Int. J. Adv. Res. Sci. Technol. Volume 12, Issue 11, 2023, pp.1182-1196

IJARST

International Journal of Advanced Research in Science and Technology

journal homepage: www.ijarst.com

ISSN 2319 – 1783 (Print)
ISSN 2320 – 1126 (Online)

ECG Signal Classification with Convolutional Neural Networks

SABIR MUHAMMAD*[1], ARSLAN[2], IMRAN KHAN[3]
Tiangong University Tianjin, China
muhammadsabir22@gmail.com

ARTICLEINFO

Article history:
Received 15 Nov 2023
Accepted 19 Nov 2023
Available online 02 Dec 2023

Keywords:
ECG,
LSTM,
Convolutional Neural Networks,
Deep Learning.

ABSTRACT

Cardiovascular waves in electrocardiograms (ECG) give significant data on heart conditions and the impacts of heart drugs. Division and analysis of ECG and its constituent cardiovascular waves are of high importance in cardiology determination and drug studies. Customarily, thoroughly prepared clinicians and cardiologists have performed ECG investigations. In any case, using clinicians in enormous scope ECG screening, for example, drug test stages or populace-based screening programs is not monetarily achievable. Consequently, an automated ECG division approach that can fragment heart waves accurately is of high significance. This thesis concentrates on Deep Learning (DL) based automated ECG division and outline techniques. Because of different shapes and irregularities in the ECG signal, the traditional straightforward component channels neglect to remove the assortment of cardiovascular wave arrangements. Convolutional Neural Networks (CNN) apply multi-facet including channels on the input to remove complex features from the sign. In this way, CNN can be used to remove various levelled highlights from ECG signals. Two Convnet structures are read up and utilized for the restriction of cardiovascular waves. Their exhibitions are contrasted with one another and different investigations in the writing too. The outcome shows that CNN is capable of extracting heart wave spatial elements with the great execution. Moreover, various long haul notwithstanding transient worldly patterns exists in ECG signals because of arrhythmia and other heart conditions. Generally speaking, momentary memory will be unable to catch the transient highlights in ECG signals, thus a Long Short-Term Memory (LSTM) network is planned and utilized to catch long haul and momentary information dependencies in ECG sequences. This technique has worked on distinguishing proof of worldly highlights, especially in P-wave. To address the crude ECG signal division issue, a hybrid DL compressing of convolutional auto encoders and LSTM networks is proposed and contemplated. Results show that it performs commonly well for crude ECG signal division.

Introduction:

Electrocardiography (ECG) is a non-invasive diagnostic tool essential for evaluating heart health and diagnosing a range of cardiovascular conditions. Despite its widespread use, accurate interpretation and analysis of ECG signals remain challenging due to the complex nature of cardiovascular diseases and the subtle variances in ECG waveforms[1]. This paper explores the application of deep learning techniques, including convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to enhance the accuracy and efficiency of ECG signal analysis[2].

The significance of this research is underscored by the global burden of cardiovascular diseases, which are leading causes of morbidity and mortality[3]. By advancing ECG interpretation through automated deep learning models[4], this study aims to contribute to the early detection, diagnosis, and management of cardiac conditions. Furthermore, the paper introduces novel approaches to ECG segmentation and arrhythmia classification, leveraging the strengths of deep learning to handle the variability and complexity inherent in ECG data[5].

ECG research is essential for the early detection,

diagnosis, and management of cardiovascular disorders. It has made a substantial contribution to the growth of cardiology and healthcare, and it is certain that additional advancements in the field will result from it son going study and development[3].

Research on electrocardiograms (ECGs) aims to better understand and diagnose cardiac diseases by examining the electrical activity of the heart as captured by an ECG (ECG). Studying the typical heartbeat patterns as well as looking for anomalies that can point to different cardiac disorders, like arrhythmias, heart attacks, and heart failure, are possible ECG research topics.

ECG research maybe conducted in the following areas:

Development of new ECG technologies: Researchers may concentrate on creating ew ECG technologies, such as wearable ECG devices or cutting-edge imaging methods, that can offer more precise and detailed information regarding cardiac activity[6].

Finding new biomarkers: One use of ECG research is to find novel biomarkers[7], or quantifiable signs of a condition, that can be found through changes in the ECG signal. This might aid in the earlier detection and better management of heart problems.

The goal of ECG study may also be to better understand how the heart works under various settings, such as stress or during physical activity [8], as well as in individuals with various health issues. Constrained by electrical flows through the heart muscle, the heart contracts musically and siphons blood all through the body. An electrocardiogram (ECG) is the recording of the heart's electrical activity to get critical in line with regards to the heart's condition. The electrical activity of the heart shifts through time. Overall, ECG is a graphical representation of heart electrical movement in milli volts (mV) [9] on the upward hub against time on the even pivot as displayed in Figure A-ECG Signal Recording.

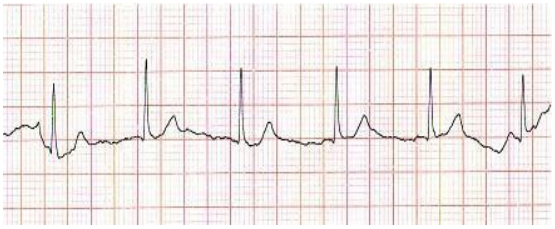

Figure 1. ECG Signal Recording

A typical pattern of heart exercises is alluded to as a cardiovascular cycle or a heart complex, known as a heartbeat. A typical cardiovascular complex is com-presented of a few heart waves. Among those waves, three of them are of high importance, to be specific, P-wave, QRS-complex, and T-wave. P-wave the a trial depolarization[10]. A normal P-wave term is not exactly equivalent to 0.11 sec-gestures and the abundance is under 0.25 mV. QRS-complex ventricular depolarization. An ordinary QRS-complex span is not exactly or equivalent to 0.12 seconds and the plenty fullness is over 0.25 mV. Further, the T-wave is the arrival of ventricular muscle to the resting state (ventricular repolarization). An ordinary T-wave term is from 0.10 to 0.25 seconds and the adequacy is at under 0.25mV[11].

Different spans and fragments in the ECG signal are subordinate to these key waves including PR Interval (i.e., from the start of the P-wave to the start of the QRS-complex), PR Segment (i.e., from the finish of the P-wave to the start of the QRS-complex section), ST-Segment (i.e., from the finish of S-wave to the start of the T-wave fragment), and QT-Interval (the finish of QRS-complex to the furthest limit of the T-wave span). Every stretch and fragment conveys information for heart analysis. In any case, finding these heart waves (P-wave, QRS-complex, and T-wave) are fundamental in tracking down the other stretches and sections. Figure 2 shows an average lead II ECG heart complex with cardiovascular stretches, sections, and pinnacle explanations.

