Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Master of Science in Artificial Intelligence

Master's Thesis

Peak Attention U-Net: Attention-Enhanced ECG Delineation Model for Peak Detection

Author: Mario Sanz Guerrero
Directors: Víctor Maojo García & Marian Díaz Vicente

Madrid, July 2024

This Master's Thesis has been deposited at the ETSI Informáticos of the Universidad Politécnica de Madrid for its defense.

Master's Thesis
Master of Science in Artificial Intelligence

Title: Peak Attention U-Net: Attention-Enhanced ECG Delineation Model for Peak Detection

July 2024

Author: Mario Sanz Guerrero

Directors: Víctor Maojo García
Biomedical Informatics Group
ETSI Informáticos
Universidad Politécnica de Madrid

Marian Díaz Vicente
Facultad de Informática
Universidad Complutense de Madrid

Acknowledgements

I would like to express my deepest gratitude to all those who have supported, guided, and inspired me throughout the completion of this Master's Thesis.

First of all, I would like to express my sincere thanks to my directors, Dr. Víctor Maojo and Dr. Marian Díaz. Their expertise, patience and support have been fundamental in shaping this work. Their insightful comments and guidance have not only improved the quality of this thesis, but also fostered my growth as a researcher.

I am very grateful to cardiology professionals Mauro Buelga (Hospital Universitario Ramón y Cajal, Madrid) and Gonzalo Alonso (Hospital Universitario de Navarra, Pamplona) for their invaluable help with the medical concepts and the validation of the results. Their contributions have ensured the medical relevance and accuracy of this work.

Thanks to Professors Joaquín Recas and Luis Piñuel for their technical expertise and dedication to this project. Their commitment and advice have greatly enriched this research.

I thank the Common Management Solutions team for their financial support. Their confidence in the potential of this project has made this research possible, and their investment in innovative healthcare solutions enables the real-world impact of this work.

To my colleague and friend, Sergio González, thank you for being there every step of the way. Your support, technical discussions and shared enthusiasm have made this journey more enjoyable.

To my parents, César and Mari Jose, and my sister Lucía, thank you for your endless love, sacrifices, and support. You have always encouraged me to pursue my dreams, and for that I will be eternally grateful.

To my partner in life, Nadia, thank you for your love, patience, and support. Your belief in me has been my greatest motivation. *Molto più di molto.*

Finally, to my grandparents. You have always been present, and I know you are very proud.

Thank you all for being part of this journey.

Abstract

According to recent statistics from the World Health Organization, cardiovascular diseases are the leading cause of death globally, highlighting the need for advanced diagnostic tools. An electrocardiogram (ECG) is a non-invasive test that records the electrical activity of the heart, providing essential information for diagnosing cardiac conditions. ECG delineation refers to the process of identifying specific points of interest (fiducial points) in the ECG signal, such as the onsets, peaks and offsets of the P, QRS, and T waves, which are crucial for accurate interpretation and diagnosis.

While traditional ECG delineation methods such as wavelet transforms and rule-based algorithms have been widely used, they often struggle with noise and morphological variations. Recent deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise, but lack accuracy in detecting smaller regions.

This Master's Thesis proposes a novel deep learning model that addresses several limitations of existing models and facilitates their future clinical use with wearable devices. The proposed attention-enhanced U-Net model (Peak Attention U-Net) incorporates an attention mechanism into the U-Net architecture, allowing it to focus on small yet clinically relevant regions of the ECG signal, such as peaks. The Peak Attention U-Net achieves state-of-the-art performance in fiducial point delineation, particularly excelling in the detection of P, R, and T peaks. Compared to recent and relevant deep learning approaches, it demonstrates a remarkable improvement of over 12.5% in F1 score for peak detection. The model's performance was evaluated using the widely recognized PhysioNet LUDB database, ensuring a robust comparison with existing methods.

Key advantages of the proposed model include its ability to detect peaks, process entire ECG signals, its lead-agnostic design, and its computational efficiency. These features make it a versatile and practical tool for various clinical scenarios, including both offline analysis and potential real-time monitoring applications.

Experimental results demonstrate the potential of the developed framework for real clinical use. By addressing the limitations of existing methods and offering improved performance, this research contributes to the advancement of ECG delineation and opens possibilities for early diagnosis beyond traditional medical settings with the use of wearable devices. Future work will focus on evaluating the model's generalizability across diverse datasets and in real-world clinical settings to confirm its clinical utility.

Keywords: ECG Delineation, Attention Mechanism, U-Net, Peak Detection, Cardiac Diagnosis, Deep Learning, Artificial Intelligence, Biomedical Signal Processing.

Resumen

Según estadísticas recientes de la Organización Mundial de la Salud, las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial, lo que destaca la necesidad de herramientas de diagnóstico avanzadas. El electrocardiograma (ECG) es una prueba no invasiva que registra la actividad eléctrica del corazón, proporcionando información esencial para el diagnóstico de afecciones cardiacas. La delineación del ECG se refiere al proceso de identificación de puntos de interés específicos (puntos fiduciales) en la señal del ECG, como los comienzos, picos y finales de las ondas P, QRS y T, que son cruciales para una interpretación y diagnóstico precisos.

Mientras que los métodos tradicionales de delineación del ECG, como las transformadas wavelet y los algoritmos basados en reglas, han sido ampliamente utilizados, a menudo tienen dificultades con el ruido y las variaciones morfológicas. Los enfoques recientes de aprendizaje profundo, incluyendo las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), han mostrado promesa, pero carecen de precisión en la detección de regiones más pequeñas.

Este Trabajo de Fin de Máster propone un modelo novedoso de aprendizaje profundo que aborda varias limitaciones de los modelos existentes y facilita su uso clínico futuro con dispositivos wearables. El modelo U-Net mejorado con atención propuesto (Peak Attention U-Net) incorpora un mecanismo de atención en la arquitectura U-Net, permitiéndole enfocarse en regiones pequeñas pero clínicamente relevantes de la señal ECG, como los picos. El modelo Peak Attention U-Net logra un rendimiento del estado del arte en la delineación de puntos fiduciales, destacando particularmente en la detección de picos P, R y T. En comparación con enfoques recientes y relevantes de aprendizaje profundo, demuestra una mejora notable de más del 12,5% en la puntuación F1 para la detección de picos. El rendimiento del modelo fue evaluado utilizando la ampliamente reconocida base de datos LUDB de PhysioNet, asegurando una comparación robusta con los métodos existentes.

Las ventajas clave del modelo propuesto incluyen su capacidad para detectar picos, procesar señales completas de ECG, su diseño agnóstico de la derivación y su eficiencia computacional. Estas características lo convierten en una herramienta versátil y práctica para varios escenarios clínicos, incluyendo tanto el análisis offline como potenciales aplicaciones de monitoreo en tiempo real.

Los resultados experimentales demuestran el potencial del marco desarrollado para su uso clínico real. Al abordar las limitaciones de los métodos existentes y ofrecer un mejor rendimiento, esta investigación contribuye al avance de la delineación de ECG y abre posibilidades para el diagnóstico precoz más allá de los entornos médicos

tradicionales con el uso de tecnología wearable. El trabajo futuro se centrará en evaluar la capacidad de generalización del modelo en diferentes conjuntos de datos y en entornos clínicos reales para confirmar su utilidad clínica.

Palabras clave: Delineación de ECG, Mecanismo de Atención, U-Net, Detección de Picos, Diagnóstico Cardíaco, Aprendizaje Profundo, Inteligencia Artificial, Procesamiento de Señales Biomédicas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recent statistics from the World Health Organization highlight cardiovascular disease as the leading cause of unnatural death, with significant mortality and disability rates [1]. The heart operates through a coordinated process of atrial depolarization followed by ventricular repolarization. To monitor this crucial activity, the electrocardiogram (ECG) is an effective and non-invasive diagnostic tool. It captures the electrical activities of the heart by detecting the propagation of electrical signals within the heart through electrodes placed on the extremities and on the thoracic surface. This method provides valuable information about the functioning of the heart, enabling early detection and treatment of cardiovascular conditions.

An ECG is traditionally composed of 12 leads, each of which provides a unique view of the electrical activity of the heart from various angles around the body. These leads are categorized into two main types: limb leads and precordial (chest) leads. The limb leads, comprising leads I, II, III, aVR, aVL, and aVF, focus on the heart's electrical activity from the frontal plane, providing a view of the superior-inferior axis of the heart. On the other hand, the precordial leads, called V1 to V6, are placed across the chest and provide a detailed view of the heart's electrical movement in the horizontal plane. The data from these 12 leads are compiled to form a complete ECG trace, as shown in Figure 1.1, which allows healthcare professionals to interpret the heart's three-dimensional electrical activity in a two-dimensional representation.

Each lead of an ECG records the electrical activity of the heart over several heartbeats, allowing accurate identification and analysis of variations in heart rhythm and

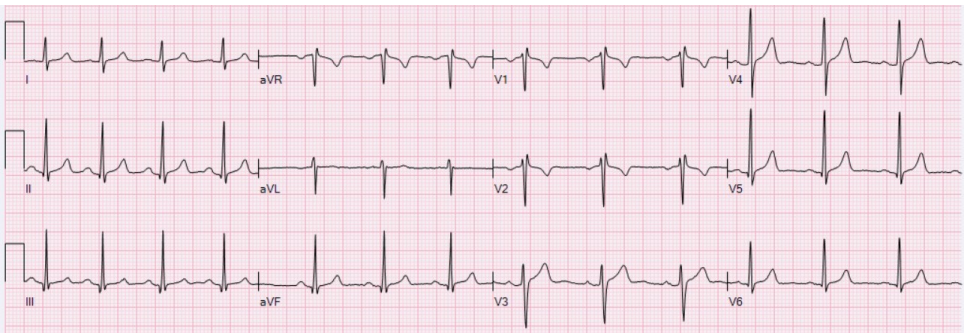


Figure 1.1: 12-lead ECG signal. Source: [2].

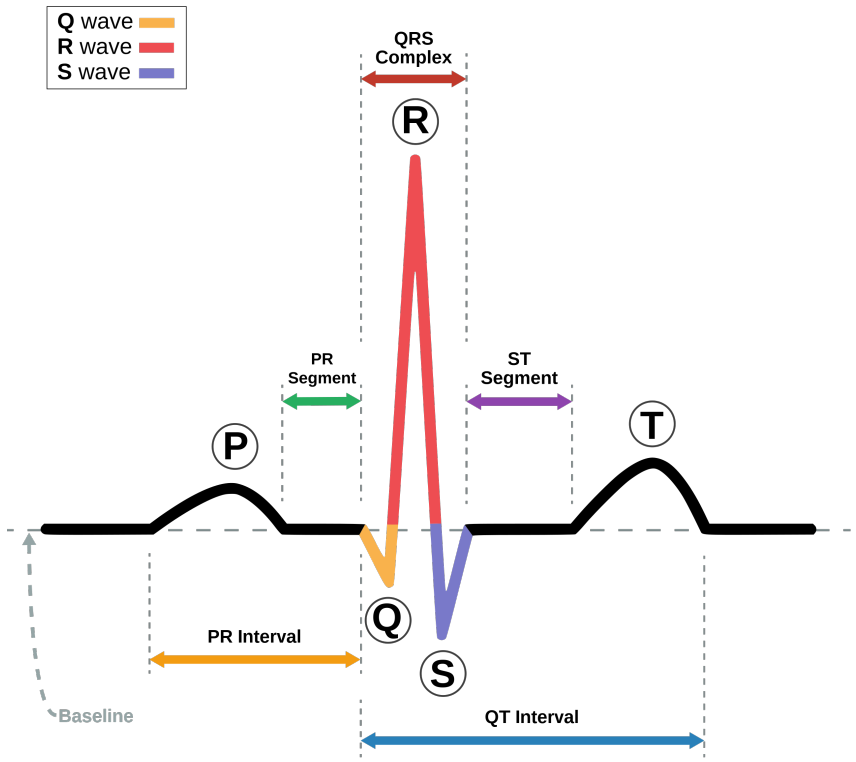


Figure 1.2: Delineated heartbeat with its waves, peaks, segments and intervals. Source: [3].

pattern. This sequential recording is vital for detecting irregularities that could indicate heart conditions such as arrhythmias, ischaemia or myocardial infarction.

The morphology of each heartbeat is composed of a series of waves and intervals, specifically the P wave, QRS complex, T wave, and the intervals between these components. The P wave represents atrial depolarization, the process by which the atria contract to push blood into the ventricles. Following the P wave, the QRS complex appears as a sharp series of peaks, representing ventricular depolarization, which is the contraction of the ventricles to push blood out to the lungs and the rest of the body. This phase is crucial, as it reflects the main pumping action of the heart. Finally, the T wave indicates ventricular repolarization, a recovery period in which the ventricles prepare for the next heartbeat. Figure 1.2 shows a heartbeat labeled with its waves, peaks, segments and intervals.

The main characterization of these waves are its onsets (beginnings), peaks, and offsets (endings), which describe the intervals between waves. These are known as fiducial points. For example, the PR interval measures the time between the onset of the P wave and the onset of the QRS complex, and it helps in detecting atrioventricular blocks [4]. The QT interval, which spans from the start of the QRS complex to the end of the T wave, reflects the total time for ventricular depolarization and repolarization, with variations indicating potential electrolyte imbalances, medications effects, or genetic conditions [5]. The morphology of each component and interval on the ECG is crucial for diagnosing and understanding a wide range of cardiac conditions.

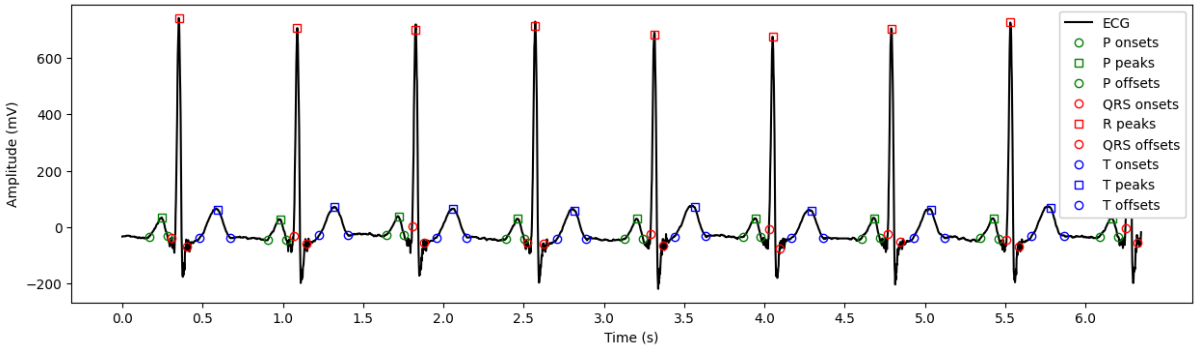


Figure 1.3: Delineated ECG with onsets, peaks and offsets of the waves.

ECG delineation refers to the detailed analysis and interpretation of these waves, intervals, and complexes recorded on an ECG trace. This process is essential for the precise assessment of heart rhythm, the identification of cardiac abnormalities, and the evaluation of the heart's overall health. ECG delineation involves the identification of the onset, peak, and offset of each wave and interval, allowing healthcare professionals to detect subtle changes that might indicate underlying cardiac conditions. Figure 1.3 shows a delineated ECG signal.

The rise of wearable devices has highlighted the importance of automatic ECG delineation as a tool for managing the growing data from continuous heart monitoring. Figure 1.4 shows the procedure to record different ECG leads using a smartwatch [6]. Automatic ECG delineation algorithms are designed to accurately identify the critical points in the ECG waveform (i.e., onset, peak, and offset of the P, QRS, and T waves) without the need for manual intervention. By automating this process, large volumes of ECG data can be processed quickly and accurately, making it easier for healthcare professionals to identify abnormal heart rhythms and diagnose conditions.

The applications of automatic ECG delineation goes beyond personal use, making cardiac monitoring feasible in settings where traditional electrocardiographs are impractical due to their size and cost. Advances in wearable technology allow for continuous heart monitoring in diverse locations, including remote clinics and ambulances, improving patient care by offering vital cardiac data anytime, anywhere.

Given the large and growing availability of ECG data and the recent advances in deep learning models, it seems appropriate to use these models to automatically delineate ECGs. This could solve the problem of the increasing volume of data generated but the lack of medical staff capacity to analyze it. However, the complex nature of ECG data poses a challenge, as even two expert cardiologists may delineate and interpret the same ECG differently. This variation in manual annotations is a challenge for the development of deep learning algorithms that require consistent data for training.

1.1 Objectives

The proposed research addresses the challenge of ECG signal delineation. The main objective is to develop a deep learning-based solution to segment ECG signals at critical points and regions, so that they can be efficiently analyzed by medical professionals. The ultimate goal is to create a robust ECG delineation framework that offers

Figure 1.4: The positions of the smartwatch to obtain 9-lead ECGs. Source: [6].

similar performance to the state-of-the-art methods. However, state-of-the-art methods are not applicable due to certain limitations that will be explained later. Instead, the framework must be able to perform effectively in real clinical scenarios.

The specific objectives of this work are:

- Develop a deep learning model for ECG delineation that can be applied to real-world clinical scenarios by addressing the limitations of current state-of-the-art models. To achieve this, the model must be able to detect the P, QRS, and T peaks, delineate entire ECG signals, and delineate ECG signals from any lead.
- Evaluate the performance of the proposed model and compare it with the state-of-the-art models. Assess the model's efficiency and generalizability across diverse patients and conditions.
- Analyze the potential clinical applications of the proposed model and its impact on the field of cardiology.

1.2 Hypothesis

This research hypothesizes that incorporating an attention mechanism into a U-Net-based deep learning model will significantly improve ECG delineation by enhancing peak detection capabilities, a limitation in current state-of-the-art deep learning approaches. Additionally, it posits that such models can be trained to accurately delineate complete ECG signals from any individual lead, thereby enabling the delineation of full 12-lead ECGs.

