-BIH arrhythmia dataset, achieving impressive marks: 99.62% accuracy, 95.09% precision, and 94.12% sensitivity initially; even surpassing these figures on a more balanced set later on. Through detailed comparisons against leading techniques and exploring different setups within their model, they convincingly illustrated its strength in tackling this task successfully—paving paths for deep learning uses within medical time-series analysis like detecting arrhythmias through ECGs.

The progress in automatic ECG classification has transitioned from basic machine learning methods to advanced deep learning architectures. With the integration of attention mechanisms and transformer models, researchers can now identify sophisticated patterns within ECG data more effectively. Nevertheless, challenges

persist, particularly regarding the clarity of model predictions, stability under varied conditions, and real-world clinical use.

Our research aims to tackle these challenges by constructing high-performing models that also emphasize transparency and the relevance of different ECG channels. By experimenting with three distinct models—CNN, CNN enhanced with attention mechanisms, and transformers—on 12-channel ECG datasets, we seek to evaluate their performances comprehensively. This investigation will reveal both their capabilities and constraints. Additionally, our innovative use of Grad-CAM techniques along with methods inspired by adversarial attacks provides insights into channel significance which contributes to developing AI systems that clinicians can trust.

This multi-faceted strategy holds promise for advancing toward more dependable AI tools in cardiology. As developments continue in analyzing ECG signals through sophisticated machine learning techniques combined with domain knowledge, it becomes evident that such integrated approaches are essential for creating systems that significantly aid clinical decision-making processes in heart-related care.

3. METHOD AND METHODOLOGY

3.1. Dataset

3.1.1. PTB Diagnostic ECG Database

The PTB Diagnostic ECG Database, version 1.0.0, released on September 25, 2004, stands as a key asset in the domain of electrocardiography and cardiovascular studies [105]. This comprehensive dataset houses an array of 549 detailed ECG recordings from a varied group of 290 individuals aged between 17 and 87 years, with an average age hovering around the late fifties. Among these participants are 209 males (averaging mid-fifties) and 81 females (averaging early sixties), with age details missing for a small fraction consisting mainly of men. Each participant contributes one to five distinct ECG files, offering a wealth of heart-related data for thorough investigation. A key feature is its outstanding recording fidelity achieved through advanced equipment capable of processing multiple inputs simultaneously. Every file captures fifteen signals concurrently: traditional twelve leads combined with three additional Frank leads which expand diagnostic capabilities. These readings are transformed at a rapid rate of 1000 samples per second, providing researchers with the accuracy needed to closely examine both rapid cardiac activities and subtle waveform details. The sensitivity further benefits from superior amplitude conversion quality over a generous range allowing for fine-grained analysis essential in identifying potentially critical cardiac anomalies that might otherwise elude detection using less refined methods.

The PTB Diagnostic ECG Database is highly regarded for its extensive annotation framework. Each ECG record comes with a comprehensive clinical summary that serves as identifiers for the collection, offering significant details such as age, sex, precise diagnosis, and when applicable, thorough data on medical history, drug treatments, surgical procedures, coronary artery irregularities, ventriculography insights, echocardiography outcomes, and hemodynamic measurements. This wealth of contextual knowledge amplifies the database's relevance in both healthcare research and in creating sophisticated machine learning systems for automated interpretation of

electrocardiograms. The diagnostic categories within this resource are varied and carefully documented across a wide range of heart-related disorders. Among them is a substantial group of 148 individuals suffering from myocardial infarction alongside 52 individuals serving as healthy controls. Other notable categories include cardiomyopathy or heart failure (18 cases), bundle branch block (15 cases), dysrhythmia (14 cases), myocardial hypertrophy (7 cases), valvular heart disease (6 cases), myocarditis (4 instances), along with 4 subjects classified under assorted cardiac anomalies. Such diversity in cardiac issues combined with the inclusion of healthy comparisons renders this dataset exceptionally useful for comparative investigations and crafting dependable diagnostic techniques.

Table 3.1: Subject Diagnoses of PTB Diagnostic ECG Database.

Diagnostic class	Number of subjects
Myocardial infarction	148
Cardiomyopathy/Heart failure	18
Bundle branch block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52

The specifications of the recording gear illuminate the exceptional caliber of the dataset. The PTB prototype device, pivotal in data gathering, is equipped with an input voltage tolerance extending to ± 16 mV and can adjust offset voltages reaching ± 300 mV. It offers an input resistance measuring 100Ω (DC) alongside a bandwidth that

spans from 0 to 1 kHz, which guarantees thorough representation of cardiac electrical patterns. Furthermore, its noise output does not exceed 10 μV peak-to-peak or reaches only 3 μV RMS under conditions where the input is shorted, thus assuring pure signal quality. Beyond this precision in measurement, there exists an innovative ability for real-time monitoring of skin resistance and ambient interference during signal acquisition—providing an added assurance of fidelity throughout data capture procedures. Despite a typical sampling rate being set at 1000 Hz, those managing this resource have mentioned that recordings might be attainable at enhanced rates reaching 10 kHz through specific requests; such flexibility potentially unlocks greater temporal insights for particularized investigative endeavors.

The PTB Diagnostic ECG Database emerges as a noteworthy collection in the field of electrocardiogram datasets, distinguished by its elevated sampling frequency, multi-channel recordings, and extensive patient details. Its broad spectrum of pathological samples makes it a valuable asset for diverse applications such as analyzing heart waveform variations across multiple cardiac disorders, refining automated tools for ECG assessment, probing into fluctuations in heartbeat patterns, and studying minute electrical changes associated with different ailments. It is equally instrumental for training computational models tasked with interpreting ECGs and categorizing cardiovascular diseases. With its finely detailed annotations and high-resolution data capture, this dataset proves exceptionally advantageous for pioneering research on detecting early heart anomalies and identifying unique electrical signals linked to various cardiac conditions.

3.1.2. PTB-XL Database

The PTB-XL collection, published on the ninth of November in 2022 (version 1.0.3), serves as an invaluable repository for ECG data, marking a significant leap forward for heart-related studies and the creation of new computational techniques. This voluminous dataset encompasses an impressive total of 21,799 clinical ECG readings derived from 18,869 distinct individuals, with each test covering a time frame of ten seconds across twelve leads [106]. One important detail is its sampling rate of 500 Hz; this means there are five hundred data points captured per second for every lead channel, facilitating meticulous examination through high-resolution temporal insights into cardiac electrical behavior and intricate waveform structures. With a generous

supply of 5,000 data points available per lead within any given ten-second interval (500 samples/second times ten seconds), it allows scholars to thoroughly study both swift transitions and extended occurrences in heart rhythms.

The 500 Hz sampling rate of the PTB-XL dataset effectively strikes a balance between comprehensive data collection and practicality. It is sufficiently elevated to encompass the entire spectrum of ECG frequencies applicable to clinical diagnosis, generally ranging from 0.05 Hz up to 150 Hz. This particular frequency surpasses the American Heart Association's guidelines for diagnostic ECG devices, which advocate a minimum of 500 samples each second. By selecting this rate, the dataset ensures precise representation of high-frequency elements within an ECG signal, including delicate indentations in the QRS complex or intricate atrial fibrillatory oscillations. Concurrently, it efficiently captures lower frequency aspects critical for evaluating alterations in ST-segment characteristics and T-wave forms.

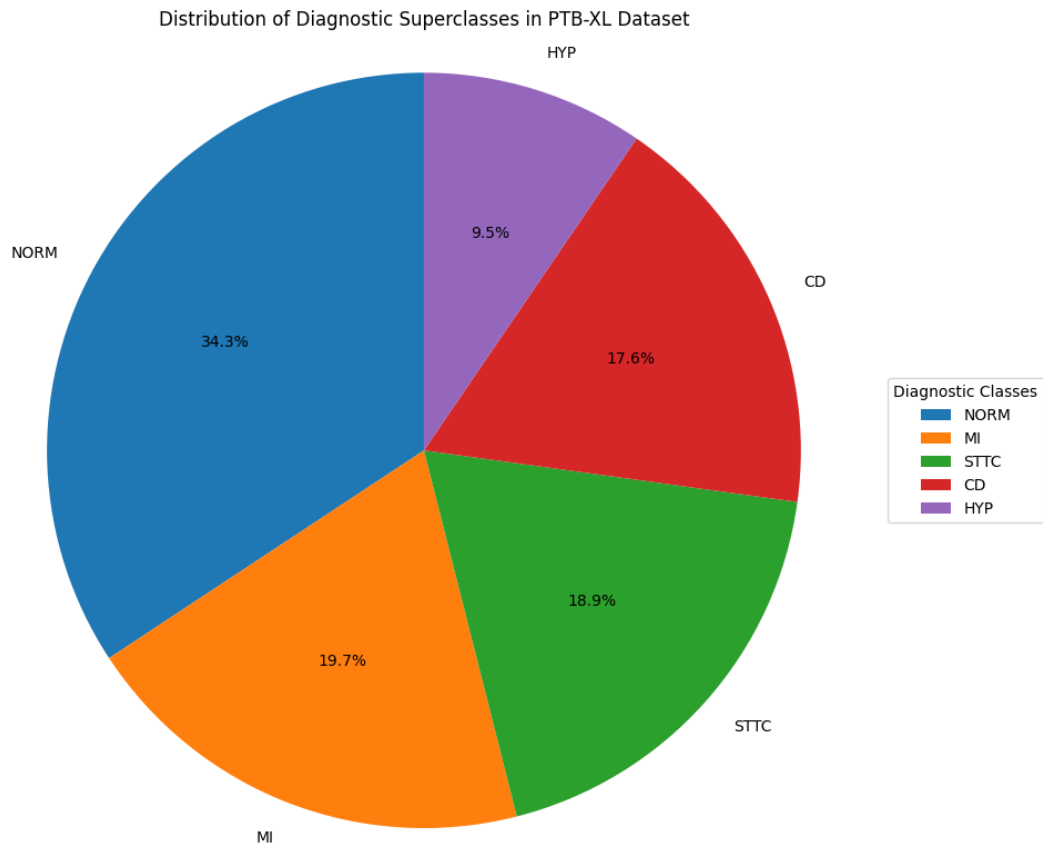


Figure 3.1 Label Distribution of PTX-XL Database.

The dataset's vast coverage is amplified by its detailed annotation framework, with each electrocardiogram record examined and marked by as many as two seasoned heart specialists. These labels adhere to the SCP-ECG guidelines and cover 71 varied ECG statements divided into categories of diagnostic nature, structural forms, and rhythmic patterns. This multifaceted strategy, paired with high-frequency data collection, offers a detailed representation of cardiac issues that frequently highlight the complex character of heart conditions where several irregularities may exist simultaneously.

The PTB-XL collection stands out due to its 500 Hz sampling rate combined with an expansive scale comprising 21,799 ECG records from a wide-ranging group of patients numbering 18,869. With extensive specialist annotations provided, it becomes an indispensable asset for electrocardiography exploration and algorithmic progression. This mix of fine temporal detail and thorough labeling supports investigators in crafting and assessing advanced ECG analytical methods which could pave the way for enhanced precision in automated diagnostics within cardiology practice.

3.1.3. St Petersburg INCART 12-lead Arrhythmia Database

St Petersburg INCART 12-lead Arrhythmia Database, introduced on the first of May in 2008, serves as an fundamental tool for those engaged in electrocardiography and heart rhythm studies. This expansive collection encompasses 75 thoroughly marked recordings, each with a duration of half an hour [107]. These samples are drawn from 32 Holter monitor records of individuals undergoing evaluations related to coronary artery disease. An essential feature is its recording frequency set at 257 cycles per second across all standard ECG leads. This rate effectively captures all pertinent heart signal frequencies while ensuring that data files remain accessible for examination.

Comprising a varied group of participants—17 males and 15 females between the ages of 18 to 80 (average age being roughly late fifties)—this database showcases a multitude of cardiac conditions. It emphasizes cases with electrocardiograms suggestive of inadequate blood supply to the heart muscle, arterial diseases impacting blood flow, electrical conduction irregularities, and other rhythm disturbances. The dataset features instances such as recent or old heart attacks, temporary interruptions in brain blood supply without lasting damage, coronary issues coupled with high blood pressure challenges, extra beats originating above the ventricles (supraventricular

ectopy), chaotic beating due to atrial fibrillation among other disorders affecting the cardiovascular system.

Table 3.2: St Petersburg INCART 12-lead Arrhythmia Database.

Diagnosis	Patients
Acute MI	2
Transient ischemic attack (angina pectoris)	5
Prior MI	4
Coronary artery disease with hypertension	7 (4 with ECGs consistent with left ventricular hypertrophy)
Sinus node dysfunction	1
Supraventricular ectopy	18
Atrial fibrillation or SVTA	3 (2 with paroxysmal AF)
WPW	2
AV block	1
Bundle branch block	3

A prominent attribute of the INCART database lies in its extensive compilation of beat annotations, boasting a collection exceeding 175,000 entries spread throughout all recordings. These annotations commence their life cycle via an automated process before being meticulously revised by human experts, conforming to the established PhysioBank standards for beat identification. This thorough approach to annotation significantly boosts the dataset's trustworthiness, providing a robust foundation for crafting and assessing algorithms that detect irregular heart rhythms.