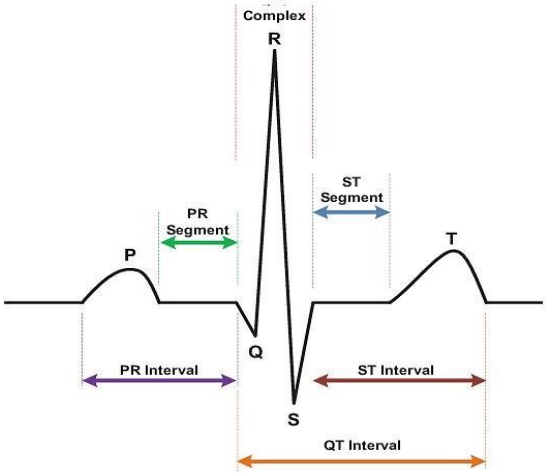

Figure 2. Typical cardiac complex with annotated cardiac waves, segments, intervals

In the realm of cardiac diagnostics, accurately identifying the QRS complex in ECG signals is crucial. The summary of the QRS detection methodologies reveals a blend of traditional algorithms and modern deep learning techniques. The Pan-Tompkins algorithm stands out for its robustness against noise, employing a series of filters for precise detection[12]. Wavelet Transform methods have also been favored for their efficacy in isolating the QRS complex amidst signal interference[13]. The advent of deep learning has introduced Convolutional Neural Networks (CNNs)[14], which, trained on extensive datasets, offer high accuracy in QRS identification[15]. Additionally, adaptive thresholding has emerged as a real-time

detection method, adaptable to the signal's amplitude variations. Collectively, these methods underscore a comprehensive study into QRS detection, underscoring the tailored selection of algorithms based on the clinical application and signal characteristics.

CNNs have emerged as a powerful tool[16] for arrhythmia classification due to their ability to autonomously extract significant features from ECG data without manual intervention. The conventional approach involves preprocessing the ECG signals into fixed-length windows and passing them through the CNN, which usually comprises several convolutional layers followed by pooling layers to reduce spatial dimensions[17]. The processed signals are then classified using fully connected layers. One of the strengths of CNNs lies in their ability to manage variable-length input signals by segmenting them into standardized windows, creating a 2D input structure for the network. This capability, along with the network's proficiency in learning local features, is essential for detecting subtle ECG signal variations indicative of various arrhythmias.

Contrasted with traditional machine learning algorithms[18] that depend on manually crafted feature extraction and selection—thus prone to over fitting—CNNs integrate these processes within the model architecture, enhancing robustness and the potential for high generalization from diverse data sources. CNNs have been successfully applied across a spectrum of scientific fields, demonstrating their versatility in signal analysis and medical image processing. The hybrid models, combining CNNs with Long Short-Term Memory (LSTM) networks[19], leverage spatial and temporal data for predictive analyses, offering a promising direction for precise and efficient clinical decision support systems in cardiac health[19, 20].

In recent years, substantial progress has been made in automated ECG analysis, particularly in China, where deep learning models have achieved high accuracy in arrhythmia classification and QRS complex detection. However, these models often require fixed-length input signals and can suffer from long training times. By proposing a hybrid deep learning model combining CNN auto encoders with LSTM networks, this study addresses these challenges, offering a promising direction for real-time, robust ECG analysis.

This paper presents a comprehensive examination of the state-of-the-art in ECG signal processing, outlines the innovations brought forth by the current research, and discusses the potential impact of these advancements on healthcare[21]. We aim to demonstrate the efficacy of our proposed models through rigorous testing and validation against existing benchmarks.

Related Work:

The evolution of Artificial Intelligence (AI) frameworks has been propelled by the imperative to assimilate vast, complex datasets, transitioning from static, hard-coded knowledge bases to dynamic learning systems. Machine Learning (ML) methodologies stand at the forefront of this transformation, with performance heavily dependent on the presentation of data. Feature representation, a pivotal aspect of ML, bifurcates into ML-derived features and hand-crafted features, each serving distinct purposes within the AI spectrum[22].

In the domain of Deep Learning (DL), a stratified, graph-based approach has emerged, capable of extrapolating abstract features through multiple layers of analysis. This hierarchical nature allows DL to harness simple, low-level data to uncover more intricate patterns. Within this framework, DL surpasses traditional ML methods by directly learning from raw data, eliminating the need for explicit feature designation. This paradigm shift has profound implications for fields that require nuanced data interpretation[23], such as Electrocardiogram (ECG) analysis, where DL approaches remain underexplored.

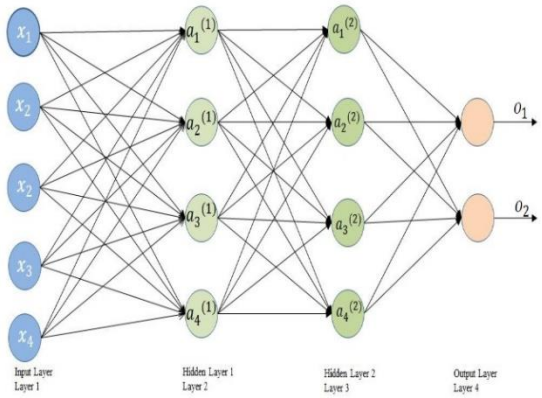

Figure 3. Multilayer perceptron model

The fundamentals of DL are grounded in statistical learning, which starkly contrasts traditional ML by employing an optimization process known as training. This process involves learning directly from the raw data, exemplified in tasks like handwriting recognition and image classification, where local features like pixel neighborhoods or RGB values serve as the data points for learning. The distinction between traditional ML and DL becomes particularly salient when considering their respective feature extraction capabilities and their applications in supervised learning, a key component in the analytical evaluation of ECG data.

Neural networks, particularly Multilayer Perceptrons (MLPs), form the architectural backbone of many DL models. These networks comprise multiple layers of interconnected nodes, each layer building upon the previous to recognize patterns and make predictions. The role of activation functions is pivotal in these

networks[24], dictating the neuron's output and enabling the network to handle nonlinear problems effectively. The sophistication of these networks is such that they can approximate any continuous function to an arbitrary degree of accuracy, provided they are equipped with sufficient layers and neuron counts.

In the context of ECG analysis, AI has attempted to address the intricate task of identifying specific patterns within ECG signals, a process traditionally reliant on deep clinical knowledge. Previous studies have used both rule-based and classic ML methods, known as traditional ECG signal processing approaches. However, there is a recognized deficiency in studies leveraging DL methodologies for ECG analysis[25], a gap that presents an opportunity for significant advancements in the field.

The review further delves into the mechanisms of the output layer and loss functions, particularly in multiclass classification problems relevant to ECG signal categorization. It also differentiates between the various learning strategies such as batch learning, mini-batch learning, and stochastic learning, which influence the updating of network weights during training.

Convolutional Neural Networks (CNNs), since their pioneering introduction by LeCun et al., have seen an astronomical rise in popularity, especially after the landmark success of Krizhevsky et al. in classifying the Image Net dataset in 2012[26]. This review focuses on the structural and functional aspects of CNNs and their transformative impact on image and video analysis, object classification, and localization tasks.