1.3 Research Methodology

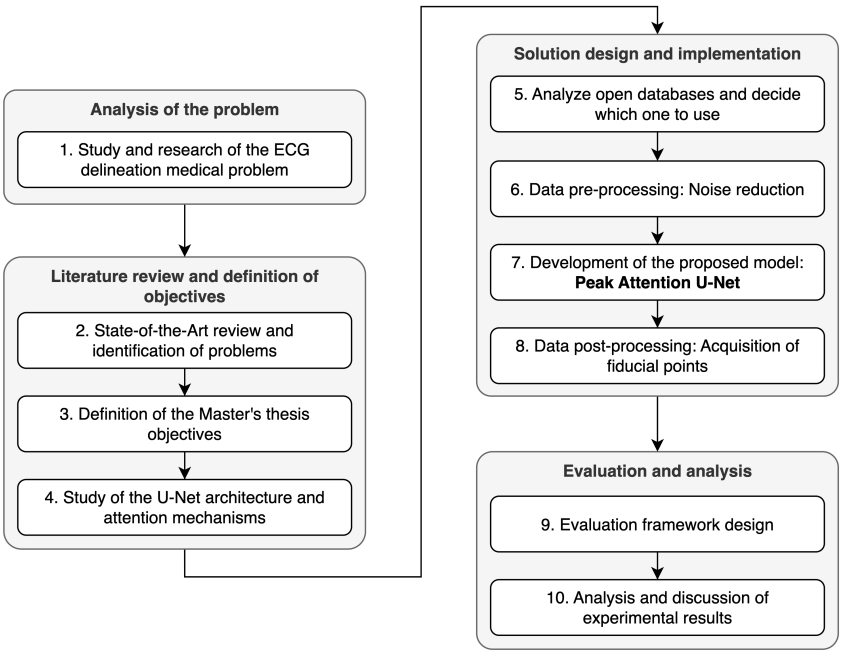


Figure 1.5: Flowchart of the employed methodology.

In order to achieve the objectives outlined in this work, a structured methodology is followed, as shown in Figure 1.5. The methodology consists of four main stages:

1. **Analysis of the problem**: Given the high medical complexity of the problem, the initial phase requires a thorough research on the subject. With the collaboration of health professionals Gonzalo Alonso[1] (Hospital Universitario de Navarra, Pamplona) and Mauro Buelga[2] (Hospital Universitario Ramón y Cajal, Madrid), the different concepts of cardiology are studied, including the delineation and the reasons for its importance and applications.

2. **Literature review and definition of the objectives**: A comprehensive review of the state-of-the-art in automatic ECG delineation is conducted to gain the necessary knowledge for addressing the problem. Through this review, unresolved issues in the literature are identified, which prevent the application of current models in real clinical scenarios. Based on these findings, the objectives of this work are defined to address these limitations and bridge the gap between research and clinical practice. A deep study of the U-Net model and the attention mechanism is carried out given their relevance to the proposed solution.

3. **Solution design and implementation**: In this stage, the proposed solution is conceived, designed and implemented. This includes investigating different databases and selecting the most suitable one for the study. The main innovation is the integration of the attention mechanism into the 1D U-Net model to solve the peak detection problem, leading to the development of the Peak Attention U-Net. The entire experimental pipeline is designed, including preprocessing and post-processing of the data.

[1]Gonzalo Alonso's ORCID: https://orcid.org/0000-0001-5914-8526
[2]Mauro Buelga's ORCID: https://orcid.org/0000-0002-6325-357X

4. **Evaluation and analysis**: After implementing the solution, the model's performance is comprehensively evaluated to verify that it achieves the proposed objectives. It is benchmarked against recent and relevant studies, with a focus on its high performance in peak detection. The efficiency and generalizability of the model in delineating ECG signals across a diverse set of patients and conditions are assessed. Finally, the results are analyzed in depth, and the potential clinical applications of the model are explored.

1.4 Organization of the Document

The document is organized as follows: Chapter 2 provides a comprehensive review of the state-of-the-art in automatic ECG delineation, identifying the limitations of current models and the rationale for this research. Chapter 3 presents the theoretical foundations needed to fully understand the research presented. Chapter 4 describes the dataset used and the proposed solution, introducing the Peak Attention U-Net model in detail. Chapter 5 discusses the results of the proposed model, comparing its performance against state-of-the-art models, and highlighting the improvements achieved by this approach. Finally, Chapter 6 concludes the study by summarizing key findings and contributions, discussing implications, and suggesting avenues for future research in the intersection of artificial intelligence and ECG analysis.

1.5 Relation with Master's Degree subjects

This Master's Thesis is primarily related to the following subjects within the master's degree program.

On a theoretical level, the most relevant subject is "Biomedical Informatics". This work revolves around electrocardiogram signals, which represent the activity of the heart and are considered biomedical data, a key topic studied in this course.

On the technical side, "Computer Vision" is the most pertinent subject. The knowledge gained in this course about convolutional neural networks has been crucial for developing the proposed model for ECG delineation. Specifically, the U-Net model, studied in the context of semantic image segmentation, served as the foundation for the proposed model in this work. This course also provided essential insights into the general workflow of machine learning projects, covering fundamental concepts such as overfitting, regularization, and model evaluation, among others. Additionally, the "Artificial Neural Networks and Deep Learning" course offered a deep dive into these fundamental concepts, which are critical for any artificial intelligence project.

The most significant improvement in the proposed model is the incorporation of an attention mechanism, the basis of which was studied in the course "Deep Learning for Natural Language Processing". The knowledge from this course inspired the idea to integrate the attention mechanism, enhancing the detection of small but crucial segments in the ECG signal, such as peaks.

In general, all the subjects of the master's degree have been valuable for this research. The constant reading of academic papers and the study of the scientific method throughout the master's program have provided me with the methodology that has led to the success of this work.

Chapter 2

Background and Rationale

This chapter provides a foundation for understanding previous work on ECG signal delineation to better frame and contextualize the objectives of this study. It begins with a review of the state-of-the-art, followed by an identification of the main unresolved issues in the literature. Finally, the specific objectives that this work aims to achieve are outlined.

2.1 State-of-the-Art Review

This section presents a comprehensive review of state-of-the-art techniques in automatic ECG delineation. It begins with an overview of classical techniques, followed by a discussion on deep learning-based methods. These include both sequence modeling approaches and those based on the encoder-decoder architecture, which has gained particular relevance in recent years for segmentation tasks.

2.1.1 Traditional approaches for ECG Delineation

The development of ECG delineation methods has historically focused on creating rule-based systems for identifying and locating the QRS complex. A notable early contribution in this area was made by Pan and Tompkins [7], who used slope, amplitude, and width information to detect the QRS complex.

Since then, more advanced techniques have been employed to identify not only the QRS complex but also the P and T waves. Digital signal processing techniques like wavelet transforms [8, 9, 10], Hilbert transforms [11, 12], and phasor transforms [13] have all been applied in this context. Additionally, classical machine learning approaches such as hidden Markov models [14, 15] and Gaussian mixture models [16] have also been used.

Wavelet-based methods have emerged as a state-of-the-art approach for ECG delineation, as demonstrated by their performance on publicly available datasets like QTDB [17]. In recent years, the wavelet-based algorithm proposed by Kalyakulina et al. [10] has been validated using the LUDB dataset [18], further solidifying its effectiveness.

A relevant study of Rincón et al. [19] proposed a real-time delineation algorithm based on digital wavelet transforms, emphasizing the importance of embedded signal

filtering to compensate noise that appears during ambulatory monitoring. Building on this method, a new modular, low-complexity algorithm for real-time ECG wave delineation was introduced by Bote et al. [20]. Optimized for resource-constrained embedded systems, this algorithm adapts its operation between an ultralow-power mode with low-frequency sampling and a high-accuracy mode with high-frequency sampling based on arrhythmic activity detection. However, such methods can be challenging to tune, are susceptible to noise, and the tuning directly affects the quality of the delineation. Additionally, these methods are not easily adaptable to different leads, as they require specific tuning for each lead. Therefore, the development of a robust ECG delineation algorithm is crucial for enhancing detection accuracy given the high data variability among patient ECG recordings, and deep learning approaches are well-suited for this task.

2.1.2 Deep Learning based ECG Delineation

Early applications of artificial intelligence in ECG signal delineation utilized machine learning algorithms such as Support Vector Machines [21] and advanced multi-agent techniques like Particle Swarm Optimization [22]. Additional studies employed a combination of wavelet transforms and fuzzy logic [23], while morphological and heartbeat interval features were leveraged to classify ECGs into various pathologies [24]. These methods were effective for their time, enhancing outcomes beyond traditional processing techniques. However, the inherent complexity and patient-to-patient variability in ECG recordings have posed significant challenges to the efficacy of these machine learning approaches.

In recent years, deep learning has emerged as a promising approach for automating the segmentation of ECG signals. Among deep learning architectures, the literature research can be divided into two subgroups, those that focus on the sequential and periodic aspects of the ECG signals, and those that adopt an encoder-decoder architecture.

2.1.2.1 Sequence-Based Deep Learning

Given the sequence nature of ECG signals, Peimankar and Puthusserypady [25] studied the effectiveness of recurrent neural networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, for automated ECG signal delineation. Given their promising results, Nurmaini et al. [26] proposed a bi-directional LSTM (BiLSTM) classifier for ECG waveform delineation. This model segmented ECG data based on individual heartbeats and achieved state-of-the-art performance. However, one notable limitation of this approach is its dependency on a predefined beat length starting from the P-wave onset, which may pose challenges in practical scenarios.

Later, Peimankar and Puthusserypady [27] introduced the DENS-ECG algorithm, a deep learning model designed for real-time heartbeat segmentation. This algorithm integrates convolutional neural networks (CNNs) with LSTM units to accurately identify the onset and offset of P, QRS, and T waves in ECG signals, yielding highly effective results.

Based on this successful combination of CNNs and LSTMs, Nurmaini et al. [28] recently introduced a novel approach by merging the CNN-LSTM framework with their earlier BiLSTM model, leading to their ConvBiLSTM model. This hybrid model

has achieved the highest performance in sequence-based deep learning techniques for ECG analysis.

2.1.2.2 Encoder-Decoder Approaches

Recent advances in deep learning for ECG signal segmentation have highlighted encoder-decoder models. These architectures excel at encoding raw ECG data to extract relevant features and then decoding them to accurately identify key waveform components, such as P, QRS and T waves. Their ability to handle variable-length sequences and adapt to physiological variations increases their robustness and generalizability. By learning to compress and decompress ECG data, these models can better handle noise and other distortions inherent in real-world data. Researchers have explored various architectures for this task.

In a pioneering study on the use of encoder-decoder architecture for ECG delineation, Moskalenko et al. [29] developed a U-Net-inspired [30] architecture that achieved state-of-the-art performance on LUDB in terms of F1-score, outperforming both deep learning [31] and wavelet-based methods [10]. Later, Jimenez-Perez et al. [32] adapted the U-Net structure to one-dimensional data, while Sereda et al. [31] employed an 8-layer convolutional network to investigate the benefits of ensemble methods versus single-network approaches.

In a similar vein, Jimenez-Perez et al. [33] modified the U-Net for segmentation, with a focus on regularization techniques to compensate for limited training data. Their model demonstrated comparable performance to digital signal processing techniques like wavelet transforms [9] when cross-validated on QTDB. More recently, Chen et al. [34] designed a 1D-U-Net model for classifying heartbeat samples into P, QRS, T, and none categories. By combining this model with a proposed post-processing strategy, the delineation algorithm outperformed other approaches in terms of sensitivity for both QTDB and LUDB.

Londhe et al. [35] proposed a hybrid model based on an encoder-decoder architecture, by combining CNNs with BiLSTM networks. This model leverages the spatial features extracted by the CNN and the temporal dynamics captured by the BiLSTM.

Recently, Liang et al. proposed the ECG_SegNet [36], a delineation model based on an encoder-decoder structure for detecting various heartbeat waveforms. They introduced a standard dilated convolution module into the encoder path to extract more informative features that are relevant to the ECG signal. Additionally, BiLSTMs were incorporated into the bottleneck to obtain a rich set of temporal features, enabling the model to accurately recognize various heartbeat waveforms.

The two latest studies [35, 36] have achieved state-of-the-art results in ECG delineation using deep learning. Consequently, they will serve as primary benchmarks for comparison and evaluation in this study, alongside the seminal work of Moskalenko et al. [29] in utilizing encoder-decoder architectures for ECG delineation. However, these studies share several characteristics that, despite being widely accepted in the literature, limit the practical use of these models in real clinical scenarios. These characteristics will be explained in the following section.

2.2 Issues in the Current Literature

Following an extensive review of the state-of-the-art, three main issues are identified in the current literature on ECG delineation using deep learning:

1. **Lack of peak delineation**. All the deep learning studies analyzed only segment the signals into their P, QRS and T waves. After post-processing, they obtain the onsets and offsets of the waves from the segment boundaries. However, there are no studies that focus on peak delineation, and it is a crucial aspect of useful and meaningful ECG delineation. Correct identification of peaks would allow the use of these models in a real clinical setting.

2. **Lack of whole signal delineation**. Existing deep learning studies for ECG delineation typically preprocess the data by looking at the labels and segmenting the entire signal into individual heartbeats, thereby training their models to delineate single heartbeats. Although this approach results in higher validation performance, it is not possible to apply it in a real clinical scenario, as the recorded signals are not easily segmented into individual heartbeats.

3. **Lead-dependent delineation**. The studies analyzed train their models using data from only one or a few ECG leads. This approach is justified by the significant morphological differences between leads. For instance, the QRS complexes in certain leads may look totally different from the QRS complexes in others (see variations in different leads shown in Figure 1.1). By selecting specific leads for model training, the models achieve better validation performance since they do not need to account for the morphological variations across different leads. However, for an automatic delineation model to be effective in real-world scenarios, it must be capable of delineating any lead, ensuring it can accurately analyze complete 12-lead ECGs.

All these issues have in common that they hinder the practical application of ECG delineation models in real clinical settings.

2.3 Rationale

The automatic delineation of ECG signals is a crucial task in cardiac diagnostics and monitoring. While significant progress has been made in this field, particularly with the advent of deep learning techniques, several critical issues remain unresolved. These include the lack of peak delineation, the need for whole signal delineation, and the limitation of lead-dependent delineation. Addressing these challenges is essential for developing a robust and clinically applicable ECG delineation algorithm.

The rationale for this study is grounded in the necessity to overcome these limitations and enhance the practical utility of ECG delineation models. By focusing on peak delineation, we aim to improve the accuracy and clinical relevance of the models. Addressing the need for whole signal delineation will ensure that the models can be applied to continuous ECG recordings without requiring segmentation into individual heartbeats. Finally, developing a lead-independent delineation model will enable the analysis of complete 12-lead ECGs, making the model more versatile and applicable in diverse clinical scenarios.

By addressing these issues, this study aims to bridge the gap between state-of-the-art

performance in controlled research settings and the practical needs of clinical ECG analysis. The goal is to develop a robust ECG delineation framework that maintains high accuracy while being directly applicable in real-world clinical scenarios, potentially improving the efficiency and accuracy of cardiac diagnostics and monitoring.

2.3.1 Objectives and Requirements of the Proposed Model

To achieve the goals outlined above, the proposed ECG delineation model must meet the following requirements:

- **Peak delineation**: The model must accurately detect the peaks of the P, R, and T waves, which are crucial for diagnosing heart conditions.
- **Whole signal delineation**: The model must delineate the entire ECG signal, avoiding the need for segmenting the signal into individual heartbeats.
- **Lead agnosticism**: The model must be able to process ECG signals from any lead, enabling the delineation of entire 12-lead ECGs.
- **Computational efficiency**: The model must be computationally efficient and lightweight, allowing for real-time processing and deployment in clinical settings with limited resources.
- **Generalizable delineation**: The model must be able to delineate ECG signals from diverse patients and conditions, ensuring its generalizability in clinical practice.

Chapter 3

Theoretical Foundations

This chapter provides an overview of the fundamental concepts necessary for a comprehensive understanding of this research. The discussion begins with an examination of artificial intelligence and neural networks in general, followed by an exploration of more specialized topics, including convolutional neural networks, the U-Net architecture, and attention mechanisms. This step-by-step foundation prepares the reader for the advanced deep learning methods used for ECG delineation in subsequent chapters.

3.1 Introduction to Artificial Intelligence

Artificial intelligence (AI) is a field of computer science focused on the development of computer systems that can perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making. Symbolic AI relies on explicit programming of rules and symbols to solve problems. This approach is prominent in fields like expert systems and knowledge representation, where logic and symbolic reasoning play crucial roles. Conversely, subsymbolic AI focuses on learning patterns and representations from data, often without explicit programming of rules. Subsymbolic AI excels in fields like computer vision, natural language processing, and robotics, where tasks are complex and require pattern recognition and adaptation.

Within subsymbolic AI, Machine Learning (ML) is a fundamental subset that focuses on the development of algorithms and statistical models that enable computers to perform specific tasks without using explicit instructions. Instead, these systems learn and make decisions based on data. ML can be further divided into three main types: supervised learning, where the model is trained on labeled data; unsupervised learning, where the model learns from data without labels; and reinforcement learning, where an agent learns to behave in an environment by performing actions and seeing the results.

Deep Learning [37], a subset of machine learning, involves neural networks with many layers. These neural networks are called “deep” because they have more than one hidden layer, allowing them to model complex patterns and perform advanced tasks like image and speech recognition more effectively than traditional machine learning approaches. Deep learning models are particularly powerful in handling

large and complex datasets, extracting patterns and features without needing manually extracted features.

3.2 Introduction to Neural Networks

A neural network consists of layers of interconnected neurons, which are basic computational units. The network typically includes an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to several neurons in the preceding and succeeding layers. These connections are weighted and represent the strength or importance of the connection in processing the input data.

The input layer receives raw data, which is then processed through successive layers where each neuron computes a weighted sum of its inputs and then applies a nonlinear function, often called an activation function. Common activation functions include the sigmoid, tanh, and ReLU (Rectified Linear Unit), each providing different properties that affect the learning dynamics of the network.

The primary goal of a neural network in processing ECG signals or any other form of data is to accurately predict an output or classify data points. This is achieved through the training process, where the network iteratively adjusts the weights of the connections to minimize the difference between the predicted output and the true output (often provided as part of labeled training data). This process is guided by an optimization algorithm, with gradient descent being one of the most common. During each iteration of training, gradients of a loss function (a measure of prediction error) are computed to update the weights in a direction that minimally decreases the error.

A key mechanism that enables learning in neural networks is backpropagation. This is a method used for efficiently calculating the gradients necessary for adjusting the weights. Essentially, errors are propagated backward through the network (from output towards input), allowing for efficient computation of gradient values at each layer. This facilitates the update of weights in a way that progressively reduces the overall error of the network.

While neural networks are powerful tools, they are also susceptible to overfitting, where a model performs well on training data but fails to generalize to unseen data. To avoid this, regularization techniques are employed. For instance, L1 and L2 regularization add a penalty for larger weights to the loss function, encouraging the model to maintain smaller weights and thus simpler models. Dropout is another technique that randomly disables neurons during training, which helps in preventing co-adaptations where neurons excessively rely on particular other neurons.