The ECG signals captured exhibit variability in their recording settings across distinct datasets, spanning from 250 up to 1100 units on an analog-to-digital converter scale

per millivolt. Such diversity permits the inclusion of both minor and major ECG signal features, thus widening the potential applications of this dataset when examining various cardiac waveform shapes. Accompanying each record is comprehensive background information including patient demographics such as age and gender alongside diagnostic data and ECG feature outlines; these components are vital in furnishing essential insights needed for academic exploration.

Comprising seventy-five thirty-minute segments extracted from thirty-two more extended Holter monitoring sessions, this database's arrangement facilitates analysis over both short-term changes in ECG readings as well as extended alterations observed within individual subjects' cardiac activities. Such a configuration proves invaluable for investigating how arrhythmias progress over time and evaluating how consistent or variable specific heart signal characteristics remain across lengthy durations.

In summary, the St Petersburg INCART 12-lead Arrhythmia Database, with its 257 Hz sampling rate, extensive beat annotations, diverse patient population, and comprehensive clinical metadata, serves as an invaluable resource for cardiac electrophysiology research. Its focus on various cardiac abnormalities, particularly arrhythmias and ischemic conditions, makes it especially suited for developing and validating algorithms for automated ECG interpretation. The database's 12-lead format provides a more complete view of cardiac electrical activity, enabling sophisticated analyses and contributing significantly to the advancement of ECG analysis techniques and the development of more accurate diagnostic tools in cardiology.

3.1.4. Merged Dataset

For this research, a broad and varied electrocardiography (ECG) dataset was assembled by incorporating data from three notable ECG databases: the PTB Diagnostic ECG Database (PTB) the PTB-XL ECG Database (PTB-XL), and the St. Petersburg INCART 12-lead Arrhythmia Database (INCART). This amalgamated dataset targets three key cardiac conditions: Hypertrophy (HYP), Normal (NORM), and Myocardial Infarction (MI). The creation of this dataset involved meticulous selection, preprocessing, and harmonization of data from these sources to maintain uniformity and reliability for subsequent analysis.

The PTB Diagnostic ECG Database is available through PhysioNet and features high-resolution 12-lead ECG recordings from 290 subjects. It encompasses various cardiac conditions with an emphasis on myocardial infarction cases. This database stands out due to its high sampling rate of 1000 Hz, offering detailed waveform information. Meanwhile, the PTB-XL database extends the original by including 21,837 clinical 12-lead ECG records from 18,885 patients at a sampling rate of 100 Hz with each recording lasting for ten seconds—an asset for machine learning applications in ECG analysis. The St. Petersburg INCART database provides valuable insights with its annotated recordings focusing on arrhythmias across diverse patient demographics; it includes seventy-five annotated recordings sampled at 257 Hz over thirty minutes each.

To create a consistent dataset, we carefully selected and prepared only the ECG recordings that showed three conditions: Hypertrophy, Normal heart patterns, and Myocardial Infarction. Ensuring consistency across different databases necessitated uniform segment lengths. Although PTB-XL database contained pre-formatted ten-second segments suitable for direct analysis, we needed to perform segmentation on both PTB and INCART databases to achieve the same interval length. The Pan-Tompkins algorithm was implemented for accurate R-peak detection, which was crucial for proper segmentation and ensuring each segment began with a complete cardiac cycle. Zero-padding addressed any R-wave interruptions, ensuring smooth transitions between segments while preserving waveform integrity throughout the study's compilation process.

In order to align the sampling rate uniformly across all datasets, INCART (257 Hz) and PTB (1000 Hz) recordings were reduced to 100 Hz, matching the PTB-XL's rate. This down sampling process utilized cubic interpolation through the `resample` function from `scipy` library in Python, ensuring that essential frequency information and morphological characteristics were retained while standardizing data formats. A bandpass filter (0.5-40 Hz) was applied to remove baseline wander and high-frequency noise while preserving clinically significant ECG components. Post-segmentation and resampling, each ECG recording in the merged dataset was standardized to contain 1000 samples, corresponding to 10 seconds of ECG data at a 100 Hz sampling rate. This uniformity in sampling rate, segment length, and signal quality is essential for maintaining consistency in input data fed into subsequent machine learning models.

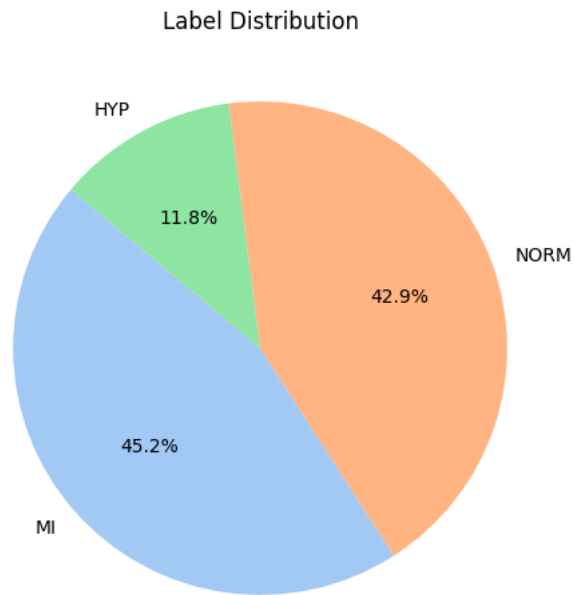


Figure 3.2 Label Distribution of the Combined Dataset.

The combined dataset, consisting of 21,868 samples, showcases distinct characteristics. Label Distribution reveals Myocardial Infarction (MI) at 45.2%, Normal (NORM) at 42.9%, and Hypertrophy at 11.8%. This balance between MI and NORM cases, alongside the significant representation of HYP cases, mirrors their prevalence in clinical practice. The Age Distribution spans a broad range from young adults to seniors, with a concentration in the 60-70 age bracket—corresponding to the higher occurrence of cardiac issues in older individuals. In terms of Sex Distribution, males constitute around 57% of the dataset while females make up approximately 43%, aligning with observed trends in cardiac condition incidences.

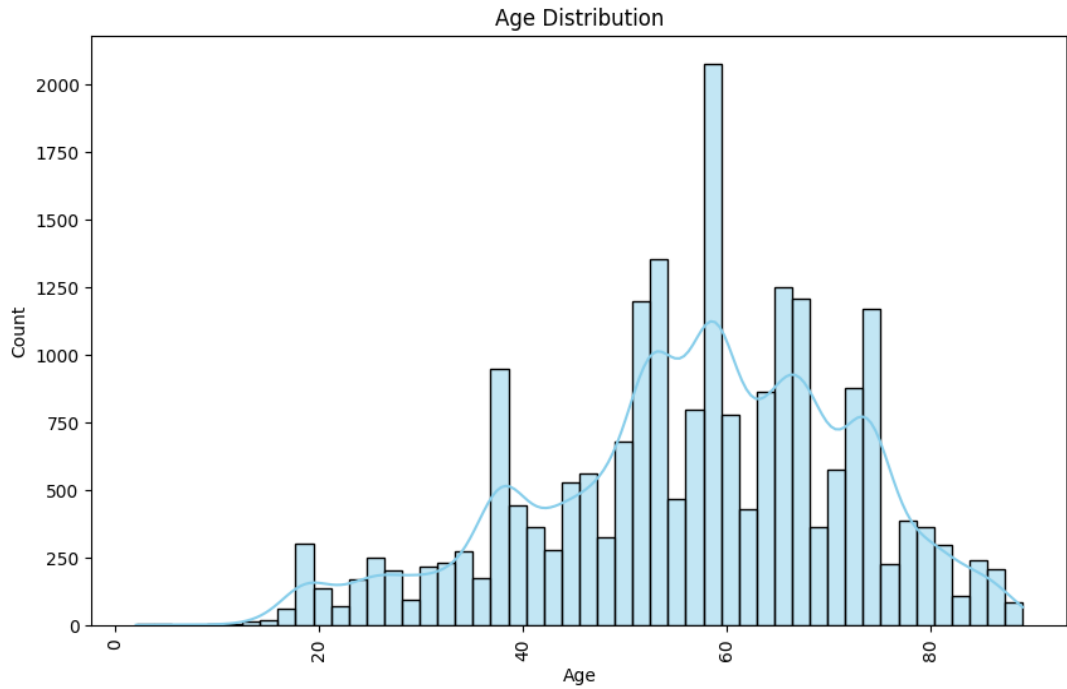


Figure 3.3 Age Distribution of Dataset.

Analyzing Age by Sex and Condition through box plots uncovers that median ages for MI patients surpass those for NORM and HYP across both genders. The wider age span seen among HYP cases is particularly notable in male subjects. Patient Count segmented by Sex and Label is depicted via stacked bar charts; here it becomes evident that while males predominantly feature among MI cases, NORM occurrences are more evenly distributed between sexes. Although HYP shows fewer instances overall, its sex-based distribution resembles that seen in MI data.

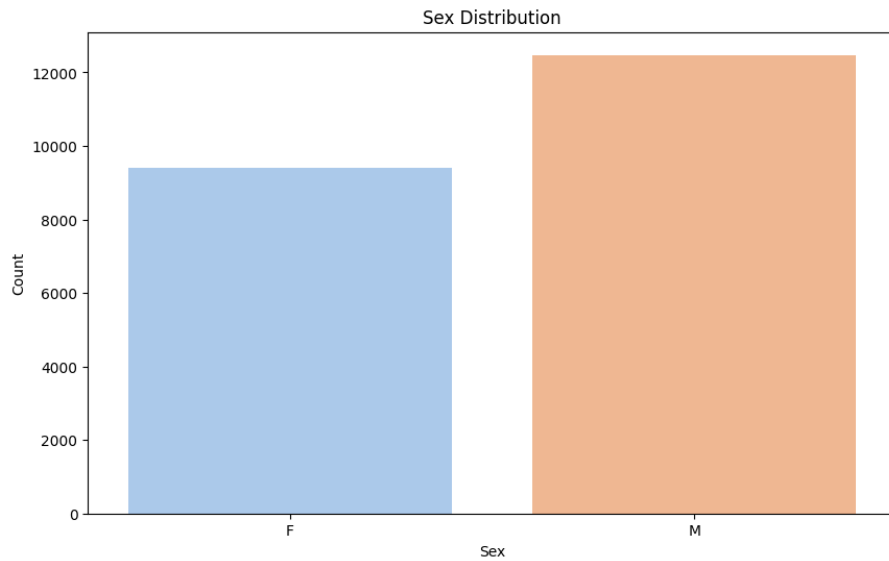


Figure 3.4 Sex Distribution of Dataset.

This meticulously curated and preprocessed dataset provides a robust foundation for the development and validation of the transformer-based interpretable algorithm for diagnosing myocardial infarction from ECG data. The diverse representation of age, sex, and cardiac conditions ensures that the subsequent analysis and model development will be based on a comprehensive and clinically relevant dataset.

3.2. Equipment

Graphics Processing Units, often abbreviated as GPUs, have become an essential part of advancing deep learning endeavors. With their unique architecture allowing for simultaneous handling of numerous calculations, these devices shine brightly in tasks involving intricate matrix and tensor computations—the very core that powers neural networks. The boost they provide can transform what once took weeks into mere hours or days, drastically improving the pace at which models learn and adapt. This swiftness not only boosts productivity but opens doors to crafting more sophisticated systems and managing vast pools of information—thereby expanding the horizons within artificial intelligence explorations.

For this research, an Apple M3 GPU with 10 cores was utilized. This GPU is part of the Apple Silicon line, designed for efficient processing of parallel computations common in deep learning tasks. The 10-core configuration provides a balance of

performance and energy efficiency, suitable for a range of machine learning applications.

Key specifications:

- GPU: Apple M3
- Core count: 10
- Architecture: Based on Apple's custom GPU design

This GPU was chosen for its ability to handle the computational demands of the deep learning models used in this study. The 10-core design allows for parallel processing of multiple tasks simultaneously, which is particularly beneficial for the matrix operations prevalent in neural network computations.

3.3. CNN Model

In this investigation, we constructed a model based on Convolutional Neural Networks to classify electrocardiogram signals into three distinct categories: Hypertrophy, Normal, and Myocardial Infarction. Our process initiated with an intricate preparation of data, utilizing a customized function designed to load ECG information from diverse file paths. This approach guaranteed that our dataset was both extensive and reflective of the conditions under study. The loading mechanism involved systematically traversing folder structures, extracting ECG waveforms from .dat files, and assigning each waveform its respective label derived directly from the folder's designation.

To maintain dataset integrity and consistency, we applied label encoding method from scikit-learn for converting categorical labels into numerical values. This conversion is essential for neural networks as they require numeric input formats. We adopted a strategic approach to splitting the data ensuring robustness in our model's reliability and generalizability. Specifically, we divided the dataset into training sets encompassing 70% of data, validation sets at 15%, and test sets also at 15%. Such division decreases risks associated with overfitting while offering realistic performance evaluation metrics within practical scenarios.

For enhancing memory efficiency alongside optimizing training processes, a TensorFlow data generator was employed. This generator creates batches dynamically

during execution—a critical feature when handling substantial datasets beyond available memory capacities. Key components of this implementation included random shuffling to introduce variability in training sequences, batching with set sizes typically around 32 instances per batch, along with prefetching mechanisms that overlap preprocessing tasks with active model execution cycles.

