International Journal of Communication Networks and Information Security
2024, 16(4)
ISSN: 2073-607X,2076-0930
https://ijcnis.org/
Research Article

A Hybrid Deep Learning Approach for ECG Arrhythmia Detection: GPT, GANs, and Triplet Loss Integration

Sultan faiz alqurashi, Yasser Mefreej Alahmadi, abdulrahman mohammed alsherbi, Faisal nawar althobaiti, Eid Ayed Alosaimi, Fawaz Matuq Almuwallad, Abdulmajeed Abdullah Alswat, Abdulaziz selmi alsaedi, Sameer mubarak alqurashi, Mohammed Abdullah Alharthi, Mohsen hassan alsrori, Ghorm Ahmad Alghamdi, Khalid Owidh Alotibi
Saudi Red Crescent Authority, Saudi Arabia
Emails: Sultan380997997@gmail.com, alahmadi_20@hotmail.com, Dhm-1411@hotmail.com, Faissal611@hotmail.com, Eid7918@hotmail.com, fawazalmowalad@gmail.com, alswat2022@gmail.com, A.z.0-1@hotmail.com, semo1515@hotmail.com, Mohammad-ams@hotmail.com, M7sin@outlook.com, alm7targ@gmail.com, kfh.1200000@gmail.com

ARTICLE INFO

Received: 17 Aug 2024
Accepted: 25 sep 2024

ABSTRACT

This paper proposes a novel deep-learning method to detect arrhythmias from the ECG data by adopting pre-trained GPT models and other powerful state-of-the-art DL algorithms. Most traditional ECG classification models face challenges in capturing complex temporal dependencies and handling class imbalances. To meet these challenges, our system leverages GPT to capture complex temporal patterns and contextual relationships within ECG signals, enabling us to better under-stand the more intricate depen-dencies in the data. Finally, the proposed system lever-ages data augmentation with Generative Adversarial Networks (GANs) to generate a wide variety ofcomplex samples, which help improve model capability and robustness. It also uses Triplet Loss, which shows it can work better on imbalanced classes and tiny differencesin different cardiac arrhythmias. Compared with other methods, our results exhibit great im-provements in classification performance, particularly for rare arrhythmias. Model Interpretability is based on SHapley Additive exPlanations(SHAP) and Gradient weighted Class Activation Map (Grad-CAM), which interpret the model decisions.

Keywords: ECG Classification, GPT, GANs, Triplet Loss, Model Interpretability, Arrhythmia Detection

1. Introduction

Electrocardiography (ECG) is one of the core approaches in modern medicine to provide important information regarding the heart's electrical activity via long-term monitoring and for diagnosing different cardiac conditions [1]. This detection is essential for early intervention and treatment, as arrhythmias and irregular heartbeats can result in fatal consequences such as stroke or even worsening of local muscle function that may possibly lead to a worsening health condition [2]. ECG interpretation is traditionally performed by expert clinicians, which entails a manual visual inspection of ECG traces that has the potential for error. Therefore, the demand for automated ECG classification tools that develop diagnostic precision and operational effectiveness has increased [3, 4].

The newly developed deep learning methods have made a significant breakthrough in the upgradation of automatic ECG classification. For the purpose of extracting spatial features from ECG signals,Convolutional Neural Networks (CNNs) were used to capture temporal dependencies Recurrent Neural Networks (RNNs) [5]. While great strides have been made over the years with these solutions, traditional models are not without their limitations. While Convolutional Neural Networks (CNNs) may not capture all long-term temporal patterns, Recurrent NN models such as LSTMs easily suffer from poor gradients with significant burdens on computational costs [6]. Furthermore, the issue of class imbalance has not been completely eliminated because some types of arrhythmias are found infrequently in clinical practice, and training dataset performance remains biased with low sensitivity for rare arrhythmia [7].

Initially created for Natural Language Processing (NLP), transformers have been proven to be very effective in sequence modeling as they keep long-term correlations and context-aware relationships [8]. While RNNs process sequences in order, Transformers employ self-attention mechanisms to assess the entire sequence simultaneously [9], compensating for certain deficiencies of RNNs. In this research, they meant a study on applying GPT to ECG arrhythmia classification [10]. The ability of GPT to comprehend and create contextually rele-vant sequences helps improve arrhythmia classification accuracy, particularly for identifying rare and subtle types [10].

Triplet Loss Also, the triplet loss alleviates imbalanced class problems by distinguishing intra-class variation and inter-class difference by focusing on relationships between examples [11]. This method enhances the classification performance, especially for underrepresented arrhythmia types [12].

Furthermore, model interpretability is crucial for clinical adoption. To facilitate understanding of the Model's decisions, we integrate interpretability tools using Grad- CAM [13]. These tools provide visual and quantitative insights into the Model's decision-making process, aiding clinicians in validating and trusting the automated diagnoses. This dual approach ensures the Model's accessibility in clinical settings with limited internet connectivity. It supports broader integration with cloud-based healthcare systems, enhancing the overall diagnostic capabilities available to healthcare providers.

The contributions of the proposed system in this work are:

- We utilize GPT's powerful temporal pattern recognition and contextual analysis capabilities within the proposed system to significantly enhance the classification of ECG arrhythmias, focusing on accurately identifying rare and complex arrhythmia types.
- By employing Generative Adversarial Networks (GANs) [14] within the proposed system for data augmentation, we generate synthetic ECG samples that address class imbalance and expand the diversity of training data, leading to improved model robustness and generalization.
- We incorporate Triplet Loss within the Proposed system to strategically manage class imbalances, focusing on optimizing the relative distances between ECG examples. This approach enhances the Model's ability to discriminate between various arrhythmias, particularly those that are underrepresented.
- We integrate interpretability tools using SHAP and Grad-CAM within the Proposed system to provide detailed insights into the Model's decision-making process, ensuring transparency and facilitating clinical validation of the predictions.

The remainder of this paper is structured as follows: Section 2 reviews the related work, including traditional machine learning approaches, deep learning models, transformers, and methods for addressing class imbalance in ECG datasets, as well as interpretability in deep learning. Section 3 provides an in-depth explanation of the proposed methodology. Section 4 presents the experimental setup, followed by an evaluation of the Model's performance in Section 5. We also benchmarked our Model against existing models and assessed the impact of triplet loss on class imbalance. Section 6 discusses the study's limitations and potential future work and concludes the paper by summarizing the key findings and contributions.

2. Related Work

The field of ECG arrhythmia classification has evolved significantly over the years, with various approaches being explored to improve accuracy, robustness, and interpretability. The present section discusses the literature and divides it into five different entities, which are as follows: Traditional machine learning techniques, deep learning solutions, Transformers & Attention mechanisms, Tackling Class Imbalance in ECG, Datasets, and Interpretable Deep Neural Networks.

2.1 Traditional Machine Learning Approaches

Machine learning models have previously been heavily utilized in ECG automated classification research, mainly concentrating on classical machine learning methods. These methods relied on handcrafted feature selection from ECG signals and classification with algorithms like Support Vector Machines (SVM) [10], K-Nearest Neighbors (KNN), and Random Forests. For example, different works have used SVMs to classify arrhythmia from heart rate variability and wavelet coefficients [15–18]. Distance metrics were computed on feature vectors extracted from ECG signals [19] so KNN classifiers could discern normal and abnormal heartbeats. An adaptive ensemble learning approach was used as Random Forests to overcome the curse of dimensionality for ECG data features. However, this approach has proven more effective than traditional techniques [20]. While these approaches provided some basis, they depended on handcrafted features, and their generalizability to other patient cohorts was limited.

Despite their early success, some common shortcomings of the traditional machine learning approach limit them from performing well in ECG classification tasks under modern setups. This manual feature extraction is labor intensive and requires domain expertise, making it non-scalable for adapting to different ECG datasets.

Additionally, their models often underperform in ECG signal classification because they are unable to handle the intricate temporal dynamics of these signals. Dealing with these challenges demonstrated the requirement for greater automation and scale, prompting the investigation of deep learning methods.

2.2 Deep Learning Models

The advent of deep learning has recently transformed the task by allowing it to extract features automatically through raw ECG signals. CNNs have become widespread applications due to capturing the spatial hierarchies in the data [21-23]. Recent studies have shown that CNNs can detect complex arrhythmias better than conventional approaches by automatically learning discriminative features directly from data [24–26]. Methods A variety of machine learning algorithms and deep neural network models have been proposed for arrhythmia detection or classification of the ECG signals, among which RNNs, especially Long Short-Term Memory (LSTM) networks, were recently introduced to capture temporal dependencies originating from representations in time series data [27–29]. Recently, some hybrid models that combine CNNs and RNNs have been proposed to use spatial features from the leads (through windowing) and sequential information about these networks [30, 31], giving best-in-class results in different ECG classification tasks.

For all their leaps forward in automatic ECG classification, deep learning models are not yet perfect. Some of these models need big annotated datasets for training that might not be easily accessible, especially in certain arrhythmias. Second, deep learning is more likely to overfit small or imbalanced datasets.These models have a certain level of complexity, making them computationally expensive to train and deploy. Therefore, it is necessary to develop more scalable models that can generalize across multiple datasets and are class-balanced agnostic, which has led to the surge in research with Transformers and attention mechanisms.

2.3 Transformers and Attention Mechanisms

Transformers have historically been created for Natural Language Processing (NLP). However, due to the self-attention mechanism that allows the modeling of long-range dependencies, it has recently gained acceptance in ECG classification tasks [32]. This level of parallelism contrasts with RNNs, which process data sequentially and make them significantly more efficient in large datasets[33]. Attention mechanisms facilitate the identification of its highly informative parts in ECG signals and, therefore, help CNNs capture subtle patterns linking with each type of arrhythmia when used as feature extractors [34]. Recent research has confirmed the applicability of pretrained Transformers, such as BERT and GPT, for ECG [35-37].