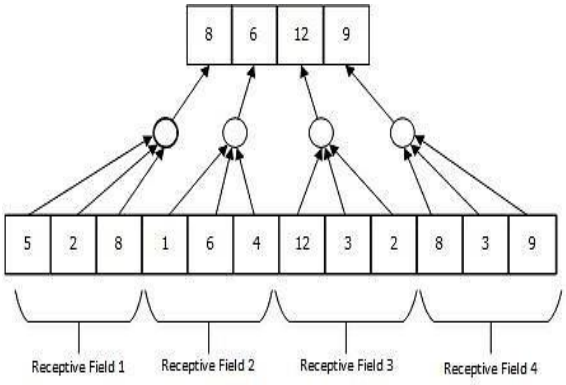

Figure 4. Convent model

CNNs distinguish themselves from traditional neural networks through their unique architecture designed to automatically and adaptively learn spatial hierarchies of features from input images[27]. Unlike Multilayer Perceptrons (MLPs) that consist of fully connected layers, CNNs utilize a more efficient structure with convolutional layers composed of numerous filters, each designed to detect local conjunctions of features from their preceding layers. These filters, or kernels, operate on overlapping local patches of the input image, producing feature maps that represent the presence of specific features at different spatial locations of the input.

A hallmark of CNNs is the utilization of max-pooling layers, which follow the convolutional layers. Max-pooling is a sample-based discretization process[28]. It aims to down sample the feature maps, reducing their dimensionality and allowing for assumptions to be made about the features contained in the sub-regions binned. This serves not only to reduce computational load and memory usage but also to provide basic translation invariance to the internal representation. Max-pooling achieves this by sliding a window across the feature map and recording the maximum value at each step, effectively zooming out and allowing the network to capture larger contextual information from the input.

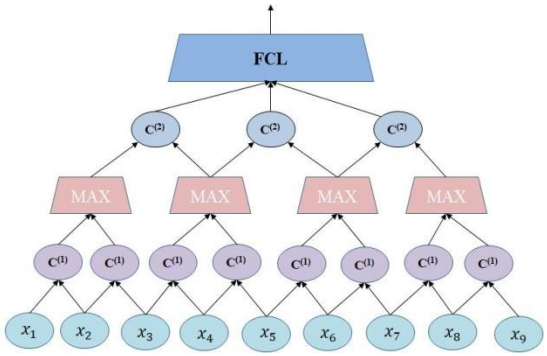

Figure 5. Max-pooling, step, and open-field models

Figure 5 represents an illustration of the max-pooling layer, step and responsive fields. In this model, the maximum pooling capacity is shown with void vehicle pieces of information. Further, step and responsive fields on an information X1×12 are clarified. In this model, step S = (1 × 3), and the size of the responsive field are additionally (1 × 3). Along these lines, the focuses of the open fields are (2, 5, 9, 11) where each middle has 2 even example distances from the following community. As it is shown, the maximum pooling activity yields the most extreme worth over each adjoining responsive field of size (1 × 3).

The stride concept is integral to the functioning of convolutional and pooling layers, defining the amount of overlap between the receptive fields as the filter moves across the image. A stride of more than one reduces the size of the subsequent feature maps, contributing to the network's ability to generalize and reducing the risk of over fitting by decreasing the resolution of the pooling operations.

Feature maps generated through convolution and pooling undergo a series of transformations, with each layer extracting more complex and abstract features.

The depth of a CNN is a key factor in its ability to perform hierarchical feature extraction[29]. This depth is achieved through the stacking of convolutional and pooling layers, where higher layers of the network can recognize complex features as combinations of the lower-level features, enhancing the network's predictive capabilities for complex tasks.

The concluding layers of a CNN are typically fully connected, serving to integrate the learned high-level features into a final output, such as a class score. The fully connected layers operate on flattened feature maps, converting the 2D features into a 1D vector that can be used for classification or regression tasks.

ECG Analysis Research Review: There are two specialized examination fields in the ECG outline. While one spotlight on the extraction of elements that can best address the construction of cardiovascular waves the other is to order the acquired elements into a particular arrangement of classes, i.e., heart wave classes or heart-condition manifestation classes A broad writing audit on ECG signal delinks-happiness can be found in As of late, specialists have applied DL strategies for ECG handling also In this segment, a foundation survey on ECG includes extraction techniques, ECG highlight characterization strategies and DL-based techniques for ECG examination will be given[30].

In ECG signal analysis, features are extracted using three main methods: filtering techniques (like smoothing and derivative filters), transformation methods (such as Discrete Wavelet Transform and Fourier Transform), and amplitude-based methods that focus on signal amplitude for classification. Filtering is used to detect key components of the ECG but may not always be reliable for abnormal signals. Transformation methods identify frequency patterns in heart waves, while amplitude-based approaches are sensitive to variations in waveforms and may require adjustments for accurate classification.

For classifying ECG heart waves, a variety of rule-based and machine learning (ML) techniques have been employed, including neural networks, Random Forest, Support Vector Machines, Naïve Bayes, and Hidden Markov Models (HMM). These methods typically analyze the peak points and other local properties of heart waves or employ transformations like wavelet analysis for feature extraction. Advances have also been made in segmenting ECG signals and diagnosing conditions such as hypertrophic cardiomyopathy (HCM) using these ML techniques[30]. However, there is ongoing research to enhance ECG delineation methods, as current approaches like HMM may fall short, particularly with raw ECG signals. Hidden Semi-Markov Models (HSMM) have shown improvement over HMM, indicating that more sophisticated temporal pattern recognition, potentially through deep learning methods like LSTM, could further refine the classification and interpretation of ECG data.

Methodology:

Data Acquisition and Sources Our research leverages data from the QT Database (QTDB), which is a well-known resource for the study of ECG signal processing. The QTDB includes ECG recordings from a diverse patient cohort[31], representing a broad spectrum of cardiac conditions. Each recording in the database is 15 minutes long, with a sampling frequency of 250Hz, providing high-resolution data for accurate waveform analysis.

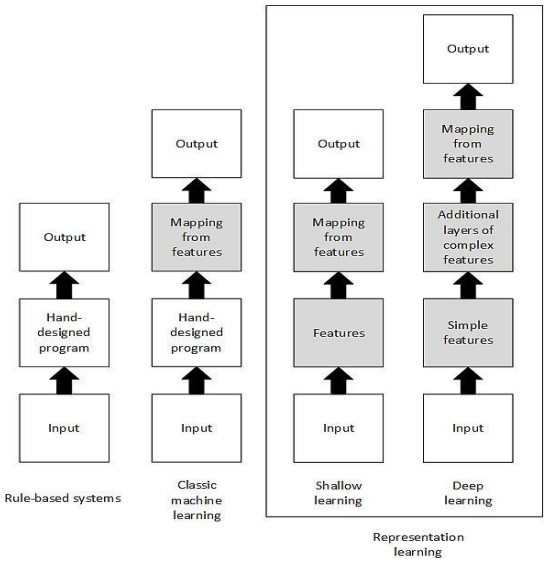

Figure 6. General interaction for the proposed heart wave

The database annotations, which include the demarcations of P, QRS, and T waves, serve as a reference for validating our cardiac wave localization algorithm.