The structure of our CNN model aims to systematically extract and interpret features from input ECG signals. Launching with an initial layer that receives the raw ECG data, the architecture then integrates three convolutional layers, each succeeded by a max pooling layer. These convolutional stages employ 32, 64, and 128 filters respectively, all set with a kernel size of 3 and ReLU activation function. This sequence ensures the capture of complex features in the ECG waveforms, transitioning from fundamental peaks and valleys to more sophisticated patterns indicative of particular cardiac anomalies.

Transitioning beyond these layers, we incorporate a global average pooling layer which condenses each feature map into a singular value. This significant reduction in parameters is crucial for minimizing overfitting risks. Subsequently, a dense layer comprising 128 units coupled with ReLU activation facilitates learning complex non-linear associations derived from previously extracted high-level features. Ending this sequence is a final dense layer equipped with softmax activation that outputs probability distributions across three distinct classes: HYP, NORM, and MI.

Our compilation strategy leverages the Stochastic Gradient Descent optimizer infused with momentum at a learning rate of $5e^{-4}$ combined with momentum factor set at 0.9 for accelerated convergence through flat terrains in the loss surface. Employing sparse categorical cross-entropy as our loss metric suits multi-class classification where class exclusivity prevails. Training spanned over 100 epochs to afford adequate iterations for recognizing inherent data patterns comprehensively. Throughout training phases conducted in batches containing 32 samples each cycle was validated against an evaluation subset post every epoch ensuring continuous performance assessment.

To measure our model's effectiveness, we employed a diverse set of evaluation methods. Tracking both training and validation accuracy throughout the process helped us observe learning progress while catching overfitting signs early on. By plotting the training and validation loss across epochs, we could better understand how

well the model converged. Post-training, a confusion matrix based on test data provided a detailed performance breakdown per class, illuminating any consistent errors. We further generated an exhaustive classification report that covered precision, recall, and F1-score for each class to offer insights into potential class imbalances.

For clarity in results interpretation, various visualization techniques were utilized. Graphs depicting training and validation accuracy as well as loss over epochs offered insight into learning patterns and highlighted optimal epoch counts. A heatmap visualizing the confusion matrix rendered an intuitive picture of classification success across all categories.

In conclusion, our CNN-based method for ECG classification underscores deep learning's promise in automating cardiac diagnosis tasks. With three convolutional layers carefully designed for precise data handling, it forms a solid groundwork for analyzing ECG readings. The comprehensive suite of evaluation metrics coupled with illustrative visualizations ensures thorough performance assessment across Hypertrophy, Normal conditions, and Myocardial Infarction classifications. This research marks substantial progress towards reliable automated systems for ECG analysis; future paths might explore sophisticated architectures or leverage larger datasets via transfer learning strategies to enhance overall capability.

3.4. CRNN Based Model

The third prototype showcased adopts a structure known as Convolutional Recurrent Neural Network (CRNN). In this model, 1-dimensional CNN and RNN are combined. This model is tailored for categorizing electrocardiogram data into three specific groups: Hypertrophy, Normal, and Myocardial Infarction. By merging the advantages of Convolutional Neural Networks (CNNs), which are skilled at drawing out features, with Bidirectional Long Short-Term Memory (BiLSTM) networks that adeptly handle time-related connections in ECG signals, it offers a pioneering method. The arrangement of this model is shaped through the process of hyperparameter adjustment, ensuring it aligns optimally with the distinct traits found within the ECG dataset.

Initially, the CRNN setup comprises several convolution layers—anywhere from one to three—and each layer follows up with processes such as batch normalization and

max pooling. Each convolutional layer's filter count ranges from 32 to 256 while kernel dimensions may vary between 3 and 7. This opening segment plays a vital role in isolating significant attributes from unprocessed ECG data. Once past these convolutional stages, there's an integration of BiLSTM layers numbering between one and three; their units fluctuate between 32 to 256 per layer. Owing to their bidirectional nature, these LSTM levels permit handling of ECG sequences in dual directions—both forwards and backwards—which enhances their capacity to identify intricate temporal trends proficiently.

In the ultimate phase of the model's design, one might incorporate up to three Dense Layers. Each layer is accompanied by a dropout mechanism, providing an opportunity for regularization. The count of units within these layers is fine-tuned, ranging from 32 to 256, and the dropout percentages fluctuate between zero and half. Such a configuration in the Dense Layer prepares the way for complex combinations of features that have been previously extracted, contributing to more effective categorization outcomes. At the end of this sequence lies an output layer employing softmax activation to yield probabilities across different cardiac states.

For fine-tuning purposes, Adam optimizer steps into play with its learning rate finely adjusted within a range from $1e^{-5}$ to $1e^{-3}$. This ensures efficient updates during training cycles. A loss function called sparse categorical cross-entropy proves useful here due to its suitability for tasks involving multiple class categories. Training processes benefit from utilizing a personalized TensorFlow data generator which efficiently manages preprocessing alongside batch handling tasks. To prevent fitting issues where model complexity overshadows actual performance on new data, early stopping serves as a helpful measure; it watches over validation loss metrics with patience extended up through five rounds before halting further iterations if improvements stall out.

The process of adjusting hyperparameters utilizes the Hyperband algorithm from Keras Tuner, which smartly investigates various configurations to identify the best model setup. This method provides a structured way to examine different architectural selections and parameter settings, aiming for optimal outcomes in classifying ECG signals.

To measure how well the model performs, several metrics are considered, such as overall accuracy and precision. Additionally, recall and F1-score for each category are

analyzed. A confusion matrix is produced to give detailed insights into how effectively the model categorizes diverse heart conditions. This thorough evaluation strategy allows us to see both what works well with the model and where there might be room for enhancements.

This CRNN-based design stands as an advanced method for interpreting ECG data by combining CNNs and BiLSTMs. The fine-tuning of hyperparameters ensures that this system is carefully adjusted to fit specific features of ECG information. Such customization potentially offers superior classification results when compared with models that have fixed structures. Moreover, it seeks not only high accuracy but also flexibility that can adapt across various datasets related to heart signal classification tasks.

3.5. CNN and Attention Model

The third model introduced here employs a Convolutional Neural Network augmented with a multi-head attention mechanism. This innovative method aims to significantly boost the model's proficiency in identifying complex temporal patterns within ECG signals, which could translate into enhanced accuracy. Initially, the architecture adopts several convolutional layers similar to traditional CNN designs. The starting layer utilizes 32 filters paired with a kernel size of 3 and ReLU activation, succeeded by max pooling with a pool size of 2. Subsequently, another convolutional layer steps in with 64 filters and identical kernel parameters, accompanied again by max pooling. Progressing further, the third convolutional layer expands to 128 filters while retaining the same kernel dimensions and activation.

The standout feature of this approach is the integration of multi-head attention following these preliminary convolutional stages. Such an attention mechanism equips the model to concurrently zero in on multiple aspects of the input sequence. This capability potentially enables it to grasp more subtle characteristics embedded within ECG signals—characteristics crucial for precise classification.

For implementing multi-head attention, two distinct one-dimensional convolutional layers each possessing 128 filters are initially applied to outputs from the third convolutional tier. These generate query and value tensors vital for attention processing. Subsequently, we employ a multi-head attention layer featuring four

heads—each characterized by key dimensions set at 128—which empowers focusing on diverse temporal intervals along with different feature subspaces present within inputs; thus possibly encapsulating both localized details alongside broad dependencies intrinsic to ECG readings effectively fostering more refined classification outcomes.

The output from the attention layer transitions through a Global Average Pooling stage, compressing spatial dimensions into a singular vector. Following this, a dense layer with 128 units and ReLU activation introduces non-linear combinations of attended features. The final dense layer employs softmax activation to generate class probabilities.

To optimize our model, we started to employ Stochastic Gradient Descent (SGD) with a learning rate set at $5e^{-4}$ and momentum fixed at 0.9. This approach aids in accelerating convergence while avoiding plateaus in the loss landscape. For our multi-class classification task, sparse categorical cross-entropy was selected as the loss function due to its effectiveness. Incorporating multi-head attention offers several advantages: it enables simultaneous focus on various parts of the ECG signal—critical for detecting complex cardiac patterns—and assists in distinguishing relevant input sequence segments from irrelevant ones for improved feature extraction.

The first training spanned across 100 epochs, granting sufficient duration for the intricate architecture to learn data patterns comprehensively. By continuously monitoring training and validation accuracy throughout this period, we could effectively gauge progress and identify potential overfitting indicators early on.

To evaluate our model's performance comprehensively, we employed a multi-faceted approach. We generated a confusion matrix using the test set, providing a detailed breakdown of the model's performance for each class. Additionally, we produced a comprehensive classification report, including precision, recall, and F1-score for each class. These metrics offer a nuanced view of the model's performance, accounting for potential class imbalances and providing insights into the model's strengths and weaknesses for each cardiac condition.

To sum up, our CNN equipped with multi-head attention marks substantial progress in ECG signal classification automation. Merging CNN's prowess in feature extraction with the focused insights derived from attention mechanisms results in a more refined

detection of complex patterns within ECG data. This hybrid approach is not just about enhancing accuracy; it also offers clearer interpretability regarding how decisions are made by the model. Future explorations could focus on fine-tuning these attention mechanisms—possibly adjusting head counts or integrating diverse forms of self-attention or cross-attention—to refine performance further.

3.6. Our Proposed Transformer Based Model

In our latest research, we developed a cutting-edge Transformer-based model for electrocardiogram classification, incorporating RR interval information to significantly enhance the model's performance. This approach represents a substantial advancement in the field of automated ECG analysis, combining the powerful sequence modeling capabilities of Transformers with crucial temporal information derived from the ECG signal.

Our methodology began with a detailed data preprocessing stage, which was critical to the success of our model. We segmented the ECG signals into 2-second windows with a 50% overlap. This segmentation strategy was carefully chosen to allow for comprehensive coverage of the signal while maintaining temporal continuity. The 2-second window size ensures that each segment contains multiple cardiac cycles, providing sufficient context for the model to learn from. The 50% overlap between adjacent segments helps in capturing transitional features and ensures that important information is not lost at the boundaries of the segments. This approach is crucial for capturing both local and global features of the ECG waveform, allowing our model to learn from a wide range of cardiac events and their temporal relationships.

A key innovation in our preprocessing pipeline is the incorporation of RR interval information. We developed a sophisticated function to calculate RR intervals using peak detection algorithms. This function employs a dynamic amplitude threshold to ensure the identification of genuine R-peaks while minimizing false positives. The resulting RR interval data was then carefully processed and aligned with the corresponding ECG segments, providing the model with additional rhythmic information that is vital for distinguishing various cardiac conditions.

Our approach to integrating RR intervals is both novel and informative. After calculating the positions of the R points, we create a multi-dimensional feature that

carries both positional and temporal information. This is achieved through a two-step process:

1. Zero-padding: We apply zero-padding to all locations in the ECG signal that are not R points. This step preserves the original signal structure while highlighting the positions of the R peaks.
2. Interval encoding: At each identified R point, we enter the time distance (in seconds) to the next R point. This temporal information captures the heart rate variability and rhythm characteristics.

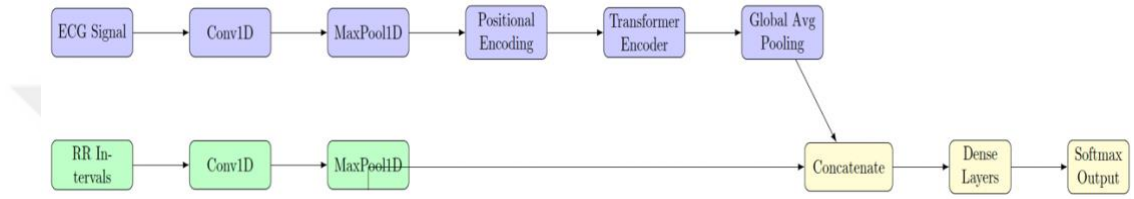


Figure 3.5 High Level Architecture of Our Model.

The resulting feature is a sequence that mirrors the length of the original ECG signal, where most values are zero (non-R points), and the non-zero values represent the RR intervals at their exact R point locations. This method effectively encodes both the timing and the relative positions of the R peaks within the signal.

This integration of RR intervals represents a significant step forward in ECG analysis, as it allows our model to consider both the morphology of the ECG waveform and the timing of cardiac cycles simultaneously. The positional aspect of this feature helps the model to correlate specific waveform characteristics with the R peak locations, while the temporal information enables the detection of rhythm abnormalities and beat-to-beat variations.

On the other hand, Our model architecture is built upon the Transformer framework, which has shown remarkable success in sequence modeling tasks across various domains. We designed a custom Transformer Encoder Layer, which includes sophisticated multi-head attention mechanisms and feed-forward networks. The multi-head attention allows the model to focus on different aspects of the input sequence simultaneously, capturing complex dependencies within the ECG signal at various time scales. This is particularly important for ECG analysis, where different features

of the waveform (such as P waves, QRS complexes, and T waves) all contribute to the overall cardiac assessment.