Challenges Remained: Even the Transformers showed promising results in ECG classification with attention mechanisms. However, these models are often too resource-hungry and hence cannot be deployed in a viable manner to the general/mobile/edge computing devices. Furthermore, despite the increased interpretability of attention mechanisms, these models are often black boxes to many clinicians, and their trust in predictions can be low. This underscores the interest in lightweight interpretable models that could fit various clinical settings.

2.4 Addressing Class Imbalance in ECG Datasets

ECG datasets often have very few samples of certain arrhythmia classes, with models trained on these data being biased towards the majority class and resulting in poor performance for minority classes, also known as an imbalanced dataset. One of the challenges has been dealing with it, and we saw several different strategies emerge. Data-level approaches such as SMOTE [38] and its variants balance the dataset to generate synthetic samples of minority classes. Moreover, Generative GANs have been used to model realistic synthetic ECG signals to enhance the models with under-represented courses [39, 40]. At an algorithmic level, cost-sensitive learning assigns different misclassification costs to each class to discourage the prediction of minority classes more frequently [41]. A modification of cross-entropy loss, adapted to focus the learning algorithm on difficult examples (often those associated with minority classes), has also been used by focal loss [42].

Despite various techniques for addressing class imbalance, challenges remain. Synthetic data generation methods like SMOTE and GANs can produce unrealistic or noisy samples, potentially degrading model performance. Cost-sensitive learning and Focal Loss are effective but require careful hyperparameter tuning, which can be challenging. These methods enhance performance for minority classes but may not ensure overall model robustness across all classes. This highlights the need for an integrated approach that combines data-level and algorithmic techniques with advanced modeling strategies, such as Transformers, to achieve balanced and reliable ECG classification.

2.5 Interpretability and Explainability in Deep Learning

As deep learning models become increasingly complex, their interpretability has become a critical concern, especially in healthcare applications where trust and transparency are paramount [43]. Various methods have been developed to make these models more interpretable. Techniques such as SHAP provide a unified approach to interpreting model predictions by attributing importance scores to each feature in the input data [44].Grad-CAM provides visual explanations showing what parts of the input ECG most affectthe Model's predictions [45]. In ECG classification, layer-wise relevance propa-gation (LRP) [46] and other saliency map techniques for the interpretability of deep neural networks have also been attempted to understand the decision-making process.

While interpretability methods have significantly improved in the past few years in explaining decisions made by deep learning models, they come with some limitations. Most of these techniques are in their infancy, and results may not be clear take-home messages for everyday clinical practice. There is also a trade-off between model accuracy and interpretability (simpler models tend to be more interpretable but less accurate, and vice versa). This suggests the continuing development of more sophisticated interpretabil-ity methods that enable a mixture of informative insights alongside good model performance but also point to what work has yet to be done. Due to this motivation, an ECG classification framework must be developed with a trade-off between accuracy and interpretability for practical deployment in clinical settings.

2.6 Motivation for the Proposed Framework

The drawbacks of the current ECG arrhyth-mia classifica-tion techniques urgently need a more complete and systematic approach. This paper proposes a framework to fill these gaps using pretrained Transformers, advanced attention mechanisms, and methods for classimbalance. The proposed framework tries to alleviate all these issues by integrating these different parts, which aims for a more reliable and comprehensive solution of ECG arrhythmia classification, increasing the accuracy, robustness, and usability of these models in clinical scenarios.

3. Proposed Methodology

3.1 Overview

Design a system for ECG classification and arrhythmia detection using deep learning. This section contains the general idea of our proposed methodology for classifying different types of cardiac arrhythmia, as shown above.It discusses how to use Deep Learning model instances to address limitations inherent with classical models that use traditional algorithms for ECG (Electrocardiography) classifications. To deal with these challenges, our method leverages pretrained GPT models[10, 47], data augmentation usingGANs, andsupervisedContrastive Loss that facilitates pushing apart samples from different classes while pulling together those belonging to the same class. We specifically use interpretability tools such as SHAP and Grad-CAM to give some idea of how the Model makes decisions.

The Framework of ECG Classification in Figure 1 for Arrhythmia upon existing solutions, the framework in Figure 2 is implemented with some technology components devoted to improving accuracy, robustness and interpretability.

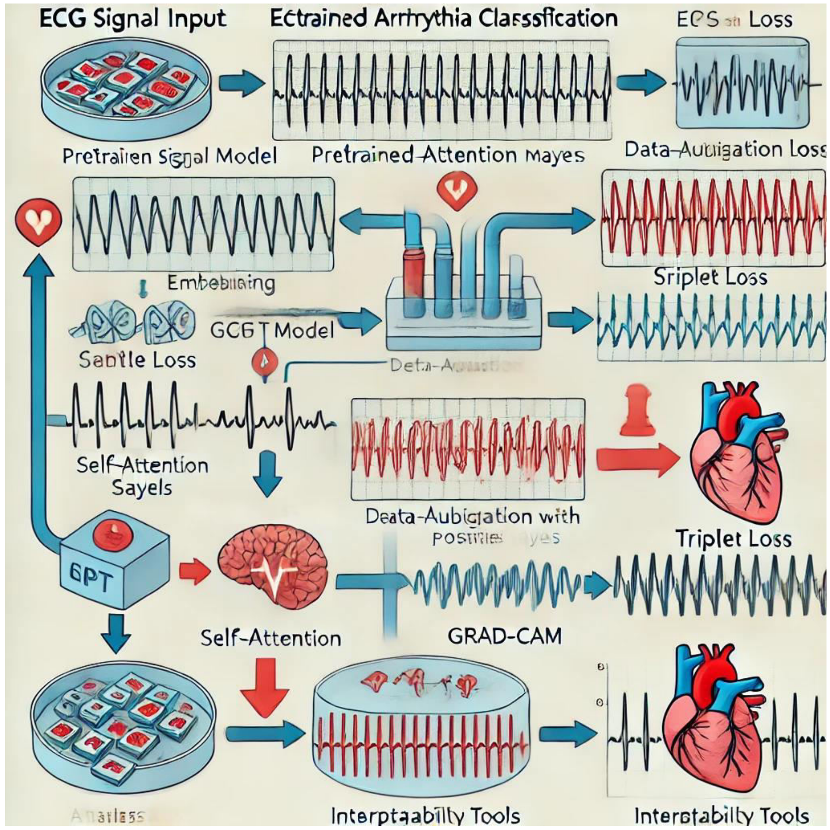

Fig. 1: Proposedsystem for ECG arrhythmia classification.

We incorporate several sophisticated modules into the ECG classification framework to further improve the accuracy and interpretability. Along with the raw ECG signal input, a pretrained GPT model is taken as an embedding layer for transforming signs to continuous vectors and self-attention layers through which temporal and context relationships are understood in relevance to that view of previously available data at each attendance. Lastly it converts its output using asynthetic generating concise representation. DIGA uses GAN-based data augmentation to synthetize ECG, and thereby fix class imbalances while also helping generalization. During training, Triplet Loss is used with anchor, positive and negative samples through which the Model learns how to separate arrhythmias of different classes. We also include interpretability tools like SHAP for quantifying feature contributions and Grad-CAM to highlight what parts of the ECG signals are influential in its decision-making process. These components are integrated into the framework in an end-to-end fashion for input to classification and provide a comprehensive technique that represents a significant step forward in automated medical diagnosis.

3.2 Pretrained GPT Models

Our proposed methodology is rooted in the notion to benefit from Pretrained GPT models for a better understanding of ECG signals, which enables the fine grained temporal pat-terns and contextual relationships captured within them. Figure 2 illustrates how the GPT model can read sequential ECG data and effectively capture rich temporal features needed for accurate arrhythmia classification. This element is beneficial for learning intricate dependencies in ECG signals that standard models would not capture. Taking advantage of the pretrained property in GPT, our approach allows effective knowledge transfer from generic domains to the target ECG classification task despite unreliable intra-domain so that better performance is obtained.

A Diagram of the Application of Pretrained GPT Models to ECG signal input is shown in Figure 2. It is the beginning of data entry from ECG signal to this system. These signals are then processed through an embedding layer, transforming the ECG signals into vector form for further analysis. This is followed by Self-Attention Layers, which enable the Model to attend over time to capture crucial temporal dependencies and contextual interactions. After that, the data goes through a Feed Forward Layer, which helps to learn higher-level features from the extracted information. The processed data finally passes to the Output Layer, in which a high-level feature representation of ECG signals is prepared for further classification process.

The GPT model works by encoding the input ECG signals into a high-dimensional space, where time information and relation-based features are embedded. Let $X = [x_1, x_2, ..., x_n]$ denote the sequence of ECG signal inputs, where each $x_i$represents a time step in the ECG signal.

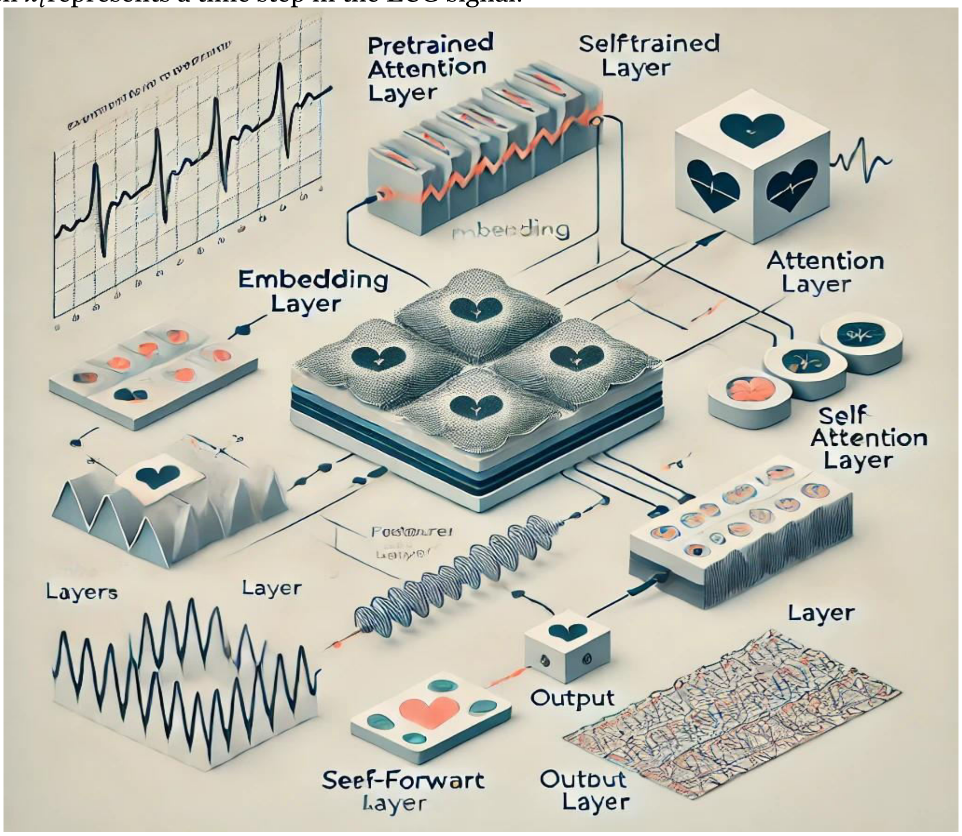

Fig. 2: The architecture of the pretrained GPT component for ECG arrhythmia classification.