Data Preprocessing:

Baseline Wander Removal: Baseline wander noise[32], which can obscure ECG signal features, is removed using median filtering. The filter's window size is set at half the sampling frequency, which, for a 250Hz sampling rate, would be 125 samples. This filter subtracts the median-filtered signal from the original, leaving behind the high-frequency components vital for analyzing cardiac activity.

$$E(WF) = E(R) - M\left(M\left(E(R)\right)\right) \quad (1)$$

Where E(WF)is sans meander ECG, Elis the crude ECG from QTDB, and M(.) operation joy is the consequence of applying a middle channel to a sign. The size of the middle channel rises to half of the examining recurrence, f req. Since ECGs from QTDB

Normalization: After baseline wander is removed, the ECG signals undergo normalization. This process adjusts the signals to a standard scale by manipulating their mean and standard deviation[33]. Such normalization is critical as it facilitates faster and more stable convergence of neural network models, which will later be used for feature extraction and classification.

$$E(norm) = \frac{E(WF) + Avg(REF)}{STD(REF)} \qquad (2)$$

Where Avg (.) is the normal capacity, Std(.) is the standard deviation capacity and E(norm) is the standardized ECG values. It's noted that while normalization does not alter the location of cardiac waves, it ensures the input values are standardized, preventing the model's gradients from becoming disproportionally large and hindering the convergence on the global minima of the loss function.

Cardiac Complex Segmentation: The ECG signals are segmented into cardiac complexes by identifying the slope variations indicative of electrical impulse activity. This method is distinct from other approaches as it does not target individual heart waves but rather the entire cardiac complex.

Following the techniques presented in subsidiary-based strategy has been created to remove cardiovascular buildings. As opposed to other subordinate-based techniques, this exploration recognizes total cardiovascular edifices rather than heart waves, i.e., while other examination attempts to find attributes of incline varieties in P-wave, QRS-complex, and T- wave to identify the cardiovascular waves, this exploration uses slant varieties to identify the areas with electrical motivation movement. Along these lines, this technique tracks down cardiovascular edifices by dissecting the high and low incline varieties of the ECG signal as displayed in Figure 7 and 8

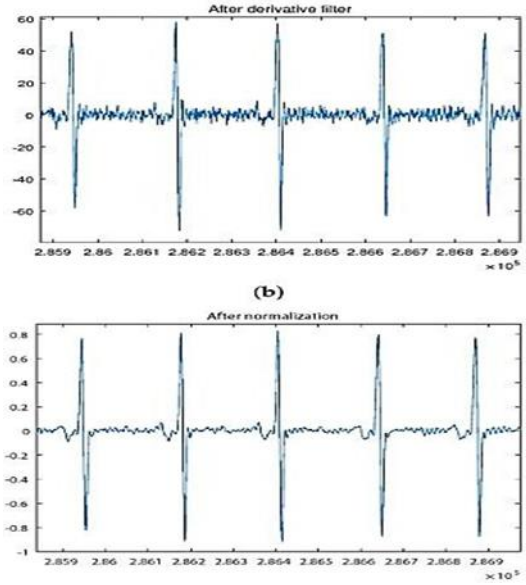

Figure 7 Normalized ECG segment

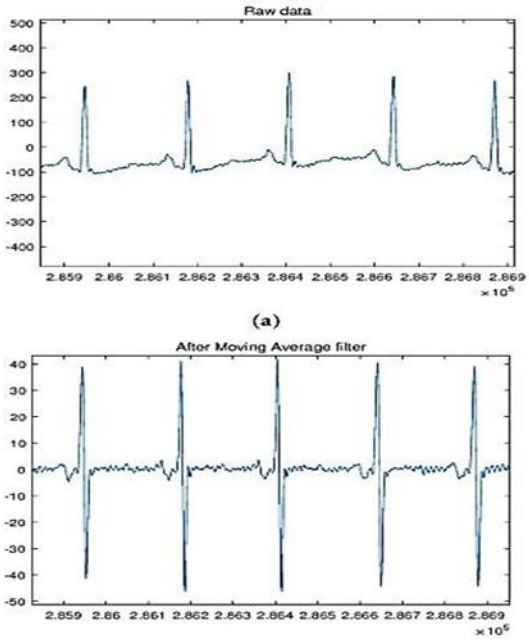

Figure 8. Separated heart buildings cardiovascular wave area

The X-axis shows accuracy and Y Axis shows the activity of ECG. In cardio vascular complex extraction, areas with a high variety in the slant of the sign (shift during the sign) suggest a heart-muscle movement, for example, a QRS complex. The most reduced incline variety movement between two high-slant variety action areas addresses the finish of one cardiovascular complex and the start of another. ECG incline variety exercises can be investigated by using a portrayal of the neighborhood of the standardized ECG second-request subsidiary bend.

The normalized ECG is smoothed and then passed

through a series of derivative operations to segment cardiac complexes

The following are performed on the normalized ECG signal.

$$E(norm) = \frac{E(WF) + Avg(REF)}{STD(REF)} \quad (3)$$

$$E_{(SM)} = E_{(norm)} \cdot \otimes \zeta_1 \quad (4)$$

Where $\otimes$ and |.|are the convolution activity and the outright worth showntonindividually.Z1= [1,1, 1] is a Smoothing bit along these lines, and E(SM) is the smoothed outright standardized ECG. The outright amplitudes of ECG signals are utilized since they can catch crossing the pattern in the ECG signal. The principal request subordinate to the E(SM) is communicated.

$$E_{(FD)} = E_{(SM)} \cdot \otimes \zeta_2 \quad (5)$$

Where ζ2 = [−1, −1, −1, 0, 1, 1, 1] is a derivative kernel. Followed by another derivative kernel, the second-order derivative of the absolute ECG signal can be expressed as:

$$E_{(SD)} = E_{(FD)} \cdot \otimes \zeta_2 \quad (6)$$

Thus, the local area under the magnified second-order derivative curve is obtained by where Ij is the jth test of E(SD), E(γ) is the neighborhood the amplified (force of six) second-request subsidiary of E curve (the second request determine the time is amplified to highlight the higher outright amplitudes,

The local maxima in the processed signal $E(\gamma)$ indicate the areas of high slope variation, corresponding to the electrical activity of cardiac complexes. A threshold based on the average of these maxima $Avg(E\Phi)$ and a minimum separation equivalent to the QT interval duration are set to differentiate between individual complexes. The end points of the complexes are determined by identifying the lowest values between successive local maxima in $E(\gamma)$. This comprehensive preprocessing ensures the ECG signal is accurately segmented into distinct cardiac complexes, ready for feature extraction and classification**.**

Ethical Considerations:

The study emphasizes the use of data from the QT Database (QTDB), which is publicly available and includes ECG signals that have been collected with proper consent and anonymization protocols[33]. This database is a well-known resource in the biomedical research community, ensuring that the data used respects the privacy and consent of the individuals from whom it was collected. By using QTDB, the study inherently adheres to ethical standards for biomedical research, avoiding any issues related to consent and privacy that would arise from collecting new data. The QTDB provides a rich dataset that negates the need for new data collection, facilitating research within ethical bounds

Detailed Methodological Steps:

Creation of Data Sets for Neural Network Training:

Cardiovascular complexes are extracted from two leads of fifteen-minute ECG recordings, except in cases where data was not suitable due to clarity issues or abnormal spikes. QTDB contains multiple datasets with varying annotations[34]; sometimes these include both the start and end of cardiovascular waves instead of a single point per wave. To standardize the data, the midpoint of the wave or the point of maximum amplitude is used if the duration is not specified. Extracted cardiac complexes vary in length, so they are padded to a fixed length of 300 samples (1.2 seconds) to create uniform input vectors for the neural network. Padding is done with repetitions of the signal's last value to avoid bias towards particular regions of the cardiac complexes, which could lead to overfitting. To further prevent overfitting, padding is randomly applied at the beginning, end, or both sides of the cardiac complexes.

