ABUAD Journal of Engineering and Applied Sciences (AJEAS)

ISSN: 2992-5584

Volume 3, Issue 1, 123-134

Stratified Normalization Technique with Long Short-Term Memory-Based Autoencoder for Anomaly Detection in Heartbeat ECG Data

Olabisi Esther JOHNSON[1,3], Felix Ola ARANUWA[2], Ezekiel Olufunminiyi OYEKANMI[3]

[1]*Department of Computer Science, Federal Polytechnic, Ile-Oluji*
[2]*Department of Informatics, School of Computing, Adekunle Ajasin University, Akungba-Akoko*
[3]*Department of Computer Science, Achievers University, Owo*

estjohnson@fedpolel.edu.ng, felix.aranuwa@aaua.edu.ng, e.oyekanmi@achievers.edu.ng

Correspondence: estjohnson@fedpolel.edu.ng Tel.: +2348035424943

Date Submitted: 21/05/2025
Date Accepted: 08/07/2025
Date Published: 31/07/2025

***Abstract*:** *Disruptions in heartbeat patterns have been identified as a critical health concern globally, often leading to serious health risks and death. Due to the subtle and infrequent nature of these irregularities, individuals may overlook early warning signs, highlighting the need for continuous monitoring and early detection systems. Traditional methods for detecting heartbeat abnormalities, while contributing significantly in the past, are often labour-intensive and lack the precision required for timely intervention. Recent advancements, particularly in wearable electrocardiogram (ECG) devices and machine learning (ML), have changed the narrative in how heartbeat data is collected and analysed. Despite the progress, existing ML models often fall short in accounting for inter-patient variability, which is essential for reliable anomaly detection across diverse populations. In addressing the limitation, this paper proposes Long Short-Term Memory implementation of Autoencoder (LSTM-AE) enhanced with Stratified Normalisation (SN), called LAESN. The LAESN model is designed to improve sensitivity to individual patient differences in ECG signals. The LAESN was trained and evaluated on the ECG5000 dataset using the 'tanh' activation function, a batch size of 32, and 50 epochs. It achieved an F1 score and AUC of 0.9660 and 0.9952, respectively, surpassing both the baseline AE and other state-of-the-art models. These results highlight the effectiveness of SN in strengthening the ECG anomaly detection model, enabling it to capture subtle variations in heartbeat signals and support a patient-centric anomaly detection system.*

Keywords: Anomalies detection, arrhythmia, electrocardiogram, heartbeat, stratified normalization.

1. INTRODUCTION

Heartbeat analysis has been widely studied in healthcare analytics because of the important role it plays in good health and well-being. The human heart is a vital organ in the circulatory system that functions as a muscular pump responsible for delivering oxygenated blood and essential nutrients to tissues throughout the body [1]. An area of interest in understanding heart functionality is called arrhythmia, which often times experiences dangerous irregularities in its rhythm. These irregularities, whether in the form of bradycardia, premature contractions, or tachycardia, pose serious health risks [2, 3]. The disruptions in heart rhythm often led to cardiovascular disorders (CVD), which are a major cause of global mortality. In 2019, for instance, CVD was reported causing an approximately 18 million deaths globally [4] and 23.6 million deaths may arise by 2030, highlighting the critical need for early detection systems to help mitigate such fatalities. Over the period of time, traditional methods of analysis of heartbeat to detect abnormality are quite cumbersome and inaccurate. Moreover, advances in technology, particularly wearable Electrocardiogram (ECG) devices, have transformed how heartbeat data is collected and analysed. Unlike traditional methods, these wearable ECGs offer precise, real-time digital data collection, allowing for more accurate interpretation and enabling personalized treatments [5].

Over the past decade, many of the heartbeat ECG analysis are approached using Machine Learning (ML) techniques, of which classifying heartbeat into different classes is more prevalent. Researchers have employed numerous ML classifiers for heartbeat classification, exploring various ML techniques, including Decision Trees (DT), Random Forest (RF) and Gradient-Boosted Trees (GBDT) [6]. Deep Learning (DL) models like Convolutional Neural Networks (CNN) [4, 7], have also shown promising results in ECG prediction. Additionally, unsupervised approaches, including Autoencoders (AE) [8] and hybrid models, have advanced the field of anomaly detection [9]. However, Anomaly Detection (AD) has emerged as a critical area in ML, gaining considerable research attention due to its essential role in modern applications like financial risk management, fraud detection, network security, and healthcare analytics [10]. With the

rapid growth of wearable devices [11], AD has become significant where the detection of deviation from normal patterns is crucial for monitoring cardiac health [11].

Unlike classification, AD focuses on identifying abnormal heart rhythms relative to the normal baseline. AE [8], particularly when combined with LSTM networks [12, 13], have shown success in this task by addressing long-term dependencies, a challenge inherent in Recurrent Neural Networks (RNNs) [14]. Despite these advances, automatic detection of anomalies in heartbeat ECG data remains difficult because of inherent variability in ECG waveform morphology and individual patient differences [15]. This variability arises from multiple factors, including physiological differences, emotional states, and even momentary fluctuations in heart rate patterns patient [5]. Specifically, variations in key ECG intervals such as RR and QRS, as well as segments like PR and ST, often reflect unique physiological processes in each patient [5]. These variations make it difficult for traditional AD systems to distinguish between normal physiological deviations and actual outliers that signal abnormal heart conditions. Detecting anomalies in ECG data, therefore, requires ML model capable of capturing subtle, patient-specific irregularities without misclassifying them as noise. In addition, past AD studies are lacking in adequately accounting for the statistical variability of ECG features [12, 16].

This paper, therefore, proposes an AD of ECG heartbeat data, leveraging an LSTM-based AE architecture, enhanced with Stratified Normalization (SN). The proposed model, termed Long Short-Term Memory-Autoencoder embedding Stratified Normalization LAESN model was evaluated using ECG 5000. The results indicate that LAESN demonstrated strong performance, effectively reusing learned features and significantly improving AD accuracy. Addressing the existing challenges, the following are contributions of the paper:

i. A novel model for improving on the inter-patient variability of the heartbeat ECG.
ii. A robust framework for detecting cardiac anomalies by adapting to patient data through the use of SN, while improving model performance.

The rest of the paper is presented in the following sections below.

2. LITERATURE REVIEW

The human heart is an essential organ in the human, regulating the delivering oxygenated blood and essential nutrients to tissues throughout the body [1]. The heart (see appendix I (a)), is muscular consisting of four compartments that works in unison. The upper compartment, the left-and-right atria, which receive blood into the heart and channel it to the lower compartment, the left and right ventricles, via a series of valves. The right-side of the heart contains deoxygenated blood, while the left side circulates oxygenated blood received from the lungs.

The chambers cooperate to sustain the cardiac cycle performing a sequence of mechanical actions from one heartbeat to the next. The cycle consists of two primary phases: diastole, during which the heart relaxes to fill with blood, and systole, denoting the heart's contraction in pumping blood to the rest of the body. During these phases, atrial and ventricular volume and pressure exhibit rhythmic fluctuations (see appendix I (b)).

ECG recordings are obtained in various formats. Meanwhile, the standard 12-lead ECG placeable on the human body is commonly used view to capture ECG heart's electrical signals [17]. The ECG recordings are frequently affected by various types of noise and artifacts, which can hinder the accurate identification of critical fiducial points, such as P, Q, R, S, and T, as well as associated intervals and offsets like P-onset, P-offset, QRS-onset, T-peaks, and T-offset [17]. There several sources of noise, including power line interference, baseline wander, and poor contact between electrodes and the skin [18, 19]. Analysing the ECG heart signal, feature extraction is usually performed to obtain the meaningful data for further analysis [6, 20].