Diagram of the ECG Signal Workflow: Going from the Input to Embedding Layer for signal feature mapping into a vector space representation; Self Attention Layers that capture temporal information; and Feed Forward layers that refine this data. A feature-rich representation in the Output Layer ensures that dataflow is stable,providing a process's normalization and residual connections. It first embeds the input sequence to a continuous vector space using embeddings function $E(x)$, which can be mathematically defined as:

$$E(x) = [E(x_1), E(x_2), ..., E(x_n)] \quad (1)$$

From there, these embedded vectors go through a stack of self-attention layers to help the Model attend to different parts of this sequence when making predictions. The self-attention mechanism can be described by the following equation:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

Here, Q is queries, K is keys, and $V$is values are all calculated using a linear transformation from the same sequence of input embedding as in equation1, $d_k$is a key dimension. The self-attention mechanism defined in equation 2 allows to learn how much different steps in a sequence contribute, effectively capturing temporal dependencies. Bypassing the self-attention layers output through as a feed-forward layer, we achieve a final representation of each token Zin the original input sequence:

$$Z = FeedForward(Attention(Q, K, V)) \quad (3)$$

This result Zwill carry useful temporal features exploited by the rest of the components in the framework to detect arrhythmia. It learns the representation between raw ECG data and its feature space, where temporal behaviors are better captured for downstream classification tasks. In our approach, the Pretrained GPT model comprises several layers that cooperate to make sense of sequences within input ECGs. Each layer in the GPT model can be decomposed into:

The essential layers and mechanisms of the GPT ECG classification model, The Embedding Layer, transform the input ECG sequence in a continuous vector space and add position encoding to maintain order over time. Next, you have Self-Attention Layers that computes the attention scores over a sequence, allowing the Model to selectively attend to relevant parts of data for each time step. The processed data then goes through the feed-forward layers that apply nonlinear transformations to fine-tune further feature representation learned from the self-attention mechanism. Normalization and Residual Connections are used after each layer in order to improve the generalization abilities of the network, making training stable with positive gradient flow. The output layer then outputs the final representation encapsulating all temporal and contextual information of the input ECG sequence for further classification. This architecture in Fig. 2 demonstrates how these components extract local and global temporal patterns within ECG data.

3.3 Data Augmentation with GANs

The Pretrained GPT Models outputs additional informed features of the more complex temporal dynamics and contextual relations within ECG signals. Although such output is essential to decipher complex dependencies in the ECG data, exposure toa wider variety of arrhythmia patterns can further strengthen the robustness ofthe classification model specifically from classes appearing sparsely within origi-nal dataset. We do so by utilizing Generative Adversarial Networks (GANs) to expand the dataset with synthetic ECG samples that represent real-world variabilities. This extra synthetic data in combination with the full-featured output of GPT models leads to a more complete training set and, hence, improved generalization of the Model.

Our GAN architecture includes the Generator and Discriminator, as shown in Figure 3; both are neural networks. The Generator (G) takes as input a random noise vector z sampled from a prior distribution b_z(z)and generates synthetic ECG data. The Discriminator (D) then receives both real ECG data x and synthetic data $G(z)$, and its task is to distinguish between the real and synthetic samples.

Figure 3 was designed to highlight the sequential nature of how each component interacts and flows in a game with respect, but not limited only to ECG processing.The process begins with the Generatorcomponent, which takes a random noise vector (z) as input and generates synthetic ECG data. This synthetic data is designed to mimic real ECG patterns. The generated ECG data, along with actual real ECG data (x), is then passed to the Dis-criminatorcomponent. The Discriminator's task is to distinguish between real and synthetic ECG data. As the GAN is trained, the Generator continuously improves its synthetic data generation to better fool the Discriminator, while the Discriminator gets better at identifying whether the data is real or generated. This adversarial process helps the GAN refine its ability to create realistic ECG data, which can be used to enhance training datasets for improved model performance in ECG arrhythmia classification.

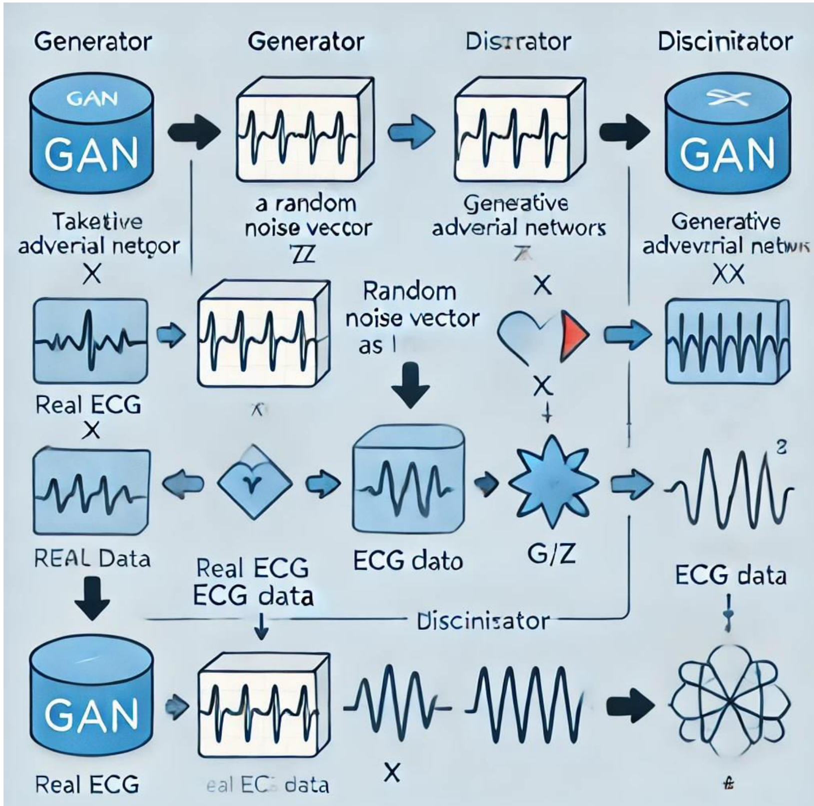

Fig. 3: Simplified diagram illustrating the process of Data Augmentation with GANs for ECG arrhythmia classification.

Here, the Pretrained GPT Model has a robust output feature of ECG which is transmitted to the GANs module as illustrated above. This module contains a generator that generates synthetic ECGs and a discriminator responsible for distinguishing real data from fake data. These artificially created 'fake' data points are then fused with the actual (real) samples to construct an augmented training dataset, strengthening the Model against vulnerability and boosting its generalization tasks.

The diagram uses arrows to show how data and feed-back flow between the Generator and Discriminator. This helps simplify the operations of a GAN.

$$\min_{G} \max_{D} V(D,G) = E_{x \sim P_{data(x)}}[\log D(x)] + E_{x \sim P_{z(z)}}[\log\left(1 - D\left(G(z)\right)\right)] \quad (4)$$

Here $p_{data}(x)$is the distribution of the real ECG data and $p_z(z)$is fed into the Generator before noise. The Generator Gin the GAN is responsible for mapping the input noise z to the synthetic data space, producing samples that resemble real ECG signals. Here is the mapping function:

$$G\left(z;\ \theta_g\right) = x' \quad (5)$$

where $x'$is the set of ECGs and $\theta_g$indicates all parameters of G. The Discriminator, meanwhile, produces a higher output probability $D(x)$of the input$x$ it sees as reals over synthetic ones. The output of the Discriminator is:

$$D(x;\ \theta_d) = P(real|x) \quad (6)$$

In the figure where $\theta_d$denotes parameters of the Discriminator. In the training process, D and G are purposely updated iteratively to become better against each other. As its name suggests, the Generator learns to generate real data with time; whereas the Discriminator becomes more accurate in identifying fake generated images. Optimally, we perform adversarial training to reach a point where the Discriminator cannot distinguish real data from synthetic data,umeaning $D(x) = 0.5$ for both real and generated samples.This is particularly useful because synthetic data from the GAN helps to extend a greater range of arrhythmia patterns into the training set, which are not accurately reflected in an unbalanced source dataset. These augmentations aim to introduce the Model with some noise and other types of images so that our classification task fine-to-the-least maintains its robustness when class imbalances occur aggressively, or there will be unseen data. Adding this variability helps the model to discriminate more easily between different arrhythmias. First, besides being exposed with incomplete samples (especially for rare arrhythmias) in Machine Learning models) we also extend the sample set by adding augmented data from GAN to avoid overfitting and exploit the full capacity of the classification pipeline with an enriched training dataset.

