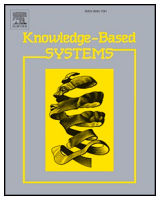
Knowledge-Based Systems 331 (2025) 114889

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Arrhythmia detection in multi-channel ECG images: vision transformer and explainable approaches

Fatma Murat Duranay [a,*], Ender Murat [b], Oğuzhan Katar [c], Yakup Demir [a], Ru-San Tan [d,e], Özal Yildirim [f], U. Rajendra Acharya [g,h]

[a] Department of Electrical and Electronics Engineering, Firat University, Elazığ, Türkiye
[b] Department of Cardiology, Health Sciences University Gülhane Training and Research Hospital, Ankara, Türkiye
[c] Department of Software Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye
[d] National Heart Centre Singapore, Singapore
[e] Duke-NUS Medical School, Singapore
[f] Department of Artificial Intelligence and Data Engineering, Faculty of Engineering, Fırat University, Elazığ, Türkiye
[g] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia
[h] Center for Health Research, University of Southern Queensland, Springfield, Australia

ARTICLE INFO

Keywords:
ECG classification
Vision transformer (ViT)
Explainable AI (XAI)
Deep learning in cardiology

ABSTRACT

The electrocardiogram(ECG) signals are usually converted to spectrogram images for analysis, but this approach has significant limitations. Firstly, it loses about time information, and secondly, it lacks resolution. In our research, we introduce an approach to directly convert ECG signals into PNG format and systematically sequentially incorporate each lead data. This method offers a precise depiction by maintaining the intricate temporal and spatial attributes of the signals through the use of Vision Transformer (ViT) models. The study evaluated the arrhythmia detection capabilities of four ViT models (ViT-B/224, ViT-L/224, ViT-B/384, ViT-L/384) for classifying ECG images. Six different Class Activation Mapping (CAM) techniques (ScoreCAM, EigenCAM, EigenGradCAM, GradCAM++, XGradCAM and LayerCAM) were employed to enhance model interpretability. Our best-performing model, ViT-B/384, achieved 96.79% accuracy and an F1-score of 96.78%, outperforming recent state-of-the-art CNN-based approaches for arrhythmia detection, such as DenseNet (F1: 98.9%, binary tasks) and ResNet-based models (Acc: 95.8%) in multi-class scenarios, while providing improved interpretability. Comparative analysis shows that our method improves multi-class arrhythmia classification accuracy by up to 1–2% over prior ViT or CNN-based methods on similar datasets. These results demonstrate that integrating high-resolution ECG image transformation with ViT models enhances diagnostic precision and model transparency, representing a significant step toward trustworthy AI-based medical diagnostic tools.

1. Introduction

The electrocardiogram (ECG) registers the surface action potentials of the heart chambers' electrical signal conduction via multiple geographically spaced electrodes placed on the chest wall and limbs. Compared with single-channel systems, multi-channel ECG systems provide more comprehensive coverage of the heart's electrical activity and plausibly confer more diagnostic differentiation. Clinicians analyse multi-channel ECG signal morphology and rhythm patterns to diagnose diverse heart conditions, including arrhythmia [1], but the manual interpretation is dependent on experience and human biases. Recent advances in machine learning hold the potential to automate multi-channel ECG signal analysis and improve both diagnostic efficiency and accuracy [2].

Yıldırım et al. [3] and Murat et al. [4] successfully combined the deep learning and single-channel ECG features for arrhythmia diagnosis underscoring the potential of deep learning for this application. Deep learning models, in ECG signal analysis, have shown processing capabilities compared to machine learning methods. Baek et al. developed a network that can detect subtle changes in ECG signals of individuals with paroxysmal atrial fibrillation during normal sinus rhythm [5]. Jeong et al. Introduced a network model that accurately categorizes

* Corresponding author.
E-mail addresses: fmurat@firat.edu.tr (F. Murat Duranay), ender.murat@sbu.edu.tr (E. Murat), okatar@firat.edu.tr (O. Katar), ydemir@firat.edu.tr (Y. Demir), tanrsnhc@gmail.com (R.-S. Tan), ozalyildirim@firat.edu.tr (Ö. Yildirim), rajendrauacharya@gmail.com (U.R. Acharya).

https://doi.org/10.1016/j.knosys.2025.114889
Received 15 July 2024; Received in revised form 18 August 2025; Accepted 9 November 2025
Available online 10 November 2025

eight types of arrhythmias using 12-lead ECG data. Baloglu et al. Presented a CNN model for detecting infarction in 12-lead ECG signals with high accuracy [6]. Baloglu et al. Presented a CNN model for detecting infarction in 12-lead ECG signals with high accuracy [7]. While CNN models have been widely successful, the Vision Transformer (ViT) has emerged as an alternative due to its structure and global connectivity, allowing for more comprehensive analysis of multi- channel ECG signals [8].

The ViT architecture includes encoder-decoder blocks that can process data simultaneously without the need for networks [9,10]. With attention mechanisms enabling the understanding of long-distance dependencies within data, ViT models excel in analyzing data [11,12], like time series information [13,14]. The application of ViT to multi-channel ECG analysis may be particularly efficacious due to ViT's ability to capture deep features essential for comprehensive image analysis. Manzari et al. devised MedViT, a CNN-transformer hierarchical hybrid medical image classification architecture capable of capturing both short- and long-term connections in visual data [15]. In their arrhythmia classification model, Che et al. used CNN to extract deep features from ECG signals, combined with ViT to analyse the temporal characteristics of the signals [16]. These models exploit ViT's capacity to integrate complex features globally, leading to significant refinements in diagnostic methods.

Explainability of the decisions of machine learning models is critical to instilling confidence in the model recommendations and, pertaining to ECG diagnosis, garnering acceptance among clinicians. Many explainable artificial intelligence (XAI) approaches have been created. Gradient-weighted Class Activation Mapping (Grad-CAM), a popular strategy, uses first-order gradients of the input signals [17,18] to elucidate and illustrate regions of the ECG signal that contribute to the decision-making process [19–21]. The utility of Grad-CAM lies in identifying these specific clinical data that most inform the predictions [22]. Shapley Additive Explanations (SHAP), another XAI approach, also elucidates the contribution of various features during the model classification. Features with high Shapley values indicate a more pronounced influence on the predictions. SHAP evaluates how individual features contribute to model predictions while considering interactions and dependencies among all features. It offers an assessment that takes into account the relationships between features [23,24]. Additionally, it can serve as a tool for extracting features, allowing for the extraction of both regional information from ECG signals [25]. Some researchers have also developed intelligent XAI methods, such as multivariate linear regression model [26], neural-backed ensemble tree [27] and frameworks tailored to their specific datasets [28,29]. These XAI approaches have enhanced the model's ability to comprehend and interpret decision-making processes, thereby improving the reliability, transparency and automation of diagnostics.

In this research, we explore the potential of ViT and XAI methods in detecting arrhythmia using 12-lead ECG recordings that have been converted into high-resolution lead images for analysis. We discuss the shortcomings of existing techniques, examine how integrating these innovative methods can enhance current diagnostic procedures, and consider their potential impact on healthcare. In this study, our primary objective is to systematically evaluate the performance of existing ViT architectures on multi-lead ECG-based arrhythmia detection and to investigate their explainability using multiple CAM techniques. Rather than proposing novel architectural modifications, we focus on a fair and reproducible comparison of standard ViT variants under consistent preprocessing and training settings. The key innovations of this study include;

- Analyzing ECG signals by transforming them into images derived from data instead of relying on conventional spectrograms. By incorporating data from each lead into the image, a precise and detailed representation is achieved. This enhancement significantly boosts the model's ability to identify arrhythmias and interpret the findings effectively.
- Employing two variations of ViT technology (ViT-B/224 and ViT-L/224) at resolutions (224 × 224 and 384 × 384) to determine the suitable model. This comparative assessment across model types and resolutions offers an examination to ascertain which combination yields optimal performance for specific arrhythmia detection tasks.
- Six various techniques, including EigenCAM, EigenGradCAM, GradCAM++, XGradCAM, LayerCAM and ScoreCAM, are applied to create models. The goal is to identify the model for explanation and conduct a comparative assessment of this model against other CAM methods.

2. Materials and methods

This paper discusses the transformation of multi-channel ECG signals into high-resolution images and the application of ViT-based deep learning models for ECG data processing and arrhythmia detection. A large dataset was analysed in detail by cardiologists and classified into 11 different rhythm classes. This classification was then reduced to four main categories for ease of analysis. The study evaluated the classification performance of the ViT model trained on the converted ECG images. The Grad-CAM method was integrated to increase the transparency and comprehensibility of the model's decision-making mechanism. This method visualises the critical clinical indicators underlying the model's predictions by showing which parts of the images the model focuses on during the classification process to yield its predictions. As shown in Fig. 1, this approach represents an important step in the development of decision support systems for the application of deep learning models in cardiology.

2.1. ECG database

The study dataset comprised a denoised ECG dataset [30] containing ECG recordings of more than 10,000 people, each 10-second recording consisting of a 12-lead ECG sampled at 500 Hz, which experts had classified into 11 different rhythm classes. Although the dataset originally included 12 leads, during the data preprocessing stage, we observed that one of the leads (Lead 12) could not be accessed due to file corruption in the released dataset. Specifically, this lead's signal data was either missing or unreadable for a large portion of the samples, which prevented its use in our analyses. Consequently, we conducted all experiments using the remaining 11 leads to ensure consistency and data integrity. Due to the small number of cases in some rhythm classes, the database authors combined groups of similar rhythms into four main categories. As shown in Fig. 2, the analyses in this current study are based on these four combined rhythm classes (Table 1).