Consequent upon the inadequacy with the traditional approach to ECG heartbeat analysis characterized by human errors, ML techniques are used to analyse the extracted features. The classifiers were designed to categories heart into various classes depending to the dataset adopted. Alarsan et al. [6] explored various ML techniques, including DT, RF, and GBDT. Their experiments on the MIT-BIH dataset revealed that RF outperformed other methods in classification accuracy. Similarly, Aziz et al. [21] developed two novel algorithms: Two-Event Related Moving Averages (TERMA) and Fractional Fourier Transform (FrFT). The TERMA algorithm identified regions of interest to locate specific peaks, while FrFT analysed ECG signals in the time-frequency domain to highlight peak locations. Features such as detected peaks, inter-peak durations, and other characteristics were then used to train Support Vector Machine (SVM) and Multilayer Perceptron (MLP) classifiers. The classifiers, trained on the MIT-BIH arrhythmia database, demonstrated robustness when tested on the INCART and SPH databases. Also, Malakouti [22] investigated various ML approaches, including Gaussian Naïve Bayes (NB), RF, Logistic Regression, Linear Discriminant Analysis (LDA), and Dummy Classifiers, to automate ECG heartbeat classification. The study employed 10-fold cross-validation to mitigate overfitting. Sinal et al. [23] had previously utilized k-Nearest Neighbours (KNN) and DT for multiclass classification of ECG signals.

While methods of ensemble [24] and heuristic optimization [25] techniques has been study for heartbeat classification, conventional ML lacks automatic feature extraction. Hence, DL techniques, particularly MLPs and CNNs, have gained popularity for their ability to automate feature extraction and classification of time-based signals. Gajendran et al. [26] studied the conversion of 1-D ECG signals into 2-D images using Continuous Wavelet Transform (CWT) to generate scalograms. These scalograms were processed using several pre-trained models, including VGGNet, Darknet, ResNet, GoogLeNet, EfficientNet, and DenseNet. However, the anomaly

approach to heartbeat detect is limited in the study. In an earlier study, Roy et al. [16] introduced ECG-NET, another LSTM-based AE model designed specifically for detecting anomalous ECG signals, with an emphasis on arrhythmia detection. Similar to Faraday's work [12], ECG-NET was trained exclusively on normal ECG signals, with anomalies identified using reconstruction loss thresholds established via manual and automated methods. The ECG-NET model's encoder processes 140 consecutive time steps of normal ECG data through three LSTM layers. The first layer, with 128 hidden neurons, captures temporal dependencies, producing hidden states that pass through a Rectified Linear Unit (ReLU) activation-function. The decoder mirrors the encoder's architecture, beginning with an LSTM layer containing 32 hidden neurons, followed by layers with 64 and 128 hidden neurons, respectively. ReLU activation and a 20% dropout rate in each layer enhance the model's ability to extract nonlinear temporal features. Nevertheless, these existing studies lacks capturing subtle, patient-specific irregularities without misclassifying them as noise. In addition, the techniques used are inadequate for accounting for the statistical variability of ECG features

3. METHODOLOGY

3.1 Proposed LAESN Architecture

The overall conceptual design of the LAESN model is depicted in Figure 1. The process begins with the preprocessing of raw ECG signals to remove artifacts and noise, ensuring cleaner and more reliable input data for subsequent analysis. Preprocessed signals undergo segmentation to extract temporal features and contextual relevance. Then, exploratory analysis is conducted on the signal, leveraging visualization techniques to uncover meaningful patterns. The ECG dataset is subsequently filtered, using normal signal subset as a training-set, while the entire data is used as test-set to develop a robust anomaly detection model.

The core of the framework utilizes an advanced LAESN (LSTM-based AE with Stratified Normalization Embedding) model. This model is specifically designed to analyze ECG signals, detect subtle variations, and distinguish between normal and abnormal patterns. The test data, is used to evaluate the model's performance, ensuring generalizability and accuracy in anomaly detection. The final outcome highlights anomalies in the heartbeat signals, contributing to the identification of potential heart-related issues.

3.2 Dataset Description

The dataset used in this study is the ECG5000 dataset. It consists 5,000 individual heartbeats extracted from a larger ECG recording. It is widely used for time series classification tasks, with each heartbeat categorized into five distinct classes representing different types of cardiac arrhythmias: Normal, Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC) and Paced Beat (PB) [16], as presented in Table 1. Each heartbeat is represented as a time series of 140 data points, sourced from the BIDMC Congestive Heart Failure Database (CHFD) on PhysioNet. The ECG signals were digitized at a sampling rate of 500Hz/s. The dataset is divided into a training set of 500 observations and a test set of 4,500 observations. Each sample consists of 141 feature points, including a target label. The ECG5000 dataset is sourced in the link at: https://www.timeseriesclassification.com/description.php?Dataset=ECG5000.

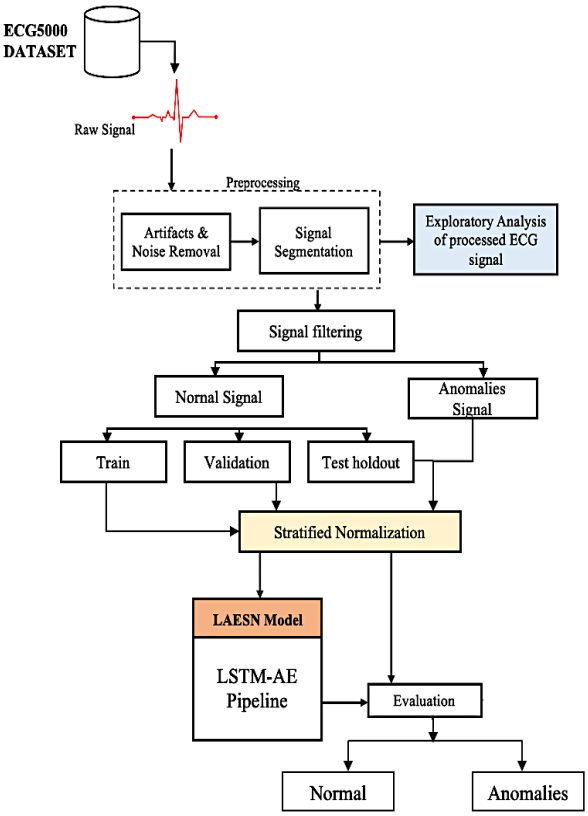


Figure 1: The conceptual framework of the LAESN model

Table 1: Overview of ECG500 dataset

Heartbeat Classes	Description
N-Normal	A normal rhythm of the heart
Left Bundle Branch Block (LBBB)	A situation where the left ventricle is activated later than normal.
Right Bundle Branch Block (RBBB)	A situation where the right ventricle activation is delayed.
Premature Ventricular Contraction (PVC)	It represents those extra heartbeats originating in the ventricles.
Paced Beat	These are the heartbeats from an artificial pacemaker. They more likely the fusion beats

3.3 Proposed Method

The proposed LAESN AD model integrates an AE with LSTM layers, with SN incorporated in the process to enhance the normalization of data. The core, AE includes

two primary components: the encoding and decoding phases expressed as shown in Equations (1) and (2).

$$e = f(Wx + b) \quad (1)$$

$$x' = f'(W'e + b') \quad (2)$$

where the activation-function is f, W denoted the weight matrix, b represents the bias vector, f', W', and b' are the parameters for the decoding phase, analogous to f, W, and b in the encoding phase.

Unlike the dense layer-based AE, LSTM layers are used to better capture the temporal dependencies inherent in ECG data, as illustrated in Figure 2. The LSTMs are particularly suitable for sequential data such as ECG signals due to the ability to process information across time steps effectively. The proposed LSTM model operates with a hidden layer comprising h units. Let $X_t \in \mathbb{R}^{n\times d}$ represent data input at time-step, t, where n is the batch size, and d is the feature-dimension. Similarly, $H_{t-1} \in \mathbb{R}^{n\times h}$ represents the hidden state from the previous time step. The LSTM model calculates the following gates for time step h [16]. The input, forget, and output gates are therefore expressed in Equations (3) – (5).