After preprocessing, the data is ready to be fed into neural networks. The extracted cardiac complexes are divided into training, validation, and test sets, with no overlap between the sets to ensure exclusivity. The targets for the neural network output are the peaks of three regions[35]: P-wave, QRS-complex, and T-wave.

Table 1. Dataset Distribution

Set	Number of Cardiac Complexes	Percentage of the Total
Training Set	100,380	60%
Validation Set	16,730	10%
Test Set	50,189	30%
Total	167,301	100%

Neural Network Training:

1. **Multilayer Perceptron Model**:

The first model is an MLP with a baseline architecture consisting of an input layer with 300 neurons for ECG segments[36], a dense hidden layer with 150 neurons, and an output dense layer with 3 neurons. The initialization of weights for the MLP is done using a Gaussian distribution with zero mean and a standard deviation of 0.1. The ReLU (Rectified Linear Unit) function is used as the activation function for neurons, represented as:

$$A = Max(0, z) \quad (7)$$

Table 2. Baseline MLP model description

Layer	Description	Size
0	Input layer	300
1	Dense layer	150
2	Output dense layer	3

Where z is the contribution to the REL and an is the result. REL work yields zero when the information is under nothing and it yields the worth of the info when the info isn't negative, in this way, a ∈[0, +∞]. The Root-Mean-Square Error (RMSE) is used as the loss function, and Adam's optimization algorithm is adopted for training due to its benefits in adapting learning rates.

2. Convolutional Neural Networks (ConvNets): Two ConvNet models are developed with alternating convolutional and max-pooling layers, followed by fully connected hidden layers. The first ConvNet model includes a dropout layer to prevent over fitting by randomly omitting some weights from updates during training[37]. The hyper parameters for the ConvNets are chosen based on the baseline MLP model performance, with adjustments in learning rates and the size of fully connected layers.

The contribution to the organization is a 1-D heart complex sign with 300 examples. All the convolutional layers have (1×5) highlight channels with the step of (1×3) and all the maximum pooling layers have non-covering (1×2) open fields with the step of (1×2). The first and second convolutional layers incorporate 16 and 32-component channels individually. This is trailed by two completely associated layers. The primary completely associated layer has 150 neurons. The second completely associated layer, the result layer of size three, shows three cardiovascular wave positions inside a heart complex

A similar activity is performed for the second convolutional layer and the subsequent max-pooling layer. Because of applying convolutional channels and max pooling, guides of (32 × 9) are produced. Then, at that point, a framework reshape is utilized to this two-layered vector and converted to one aspect. Thusly, the repetitive aspects are disposed of, turning the vector into a288-highlightpointvector.

The288-include point vector is trailed by a completely associated layer with 150 neurons. The following layer is the result layer, which is the last completely associated layer.

3. ConvNet with Dropout Layers:
The second ConvNet model is similar to the first but includes dropout layers after each max-pooling layer with a dropout rate of 0.5 to further combat over fitting.

Table 3. The proposed Convent description

Layer	Description	Size	Receptive Field Size	Stride
0	Input	(1×300)	-	-
1	First convolutional layer	(16×100)	(1×5)	(1×3)
2	First max-pooling layer	(16×50)	(1×2)	(1×2)
3	Second convolutional layer	(32×18)	(1×5)	(1×3)
4	Second max-pooling layer	(32×9)	(1×2)	(1×2)
None	Reshape	(288)	-	-
5	First dense layer	(150)	-	-
6	Output dense layer	(3)	-	-

Table 4. The proposed Convent with dropout layers description

Layer	Description	Size	Receptive Field Size	Stride
0	Input	(1×300)	-	-
1	First convolutional layer	(16×100)	(1×5)	(1×3)
2	First max-pooling layer	(16×50)	(1×2)	(1×2)
3	First dropout layer	(16×50)	-	-
4	Second convolutional layer	(32×18)	(1×5)	(1×3)
5	Second max-pooling layer	(32×9)	(1×2)	(1×2)
6	Second dropout layer	(32×9)	-	-
None	Reshape	(288)	-	-
7	First dense layer	(150)	-	-
8	Output dense layer	(3)	-	-

Evaluation Metrics:

The performance of the neural network models is evaluated using metrics such as Root Mean Square Error (RMSE) and accuracy. These metrics help in quantifying the precision of the cardiac wave localization relative to the annotations provided in the QTDB.

In assessing classification algorithms, especially in the context of machine learning for ECG signal analysis, the performance is quantified using four primary metrics that stem from the outcomes of the prediction: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

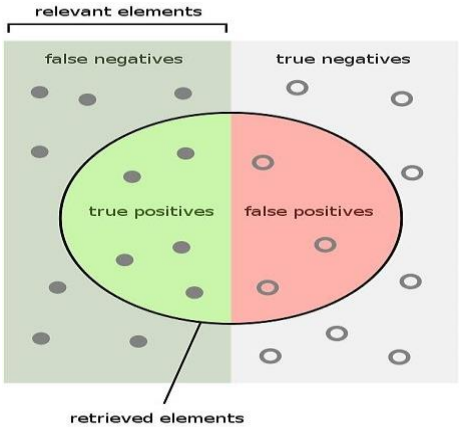

Figure 9. Categorization states

- **True Positives (TP)**: Instances where the algorithm correctly identifies a positive class, such as correctly classifying a P-wave segment in an ECG as a P-wave.

- **True Negatives (TN)**: Instances where the algorithm correctly identifies the absence of a positive class, such as correctly recognizing that a segment is not a P-wave.

- **False Positives (FP)**: Instances where the algorithm incorrectly predicts a positive class, such as wrongly classifying a segment as a P-wave when it is not.

- **False Negatives (FN)**: Instances where the algorithm fails to identify a positive class, such as missing a P-wave.