2.2. Processing of ECG raw data

A significant advancement in this study is how the one-dimensional input data is structured as channel images. Initially, each lead of the raw ECG signal was transformed into a single-channel normalized image with a resolution of 300 × 600 pixels. The resulting 11 lead images were concatenated vertically according to the lead order to generate a single composite image. The composite image was then resized to 224 × 224 pixels and 384 × 384 pixels, each with three color channels (RGB), to match the input specifications of the ViT models employed in this study. The process is shown in Fig. 3.

2.3. ViT architecture

The ViT model applies the Transformer architecture, known for its success in natural language processing (NLP), to tasks involving image classification [8]. The basic method of the ViT model is based on dividing images into uniformly sized patches and using these patches as

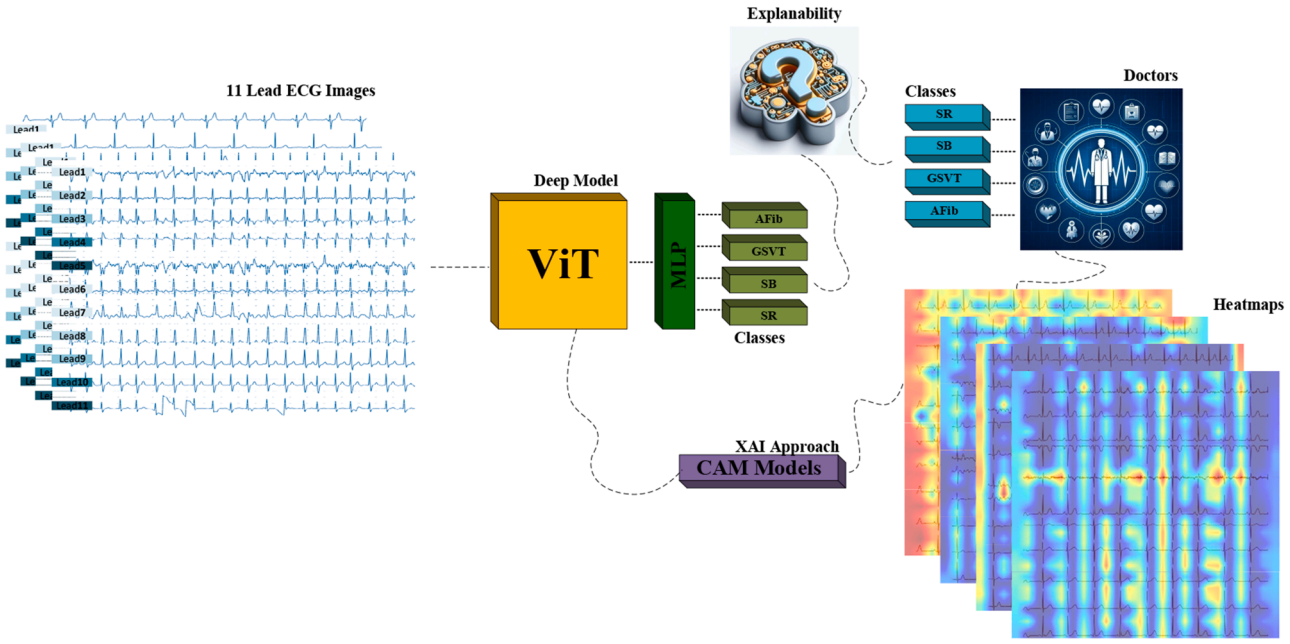


Fig. 1. Study framework: 11-lead ECG images processed by the ViT model, classified into four classes with CAM-generated heatmaps.

Table 1
Information about the values and labels for the combined ECG rhythm categories.

Merged Rhythms	Rhythm Classes	Class Name	Number
AF, AFIB	atrial flutter, atrial fibrillation	"AFib"	2218
SVT, AT, SAAWR, SINT, AVNRT, AVRT	supraventricular tachycardia, atrial tachycardia, sinus atrium to atrial wandering rhythm, sinus tachycardia, atrioventricular node reentrant tachycardia, atrioventricular reentrant tachycardia	"GSVT"	2260
SB	sinus bradycardia	"SB"	3888
SR, SI	sinus rhythm, sinus irregularity	"SR"	2222

inputs to the Transformer architecture treating each patch as a word. This process begins with the generation of image patches. ViT processes an image by dividing it into uniform patches of a given size, e.g., 16×16, 32×32. This process converts the image into a series of flat vectors, where each vector contains the smoothed pixel values of a patch. Each patch vector is matched to a Transformer by adding a unique position code. This encoding allows the model to understand patch order and position, as the Transformer architecture is not sensitive to input order. ViT uses a "class token" (often referred to as a [CLS] token) for classification tasks. This token represents the entire image and is used in the training process for the model to transform it into a learned vector representing the image. Patch vectors and the class token are passed through an array of Transformer blocks. Each block contains multiple header attention mechanisms and successive fully connected layers. This structure aims to learn the relationships and contexts between patches and to obtain an overall representation of the image. At the end of the training, the class token output is fed into a classification layer (usually a fully connected layer) and the class to which the image belongs is predicted. The novelty of the ViT model lies in how it processes images, i.e., by processing them directly into patches and feeding these patches into the Transformer model, which differs from the way traditional CNNs learn local features and structures. This approach exploits the Transformer's ability to learn long-range connections and complex patterns, leading to impressive results in image classification tasks.

There are ViT-Base and ViT-Large models of the ViT structure, built

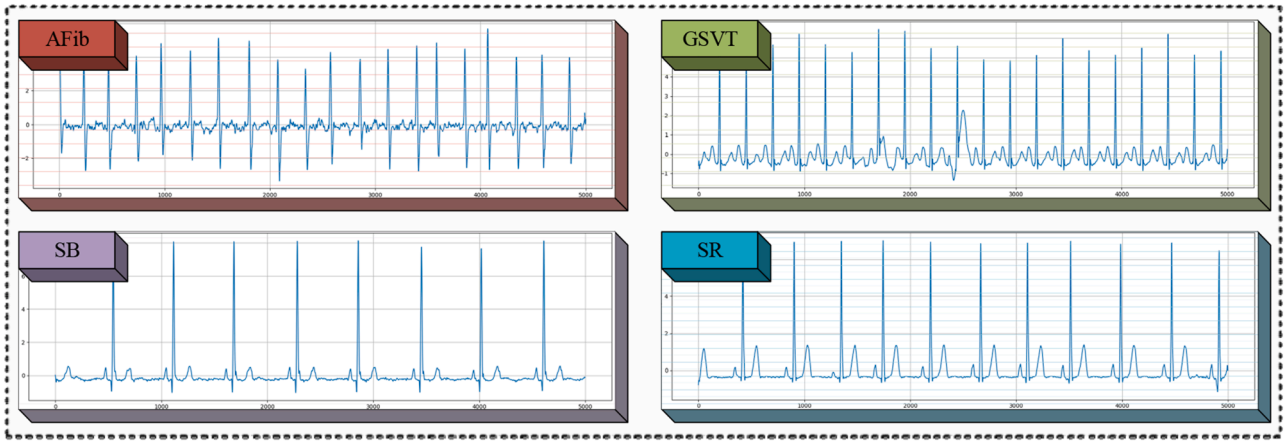


Fig. 2. Example ECG signals for four classes: AFib, GSVT, SB, and SR.

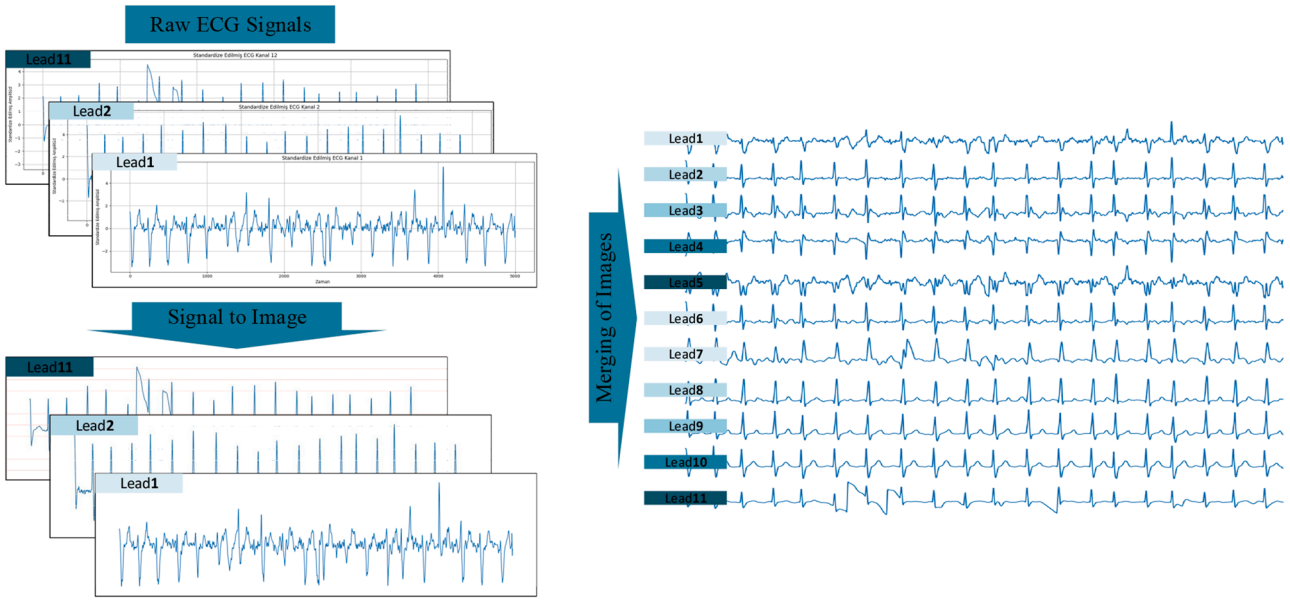


Fig. 3. Process of transforming raw ECG signals into a merged image format.

in various configurations and sizes. Different configurations vary according to the input patch size of the model. For example, ViT-L/16 indicates that the "Large" model has an input patch size of 16 × [8] There are also hybrid models combined with ResNet structures, for example, configurations such as 'R50+ViT-B/32′ and 'R50+ViT-L/16′ This diversity ensures the scalability of the model for different tasks and requirements. Larger models generally perform better, while smaller models may require faster inference and fewer computational resources. The model configurations used in this study are detailed in Table 2.