3.4 Triplet Loss for Class Imbalance

We integrate Triplet Loss into our training framework to effectively address class imbalances, particularly for underrepresented classes like rare arrhythmias.Triplet Loss is used to pull ECG samples from the same class closer together in feature space while pushing those of different classes farther apart. Once the Pretrained GPT Model and Data Augmentation process ECG signals using GANs, they undergo a classification layer with Triplet Loss. Image SourceThe Model takes in 3 outputs simultaneously: an anchor sample, a positive sample from the similar class and a negative sample from different classes. We want to close the anchor and positive samples while making them far from the negative ones. This can be observed in Figure 4, where Triplet Loss reduces the intra-class distances between ECG representations while increasing inter-class distance, allowing the Model to learn better feature separators for common and rare arrhythmias. The Triplet Loss function is mathematically defined as:

$$\mathcal{L}_{triplet} = \max(0, ||f(x_{a)}||^2 + \alpha) \quad (7)$$

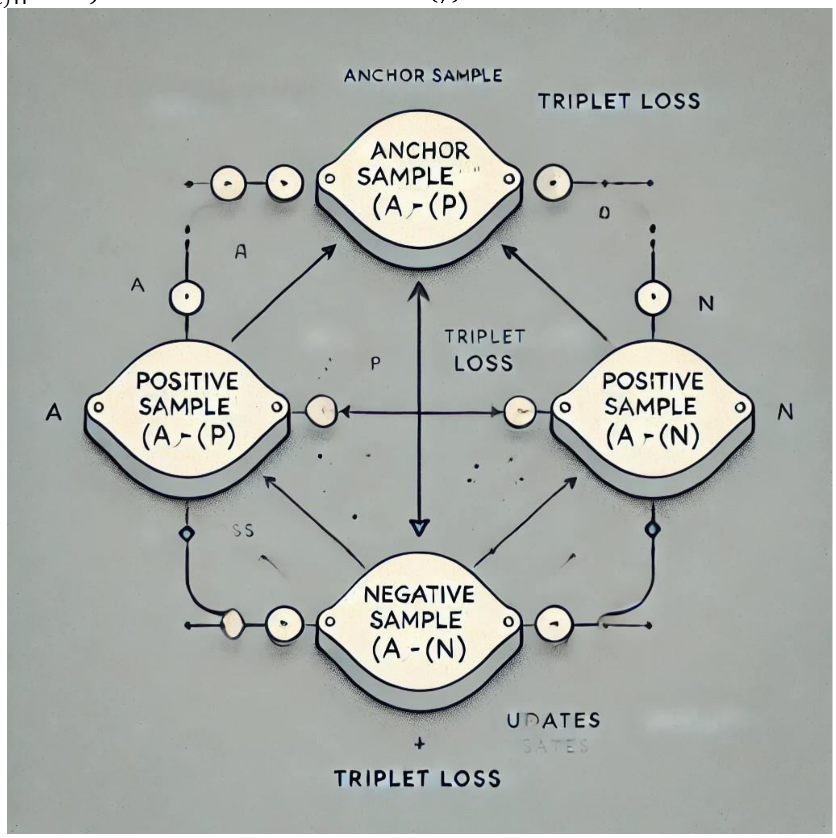

Fig. 4: A simple diagram illustrating the basic steps of how Triplet Loss works.

The relationship of an Anchor sample (A) with a Positive sample (P) of the same class and Negative samples (N) from different classes, for example, as shown in this diagram below. Our aim is to reduce the distance between Anchor and Positive samples while increasing that of Anchor with Negative ones which ultimately results in a better classification accuracy due to more discriminative feature space.

Where $f(x)$in equation(7) denotes the embedding function of our Model that maps an input ECG sampleto a high-dimensional feature space Here,x_a, x_p, and x_n are the anchor, positive and negative samples, respectively. The margin $\alpha$is just a threshold determining how far should the negative sample be compared to the positive against the anchor, where$\|\cdot\ ||^2$is the squared Euclidean distance between feature vectors. Triplet loss is designed so that the distance between the anchor and positive sample is minimized while at least a margin α away from negative samples. If the difference between them is more significant than this margin, the loss will be 0, meaning we have a well-separated triplet. In contrast to conventional loss functions such as cross-entropy, which consider all classes at par and do not work well with the minority set of examples, Triplet Loss results in a more discriminative feature space. This optimizes the Model to separate similar and dissimilar samples in their respective inter-sample distances, an advantage when handling class imbalances within ECG arrhythmia classification. The algorithm generates a feature space in which samples from the same class are gathered together, and different classes of objects stay separately so it can perform better classification tasks such as detecting rare arrhythmia.

3.5 Model Interpretability with SHAP and Grad-CAM

To make our models more transparent and easier to understand, we use two powerful interpretability tools: SHAP and Grad-CAM. These tools are essential to prove the model's predictions and evaluate whether a classification mechanism matches clinical intuition.

3.5.1 SHAP

This is a game-theoretic approach to explain the output of any machine learning model. SHAP then computes for every input ECG signal it is fed to the Model how much each feature (e.g., heart rate variability, specific components of an ECG waveform) plays a role in ultimately obtaining that classification. The sum of these contributions is the difference between what was output by the Model and what an overall average production would be. This means that with SHAP you can see why the Model made a specific decision and what feature(s) were driving it. SHAP value is calculated as:

$$SHAP\ value\ for\ feature\ i = \sum_{S\subseteq F\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\ [f(S\cup\{i\}-f(S)] \quad (8)$$

Where$F$ is the set of all features, S is a subset of Fnot including feature i in it, and $f(S)$is the model output when only featuresin subset $S$are active. This equation represents the change in model prediction when a new feature is added to that taken over all other possible subsets of features.

3.5.2 Grad-CAM

In contrast, Grad-CAM (2) is oriented in a more visual perspective generating Heatmaps that indicate which regions of ECG signal are the most important for model decision. More precisely, Grad-CAM exploits gradients of the Model's output concerning the final convolutional layer to produce a localization map overlaid on input ECG signals. The map shows which areas of the ECG were most important in determining if a patient had an arrhythmia, giving clinicians insight into what aspects of trace data led to the Model predicting any one specific irregular rhythm.

Grad-CAM starts by calculating the gradient of a score for class c concerning feature maps last convolutional layer. Then, these gradients are aggregated to obtain the importance weights of each feature map. Once the weights have been determined, you take a weighted sum of these weights with their associated feature maps. After that, we apply the ReLU function to get the Heatmap. The resulting Heatmap overlaps the ECG signal, highlighting portions of the ECG that the Model finds most important when making a prediction.

3.5.3 Integration and Impact

Sequential steps of using SHAP and Grad-CAM for model interpretability in ECG arrhythmia classificationare shown in Figure 5. For the process, ECG ones come under input, in which raw data of ECG is given. Next, in Step 2, the Model takes this input and provides a model witha prediction as output — predicting perhaps what kind of arrhyth-mia is present. Moving to Step 3, the SHAP Explanation explains details of this model prediction — it tells us how much each feature has contributed towards changing the average output from one value to another and gives significant insights regarding how a particular feature affects your predicted outcome. Grad-CAM Visu-alization overlays a heatmap on the ECG waveform to identify parts with the highest impact upon which it bases its decision, thus completing Step 4. This elaborate method guarantees that the predictions of our model are not only precise but also interpretable, which helps clinicians understand and believe the results.

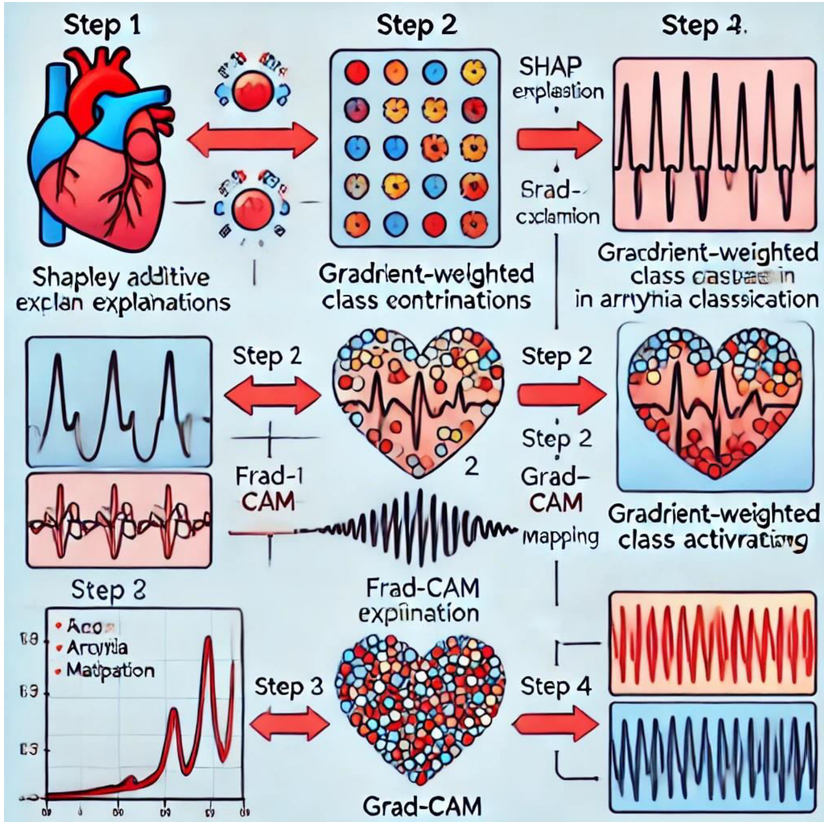

Fig. 5: Diagram illustrating the use of SHAP and Grad-CAM for model interpretability in ECG arrhythmia classification.

The process starts with ECG Input, which is processed in the Model to predict something. SHAP elucidates the prediction by detailing the contribution of each feature, and Grad-CAM offers a visual heatmap to accentuate regions in an ECG that most impacted the model decision. Combined, they help validate the Model's predictions to meet clinical expectations.

Such interpretability tools are crucial to bridge the gap between complex machine learning models and their real-world clinical applications. The added value of SHAP and Grad-CAM lies in providing insights into the decision-making process of a given model, empowering clini-cians to understand and validate those decisions by making predictions more transparent while aligning them with clinical expectations during reading/interpretation workflow for classification outcomes.