3.3 Convolutional Neural Networks

Following this foundational understanding of general neural networks, for the purpose of this research it is essential to explore convolutional neural networks (CNNs), which introduce the concept of convolutions to neural networks. CNNs are particularly suited for data that have a grid-like topology, such as images or time-series data like ECG signals. The convolutional layers in CNNs exploit the spatial correlations present in the input data, making them highly efficient for tasks involving images, videos, and medical signal data, such as ECG analysis.

Convolutions are mathematical operations used in CNNs that involve sliding a filter or kernel over the input data. This kernel is a small matrix used to extract features such as edges, shapes, and textures from the input by performing element-wise multiplication with the part of the input it covers, and then summing the results into a single output pixel. This process is repeated across the entire input, creating a feature map that highlights important attributes in the data.

The effectiveness of CNNs in feature detection and extraction is enhanced by stacking multiple convolutional layers. Each layer typically performs a different convolution, focusing on increasingly abstract aspects of the input. After convolutional layers, pooling layers are often used to reduce the dimensionality of the data, helping to decrease the computational load and control overfitting. The most common types of pooling include max pooling and average pooling, which downsample the input by selecting the maximum or average value respectively from a set of values in a region defined by the pooling filter.

The architectural sophistication of CNNs has given rise to more complex architectures, each designed for specific types of problems or to improve upon the limitations of earlier models. Notable examples include:

1. AlexNet [38]: This was one of the first deep architectures that showed the power of CNNs in large scale image recognition tasks. It significantly outperformed traditional methods in the 2012 ImageNet competition.
2. VGGNet [39]: Known for its simplicity, VGGNet consists of multiple convolutional layers followed by three fully-connected layers. It emphasizes depth by using a large number of convolutional layers with small-sized kernels.
3. ResNet [40]: Short for Residual Network, ResNet introduced the concept of skip connections allowing gradients to flow through an alternative shortcut path to combat the vanishing gradient problem in very deep networks.
4. GoogLeNet (Inception) [41]: This model introduced the idea of an 'Inception module', which uses parallel convolutional layers with different sizes to capture information at various scales.

These architectures exemplify the evolution and diversification of convolutional neural networks, showing how each iteration improves on the previous one to address specific challenges in image processing and pattern recognition.

The following section will delve into U-Net, a significant architecture derived from CNNs, specifically designed for biomedical image segmentation, showing how specialized network designs can further optimize performance for specific tasks.

3.4 The U-Net Architecture

The U-Net architecture was introduced by Ronneberger et al. in their paper titled "U-Net: Convolutional Networks for Biomedical Image Segmentation" [30]. It is characterized by its symmetric structure, which enables precise localization and uses of context. This architecture is particularly suited for tasks where the output requires exact alignment with the input, such as semantic segmentation tasks, where the goal is to classify each pixel of an image into a specific category.

The U-Net architecture comprises two main parts: the contracting (downsampling) path and the expansive (upsampling) path, which gives it a characteristic "U" shape. Here is a breakdown of each component:

- Contracting Path: The contracting path follows the typical architecture of a convolutional network. It consists of repeated application of two 3x3 convolutions (each followed by a rectified linear unit, or ReLU), and a 2x2 max pooling operation with stride 2 for downsampling. With each downsampling, the number of feature channels doubles (i.e., the feature maps increase in depth) to capture increasingly abstract features and reduce spatial dimensions. This path is responsible for capturing the context in the input image, which helps in recognizing features at various scales.
- Bottleneck: This part of the network bridges the contracting and expansive paths. It typically consists of two 3x3 convolutions, each followed by a ReLU. It helps the network transition from feature encoding to feature decoding, which is crucial for precise localization in the output.
- Expansive Path: The expansive path mirrors the architecture of the contracting path, but in reverse order. It includes a series of upsampling operations followed by 2x2 transposed convolutions (or upsampling and a 3x3 convolution) that increase the dimensions of the feature maps. After each upsampling, the number of feature channels is halved. This is followed by a concatenation with the corresponding cropped feature map from the contracting path (skip connection). This process of concatenation is critical as it allows the network to use context gathered during the contracting path to enable precise localization. Following the concatenation, two 3x3 convolutions are applied, each followed by a ReLU activation function. These convolutions help refine the features for more accurate output.
- Final Layer: The final layer of the U-Net is a 1x1 convolution that is used to map the feature vectors to the desired number of classes in each pixel, outputting the segmented image.

Figure 3.1 shows the diagram of the U-Net architecture presented in the original paper.

In the context of ECG delineation, the U-Net architecture can be used to perform precise segmentation of the ECG signal into its waveforms—P wave, QRS complex, and T wave. The ability of the U-Net to capture both local and broader contextual information makes it suitable for this task, as ECG delineation requires both high precision in waveform boundary detection and an understanding of the overall structure of the ECG signal.

The U-Net architecture offers several advantages for ECG delineation. One of the main benefits is its precision. Thanks to its symmetric expansion path, U-Net can achieve precise localization, making it ideal for accurately delineating ECG waveform boundaries. Also, U-Net can be trained end-to-end from very few images while still yielding high segmentation accuracy. This is particularly advantageous in medical applications where annotated examples are often scarce. Furthermore, although U-Net was originally designed for segmentation tasks, its versatility allows it to be successfully adapted for other applications requiring high-resolution output from limited input data, such as ECG delineation.

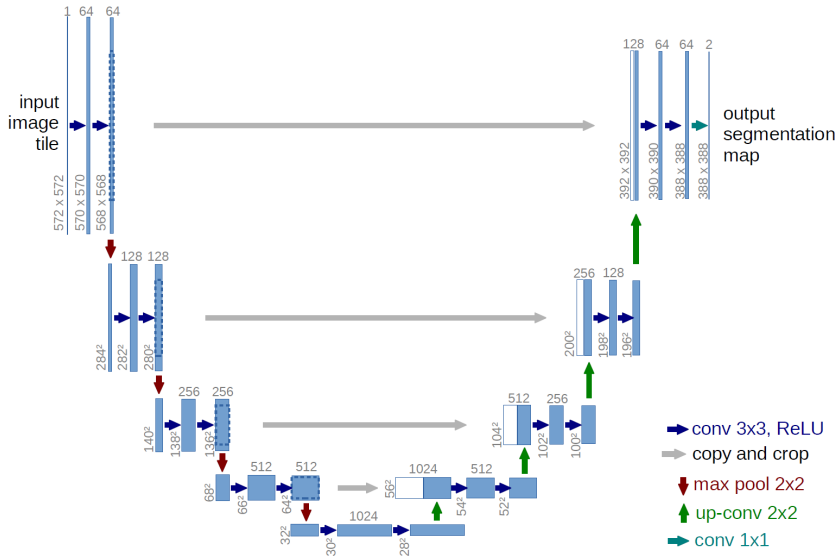


Figure 3.1: Original U-Net architecture. The contracting path (left) captures context through successive downsampling operations, while the expansive path (right) enables precise localization via upsampling and concatenation with high-resolution features from the contracting path. Source: [30].

3.5 Attention Mechanism

The attention mechanism has become one of the most influential recent innovations in the field of artificial intelligence, especially in the area of deep learning. Inspired by human perceptual processes, this mechanism enhances the ability of models to focus on specific parts of an input when performing tasks, thus improving the efficiency and effectiveness of learning processes. It has been adopted mainly in domains like natural language processing (NLP) and image recognition.

Introduced by Bahdanau et al. (2014) [42] in sequence-to-sequence (seq2seq) models, the attention mechanism was developed to overcome the limitations of traditional fixed-sized context vectors in handling long sequences, thereby improving performance in tasks such as machine translation. Initial work explored attention-maps by interpreting gradient of output class scores with respect to the input image. On the other hand, trainable attention is enforced by design, and can be categorized into hard-, soft- and self-attention.

Hard attention, as described by Mnih et al. (2014) [43], is typically non-differentiable and relies on reinforcement learning for parameter updates, making the training process more challenging. This type of attention focuses on specific parts of the inputs while ignoring others, which is beneficial for tasks requiring explicit decisions, such as iterative region proposal and cropping. Recursive hard attention has been applied to detect anomalies in chest X-ray scans [44].

In contrast, soft attention is probabilistic and employs standard back-propagation without the need for Monte Carlo sampling [45]. It assigns weights to all input parts using a softmax function to compute these weights. Examples include additive soft attention in machine translation [42, 46] and image classification [47, 48]. Squeeze-and-Excitation networks [49] use channel-wise attention to emphasize important fea-

tures, winning the ILSVRC 2017 image classification challenge [50].

Self-attention [47, 51] analyzes the importance of input parts relative to each other, removing the need for external gating. It captures long-range dependencies [51] and improves image classification through class-specific pooling [47, 48]. Self-attention is a key component of the widely-used Transformer model.

The Transformer model, introduced by Vaswani et al. (2017) [52], relies entirely on self-attention to capture global dependencies between input and output. The model's ability to handle long-range dependencies made it significantly more powerful than models based on earlier seq2seq architectures. This model introduced the concept of multi-head attention, where multiple attention mechanisms are utilized in parallel, enabling the model to attend to different aspects of the input. Since its introduction, Transformers have dominated the field of NLP with adaptations such as BERT [53], GPT [54], and others, demonstrating the effectiveness of the attention mechanism.

The attention mechanism also improves model interpretability, as its attention weights can be analyzed to understand what the model is focusing on, making it easier to interpret the model's decisions. It also enhances performance due to its focus on relevant parts of the data, allowing for more effective handling of complex datasets. Moreover, its flexibility makes it versatile, integrating well with various neural network architectures, and enhancing their applicability across different AI domains.

3.6 Attention U-Net

Building on the concepts from previous sections on U-Net and attention mechanisms, Oktay et al. proposed a novel approach that leverages both to address challenges in medical image segmentation [55]. Their work introduces Attention U-Net, which improves on the base U-Net architecture, particularly for complex anatomical structures with high variability in shape, size, and appearance.

The Attention U-Net addresses this limitation by incorporating attention gates into the network architecture. Specifically, the attention gates are applied to the concatenation of skip connections from the contracting path (encoder) and the upsampled feature maps from the expansive path (decoder). This architectural modification allows the network to selectively emphasize or suppress relevant features during the decoding process, ultimately leading to improved segmentation performance.

The key difference between the standard U-Net and the Attention U-Net is the introduction of these attention gates. Each attention gate is a small convolutional block that generates an attention map corresponding to the input feature maps. This attention map is then element-wise multiplied with the input feature maps, modulating the feature activations based on their relevance to the target segmentation task.

By focusing on the relevant parts of the image, Attention U-Net achieves higher segmentation accuracy, especially in complex cases where traditional models might fail. Attention gates also help reduce computational complexity by allowing the model to focus on the areas of interest, potentially reducing the need for a larger number of parameters. In addition, this model better handles variations in medical images, making it more robust and generalizable across different datasets and medical conditions. Moreover, the attention maps provide interpretability, allowing visualisation of which regions the network is focused on during the segmentation process.

Chapter 4

Methods

This chapter provides details on the data and methods utilized in this work. It begins by describing the datasets employed, including their source, size, and characteristics along with an exploratory data analysis to gain insights into the data. The data pre-processing steps, such as filtering, are then explained. The architecture and implementation of the proposed deep learning model are outlined. Finally, the chapter covers the post-processing methods to refine outputs and the evaluation metrics for assessing performance. This overview establishes the foundation for the presentation of experimental results and analyses in subsequent chapters.

4.1 Databases

Training and validating ECG delineation models requires standardized databases with complexes and waves, manually annotated by specialists. Several widely used collections are available, such as the MIT-BIH Arrhythmia Database [56], the European ST-T Database [57], and the QT Database [17]. However, their annotations have limitations. For instance, the MIT-BIH Arrhythmia Database and the European ST-T Database only provide annotations for QRS complexes. The QT Database includes annotations for P, QRS, and T waves, but several complexes remain unmarked, and it only contains leads I and II. To address these limitations, the Lobachevsky University Electrocardiography Database (LUDB) [18] was developed.

4.1.1 LUDB Database

The Lobachevsky University Electrocardiography Database (LUDB) [18] is a public dataset developed for the purpose of ECG analysis and the testing of delineation algorithms. It comprises 200 12-lead ECG recordings sampled at a frequency of 500 Hz with a 10-second duration. Each ECG recording in LUDB is annotated with P-wave, QRS complex, and T-wave boundaries and peaks identified and validated by expert cardiologists, providing high-quality ground truth for training and validating deep learning models. The dataset includes ECGs from patients with various arrhythmias, thus presenting a diverse set of cardiac rhythms for comprehensive analysis. Additionally, LUDB provides demographic information for each subject, allowing researchers to consider age-related and gender-specific characteristics in their studies.

4.1.2 Expanded Classification Approach in This Study

Most works in the deep learning literature consider ECG delineation as a four-class segmentation problem, distinguishing between P waves, QRS complexes, T waves, and isoelectric segments. However, while classifying these waveforms provides significant insights, the precise delineation of the peaks within these waves (specifically, P, R, and T peaks) is also crucial. Precisely locating the peaks enhances temporal precision, crucial for measuring intervals and durations within the ECG that diagnose conditions like arrhythmias. Furthermore, peak characteristics, such as the height and sharpness of the R peak, can yield insights into cardiac functions, indicating conditions like ventricular hypertrophy. Peaks can also reveal subtle changes in heart's electrical activity that broader wave classifications might overlook, aiding in the detection of early or mild cardiac dysfunctions.

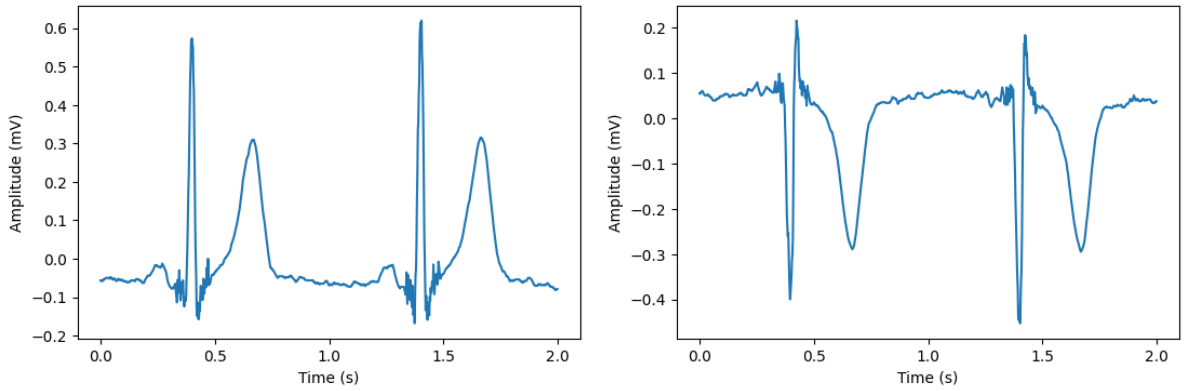
Given the importance of both waveforms and their peaks, this study extends beyond the traditional four classes typically considered in ECG analysis. An expanded seven-class system is incorporated to include the three additional peak classes: P peak, R peak, and T peak. This expanded classification scheme enables the proposed deep learning model to not only recognize and categorize the main waveforms but also to accurately identify and analyze the critical peaks within these waves. By training the model on this refined class system, a higher level of detail and accuracy in ECG signal analysis is aimed for, potentially leading to improved diagnostic capabilities.

4.1.3 Training on Multi-Lead Data for Enhanced Model Robustness

Another key aspect that sets this study apart from the existing literature is the approach to training data. Commonly, deep learning models for ECG delineation are trained and evaluated using signals from a single lead, typically Lead II (see [29, 36, 34]). However, one of the main objectives of this work is to develop a model that can be applied in real-world scenarios, where ECG signals may be acquired from any lead depending on clinical settings and patient characteristics. This is particularly relevant with the increasing popularity of wearable devices, which often allow users to select the number and placement of leads based on their preferences and the device's capabilities.

To achieve this "lead agnosticism", the model is trained using all available leads from the LUDB database. Each of the 12 leads in an ECG captures electrical signals from different angles across the heart, resulting in varied waveforms or morphologies. These differences in morphology across leads can include variations in wave amplitude, shape, and timing. For instance, the R peak might appear sharper and more pronounced in one lead while being flatter in another, as illustrated in Figure 4.1. This variability poses a significant challenge even for classical algorithms not based on deep learning, which is why they are also typically tuned and validated for one or a few leads (see [9, 19, 20]). Similarly, deep learning models trained on a single lead may struggle to generalize to other leads, potentially leading to suboptimal performance in real-world applications.

The implication of these morphological variations for a deep learning model is twofold. Firstly, it increases the complexity of the learning task as the model must recognize and interpret the same waveform patterns presented in different ways. Secondly, and more importantly, it enhances the model's robustness and adaptability. By exposing



(a) Lead I morphology.

(b) Lead III morphology.

Figure 4.1: Comparison of ECG morphology between leads during the same cardiac cycle, captured simultaneously.

the model to diverse waveform representations during training, the aim is to make it more versatile and capable of handling ECG signals from any lead. This "lead agnosticism" is particularly valuable in clinical practice, where the choice of lead may vary or when dealing with patients who have unconventional lead placements due to specific medical conditions or anatomical variations.

4.2 Exploratory Data Analysis

Table 4.1 presents the distribution of sample point classes across the seven predefined categories: P, QRS, and T waves, along with their respective peaks, and the isoelectric class. The table reveals a significant class imbalance, particularly between the peak classes and the other categories. This imbalance is expected given that a peak represents a singular, discrete sample point within a wave, which itself consists of a contiguous segment of sample points.

Table 4.1: Sample point distribution of the entire dataset.

Class	Number of sample points	% of total
Isoelectric	7,667,846	68.45%
P wave	770,854	8.84%
QRS wave	989,842	14.81%
T wave	1,659,130	6.88%
P peak	33,078	0.38%
R peak	42,530	0.34%
T peak	38,048	0.30%

A closer examination of Table 4.1 also uncovers imbalances among the wave classes themselves. The isoelectric class, representing transitions between waves in the ECG signal, constitutes the majority of the data. This prevalence is understandable as these transitions occupy most of the ECG signal. The P wave exhibits fewer sample points compared to the QRS and T waves, which can be attributed to cases of atrial

Table 4.2: Distribution of ECG conditions in the LUDB database.