Thinking about TP, TN, FP, and FN, estimating the presentation of ML Algo-rhythms depends on four measurements that consolidate the classification states. These four measurements are exactness, accuracy, review, andf1-score. These measurements are utilized to gauge the exhibition of the division calculation too. The following will characterize the measurements and talk about their significance of them[37]. Precision: Accuracy is the most utilized measurement to gauge the performance of ML and arrangement techniques. This measurement measures the proportion of right perceptions over complete perceptions. The precision recipe is communicated as:

Table 5. Categorization output states

		Predicted Class	
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

- **Accuracy**: This metric assesses the overall correctness of the algorithm by dividing the sum of TP and TN by the total number of predictions (TP + TN + FP + FN). High accuracy indicates that the algorithm is correctly classifying the majority of the instances. However, accuracy alone can be misleading, especially

if the dataset is imbalanced (i.e., the number of instances in each class varies significantly).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

- **Precision**: Precision measures the ratio of TP to the total number of instances predicted as positive (TP + FP). It reflects how reliable the algorithm's positive predictions are. High precision implies that when the algorithm predicts an instance as positive, it is likely to be correct.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

- **Recall (Sensitivity)**: Recall calculates the ratio of TP to all instances that are actually positive (TP + FN). It evaluates the algorithm's ability to detect all relevant instances. High recall indicates that the algorithm successfully identifies most of the positive instances.

$$\text{Re}call = \frac{TP}{FP + FN} \quad (10)$$

- **F1-score**: The F1-score is the harmonic mean of precision and recall. It is particularly useful because it balances the contribution of precision and recall, providing a single metric for situations where the trade-off between FP and FN is important.

These metrics collectively give a robust evaluation of a classification algorithm's performance, taking into account not only the accuracy but also the balance between the types of correct and incorrect predictions. This is essential in medical applications like ECG analysis, where the cost of FP and FN can have significant implications for patient care.

By following this detailed methodology, we aim to develop an accurate and reliable system for cardiac wave localization in ECG signals, which is an important aspect of automated cardiac diagnostics.

Results and Discussion:

Supervised Cardiac Wave Segmentation Using LSTM RNN

The objective of supervised segmentation in our study is to model the intricacies of cardiac waveforms, particularly leveraging the LSTM RNN's ability to assimilate and propagate information across time series data[38]. Cardiologists' expertise in identifying the sequential nature of heart waves – where one wave follows another in a predictable pattern – inspired our approach. Using an LSTM RNN model, we have successfully captured both long-term and short-term dependencies present in ECG signals.

The LSTM RNN model showed improved efficacy in identifying the temporal features characteristic of P-waves. This advancement is credited to the model's capability to classify each ECG signal sample into one of four categories: neutral, P-wave, QRS-complex, and T-wave. Through this classification, the segmentation of ECG into its constitutive waves is achieved, which is instrumental for subsequent diagnostic evaluations.

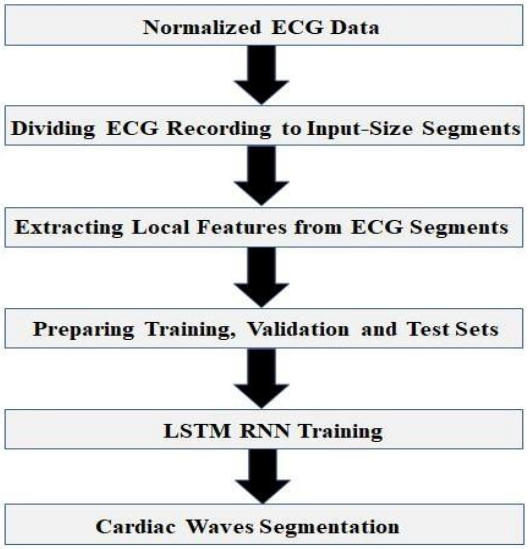

Figure 10. Overall process for cardiac wave segmentation using an LSTM RNN model

Data Preparation and Feature Extraction

Prior to feature extraction, we subjected the ECG recordings to preprocessing steps that included baseline wander removal and normalization. Distinct from the previous research where entire ECG recordings were processed, in this study, we partitioned the ECG data into 500-sample segments. Each segment, corresponding to 2 seconds of ECG data[39], typically contained more than one complete cardiac complex. This segmentation strategy allowed the LSTM model to learn from both complete and partial cardiac complexes, thereby enhancing its predictive performance.

From each ECG segment, local features were extracted using various filters, including smoothing and derivative kernels[40]. Consequently, the LSTM RNN model received a four-feature input for each time step, encapsulating the original signal and its derived attributes.

Feature Sets

The extracted ECG segments were classified into three sets: training, validation, and test sets. To avoid over fitting and ensure model generalizability, these sets were kept mutually exclusive and patient-independent. Below table records the preparation, approval, and test information.

Table 6. Categorization output states

Dataset	Number of segments	Percentage
Training set	51,419	55%
Validation set	9,350	10%
Test set	32,721	35%
Total	93,490	100%

Bidirectional Long Short-Term Memory Recurrent Neural Network Review

In traditional RNNs, the information flow is unidirectional, thereby utilizing only prior data. However, ECG signals contain valuable future context which can be harnessed using Bidirectional RNNs. Schuster et al. presented a Bidirectional Recurrent Neural Network (BRNN)[4], which utilizes the two headings of the information, earlier and future examples, in two separate secret layers, and the initiations are passed to the layer over (one more secret layer or a result layer). Utilizing the contribution at time t, the forward secret layer actuation vector, and the regressive secret layer AC Ti- applause vector, at, are portrayed as[3]. Like BRNN, the regressive secret layer can catch future data and the forward secret layer can catch the previous data.

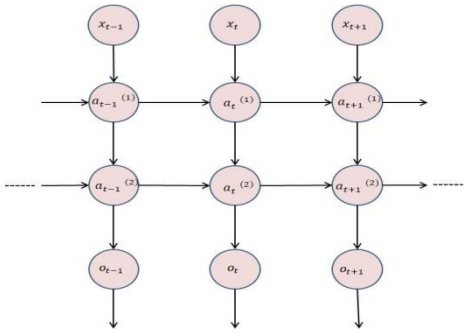

Figure 11. Traditional deep RNN architecture

Utilizing the contribution at time t, the forward secret layer actuation vector, and the regressive secret layer AC Ti-applause vector, at, are portrayed as:

$$a_{(t)} = G\theta x, \rightarrow hxt \ + \ \theta \rightarrow h, \rightarrow h \rightarrow at - 1 \ + \ b \rightarrow h(11)$$

$$a_{(t)} = G\theta_{x\ ,\ \rightarrow\ h}{}^{Tx + \theta_{he}} at + 1 + b_{h(12)}$$

Where $\overrightarrow{h}$ refers to forwarding layer neurons and h refers to the backward layer neurons. Correspondingly, the following equation computes the output in a BRNN layer

$$t_o = \theta_{\rightarrow hike\ \rightarrow\ at + \theta_{hake}} at + bk(13)$$

Where k refers to output neurons.

Figure 12 shows an example of BRNN with only one hidden layer with its backward and forward hidden layers.

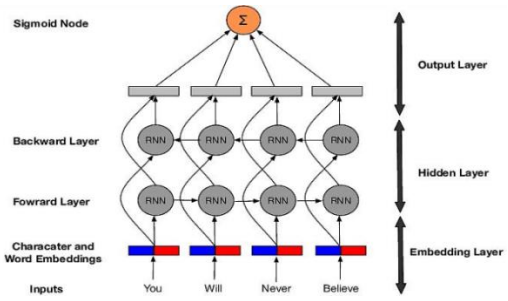

Figure 12. Bidirectional RNN LAYER

ECG-Senet Architecture Model

Our proposed ECG-Senet architecture builds upon the BLSTM framework, adding a fully connected time-distributed output layer with Soft Max function, which allows for the classification of ECG samples into respective wave categories. This novel architecture was found to be highly effective in capturing the temporal dynamics of ECG data.