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \quad (3)$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \quad (4)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \quad (5)$$

where $W_{xi}, W_{xf}, W_{xo} \in \mathbb{R}^{d\times h}$, $W_{hi}, W_{hf}, W_{ho} \in \mathbb{R}^{h\times h}$, and $b_i, b_f, b_o \in \mathbb{R}^{1\times h}$ denote the weight matrices and bias vectors for the respective gates.

The activation function σ is typically the sigmoid function, which maps the input values to a range between 0 and 1. Memory Cell Update is then computed next, which is defined as the candidate memory cell $\hat{C}_t$ calculated using the hyperbolic tangent tanh function, producing values within the range [-1, 1]. This ensures smooth activation transitions and maintains stability during the learning process. The mathematical formulation for $\hat{C}_t$ is as in Equation (6).

$$\hat{C}_t = tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \quad (6)$$

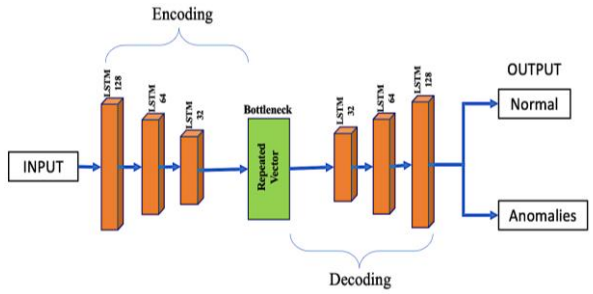


Figure 2: The proposed LSTM-AE model

Whereas, the Stratified Normalization (SN) shown in Figure 3 involves normalizing weights within each class using the class distribution as a prior and then performing normalization across classes. This approach, introduced by Song et al. [27], is grounded in the principle of stratified sampling. Earlier study to LogitBoost algorithm shows that SN is equally relevant for DL training, as demonstrated by [28]. Unlike instance normalization, which is typically restricted to image data and does not normalize across the batch, SN leverages participant labels as additional information to normalize features. This makes SN versatile and extendable to various types of data, beyond just image-based datasets. Mathematically, SN can be expressed for an input feature vector x belonging to a stratum k. The normalized value x_i' is computed using Equation (7).

$$\hat{x}_{i,k} = \frac{x_{i,k} - \mu_k}{\sigma_k + \epsilon} \quad (7)$$

where x_i is the i-th feature value of the input, μ_k denotes the mean and σ_k represents the standard deviation of the features for stratum k, ϵ is a small constant to ensure non-zero division, and x_i' represents the normalized feature value.

The mean, μ_k is computed as shown in Equation (8).

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_{i,k} \quad (8)$$

While the standard deviation, σ_k is expressed as in Equation (9).

$$\sigma_k = \sqrt{\frac{1}{N_k} \sum_{i=1}^{N_k} (x_{i,k} - \mu_k)^2} \quad (9)$$

The SN step-by-step process is defined as follows:

i. Identify which stratum k each input sample x_i belongs to
ii. For each stratum k, compute the mean μ_k and standard deviation σ_k based only on the samples belonging to that stratum.
iii. For every sample x_i in stratum k, apply the normalization formula defined for x_i'.

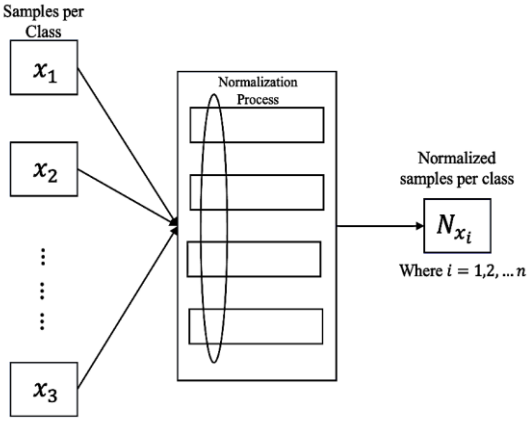


Figure 3: Stratified normalization process

3.2 Proposed Method

The following performance metrics were used in the study to evaluate the proposed LAESN model, including Mean Square Error (MSE), recall, F1-score, precision, Youden Index, and AUC. Accuracy was not primarily focused on, since it will not provide right judgement for an imbalance data. The brief description of the metrics is provided in Table 2 as follows:

Table 2: Overview of the performance metrics

Metrics	Formular
MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Accuracy	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$
Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{(TP + FP)}$
F1-score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$
Youden Index (J)	$Sensitivity + Specificity - 1$

where True Positives (TP) represent the number of cases correctly identified as positive. False Positives (FP) refer to instances that are incorrectly classified as positive when they are actually negative. True Negatives (TN) denote cases that are accurately identified as negative. False Negatives (FN) occur when positive cases are mistakenly labeled as negative, True Positive Rate (TPR), False Positive Rate (FPR), and True Negative Rate (TNR).

The AUC, which denotes the area under the curve, explains the TPR against the False Positive Rate (FPR). The AUC quantifies the model's ability to differentiate between classes—in this case, normal and anomalous heartbeats, as illustrated in Figure 4. A higher AUC indicates a stronger ability to correctly classify normal heartbeats as normal (0) and anomalies as anomalies (1). In the context of AD for heartbeats, a high AUC signifies a more effective model in distinguishing between patients with and without heart disease.

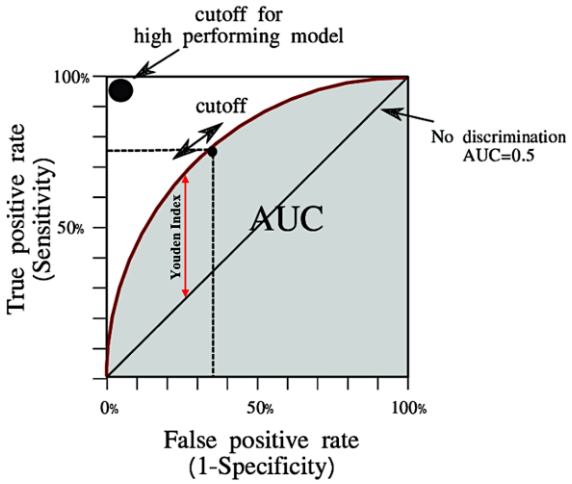


Figure 4: AUC and Youden Index metrics [29]

3.3 Experimental Setup

Experiments were conducted on the ECG5000 dataset, which was already pre-segmented. Further processing was performed to include assessment of noise removal using wavelet decomposition and data slicing to extract labels and features. While, the dataset consists of 400 training and 4,500 test samples across multiple classes, for this study, both sets were merged and reclassified into two categories: normal and abnormal heartbeat signals through filtering. The normal class was subsequently partitioned into training and test subsets in an 80:10 ratio, with an additional 10% of the test set held out for validation. In this study, imbalance inherent in the dataset is irrelevant during training, since the model is trained on normal signals. While it is understandable that the test set consisting the test holdout plus the anomalies might exhibit imbalance, a careful attention was paid to choose threshold for anomalies detection during evaluation. This threshold choosing criteria is further discussed in the result section.

As a baseline, an AE with Dense layers was trained exclusively on normal signals and evaluated on the validation set alongside abnormal samples. The proposed LAESN extended this baseline by replacing Dense layers with LSTM units to capture temporal dependencies and inter-individual variability, while incorporating SN at the input preprocessing stage. SN was applied separately within each stratum of class labels, addressing amplitude heterogeneity across heartbeat categories. Crucially, SN was performed strictly before model training and did not interact with internal layers, thereby preventing label leakage while improving stability through consistent feature scaling.