Condition		Number of ECGs	% of total
Rhythm	Sinus rhythm	1,716	71.5%
	Sinus tachycardia	48	2%
	Sinus bradycardia	300	12.5%
	Sinus arrhythmia	96	4%
	Irregular sinus rhythm	24	1%
	Abnormal rhythm	228	9.5%
Conduction abnormalities (blocks)		792	33.00%
Extrasystoles		168	7.00%
Hypertrophy		1,704	71.00%
Ischemia		612	25.50%
Pacemaker		120	5.00%
Non-specific repolarization abnormalities		588	24.50%
Other		108	4.50%

fibrillation—an ECG pathology characterized by the absence of P waves. Additionally, the T wave has a larger representation than the QRS wave due to its longer duration within a heartbeat.

Similar imbalances are observed within the peak classes. The lower representation of P peaks can be linked to the presence of atrial fibrillation cases in the data. R peaks, however, appear more frequently than T peaks due to the labeling convention in the LUDB database, where many samples end with an R label, omitting the subsequent T wave for the last heartbeat.

The LUDB database also includes a wide range of cardiac conditions, providing a diverse set of ECG signals for training and evaluation. Table 4.2 displays the distribution of ECG conditions in the database. It can be observed that the majority of ECGs exhibit sinus rhythm, which is the normal heart rhythm originating from the sinus node. Notably, almost 10% of the ECGs in the database exhibit an abnormal rhythm, indicating the presence of arrhythmias or other cardiac abnormalities. Additionally, the dataset includes ECGs with conduction abnormalities, extrasystoles, hypertrophy, ischemia, and other conditions. The age of the patients in the database ranges from 11 to 89 years, with an average age of 52 years, and a gender distribution of 57.5% men and 42.5% women. This diversity of conditions and demographics provides a rich and comprehensive dataset for training and evaluating the deep learning model, enabling it to learn patterns across a wide range of cardiac rhythms and abnormalities and enhancing its generalization capabilities.

4.3 Data Pre-processing

ECG signals are inherently noisy and prone to various artifacts due to the sensitive nature of cardiac measurements and the dynamic environment in which they are acquired. The quality of the ECG signal is crucial for accurate delineation and interpretation, therefore, robust pre-processing techniques are essential. This section discusses the pre-processing steps employed to enhance the ECG signals and

improve the accuracy of delineation with deep learning algorithms.

4.3.1 Noise in ECG Signals

Noise in ECG signals refers to any unwanted electrical interference or artifacts that distort the true cardiac electrical activity. These distortions can obscure important diagnostic features of the ECG waveform, such as the P wave, QRS complex, and T wave, leading to misinterpretation of the signal and potentially incorrect clinical decisions. Several types of noise commonly affect ECG signals, including:

- Powerline interference: This noise originates from the AC (alternating current) electrical power supply and typically appears as a 50Hz or 60Hz interference, depending on the region's power standard. It manifests as a prominent, consistent frequency component in the ECG signal.
- Baseline Wander: This type of noise refers to low-frequency variations in the isoelectric line of the ECG signal. It can be caused by respiration, poor electrode contact, or patient movement, and may interfere with accurate delineation of the signal.
- Electromyographic (EMG) noise: EMG noise arises from the electrical activity of skeletal muscles. It is often observed in ECG recordings due to movement or muscle contractions, appearing as high-frequency noise or baseline wander.
- Motion artifacts: Patient movement or vibration of the recording equipment can introduce noise into the ECG signal. This type of noise can vary in frequency and amplitude, often requiring adaptive filtering techniques for effective removal.

Effective noise reduction is essential for accurate ECG delineation, as it enhances the visibility of the waveform components and improves the model's ability to detect and classify the fiducial points accurately. The following pre-processing steps are applied to the ECG signals to remove noise and prepare the data for input to the deep learning model.

4.3.2 Band-pass Filtering

To remove unwanted noise and emphasize the frequency components of interest, a band-pass filter is applied to the ECG signals. This filter allows only a specific range of frequencies to pass through while attenuating frequencies outside this range. The band-pass filter is characterized by three key parameters: the lowcut frequency, which defines the lower boundary of the passband; the highcut frequency, which sets the upper boundary; and the order of the filter, which influences the sharpness of the transition between the passband and the stopband[1].

In this study, a third-order Butterworth band-pass filter [58] was employed with a lowcut frequency of 0.5Hz and a highcut frequency of 40Hz. The lowcut frequency of 0.5Hz was chosen to remove baseline wander and very low-frequency artifacts, while the highcut frequency of 40Hz was selected to eliminate high-frequency noise, including EMG noise and powerline interference. The order of 3 was chosen to provide

[1] A higher order filter results in a steeper transition, providing a more abrupt cutoff of frequencies outside the desired range.

a balance between effective noise reduction and preservation of the ECG waveform morphology.

The effect of the band-pass filter on the ECG signal is illustrated in Figure 4.2, which zooms in on two heartbeats from a longer ECG recording. The blue signal represents the raw ECG, which includes noise, especially around the QRS wave. The orange signal displays the filtered output after applying the band-pass filter, which demonstrates a clearer waveform morphology, with reduced high-frequency noise and a more stable baseline.

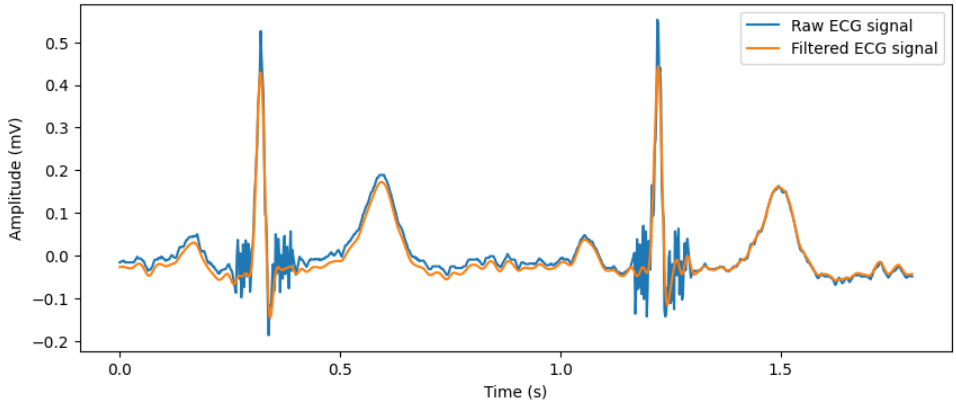


Figure 4.2: Effect of band-pass filtering (0.5-40Hz) on the ECG signal. The blue signal represents the raw ECG, while the orange signal shows the filtered output.

4.4 Proposed Model Architecture: Peak Attention U-Net

This section introduces the proposed deep learning model for ECG delineation, the Peak Attention U-Net. The model architecture is designed to accurately classify each sample point of an ECG signal into one of seven predefined classes: P, QRS, and T waves, along with their respective peaks (P peak, R peak, and T peak), and the isoelectric class. The model leverages the U-Net architecture, a popular choice for image segmentation tasks, and incorporates attention gates to enhance the model's ability to focus on salient features and suppress irrelevant information. The Peak Attention U-Net is designed to capture the complex temporal patterns and subtle morphological features present in ECG signals, enabling precise delineation of the fiducial points.

The proposed ECG delineation method classifies individual sample points from the signal data, similar to pixel classification in image semantic segmentation tasks. The U-Net architecture [30], known for its success in these tasks, serves as the foundation for our model. However, given that the original U-Net is tailored for 2-dimensional image segmentation, we adapt it by substituting 2D convolutions with 1D convolutions, as described by Equation 4.1, thus creating the 1D U-Net.

$$y(t) = (x * w)(t) = \sum_{n=-m}^{m} x(n) \cdot w(t-n) \tag{4.1}$$

where x is the input data sequence, y is the output of the convolution, and w is the

convolution kernel of size m. This kernel w slides over the input signal x to generate feature maps. To effectively capture the characteristics of different waveforms in the ECG signal, the kernel size m is set to 9 after testing different kernel sizes [59].

4.4.1 Encoder-Decoder Architecture

The proposed model in this work is based on the well-known U-Net architecture, which employs an encoder-decoder structure. This design allows the model to effectively capture and utilize both high-level representations and fine-grained details for the task of ECG delineation.

Encoder The encoder component of the model is responsible for extracting relevant features from the raw ECG signals. It is composed of three blocks, each designed to learn increasingly complex and abstract representations of the input data. The detailed structure of each encoder block is as follows:

1. First Convolutional Layer: The first layer of the encoder begins with a 1D convolutional layer using 16 filters. This layer applies learnable filters across the temporal dimension of the ECG signal to extract local patterns and features. The number of filters in the convolutional layers doubles with each subsequent block.

   All the convolutional layers employ a ReLU (Rectified Linear Unit) activation function to introduce non-linearity into the model while maintaining computational efficiency. The ReLU activation function is defined as $\mathrm{ReLU}(x) = \max(0, x)$, which helps in mitigating the vanishing gradient problem and accelerates convergence.

2. Batch Normalization: Following the convolution, batch normalization is applied. This technique normalizes the activations of the previous layer for each mini-batch, stabilizing and accelerating the training process. It also reduces the internal covariate shift and allows for higher learning rates by maintaining the mean output close to zero and the output standard deviation close to one.

3. Spatial Dropout: To prevent overfitting and improve the generalization capability of the model, spatial dropout is employed. Unlike standard dropout, which drops individual neurons randomly, spatial dropout drops entire feature maps. This is particularly effective in 1D convolutional networks, as it encourages the model to develop redundant representations and prevents co-adaptation of adjacent features.

4. Second Convolutional Layer: Another 1D convolutional layer with the same number of filters as the first layer of the current block is applied. This layer further refines the features extracted by the previous layer, capturing more complex patterns and relationships in the ECG signal.

5. Max-pooling: Finally, a max-pooling operation is performed, which reduces the spatial (in this case, temporal) dimensionality of the feature maps and increases the receptive field of subsequent layers. Max-pooling selects the maximum value within small windows of the feature maps, thereby providing a form of translation invariance. This operation downsamples the input, allowing the model to learn hierarchical representations at different scales.

As the input progresses through the three encoder layers, the temporal resolution decreases while the feature depth increases. These layers have 16, 32, and 64 filters, respectively, capturing both local and global patterns in the ECG signal.

Bottleneck The bottleneck layer serves as the bridge between the encoder and decoder. It captures the most abstract and high-level features extracted by the encoder and provides a compact representation of the input signal. The bottleneck layer consists of two 1D convolutional layers with 128 filters, each followed by batch normalization, spatial dropout, and ReLU activation.

Decoder The decoder component of the 1D U-Net is responsible for reconstructing the segmented ECG signals by upsampling the feature maps obtained from the encoder. This process involves combining high-level semantic information with fine-grained spatial details to generate accurate delineations. The decoder consists of three blocks, each designed to restore the spatial resolution of the feature maps. The detailed structure of each decoder block is as follows:

1. Upsampling: The decoder starts with an upsampling operation, implemented as transposed 1D convolution. This learnable upsampling operation increases the spatial resolution of the feature maps, allowing the decoder to recover the fine-grained spatial details lost during the downsampling process in the encoder. The transposed convolutional layer uses a kernel size of 2 and a stride of 2 to double the spatial dimensions of the input.
2. Concatenation: A key aspect of the U-Net architecture is the use of skip connections, which concatenate the upsampled feature maps from the decoder with the corresponding feature maps from the encoder. This enables the decoder to access detailed spatial information captured at different levels of the encoder hierarchy. The skip connections facilitate the flow of fine-grained details from the encoder to the decoder, enhancing the model's ability to generate accurate delineations.
3. First Convolutional Layer: Following the concatenation, a 1D convolutional layer with the same number of filters as the corresponding encoder block is applied. This layer refines the combined feature maps, capturing the spatial dependencies and morphological characteristics.
4. Batch Normalization: Similar to the encoder, batch normalization is applied to stabilize and normalize the feature maps, improving the training process and accelerating convergence.
5. Spatial Dropout: Again, spatial dropout is used to prevent overfitting and enhance the generalization capability of the model.
6. Second Convolutional Layer: Another 1D convolutional layer is applied to further refine the features and generate more detailed representations. This layer captures complex patterns and relationships in the feature maps, enhancing the model's ability to delineate the ECG signals accurately.

The decoder gradually restores the temporal resolution while reducing the feature depth. The three decoder blocks have 64, 32, and 16 filters, respectively, mirroring the encoder's structure.

Output Layer The final layer of the model outputs a feature map with the same temporal dimensions as the input ECG signal but with 7 channels, corresponding to the 7 classes of the delineation task. A softmax activation function is applied to obtain a probability distribution over the classes for each sample point, as shown in Equation 4.2. This distribution indicates the likelihood of each class being the correct classification at each time step of the ECG signal. The class with the highest probability is taken as the predicted classification for that sample point.

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{7} e^{z_j}} \tag{4.2}$$

In this equation, z_i represents the input to the softmax function for the i-th class, which is the output of the convolutional layer for that class. The numerator e^{z_i} is the exponential of the i-th class score, and the denominator $\sum_{j=1}^{7} e^{z_j}$ is the sum of the exponentials of all class scores. The softmax function normalizes the inputs into a probability distribution, where the sum of the probabilities for all classes equals 1. The output of the model is a delineation of the input ECG signal, where each data point is assigned a probability for each class. The class with the highest probability is taken as the predicted classification for that data point. This enables accurate and precise classification of each segment of the ECG signal into one of the 7 predefined classes.

4.4.2 Incorporating Attention Gates

Attention mechanisms have shown remarkable improvements in various deep learning tasks. A later study proposed the Attention U-Net [55], which incorporated attention gates to highlight salient features and suppress irrelevant ones. Motivated by the enhanced results of this study in the context of medical image segmentation, we extend this concept to the 1D ECG signals, aiming to achieve similar improvements in performance.

The intuition behind the attention gate is to provide the network with a mechanism to selectively emphasize or suppress different parts of the input dynamically (i.e., weight the importance of different parts of the input). This adaptive behavior enables the model to focus on the most informative regions of the ECG signal, potentially improving its ability to classify sample points accurately. This is particularly beneficial for tasks where the relevant features are sparse or localized, such as the peaks in an ECG wave.

The mathematical framework for implementing attention gates for 1D signals is discussed below.

1. **Convolutional Transformation**: The attention gate takes two inputs: the upsampled feature map ($x_1(t)$) and the corresponding feature map (skip connection) from the encoder path of the U-Net ($x_2(t)$). The first step in the attention gate is to project these inputs into an intermediate feature space using 1D convolution operations with a kernel size of 1 (point-wise convolution). This transformation is described by:

$$y_i(t) = \sum_{k=0}^{F_i-1} w_{i,k} \cdot x_i(t) + b_i \quad \text{for } i = 1, 2 \tag{4.3}$$

where F_i is the number of filters for the i-th input, $w_{i,k}$ are the weights of the k-th filter, and b_i is the bias. In our implementation, $F_1 = F_2 = n$ (number of intermediate filters), and the transformation aims to reduce the channel dimensions to n for both inputs without altering the spatial dimensions.

2. **Feature Combination and Activation**: After obtaining the transformed features $y_1(t)$ and $y_2(t)$, where $y_2(t)$ acts as the context feature map, the next step is to combine these features. This is achieved using element-wise addition followed by a ReLU activation function, as shown in Equation 4.4:

$$z(t) = \text{ReLU}(y_1(t) + y_2(t)) \tag{4.4}$$

3. **Attention Coefficients**: Subsequently, the combined signal $z(t)$ is processed by another convolutional layer to compute the attention coefficients (Equation 4.5). This convolutional operation reduces the number of filters to 1, aiming to generate a single attention signal $g(t)$ which is passed through a sigmoid activation to form the final attention coefficients $\alpha(t)$ (Equation 4.6):

$$g(t) = \psi \cdot z(t) + b \tag{4.5}$$

$$\alpha(t) = \sigma(g(t)) \tag{4.6}$$

where ψ and b are the weights and bias of this convolutional filter, and σ denotes the sigmoid function given by $\sigma(x) = \frac{1}{1+e^{-x}}$. The sigmoid function ensures that the output attention coefficients $\alpha(t)$ are normalized to a range between 0 and 1, indicating the importance of the corresponding input features.

4. **Application of Attention Coefficients**: Finally, the attention coefficients are applied to the original input $x_1(t)$ by performing an element-wise multiplication, which modulates the input based on the learned attention weights. This operation is expressed as in Equation 4.7:

$$\hat{x}_1(t) = \alpha(t) \odot x_1(t) \tag{4.7}$$

Here, $\hat{x}_1(t)$ represents the attention-modulated output of the gate. This output selectively emphasizes features in $x_1(t)$ based on the contextual cues captured from $x_2(t)$.

By adding attention gates to the decoder blocks, the model can focus on relevant features during upsampling, resulting in more precise and robust ECG delineation and significantly better peak detection. Figure 4.3 shows the structure of the proposed attention gate.

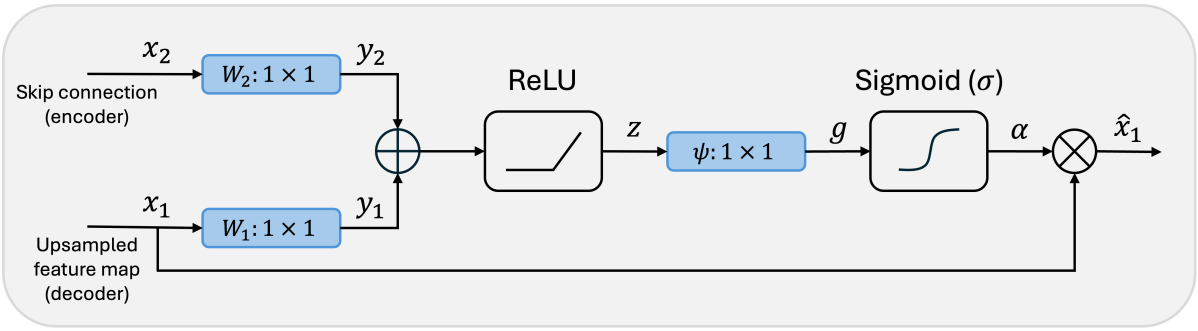


Figure 4.3: Structure of the proposed attention gate.

Putting it all together, the proposed model architecture, Peak Attention U-Net, combines the 1D U-Net structure with attention gates to enhance the model's ability to capture salient features and suppress irrelevant information. The model is designed to accurately delineate ECG signals by classifying each sample point into one of the seven predefined classes. Figure 4.4 illustrates the architecture of the Peak Attention U-Net, highlighting the encoder-decoder structure, the incorporation of attention gates, and the flow of information through the network.