In this research, we utilized four ViT models. Two ViT models, ViT Base and ViT Large, were employed with images of two sizes. 224 × 224 and 384 × 384 pixels. Comparisons were made based on both the model type and image size used.

- The ViT Base model, with an input resolution of 224 × 224 pixels, is designed for images. It consists of 12 transformer blocks, each with 12 attention heads. The total parameter count for this model is around 86 million.
- On the other hand, the ViT Large model also operates with an input resolution of 224 × 224 pixels. Is larger in scale. It comprises 24 transformer blocks, each with 16 attention heads totaling 307 million parameters optimized for images.
- Moving on to the high resolution category, the ViT-B/384 model is tailored for images at a resolution of 384 × 384 pixels. Similar to its counterpart, it features 12 transformer blocks with 12 attention heads and around 86 million parameters.
- Lastly, the ViT-L/384 model caters to high-resolution images well but boasts an extensive structure. It encompasses 24 transformer blocks with 16 attention heads and approximately 307 million parameters.

The performance of these four models was compared, and the performance of each model at different image resolutions and parameter configurations was analysed. The results are used to evaluate the ability of the models to discriminate arrhythmia classes.

3. Experimental results

The ViT models we proposed were trained on a high-performance PC workstation. Our workstation features an Intel i9 14900 K processor paired with an NVIDIA GeForce RTX 4090 graphics card, along with 64 GB of RAM. We utilized software versions such as timm==0.9.2 and torch≥1.7, running the training process in Python using the PyTorch framework and TIMM library. Thanks to this hardware setup and optimized software tools, the models were trained efficiently and swiftly.

The dataset used in our study was carefully partitioned for the model training, validation and testing phases. The data partitioning process was done to evaluate the performance of the models fairly and to test their generalisation capabilities. Fig. 4 shows how the dataset is divided into training, validation and testing datasets. As can be seen from the graph, the dataset is divided into three main categories: Training data (blue), 80 %, Validation data (red), 10 % and Testing data (green), 10 %. This split provides a balanced and comprehensive approach to training and evaluating the performance of the models. In the testing dataset, there are 222 AFib, 222 GSVT, 222 SR, and 388 SB samples, ensuring a balanced class distribution that reduces potential bias and allows for a fair evaluation of classification performance. For all experiments, the models were trained using a fixed set of optimization hyperparameters, including a learning rate of 2e-5, batch size of 16, AdamW optimizer, and 100 training epochs, with learnable positional embeddings for all patch sizes (16 × 16). The full list of optimization hyperparameters used in this study is presented in Table 3.

In our study, we trained four different ViT models and recorded 100 epochs to evaluate their training and validation performance. We found the epoch number with the best test validation of the models and marked it on the accuracy graphs. While the maximum test accuracy value for

Table 2
Specifications of the four ViT models used in the study.

Model	Input Resolution	Patch Count	Transformer Blocks	Attention Heads	Parameters (Approx.)	Embedding Dimension
ViT-B/224	224 × 224	196	12	12	86M	768
ViT-L/224	224 × 224	196	24	16	307M	1024
ViT-B/384	384 × 384	576	12	12	86M	768
ViT-L/384	384 × 384	576	24	16	307M	1024

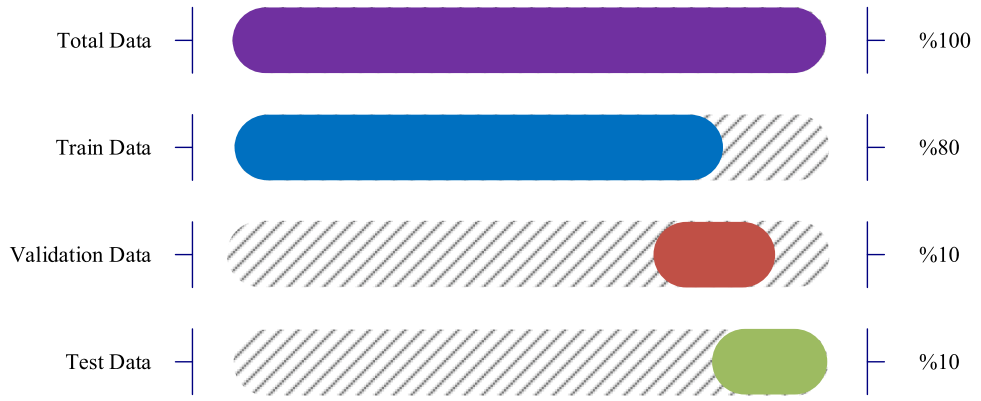


Fig. 4. Data split distribution used in the study, showing the allocation of the total data into training (80 %), validation (10 %), and test (10 %) sets.

Table 3
Optimization hyperparameters used in this study.

Patch Size	16 × 16
Positional Encoding	Learnable positional embeddings
Learning Rate	2e-5
Batch Size	16
Optimizer	AdamW
Epochs	100

the ViT-B/224 model was obtained at epoch 47, it was found to be 82 for the ViT-B/384 model, 38 for the ViT-L/224 model and 98 for the ViT-L/384 model.

As can be seen in Fig. 5, the training and validation accuracies of each model increased over time and stabilised at a certain level. The accuracy curves in the graph show that the validation accuracy of all four models is close to the training accuracy, indicating that the models have good generalisation capabilities. The ViT-B/224 and ViT-B/384 models show a rapid increase in accuracy, especially in the early epochs, while the ViT-L/224 and ViT-L/384 models show a slower but steady growth. These results suggest that the larger models perform better, especially for high-resolution images. However, the performance differences between the models were relatively small, and all four models successfully classified ECG signals.

Once the training was complete, the models were tested on never-before-seen test data to assess their performance. The test data was used to determine how the model performed in real-world scenarios. In this dataset, where all data is separated on a patient basis, there is only one data for each patient. This patient-based structure prevented data leakage between the training and testing phases and allowed us to assess the generalisation capabilities of the models accurately.

In the testing process, the models were applied to the test dataset and the multi-channel ECG signal images of each patient were analysed and the appropriate classifications were made. At this stage, the performance of the models was evaluated using various metrics. Performance metrics were used to determine the disease classification capabilities of the models. The test results shown in Table 3 demonstrate the ability of the models to generalise what they have learned during the training process and to make accurate predictions on unseen data.

When analysing the confusion matrices of the four ViT models (Fig. 6), it can be seen that all models have generally high accuracy

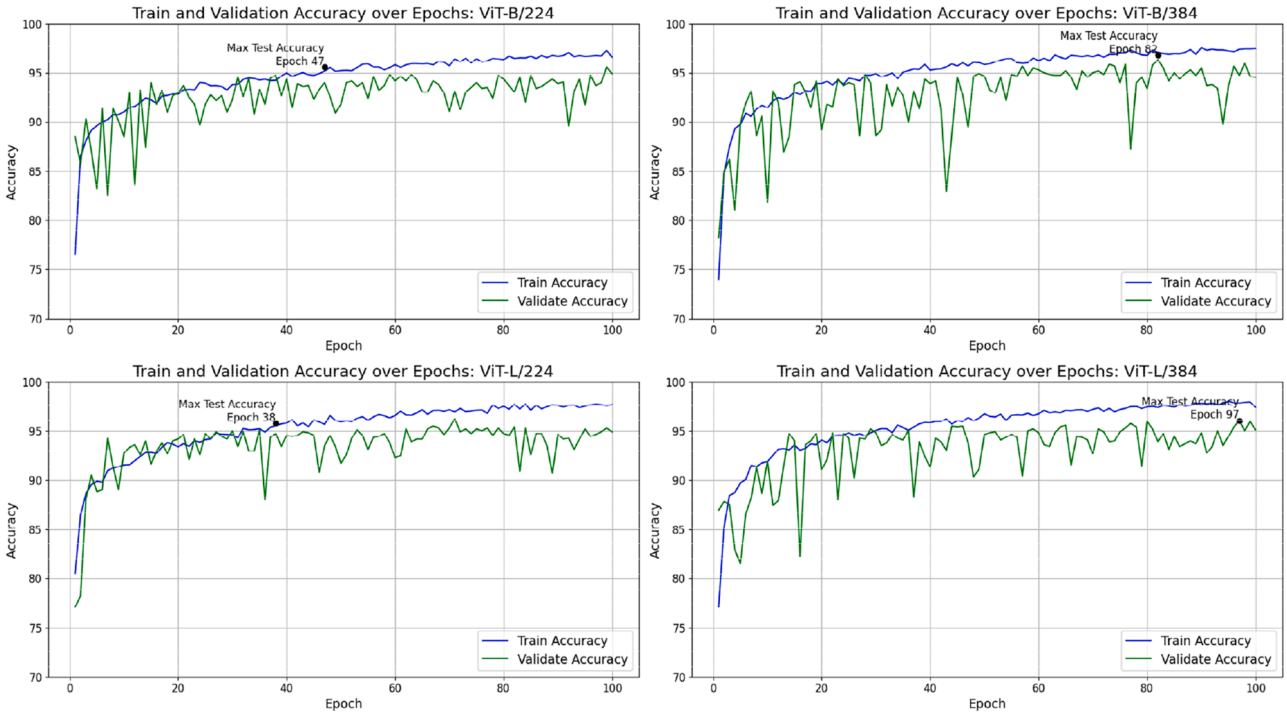


Fig. 5. Training and validation accuracy over 100 epochs for the four ViT models (ViT-B/224, ViT-B/384, ViT-L/224, ViT-L/384).