To account for the sequential nature of ECG data, we implemented a Positional Encoding layer. This layer adds position-dependent information to the input, enabling the model to identify and utilize the order of the sequence effectively. In the context of ECG analysis, this positional information is crucial as it helps the model understand the temporal progression of the cardiac cycle, which is essential for accurate diagnosis.

The core of our model processes the ECG signal through multiple Transformer Encoder Layers. Each layer refines the representation of the input, allowing the model to capture increasingly abstract features of the ECG signal. Concurrently, we process the RR interval data through a separate convolutional neural network branch. This CNN branch is designed to extract relevant features from the RR interval data, such as heart rate variability and rhythm irregularities. This dual-stream architecture allows the model to learn from both the raw ECG waveform and the derived RR interval features independently before combining them for final classification.

Our method signifies a considerable leap forward in the domain of computerized ECG examination, setting itself apart from earlier works through distinct enhancements. While conventional techniques might focus solely on ECG shapes or handle RR intervals independently, our framework merges these components smoothly into one cohesive system. This innovative integration of the Transformer-oriented ECG evaluation with a CNN-based approach for RR intervals stands as a groundbreaking addition to scholarly discussions. This combination strategy facilitates an all-encompassing analysis that captures intricate waveform details alongside comprehensive rhythmic trends simultaneously. Unlike older models that may disregard chronological context, our technique retains and leverages accurate timing and spatial information related to R peaks using advanced RR interval encoding mechanisms. Additionally, employing the Transformer design for ECG study paired with uniquely crafted layers and integration methods expands the potential horizons of automatic heart function assessments. Such a comprehensive strategy boosts detection sensitivity for subtle irregularities while offering an enhanced classification scheme that's both reliable and medically significant—overcoming challenges identified in prior research efforts while establishing new benchmarks in cardiac interpretation systems. This is a novel fusion strategy where the outputs of the Transformer-based

ECG processing stream and the CNN-based RR interval processing stream are concatenated. This combined representation undergoes further processing through dense layers, allowing the model to learn complex interactions between the ECG morphology and the rhythm information. This fusion approach enables our model to make more informed decisions, considering both the shape of the ECG waves and the timing between heartbeats.

$$Y_{combined} = [Y_{ECG}^{final}; Y_{RR}^{final}] \quad (3.1)$$

Where $Y_{combined}$ represents the combined output, Y_{ECG}^{final} represents the final ECG output, Y_{RR}^{final} represents the final RR output, and $[\cdot]$ denotes concatenation of these outputs.

Our training process utilized the Stochastic Gradient Descent (SGD) optimizer with momentum and gradient clipping to ensure stable training. The choice of SGD with momentum allows for efficient navigation of the loss landscape, while gradient clipping prevents extreme parameter updates that could destabilize the training process. We implemented a custom callback for managing long training times, automatically saving checkpoints and allowing for training resumption in case of interruptions. This robust training setup ensures that our model can be trained effectively on large datasets, which is crucial for capturing the wide variability present in ECG signals across different patients and conditions.

To evaluate our model's performance, we employed a record-level majority voting scheme. This approach aggregates predictions from individual segments of a full ECG record, providing a more robust classification at the patient level. For each full ECG recording, our model makes predictions on multiple 2-second segments, and the final classification is determined by the majority vote across these segments. This strategy is particularly important in clinical settings where diagnosis is typically made on full ECG recordings rather than short segments. It helps to mitigate the impact of potential noise or artifacts in individual segments and provides a more holistic view of the patient's cardiac condition.

3.7. Hyperparameter Optimization

Employing the Keras-Tuner platform, we delved into an array of 128 diverse hypermodels per model, adopting the hyperband approach for fine-tuning our model parameters. This extensive exploration played a key role in enhancing the performance metrics of our models. The hyperband strategy is renowned for its skill at harmonizing discovery with resource allocation, allowing us to prudently distribute computational resources to more promising configurations and swiftly discard less efficient options. Presented below are tables that detail the ranges explored for each model's hyperparameters.

Table 3.3: CNN Model Hyperparameter Search Ranges.

Hyperparameter	Search Range
Convolutional Layers	2 - 5
Number of Filters	32 - 256
Kernel Size	3 - 7
Dense Layer Units	64 - 512
Learning Rate	$1e^{-4} - 1e^{-2}$

Table 3.4: CRNN Based Model Hyperparameter Search Ranges.

Hyperparameter	Search Range
Convolutional Layers	1 - 3
Convolutional Filters	32 - 256
Convolutional Kernel Size	3, 5, 7
BiLSTM Layers	1 - 3
BiLSTM Units	32 - 256
Dense Layers	1 - 3
Dense Layer Units	32 - 256 (step 32)
Dropout Rate	0.0 - 0.5
Learning Rate	$1e^{-5}$ - $1e^{-3}$

Table 3.5: CNN with Attention Model Hyperparameter Search Ranges.

Hyperparameter	Search Range
Convolutional Layers	2 - 5
Number of Filters	32 - 256
Attention Heads	2 - 8
Attention Dimension	64 - 256
Dense Layer Units	64 - 512
Learning Rate	$1e^{-4} - 1e^{-2}$

Table 3.6: Transformer-Based Model Hyperparameter Search Ranges.

Hyperparameter	Search Range
Transformer Layers	2 - 6
Attention Heads	4 - 16
Hidden Dimension	128 - 512
Feed-Forward Size	256 - 1024
Learning Rate	$1e^5 - 1e^3$

The hyperparameter optimization process played a critical role in determining the optimal configurations for each model. By applying the hyperband method, we were able to:

1. Efficiently explore a large hyperparameter space
2. Quickly identify and focus on the most promising configurations
3. Allocate more computational resources to the best-performing models
4. Achieve better results in less time compared to traditional grid or random search methods

Our study included a comprehensive examination of a range of 128 varied hypermodels. By employing the highly efficient hyperband method, we significantly boosted our models' capabilities. This strategic effort is particularly critical for tasks demanding precision, such as classifying electrocardiogram (ECG) data, where accuracy is essential in medical settings.

The adaptive resource allocation inherent in the hyperband strategy enabled us to rapidly identify superior configurations, a necessity when grappling with intricate models and extensive datasets that are characteristic of ECG examinations. Through

this technique, not only did we elevate our models' proficiency but also fine-tuned the use of computational resources during training sessions.

Across different model frameworks, including convolutional neural networks, convolutional recurrent neural networks, CNNs with attention mechanisms, and Transformer-based architectures, the hyperband technique served as a base strategy for effectively exploring diverse sets of hyperparameters. This approach allowed us to detect which parameters had the greatest influence on each architecture's performance. For instance, determining optimal values such as the count of layers in a convolutional setup or learning rates was crucial for CNNs. Similarly, understanding the impact of varying attention heads and hidden dimensions proved essential for Transformer models.

Applying hyperband in our process of refining model settings significantly contributed to developing high-caliber systems with greater efficiency. Through this methodology, not only did we advance our existing ECG classification systems but we also gained substantial knowledge that holds promise for forthcoming investigations within this key domain of medical analysis and assessment. valuable insights for future research in this critical area of medical diagnostics.

4. RESULTS AND DISCUSSIONS

This section presents the results obtained from our three developed models: CNN, CNN with Attention Mechanism, and Transformer-based model. For each model, we analyzed confusion matrices and calculated performance metrics. Additionally, we employed Keras-Tuner to explore 128 different hyperparameter combinations for each model to determine the optimal configurations.

4.1. Result

Table 4.1: Comparative Performance Metrics for All Models.

Model	Accuracy	Precision	Recall	F1-Score
CNN	0.91	0.92	0.91	0.91
CRNN Based Model	0.916	0.93	0.90	0.91
CNN with Attention	0.92	0.93	0.91	0.92
Our Transformer Based Model	0.93	0.94	0.92	0.93
CapsNet [108]	0.92	0.92	0.90	0.91
CSA-MResNet [109]	0.92	0.94	0.90	0.92
SE-ECG Net [110]	0.95	0.96	0.92	0.94

Note: Precision, Recall, and F1-Score are macro-averaged across all classes

Analyzing the performance outcomes, it is clear that the model based on the Transformer architecture excels significantly beyond their counterparts within our internal 4 models. Achieving a notable accuracy rate of 0.93, it exceeds both CNN at 0.91 and CRNN models registering at 0.916, with even the CNN with Attention model trailing behind at 0.92. Furthermore, Transformers exhibit outstanding capabilities in precision, scoring a commendable 0.94 and an F1-score reaching 0.925; these metrics further emphasize their proficiency compared to alternative strategies.

In contrast, while CNN demonstrates solid functionality overall with its lowest relative accuracy of 0.91 amongst these methodologies, it still manages to offer consistent outputs as evidenced by its equal precision and recall figures of approximately 0.92 and slightly below that at about 0.91 respectively for recall tasks. This indicates that

despite CNNs being capable of identifying localized signal features within ECG data effectively, they potentially encounter limitations when addressing more extended relational dependencies vital for achieving optimal classification outcomes.

Moreover, CRNN model demonstrates notable success, securing the third position in accuracy with a score of 0.916 and second place in precision with 0.93 among the models assessed. This performance underscores how blending layers that process data sequentially with those analyzing spatial elements can proficiently discern intricate patterns over time within ECG data.

Conversely, the CNN integrated with an attention mechanism exceeds its simpler counterpart by achieving enhanced accuracy at 0.92. This advancement stems from the model's capacity to home in on critical sections of input sequences effectively; however, this progression is limited when compared to the significant improvement observed with employing a Transformer framework for processing tasks.

The Transformer model's exceptional performance can be attributed to its ability to recognize complex patterns and long-range dependencies in ECG signals. Its self-attention component grants the model the capacity to assess, in a dynamic manner, the significance of various segments of input data. This results in refined interpretations and more precise categorizations that surpass simpler analytical methods.

Our experimental results demonstrate the competitive performance of various deep learning architectures for ECG classification. While CapsNet [108] and CSA-MResNet [109] both achieved respectable accuracy scores of 0.92, our transformer-based model showed superior performance with an accuracy of 0.93 and an F1-score of 0.93. The SE-ECG Net [110] emerged as the top performer with an accuracy of 0.95 and an F1-score of 0.94. However, SE-ECG Net's architecture shares fundamental principles with transformer-based approaches through its use of attention mechanisms, specifically the squeeze and excitation blocks that capture channel-wise dependencies and spatial attention mechanisms that focus on lead-wise importance. This architectural similarity suggests that attention-based feature weighting, whether implemented through traditional transformer blocks or specialized attention mechanisms like squeeze and excitation blocks, plays a crucial role in achieving superior ECG classification performance. The effectiveness of both our transformer-

based model and SE-ECG Net underscores the importance of attention mechanisms in capturing complex temporal and spatial relationships in ECG signals.

Table 4.2: Class-Wise Performance Metrics for All Models.

Model	Class	Precision	Recall	F1-Score
CNN	HYP	0.95	0.88	0.91
	MI	0.91	0.93	0.92
	NORM	0.89	0.93	0.91
CRNN Based Model	HYP	0.97	0.84	0.90
	MI	0.90	0.96	0.93
	NORM	0.92	0.90	0.91
CNN with Attention	HYP	0.96	0.88	0.92
	MI	0.90	0.93	0.92
	NORM	0.90	0.89	0.90
Transformer-based	HYP	0.97	0.89	0.93
	MI	0.96	0.93	0.94
	NORM	0.90	0.95	0.92
CapsNet [108]	HYP	0.93	0.84	0.88
	MI	0.95	0.91	0.93
	NORM	0.89	0.95	0.92
CSA-MResNet [109]	HYP	0.99	0.83	0.90
	MI	0.95	0.91	0.93
	NORM	0.88	0.96	0.91
SE-ECG Net [110]	HYP	1.0	0.84	0.91
	MI	0.97	0.95	0.96
	NORM	0.91	0.98	0.94

Table 4.2 presents a comprehensive analysis of class-specific performance metrics across various model architectures for ECG classification, encompassing Hypertrophy (HYP), Myocardial Infarction (MI), and Normal (NORM) categories. This detailed examination reveals distinctive patterns in model effectiveness and specialization across different cardiac conditions.

The SE-ECG Net [117] exhibits exceptional performance across all categories, particularly excelling in precision metrics. It achieves perfect precision (1.0) for HYP classification and an impressive precision score of 0.97 for MI, coupled with the highest recall rate for NORM cases (0.98). These results translate into outstanding F1-scores, notably 0.96 for MI and 0.94 for NORM classifications, establishing it as the top-performing model in this comparison.

The CSA-MResNet [116] architecture demonstrates remarkable precision in HYP detection (0.99); however, its recall rate for HYP (0.83) suggests a more conservative approach to hypertrophy classification. It shows balanced performance in MI detection, with both precision and recall at 0.91, and achieves a high recall rate (0.96) for NORM cases, indicating reliable recognition of normal ECG patterns. Consistent F1-scores across categories (0.90 for HYP, 0.93 for MI, and 0.91 for NORM) suggest stable performance across different conditions.