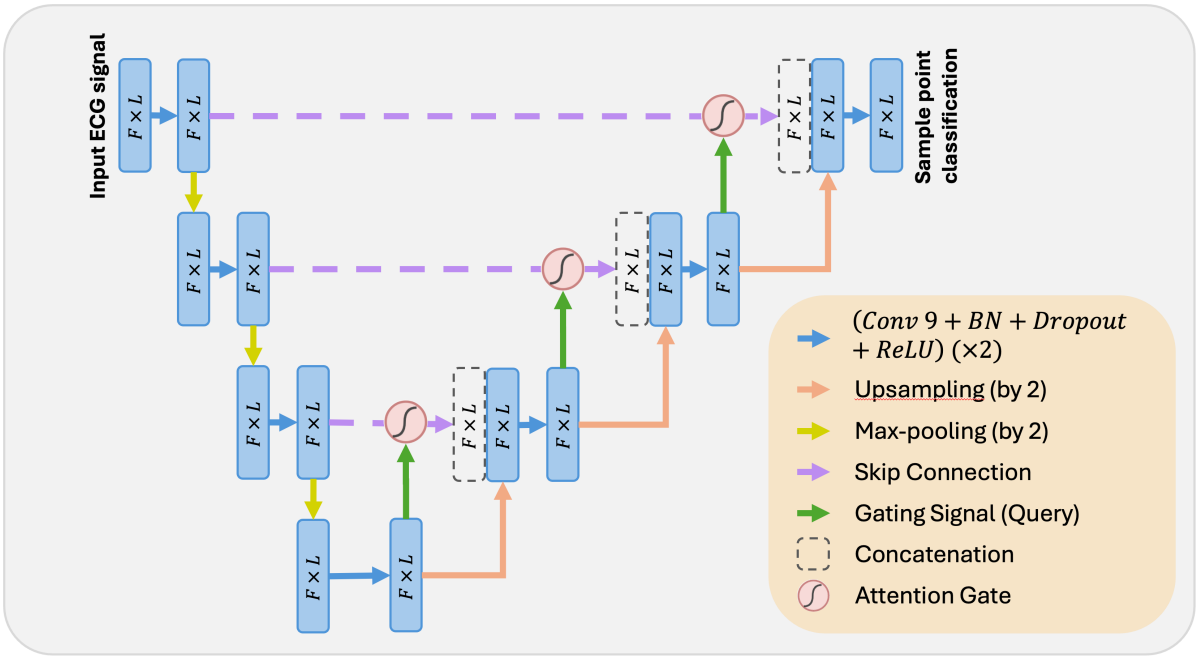


Figure 4.4: Architecture of the proposed Peak Attention U-Net.

Table 4.3 presents the size in MB and the number of parameters for the proposed model, both with and without attention gates. Incorporating the attention gates results in an approximate increase of 2.3% in both model size and number of parameters. However, as will be demonstrated in Chapter 5, this marginal increase in model complexity leads to a significant improvement in ECG peak detection. Additionally, the relatively small model size and low number of parameters make the model feasible to use in real-world applications.

Table 4.3: Model sizes and number of parameters.

Model	Size (MB)	Number of parameters
U-Net (this work w/o Attention)	1.76	460,343
Peak Attention U-Net (this work)	1.8	471,434

4.5 Proposed Model Implementation: Peak Attention U-Net

This section delves into the specifics of the proposed Peak Attention U-Net model implementation, including the choice of loss function, optimization algorithm, batch size, training epochs, and the use of activation functions. All these choices were made after extensive hyperparameter optimization to ensure the model's optimal performance.

First, the loss function employed in the model is discussed. The cross-entropy loss function, as defined in Equation 4.8, was utilized to measure the discrepancy between the predictions of the model and the actual labels (ground truth).

$$L(y, \hat{y}) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{4.8}$$

In this equation, y represents the ground truth as a one-hot encoded vector, where the correct class label is assigned a value of 1, and all other elements are set to 0. $\hat{y}$ denotes the model's predicted probability distribution, indicating the assigned probability of each instance belonging to each class. The terms y_c and $\hat{y}_c$ represent the true label and the predicted probability for class c, respectively. The summation $\sum$ extends over all C classes, representing the total loss computed from the discrepancies across all predicted classes.

This function L quantifies the degree to which the model's predictions deviate from the true classifications. The logarithmic component of the loss function penalizes incorrect classifications more heavily when the model is confident in its incorrect prediction (i.e., when $\hat{y}_c$ is high for the wrong class), emphasizing the cost of being confidently incorrect. The negative sign ensures that a lower loss score corresponds to a higher model accuracy, aligning the objective of loss minimization with the goal of maximizing predictive accuracy.

To optimize the model and minimize the cross-entropy loss, the Adam optimization algorithm [60] is employed. Adam, short for Adaptive Moment Estimation, is an extension of the classic stochastic gradient descent and is characterized by its use of momentum [61] to compute individual adaptive learning rates for different parameters. This method combines the advantages of two other extensions of stochastic gradient descent: Adaptive Gradient Algorithm (AdaGrad) [62], which works well with sparse gradients, and Root Mean Square Propagation (RMSProp) [63], which handles non-stationary objectives effectively. The Adam optimizer is widely recognized for its fast convergence and robustness to initial learning rate settings, making it a standard choice for many deep learning applications.

During the training process, a batch size of 32 is used, which means that the model processes 32 training samples at a time before updating the model parameters. This

batch size is a trade-off between computational efficiency and model accuracy, as larger batch sizes can lead to faster training but may also result in less accurate models.

The number of training epochs is set to 300, with an early stopper of 10 epochs. An epoch refers to one complete pass through the entire training dataset, where the model sees every training sample once. The early stopper is a mechanism that monitors the model's performance on the validation set and stops the training process if the model's performance does not improve for a specified number of epochs (in this case, 10). This prevents overfitting and saves computational resources.

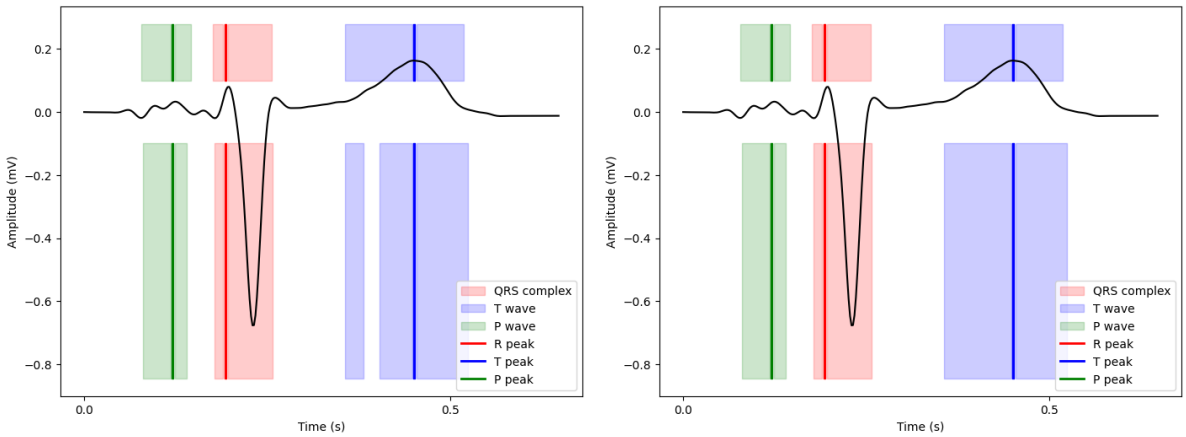
In addition to the early stopper, two callbacks are also employed to further optimize the training process. The first callback reduces the learning rate on a loss plateau of 5 epochs, which means that if the model's loss on the validation set does not decrease for 5 consecutive epochs, the learning rate is reduced to encourage the model to explore different regions of the parameter space. The second callback saves in memory the model with the lowest loss in validation, ensuring that the best-performing model is kept for further testing and evaluation.

The proposed model was implemented using the Keras framework [64] with TensorFlow [65] as the backend engine. The experiments were conducted on an Ubuntu machine with an Intel i9-12900KS CPU, 128 GB of RAM, and an NVIDIA RTX 4090 GPU with 24 GB of memory. The model structure required 1.8 MB of memory to store its 471 thousand parameters, making it feasible for deployment on a variety of platforms. Regarding the computational time for inference, the model processes a 10-second ECG signal in approximately 0.24 milliseconds when using the GPU, and 4.04 milliseconds when using a low-end CPU (Intel i5-12450H). The model's efficiency and speed make it suitable for real-time applications and clinical settings.

4.6 Post-processing

In this study, the proposed Peak Attention U-Net model is trained to classify each sample point of an ECG signal into one of seven predefined classes: P, QRS, and T waves; P, R, and T peaks; and Isoelectric (none). The output of the model is a sequence of labels with the same length as the input ECG signal, where each label represents the predicted class for the corresponding sample point. The following post-processing steps are designed to convert this sequence of labels into the fiducial points of each waveform.

The first step in the post-processing pipeline is to smooth the model's output using a moving average filter. This filter assigns to each sample point the average value of its neighboring points, reducing the impact of sporadic misclassifications. The inherent noise and variability present in ECG signals, coupled with the model's sensitivity, can lead to isolated misclassifications that deviate from the overall pattern. By applying the moving average filter, the impact of these outlier predictions is mitigated, and a more robust representation of the ECG waveform structure is obtained. Figure 4.5 provides a visual example of how the moving average filter improves the model's output by avoiding a misclassification while preserving the overall waveform morphology. This smoothing step is a common practice in time series segmentation tasks, where maintaining the continuity and smoothness of the output is crucial.



(a) Before moving average.

(b) After moving average.

Figure 4.5: Delineated heartbeat before and after applying the moving average filter.

After smoothing the signal, the onset and offset of each waveform are identified. This process relies on detecting transitions between the 'Isoelectric' class and the waveform classes (P, QRS, and T waves). For example, a transition from the 'Isoelectric' class to the 'P wave' class indicates the onset of a P wave, while a transition from the 'P wave' class back to the 'Isoelectric' class or another waveform class marks the offset of the P wave. Accurate onset and offset delineation is essential for subsequent measurements and analysis of ECG signals, providing critical information about cardiac intervals and waveform morphology.

Finally, the delineation results are validated and refined using domain knowledge and predefined rules. This step ensures that the identified waveform boundaries adhere to physiological constraints and fall within acceptable ranges. For example, constraints on the minimum and maximum duration of each waveform type are enforced, eliminating outlier delineations that deviate significantly from expected norms [66, 67, 68]. This refinement step enhances the robustness and clinical relevance of the final delineation results.

This data post-processing pipeline transforms the sample point classification output of the Peak Attention U-Net model into precise ECG delineation results. Figure 4.6 shows the result of applying these steps to two heartbeats, obtaining the fiducial points. These steps are crucial for ensuring the reliability and interpretability of ECG delineation using deep learning models and enable subsequent quantitative evaluation and clinical decision-making based on the extracted waveform information.

4.7 Description of Evaluation Metrics

Evaluating the performance of the proposed Peak Attention U-Net model involves assessing its effectiveness in two distinct tasks: sample point classification (before the post-processing step) and fiducial point delineation (after the post-processing step). Each task presents unique challenges, and specific evaluation metrics are employed to quantify the model's accuracy and reliability.

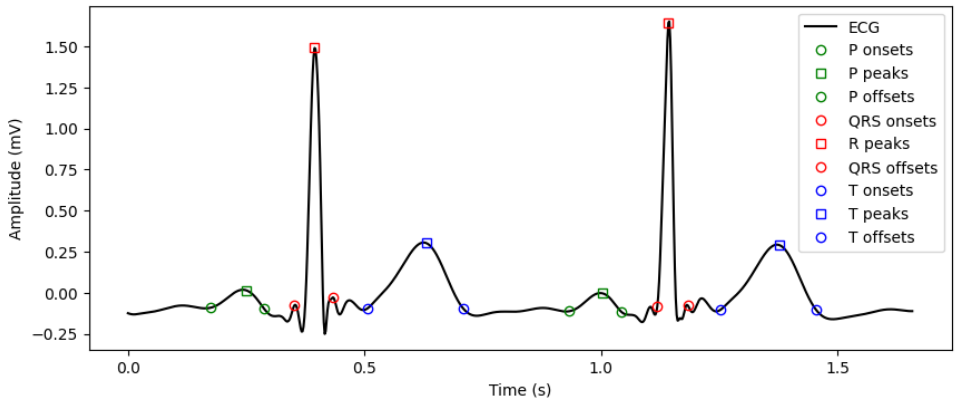


Figure 4.6: Fiducial point delineation after post-processing steps.

4.7.1 Sample Point Classification

Sample point classification evaluates the model's ability to classify each data point in the ECG signal accurately into one of the predefined classes. In the proposed framework, these classes include the different segments of the ECG waveform: P-wave, QRS complex, T-wave, and the isoelectric segments between waves (classified as 'None'). This evaluation step occurs before the post-processing stage, where the raw output of the model is analyzed. The effectiveness of this classification is quantified using precision, recall, and F1 score for each class.

- Precision calculates the proportion of correctly predicted positive instances out of all predicted positive instances. In the context of ECG delineation, precision for a specific class indicates the proportion of correctly classified points belonging to that class out of all points predicted as belonging to that class. A high precision indicates low false positive rate. Mathematically, it is expressed as in Equation 4.9.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.9}$$

  where TP represents the number of true positive instances (instances correctly classified as positive), and FP represents the number of false positive instances (instances incorrectly classified as positive).

- Recall, also known as sensitivity or true positive rate, calculates the proportion of correctly predicted positive instances out of all actual positive instances. In ECG delineation, recall for a class represents the proportion of correctly classified points belonging to that class out of all points that truly belong to that class. High recall indicates low false negative rate. It is calculated as in Equation 4.10.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.10}$$

  where FN represents the number of false negative instances (instances incorrectly classified as negative).

- F1 Score is the harmonic mean of precision and recall, providing a single metric that balances both the false positives and false negatives. It is useful when there is an imbalance between classes, as it gives equal importance to both precision

and recall. An F1-score of 1 indicates perfect precision and recall, while a score of 0 indicates that either precision or recall is zero. Its mathematical expression is shown in Equation 4.11.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.11}$$

4.7.2 Fiducial Point Delineation

Fiducial point delineation focuses on identifying the key points on the ECG waveform, specifically the onsets, peaks, and offsets of the P, QRS, and T waves. These fiducial points are crucial for clinical interpretation and subsequent measurement of intervals and amplitudes. This evaluation task occurs after the post-processing step, where the raw output of the model has been transformed into structured waveform information. According to the ANSI/AAMI-EC57:1998-2012 standards [69], a fiducial point is considered correctly delineated if the difference between the predicted and actual positions is within a 150 ms tolerance threshold. To evaluate this aspect, Sensitivity (SE), Positive Predictive Value (PPV), F1 score, and the mean and standard deviation of errors in fiducial point placement are employed.

- Sensitivity (SE) for fiducial point delineation measures the proportion of actual fiducial points that are correctly identified by the model. This metric is similar to recall in classification tasks, as it represents the algorithm's ability to correctly detect the true fiducial points. It is also expressed as in Equation 4.10.
- Positive Predicted Value (PPV) is similar to precision, reflecting the proportion of identified fiducial points that are true positive identifications. It indicates the algorithm's ability to correctly classify predicted fiducial points, and it is calculated as in Equation 4.9.
- F1 Score for fiducial points combines SE and PPV to provide a balanced measure of the model's accuracy in identifying critical ECG landmarks.
- Additionally, the mean and standard deviation ($\mu \pm \sigma$) of the predicted points are reported, which provides an indication of the average error in the delineation. The mean error (μ, Equation 4.12) is the central value of misplacements, while the standard deviation (σ, Equation 4.13) shows the variability of these errors. Lower values of mean and standard deviation indicate more accurate delineation performance.

$$\mu = \frac{1}{n} \sum_{i=1}^{n} |P_i - T_i| \tag{4.12}$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(|P_i - T_i| - \mu\right)^2} \tag{4.13}$$

where n is the total number of fiducial points identified, P_i is the position of the i-th predicted fiducial point, and T_i is the true position of the i-th fiducial point.

These evaluation metrics are widely used in the literature for ECG delineation studies, allowing for a direct comparison of the proposed algorithm's performance with existing methods.

Chapter 5

Experimental Results

The purpose of this chapter is to present and analyze the experimental results of the proposed deep learning model for ECG delineation. This chapter aims to demonstrate the efficacy of the model through a detailed comparison with existing state-of-the-art models. The results are presented in two main sections: sample point classification and fiducial point delineation, each supported by in-depth tables summarizing the performance metrics. First, the experimental setup is described, including data management and training details. The results of the proposed model are then presented and discussed in comparison with other state-of-the-art methods. Finally, a detailed analysis and discussion of the results is provided, highlighting the strengths and limitations of the approach.

5.1 Training Details

5.1.1 Data Management

To train and evaluate the proposed deep learning model, the dataset was divided into three subsets: training, validation, and test sets. An 80/10/10 split is used to ensure a robust evaluation of the model's performance. Each subset is described in detail below:

- **Training Set (80%)**: This subset is used to train the model by adjusting the weights of the network to minimize the loss function. The goal is to identify a set of model parameters that best capture the patterns and relationships within the data, allowing the model to make accurate predictions. It contains 80% of the total available data, providing a diverse collection of ECG recordings to facilitate effective learning of the model parameters.
- **Validation Set (10%)**: This subset is utilized to tune hyperparameters and prevent overfitting during the training process. By evaluating the model on this subset at each epoch, we can monitor its performance on unseen data and make necessary adjustments to improve generalization. It contains 10% of the total data and serves as a checkpoint for the model's performance during training.
- **Test Set (10%)**: This subset is reserved for the final evaluation of the model. It contains 10% of the total data and is not used during the training or validation

Table 5.1: Sample point distribution of the different data subsets.

Class	Training (80%)	Validation (10%)	Test (10%)
Isoelectric	6,132,290 (68.45%)	761,488 (68.04%)	774,068 (68.87%)
P wave	612,554 (6.84%)	79,983 (7.15%)	78,317 (6.97%)
QRS wave	794,114 (8.86%)	97,715 (8.73%)	98,013 (8.72%)
T wave	1,328,362 (14.83%)	168,414 (15.05%)	162,354 (14.44%)
P peak	26,374 (0.29%)	3,398 (0.30%)	3,306 (0.29%)
R peak	33,996 (0.38%)	4,310 (0.39%)	4,224 (0.38%)
T peak	30,478 (0.34%)	3,860 (0.34%)	3,710 (0.33%)

phases. This set provides an unbiased assessment of the model's performance on unseen data.