rates. The ViT-B/224 and ViT-B/384 models showed high accuracy in the SB and SR classes. The ViT-L/224 and ViT-L/384 models, despite having a larger parameter set, achieved similar accuracy rates and had slightly higher false positive classification rates in the AFib and GSVT classes. In the AFib class, the ViT-L/384 model had the highest accuracy rate. In the GSVT class, ViT-B/224 and ViT-L/224 were the models with the fewest errors. The SB class stands out as the class with the highest accuracy rate for all models, while in the SR class, all models gave very successful results. In general, the higher resolution ViT models (ViT-B/384 and ViT-L/384) performed better, but all models successfully classified ECG signals.

We used measures of success to assess the performance of the models in our study. These measurements helped us determine the accuracy, precision, recall and F1 score of the models. Accuracy is calculated by comparing the predicted samples to the number of samples. Precision indicates the percentage of predicted positive instances, out of all instances identified as positive by the model. Recall measures how effectively the model identifies positive situations. The F1 Score combines precision and recall to evaluate overall model performance. Below are the formulas we used to calculate these metrics [7]:

$$Accuracy = \frac{True\ Positives\ + True\ Negatives}{Total\ Number\ of\ Samples} \tag{1}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3}$$

$$F1 - Score = 2\ \times \frac{Precision\ \times Recall}{Precision\ + Recall} \tag{4}$$

These metrics are employed to assess and contrast the effectiveness of the models thoroughly. Each metric reveals a different aspect of the model's classification capabilities, allowing for a more comprehensive performance analysis.

Table 4 shows the performance metrics of four different ViT models (ViT-B/224, ViT-B/384, ViT-L/224, ViT-L/384) on four classes (AFib, GSVT, SB, SR). The overall performance of the models was evaluated using the metrics of accuracy, precision, recall and F1 score.

The ViT-B/384 model seems to outperform the models in terms of overall performance. It excelled in the AFib category, achieving an

Table 4
Performance metrics for the four ViT models across four classes: AFib, GSVT, SB, and SR.

Model	Classes	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
ViT-B/224	AFib	97.17	96.15	90.09	93.02
	GSVT	97.36	92.31	95.58	93.91
	SB	99.06	97.73	**99.74**	98.73
	SR	97.54	94.55	93.69	94.12
	Overall	95.56	95.58	95.56	95.54
ViT-B/384	AFib	98.39	95.96	96.40	96.18
	GSVT	97.83	96.77	92.92	94.81
	SB	**99.24**	**98.47**	99.49	**98.98**
	SR	98.11	94.69	96.40	95.54
	Overall	**96.79**	**96.79**	**96.79**	**96.78**
ViT-L/224	AFib	97.26	92.51	94.60	93.54
	GSVT	97.64	93.89	95.13	94.51
	SB	98.77	98.21	98.46	98.33
	SR	97.83	96.71	92.79	94.71
	Overall	95.75	95.78	95.75	95.75
ViT-L/384	AFib	98.49	96.40	96.40	96.40
	GSVT	97.83	96.35	93.36	94.83
	SB	98.49	97.46	98.46	97.95
	SR	97.26	92.89	94.14	93.51
	Overall	96.03	96.04	96.03	96.03

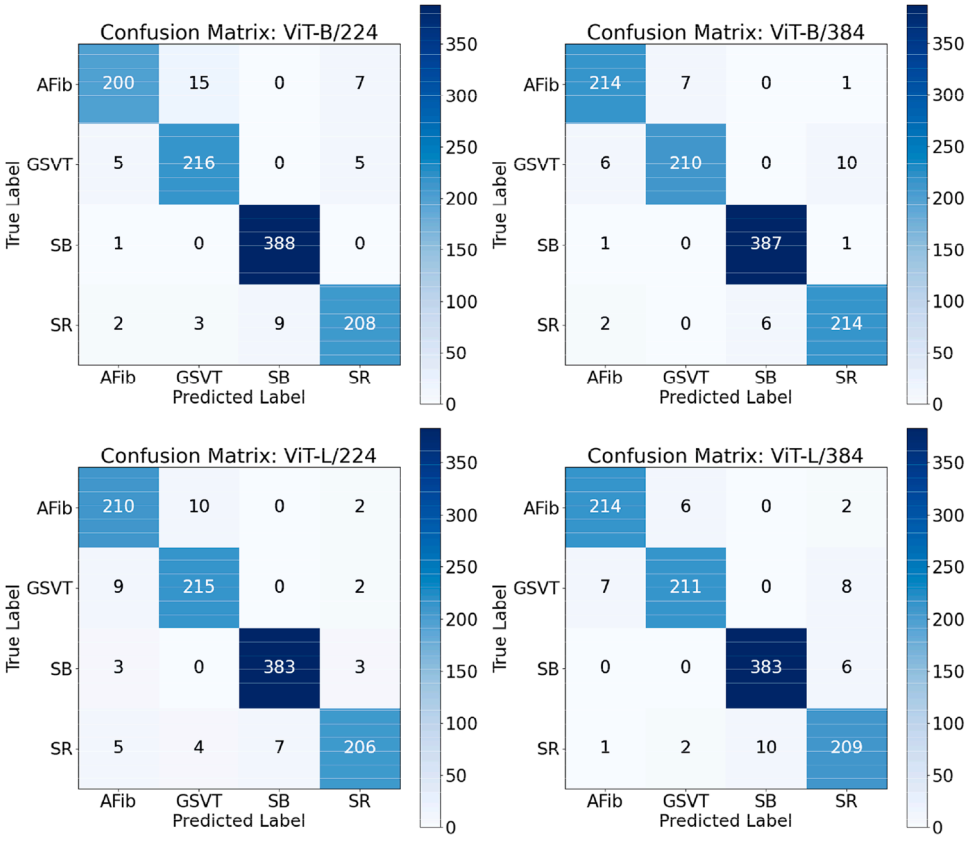


Fig. 6. Confusion matrices for the four ViT models showing classification performance across four classes: AFib, GSVT, SB, and SR.

accuracy of 98.39 %, precision of 95.96 % recall of 96.40 % and an F1 score of 96.18 %. In the GSVT category, it achieved an accuracy of 97.83 %, precision of 96.77 %, recall of 92.92 % and an F1 score of 94.81 %. The SB category showed the performance with an accuracy of 99.24 %, precision of 98.47 % recall of 99.49 % and an F1 score of 98.98 %. The SR category reached an accuracy of 98.11 %, precision of 94.69 % recall of 96.40 % and an F1 score of 95.54 %. Overall, this model delivered the performance with accuracies, precisions, recalls and F1 scores all around the mark at approximately 96 %.

The impressive results indicate that the ViT-B/384 model stands out as the dependable and efficient choice, delivering a combination of high accuracy along with balanced precision and recall metrics. As a result, further tasks were carried out using the ViT-B/384 model, which consistently demonstrated performance in classifying ECG signals. This decision was taken to enhance the accuracy, precision and generalization capacity of the models.

3.1. XAI approaches and Grad-CAM

GradCAM provides visual descriptions of model decisions without model architecture changes or model re-training [17]. By combining class-specific localisation with high-resolution visualisation, GradCAM produces interpretable and faithful visualisations representing the model's decision-making process. This technique helps us understand why the model fails, discover biases in the dataset, and increase confidence in the model's predictions. The importance of this framework is that it sheds light on the 'black box' nature of AI models, helping to make these models more understandable and reliable.

In the GradCAM technique, the initial step involves calculating the gradients to assess how much the activation maps in the convolution layer of the model impact the target class. This process entails determining the gradients of the convolution layer with respect to the output values of the target class. Through mean pooling applied to these gradients, weight factors are derived for each activation map, indicating their significance in predicting the target class. These weight factors are then multiplied by the activation maps in the convolutional layer to create a CAM. This visual representation highlights areas that play a role in predicting the target class. By applying the ReLU function to this CAM, negative impacting activations on the target class are suppressed, showcasing features that positively contribute to the map. Finally, the resulting class activation map is integrated with the original image to provide a final visualisation of which areas are important in the model's decision-making process. As shown in Fig. 7, by revealing the internal dynamics and decision-making mechanism of the model more transparently, the Grad-CAM method helps to understand model mispredictions, make improvements and detect potential biases in the dataset.