Model training employed the Adam optimizer with a batch size of 32 for 50 epochs, and LSTM layers utilized the 'Tanh' activation function to ensure stable gradient flow. All experiments were conducted in Python 3.9 using Keras and TensorFlow on an Ubuntu 22.04 system equipped with a Ryzen 7 5700G CPU, 64 GB RAM, and an RTX 4060 Ti 16G GPU. Model hyperparameters are summarized in Table 3.

Table 3: Hyperparameter setting for LAESN model

Hyperparameter	Value
Number of Hidden Units	64, 32
Dropout	0.3
AF	tanh
Optimizer	Adam
Batch Size	32
Sequence length	140
Epoch	50
Early stopping	Monitor='loss', patience=5
Model loss	MSE
Seed	42
Model runtime	3ms/step

4. RESULTS AND DISCUSSION

4.1 Exploratory Analysis

On visualizing the ECG5000 heartbeat signals, as shown in Figure 5, it was observed that the heartbeat categories exhibit distinguishable temporal and amplitude characteristics across different classes. The Normal class (depicted in bold blue) shows a well-defined waveform with sharp peaks and troughs, consistent with typical ECG patterns. In contrast, the LBBB and RBBB signals in green dotted and purple dashed lines, respectively revealed critical deviations in their QRS complexes. This indicates delays in electrical conduction. The PVC class in red dashed line exhibits the most pronounced anomaly, with an early and widened QRS complex and an irregular shape, deviating significantly from the Normal waveform, while the Paced Beat in orange line presents a distinct waveform altogether, characterized by flatter segments and an absence of the typical ECG morphology. The visualization provides the insight into the challenge posed by intra-class variability and inter-class similarities, especially between LBBB, RBBB, and Normal classes, which can affect the performance of traditional ML classifiers.

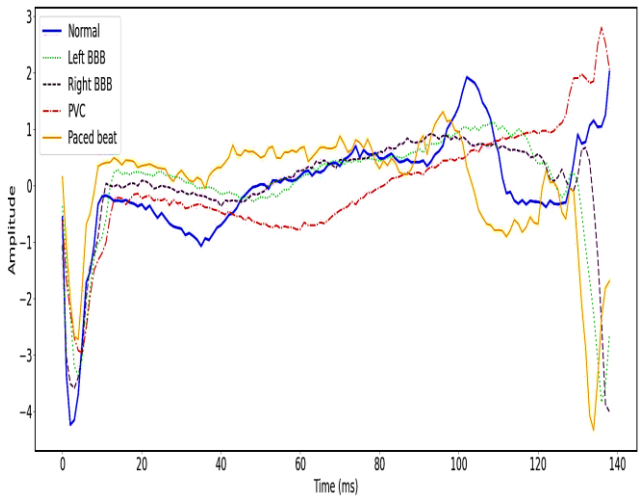


Figure 5: Temporal and amplitude characteristics across different classes of ECG5000 dataset

4.2 Class Distribution, Reconstruction Errors, and Thresholding

Resulting from the class distribution, depicted in Figure 6, it was observed that normal heartbeats constitute the majority of the recorded signals, accounting for 2919 (58.38%) followed by the LBBB beats representing 1767 (35.34%). The remaining categories RBBB, PVC, and Paced beats, account for smaller proportions, at 194 (3.88%), 96 (1.92%), and 24 (0.48%) respectively. Leaning on this perspective, the distribution highlights the inherent class imbalance within the ECG5000 dataset. The dominance of normal heartbeats is potential to the models being biased towards this class, hence exhibiting high accuracy on normal beats but performing poorly on the less frequent but clinically significant abnormal beats.

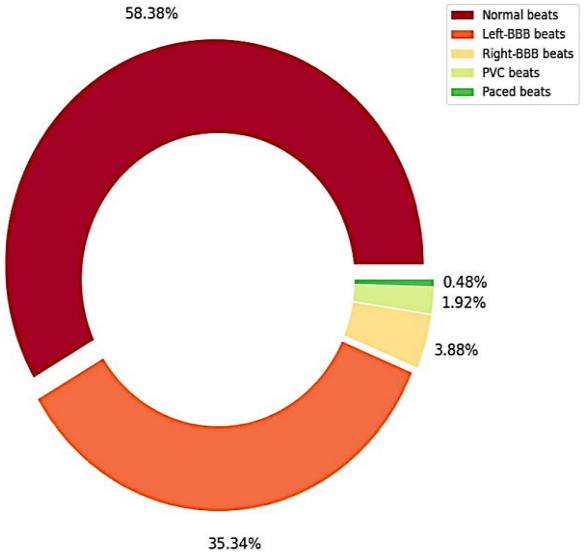


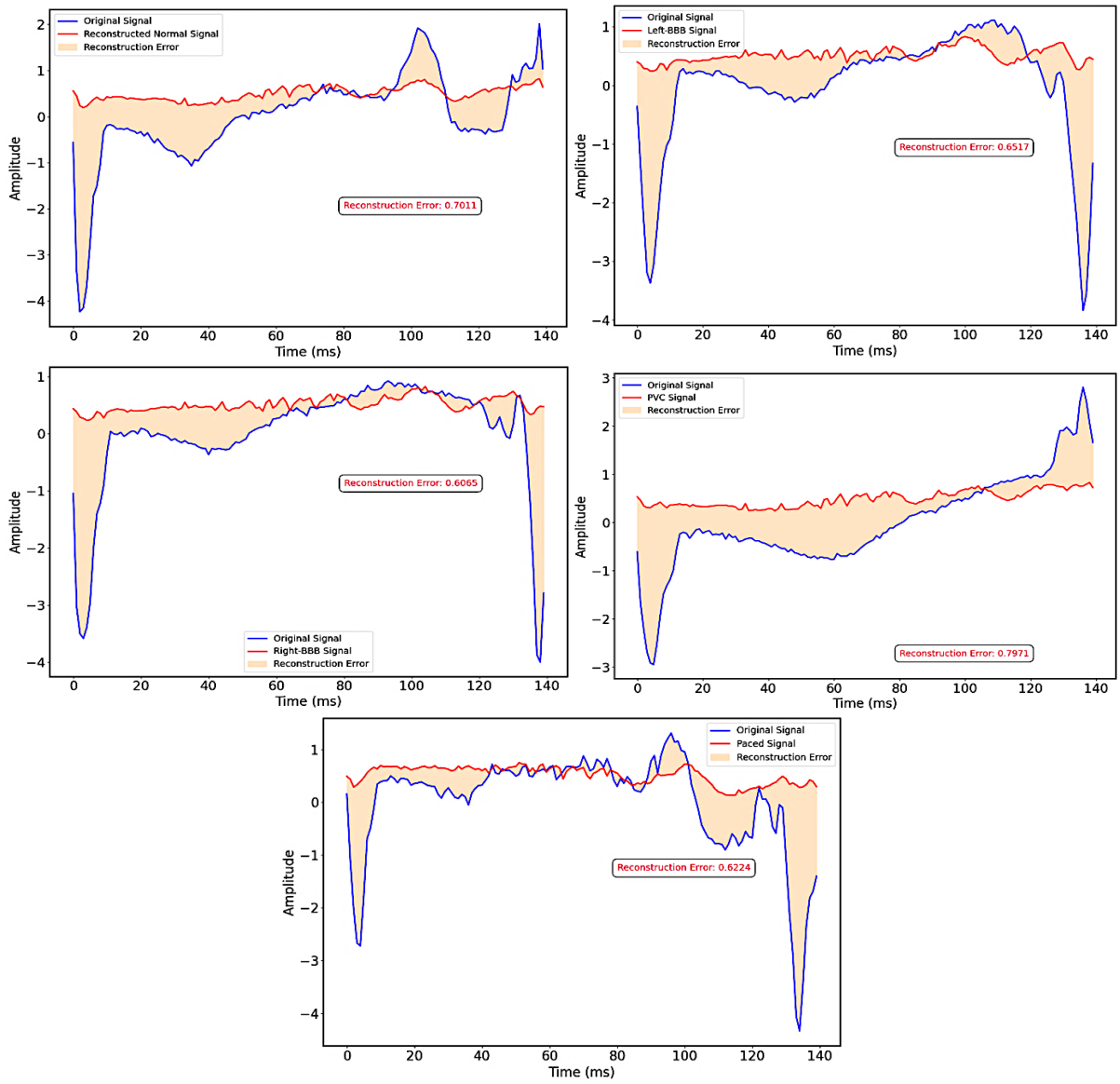
Figure 6: Distribution of classes in ECG5000 dataset