The CapsNet [115] architecture presents noteworthy characteristics, particularly in handling normal cases with a high recall rate (0.95) and strong MI detection capabilities (precision 0.95, F1-score 0.93). However, its performance in HYP detection requires improvement, as indicated by lower recall (0.84) and precision (0.93) rates compared to other models, resulting in the lowest F1-score (0.88) for HYP classification among all models evaluated.

The Transformer-based model, while not achieving the highest overall metrics, exhibits promising and consistent performance across all categories. Its balanced effectiveness is evident in strong precision scores for both HYP (0.97) and MI (0.96), along with robust recall rates (0.89 for HYP, 0.93 for MI, and 0.95 for NORM). This consistency, reflected in F1-scores of 0.93 for HYP and 0.94 for MI, underscores the model's potential for reliable clinical application.

The CNN with Attention architecture shows notable improvements over the basic CNN, particularly in HYP detection precision (0.96 vs. 0.95) and F1-score (0.92 vs.

0.91). However, its performance in NORM classification shows slightly diminished results, with a recall of 0.89 compared to the basic CNN's 0.93. This suggests that the attention mechanism, while beneficial for detecting specific conditions, might occasionally overemphasize potential abnormalities in normal ECG patterns.

The CRNN-based model excels in specific areas, matching the Transformer model's precision for HYP (0.97) and achieving high recall for MI cases (0.96). Its balanced performance across categories is reflected in consistent F1-scores (0.90 for HYP, 0.93 for MI, and 0.91 for NORM), though its recall rate for HYP (0.84) indicates potential limitations in sensitivity for hypertrophy detection.

The baseline CNN model, despite being the simplest architecture, maintains surprisingly consistent performance across categories. Its balanced recall rates (0.88 for HYP, 0.93 for MI, and 0.93 for NORM) and precision scores (0.95 for HYP, 0.91 for MI, and 0.89 for NORM) demonstrate that even basic deep learning architectures can provide reliable ECG classification capabilities. The model's consistent F1-scores around 0.91–0.92 across all categories suggest it serves as a solid baseline for comparison.

Ultimately, these findings highlight the evolution and capabilities of various deep learning architectures in ECG classification. While the SE-ECG Net demonstrates superior performance in most metrics, the consistent and promising results of the Transformer-based model suggest significant potential for future development and application in clinical settings. Additionally, the analysis reveals that while newer architectures generally show improvements in specific areas, even baseline models maintain reliable performance levels. Therefore, the choice of model architecture should consider both performance metrics and practical implementation requirements in clinical settings.

4.2. Interpretability

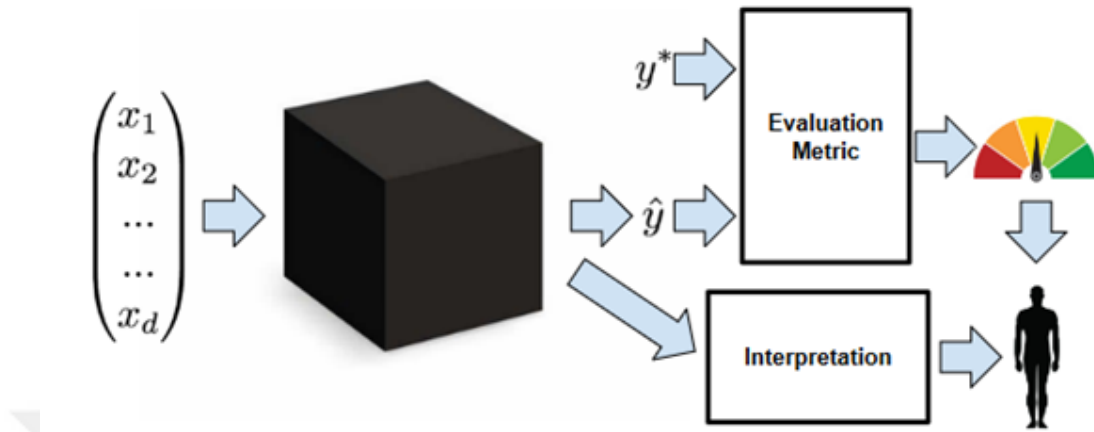


Figure 4.1 Blackbox Diagram of Machine Learning Models [111].

With the rise of machine learning tools in clinical settings, the quest for making these systems clear and understandable has taken center stage. In the world of ECG analysis, gaining healthcare workers' trust is closely tied to how easily they can grasp a model's rationale. While models like deep neural networks shine in handling ECG tasks, their complex inner workings make it tough to spell out individual outcomes. This lack of transparency can make medical experts wary about relying on AI-powered diagnostics, possibly slowing down broader acceptance of these innovative instruments in health care. Focusing on techniques that bring clarity to ECG classification systems highlights their role in turning results into practical knowledge for doctors and nurses alike. By applying different methods—from identifying which features matter most to delving deeper with strategies like layer-wise relevance mapping—we can highlight what parts of an ECG signal play important roles in reaching a decision. This effort not only aids practitioners in understanding why a certain diagnosis was made but also acts as a helpful teaching aid for those learning the basics of interpreting heart signals. Ultimately, crafting ECG classifiers that rival or surpass human skills while delivering easy-to-follow justifications promises big changes ahead for cardiac treatment practices.

4.2.1 GRAD-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) has emerged as a powerful technique for visualizing the decision-making process of deep learning models, particularly in the domain of medical signal analysis such as electrocardiogram (ECG) classification. This technique is especially valuable in the context of our transformer-based ECG classification model, as it provides crucial insights into which temporal regions of the ECG signal are most influential in determining the model's predictions for different cardiac conditions.

Grad-CAM's significance in ECG interpretation stems from its ability to highlight the specific portions of the input signal that are most salient for the model's decision. This visual feedback serves multiple purposes, ranging from enhancing model interpretability by bridging the gap between the black-box nature of deep learning models and the need for explainable AI in medical applications, to enabling clinical validation by allowing medical professionals to verify if the model is focusing on clinically relevant ECG features. Furthermore, it functions as an invaluable educational tool for training medical students by highlighting key diagnostic features in ECGs, while simultaneously serving as a crucial debugging mechanism that helps identify potential biases or errors in the model's learning process.

Before delving into the Grad-CAM results, it's crucial to understand the typical ECG morphology and how it relates to different cardiac conditions. A normal (NORM) ECG is characterized by regular P waves, QRS complexes, and T waves with consistent intervals, while Myocardial Infarction (MI) often presents with ST-segment elevation or depression, Q waves, and T wave inversions. In cases of Hypertrophy (HYP), the ECG may show increased QRS complex amplitude and duration, along with potential ST-segment and T wave abnormalities.

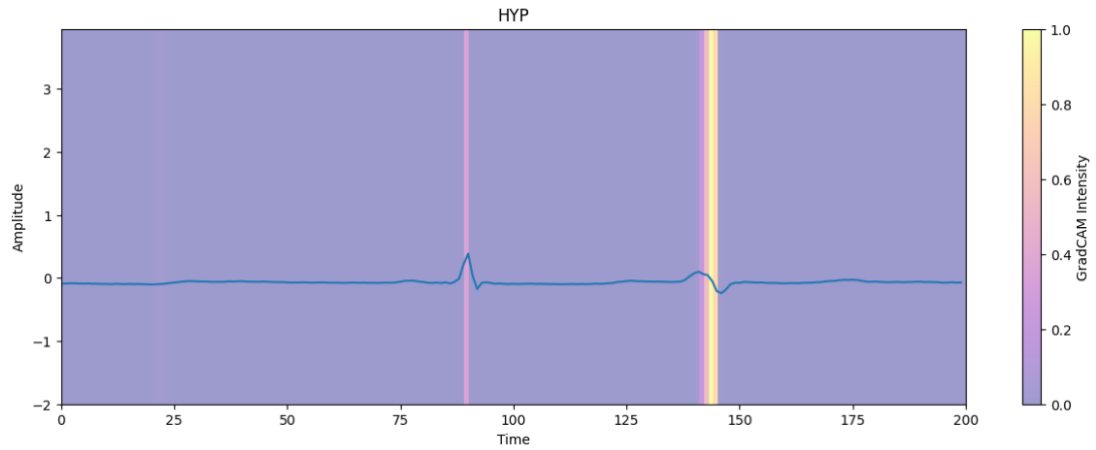


Figure 4.2 Grad-CAM of HYP Example.

Within the Grad-CAM depiction of Hypertrophy (HYP) displayed in Figure 4.2, two pronounced areas with heightened activity stand out distinctly. Initially, one observes a significant peak near time point 85, aligning with the QRS complex of an ECG. Additionally, another area showing notable activation is found around time point 140, likely connected to the T wave phenomena. This pattern of engagement corresponds well with typical alterations seen in hypertrophic states within electrocardiograms. It appears that the model concentrates on amplified amplitude and prolonged duration within the QRS complex — traits indicative of ventricular enlargement conditions. Furthermore, attention towards the T wave zone may suggest focus on abnormalities related to repolarization processes frequently linked with hypertrophic scenarios.

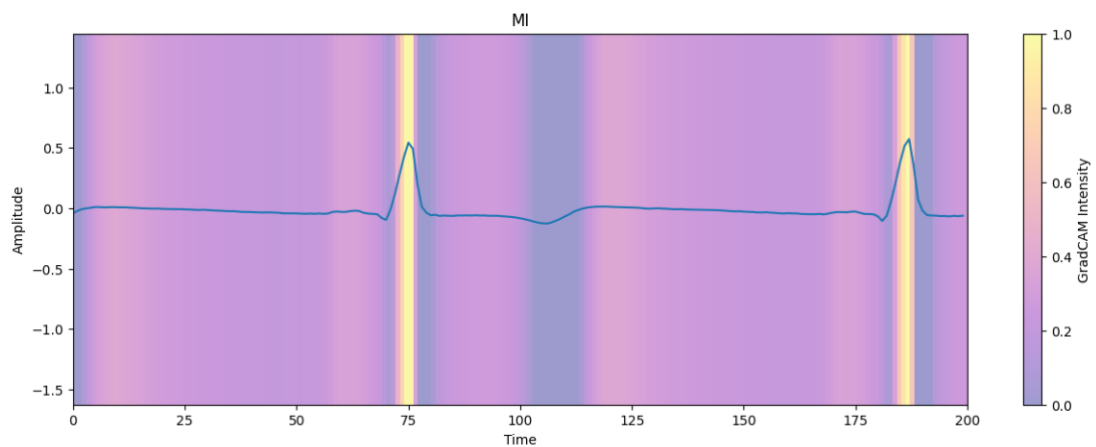


Figure 4.3 Grad-CAM of MI Example.

In examining the Grad-CAM visualization for Myocardial Infarction (MI) displayed in Figure 4.3, one can observe a more intricate activation scheme characterized by two pronounced peaks—one located approximately at time point 75 and another near 185—and an extensive zone of moderate activity bridging these peaks. This pattern intimates that the model is considering various segments of the ECG signal when identifying MI. The initial peak likely aligns with the QRS complex, highlighting potential irregularities in Q waves, whereas the expansive region of activity might denote attention to ST-segment deviations, which are pivotal indicators for diagnosing MI. The subsequent peak may be associated with anomalies in T waves or perhaps signifies the onset of a new cardiac cycle where reversed changes typical of MI could appear.

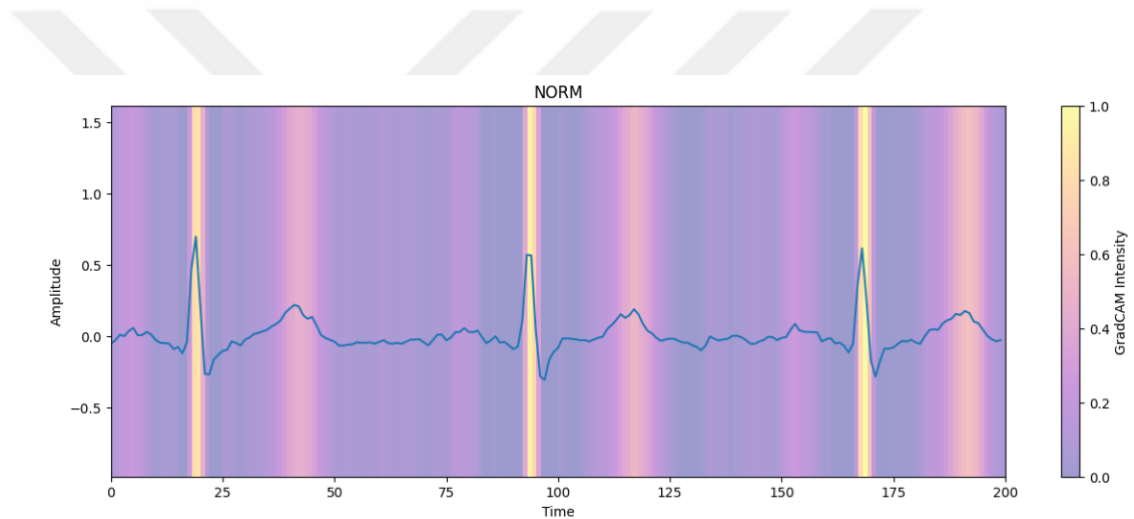


Figure 4.4 Grad-CAM of NORM.