In our study, several CAM methods were used to improve the interpretability of the ViT-B/384 model predictions on the test data set. These CAM methods include ScoreCAM, EigenCAM, EigenGradCAM, GradCAM++, XGradCAM and LayerCAM. Each CAM method produces heat maps highlighting regions of the ECG signals that the model focuses on when making predictions. ScoreCAM directly shows the regions to which the model is paying attention, while EigenCAM uses eigenvalue decomposition of activation maps to highlight important components. EigenGradCAM combines eigenvalue and gradient information to produce more detailed maps. GradCAM++ provides more detailed activation maps by taking into account the second-order effects of gradients. XGradCAM combines gradient information from different layers and LayerCAM evaluates activation information from each layer. Understanding representations is essential for grasping how deep learning models make decisions and guaranteeing their trustworthiness in medical environments. The experimental results from these CAM methods visually represent the regions that the model is paying attention to and help clinicians validate the model's predictions and ensure that the model is making decisions based on clinically important features. This approach bridges the gap between model interpretability and practical applicability in medical diagnosis.

In terms of evaluating the best CAM method for AFib diagnosis, Fig. 8 shows that the EigenCAM and ScoreCAM methods perform remarkably well. EigenCAM uses the eigenvalue decomposition of the activation maps to highlight the important components, clearly showing the regions of focus of the model. This method clearly shows the absence of irregular R-R intervals and P-waves in ECG signals. On the other hand, ScoreCAM highlights clinically relevant findings by directly showing the model's regions of interest. The visual descriptions provided by ScoreCAM clearly show which areas are important in the model's decision-making process. Although both methods highlighted the characteristic features of AFib, EigenCAM was used to interpret the patient data to present EigenCAM as an alternative.

The model output of a patient belonging to the AFib class with Eigencam and the locations where it is concentrated are shown in Fig. 9. Accordingly, the heat map of the model classifying a patient with AFib with 99.78 % accuracy is analysed. The heat map shows that the model

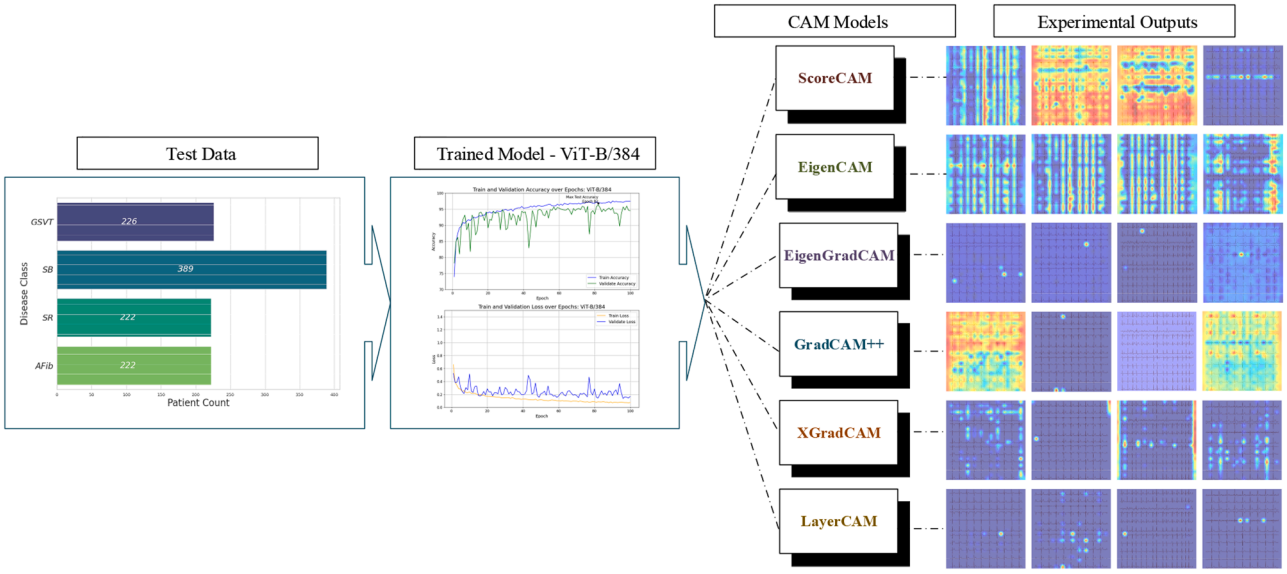


Fig. 7. Visualization of the testing process and results for the ViT-B/384 model.

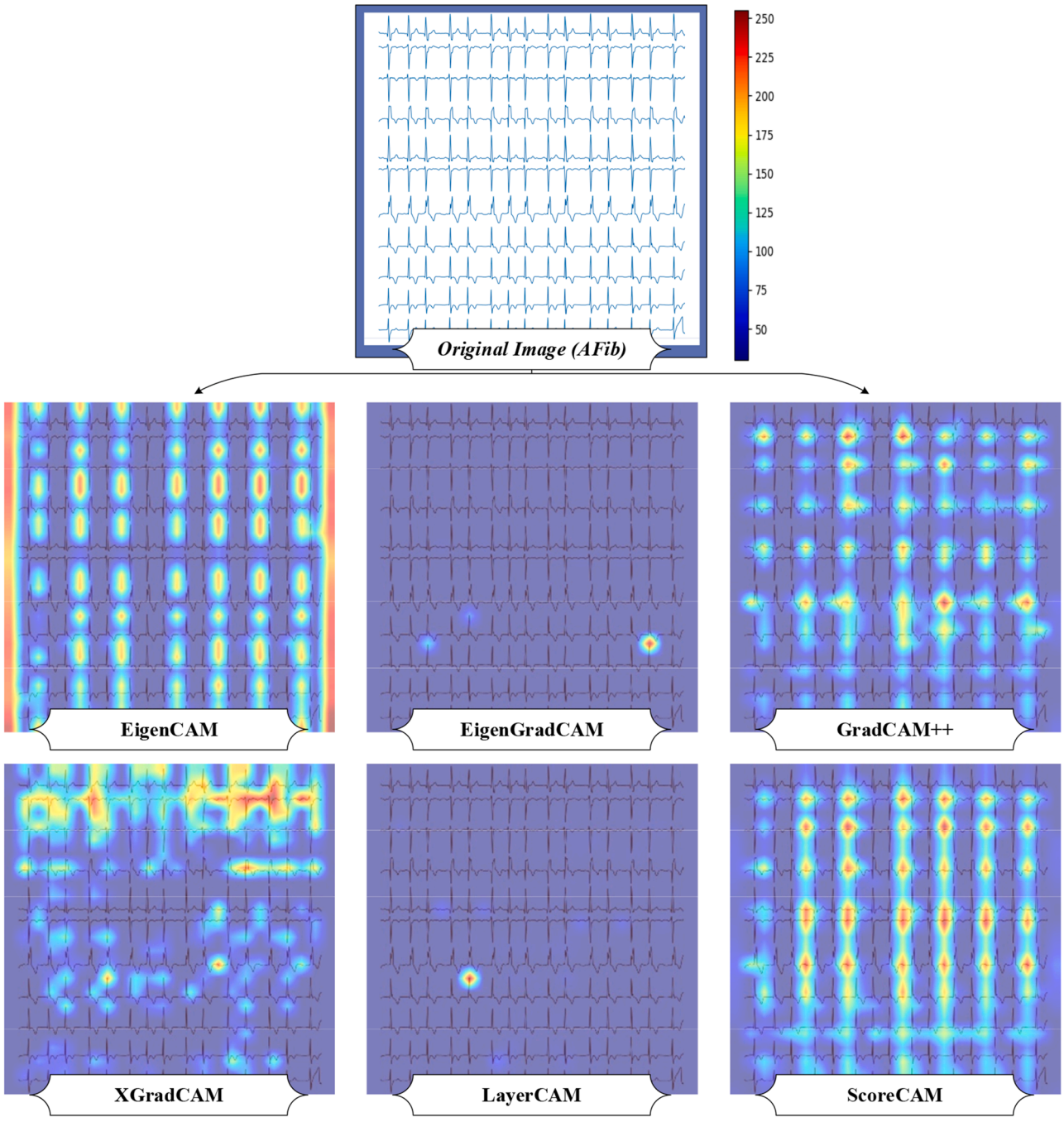


Fig. 8. Visualisation of different CAM methods applied to an ECG image.

emphasises the characteristic features of AFib, such as irregular R-R intervals and the absence of P-waves in the ECG signals. The model's attention to fibrillation waves and irregular ventricular response areas is crucial for detecting AFib. This focus aligns with the clinical and electrophysiological characteristics of AFib, endorsing the model as a reliable diagnostic tool. Thus, the regions highlighted by the model in the heatmap can be linked to AFib, indicating that the model is appropriate for applications.

In Fig. 10, the model's output for a patient classified under the GSVT category using Eigencam and its concentrations is displayed. The heatmap illustrates where the model concentrated its efforts to identify this category. The GSVT category is distinguished by a consistent ventricular response. The heatmap reveals that the model pays attention to characteristics like the absence of p waves and regularity of R R intervals in this class. Nonetheless, the NSR category also exhibits a rhythm and distinct P waves. These similarities may confuse the model in distinguishing the SR class from GSVT.

The main reason the model has difficulty distinguishing between GSVT and SR classes is that both classes have regular rhythm patterns. In SR, the p-waves are regular and evenly spaced, whereas in GSVT, there is some regularity between atrial and ventricular activity. The model's focus on the location of p-waves and R-R intervals in the heat map makes it difficult to distinguish between these two classes. Therefore, the model confuses GSVT with the SR class because both types of arrhythmia are regular rhythms. This can be resolved with more training data and improved algorithms to increase the accuracy of the model in clinical use.