The reconstruction errors plot for the models training, are as shown, in Figure 7 (a) and (b). Each subplot displays the original ECG signal in blue overlaid with the reconstructed signal from the model (in red). The shaded orange region visually represents the reconstruction error, along with the calculated RMSE for each representative heartbeat. It was observed that each model in (a) and (b) provides varying ability to capture and reproduce the characteristics of different heartbeat morphologies. For instance, some heartbeat types appear to be reconstructed with higher fidelity (smaller error area and lower RMSE value) compared to others. Higher reconstruction errors shown in certain heartbeat types suggest that the model struggles to learn the underlying patterns of these specific morphologies as effectively as others. The highlights the need to address the lower prevalence of these heartbeat types in the training data or the presence of unique features that are not well captured by the model's latent space.

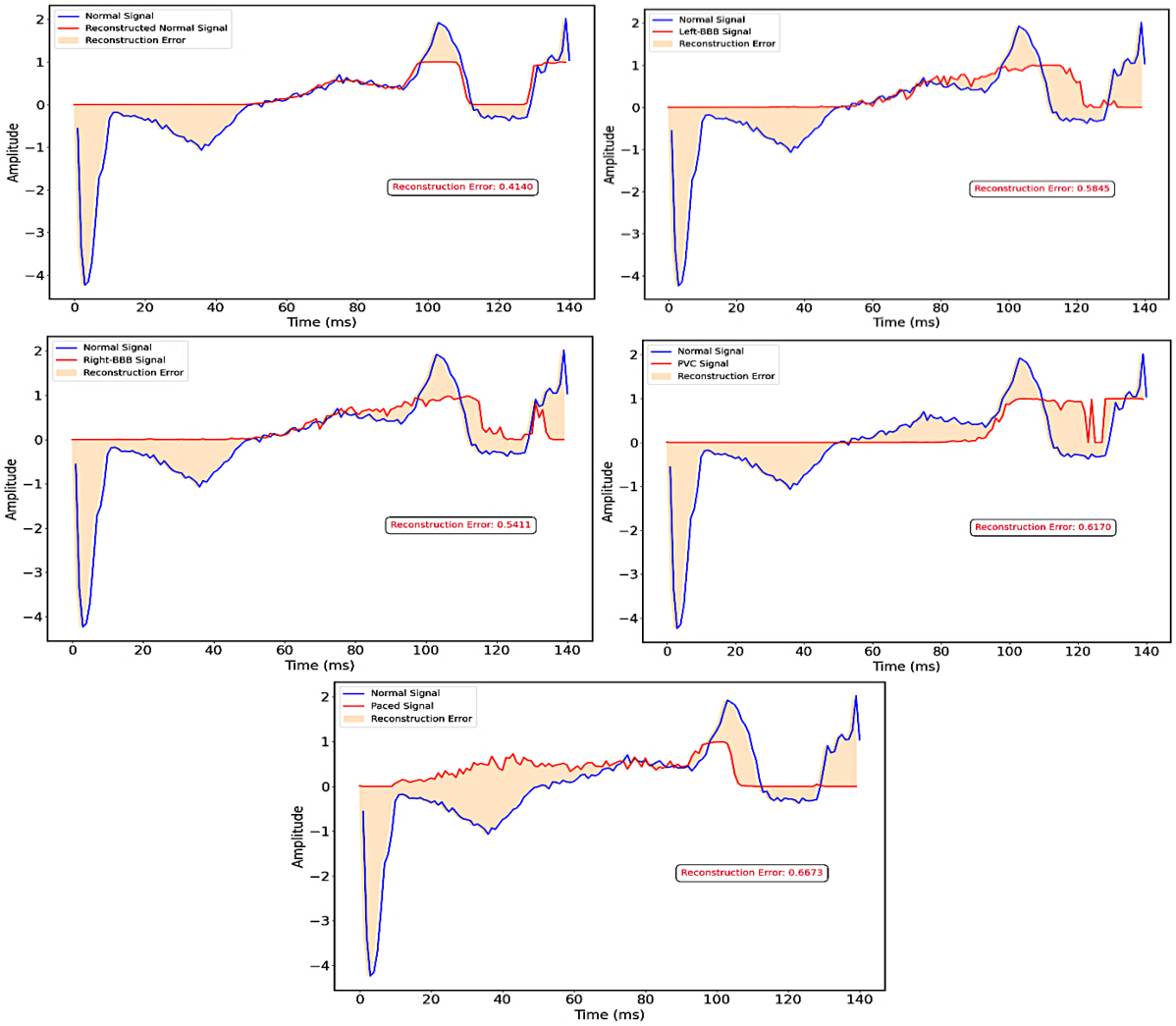
Comparatively, the reconstruction errors and the RMSE values obtained across the corresponding heartbeat types in both (a) and (b) show the proposed LAESN model offers

improvements in capturing the nuances of the ECG signals than the baseline AE. The observation suggests that the LAESN model incorporating temporal dependencies and a more sophisticated latent space, is better at reconstructing those particular morphologies than the baseline AE. Conversely, there is also observed where the reconstruction errors are comparable or even higher for certain heartbeat types in the LAESN model, indicating areas where the baseline AE performs just as well or even better.

Anomaly detection was conducted by defining the threshold at the 95th percentile of the validation reconstruction error distribution. Threshold determination was guided by identifying an optimal cut-off point that mitigates the effects of class imbalance during evaluation. In the study, two strategies are employed: a non-parametric approach, where thresholds are set at specific percentiles of the reconstruction error distribution (e.g., 95th, 97.5th, or 99th), and a parametric approach, where the threshold is computed as $\mu + k \times \sigma$ (with k = 2 or 3) [30]. Given the highly imbalanced nature of the dataset, greater emphasis was placed on reducing false negatives, since missed anomalies are typically more costly than false positives. The fitted normal distribution (μ = 0.0151, σ = 0.0101) reinforces this rationale, as reconstruction errors for normal samples are tightly clustered around the mean, forming a sharp peak with rapid decay, as shown in Figure 8. By contrast, anomalous cases are more likely to appear in the distribution's tail, where higher reconstruction errors are concentrated. This distinct separation between the dense cluster of normal errors and the sparse anomalous tail provided a principled justification for selecting the non-parametric threshold. Accordingly, the threshold was deliberately biased toward maximizing recall.