Figure 4.4 illustrates the NORM Grad-CAM visualization, highlighting a widespread pattern of neural activation that exhibits multiple peaks at intervals such as time points 20, 90, and 160. Between these major activation peaks lie broader regions of more subtle, distributed activity. This particular arrangement aligns with what one would expect from a typical ECG, suggesting that the model is validating the steadiness and healthiness of several cardiac cycles. The uniformly timed activations seem to align with regular QRS complexes, whereas the dispersed signals between these peaks may reflect how the model focuses on usual P waves, ST segments, and T waves within an ECG context.

Upon examining the three Grad-CAM visuals, several significant findings become apparent. The specificity of focus is notable, as the HYP and MI visualizations exhibit regions of high activation that are tightly concentrated, suggesting a focused detection of specific ECG abnormalities by the model. In contrast, the NORM visualization displays a more dispersed pattern of attention, indicative of the model assessing overall regularity in the ECG readings. Additionally, the MI visualization spans wider areas over time, hinting at a comprehensive review by the model over extended ECG segments for its diagnosis process. This approach mirrors clinical practices where changes in ST-segments over durations are scrutinized during myocardial infarction assessments. Furthermore, the amplitude of activation is particularly pronounced in the HYP visualization, especially for the QRS complex, corroborating the model's attention to increased QRS amplitude characteristic of hypertrophy. Lastly, the NORM visualization most clearly demonstrates the model's attention to the cyclic nature of the ECG, with regular, repeating patterns of activation.

Ultimately, these Grad-CAM visualizations offer significant insights into our transformer-based model's decision-making process for ECG classification. They reveal the model's focus on clinically relevant ECG features specific to various cardiac conditions, aligning well with established medical knowledge. This interpretability validates the model's learning and boosts its potential integration into clinical practice, where transparency and explainability are essential. Future research could correlate these activation patterns with specific ECG measurements and investigate their evolution throughout the training process.

4.2.3 Lead Importance

To enhance the interpretability of ECG classification, we conducted a lead importance analysis. This analysis quantifies the relative importance of each ECG lead in the model's decision-making process by measuring the impact of noise injection on classification accuracy.

We employed a noise injection technique, adding Gaussian noise to each of the 12 standard ECG leads (I, II, III, AVR, AVL, AVF, V1-V6) at 10 different noise levels (sigma values) ranging from 0.1 to 1.0. For each lead and noise level combination, we performed 50 iterations to ensure statistical robustness. The average decrease in

classification accuracy compared to the original, noise-free ECG was used as a proxy for lead importance.

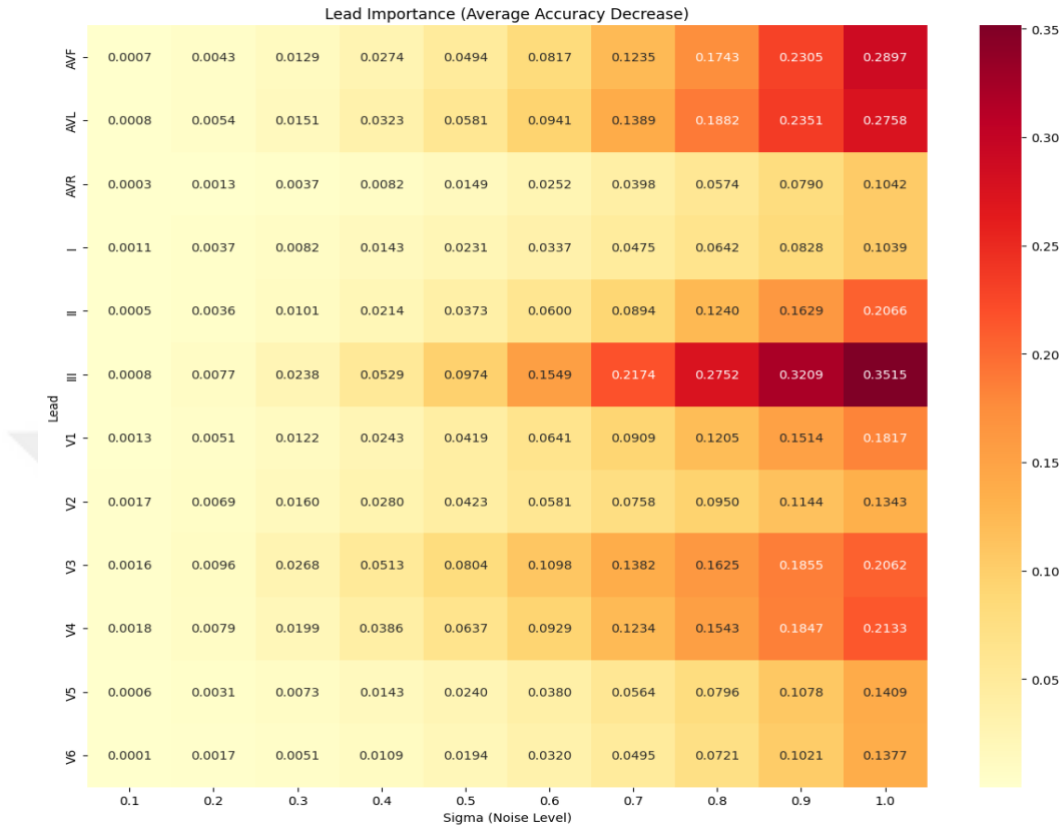


Figure 4.5 Transformer Model’s Heatmap of Accuracy Decrease Per Lead With Respect to Noise

The transformer architecture demonstrates a pronounced sensitivity to Lead III, with the accuracy decrease reaching 0.3515 at $\sigma=1.0$, while maintaining significant importance for AVF and AVL leads. This architecture shows a notably lower reliance on chest leads (V1-V6), suggesting a preference for information contained in limb and augmented leads. The pattern indicates that the transformer’s self-attention mechanisms may be particularly adept at capturing long-range dependencies in these specific leads, potentially due to their representation of cardiac electrical activity from complementary angles.

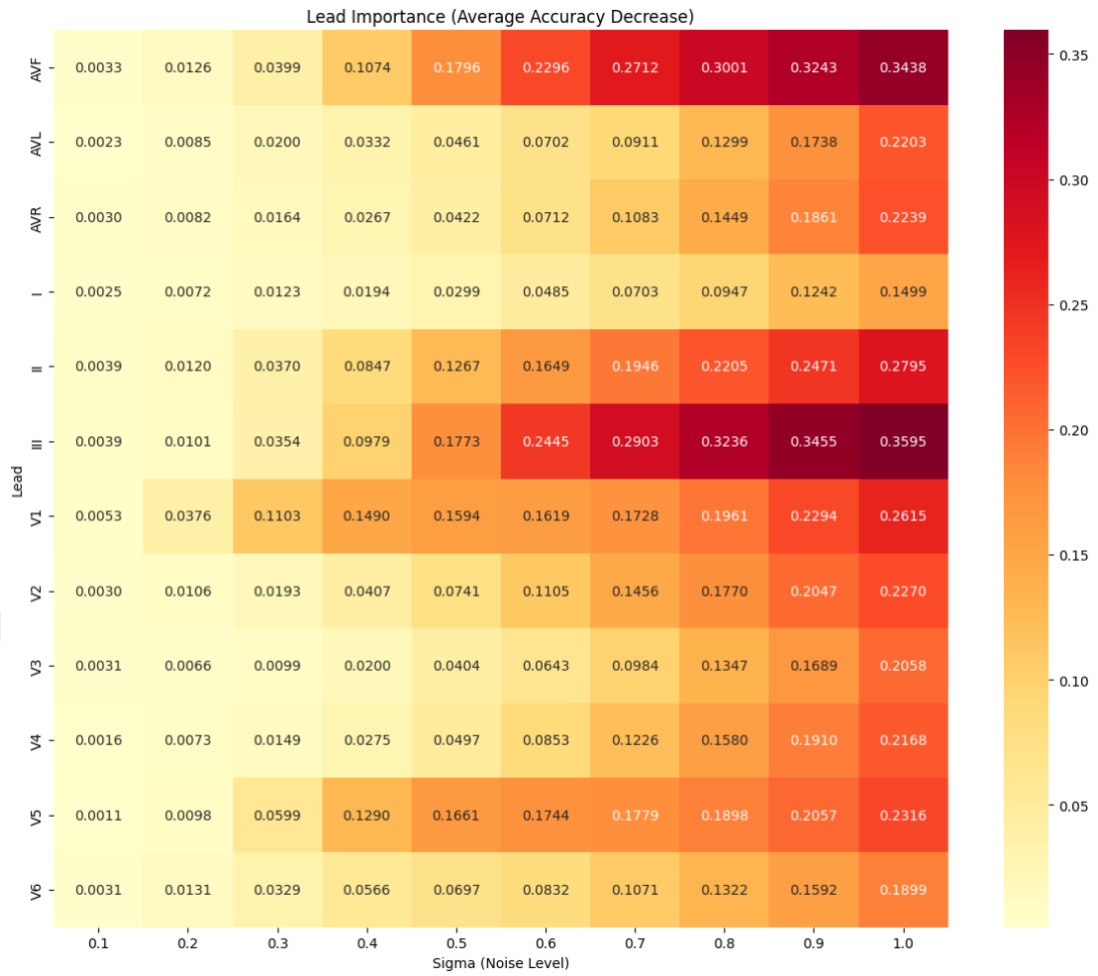


Figure 4.6 CNN and Attention Model’s Heatmap of Accuracy Decrease Per Lead With Respect to Noise

In the CNN with attention architecture, there's a shift in lead importance, with AVF emerging as the most crucial lead, showing an accuracy decrease of up to 0.3438 at $\sigma=1.0$. The model maintains significant importance for Lead III while demonstrating a more uniform distribution of importance across the remaining leads. The addition of the attention mechanism facilitates a more balanced approach to feature extraction, enabling the model to capture subtle interdependencies between different leads that might be overlooked by simpler architectures.

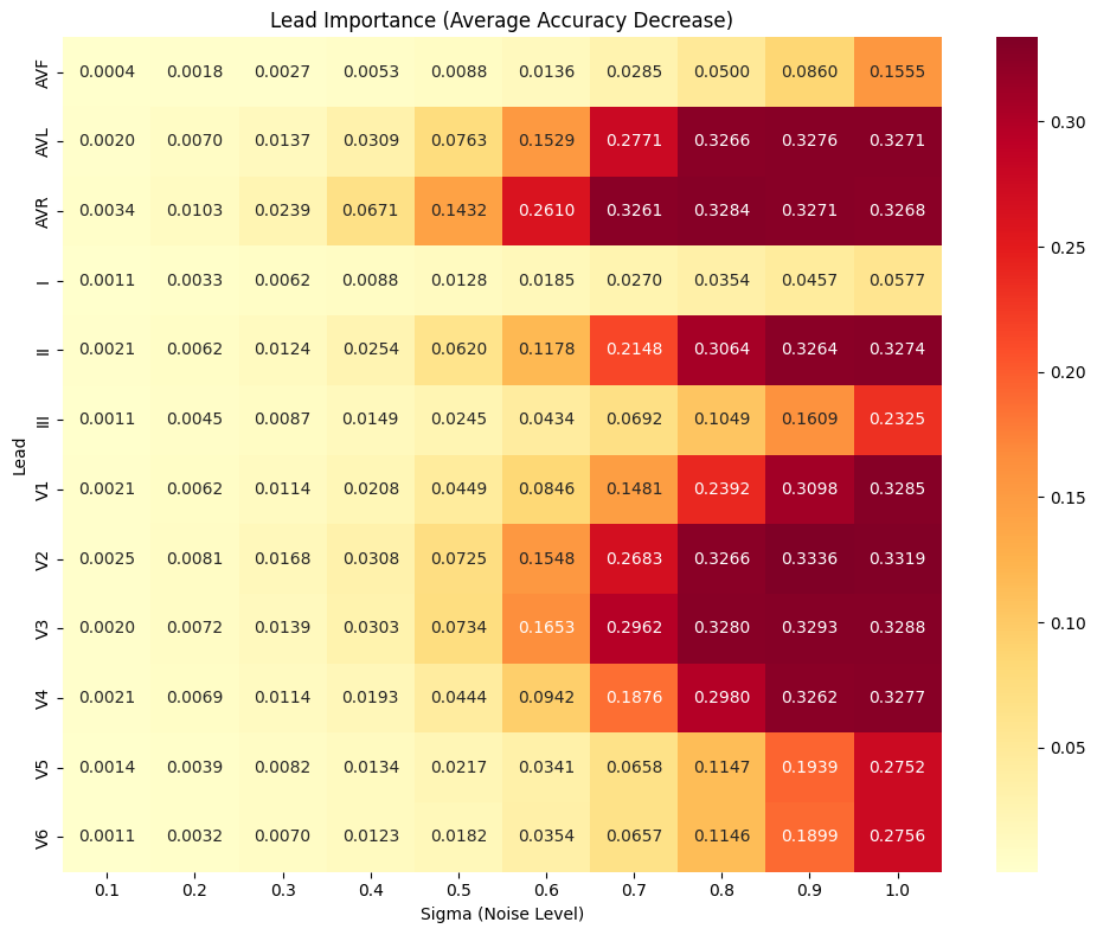


Figure 4.7 CRNN Model’s Heatmap of Accuracy Decrease Per Lead With Respect To Noise

The CRNN architecture presents a distinctly different pattern, characterized by a balanced distribution of lead importance and notably lower sensitivity to noise in peripheral leads. This architecture shows increased sensitivity to information from leads AVR and AVL, departing significantly from the patterns observed in other architectures. The combination of convolutional and recurrent components enables the CRNN to effectively capture both spatial and temporal dependencies in the ECG signal.