We have evaluated the performance of our proposed methodology on extensive experiments over multiple real-world ECG-arrhythmia datasets. Our primary goal with these experiments was to assess the efficacy of our end-to-end approach consisting of pretrained GPT models, data augmentation via GANs,and Triplet Loss classification regardingoverall accuracy and robustness, especially when identifying rare arrhythmias.

4. Experiment Setup

4.1.1 Datasets

Null To evaluate the performance of our proposed system, we performed experiments on two popular ECG arrhythmia datasets, namely MIT-BIH Arrhythmia Database [48] and PTB Diagnostic ECG Database [49]. The diversity in the arrhythmia types of these datasets and their wide application within the research community characterize them as ideal for assessing our method's robustness and generalization [31].

4.1.2 MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database [44],a landmark in the field of arrhythmias detection research) has 48 half-hour ECG recordings from 47 subjects. These recordings of more than 110,000 annotated beats are organized in 16 arrhythmia classes, including standard ECG patterns and select examples of premature ventricular contractions (PVCs)and atrial fibrillation (AF). This dataset has a sampling rate of 360 Hz. Itis digitized in an amplitude range of 11-bit resolutionwhere ECG features are well represented to be used both for training or validation purposes on ECG classification systems. Our study particularly aimed at the specific arrhythmia categories such as Normal Sinus

Rhythm (N), Supraventricular Pre-mature or Ectopic Beat (S), Ventricular Premature or Ectopic Beat (V), Fusion of Ventricular, Narmal beat, and Unknown Beats(Q) [31].

4.1.3 PTB Diagnostic ECG Database

PTB has one of the most popular ECG databases [49], which is the PTB Diagnostic ECG Database collected at University Hospital Bonn, Ger-many contains 5,388 recordings from 549 patients. All were recorded with a 12-lead ECG system at a sampling rate of 1,000 Hz and included myocardial infarction and various forms of cardiac hypertrophy. Each recording is associated with diagnostic labels provided by cardiology experts. For the purpose of this study, we focused on two primary classes: Normal (N) and Myocardial Infarction (M). Table1Offers a detailed overview of the sample counts for each class within both datasets.

Table 1: Comprehensive Overview of the Utilized Datasets

Database	**MIT-BIH Arrhythmia**						**PTB Diagnostic ECG**		
Classes	N	S	V	F	Q	Total	N	M	Total
Training	72,471	2,223	5,788	641	6,431	87,554	3,236	8,400	11,636
Testing	18,118	556	1,448	162	1,608	21,892	809	2,100	2,909

4.1.4 Datasets Preprocessing:

Preprocessing of ECG signals is crucial, as these signals form the primary input for our system. To optimize the effectiveness of our approach, we implemented the following preprocessing steps for ECG signals and the subsequent extraction of cardiac beats, as outlined in [50]. The procedure includes:

1. **ECG Signal Partitioning**: The continuous ECG signal is first divided into 10-second segments, with a particular 10-second segment selected for subsequent processing.
2. **Amplitude Scaling**: The amplitude within the selected segment is rescaled to lie between zero and one, ensuring a consistent representation of the ECG signal.
3. **Detection of Local Peaks**: Local peaks are identified by analyzing where the first derivative of the signal crosses zero.
4. **R-Peak Identification**: A crucial part of the process is the identification of R-peak candidates, which is done by setting a threshold of 0.9 on the normalized local peak values.
5. **Estimation of Typical Heartbeat Period**: The median value of the R-R intervals is calculated to estimate the typical heartbeat period within the selected 10-second segment.
6. *Selection of Signal Segments:* For each identified R-peak, a segment of the signal, measuring 1.2 times the estimated heartbeat period (1.2T), is selected.
7. **Zero Padding**: Each chosen signal segment is padded with zeros to achieve a uniform and predetermined length.

4.2 Implementation Details

Our proposed system employs several advanced techniques with specific configurations to optimize ECG classification performance. It utilizes a fine-tuned GPT-2 model with 12 transformer layers, 768 hidden units, and 12 attention heads, processing sequences up to 1024 tokens. Training is conducted with a learning rate 5e-5, the AdamW optimizer with 0.01 weight decay, and a batch size 16 over 3 epochs, effectively capturing complex temporal patterns. For data augmentation, a GAN architecture with a generator and Discriminator, each featuring 4 convolutional layers with LeakyReLU activation functions, is used to address class imbalance by generating diverse synthetic ECG samples. The GAN is trained for 200 epochs with a latent vector size of 100 and a batch size of 64. Triplet Loss, configured with a margin of 0.2, uses semi-hard negative mining to optimize the feature space by clustering similar ECG patterns and distancing dissimilar ones. This approach, with an embedding dimension of 256, is trained over 50 epochs using an Adam optimizer (learning rate 1e-4) and a batch size of 32. To enhance model transparency, SHAP calculates feature importance with 100 background examples and 1000 samples, while Grad-CAM generates Heatmaps (224x224 resolution, 0.5 opacity) to visualize influential regions of ECG signals. The training procedure involves partitioning the dataset into 80% for training and 20% for testing, with 20% of the training data reserved for validation, and conducting training over 60 epochs with the Adam optimizer (learning rate 1e-3, weight decay 1e-4) and a batch size of 128, ensuring model convergence and stability.

4.3 Evaluation Metrics

To rigorously evaluate our Model's performance, we utilized a set of key metrics:

• **Accuracy (ACC):** Measures the proportion of correctly classified instances out of the total number of cases. The accuracy is given by:

$$\mathbf{ACC} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

Macro Average F1-Score (MF1): Provides a balance between precision and Recall, particularly valuable for imbalanced datasets. The Macro Average F1-Score is calculated as the mean of $F1$-scores for each class:

$$MF1 = \frac{1}{n}\sum_{i=1}^{n} F1_i \quad (10)$$

where $n$is the number of classes and $F1_i$is the F1-score of the $i-th$ class.

• **Precision:** Reflects the accuracy of positive predictions made by the Model. It is defined as:

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Recall: Indicates the Model's ability to identify all relevant instances. The Recall is calculated as:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

5. Results and Analysis

In the results section, we showcase the findings from multiple experiments that evaluate our Model's performance across various configurations.

5.1 Performance Analysis: In-depth Examination

The proposed system achieved high classification performance on the typical class of ECG arrhythmias, especially for those handling the most prevalent and well-defined classes. The Model obtains almost perfect precision-recall and F1scores on Normal Sinus Rhythm and Unknown Beat (Table 2), scoring more than 99% in these metrics. The specificity and sensitivity values of the proposed Model show that it can differentiate normal from abnormal heart rhythms with very few errors, which means this Model is exceptionally robust for clinical applications. The ability of GPT to generate data and fine-grain control over what can be generated from inference through training on similar but more limited synthetic datasets expands the dataset space and speculatively allows for more efficient memorization. The high macro-averaged preci-sion of 96.16% and Recall of 93.62 support the discrimina-tory ability for various arrhythmia types with good consistency across different classes.

Additionally, the proposed system demonstrates its power even in challenging scenarios such as class imbalances and subtlearrhythmia variations. When Triplet Loss is included, the Model focuses on relative differences between classes, leading to significant class imbalance improvement for underrepresented arrhythmias. While the performance for more complex classes like Supraventricular Premature Beat (S) and Fusion of Ventricular and Normal Beat (F) is slightly lower, the Model still maintains strong Fl-scores, as indicated in Table2, demonstrating its capability to handle complex classification tasks. The overall robustness of the proposed system, combined with its high interpretability through tools like SHAP and Grad-CAM, makes it a powerful tool for accurate and transparent ECG arrhythmia detection in both clinical and research settings.

Table 2: Performance metric evaluation across different classes in the MIT-BIH dataset using the proposed system.

Metric	**N**	**S**	**V**	**F**	**Q**	**Macro-Avg**
Precision	99.48% ± 0.08	92.45% ± 1.95	96.32% ± 1.35	92.12% ± 5.10	99.41% ± 0.24	96.16% ± 1.45
Recall	99.54% ± 0.23	87.75% ± 2.20	98.12% ± 0.15	83.45% ± 3.62	99.22% ± 0.03	93.62% ± 1.05
F1-Score	99.51% ± 0.07	90.03% ± 0.20	97.40% ± 0.68	87.65% ± 1.15	99.30% ± 0.13	94.78% ± 0.26

The evaluation of the proposed system on the PTB Diagnostic ECG dataset, as shown in Table 3, highlights its exceptional capability in distinguishing between different classes of ECG signals.It attains excellent precision scores of 99.10% for Normal (N) and an even higher score larger than 99.82% in the case of Myocardial Infarction, so the Model makes very few mistakes on predictions here. The high precision of the model also highlights its ability to detect true positives with few false positives, which is a vital component for clinical use as this directly determines patient outcomes. Macro-averaged precision 99.46% also suggested the performance froth of this Model that pushed average recognition to over between classes; however,it's another fact shows how well it is going with all other classes data, so calling final micro F1-score would be very difficult as both sounds equally goodo

As the Model is trained on each feature of ECG and in terms of Recall, it does very well, with 99.43% for Normal and 99.67% for Myocardial Infarction. The extremely high recall values mean that almost all authentic cases of other conditions are contracted by the Model, and very few actual patients who are positive for one condition are wrongly predicted as having another.

The high F1-scores of 99.27% for Normal and 99.74% for Myocardial Infarction, with a macro, averaged F1-score of 99.51%, balances both precision and recall appropriately, thus covering every type in the positive class. This trade-off is crucial for diagnostic medicine, where detecting true positives and minimizing false negatives are equally important. The results shown in Table 3 show that this system can classify ECG signals with better performance than state-of-the-art models, especially for classes related to critical conditions such as myocardial infarction. It might be an interesting tool capable of safely deploying in the clinical environment.