Table 5.1 presents the number of sample points for each class across the training, validation, and test subsets. While it is challenging to achieve a perfectly proportional division of subsets due to the inherent variability of ECG waveforms, the class distributions across the subsets are well-balanced. This approach ensures that each subset is representative of the overall dataset.

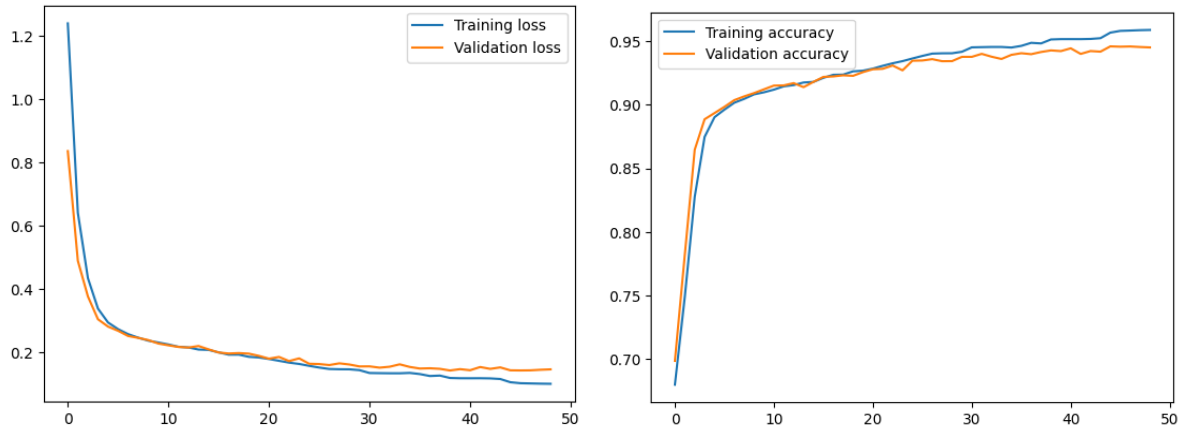
5.1.2 Training Curves

The training process was monitored using two key metrics: loss and accuracy. These metrics were recorded for both the training and validation sets at each epoch to track the model's learning progress and generalizability. The training curves for loss and accuracy are presented below.

The loss function, which measures the discrepancy between predicted and actual values, was minimized during the training process. Figure 5.1a illustrates the training and validation loss curves. The decreasing training loss indicates that the model was effectively learning the underlying patterns in the training data, while the decreasing validation loss suggests that the model was not overfitting and was generalizing well to unseen data.

On the other hand, Figure 5.1b shows the accuracy of the model during training, representing the proportion of correctly predicted samples. The increasing training accuracy indicates the model's improved ability to correctly delineate ECG features over time. Additionally, the increasing validation accuracy demonstrates the model's robustness and generalization capability.

As mentioned in Section 4.5, the number of epochs was set to 300 with an early stopping mechanism triggered after 10 epochs without improvement in the validation loss value. This early stopping halts the training process if the loss of the validation subset does not decrease for ten consecutive epochs. In our case, the lowest loss value was achieved at epoch 39, leading to the training process stopping at epoch 49. This approach helps prevent overfitting, as epoch 39 represents the point of best generalizability. Beyond this epoch, the model would begin to overfit to the training data, losing its ability to generalize effectively.



(a) Loss curves.

(b) Accuracy curves.

Figure 5.1: Training and validation curves.

5.2 Sample Point Classification

In this section, the performance of the proposed model is evaluated on the sample point classification problem, which involves segmenting the different waveforms in ECG signals. This task is the raw output of the model before any post-processing techniques are applied.

The model's performance was evaluated by comparing it to three recent and relevant approaches in the literature:

- Moskalenko et al. (2019) [29]: This study adapted the original U-Net architecture, initially designed for image segmentation, into a 1D U-Net suitable for ECG signals. Their 1D U-Net model was pioneering in demonstrating the feasibility of using convolutional networks for ECG delineation. The proposed model improves on this by incorporating attention gates at each stage of the decoder. The attention gates apply dynamic attention coefficients to each feature map of the encoder, based on the upsampled output from the previous stage of the decoder.
- Londhe et al. (2021) [35]: They introduced a hybrid model combining CNNs with BiLSTM networks. This model leverages the spatial features extracted by the CNN and the temporal dynamics captured by the BiLSTM. Compared to their hybrid approach, our model employs a simpler architecture that reduces computational complexity while maintaining high accuracy, specifically focusing on enhancing peak detection.
- Liang et al. (2022) [36]: This work introduced a complex architecture integrating a U-Net-like design with a BiLSTM bottleneck and additional convolutional blocks in the decoder that include parallel dilated convolutions to capture both global and local contextual information. Our model, while less complex, aims to achieve superior performance in peak detection by focusing on essential features.

Table 5.2 presents the performance comparison of the proposed model against the three aforementioned approaches in terms of commonly used evaluation metrics for segmentation tasks: precision, recall, and F1-score.

Table 5.2: Performance in sample points classification.

Model	Metric	Classes			
		Isoelectric	P	QRS	T
Moskalenko et al. [29]	Precision	0.9656	0.9077	0.9340	0.9371
	Recall	0.9678	0.8910	0.9448	0.9294
	F1-score	0.9667	0.8993	0.9394	0.9333
Londhe et al. (2021) [35]	Precision	0.9684	0.9029	0.9389	0.9330
	Recall	0.9670	0.9039	0.9429	0.9365
	F1-score	0.9677	0.9034	0.9409	0.9347
ECG_SegNet (2022) [36]	Precision	0.9693	0.9063	0.9413	0.9352
	Recall	0.9679	0.9117	0.9405	0.9393
	F1-score	**0.9686**	**0.9090**	0.9409	**0.9372**
1D U-Net (this work w/o Attention)	Precision	0.9628	0.9149	0.9420	0.9402
	Recall	0.9707	0.8851	0.9430	0.9196
	F1-score	0.9668	0.8998	**0.9425**	0.9298
Peak Attention U-Net (this work)	Precision	0.9617	0.9124	0.9495	0.9405
	Recall	0.9716	0.8892	0.9331	0.9181
	F1-score	0.9666	0.9007	0.9412	0.9292

The results are similar to those of other studies, which is reasonable given that the incorporated attention gates do not have a direct impact on the classification of all sample points. Instead, these attention gates affect small but relevant segments of the signal, such as peaks. This will be discussed in more detail in the next section.

In fact, as shown in Table 5.2, simpler models (e.g., Moskalenko et al. [29]) have very similar metrics to more complex models. This is because the sample point classification problem to segment the ECG signal into waveforms is a relatively easy task that can be solved with simple models.

The best overall performance across all classes is achieved by the ECG_SegNet model [36]. This makes sense because their use of BiLSTM in the bottleneck, combined with parallel dilated convolutions, effectively mitigates sporadic misclassifications of waveforms, as depicted in Figure 4.5. The integration of BiLSTMs helps in capturing the temporal dependencies within the ECG signals, while the parallel dilated convolutions enhance the model's ability to capture both global and local contextual information, thereby reducing the likelihood of errors in waveform segmentation.

5.3 Fiducial Point Delineation

This section presents the performance metrics of the proposed model for the fiducial point delineation task, which involves identifying specific points in ECG signals. These points include the onset (beginning of the wave), peak, and offset (end of the wave) of the various waveforms (P, QRS, and T). Thus, the delineation performance is evaluated for a total of nine different points.

Table 5.3 compares the performance of the proposed model with the three studies discussed in the previous section. Commonly used metrics in the literature for fiducial point delineation are utilized: sensitivity (SE), positive predictive value (PPV),

Table 5.3: Performance in ECG fiducial points delineation (with a tolerance of 150 ms).

Model	Metric	Classes								
		Pon	Ppeak	Poff	QRSon	Rpeak	QRSoff	Ton	Tpeak	Toff
Moskalenko et al. [29]	SE	0.8950	0.0000	0.8957	1.0000	0.0000	1.0000	0.9883	0.0000	0.9877
	PPV	0.8757	0.0000	0.8765	0.9760	0.0000	0.9760	0.9622	0.0000	0.9618
	F1-score	0.8832	0.0000	0.8840	0.9866	0.0000	0.9866	0.9735	0.0000	0.9730
	$\mu \pm \sigma$ (ms)	8.16±6.01	0.00±0.00	8.76±6.93	5.26±3.69	0.00±0.00	6.98±6.07	12.20±9.20	0.00±0.00	10.13±8.76
Londhe et al. (2021) [35]	SE	0.8948	0.5660	0.8947	0.9989	0.8543	0.9989	0.9898	0.6591	0.9892
	PPV	0.8766	0.7715	0.8764	0.9876	0.9379	0.9876	0.9544	0.8531	0.9539
	F1-score	0.8841	0.6224	0.8839	0.9926	0.8753	0.9926	0.9697	0.7098	0.9692
	$\mu \pm \sigma$ (ms)	8.02±6.19	2.47±1.38	8.61±6.97	4.99±3.69	2.47±1.31	6.20±5.13	12.23±10.31	1.76±1.30	10.17±9.14
ECG_SegNet (2022) [36]	SE	0.9010	0.6940	0.9009	1.0000	0.8344	1.0000	0.9909	0.8136	0.9903
	PPV	0.8426	0.8634	0.8427	0.9443	0.9155	0.9443	0.9011	0.9673	0.9008
	F1-score	0.8665	0.7459	0.8665	0.9669	0.8534	0.9669	0.9366	0.8606	0.9362
	$\mu \pm \sigma$ (ms)	7.87±6.22	3.29±2.71	9.52±9.59	5.22±3.85	2.37±1.41	8.53±8.10	12.41±9.73	2.90±3.06	11.86±10.89
1D U-Net (this work w/o Attention)	SE	0.8898	0.6887	0.8911	1.0000	0.8263	1.0000	0.9863	0.6985	0.9849
	PPV	0.8475	0.8685	0.8492	0.9666	0.8828	0.9666	0.9182	0.9270	0.9171
	F1-score	0.8638	0.7426	0.8653	0.9807	0.8409	0.9807	0.9461	0.7628	0.9450
	$\mu \pm \sigma$ (ms)	8.23±6.62	4.11±2.38	10.56±9.07	4.81±3.59	1.74±1.24	7.23±6.42	12.55±10.06	2.09±1.44	12.72±11.87
Peak Attention U-Net (this work)	SE	0.8880	0.8877	0.8893	1.0000	0.9996	1.0000	0.9837	0.9826	0.9830
	PPV	0.9008	0.9014	0.9028	0.9970	0.9953	0.9970	0.9719	0.9704	0.9712
	F1-score	**0.8925**	**0.8926**	**0.8940**	**0.9984**	**0.9972**	**0.9984**	**0.9764**	**0.9751**	**0.9757**
	$\mu \pm \sigma$ (ms)	8.10±6.17	8.75±7.41	7.70±5.37	4.97±3.65	6.44±2.53	5.86±4.20	12.51±10.32	7.18±8.98	10.21±8.45

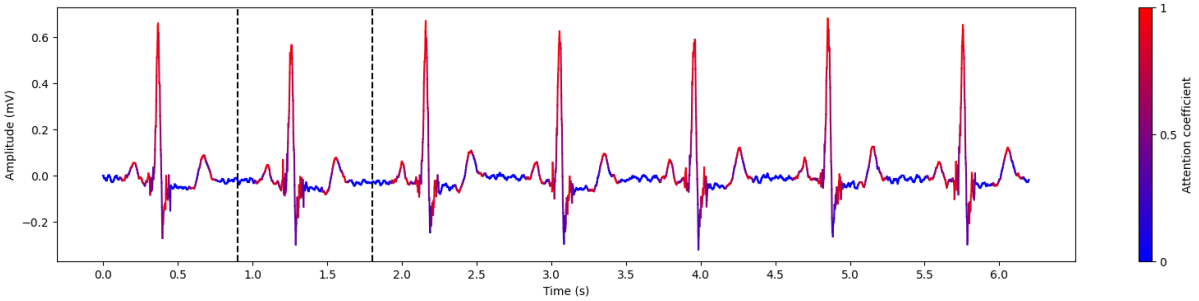
F1-score, and the mean and standard deviation of the distance between the predicted and actual points ($\mu \pm \sigma$). As detailed in Section 4.7.2, a predicted point is considered correct if its distance to the actual delineated point (label) is less than 150 milliseconds. This tolerance is established according to the ANSI/AAMI-EC57:1998-2012 standards [69], which is a widely accepted practice in the literature.

The results of the proposed model in fiducial point delineation surpass those of other studies significantly. This improvement is attributed to the incorporation of attention gates, which enable the model to focus on small but critical parts of the ECG signal for delineating these points.

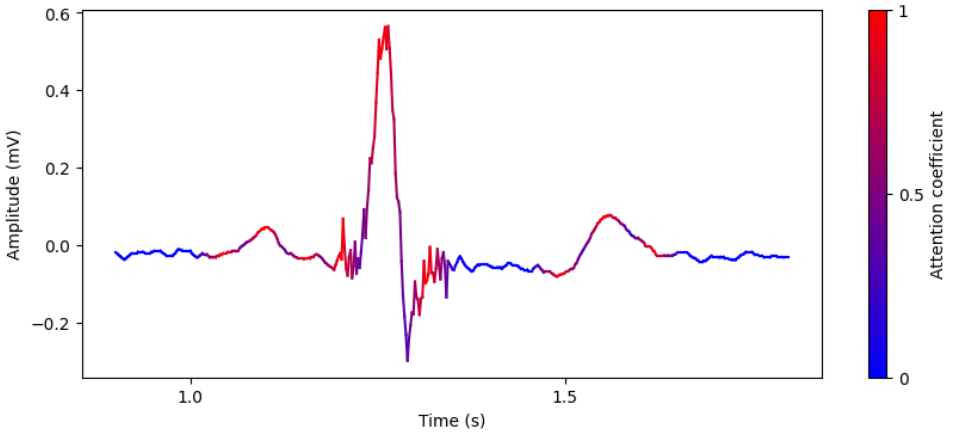
As shown in Table 5.3, the most basic model (the one proposed by Moskalenko et al. [29]), fails to detect any peaks. However, it can detect wave onsets and offsets, as these are derived from the boundaries of the predicted sample point segments, as discussed in the previous section.

More sophisticated models, like those of Londhe et al. [35] and Liang et al. [36], can detect peaks to some extent. In fact, their results are comparable to those of our proposed 1D U-Net model without attention gates.

However, the addition of attention gates to our model (Peak Attention U-Net) leads to a significant improvement in fiducial point delineation performance. Our model achieves the highest F1-scores across all classes. While the detection of onsets and offsets shows little difference compared to other models (since they are extracted from the boundaries of the predicted sample point segments), the delineation performance for peaks is greatly improved. Our model attains F1-scores of 89.26%, 99.72%, and 97.57% for the P peaks, R peaks, and T peaks, respectively. The variation in performance among the peaks is due to the differing complexities in delineating each type. The R peak is easily characterized by its high-frequency shape, and the T peak by its larger and broader waveform. In contrast, the P peak is more challenging to detect due to its short and subtle waveform, which can sometimes be indistinguishable from noise in the signal.



(a) Attention coefficients of an ECG signal.



(b) Zoomed-in view of the attention coefficients of a single heartbeat (segment between lines in Figure 5.2a).

Figure 5.2: Attention coefficients of the last attention gate in the decoder.

To provide further insight into the model's performance and the impact of attention gates, Figure 5.2 displays the attention coefficients of the last attention gate in the decoder applied to an ECG signal of the test set (i.e., the model has not seen this example during training). This visualization illustrates the model's focus on specific parts of the signal, with red regions indicating high attention and blue regions indicating low attention. Figure 5.2a shows the attention coefficients of the entire ECG signal, and it can be observed that the model focuses on key points of the signal around the waves, while the isoelectric segments receive less attention. Figure 5.2b provides a zoomed-in view of the attention coefficients for a single heartbeat, highlighting the model's ability to focus on the onset, peak, and offset of the waveforms. On the other hand, intra- and inter-wave segments receive less attention, as they are less critical for delineation. This visualization demonstrates the effectiveness of attention gates in enhancing the model's focus on important parts of the signal, leading to improved delineation performance.

5.4 Discussion

The proposed model's performance in the sample point classification task, as evaluated on the LUDB dataset, is comparable to that of other relevant models in the literature. Despite the introduction of attention gates, the overall performance metrics remain similar to those of the other models. This suggests that the task of segment-

ing ECG signals into waveforms is sufficiently straightforward, allowing even simpler models to achieve high performance. Consequently, while more complex architectures can offer marginal improvements in specific aspects of the segmentation task, the fundamental classification of sample points in ECG signals does not necessarily require highly sophisticated models.

In contrast, the fiducial point delineation task highlights the strengths of our model. The introduction of attention gates significantly enhances the model's ability to focus on small but important parts of the signal (i.e., peaks), leading to superior delineation performance. Precise identification of peaks is crucial for extracting signal features such as heart rate, PR interval, QT interval, and ST segment duration, which are essential for meaningful and useful delineation.

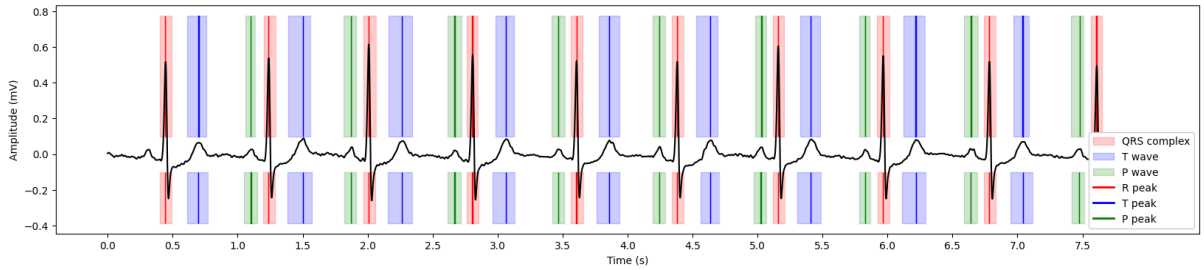
To illustrate the effectiveness of attention gates in the peak detection task, let us consider an analogy with a hypothetical 2D image segmentation task: detecting peaks in ECG signals is similar to segmenting a very specific region of a single pixel based on the surrounding context. Attention gates are particularly beneficial in this scenario, as they have been shown to effectively improve the model's focus on these critical details [55], as demonstrated by the experimental results.