The model output and concentrations of a patient belonging to the SB

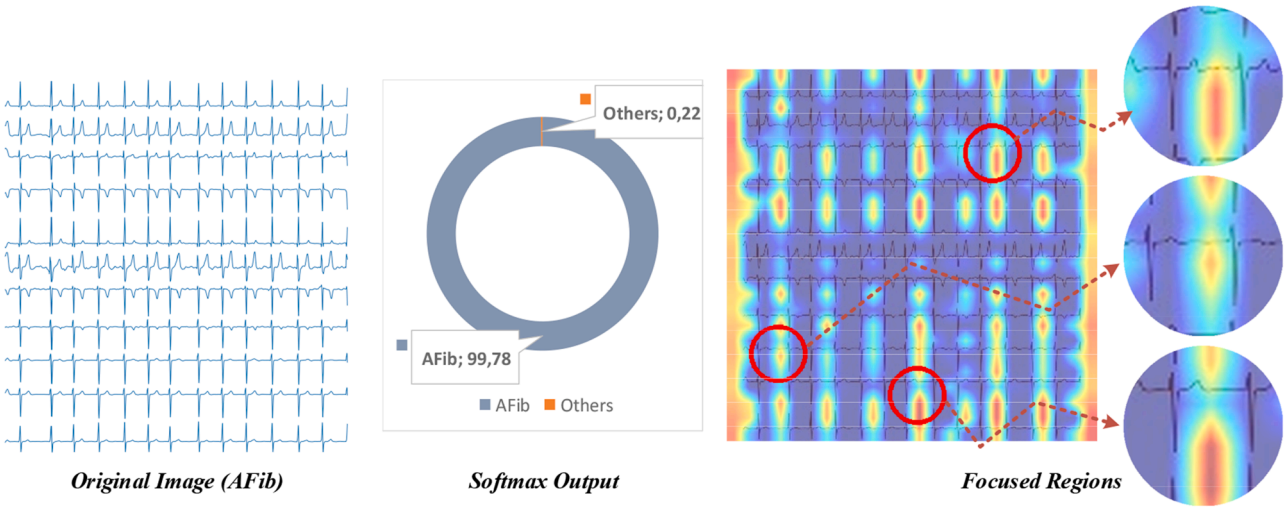


Fig. 9. Visualization of the original ECG image, the softmax output indicating the classification confidence (99.78 % for AFib and 0.22 % for Others), and the heatmap with highlighted focus areas.

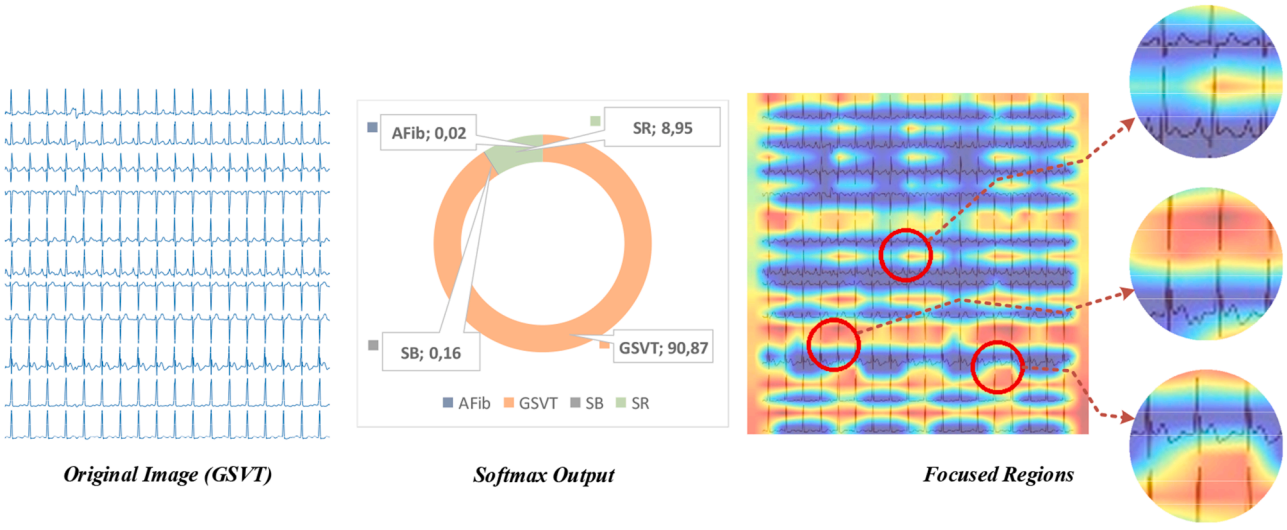


Fig. 10. Visualisation showing the original ECG image, the softmax output indicating classification probabilities (AFib: 0.02 %, SB: 0.16 %, SR: 8.95 %, GSVT: 90.87 %), and the heatmap indicating the model's focus areas.

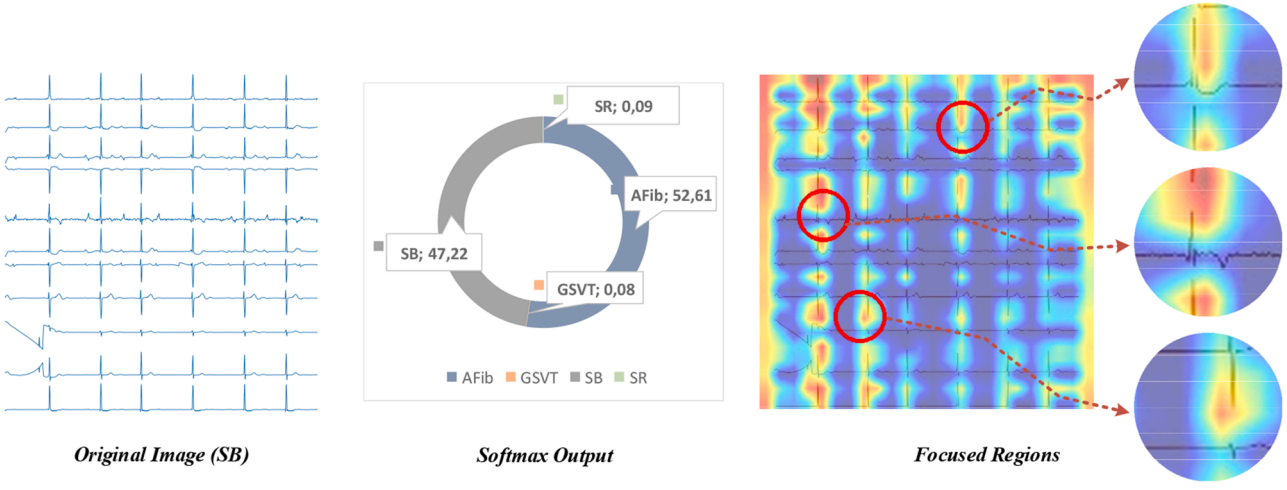


Fig. 11. Visualization of the original ECG image, the softmax output showing classification probabilities (AFib: 52.61 %, SB: 47.22 %, SR: 0.09 %, GSVT: 0.08 %), and the heatmap indicating the model's focus areas.

class with Eigencam are shown in Fig. 11. Softmax outputs show that the model identifies SB with 47.22 % accuracy and classifies AFib with 52.61 % accuracy. The possible reason for the model confusing SB and AFib classes is that low heart rates can characterise both rhythm types. SB has regular P-waves and R-R intervals, whereas AFib waves are irregular and chaotic. These irregularities and chaos are features that strengthen the diagnosis of AFib. Although the model sometimes gives false positive results, these features can be crucial in diagnosing AFib. In addition, as the heart rate decreases in AFib patients, the R-R intervals become more regular and can be confused with sinus bradycardia, leading to false negative results. Therefore, the model's difficulty distinguishing between AFib and sinus bradycardia is understandable, as low heart rate can be a common feature of both conditions.

The model output and concentrations of a patient in the SR class with Eigencam are shown in Fig. 12. The model identified SR with an accuracy of 86.45 % but also predicted AFib with a probability of 11.26 %. The areas of focus of the model in the heat map highlight the characteristic features of sinus rhythm, such as regular P waves and equal R-R intervals. However, the 11.26 % probability of AFib may be due to the similarity of some slow AFib episodes to the regular rhythm pattern. Although the waves of AFib are irregular and chaotic, there are occasions when the ventricular response appears regular. These similarities may cause the model to confuse AFib with SR.

4. Discussion

When compared with state-of-the-art CNN-based methods reported in prior works (Table 5), our proposed ViT-B/384 model demonstrated superior performance in multi-class arrhythmia detection. For instance, DenseNet-based models achieved an F1-score of 98.9 % in binary classification tasks, while ResNet-based methods reported an overall accuracy of 95.8 % in multi-class scenarios. In contrast, our best-performing model (ViT-B/384) attained 96.79 % accuracy and 96.78 % F1-score in multi-class classification, while also offering improved interpretability via CAM methods. These results indicate that, although CNN models remain strong baselines, ViTs provide a more balanced and transparent framework for complex ECG classification tasks. The direct benchmarks with prior CNN and spectrogram-based methods confirm the advantages of integrating ViTs with image-transformed ECG data.

The innovations we present in this paper provide significant improvements in the processing and analysis of ECG signals. Traditionally, ECG signals are analysed by converting them into spectrograms. However, this approach can lead to the loss of time and frequency information and may not reflect the full dynamics of ECG signals. Therefore, in our study, we converted ECG signals directly from raw data into images in PNG format. Data from each lead was added sequentially to create a more detailed and accurate representation. This method allows a more comprehensive analysis while preserving all the temporal and spatial characteristics of the signals. This approach represents a breakthrough in the processing and visualisation of raw ECG data, overcoming the limitations of conventional spectrum analysis.