(a)



(b)

Figure 7: Reconstruction error plots for (a) proposed LAESN and (b) baseline AE models

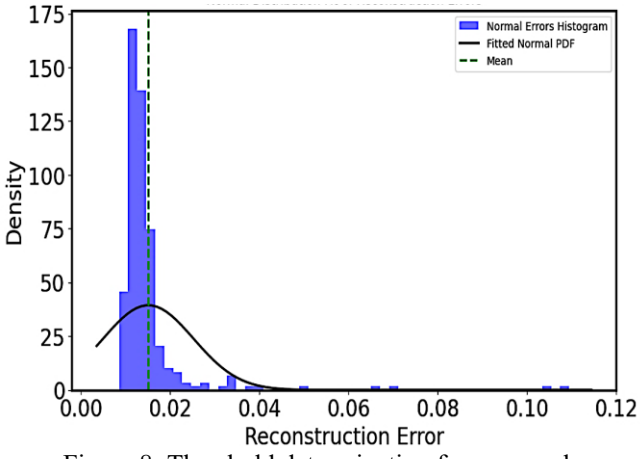


Figure 8: Threshold determination from normal distribution

4.2 Model Results

The comparative analysis of model performance in aggregated recording, illustrated in Figure 9, demonstrates that the proposed LAESN model achieves an accuracy of 97.08%, outperforming the Baseline-AE, which attains 92.92%. With respect to the F1-score, LAESN records 96.60% compared to 91.40% for the Baseline-AE. Similarly, LAESN achieves an AUC of 99.52% and a Youden Index (J) of 95.00%, exceeding the corresponding values of 95.75 % and 85.00% for the Baseline-AE. These results affirm the superior discriminative capacity and more optimal decision threshold of LAESN relative to the Baseline-AE. Importantly, both models report comparable specificity of 95.00%, suggesting that the substantial gain in sensitivity observed in LAESN does not compromise its ability to correctly classify normal heartbeats.

When benchmarked against conventional anomaly detection methods such as Isolation Forest (IF) and One-Class SVM (OC-SVM), LAESN consistently outperforms or closely matches their performance across multiple metrics. For instance, while IF and OC-SVM achieve recalls of 97.93% and 99.09% respectively, their corresponding F1-scores of 93.49% and 95.93% remain below the 96.60% attained by LAESN. Furthermore, the AUC values indicate that Baseline-AE achieves 95.75%, IF achieves 95.15%,

and OC-SVM achieves 99.23%, whereas LAESN maintains superiority at 99.52%. The Youden Index further reinforces this trend, with LAESN scoring J of 95.00, surpassing IF of 89.68 and OC-SVM 93.74, thereby reflecting a more favourable balance between sensitivity and specificity.

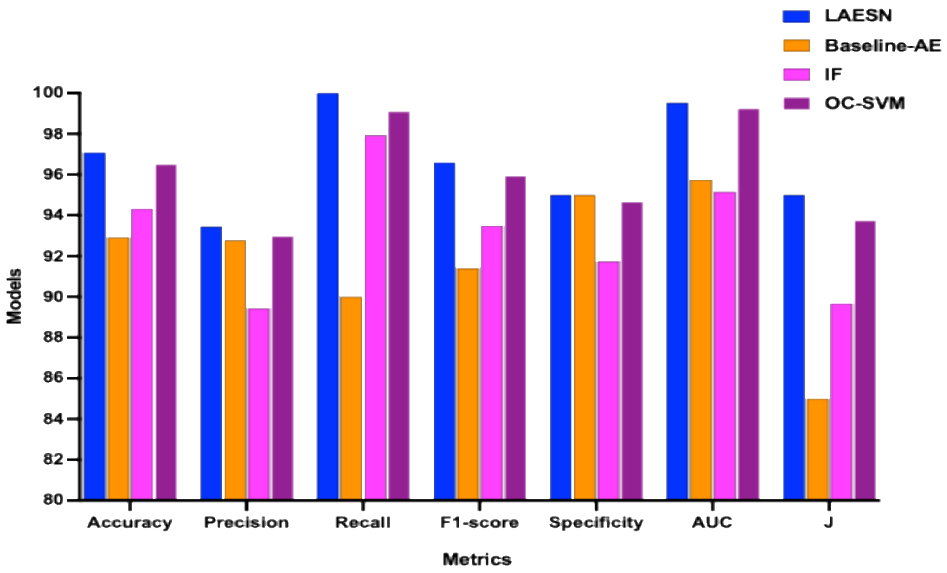


Figure 9: Comparative analysis between LAESN and other ML anomaly models

Further investigation using the TPR–FPR curves, as shown in Figure 10, reveals that the baseline-AE achieves an AUC of 0.9575, while the LAESN model attains a markedly higher value of 0.9952, reflecting its stronger capacity to distinguish between different heartbeat classes. The curve in panel (b) highlights limitations in the baseline-AE, whereas panel (a) demonstrates the improvement achieved by LAESN. In comparison, conventional anomaly detection models such as IF achieving 95.15% and OC-SVM of 99.23% yield competitive AUC scores, yet LAESN remains superior.

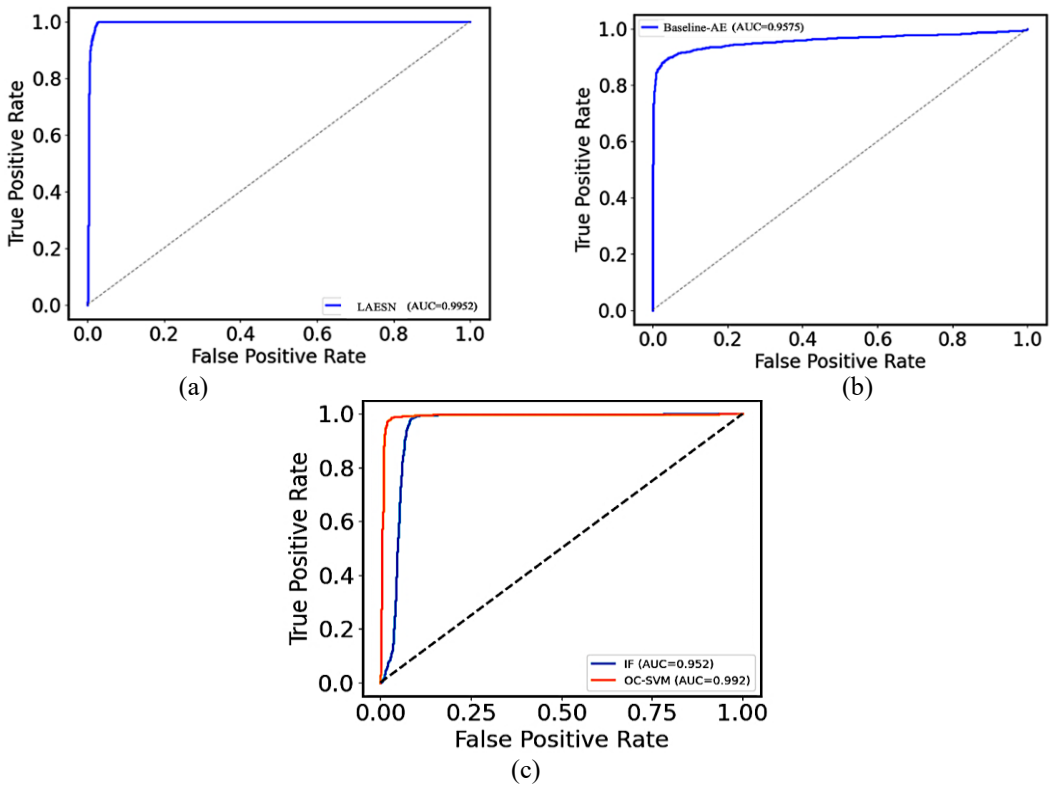


Figure 10: TRP-FPR curve of (a) LAESN, (b) Baseline-AE, and (c) IF and OC-SVM models

From a clinical perspective, the implications of these findings are significant. LAESN's recall of 100% indicates that no pathological heartbeats were missed, although such outcomes may require further investigation and careful threshold calibration. Its precision of 93.44% further ensures that the majority of detected abnormalities correspond to true positive cases. Crucially, the F1-score underscores the balance achieved between precision and recall, demonstrating LAESN's ability to minimize both false positives and false negatives. This balance is essential in cardiology, ensuring models that maximize recall at the expense of precision risk overwhelming clinicians with false alarms, while models biased toward precision may fail to detect life-threatening arrhythmias. Another notable trade-off lies in threshold sensitivity. Models tuned for maximal recall (as in the case of LAESN) may achieve perfect sensitivity but risk reductions in precision if thresholds are not carefully managed. Conversely, Baseline-AE, with a recall of 90%, avoids over-detection but misses a clinically significant proportion of abnormal heartbeats, an outcome that is less desirable in medical screening contexts.