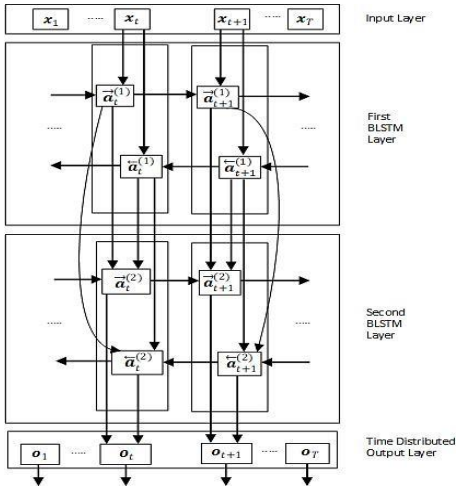

Figure 13. The proposed BLSTM RNN

This model contains two BLSTM layers and a completely associated time-dispersed result layer with Soft Max work as its actuation work. A completely associated time disseminated layer applies completely associated layer to each timestamp. The hyper parameters, for example, the quantity of stowed away layers and the quantity of LSTM cells are picked considering observing the combination of the BLSTM RNN model.

The primary layer is the info layer which requires some investment series 4×500 as clarified in the below Section. The following layer is a secret BLSTM layer. In this layer, each back-wards towed away layer and forward secret layer have250LSTMcells.

Table 7. Deep BLSTM RNN for ECG segmentation.

Layer	Description of the Layer	Size
Input	ECG raw signa land its features	(4×500)
Layer1	BLSTM with 500 LSTM cells	(2×250)
Layer2	BLSTM with 250 LSTM cells	(2×125)
Output	Time-distributed dense layer	(4×500)

This is trailed by one more BLSTM of size 250 and followed by the result layer which is a completely associated time-conveyed layer of size K to sort each example into one of the K classifications. Working out the misfortune for successive models is like non-consecutive models aside from that the misfortune in consecutive models aggregates over all the timestamps and afterward back propagates to the loads.

Training Experiment

The training of ECG-Senet employed the Adam optimizer, which is known for its efficient handling of sparse gradients. The model was trained through sixty-eight ages utilizing the small-scale group method of cluster size 250 which is picked considering assembly movement. Early stopping was implemented to prevent over fitting.

In the wake of preparing, the outcomes showed 94.6%exactness for the preparing set, 93.8% precision for the approval set, and 93.7% precision for the test set. Figure 14 shows Accuracy measure test. And Figure 15 shows the loss curve.

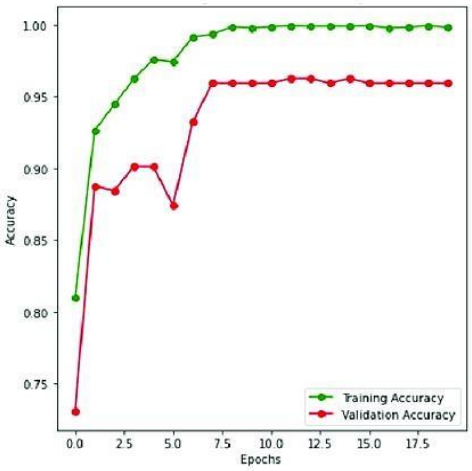

Figure 14. Accuracy test measure

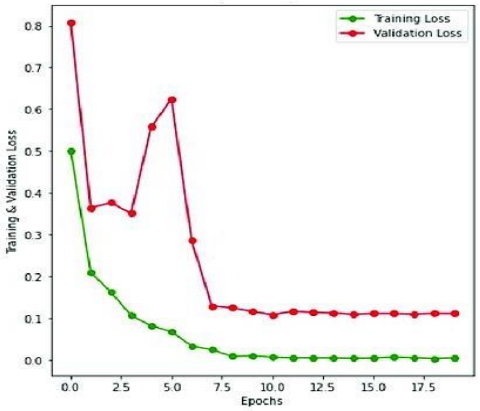

Figure 15. Loss Curve

Results

The ECG-Senet model achieved high accuracy rates across the training, validation, and test sets, demonstrating its robustness in ECG wave segmentation. Precision, recall, and F1-score metrics further validated the model's performance, with QRS-complexes being identified with the highest precision and recall.

The accuracy, review, and 1 score connected with each heart wave classification and neutral test are accounted for in Table 8. The most elevated accuracy in the heart waves has a place with the QRS-complex at 0.94 and the least has a place with T-wave at 0.90.The most elevated review rate has a place with the QRS-complex with a 0.95 review rate and the least review rate has a place with the P-wave at 0.90. Further, the most elevated 1score be-yearns to the QRS-complex at 0.94, and the least is tied between the P-wave and theT-wave classes with 0.91.

Table 8. ECG segmentation results

	Precision	Recall	1Score
Neutral	0.95	0.95	0.95
P-wave	0.92	0.90	0.91
QRS-complex	0.94	0.95	0.94
T-wave	0.90	0.92	0.91
Average	0.94	0.94	0.94

For relationship purposes, two distinct philosophies are used. Hughes et al. used HMM to handle ECG division with two approaches. The essential approach used unrefined ECG signals and the resulting strategy used wavelet-encoded ECG. Assessment of ECG-Senet and both HMM approaches are given in Table 9. It was shown that ECG-Senet performs better in terms of accuracy. Well, wavelet encoding gives better accuracy

in partitioning QRS-complex and T-wave. In any case, in any left over cases and overall results, ECG-Senet defeats HMM approaches.

Table 9. Segmentation accuracy comparison

Method	P-wave (%)	QRS-complex (%)	T-wave (%)	Overall (%)
ECG-Senet	92.0%	94.0%	90.0%	92.0 %
HMM on raw ECG	5.5%	79.0%	83.6%	56.03%
HMM on wavelet end-coded ECG	74.2%	94.4%	96.1%	88.23%

An enormous part of the other assessment is based on finding the cardio vascular complex fid-coal centers and not separating every single illustration of ECG unreservedly. Notwithstanding the way that the ECG-Senet task is not exactly equivalent to observing ECG heart waves identification, it gives vicious accuracy in perceiving cardio vascular waves. Shows the precision of observing waves paying little notice to division using ECG-Senet. More relationships of heart wave recognizing confirmation with various strategies will be discussed in the accompanying area.