An essential aspect of our research involves choosing a suitable model by comparing two variations of ViT technology (ViT-B/224 and ViT-L/224) at different image resolutions (224 × 224 and 384 × 384). Through assessing how various combinations of models and resolutions perform, this comparative study aims to identify which setup yields the results for specific arrhythmia detection tasks. This approach enables us to enhance model efficiency and select the optimal deep learning framework for diverse clinical scenarios. The findings indicate that resolution ViT models, particularly ViT-B/384, exhibit superior performance, suggesting that larger models and high-resolution images are more effective in detecting arrhythmias.

Furthermore, our research utilized six CAM techniques to develop interpretable models: EigenCAM, EigenGradCAM, GradCAM++, XGradCAM, LayerCAM and ScoreCAM. Each CAM method reveals the areas within ECG images that are focal points for the model's attention and influential in classification decisions. Following an assessment, the ScoreCAM method emerged as the top performer in elucidating how models make decisions. This thorough evaluation aims to deepen insights into model decision-making processes and enhance the interpretability of outcomes..

This study was conducted with comments from cardiologists:

4.1. Expert commentary

In this part, the analysis focuses on areas of the model when examining rhythm patterns. Through the analysis, significant insights emerged regarding these areas;

- When assessing patient's rhythms, it was noted that they often pay more attention to T P intervals rather than R R intervals. This difference could help differentiate between irregular rhythms. However, variations in breathing or additional beats might alter T P intervals, leading to a rhythm appearing irregular in such cases. This section analyses the focal regions of the model during rhythm analysis.
- As the rhythms compared with AFib are all regular rhythms (GSVT, SB, SR), it is possible for the model to diagnose AFib based on rhythm

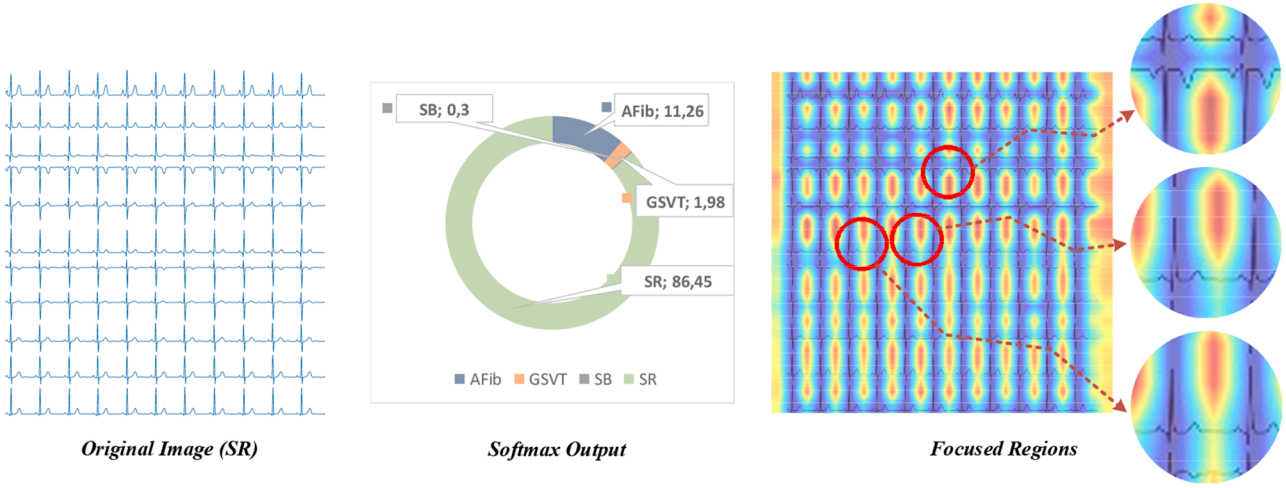


Fig. 12. Visualization of the original ECG image, the softmax output showing classification probabilities (AFib: 11.26 %, GSVT: 1.98 %, SB: 0.3 %, SR: 86.45 %), and the heatmap indicating the model's focus areas.

Table 5
Comparison of our method with recent XAI-based deep learning approaches.

Authors, Year	Input Types	XAI Method	Deep Method	Tasks	Performance (%)
Hicks et al. [31], 2021	ECG	Grad-CAM	CNN	Sex prediction	Acc:89
Lee et al. [32], 2019	Computed-tomography	Attention map	CNN	detection of acute intracranial haemorrhage	S1: sen:98, spec:95 S2: sen:92, spec:95
Kim et al. [33], 2022	ECG	Grad-CAM	Visual-DenseNet	Arrhythmia detection	Acc: 98.7, sen:98.7, F1: 98.9
Elul et al. [34], 2021	ECG	STA	CNN	Identification of underlying cardio-pathology	Acc: 96
Jahmunah et al. [35], 2022	ECG	Grad-CAM	DenseNet, CNN	Detection of myocardial infarction	Acc: 98.9
Alhusseini et al. [36], 2020	Image Grids	Grad-CAM	CNN (5 convolutional layers)	Classify Intracardiac Electrical Patterns During Atrial Fibrillation	Acc: 95
Chang et al. [37], 2021	Retina Image	AEs	ResNet-50	Glaucoma Decisions	AUC: 90
Kuo et al. [38], 2020	Image	Grad-CAM	VGG16Net InceptionV3 ResNet152	Identifying patients with keratoconus	Acc: 93, sen:91.7, spec: 94.4 Acc: 93, sen:91.7, spec: 94.4 Acc: 95.8, sen:94.4, spec: 99.5
Brunese et al. [39], 2020	X-Ray	Grad-CAM	VGG16Net	Pulmonary Disease and Coronavirus COVID-19 Detection	Acc: 97
Ozturk et al. [40], 2020	Chest X-Ray	Grad-CAM	DarkCovidNet	to detect and classify COVID-19 cases	Acc: 98.08 (binary) Acc: 87.02 (multi)
Shi et al. [41], 2021	CT image X-Ray image	DAM	Attention transfer network	to improve the efficiency and accuracy of COVID-19 diagnosis	F1-score: 88.28, Acc: 87.98
Our study	**ECG Images**	**EigenCAM**	**ViT-B/384**	**Arrhythmia classification**	**F1-score: 96.78, Acc:96.79**
Authors, Year	**Input Types**	**XAI Method**	**Deep Method**	**Tasks**	**Performance (%)**
Hicks et al. [31], 2021	ECG	Grad-CAM	CNN	Sex prediction	Acc:89
Lee et al. [32], 2019	Computed-tomography	Attention map	CNN	detection of acute intracranial haemorrhage	S1: sen:98, spec:95 S2: sen:92, spec:95
Kim et al. [33], 2022	ECG	Grad-CAM	Visual-DenseNet	Arrhythmia detection	Acc: 98.7, sen:98.7, F1: 98.9
Elul et al. [34], 2021	ECG	STA	CNN	Identification of underlying cardio-pathology	Acc: 96
Jahmunah et al. [35], 2022	ECG	Grad-CAM	DenseNet, CNN	Detection of myocardial infarction	Acc: 98.9
Alhusseini et al. [36], 2020	Image Grids	Grad-CAM	CNN (5 convolutional layers)	Classify Intracardiac Electrical Patterns During Atrial Fibrillation	Acc: 95
Chang et al. [37], 2021	Retina Image	AEs	ResNet-50	Glaucoma Decisions	AUC: 90
Kuo et al. [38], 2020	Image	Grad-CAM	VGG16Net InceptionV3 ResNet152	Identifying patients with keratoconus	Acc: 93, sen:91.7, spec: 94.4 Acc: 93, sen:91.7, spec: 94.4 Acc: 95.8, sen:94.4, spec: 99.5

alone. However, it can be compared with rhythms such as multifocal atrial tachycardia, which may also be irregular but have p-waves, to assess the presence of p-waves.

- The speed variability of AFib ECGs allowed the model to perform well in terms of speed compared to GSVT, SB and SR, but reduced the need to focus on p-waves. To facilitate the training process, a comparison of AFib, SR and accelerated junctional rhythm with a rate between 60 and 80 may be considered. This is because although SR and axial junctional rhythm are regular rhythms, SR has prominent p-waves, whereas axial junctional rhythm does not have prominent p-waves.
- In general, we would expect the enhancement pattern to be along the vertical line because the impulse occurs simultaneously in each lead. It would be more logical to concentrate on each of these pulses similarly. In particular, if the concentration is on a wave that creates a difference from pulse to pulse, concentrations along the vertical line (|||||) are expected. However, this concentration pattern may not be valid for waves with different amplitudes in each lead, such as the p-wave. In this case, a horizontal condensation pattern can be expected in the best- observed lead. These findings provide important clues for improving the model and making more accurate diagnoses.

5. Conclusion

In summary, this study illustrates how ViT models show promise in categorizing various ECG rhythms like AFib, GSVT, SB, and SR. The diverse CAM methods utilized offer insights into the model's decision-making process, enhancing the interpretability and dependability of the findings. While the models generally performed well, there were limitations noted when focusing on T–P intervals in cases involving respiratory variations or ectopic beats.

Our best-performing model, ViT-B/384, achieved an overall accuracy of 96.79 %, precision of 96.79 %, recall of 96.79 %, and an F1-score of 96.78 % across the four arrhythmia classes. For specific classes, the model obtained accuracy rates of 98.39 % (AFib), 97.83 % (GSVT), 99.24 % (SB), and 98.11 % (SR), demonstrating robust classification performance. These results indicate the effectiveness of integrating high-resolution ECG image transformation with ViT architectures in multi-class arrhythmia detection.