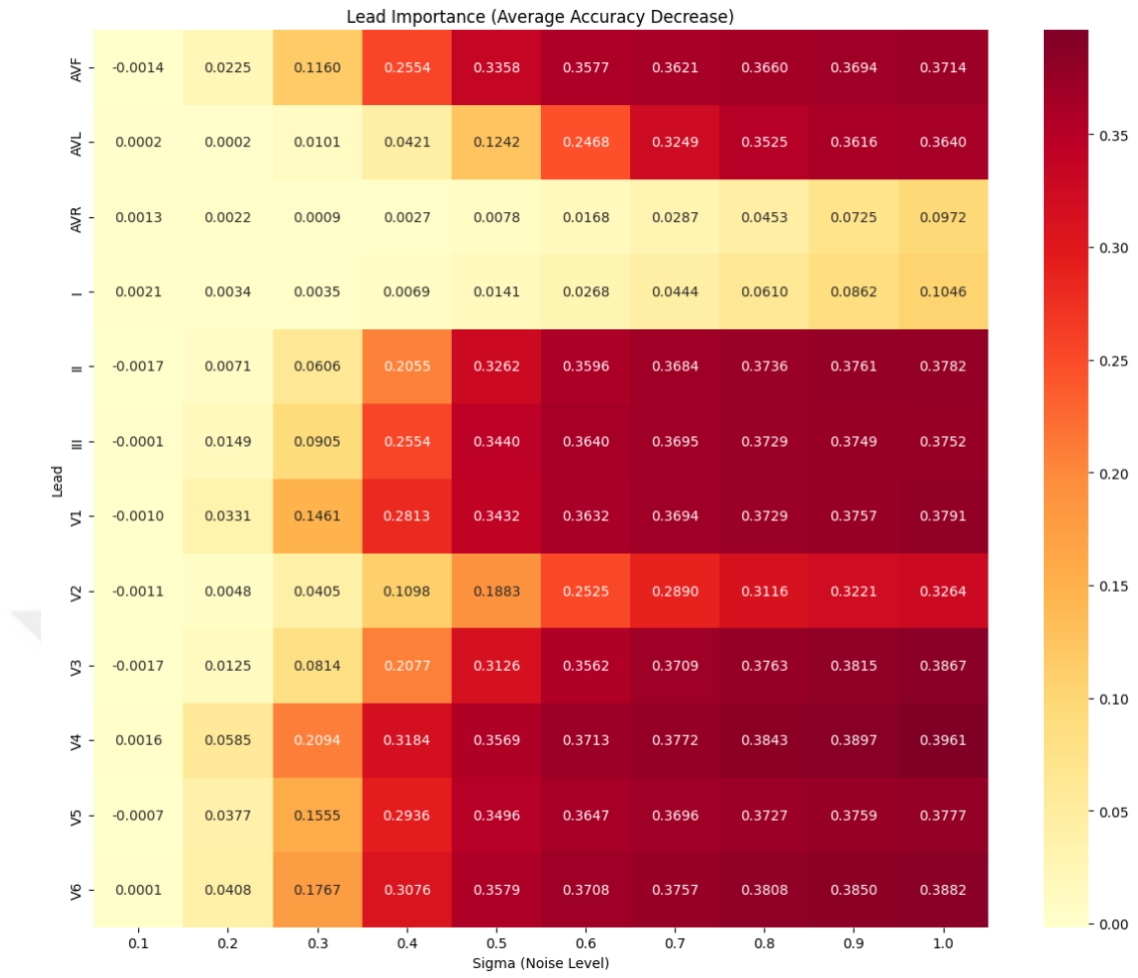


Figure 4.8 CNN Model’s Heatmap of Accuracy Decrease Per Lead With Respect to Noise

The standard CNN architecture exhibits the highest overall sensitivity to noise perturbations, with accuracy decreases approaching 0.39 across multiple leads, including III, AVF, V1, and V4. This heightened sensitivity manifests in a more pronounced nonlinear response to noise, particularly at higher noise levels. The pattern suggests a distributed approach to feature extraction but potentially at the cost of reduced noise resilience compared to attention-based architectures. The similar importance levels across multiple leads indicate that the CNN’s convolutional operations capture spatially local features uniformly across the ECG signal.

Across all architectures, the relationship between noise level and accuracy decrease is nonlinear, though the patterns of degradation vary. The transformer and CNN with attention demonstrate more gradual performance degradation with increasing noise, while the standard CNN shows abrupt transitions at moderate noise levels. The CRNN

exhibits the most uniform degradation pattern, suggesting robust feature extraction mechanisms that maintain stability across different noise conditions.

These architectural variations in lead importance patterns have significant implications for the development and deployment of AI-based ECG analysis systems. The distinct preferences and sensitivities of each architecture suggest that different models might be optimal for various clinical scenarios, depending on specific requirements for lead analysis and noise resilience. The complexity of feature extraction in ECG analysis underscores the importance of architectural choices in determining which aspects of the signal are emphasized during processing.

Our findings highlight the intricate relationship between neural network architecture and ECG lead importance, revealing how structural choices in model design can fundamentally alter the processing and interpretation of cardiac electrical information. This understanding is crucial for developing robust and clinically relevant AI systems for ECG analysis, as it enables informed decisions about architectural choices based on clinical requirements and operating conditions. The variations in lead importance patterns across architectures also suggest potential avenues for ensemble approaches that leverage the complementary strengths of different architectural designs.

To sum up, these results provide valuable insights for future research in AI-based ECG analysis, particularly in the development of more robust and interpretable models. The clear differences in lead importance patterns across architectures emphasize the need for careful model selection tailored to specific clinical applications. Different architectural approaches may be better suited for varied diagnostic scenarios or clinical requirements, and leveraging these insights can enhance the effectiveness of AI-based ECG analysis systems.

5. CONCLUSIONS

5.1. Discussions

In this study, we developed a Transformer-based model enhanced with RR interval information for the classification of ECG signals, aiming to improve the diagnosis of myocardial infarction. The integration of RR intervals provided the model with vital temporal and rhythmic information, complementing the morphological features extracted from the ECG waveforms. The model demonstrated superior performance compared to traditional CNN, CRNN based model and CNN with attention model, achieving an accuracy of 93% and an F1-score of 0.925. This performance gain underscores the effectiveness of combining Transformer architectures with additional temporal features in the context of ECG analysis.

The lead importance analysis across different deep learning architectures revealed distinct patterns in how each model processes and prioritizes ECG lead information. The transformer architecture demonstrated particular sensitivity to Lead III and AVF, while the CNN with attention showed a more balanced distribution with AVF emerging as the most crucial lead. This architectural variation in lead importance provides interesting insights into how different neural network structures interpret ECG signals. The standard CNN exhibited higher overall sensitivity to noise perturbations across multiple leads (III, AVF, V1, V4), suggesting a more distributed but potentially less robust feature extraction mechanism. In contrast, the CRNN architecture showed a uniquely balanced distribution of lead importance with increased sensitivity to AVR and AVL, indicating its capability to capture both spatial and temporal dependencies effectively. These architectural differences in lead importance not only enhance our understanding of each model's decision-making process but also suggest that different architectures might be optimal for specific clinical scenarios. The varying patterns of noise resilience across leads and architectures provide valuable insights for model selection in clinical applications, where signal quality and specific lead information may be of particular importance. This analysis of lead importance across architectures suggests potential connections between theoretical deep learning models and clinical applications, which may help

inform the selection of architectural approaches based on specific diagnostic needs and operating conditions.

Our approach addresses a critical need for transparency in AI-driven medical diagnostics. By incorporating techniques like Grad-CAM and perturbation-based lead importance analysis, we not only improved the model's performance but also made strides toward making its predictions more understandable to clinicians. This is essential for building trust in AI systems used in healthcare, where understanding the rationale behind a diagnosis is as important as the diagnosis itself.

However, the study has limitations. The use of Gaussian noise in the perturbation analysis may not fully capture the complexities of real-world ECG signal distortions caused by patient movement, electrode misplacement, or electrical interference. Additionally, while the model accounts for RR intervals, it does not explicitly model other physiological factors like heart rate variability in different pathological conditions. The datasets used, although comprehensive, may still lack representations of certain rare cardiac events, potentially limiting the model's generalizability.

5.2. Future Work

Future research can build upon this work in several directions. First, incorporating more sophisticated noise models that simulate realistic ECG artifacts could provide a deeper understanding of the model's robustness and lead importance under practical conditions. This could involve using data augmentation techniques to simulate common issues like baseline wander, muscle artifacts, and electrode motion artifacts.

Second, extending the model to consider additional physiological signals and patient metadata could enhance its diagnostic capabilities. Integrating parameters such as blood pressure, oxygen saturation, and patient history could allow the model to make more nuanced predictions, especially in complex clinical scenarios where ECG findings alone may be insufficient.

Third, exploring the use of multi-modal approaches that combine imaging modalities like echocardiograms with ECG data could provide a more comprehensive assessment of cardiac function. Such approaches could be particularly valuable in diagnosing conditions that have subtle ECG presentations but are more apparent in imaging studies.

Moreover, adapting the model for real-time analysis and deployment on edge devices could facilitate its integration into clinical workflows. Implementing efficient algorithms that maintain accuracy while reducing computational requirements would make the model more accessible in settings with limited resources.

Lastly, conducting extensive validation studies in diverse clinical environments is essential to assess the model's performance across different populations. Collaborations with hospitals to carry out prospective studies would help in evaluating the model's effectiveness in routine clinical practice and could provide opportunities to gather feedback from clinicians to further refine the model.

5.3. Conclusion

Our findings suggest that this study contributes to the field of automated ECG analysis through a Transformer-based model that brings together morphological and temporal features by incorporating RR interval information. The model not only achieves high accuracy in diagnosing myocardial infarction but also enhances interpretability through lead importance analysis. By aligning the model's learned features with clinical insights, we have taken a crucial step toward creating AI systems that are both accurate and trustworthy.

The findings demonstrate that advanced deep learning models, when carefully designed and augmented with domain-specific knowledge, can outperform traditional methods in complex tasks like ECG classification. The integration of RR intervals proves to be a valuable enhancement, capturing the temporal dynamics of cardiac rhythms that are essential for accurate diagnosis.

This work aims to contribute to research on developing interpretable AI models in cardiology. By addressing both the performance and transparency of the model, we contribute to the ongoing efforts to safely and effectively incorporate AI into healthcare settings. The ultimate goal is to assist clinicians in making timely and accurate diagnoses, thereby improving patient outcomes in cardiovascular diseases.

REFERENCES

- [1] E. J. Benjamin *et al.*, “Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association,” *Circulation*, vol. 139, no. 10, pp. e56–e528, Mar. 2019, doi: 10.1161/CIR.0000000000000659.
- [2] “Global health estimates: Leading causes of death.” Accessed: Oct. 08, 2024. [Online]. Available: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghel-leading-causes-of-death>
- [3] B. Vogel *et al.*, “ST-segment elevation myocardial infarction,” *Nature Reviews Disease Primers* 2019 5:1, vol. 5, no. 1, pp. 1–20, Jun. 2019, doi: 10.1038/s41572-019-0090-3.
- [4] “12-Lead ECG Placement Guide - CardiacDirect.” Accessed: Dec. 25, 2024. [Online]. Available: <https://www.cardiacdirect.com/12-lead-ecg-placement-guide/>
- [5] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Future Healthc J*, vol. 6, no. 2, p. 94, Jun. 2019, doi: 10.7861/FUTUREHOSP.6-2-94.
- [6] “(1) (PDF) A survey of noise removal techniques for ecg signals.” Accessed: Apr. 22, 2024. [Online]. Available: https://www.researchgate.net/publication/303155606_A_survey_of_noise_removal_techniques_for_ecg_signals
- [7] “Artificial Intelligence in Healthcare Market Size, Report 2034.” Accessed: Oct. 08, 2024. [Online]. Available: <https://www.precedenceresearch.com/artificial-intelligence-in-healthcare-market>
- [8] A. Rajkomar *et al.*, “Scalable and accurate deep learning with electronic health records,” *npj Digital Medicine* 2018 1:1, vol. 1, no. 1, pp. 1–10, May 2018, doi: 10.1038/s41746-018-0029-1.
- [9] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nat Med*, doi: 10.1038/s41591-018-0300-7.
- [10] “521 FDA Authorized AI-Enabled Medical Devices | LinkedIn.” Accessed: Oct. 08, 2024. [Online]. Available: <https://www.linkedin.com/pulse/fda-publishes-updated-list-521-authorized-aiml-margaretta-colangelo/>
- [11] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, “Explainable AI for the Medical Domain What do we need to build explainable AI systems for the medical domain?”, Accessed: Apr. 22, 2024. [Online]. Available: <https://www.computerworld.com.au/article/617359>
- [12] E. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 11, p. 4793, 2021, doi: 10.1109/TNNLS.2020.3027314.
- [13] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
- [14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 2018.