In comparison with state-of-the-art models, it was observed that the LAESN model demonstrates an AUC of 0.9952, showing comparative advantage with models in existing studies. When compared with [9], who employed a VRAE+SVM approach obtaining an accuracy of 0.9843, LAESN records a slightly lower accuracy of 0.9708. However, the proposed LAESN with AUC score of 0.9952 better surpasses their AUC score of 0.9836, indicating better discriminatory power. Similarly, while [16] achieved a marginally higher accuracy of 0.9842 using Deep LSTM-AE, the proposed LAESN offers a less accuracy of 0.9708. However, LAESN demonstrated a more balanced performance by achieving both a high AUC 0.9952 and an F1-score of 0.9661 (notably absent in Roy et al.'s work). The inclusion of F1-score in this work showcased transparency and a stronger focus on imbalanced data scenarios, where F1-score is known to provide a more holistic view of classification performance than accuracy alone. Additionally, Time Series Memory Augmented Autoencoder (TSMAE) model in [31] achieved an AUC of 0.9516, which is significantly lower than that of LAESN, further affirming LAESN's competitive detection capability.

Moreover, looking at issues relating to deployment perspective, especially in resource-constrained or real-time settings, VRAE+SVM offers computational efficiency, ease of implementation, and better interpretability, as VRAE are typically more lightweight compared to sequential models like LSTM-based architectures. However, computational efficiency and interpretability are of less concerns in the current wave of technology with vast high-end resources and deep interpretable networks framework easily accessible. In addition, empirical evidence obtained from the motivation behind adopting LSTM-AE and supporting with the integration of SN to enhance sensitivity to inter-patient variability has made a significant performance contribution to anomalous heartbeat detection.

4.3 Limitations

Despite the performance exhibited by the LAESN model, several limitations warrant consideration. First, the experimental evaluation was conducted on a single benchmark dataset, which restricts the scope of generalization. While this dataset is widely used in the anomaly detection community, its relatively constrained variability in heartbeat patterns may not fully reflect the heterogeneity present in real-world clinical populations. Consequently, the limitation underscores the importance of extending evaluation to larger and more diverse datasets. Secondly, the current framework implemented a data split to avoid unintended information leakage during. However, training and testing on overlapping patient data can artificially inflate performance metrics, further research interest may include rigorous evaluation protocols, such as leave-one-subject-out validation or stratifying validation, to ensure that the reported performance reflects genuine generalization rather than dataset-specific bias.

Model's reliance on reconstruction error thresholding introduces sensitivity to threshold selection. As highlighted in the preceding discussion, thresholds set at the distribution tail strongly influence the trade-off between recall and precision. While recall-biased thresholding proved advantageous in minimizing false negatives in this study, it may not be universally optimal across datasets with different class distributions. An overly aggressive threshold could increase false positives, potentially reducing clinical trust in the system. This challenge points toward the need for adaptive or dynamic thresholding strategies, possibly guided by cost-sensitive learning or Bayesian uncertainty quantification, to provide more robust decision-making across varying clinical contexts. These limitations suggest several directions for future research. Expanding evaluations to broader and more heterogeneous datasets would test the scalability and clinical robustness of LAESN. Incorporating subject-level validation protocols would mitigate risks of information leakage, while adaptive thresholding methods could refine sensitivity-specificity trade-offs in practice. Moreover, integration with real-time monitoring systems and prospective clinical trials could provide crucial insights into the translational potential of LAESN.

5. CONCLUSION

This study explored the detection of anomalies in heartbeat signals, discussing the critical role of the heart in sustaining life and the danger posed by abnormal rhythms, which has been recorded leading to severe health complications or even death. The study also highlighted early detection systems being vital for timely intervention. Moreover, traditional methods for analyzing heartbeat patterns are often limited by complexity and lack of precision. However, technological advancements, particularly in ECG monitoring and ML, have change the narrative both the timely collection and accurate interpretation of cardiac signals. While previous studies have introduced various ML models for heartbeat anomaly

detection, a major limitation has been their insufficient sensitivity to inter-patient variability. To address this gap, the proposed LAESN model was developed and evaluated using the ECG5000 dataset. The model achieved a superior AUC score of 99.52%, outperforming both the baseline AE and several state-of-the-art models. This highlights LAESN's robust capability to detect anomalous heartbeats by effectively capturing temporal dependencies and subtle signal variations through its LSTM-based architecture that leans on the SN. The results underscore the importance of adopting advanced feature extraction and modeling techniques, LAESN, that are sensitive to diverse patient-specific patterns. For future work, it is recommended to evaluate the model on additional ECG datasets to assess its generalizability to unseen cases. Furthermore, the integration and optimization of SN within the LSTM-AE framework as layer-embedding can be further explored to enhance performance and adaptability in real-world clinical applications.

REFERENCES

[1] Saini, S.K. and Gupta, R., 2022. Artificial intelligence methods for analysis of electrocardiogram signals for cardiac abnormalities: state-of-the-art and future challenges. Springer Netherlands.

[2] Mathews, S.M., Kambhamettu, C. and Barner, K.E., 2018. A novel application of deep learning for single-lead ECG classification. Computers in Biology and Medicine, 99, pp.53–62. https://doi.org/10.1016/j.compbiomed.2018.05.013

[3] Murat, F., Yildirim, O., Talo, M., Baloglu, U. B. and Acharya, U. R., 2020. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. Computers in Biology and Medicine, 120, p.103726. https://doi.org/10.1016/j.compbiomed.2020.103726

[4] Johnson, O.V., XinYing, C., Khaw, K.W. and Ming, L.H., 2023. A cost-based dual ConvNet-attention transfer learning model for ECG heartbeat classification. Journal of Information and Web Engineering, 2, pp.90–110. https://doi.org/10.33093/jiwe.2023.2.2.7

[5] Alqudah, A.M., Qazan, S., Al-Ebbini, L., Alquran, H. and Abu-Qasmieh, I., 2022. ECG heartbeat arrhythmias classification: a comparison study between different types of spectrum representation and convolutional neural networks architectures. Springer Berlin Heidelberg.

[6] Alarsan, F.I. and Younes, M., 2019. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. Journal of Big Data, 6. https://doi.org/10.1186/s40537-019-0244-x

[7] Johnson, O.V., XinYing, C., Khaw, K.W. and Lee, M.H., 2023. ps-CALR: Periodic-Shift Cosine Annealing Learning Rate for Deep Neural Networks. IEEE Access, 11, pp.139171–139186. https://doi.org/10.1109/ACCESS.2023.3340719

[8] Park, J., Kim, D.Y., Kim, Y., Yoo, J. and Kim, T. J., 2024. EB-GAME: A Game-Changer in ECG Heartbeat Anomaly Detection. arXiv preprint.