Table 3: Performance metric evaluation across different classes in the PTB Diagnostic ECG dataset using the proposed system

Metric	**N**	**M**	**Macro-Avg**
Precision	99.10% ± 0.07	99.82% ± 0.05	99.46% ± 0.03
Recall	99.43% ± 0.15	99.67% ± 0.02	99.55% ± 0.07
F1-Score	99.27% ± 0.07	99.74% ± 0.02	99.51% ± 0.04

5.2 Comparative Study: Benchmarking Against Established Baseline Models

The comparison of different methods on the MIT-BIH dataset, as presented in Table4, highlights the superior performance of theproposed system, which integrates GPT, GANs, and Triplet Loss, demonstrates the highest accuracy (ACC) and macro-average Fl-score (MF1) among all evaluated methods. Specifically, the proposed system achieves an ACC of 99.50% and an MF1 of 95.10%, outperforming the previously best-performing method, ECGTransForm, which achieved 99.35% ACC and 94.26% MF1.

This minimal yet prominent increase in those performance metrics highlights the potential of our Model for solving real-world arrhythmia detection tasks, given the complexity and diversity of MIT-BIH data.

GPT can be flexibly applied to better capture the intricate temporal dependencies and contextual relation-ships within ECG signals, leading to performance improvements as those obtained by our Proposed system. Due to the limitations of traditional methods such as ResNet or biLSTM regarding long-range dependencies and rich contextual cue parsing, GPTs transformer structure can handle sequenced data relatively well. This ability gives the proposed system an edge over other models when it comes to the exact classification of not just those common arrhythmias but some obscure and lesser-common ones that usually go amiss. Table 4 shows that incorporating GPT improves F1scores with better accuracy and broadens its range in comprehending ECGplots.

Another important reason is the synergetic effect of GAN-based data augmentation and Triplet Loss in the proposed system. Alleviation of class imbalance: GANs generate realistic synthetic ECG samples so the machine learning model can train well across various arrhythmias. At the same time, Triplet Loss is responsible for cleaning up our feature space so that similar ECG patterns are tightly grouped while dissimilar ones lie far apart. The method increases the discrimina-tive ability of our Model, particularly in separating classes that are close to each other, and yields a record high MF1 for the comparison. The proposed system,combined with these advanced tech-niques, provides a robust and efficient solution for ECG classifica-tion, as summarized in Table 4.

Table 4: Comparison of our proposed framework against baseline methods on the MIT-BIH dataset

Existing Models	**Method**	**ACC (%)**	**MF1 (%)**
[23]	ResNet + LSTM + GA	98.00	89.70
[25]	AE + DNN	99.34	91.44
[26]	ResNet + SE block + biLSTM	99.20	91.69
[28]	Seq2Seq + CRNN	97.60	89.00
[29]	DLA + CLSTM	88.76	80.54
[30]	AE + Transformer	97.66	N/A
[31]	MSC + CRM + BiTrans + CAL	99.35	94.26
Proposed System	**GPT + GAN + Triplet Loss**	**99.50**	**95.10**

5.3 Component Analysis: Ablation Study and Comprehensive Insights

Consequently, Figure 6 shows our improvement relative to the Macro F1-score (MF1) concerning each key compo-nent integration called GPT, GANs, and Triplet Loss. The simplest CNN model, which results in an accuracy of 87.00%, is named the baseline performance, and it establishes a floor on how many improvements we could get by going beyond this machine learning or deep learning techniques.

The most significant increase (compared to other types of improvements) comes in when GPT is integrated into the Model, bumping up the MF1 score from 87.00% →90.50%. The improvement, a 3.5% boost, is based on the success of GPT in cap-tur-ing and mod-el-ing com-plex tem-po-ral depen-den-cies and contextual rela-tionships embedded inside ECG data. Given that ECG signals, especially those corresponding to...arrhythmias, involve complex timing patterns that are not discernible from the waveforms, GPT helps in learning these subtle differences by functioning as an improved sequence processor for ulterior tasks.

The final significant improvement, which takes advantage of Generative Adversarial Networks (GANs), pushes the MF1 score up even more to 93.20%. GANs help generate artificial ECG samples, which is specifically handy when dealing with the class imbalance that most medical datasets face. The result is that the training set now contains even more synthetic data, increasing its variance and enabling ML algorithms to recognize other arrhythmias that are not adequately represented. We note that a 2.7% gain in the MF1 score is substantial and reflects the importance of data augmentation for robustness and adaptability, which are required properties of such a system before being deployed into practice.

The MF1 score reaches 94.30% from the extended formulation proposed before due to additional fine-tuning of the feature space using it. Triplet Loss is very good at optimizing distances between distributions of our data points, thus enabling better generalization in discerning various arrhythmias. This 1.1% improvement emphasizes the need for feature space management, especially in scenarios with imbalanced classes since accurate classification between arrhythmias that are tough to distinguish from one another is critical.

With all components integrated, the final proposed systemachieves an impressive MF1 score of 95.10%. This cumulative improvement reflects the synergistic effect of combining GPT's deep temporal understanding, GANs' data augmentation, and Triplet Loss's feature space optimization. Each component plays a distinct role in enhancing the Model's performance. This results in a highly effective and robust ECG arrhythmia classification system that outperforms the baseline and demonstrates substantial improvements with each added component. Figure 6 showcases the stepwise gains in performance, validating the proposed system design's effectiveness and each component's contributions.

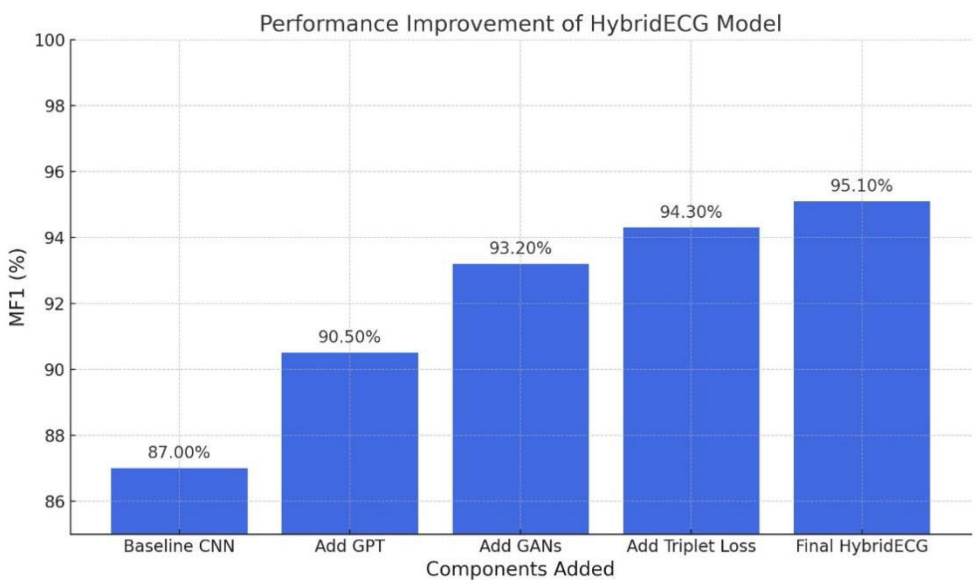

Fig. 6: Performance Improvement of the proposed system. The bar chart illustrates the MF1 scores (%) achieved after adding each component, highlighting the incremental benefits of GPT, GANs, and Triplet Loss.

5.4 Addressing Class Imbalance: Evaluating Triplet Loss

In the radar chart provided in Figure7, some classes, such as F (Fusion of Ventricular and Normal Beat) and S (Supraventricular Premature Beat), exhibit relatively lower performance across all imbalance handling techniques, including the Triplet Loss. The challenges associated with class F stem from its ambiguous nature, as it represents a mix of normal and abnormal beats. This overlap in features between normal and ventricular beats makes it difficult for the Model to distinguish it from other classes, leading to confusion during classification. Consequently, even with advanced techniques like Triplet Loss, the Model struggles to create distinct boundaries in the feature space, resulting in decreased classification accuracy for this class.

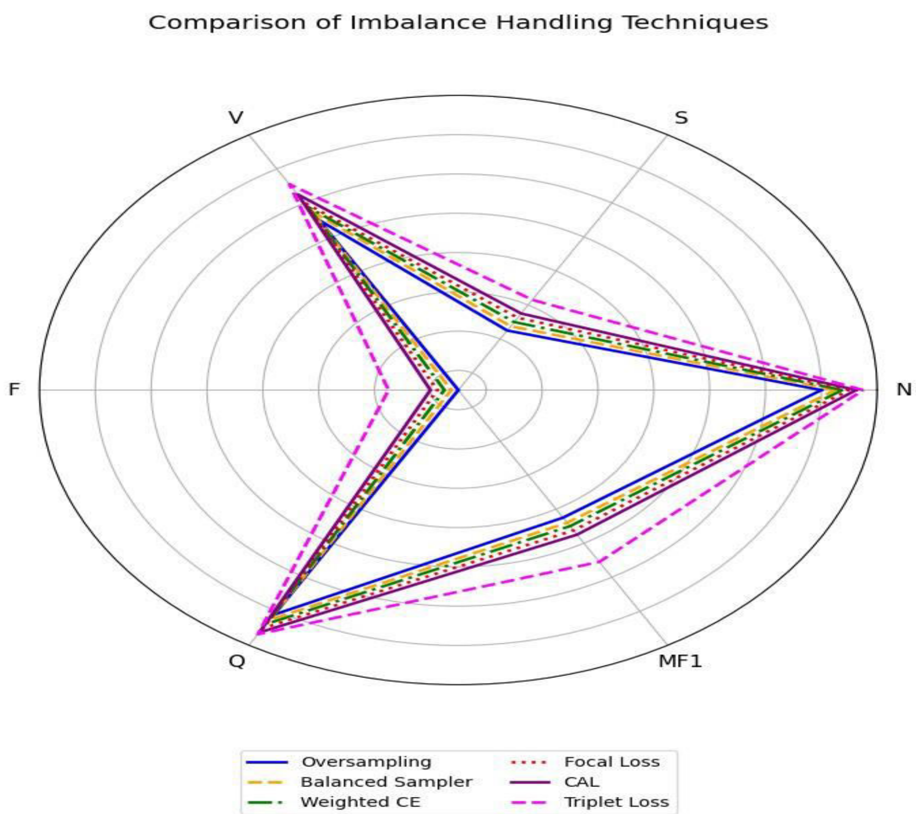

Fig. 7: Comparison of Imbalance Handling Techniques, including Triplet Loss, CAL, Oversampling, Balanced Sampler, Weighted CE, and Focal Loss across various ECG classes.