Examples of the proposed model output for different cases and pathologies are presented below. These examples highlight the model's ability to delineate full ECG signals, which was a primary goal of this study. Unlike other studies that preprocess the data by splitting the signal into individual heartbeats—a simpler task since it involves delineating single heartbeats into their three characteristic waves—the proposed approach directly handles complete ECG signals. This capability is crucial for real-world applications, where ECG signals are not easily divided into individual heartbeats, thus improving the practical utility and robustness of the model.

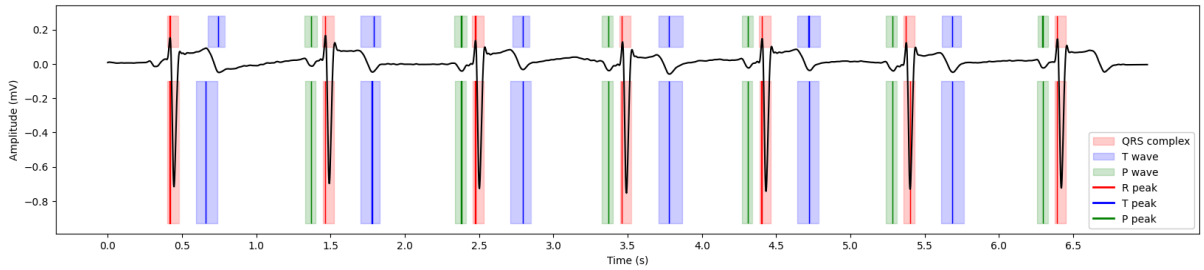
Additionally, these examples demonstrate our model's ability to detect waveform peaks. This was another key objective, as many studies only segment waveforms due to the challenge of accurately segmenting such small regions as peaks, which consist of a single sample point. This accuracy is achieved through the incorporation of attention gates, which enable the model to focus on these small but significant parts of the signal.

In the figures that follow, the labels above the ECG signal represent the ground truth, as annotated by expert cardiologists. The labels below the ECG signal represent the predictions made by the model. Also note that all these examples are extracted from the test set, so the model has not seen these examples while training.

The first example, presented in Figure 5.3, includes two instances of a healthy ECG, both of which are accurately delineated with precisely detected peaks. Figure 5.3a displays an ECG with positive waves, while Figure 5.3b shows an ECG with negative waves. These examples demonstrate the model's capability to delineate varying waveforms from different ECG leads. Accurately delineating an entire ECG, regardless of the lead, was a key objective of this study as it facilitates the practical application of the model in various clinical settings.



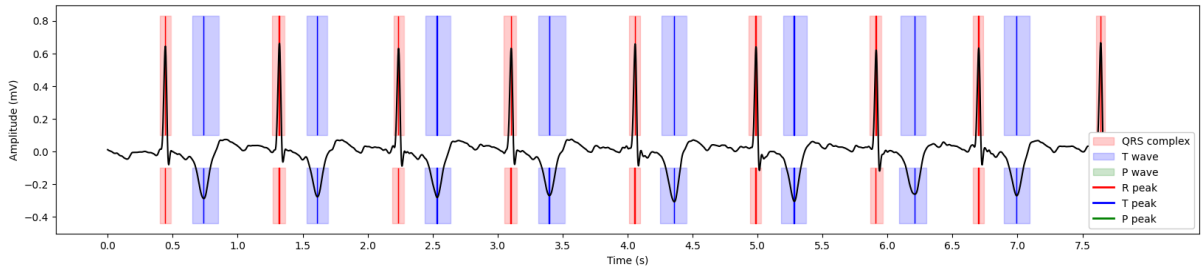
(a) Healthy with positive waves.



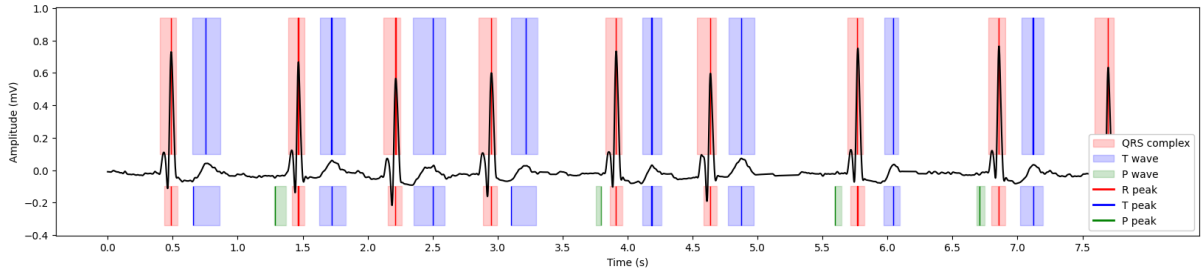
(b) Healthy with negative waves.

Figure 5.3: Healthy ECG examples with different waveform morphologies.

Figure 5.4 presents two examples of ECGs with atrial fibrillation, a cardiac pathology characterized by the absence of P waves. In Figure 5.4a, the example is correctly delineated, with no P waves predicted by the model. However, Figure 5.4b shows an example with some delineation errors, as three heartbeats erroneously include P waves. Most of the false positive P peaks reported in Table 5.3 can be attributed to this issue of detecting P waves in atrial fibrillation cases.



(a) Atrial fibrillation case correctly delineated.



(b) Atrial fibrillation case wrongly delineated.

Figure 5.4: Atrial fibrillation examples.

Figure 5.5 illustrates an extreme case. Examining the third heartbeat around the 2.5-second mark, we observe it is unusually close to the preceding heartbeat and significantly distant from the subsequent one. Additionally, the P wave of the third heartbeat is negative, in contrast to the other P waves. This cardiac condition is known as supraventricular extrasystole with compensatory pause. Despite the complexity and rarity of this example, the proposed model accurately delineates it. This demonstrates the model's robustness and ability to handle challenging cases, even those with unique and atypical characteristics.

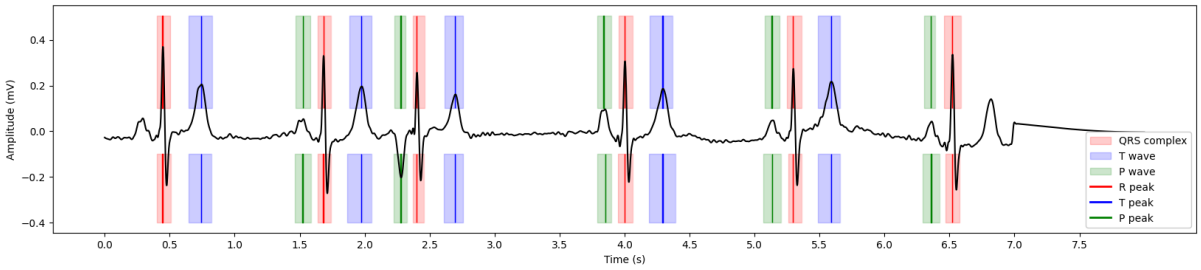
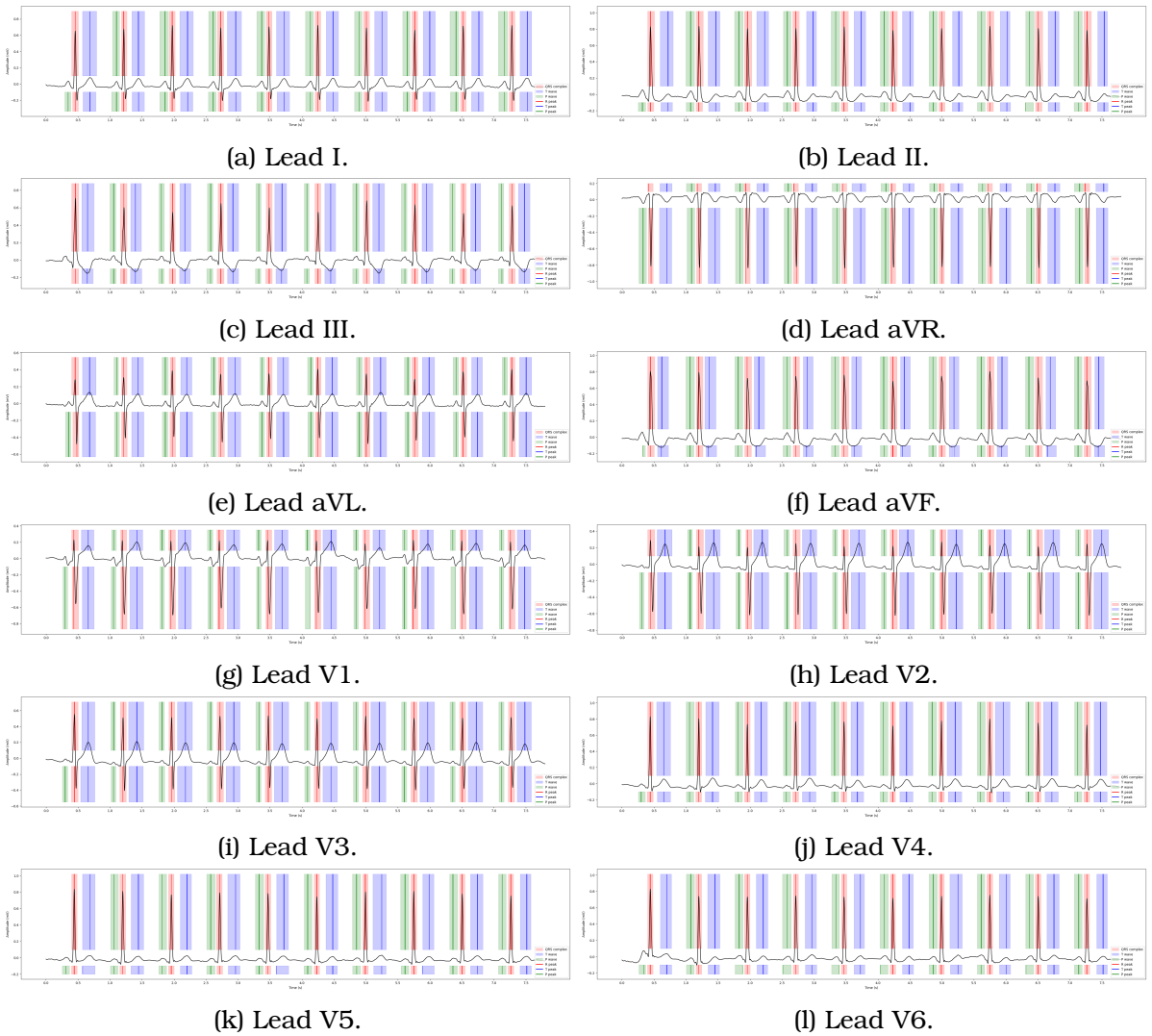


Figure 5.5: Supraventricular extrasystole with compensatory pause case correctly delineated.

In light of the results and examples presented, we can confidently confirm our initial hypothesis. The incorporation of an attention mechanism into our Peak Attention U-Net has indeed significantly improved ECG delineation, particularly in enhancing peak detection capabilities. This addresses a key limitation observed in current state-of-the-art deep learning approaches. The superior performance in the fiducial point delineation task, as discussed earlier, directly supports this aspect of our hypothesis. Furthermore, our results confirm the second part of our hypothesis: our model has successfully been trained to accurately delineate complete ECG signals from individual leads, enabling the delineation of full 12-lead ECGs.

Figure 5.6 illustrates all the novel capabilities of the Peak Attention U-Net, serving as a comprehensive validation of our hypothesis. It showcases the model's ability to delineate peaks, process entire signals, and handle complete 12-lead ECGs. This achievement addresses the limitations identified in the current literature (see Section 2.2) and represents a significant advancement in ECG analysis, with direct implications for clinical applications.



(a) Lead I.

(b) Lead II.

(c) Lead III.

(d) Lead aVR.

(e) Lead aVL.

(f) Lead aVF.

(g) Lead V1.

(h) Lead V2.

(i) Lead V3.

(j) Lead V4.

(k) Lead V5.

(l) Lead V6.

Figure 5.6: Complete 12-lead ECG delineated by the proposed model, which demonstrates its ability to: (1) delineate peaks; (2) delineate entire signals; and (3) delineate the complete 12-lead ECG.

Chapter 6

Conclusions

This Master's Thesis proposes and validates a deep learning model for ECG delineation (Peak Attention U-Net) that successfully addresses several limitations that persisted in the deep learning literature and limited the applicability of these models in real clinical scenarios. The primary objective of this work was to overcome these challenges, thereby facilitating the clinical use of automatic ECG delineation models.

One of the key contributions of this work is the successful solution to the problem of peak detection, which remained a challenge in the literature. The proposed Peak Attention U-Net model incorporates an attention mechanism into the decoder path of the U-Net, allowing it to selectively focus on crucial segments of the ECG signal, such as peaks. This targeted attention enhances the fiducial point delineation performance of the model, particularly in the detection of P, R, and T peaks, which are essential for diagnosing various cardiac conditions. Compared to recent and relevant studies in the literature, the Peak Attention U-Net demonstrates a significant improvement of over 12.5% in F1 score for peak detection. This is a crucial advancement as accurate peak detection is essential for meaningful ECG interpretation and clinical decision-making, suggesting the model's potential for real-world clinical applications, pending further validation and clinical trials.

Furthermore, a robust delineation framework has been proposed, capable of processing entire ECG signals rather than individual heartbeats. This approach eliminates the need for prior segmentation of the signal, as typically done in other studies of the literature. This simplification of the overall delineation process could potentially enhance its accessibility for clinical application, subject to rigorous testing in clinical settings.

Moreover, the Peak Attention U-Net is designed to be lead-agnostic, meaning it can delineate ECG signals from any lead, enabling the delineation of complete 12-lead ECGs. This flexibility could broaden the applicability of the model, potentially making it a versatile tool for various clinical scenarios, though further research is needed to confirm its efficacy across different clinical contexts.

The proposed model is not only reasonably sized, but also highly efficient in terms of computational requirements and inference speed. Specifically, the Peak Attention U-Net model occupies only 1.8 MB of disk space and can delineate a 10-second ECG signal in just 4 milliseconds on a standard laptop. This makes it particularly suitable for real-world clinical applications, where fast response is often critical for effective

patient care. With the growing popularity of wearable health monitoring devices and the resulting increase in data volume, the efficiency of the model is especially important in today's era of big data and the Internet of Medical Things (IoMT), where efficient and accurate interpretation of data is essential to improve healthcare outcomes [70].

In conclusion, the contributions of this work can be summarized as follows:

- The development of a **novel attention-enhanced U-Net model**, named **Peak Attention U-Net**, which significantly improves the performance of fiducial point delineation, particularly in the **detection of P, R, and T peaks**.
- The proposal of a **robust delineation framework** capable of **processing entire ECG signals**, eliminating the need for prior segmentation into individual heartbeats.
- The design of a **lead-agnostic model** that can **delineate ECG signals from any lead**, enabling the **delineation of complete 12-lead ECGs**.
- The development of a **reasonably sized and highly efficient model** suitable for **real-world clinical applications**.
- The validation of the model across **diverse patient demographics and diseases**, demonstrating its **robustness and generalizability**.

This thesis presents a significant advancement in the field of ECG delineation, offering a model that is accurate, efficient, and applicable to a wide range of clinical scenarios. The proposed Peak Attention U-Net model has the potential to revolutionize the field of ECG analysis, enabling faster and more accurate interpretation of ECG signals, which is essential for the early detection and diagnosis of various cardiac conditions. We believe that the contributions made in this work will pave the way for further research, leading to improved patient care and better health outcomes through the effective use of artificial intelligence.

6.1 Limitations

While the proposed Peak Attention U-Net model demonstrates significant advancements in ECG delineation, it is important to acknowledge certain limitations of this research.

The primary limitation of this study is the lack of large, accurately delineated ECG databases for model training. Creating such databases requires a substantial investment of time and effort from highly qualified cardiologists, making it a challenging and costly process. In this study, we utilized a dataset collected from a single hospital, which may introduce potential biases in terms of population demographics, prevalent cardiac conditions, and ECG recording practices specific to that institution. This limitation could affect the model's generalizability across diverse patient populations, varied healthcare settings, and different geographical regions. Furthermore, the dataset may not fully represent the wide spectrum of ECG morphologies and rare cardiac conditions encountered in global clinical practice. The single-source nature of the data also means that any systematic biases in annotation practices or equipment calibration specific to that hospital could influence the model's learning and

performance. These factors collectively highlight the need for more diverse, multi-center ECG databases to enhance the robustness and applicability of deep learning models in ECG analysis.

Moreover, while the model shows promising results in controlled research settings, extensive testing in real-world clinical environments is necessary to fully assess its performance and practical utility. The model's performance may be influenced by factors such as signal noise, artifacts, and variations in ECG recording quality, which are common in clinical practice.

Finally, while the model shows improved performance compared to existing studies, a comprehensive comparison with human expert delineation across a large and diverse dataset would provide additional validation of the model's clinical utility.

6.2 Future Work

Given the promising results obtained in this work, several avenues for future research can be explored to further enhance the proposed model and its applicability in real-world clinical scenarios.

- **Publication of Results**. The first step after this Master's Thesis is to prepare a paper for publication in a scientific journal or conference, so that the research reaches a broader audience and contributes to the existing literature in the field of ECG delineation.
- **Validation Across Future Datasets**. While our model performs exceptionally well in delineating fiducial points, it is essential to consider the variations in ECG signals across different populations and conditions. Future work should focus on validating the model across diverse datasets to ensure its robustness and generalizability. However, obtaining annotated datasets for ECG delineation is a complex task. This is because the process of annotating ECG signals is time-consuming and requires the expertise of cardiologists, making it a costly and challenging endeavor. The complexity of ECG signals and the need for precise annotations contribute to the difficulty in creating large, accurately labeled datasets.
- **Model Enhancements**. Additionally, exploring further enhancements in attention mechanisms and hybrid architectures could lead to even more accurate and efficient ECG analysis models.
- **Mobile Application**. As for the real-world applicability of our model, it could be integrated into mobile applications that record and delineate ECG signals with the use of wearable devices. Future work could also explore the implementation of this application to conduct a clinical study involving the proposed method.
- **Federated Learning**. Given the lack of delineated datasets and the mobile applicability, it would also be interesting to explore the concept of federated learning to further train the Peak Attention U-Net model. Federated learning [71] is a machine learning approach that allows a model to be trained across multiple decentralized devices or servers holding local data samples, without exchanging them. This method enhances privacy and security, as the data remains on the user's device while only the model updates are shared. By utilizing federated

learning, data from multiple mobile users can be leveraged to iteratively improve our ECG delineation model. This decentralized approach not only addresses the challenge of obtaining large annotated datasets but also ensures that the model evolves and adapts to a wide variety of ECG signals encountered in real-world scenarios accross different populations and conditions.