Table 10. Cardiac wave identification results

Cardiac wave	**Accuracy**
P-wave	95%
QRS-complex	98%
T-wave	97%

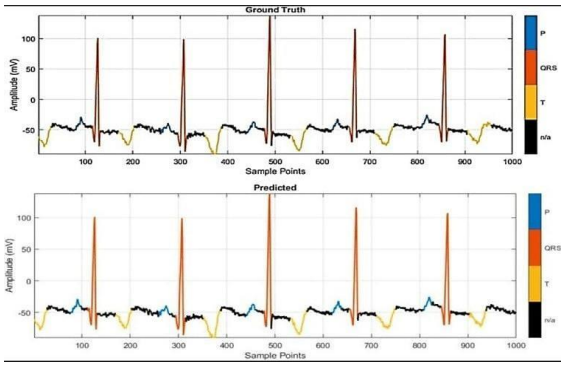

Figure 16. Sample Results

Figure 16 shows an example from the test set and its connected outcome, and it shows the precision of the ECG-Senet. P-wave, QRS-complex, and T-wave regions are represented by red, blue, and green districts.

Conclusion and Future Work:

This study presented a hybrid deep learning methodology for end-to-end ECG signal segmentation, addressing the critical need for automated cardiac diagnostics. The proposed approach combined Convolutional Neural Networks (CNNs) for detailed spatial feature extraction with Bidirectional Long Short-Term Memory (BLSTM) neural networks to effectively capture the temporal dynamics of ECG signals.

The convolutional auto encoder architecture developed was particularly adept at delineating the hierarchical structure of ECG signals. It demonstrated robust performance across datasets with varied cardiac conditions, including those with abnormal T-wave formations and diverse ST-segment presentations. The LSTM-based sequence learner outperformed conventional methods like Hidden Markov Models, showcasing the potential for deep learning to advance the field of cardiology significantly.

The success of this study opens several avenues for future research:

- Dataset Expansion: To enhance the generalizability of the model, future studies should incorporate a broader range of datasets, especially those containing cardiac conditions not represented in the QT Database (QTDB).

- Diagnostic Application: The ultimate utility of E2E ECG segmentation is in reaching diagnostic conclusions. Applying the proposed segmentation approach to diagnostic tasks, such as anomaly detection using criteria like the Seattle Criteria, could yield insights into the model's practical utility.

- Exploration of DL Architectures: With the rapid evolution of deep learning, new architectures and topologies should be investigated. Attention models and capsule networks may provide innovative ways to approach the ECG segmentation problem.

- Application to Clinical Systems: Applying the proposed model to clinical applications, such as wearable technology for continuous health monitoring, could significantly impact patient care.

- Energy Efficiency Assessment: Particularly for wearable devices, the energy efficiency of deep learning algorithms is paramount. Future research should address the energy consumption of DL methods to ensure the practicality of portable medical devices.

In summary, while this study has taken significant strides in ECG segmentation using deep learning, the

potential for further improvements is vast. The continued exploration of deep learning techniques in ECG analysis could lead to breakthroughs in automated, high-throughput cardiac diagnostic systems, ultimately aiding in large-scale disease prevention and management.

References

1. Deo, R., et al., Electrocardiographic measures and prediction of cardiovascular and noncardiovascular death in CKD. Journal of the American Society of Nephrology: JASN, 2016. **27**(2): p. 559.
2. Ansari, Y., et al., Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. Frontiers in Physiology, 2023. **14**.
3. Lin, T.-Y., et al. Feature pyramid networks for object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
4. Lin, T.-Y., et al. Focal loss for dense object detection. in Proceedings of the IEEE international conference on computer vision. 2017.
5. Liu, W., et al. Ssd: Single shot multibox detector. in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. 2016. Springer.
6. Tan, M., R. Pang, and Q.V. Le. Efficientdet: Scalable and efficient object detection. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
7. Mason, L., et al., Boosting algorithms as gradient descent. Advances in neural information processing systems, 1999. **12**.
8. Opelt, A., et al. Weak hypotheses and boosting for generic object detection and recognition. in Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part II 8. 2004. Springer.
9. Prosser, B.J., et al. Person re-identification by support vector ranking. in Bmvc. 2010.
10. Qi, Z., et al., Online multiple instance boosting for object detection. Neurocomputing, 2011. **74**(10): p. 1769-1775.
11. Chen, H., A. Gallagher, and B. Girod. Describing clothing by semantic attributes. in Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12. 2012. Springer.
12. Rakibe, R.S. and B.D. Patil, Background subtraction algorithm based human motion detection. International Journal of scientific and research publications, 2013. **3**(5): p. 2250-3153.
13. Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
14. Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. **28**.
15. Choi, S.-W., C.H. Lee, and I.K. Park. Scene classification via hypergraph-based semantic attributes subnetworks identification. in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13. 2014. Springer.
16. Sermanet, P., et al., Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.
17. Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
18. Mondal, S. and N. Mahfuz. Convolutional neural networks based bengali handwritten character recognition. in Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2. 2020. Springer.
19. Sun, Z., et al., Iris image classification based on hierarchical visual codebook. IEEE Transactions on pattern analysis and machine intelligence, 2013. **36**(6): p. 1120-1133.
20. Dalal, N. and B. Triggs. Histograms of oriented gradients for human detection. in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). 2005. Ieee.
21. Farenzena, M., et al. Person re-identification by symmetry-driven accumulation of local features. in 2010 IEEE computer society conference on computer vision and pattern recognition. 2010. IEEE.
22. Farhadi, A., et al. Describing objects by their attributes. in 2009 IEEE conference on computer vision and pattern recognition. 2009. IEEE.
23. Felzenszwalb, P., D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable

part model. in 2008 IEEE conference on computer vision and pattern recognition. 2008. Ieee.

24. Akata, Z., et al. Label-embedding for attribute-based classification. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.
25. Freund, Y. and R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 1997. **55**(1): p. 119-139.
26. Fu, Y., et al. Attribute learning for understanding unstructured social activity. in Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part IV 12. 2012. Springer.
27. Fukui, A., et al., Multimodal compact bilinear pooling for visual question answering and visual grounding. arXiv preprint arXiv:1606.01847, 2016.
28. Harzallah, H., et al. Classification aided two stage localization. in PASCAL Visual Object Classes Challenge Workshop, in conjunction with ECCV. 2008.
29. Gan, C., T. Yang, and B. Gong. Learning attributes equals multi-source domain generalization. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
30. Girshick, R. Fast r-cnn. in Proceedings of the IEEE international conference on computer vision. 2015.
31. Girshick, R., et al., Region-based convolutional networks for accurate object detection and segmentation. IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(1): p. 142-158.
32. He, K., et al., Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2015. **37**(9): p. 1904-1916.
33. He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
34. Hearst, M.A., et al., Support vector machines. IEEE Intelligent Systems and their applications, 1998. **13**(4): p. 18-28.
35. Kanwisher, N. and J. Driver, Objects, attributes, and visual attention: Which, what, and where. Current Directions in Psychological Science, 1992. **1**(1): p. 26-31.
36. Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
37. Lafferty, J., A. McCallum, and F.C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
38. Lampert, C.H., H. Nickisch, and S. Harmeling, Attribute-based classification for zero-shot visual object categorization. IEEE transactions on pattern analysis and machine intelligence, 2013. **36**(3): p. 453-465.
39. Layne, R., et al. Person re-identification by attributes. in Bmvc. 2012.
40. LeCun, Y., et al., Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, 1989. **2**.