Class S shows lower performance due to its limited representation in the dataset, which results in fewer training examples and exacerbates the class imbalance problem. This scarcity makes it challenging for the Model to learn enough patterns to classify these rare and subtle arrhythmias accurately.While the introduction of Triplet Loss did increase model performance by addressing class imbalances, discrimination between rare arrhythmias such as S is still more challenging relative to more common classes like N (Normal Sinus Rhythm). These challenges reveal the intricate nature of ECG classification and hint at a requirement for customized loss functions to improve model efficacy. Otherwise, Triplet Loss proved to be the most successful in multi-class and maintained class balance performance (even if not excellent), thus showing how it can enhance classification accuracy, especially on imbalanced datasets.

5.5 Interpretability and Clinical Relevance

In order to have a confident model that predicts accurately and can interpret the reason behind it, we integrated two tools, SHAP & Grad-CAM. SHAP values provide a way to evaluate the impact of each feature on model predictions and render feature importance from a global perspective. This is understandable, as shown in Fig. 8, as the QRS complex coupled with RR Interval and Heart Rate Variability are the essential features that agree with a prior clinical understanding of meaningful arrhythmia indices. As shown in Figure 9, Grad-CAM generates Heatmaps to localize the representative areas from ECG signals, which impact model decisions. Brighter colors on the Heatmap represent regions more prone to denoting a relevant factor such as this P-wave, QRS complex,or T-wave. When we overlay the Heatmap on top of our original ECG signal, it is evident that these further include clinically relevant features — such as the QRS complex in ventricular arrhythmias — which make them especially effective for feature attribution. Combining SHAP and Grad-CAM yields a complete interpretable framework where SHAP provides quantitative feature importance, while Grd-CAM shows visual salient signal regions. This dualism facilitates model interpretability and trust, thus increasing clinical alignment of the model's predictions and downstream integration in healthcare pipelines.

6. Discussion

In this section, we discuss the outcomes of our findings, which point towards performance analysis of the proposed system and its close relevance to state-of-the-art ECG arrhythmia classification. The proposed system combines GPT models and GANs for data augmentation and triplet loss to tackle class imbalances. Such combination has resulted in a substantial increase in classification accuracy, especially in recognizing rare and very complex arrhythmias, as evidenced by the performance metrics.

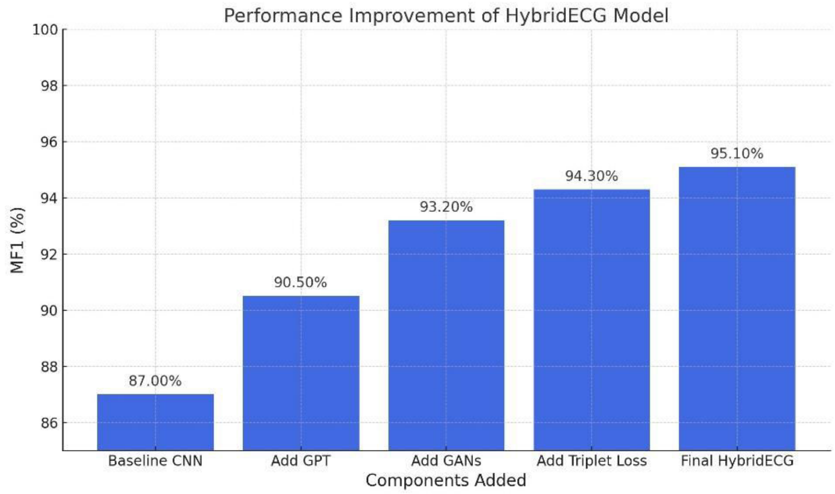

Fig. 8:SHAP Value Distribution for ECG Features This bar chart represents the mean effect of top vital features related to our Model's decision on Eg1-QRS, and the RR interval shows the highest impact.

The performance improvement model for integrating GPT in the proposed system is evident. Therefore, it demonstrates its valuable ability for more accurate arrhythmia pattern classifications, where traditional models struggle to capture the intricate temporal dependencies and contextual relationships manifesting on ECG signals. The GPT's capability to model long-range dependencies has led to noticeable improvements in performance, as shown in the comparison with baseline methods (Table 4). Additionally, using GANs for data augmentation has been instrumental in addressing class imbalance by generating synthetic ECG samples, which exposes the Model to a broader variety of arrhythmia patterns and improves its generalization and robustness, as indicated by improvements in Fl-score and MF1 metrics. The integration of Triplet Loss further refines the feature space, clustering similar ECG patterns together and distancing dissimilar ones, which enhances the Model's ability to differentiate between arrhythmias, particularly those that are underrepresented. The superior performance of the proposed system with Triplet Loss is evident from the radar chart (Figure 7), highlighting its effectiveness compared to other loss functions.

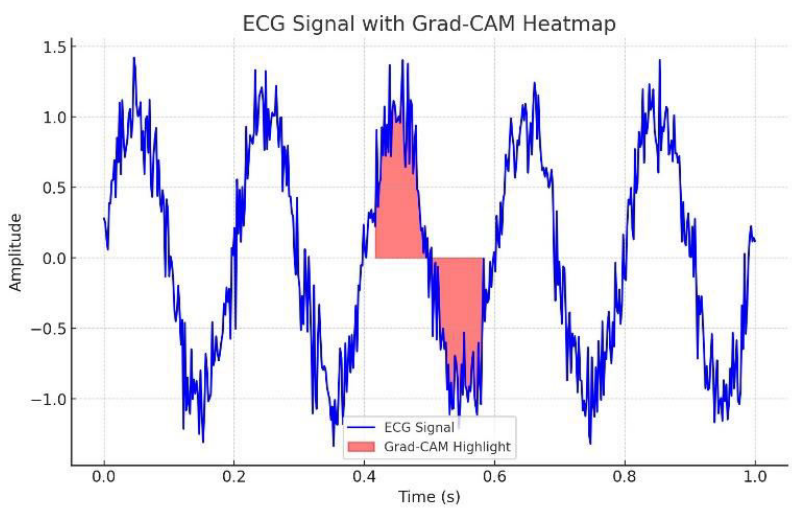

Fig. 9: Grad-CAM Heatmap Overlay on ECG Signal: The Heatmap highlights the regions of the ECG signal that the proposed system identified as most influential in its classification decisions. Warmer colors indicate higher importance.

Using SHAP and Grad-CAM values, we also improved the model interpretations. These tools offer insight into our model's decision, ensure transparency, and help with clinical validation.

6.1 Limitations and Future Work

Though the classification further increases the efficiency in ECG arrhythmia classification, there are still drawbacks. How well the Model works relies on how good and representative your training datasets are, which can arouse concerns about generalizing ability to other populations that may have utterly unseen arrhythmia types. In the future, we will continue working on more complex datasets and introduce real noise to improve robustness.

Using components such as GPT adds a part of this computational complexity. GANs and Triplet Loss are beneficial for their level of accuracy but may restrict these models from being deployed in real-time due to resource limitations. This work can be extended in the future by exploring optimization techniques and hardware acceleration to balance high accuracy with high efficiency. While SHAP and Grad-CAM help understand the model decisions, interpretability is still vital for clinical trust. Improving these tools to derive a more natural approach by embracing the feedback from their user base will be crucial for adoption in clinical practices.

While the Model improves the detection of rare arrhythmias, challenges remain in distinguishing similar classes and detecting infrequent conditions. Future efforts should include advanced data augmentation, transfer learning, and the integration of domain-specific knowledge to address these challenges and further enhance the Model's capabilities.

7. Conclusion

This paper presentsthe proposed system, a cutting-edge deep learning framework for ECG arrhythmia classification that integrates GPT, GANs, and Triplet Loss. The proposed system significantly enhances classification accuracy by harnessing GPT's ability to capture intricate temporal dependencies, GANs' data augmentation to tackle class imbalances, and Triplet Loss to refine feature space, particularly for rare and complex arrhythmias. Our extensive evaluations of the MIT-BIH and PTB Diagnostic ECG datasets demonstrate that the proposed system outperforms existing state-of-the-art methods, achieving superior accuracy and macro-FI scores. Furthermore, integrating interpretability tools like SHAP and Grad-CAM ensures transparency in model predictions, making the proposed system highly effective and reliable for clinical applications. This work underscores the potential of combining advanced AI techniques to improve healthcare diagnostic accuracy and sets the stage for future research in extending these methods to other biomedical signal classification tasks.

References

[1] Aggarwal, Geetika, and Yang Wei. "Non-invasive fetal electrocardiogram monitoring techniques: Potential and future research opportunities in smart textiles." *Signals* 2, no. 3 (2021): 392-412.

[2] Fu, Du-guan. "Cardiac arrhythmias: diagnosis, symptoms, and treatments." *Cell biochemistry and biophysics* 73, no. 2 (2015): 291-296.

[3] Merdjanovska, Elena, and Aleksandra Rashkovska. "Comprehensive survey of computational ECG analysis: Databases, methods and applications." *Expert Systems with Applications* 203 (2022): 117206.

[4] Ansari, Yaqoob, Omar Mourad, Khalid Qaraqe, and Erchin Serpedin. "Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023." *Frontiers in Physiology* 14 (2023): 1246746.

[5] Ebrahimi, Zahra, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi. "A review on deep learning methods for ECG arrhythmia classification." *Expert Systems with Applications: X* 7 (2020): 100033.

[6] Din, Sadia, Marwa Qaraqe, Omar Mourad, Khalid Qaraqe, and Erchin Serpedin. "ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial–temporal and long-range dependency features." *Artificial Intelligence in Medicine* 150 (2024): 102818.