- [15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, 2016.
- [17] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychol Rev*, vol. 65, no. 6, p. 386, 1958.
- [18] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.
- [19] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [22] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med Image Anal*, vol. 42, pp. 60–88, 2017.
- [23] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [24] D. Castelvechi, "Can we open the black box of AI?," *Nature News*, vol. 538, no. 7623, p. 20, 2016.
- [25] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer Nature, 2019.
- [26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [27] Y. LeCun and others, "Backpropagation applied to handwritten zip code recognition," *Neural Comput*, vol. 1, no. 4, pp. 541–551, 1989.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [29] G. Litjens and others, "A survey on deep learning in medical image analysis," *Med Image Anal*, vol. 42, pp. 60–88, 2017.
- [30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [31] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.
- [32] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

- [33] S. Mallat, *A wavelet tour of signal processing: The sparse way*. Academic press, 2008.
- [34] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning*, 2010, pp. 807–814.
- [35] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *Proceedings of 2010 IEEE international symposium on circuits and systems*, 2010, pp. 253–256.
- [36] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [38] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [39] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*, 2015, pp. 448–456.
- [40] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [41] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” *IEEE Trans Biomed Eng*, vol. 63, no. 3, pp. 664–675, 2016.
- [42] O. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Comput Biol Med*, vol. 96, pp. 189–202, 2018.
- [43] U. R. Acharya and others, “A deep convolutional neural network model to classify heartbeats,” *Comput Biol Med*, vol. 89, pp. 389–396, 2017.
- [44] M. Zhu and others, “ECG-based heartbeat classification using convolutional neural networks,” *Future Generation Computer Systems*, vol. 102, pp. 569–576, 2020.
- [45] P. Rajpurkar and others, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat Med*, vol. 25, no. 1, pp. 65–69, 2019.
- [46] A. Y. Hannun and others, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat Med*, vol. 25, no. 1, pp. 65–69, 2019.
- [47] J. H. Park and others, “Automated detection and classification of the proximal LAD artery stenosis using a deep learning approach on 12-lead electrocardiograms,” *Sci Rep*, vol. 11, no. 1, pp. 1–11, 2021.
- [48] S. Mousavi and F. Afghah, “Inter-and intra-patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach,” in *ICASSP 2019-*

2019 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1308–1312.

[49] H. Duan and others, “Automatic sleep stage classification with single-channel EEG using convolutional neural networks,” in *2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2017, pp. 2844–2847.

[50] C. Chen and others, “Deep learning for cardiac image segmentation: A review,” *Front Cardiovasc Med*, vol. 7, p. 25, 2020.

[51] Y. Xia and others, “An automatic cardiac arrhythmia classification system with wearable electrocardiogram,” *IEEE Access*, vol. 6, pp. 16529–16538, 2018.

[52] B. Hou and others, “FUAD-Net: Fully unsupervised adversarial domain adaptation network for ECG signal classification,” *IEEE J Biomed Health Inform*, vol. 25, no. 6, pp. 2041–2051, 2021.

[53] G. D. Clifford and others, “AF classification from a short single lead ECG recording: The PhysioNet/Computing in Cardiology Challenge 2017,” in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[54] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/NECO.1997.9.8.1735.

[55] A. Graves, A. R. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 6645–6649, Oct. 2013, doi: 10.1109/ICASSP.2013.6638947.

[56] F. A. Gers and J. Schmidhuber, “Recurrent nets that time and count,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 3, pp. 189–194, 2000, doi: 10.1109/IJCNN.2000.861302.

[57] K. Cho *et al.*, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734, Jun. 2014, doi: 10.3115/v1/d14-1179.

[58] D. Amodei *et al.*, “Deep Speech 2: End-to-End Speech Recognition in English and Mandarin,” *33rd International Conference on Machine Learning, ICML 2016*, vol. 1, pp. 312–321, Dec. 2015, Accessed: Oct. 15, 2024. [Online]. Available: <https://arxiv.org/abs/1512.02595v1>

[59] A. Vaswani *et al.*, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[60] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[61] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv preprint arXiv:1409.0473*, 2014.

[62] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” *arXiv preprint arXiv:1607.06450*, 2016.

- [63] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [64] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [65] S. Wang, B. Li, M. Khabsa, H. Fang, and H. Ma, “Linformer: Self-attention with linear complexity,” *arXiv preprint arXiv:2006.04768*, 2020.
- [66] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V Le, and R. Salakhutdinov, “Transformer-XL: Attentive language models beyond a fixed-length context,” *arXiv preprint arXiv:1901.02860*, 2019.
- [67] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018.
- [68] S. Sukhbaatar, E. Grave, P. Bojanowski, and A. Joulin, “Adaptive attention span in transformers,” *arXiv preprint arXiv:1905.07799*, 2019.
- [69] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontanon, “FNet: Mixing tokens with Fourier transforms,” *arXiv preprint arXiv:2105.03824*, 2021.
- [70] Z. Liu *et al.*, “Swin transformer: Hierarchical vision transformer using shifted windows,” *arXiv preprint arXiv:2103.14030*, 2021.
- [71] X. Li, T. K. L. Wong, R. T. Q. Chen, and D. Duvenaud, “Scalable gradients for stochastic differential equations,” *arXiv preprint arXiv:2001.01328*, 2020.
- [72] P. Laguna, R. Jané, and P. Caminal, “Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database,” *Computers and Biomedical Research*, vol. 27, no. 1, pp. 45–60, 1994, doi: 10.1006/cbmr.1994.1006.
- [73] N. V. Thakor and Y. S. Zhu, “Applications of Adaptive Filtering to ECG Analysis: Noise Cancellation and Arrhythmia Detection,” *IEEE Trans Biomed Eng*, vol. 38, no. 8, pp. 785–794, 1991, doi: 10.1109/10.83591.
- [74] S. Celin and K. Vasanth, “ECG Signal Classification Using Various Machine Learning Techniques,” *J Med Syst*, vol. 42, no. 12, pp. 1–11, Dec. 2018, doi: 10.1007/S10916-018-1083-6/FIGURES/19.
- [75] F. M. Dias, H. L. M. Monteiro, T. W. Cabral, R. Naji, M. Kuehni, and E. J. da S. Luz, “Arrhythmia classification from single-lead ECG signals using the inter-patient paradigm,” *Comput Methods Programs Biomed*, vol. 202, p. 105948, Apr. 2021, doi: 10.1016/J.CMPB.2021.105948.
- [76] U. R. Acharya *et al.*, “A deep convolutional neural network model to classify heartbeats,” *Comput Biol Med*, vol. 89, pp. 389–396, Oct. 2017, doi: 10.1016/J.COMPBIOMED.2017.08.022.

- [77] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Arrhythmia detection and classification using morphological and dynamic features of ECG signals," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1918–1921, 2012.
- [78] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed Signal Process Control*, vol. 8, no. 5, pp. 437–448, 2013.
- [79] Y. Li, Y. Pei, and Y. Shen, "Deep learning for ECG segmentation," *arXiv preprint arXiv:1905.04288*, 2019.
- [80] W. Ullah, I. Siddique, R. M. Zulqarnain, M. M. Alam, I. Ahmad, and U. A. Raza, "Classification of Arrhythmia in Heartbeat Detection Using Deep Learning," *Comput Intell Neurosci*, vol. 2021, no. 1, p. 2195922, 2021, doi: <https://doi.org/10.1155/2021/2195922>.
- [81] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat Med*, vol. 25, no. 1, pp. 65–69, 2019.
- [82] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," *Knowl Based Syst*, vol. 132, pp. 62–71, 2017.
- [83] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 443–444.
- [84] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput Biol Med*, vol. 102, pp. 278–287, 2018.
- [85] P. Warrick and M. N. Homsy, "Ensembling convolutional and long short-term memory networks for electrocardiogram arrhythmia detection," *Physiol Meas*, vol. 39, no. 11, p. 114002, 2018.
- [86] A. Vaswani *et al.*, "Attention Is All You Need," *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: Dec. 12, 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762v7>
- [87] G. Yan, S. Liang, Y. Zhang, and F. Liu, "Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification," *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pp. 898–905, Nov. 2019, doi: 10.1109/BIBM47256.2019.8983326.
- [88] X. Li, W. Cai, X. He, H. Ma, and X. Zhang, "Multi-Lead ECG Classification with Attention-Based Hierarchical Transformer," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 5766–5769.
- [89] C. Che, C. Xiao, J. Liang, B. Jin, J. Zho, and F. Wang, "A hybrid transformer network for ECG arrhythmia classification," in *2021 Computing in Cardiology (CinC)*, 2021, pp. 1–4.

- [90] S. Raghunath *et al.*, “Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network,” *Nat Med*, vol. 26, no. 6, pp. 886–891, 2020.
- [91] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, “Deep learning for ECG analysis: Benchmarks and insights from PTB-XL,” *IEEE J Biomed Health Inform*, vol. 25, no. 5, pp. 1519–1528, 2021.
- [92] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, “Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals,” *Inf Sci (N Y)*, vol. 415, pp. 190–198, 2017.
- [93] S. Mousavi, A. Fotoohinasab, and F. Afghah, “Multi-task learning for ECG-based biometrics,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 2419–2423.
- [94] P. Sarkar and A. Etemad, “Self-supervised learning for ECG-based emotion recognition,” in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 3217–3221.
- [95] J. Liu, X. Yuan, and C. Wang, “Federated learning for ECG classification with deep neural networks,” *Comput Biol Med*, vol. 135, p. 104569, 2021.
- [96] C. Sirocchi, A. Bogliolo, and S. Montagna, “Medical-informed machine learning: integrating prior knowledge into medical decision systems,” *BMC Med Inform Decis Mak*, vol. 24, no. Suppl 4, pp. 1–18, Dec. 2024, doi: 10.1186/S12911-024-02582-4/TABLES/3.
- [97] G. H. Tison *et al.*, “Passive detection of atrial fibrillation using a commercially available smartwatch,” *JAMA Cardiol*, vol. 3, no. 5, pp. 409–416, 2018.
- [98] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nature Medicine* 2019 25:1, vol. 25, no. 1, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.
- [99] Z. I. Attia *et al.*, “An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction,” *The Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.
- [100] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, “ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network,” *Artif Intell Med*, vol. 106, p. 101856, 2020.
- [101] S. Nemati *et al.*, “Monitoring and detecting atrial fibrillation using wearable technology,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3394–3397.
- [102] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, “Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network,” *Information Fusion*, vol. 53, pp. 174–182, 2020.
- [103] Y. Li, H. Chen, L. Zheng, and H. Li, “Quantum machine learning algorithm for knowledge graphs,” *Phys Rev A (Coll Park)*, vol. 103, no. 3, p. 32419, 2021.

- [104] G. Yan, S. Liang, Y. Zhang, and F. Liu, "Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification," *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pp. 898–905, Nov. 2019, doi: 10.1109/BIBM47256.2019.8983326.
- [105] "PTB Diagnostic ECG Database v1.0.0." Accessed: Jul. 29, 2024. [Online]. Available: <https://www.physionet.org/content/ptbdb/1.0.0/>
- [106] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1038/S41597-020-0495-6.
- [107] "St Petersburg INCART 12-lead Arrhythmia Database v1.0.0." Accessed: Jul. 29, 2024. [Online]. Available: <https://www.physionet.org/content/incartdb/1.0.0/>
- [108] T. Neela and S. Namburu, "ECG signal classification using capsule neural networks," 2021, doi: 10.1049/ntw2.12018.
- [109] S. Wang, R. Li, X. Wang, S. Shen, B. Zhou, and Z. Wang, "Multiscale Residual Network Based on Channel Spatial Attention Mechanism for Multilabel ECG Classification," *J Healthc Eng*, vol. 2021, p. 6630643, 2021, doi: 10.1155/2021/6630643.
- [110] H. Zhang, W. Zhao, and S. Liu, "SE-ECGNet: A Multi-scale Deep Residual Network with Squeeze-and-Excitation Module for ECG Signal Classification," *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, pp. 2685–2691, Dec. 2020, doi: 10.1109/BIBM49941.2020.9313548.
- [111] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Feb. 2017, Accessed: Oct. 15, 2024. [Online]. Available: <http://arxiv.org/abs/1702.08608>

BIOGRAPHY

Barış ÖZCAN completed his high school education at Özel Bornova Anadolu Lisesi in 2014. He received his B.Sc. in Electrical/Electronics Engineering from İzmir University of Economics in 2019. He is currently pursuing his M.Sc. in Computer Engineering at Gebze Technical University, with an expected graduation in 2024. During his graduate studies, he worked as a Project Researcher and later as a Research Assistant at Gebze Technical University. Currently, he is working as a Large Language Model Research Engineer at Huawei Türkiye.



PUBLICATIONS AND PRESENTATIONS FROM THE THESIS

[1] Barış Özcan, Yakup Genç, Yusuf S. Akgül, "Interpretable Algorithm for Diagnosing Myocardial Infarction from Electrocardiography Data", Sinyal İşleme ve İletişim Uygulamaları Kurultayı(2024).