- **Feature-based Classification**. Further research should also consider the high quality fiducial point delineation achieved in this work to train feature-based classification models to classify ECGs according to their diagnosis. Numerous works have studied the ECG classification based on deep learning [72, 73, 74, 75]. A promising research avenue involves comparing the performance of raw ECG signal classification with deep learning against a feature-based machine learning classification with the features extracted from the delineation (i.e., variations in RR, PR, or QT intervals). This would significantly enhance the medical support of the proposed method, as it would not only include the delineated ECG signal but also a classification serving as a pre-diagnostic tool to assist health professionals.

Bibliography

[1] World Health Organization. *Cardiovascular diseases (CVDs)*. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] Wikipedia. *Arritmia sinusal respiratoria*. Accessed: 2024-04-04. 2024. URL: https://es.wikipedia.org/wiki/Arritmia_sinusal_respiratoria.

[3] Wikipedia. *Electrocardiography*. Accessed: 2024-04-04. 2024. URL: https://en.wikipedia.org/wiki/Electrocardiography.

[4] Maurice Lev. "Anatomic basis for atrioventricular block". In: *The American Journal of Medicine* 37.5 (Nov. 1964), pp. 742–748. ISSN: 0002-9343. DOI: 10.1016/0002-9343(64)90022-1. URL: http://dx.doi.org/10.1016/0002-9343(64)90022-1.

[5] Martin M Bednar et al. "The QT interval". In: *Progress in cardiovascular diseases* 43.5 (2001), pp. 1–45.

[6] Carmen Anna Maria Spaccarotella et al. "Multichannel Electrocardiograms Obtained by a Smartwatch for the Diagnosis of ST-Segment Changes". In: *JAMA Cardiology* 5.10 (Oct. 2020), pp. 1176–1180. ISSN: 2380-6583. DOI: 10.1001/jamacardio.2020.3994. URL: https://doi.org/10.1001/jamacardio.2020.3994.

[7] Jiapu Pan and Willis J. Tompkins. "A Real-Time QRS Detection Algorithm". In: *IEEE Transactions on Biomedical Engineering* BME-32.3 (Mar. 1985), pp. 230–236. ISSN: 0018-9294. DOI: 10.1109/tbme.1985.325532. URL: http://dx.doi.org/10.1109/TBME.1985.325532.

[8] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. "Detection of ECG characteristic points using wavelet transforms". In: *IEEE Transactions on Biomedical Engineering* 42.1 (1995), pp. 21–28. ISSN: 0018-9294. DOI: 10.1109/10.362922. URL: http://dx.doi.org/10.1109/10.362922.

[9] J.P. Martinez et al. "A Wavelet-Based ECG Delineator: Evaluation on Standard Databases". In: *IEEE Transactions on Biomedical Engineering* 51.4 (Apr. 2004), pp. 570–581. ISSN: 0018-9294. DOI: 10.1109/tbme.2003.821031. URL: http://dx.doi.org/10.1109/TBME.2003.821031.

[10] A. I. Kalyakulina et al. "Finding Morphology Points of Electrocardiographic-Signal Waves Using Wavelet Analysis". In: *Radiophysics and Quantum Electronics* 61.8–9 (Jan. 2019), pp. 689–703. ISSN: 1573-9120. DOI: 10.1007/s11141-019-09929-2. URL: http://dx.doi.org/10.1007/s11141-019-09929-2.

[11] D.S. Benitez et al. "A new QRS detection algorithm based on the Hilbert transform". In: *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*. CIC-00. IEEE. DOI: 10.1109/cic.2000.898536. URL: http://dx.doi.org/10.1109/CIC.2000.898536.

[12] S. K. Mukhopadhyay, M. Mitra, and S. Mitra. “Time plane ECG feature extraction using Hilbert transform, variable threshold and slope reversal approach”. In: *2011 International Conference on Communication and Industrial Application*. IEEE, Dec. 2011. DOI: 10.1109/iccinda.2011.6146675. URL: http://dx.doi.org/10.1109/ICCIndA.2011.6146675.

[13] A Martínez, R Alcaraz, and J J Rieta. “Automatic electrocardiogram delineator based on the Phasor Transform of single lead recordings”. In: *2010 Computing in Cardiology*. 2010, pp. 987–990.

[14] S. Graja and J.-M. Boucher. “Hidden Markov Tree Model Applied to ECG Delineation”. In: *IEEE Transactions on Instrumentation and Measurement* 54.6 (Dec. 2005), pp. 2163–2168. ISSN: 0018-9456. DOI: 10.1109/tim.2005.858568. URL: http://dx.doi.org/10.1109/TIM.2005.858568.

[15] Mahsa Akhbari et al. “ECG segmentation and fiducial point extraction using multi hidden Markov model”. In: *Computers in Biology and Medicine* 79 (Dec. 2016), pp. 21–29. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2016.09.004. URL: http://dx.doi.org/10.1016/j.compbiomed.2016.09.004.

[16] Rémi Dubois et al. “Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators”. In: *Computer Methods and Programs in Biomedicine* 88.3 (Dec. 2007), pp. 217–233. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2007.09.005. URL: http://dx.doi.org/10.1016/j.cmpb.2007.09.005.

[17] Pablo Laguna et al. *The QT Database*. 1997. DOI: 10.13026/C24K53. URL: https://physionet.org/content/qtdb/.

[18] Alena Kalyakulina et al. *Lobachevsky University Electrocardiography Database*. 2021. DOI: 10.13026/EEGM-H675. URL: https://physionet.org/content/ludb/1.0.1/.

[19] F. Rincón et al. “Development and Evaluation of Multilead Wavelet-Based ECG Delineation Algorithms for Embedded Wireless Sensor Nodes”. In: *IEEE Transactions on Information Technology in Biomedicine* 15.6 (Nov. 2011), pp. 854–863. ISSN: 1558-0032. DOI: 10.1109/titb.2011.2163943. URL: http://dx.doi.org/10.1109/TITB.2011.2163943.

[20] Jose Manuel Bote et al. “A Modular Low-Complexity ECG Delineation Algorithm for Real-Time Embedded Systems”. In: *IEEE Journal of Biomedical and Health Informatics* 22.2 (Mar. 2018), pp. 429–441. ISSN: 2168-2208. DOI: 10.1109/jbhi.2017.2671443. URL: http://dx.doi.org/10.1109/JBHI.2017.2671443.

[21] Qibin Zhao and Liqing Zhang. “ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines”. In: *2005 International Conference on Neural Networks and Brain*. Vol. 2. 2005, pp. 1089–1092. DOI: 10.1109/ICNNB.2005.1614807.

[22] F. Melgani and Y. Bazi. “Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization”. In: *IEEE Transactions on Information Technology in Biomedicine* 12.5 (Sept. 2008), pp. 667–677. ISSN: 1089-7771. DOI: 10.1109/titb.2008.923147. URL: http://dx.doi.org/10.1109/TITB.2008.923147.

[23] L.-Y. Shyu, Y.-H. Wu, and W. Hu. “Using Wavelet Transform and Fuzzy Neural Network for VPC Detection From the Holter ECG”. In: *IEEE Transactions on Biomedical Engineering* 51.7 (July 2004), pp. 1269–1273. ISSN: 0018-9294. DOI:

10.1109/tbme.2004.824131. URL: http://dx.doi.org/10.1109/TBME.2004.824131.

[24] P. deChazal, M. O'Dwyer, and R.B. Reilly. "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features". In: *IEEE Transactions on Biomedical Engineering* 51.7 (July 2004), pp. 1196–1206. ISSN: 0018-9294. DOI: 10.1109/tbme.2004.827359. URL: http://dx.doi.org/10.1109/TBME.2004.827359.

[25] Abdolrahman Peimankar and Sadasivan Puthusserypady. "An Ensemble of Deep Recurrent Neural Networks for P-wave Detection in Electrocardiogram". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2019. DOI: 10.1109/icassp.2019.8682307. URL: http://dx.doi.org/10.1109/icassp.2019.8682307.

[26] Siti Nurmaini et al. "Electrocardiogram signal classification for automated delineation using bidirectional long short-term memory". In: *Informatics in Medicine Unlocked* 22 (2021), p. 100507. ISSN: 2352-9148. DOI: 10.1016/j.imu.2020.100507. URL: http://dx.doi.org/10.1016/j.imu.2020.100507.

[27] Abdolrahman Peimankar and Sadasivan Puthusserypady. "DENS-ECG: A deep learning approach for ECG signal delineation". In: *Expert Systems with Applications* 165 (Mar. 2021), p. 113911. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2020.113911. URL: http://dx.doi.org/10.1016/j.eswa.2020.113911.

[28] Siti Nurmaini et al. "Robust electrocardiogram delineation model for automatic morphological abnormality interpretation". In: *Scientific Reports* 13.1 (Aug. 2023). ISSN: 2045-2322. DOI: 10.1038/s41598-023-40965-1. URL: http://dx.doi.org/10.1038/s41598-023-40965-1.

[29] Viktor Moskalenko, Nikolai Zolotykh, and Grigory Osipov. "Deep Learning for ECG Segmentation". In: *Advances in Neural Computation, Machine Learning, and Cognitive Research III*. Springer International Publishing, Sept. 2019, pp. 246–254. ISBN: 9783030304256. DOI: 10.1007/978-3-030-30425-6_29. URL: http://dx.doi.org/10.1007/978-3-030-30425-6_29.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241. ISBN: 9783319245744. DOI: 10.1007/978-3-319-24574-4_28. URL: http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[31] Iana Sereda et al. *ECG Segmentation by Neural Networks: Errors and Correction*. 2018. DOI: 10.48550/ARXIV.1812.10386. URL: https://arxiv.org/abs/1812.10386.

[32] Guillermo Jimenez-Perez, Alejandro Alcaine, and Oscar Camara. "U-Net Architecture for the Automatic Detection and Delineation of the Electrocardiogram". In: *Computing in Cardiology Conference (CinC)*. CinC2019. Computing in Cardiology, Dec. 2019. DOI: 10.22489/cinc.2019.284. URL: http://dx.doi.org/10.22489/CinC.2019.284.

[33] Guillermo Jimenez-Perez, Alejandro Alcaine, and Oscar Camara. "Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks". In: *Scientific Reports* 11.1 (Jan. 2021). ISSN: 2045-2322. DOI: 10.1038/s41598-020-79512-7. URL: http://dx.doi.org/10.1038/s41598-020-79512-7.

[34] Zhenqin Chen et al. "Post-processing refined ECG delineation based on 1D-UNet". In: *Biomedical Signal Processing and Control* 79 (Jan. 2023), p. 104106.

ISSN: 1746-8094. DOI: 10.1016/j.bspc.2022.104106. URL: http://dx.doi.org/10.1016/j.bspc.2022.104106.

[35] Aboli N. Londhe and Mithilesh Atulkar. "Semantic segmentation of ECG waves using hybrid channel-mix convolutional and bidirectional LSTM". In: *Biomedical Signal Processing and Control* 63 (Jan. 2021), p. 102162. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2020.102162. URL: http://dx.doi.org/10.1016/j.bspc.2020.102162.

[36] Xiaohong Liang et al. "ECG_SegNet: An ECG delineation model based on the encoder-decoder structure". In: *Computers in Biology and Medicine* 145 (June 2022), p. 105445. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2022.105445. URL: http://dx.doi.org/10.1016/j.compbiomed.2022.105445.

[37] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (May 2015), pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. URL: http://dx.doi.org/10.1038/nature14539.

[38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).

[39] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. DOI: 10.48550/ARXIV.1409.1556. URL: https://arxiv.org/abs/1409.1556.

[40] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. DOI: 10.48550/ARXIV.1512.03385. URL: https://arxiv.org/abs/1512.03385.

[41] Christian Szegedy et al. *Going Deeper with Convolutions*. 2014. DOI: 10.48550/ARXIV.1409.4842. URL: https://arxiv.org/abs/1409.4842.

[42] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2014. DOI: 10.48550/ARXIV.1409.0473. URL: https://arxiv.org/abs/1409.0473.

[43] Volodymyr Mnih et al. *Recurrent Models of Visual Attention*. 2014. DOI: 10.48550/ARXIV.1406.6247. URL: https://arxiv.org/abs/1406.6247.

[44] Petros-Pavlos Ypsilantis and Giovanni Montana. *Learning what to look in chest X-rays with a recurrent visual attention model*. 2017. DOI: 10.48550/ARXIV.1701.06452. URL: https://arxiv.org/abs/1701.06452.

[45] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. ISSN: 0006-3444. DOI: 10.1093/biomet/57.1.97. URL: http://dx.doi.org/10.1093/biomet/57.1.97.

[46] Tao Shen et al. *DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding*. 2017. DOI: 10.48550/ARXIV.1709.04696. URL: https://arxiv.org/abs/1709.04696.

[47] Saumya Jetley et al. *Learn To Pay Attention*. 2018. DOI: 10.48550/ARXIV.1804.02391. URL: https://arxiv.org/abs/1804.02391.

[48] Fei Wang et al. *Residual Attention Network for Image Classification*. 2017. DOI: 10.48550/ARXIV.1704.06904. URL: https://arxiv.org/abs/1704.06904.

[49] Jie Hu et al. *Squeeze-and-Excitation Networks*. 2017. DOI: 10.48550/ARXIV.1709.01507. URL: https://arxiv.org/abs/1709.01507.

[50] Olga Russakovsky et al. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3 (2017), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

[51] Xiaolong Wang et al. *Non-local Neural Networks*. 2017. DOI: 10.48550/ARXIV.1711.07971. URL: https://arxiv.org/abs/1711.07971.

[52] Ashish Vaswani et al. *Attention Is All You Need*. 2017. DOI: 10.48550/ARXIV.1706.03762. URL: https://arxiv.org/abs/1706.03762.

[53] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. DOI: 10.48550/ARXIV.1810.04805. URL: https://arxiv.org/abs/1810.04805.

[54] Alec Radford et al. *Improving Language Understanding by Generative Pre-Training*. 2018.

[55] Ozan Oktay et al. *Attention U-Net: Learning Where to Look for the Pancreas*. 2018. DOI: 10.48550/ARXIV.1804.03999. URL: https://arxiv.org/abs/1804.03999.

[56] George B Moody and Roger G Mark. *MIT-BIH Arrhythmia Database*. 1992. DOI: 10.13026/C2F305. URL: https://physionet.org/content/mitdb/.

[57] Alessandro Taddei et al. *European ST-T Database*. 2000. DOI: 10.13026/C2D59Z. URL: https://physionet.org/content/edb/.

[58] Stephen Butterworth. "On the Theory of Filter Amplifiers". In: *Experimental Wireless & the Wireless Engineer* 7.6 (Oct. 1930), pp. 536–541.

[59] Wenjie Luo et al. *Understanding the Effective Receptive Field in Deep Convolutional Neural Networks*. 2017. DOI: 10.48550/ARXIV.1701.04128. URL: https://arxiv.org/abs/1701.04128.

[60] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980. URL: https://arxiv.org/abs/1412.6980.

[61] Ning Qian. "On the momentum term in gradient descent learning algorithms". In: *Neural Networks* 12.1 (Jan. 1999), pp. 145–151. ISSN: 0893-6080. DOI: 10.1016/s0893-6080(98)00116-6. URL: http://dx.doi.org/10.1016/S0893-6080(98)00116-6.

[62] John Duchi, Elad Hazan, and Yoram Singer. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *Journal of Machine Learning Research* 12.61 (2011), pp. 2121–2159. URL: http://jmlr.org/papers/v12/duchi11a.html.

[63] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. *Neural Networks for Machine Learning: Lecture 6e - RMSProp: Divide the gradient by a running average of its recent magnitude*. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. 2012.

[64] François Chollet et al. *Keras*. https://keras.io. 2015.

[65] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[66] Steven A. Guidera and Jonathan S. Steinberg. "The signal-averaged P wave duration: A rapid and noninvasive marker of risk of atrial fibrillation". In: *Journal of the American College of Cardiology* 21.7 (June 1993), pp. 1645–1651. ISSN: 0735-1097. DOI: 10.1016/0735-1097(93)90381-a. URL: http://dx.doi.org/10.1016/0735-1097(93)90381-a.

[67] C Saritha, V Sukanya, and Y Narasimha Murthy. "ECG signal analysis using wavelet transforms". In: *Bulg. J. Phys* 35.1 (2008), pp. 68–77.

[68] Robert B Northrop. *Non-invasive instrumentation and measurement in medical diagnosis*. CRC press, 2017.

[69] Association for the Advancement of Medical Instrumentation. *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. ANSI/AAMI EC57. Arlington, VA, USA, 2012.

[70] Gulraiz J. Joyia et al. "Internet of Medical Things (IOMT): Applications, Benefits and Future Challenges in Healthcare Domain". In: *Journal of Communications* (2017). ISSN: 1796-2021. DOI: 10.12720/jcm.12.4.240-247. URL: http://dx.doi.org/10.12720/jcm.12.4.240-247.

[71] Nicola Rieke et al. "The future of digital health with federated learning". In: *npj Digital Medicine* 3.1 (Sept. 2020). ISSN: 2398-6352. DOI: 10.1038/s41746-020-00323-1. URL: http://dx.doi.org/10.1038/s41746-020-00323-1.

[72] Awni Y. Hannun et al. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". In: *Nature Medicine* 25.1 (Jan. 2019), pp. 65–69. ISSN: 1546-170X. DOI: 10.1038/s41591-018-0268-3. URL: http://dx.doi.org/10.1038/s41591-018-0268-3.

[73] Antônio H. Ribeiro et al. "Automatic diagnosis of the 12-lead ECG using a deep neural network". In: *Nature Communications* 11.1 (Apr. 2020). ISSN: 2041-1723. DOI: 10.1038/s41467-020-15432-4. URL: http://dx.doi.org/10.1038/s41467-020-15432-4.

[74] Jikuo Wang et al. "Automated ECG classification using a non-local convolutional block attention module". In: *Computer Methods and Programs in Biomedicine* 203 (May 2021), p. 106006. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2021.106006. URL: http://dx.doi.org/10.1016/j.cmpb.2021.106006.

[75] Khiem H. Le et al. "LightX3ECG: A Lightweight and eXplainable Deep Learning System for 3-lead Electrocardiogram Classification". In: *Biomedical Signal Processing and Control* 85 (Aug. 2023), p. 104963. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2023.104963. URL: http://dx.doi.org/10.1016/j.bspc.2023.104963.