[7] El-Ghaish, Hany, and Emadeldeen Eldele. "ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer." *Biomedical Signal Processing and Control* 89 (2024): 105714.

[8] Patwardhan, Narendra, Stefano Marrone, and Carlo Sansone. "Transformers in the real world: A survey on nlp applications." *Information* 14, no. 4 (2023): 242.

[9] Shi, Dapai, Jingyuan Zhao, Zhenghong Wang, Heng Zhao, Junbin Wang, Yubo Lian, and Andrew F. Burke. "Spatial-temporal self-attention transformer networks for battery state of charge estimation." *Electronics* 12, no. 12 (2023): 2598.

[10] Nagamani, T., S. Jaikanth, S. Jayakumar, and V. Manikandan. "Enhancing Diagnostic Accuracy in Echocardiography with Deep Learning Techniques." In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 941-948. IEEE, 2023.

[11] Chen, Kanghao, Weixian Lei, Shen Zhao, Wei-Shi Zheng, and Ruixuan Wang. "PCCT: Progressive class-center triplet loss for imbalanced medical image classification." *IEEE Journal of Biomedical and Health Informatics* 27, no. 4 (2023): 2026-2036.

[12] Chen, Tsai-Min, Yuan-Hong Tsai, Huan-Hsin Tseng, Kai-Chun Liu, Jhih-Yu Chen, Chih-Han Huang, Guo-Yuan Li, Chun-Yen Shen, and Yu Tsao. "Srecg: Ecg signal super-resolution framework for portable/wearable devices in cardiac arrhythmias classification." *IEEE Transactions on Consumer Electronics* 69, no. 3 (2023): 250-260.

[13] Quach, Luyl-Da, Khang Quoc Nguyen, Anh Quynh Nguyen, Nguyen Thai-Nghe, and Tri Gia Nguyen. "Explainable Deep Learning Models With Gradient-Weighted Class Activation Mapping for Smart Agriculture." *IEEE Access* 11, no. August (2023): 83752-83762.

[14] Dash, Ankan, Junyi Ye, and Guiling Wang. "A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: from medical to remote sensing." *IEEE Access* (2023).

[15] Mandal, Saurav, Pulak Mondal, and Anisha Halder Roy. "Detection of Ventricular Arrhythmia by using Heart rate variability signal and ECG beat image." *Biomedical Signal Processing and Control* 68 (2021): 102692.

[16] Mandal, Saurav, and Nabanita Sinha. "Prediction of atrial fibrillation based on nonlinear modeling of heart rate variability signal and SVM classifier." *Research on Biomedical Engineering* 37 (2021): 725-736.

[17] Hirsch, Gerald, Søren H. Jensen, Erik S. Poulsen, and Sadasivan Puthusserypady. "Atrial fibrillation detection using heart rate variability and atrial activity: A hybrid approach." *Expert Systems with Applications* 169 (2021): 114452.

[18] Khairuddin, Adam Mohd, and K. N. F. Ku Azir. "Using the HAAR wavelet transform and K-nearest neighbour algorithm to improve ECG detection and classification of arrhythmia." In *Advances and Trends in Artificial Intelligence. From Theory to Practice: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part II 34*, pp. 310-322. Springer International Publishing, 2021..

[19] Sinha, Nabanita, Rajesh Kumar Tripathy, and Arpita Das. "ECG beat classification based on discriminative multilevel feature analysis and deep learning approach." *Biomedical Signal Processing and Control* 78 (2022): 103943.

[20] Gajowniczek, Krzysztof, Iga Grzegorczyk, Tomasz Ząbkowski, and Chandrajit Bajaj. "Weighted random forests to improve arrhythmia classification." *Electronics* 9, no. 1 (2020): 99.

[21] Eltrass, Ahmed S., Mazhar B. Tayel, and Abeer I. Ammar. "Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures." *Neural Computing and Applications* 34, no. 11 (2022): 8755-8775.

[22] Islam, Md Shofiqul, Khondokar Fida Hasan, Sunjida Sultana, Shahadat Uddin, Julian MW Quinn, and Mohammad Ali Moni. "HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN." *Neural Networks* 162 (2023): 271-287.

[23] M. Hammad, A.M. Iliyasu, A. Subasi, E.S.L. Ho, A.A.A. El-Latif, A multitier deep learning model for Arrhythmia detection, IEEE Trans. Instrum. Meas. 70 (2021)1–9.

[24] Degirmenci, Murside, Mehmet Akif Ozdemir, Elif Izci, and Aydin Akan. "Arrhythmic heartbeat classification using 2d convolutional neural networks." *Irbm* 43, no. 5 (2022): 422-433.

[25] S. Nurmaini, A. Darmawahyuni, A.N. Sakti Mukti, M.N. Rachmatullah, F. Firdaus, B. Tutuko, Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification, Electronics 9 (1) (2020) 135.

[26] Y.K. Kim, M. Lee, H.S. Song, S.-W. Lee, Automatic cardiac Arrhythmia classification using residual network combined with long short-term memory, IEEE Trans. Instrum. Meas. 71 (2022) 1–17.

[27] Ma, Shuai, Jianfeng Cui, Weidong Xiao, and Lijuan Liu. "Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 1577778.

[28] T. Pokaprakarn, R.R. Kitzmiller, R. Moorman, D.E. Lake, A.K. Krishnamurthy, M.R. Kosorok, Sequence to sequence ECG cardiac rhythm classification using convolutional recurrent neural networks, IEEE J. Biomed. Health Inf. 26 (2). (2022) 572–580.

[29] Y. Jin, J. Liu, Y. Liu, C. Qin, Z. Li, D. Xiao, L. Zhao, C. Liu, A novel interpretable method based on dual-level attentional deep neural network for actual multilabel Arrhythmia detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.

[30] Y. Xia, Y. Xiong, K. Wang, A transformer model blended with CNN anddenoising autoencoder for inter-patient ECG Arrhythmia classification, Biomed. Signal Process. Control 86 (2023) 105271.

[31] El-Ghaish, H. and Eldele, E., 2024. ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer. Biomedical Signal Processing and Control, 89, p.105714.

[32] Meng, Lingxiao, Wenjun Tan, Jiangang Ma, Ruofei Wang, Xiaoxia Yin, and Yanchun Zhang. "Enhancing dynamic ECG heartbeat classification with lightweight transformer model." *Artificial Intelligence in medicine* 124 (2022): 102236.

[33] Hutchins, DeLesley, Imanol Schlag, Yuhuai Wu, Ethan Dyer, and Behnam Neyshabur. "Block-recurrent transformers." *Advances in neural information processing systems* 35 (2022): 33248-33261.
[34] Omarov, Batyrkhan, Meirzhan Baikuvekov, Daniyar Sultan, Nurzhan Mukazhanov, Madina Suleimenova, and Maigul Zhekambayeva. "Ensemble Approach Combining Deep Residual Networks and BiGRU with Attention Mechanism for Classification of Heart Arrhythmias." *Computers, Materials & Continua* 80, no. 1 (2024).
[35] Huang, Youhe, Hongru Li, and Xia Yu. "A novel time representation input based on deep learning for ECG classification." *Biomedical Signal Processing and Control* 83 (2023): 104628.
[36] Nerella, Subhash, Sabyasachi Bandyopadhyay, Jiaqing Zhang, Miguel Contreras, Scott Siegel, Aysegul Bumin, Brandon Silva et al. "Transformers and large language models in healthcare: A review." *Artificial Intelligence in Medicine* (2024): 102900.
[37] Xia, Yi, Yangyang Xu, Peng Chen, Jun Zhang, and Yongliang Zhang. "Generative adversarial network with transformer generator for boosting ECG classification." *Biomedical Signal Processing and Control* 80 (2023): 104276.
[38] Pandey, Saroj Kumar, and Rekh Ram Janghel. "Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE." *Australasian physical & engineering sciences in medicine* 42, no. 4 (2019): 1129-1139.
[39] Golany, Tomer, and Kira Radinsky. "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 557-564. 2019.
[40] Hazra, Debapriya, and Yung-Cheol Byun. "SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation." *Biology* 9, no. 12 (2020): 441.
[41] Zheng, Huilin, Syed Waseem Abbas Sherazi, and Jong Yun Lee. "A cost-sensitive deep neural network-based prediction model for the mortality in acute myocardial infarction patients with hypertension on imbalanced data." *Frontiers in Cardiovascular Medicine* 11 (2024): 1276608.
[42] Hossain, Md Sazzad, John M. Betts, and Andrew P. Paplinski. "Dual focal loss to address class imbalance in semantic segmentation." *Neurocomputing* 462 (2021): 69-87.
[43] Vellido, Alfredo. "The importance of interpretability and visualization in machine learning for applications in medicine and health care." *Neural computing and applications* 32, no. 24 (2020): 18069-18083.
[44] Sahlaoui, Hayat, Anand Nayyar, Said Agoujil, and Mustafa Musa Jaber. "Predicting and interpreting student performance using ensemble models and shapley additive explanations." *IEEE Access* 9 (2021): 152688-152703.
[45] Ojha, Jaya, Hårek Haugerud, Anis Yazidi, and Pedro G. Lind. "Exploring Interpretable AI Methods for ECG Data Classification." In *Proceedings of the 5th ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*, pp. 11-18. 2024.
[46] Ayano, Y., F. Schwenker, B. D. Dufera, and T. G. Debelee. "Interpretable machine learning techniques in ECG-based heart disease classification: a systematic review. Diagnostics. 2023, 13 (1): 111." (2022).
[47] YENDURI, Gokul, et al. Gpt (generative pre-trained transformer)–a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 2024.
[48] G. Moody, R. Mark, The impact of the MIT-BIH Arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[49] R. Bousseljot, D. Kreiseler, A. Schnabel, Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet, Walter de Gruyter, Berlin/New York Berlin, New York, 1995.
[50] M. Kachuee, S. Fazeli, M. Sarrafzadeh, ECG heartbeat classification: A deep transferable representation, in: IEEE International Conference on Healthcare Informatics, ICHI, 2018, pp. 443–444,