[9] Rawi, A.A., Elbashir, M.K. and Ahmed, A.M., 2022. ECG Heartbeat Classification Using CONVXGB Model. Electronics (Switzerland), 11. https://doi.org/10.3390/electronics11152280

[10] Pereira, J. and Silveira, M., 2019. Learning Representations from Healthcare Time Series Data for Unsupervised Anomaly Detection. 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), pp.1–7. https://doi.org/10.1109/BIGCOMP.2019.8679157

[11] Sivapalan, G., Nundy, K.K., Dev, S., Cardiff, B. and John, D., 2022. ANNet: A lightweight neural network for ECG anomaly detection in IoT edge sensors. IEEE Transactions on Biomedical Circuits and Systems, 16, pp.24–35.

[12] Farady, I., Patel, V., Kuo, C.C. and Lin, C.Y., 2024. ECG Anomaly Detection with LSTM-Autoencoder for Heartbeat Analysis. Digest of Technical Papers - IEEE International Conference on Consumer Electronics, pp.1–5. https://doi.org/10.1109/ICCE59016.2024.10444327

[13] Liu, P., Sun, X., Han, Y., He, Z., Zhang, W. and Wu, C., 2022. Arrhythmia classification of LSTM autoencoder based on time series anomaly detection. Biomedical Signal Processing and Control, 71, p.103228. https://doi.org/10.1016/j.bspc.2021.103228

[14] Kieu, T., Yang, B., Guo, C. and Jensen, C.S., 2019. Outlier detection for time series with recurrent autoencoder ensembles. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2019, pp.2725–2732. https://doi.org/10.24963/ijcai.2019/378

[15] Ou, Y., Li, X., Guo, Z. and Wang, Y., 2020. Anobeat: Anomaly detection for electrocardiography beat signals. Proceedings - 2020 IEEE 5th International Conference on Data Science in Cyberspace (DSC), pp.142–149.

[16] Roy, M., Majumder, S., Halder, A. and Biswas, U., 2023. ECG-NET: A deep LSTM autoencoder for detecting anomalous ECG. Engineering Applications of Artificial Intelligence, 124, p.106484. https://doi.org/10.1016/j.engappai.2023.106484

[17] Berkaya, K.S., Uysal, A.K., Gunal, S. E., Ergin, S., Gunal, S. and Gulmezoglu, M. B., 2018. A survey on ECG analysis. Biomedical Signal Processing and Control, 43, pp.216–235. https://doi.org/10.1016/j.bspc.2018.03.003

[18] Terada, T., Toyoura, M., Sato, T. and Mao, X., 2021. Noise-reducing fabric electrode for ECG measurement. Sensors, 21, pp.1–17. https://doi.org/10.3390/s21134305

[19] Boda, S., Mahadevappa, M. and Dutta, P.K., 2021. A hybrid method for removal of power line interference and baseline wander in ECG signals using EMD and

EWT. Biomedical Signal Processing and Control, 67, p.102466. https://doi.org/10.1016/j.bspc.2021.102466

[20] Berwal, D., Vandana, C.R., Dewan, S., Jiji, C. V. and Baghini, M., 2019. Motion Artifact Removal in Ambulatory ECG Signal for Heart Rate Variability Analysis. IEEE Sensors Journal, 19,pp.12432–12442. https://doi.org/10.1109/JSEN.2019.2939391

[21] Aziz, S., Ahmed, S. and Alouini, M.S., 2021. ECG-based machine-learning algorithms for heartbeat classification. Scientific Reports, 11, pp.1–14. https://doi.org/10.1038/s41598-021-97118-5

[22] Malakouti, S.M., 2023. Heart disease classification based on ECG using machine learning models. Biomedical Signal Processing and Control, 84, p.104796.

[23] Bin Sinal, M.S. and Kamioka, E., 2018. Early abnormal heartbeat multistage classification by using decision tree and K-nearest neighbor. ACM International Conference Proceeding Series, pp.29–34. https://doi.org/10.1145/3299819.3299848

[24] Alfaras, M., Soriano, M.C. and Ortín, S., 2019. A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection. Frontiers in Physics, 7, pp.1–11. https://doi.org/10.3389/fphy.2019.00103

[25] Hassaballah, M., Wazery, Y.M., Ibrahim, I. E. and Farag, A., 2023. ECG heartbeat classification using machine learning and metaheuristic optimization for smart healthcare systems. Bioengineering, 10, p.429.

[26] Gajendran, M.K., Khan, M.Z. and Khattak, M.A.K., 2021. ECG classification using deep transfer learning. Proceedings - 2021 4th International Conference on Information and Computer Technologies (ICICT), pp.1–5.

[27] Song, J., Lu, X., Liu, M. and Wu, X., 2011. Stratified normalization logitboost for two-class unbalanced data classification. Communications in Statistics - Simulation and Computation, 40, pp.1587–1593. https://doi.org/10.1080/03610918.2011.589332

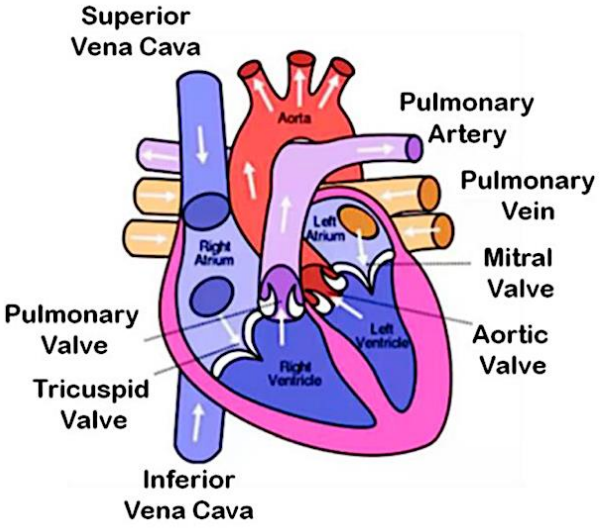
[28] Fdez, J., Guttenberg, N., Witkowski, O. and Pasquali, A., 2021. Cross-Subject EEG-Based Emotion Recognition Through Neural Networks With Stratified Normalization. Frontiers in Neuroscience, 15. https://doi.org/10.3389/fnins.2021.626277

[29] Nahm, F.S., 2022. Receiver operating characteristic curve: overview and practical use for clinicians. Korean Journal of Anesthesiology, 75(1), pp.25–36.

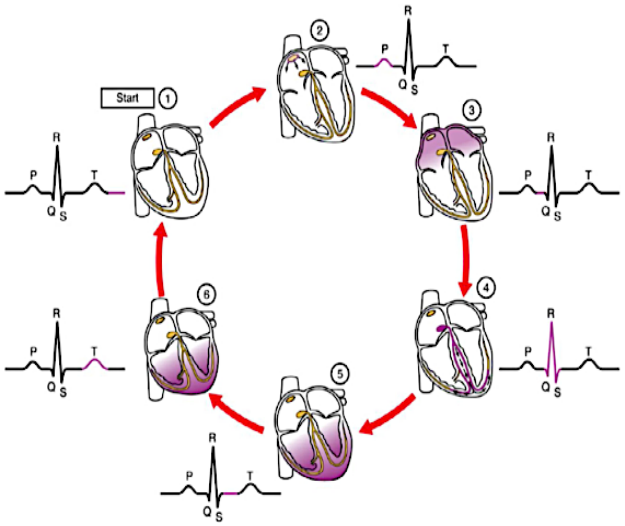
[30] Blázquez-García, A., Conde, A., Mori, U. and Lozano, J.A., 2021. A review on outlier/anomaly detection in time series data. ACM Computing Surveys (CSUR), 54(3), pp.1–33.

[31] Gao, H., Qiu, B., Barroso, R.J.D., Hussain, W., Xu, Y. and Wang, X., 2023. TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder. IEEE Transactions on Network Science and Engineering, 10, pp.2978–2990. https://doi.org/10.1109/TNSE.2022.3163144.

APPENDIX I



(a) A view of the human heart [1]



(b) Cardiac cycle process [17]