This research underscores the significance of model interpretability in clinical settings and sets the stage for advancements in automated ECG analysis. By emphasizing p-waves and rhythm patterns, the model's ability to deliver reliable and understandable results can aid in clinical diagnosis processes. This proves advantageous when differentiating between complex rhythms with similar characteristics, assisting clinicians in making precise diagnoses and determining suitable treatments.

Future studies could explore comparisons between AFib and rhythms

like atrial tachycardia (MAT), both exhibiting p-waves. The focus of comparisons should center on P-waves irrespective of rate. A comprehensive comparative analysis might involve SR, AFib, MAT, and accelerated junctional rhythm with rates ranging from 60–100. In these comparisons, selection should be based on pattern rather than rate considerations. While AFib is the most irregular rhythm, MAT is also irregular, but these two rhythms are different. It would not be correct to comment on the pattern alone. Therefore, it should be ensured that the model tends to focus on p-waves in particular, or this tendency should be taught in the model. SR and axial junctional rhythm are also regular rhythms, but they are different from each other. The main difference is that SR has a P-wave, whereas axial junctional rhythm has no P-wave. The most important criterion for distinguishing these four rhythms will be the P-wave, making machine learning easier and more accurate.

However, this study has certain limitations. First, the dataset used originates from a single publicly available ECG database, which may not fully capture the diversity of real-world clinical data. Second, while the model demonstrated strong overall performance, minor misclassifications occurred between rhythm types with similar morphological characteristics (e.g., SR vs. GSVT), suggesting that further refinement with additional annotated data could improve discriminatory capability. Lastly, the study focused on 11-lead ECG data, and performance may vary with other lead configurations or lower-quality signals. Also, incorporating customized modules, hybrid architectures, or attention enhancement mechanisms into the ViT framework will be explored to potentially further improve classification performance and interpretability. Future work will address these limitations by incorporating multi-source clinical datasets, exploring domain adaptation techniques, and integrating additional rhythm classes for broader diagnostic applicability.

CRediT authorship contribution statement

Fatma Murat Duranay: Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Data curation. **Ender Murat:** Validation. **Oğuzhan Katar:** Software. **Yakup Demir:** Supervision. **Ru-San Tan:** Writing – review & editing, Validation. **Özal Yildirim:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology. **U. Rajendra Acharya:** Visualization, Validation, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396, https://doi.org/10.1016/j.compbiomed.2017.08.022.

[2] F. Murat, O. Yildirim, M. Talo, U.B. Baloglu, Y. Demir, U.R. Acharya, Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review, Comput. Biol. Med. 120 (2020) 103726, https://doi.org/10.1016/j.compbiomed.2020.103726.

[3] Ö. Yildirim, A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification, Comput. Biol. Med. (2018), https://doi.org/10.1016/j.compbiomed.2018.03.016.

[4] F. Murat, O. Yildirim, M. Talo, Y. Demir, R.S. Tan, E.J. Ciaccio, U.R. Acharya, Exploring deep features and ECG attributes to detect cardiac rhythm classes, Knowledge-Based Syst (2021), https://doi.org/10.1016/j.knosys.2021.107473.

[5] Y.S. Baek, S.C. Lee, W. Choi, D.H. Kim, OPEN A new deep learning algorithm of 12 - lead electrocardiogram for identifying atrial fibrillation during sinus rhythm, Sci. Rep. (2021) 1–10, https://doi.org/10.1038/s41598-021-92172-5.

[6] D.U. Jeong, K.M. Lim, Convolutional neural network for classification of eight types of arrhythmia using 2D time–frequency feature map from standard 12-lead electrocardiogram, Sci. Rep. 11 (2021) 1–9, https://doi.org/10.1038/s41598-021-99975-6.

[7] U.B. Baloglu, M. Talo, O. Yildirim, R.S. Tan, U.R. Acharya, Classification of myocardial infarction with multi-lead ECG signals and deep CNN, Pattern Recognit. Lett. 122 (2019) 23–30, https://doi.org/10.1016/j.patrec.2019.02.016.

[8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, an image is worth 16X16 words: transformers for image recognition At scale, in: ICLR 2021 - 9th Int. Conf. Learn. Represent, 2021.

[9] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, L. Wang, Vision transformers for image classification: a comparative survey, Technologies 13 (2025), https://doi.org/10.3390/technologies13010032.

[10] M.M. Al Rahhal, Y. Bazi, R.M. Jomaa, A. Alshibli, N. Alajlan, M.L. Mekhalfi, F. Melgani, COVID-19 detection in CT/X-ray imagery using vision transformers, J. Pers. Med. 12 (2022), https://doi.org/10.3390/jpm12020310.

[11] V. Mubonanyikuzo, H. Yan, T.E. Komolafe, L. Zhou, T. Wu, N. Wang, Detection of Alzheimer disease in neuroimages using vision transformers: systematic review and meta-analysis, J. Med. Internet Res. 27 (2025) e62647, https://doi.org/10.2196/62647.

[12] P. Bing, Y. Liu, W. Liu, J. Zhou, L. Zhu, Electrocardiogram classification using TSST-based spectrogram and ConViT, Front. Cardiovasc. Med. 9 (2022), https://doi.org/10.3389/fcvm.2022.983543.

[13] S. Aburass, O. Dorgham, J. Al Shaqsi, M. Abu Rumman, O. Al-Kadi, Vision Transformers in medical imaging: a comprehensive review of advancements and applications across multiple diseases, J. Imaging Informatics Med. (2025), https://doi.org/10.1007/s10278-025-01481-y.

[14] R. Hu, J. Chen, L. Zhou, A transformer-based deep neural network for arrhythmia detection using continuous ECG signals, Comput. Biol. Med. 144 (2022) 105325, https://doi.org/10.1016/j.compbiomed.2022.105325.

[15] O.N. Manzari, H. Ahmadabadi, H. Kashiani, S.B. Shokouhi, A. Ayatollahi, MedViT: A robust vision transformer for generalized medical image classification, Comput. Biol. Med. 157 (2023), https://doi.org/10.1016/j.compbiomed.2023.106791.

[16] C. Che, P. Zhang, M. Zhu, Y. Qu, B. Jin, Constrained transformer network for ECG signal processing and arrhythmia classification, BMC Med. Inform. Decis. Mak. 21 (2021) 1–13, https://doi.org/10.1186/s12911-021-01546-2.

[17] S.R. R, C. Michael, D. Abhishek, V. Ramakrishna, P. Devi, B. Dhruv, Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proc. IEEE Int. Conf. Comput. Vis, 2017.

[18] M. Ennab, H. Mcheick, Advancing AI interpretability in medical imaging: A comparative analysis of pixel-level interpretability and grad-CAM models, Mach. Learn. Knowl. Extr. 7 (2025) 12, https://doi.org/10.3390/make7010012.

[19] Y. Cho, J. myoung Kwon, K.H. Kim, J.R. Medina-Inojosa, K.H. Jeon, S. Cho, S. Y. Lee, J. Park, B.H. Oh, Artificial intelligence algorithm for detecting myocardial infarction using six-lead electrocardiography, Sci. Rep. 10 (2020) 1–10, https://doi.org/10.1038/s41598-020-77599-6.

[20] M.M. M, M.T. R, V.K. V, S. Guluwadi, Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50, BMC Med. Imaging. 24 (2024) 107, https://doi.org/10.1186/s12880-024-01292-7.

[21] J.A. Marmolejo-Saucedo, U. Kose, Numerical grad-cam based explainable convolutional neural network for brain tumor diagnosis, Mob. Networks Appl. 29 (2024) 109–118, https://doi.org/10.1007/s11036-022-02021-6.

[22] V. Sangha, B.J. Mortazavi, A.D. Haimovich, A.H. Ribeiro, C.A. Brandt, D.L. Jacoby, W.L. Schulz, H.M. Krumholz, A.L.P. Ribeiro, R. Khera, Automated multilabel diagnosis on electrocardiographic images and signals, Nat. Commun. 13 (2022), https://doi.org/10.1038/s41467-022-29153-3.

[23] N. Alamatsaz, L. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz, F. Nasimi, A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection, Biomed. Signal Process. Control. 90 (2024) 105884, https://doi.org/10.1016/j.bspc.2023.105884.

[24] C. van Zyl, X. Ye, R. Naidoo, Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP, Appl. Energy. 353 (2024) 122079, https://doi.org/10.1016/j.apenergy.2023.122079.

[25] A. Anand, T. Kadian, M.K. Shetty, A. Gupta, Explainable AI decision model for ECG data of cardiac disorders, Biomed. Signal Process. Control. 75 (2022) 103584, https://doi.org/10.1016/j.bspc.2022.103584.

[26] T. Lindow, I. Palencia-Lamela, T.T. Schlegel, M. Ugander, Heart age estimated using explainable advanced electrocardiography, Sci. Rep. 12 (2022) 1–10, https://doi.org/10.1038/s41598-022-13912-9.

[27] Y.Y. Jo, Y. Cho, S.Y. Lee, J. myoung Kwon, K.H. Kim, K.H. Jeon, S. Cho, J. Park, B. H. Oh, Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram, Int. J. Cardiol. 328 (2021) 104–110, https://doi.org/10.1016/j.ijcard.2020.11.053.

[28] B.M. Maweu, S. Dakshit, R. Shamsuddin, B. Prabhakaran, CEFEs: A CNN explainable framework for ECG signals, Artif. Intell. Med. (2021), https://doi.org/10.1016/j.artmed.2021.102059.

[29] I. Neves, D. Folgado, S. Santos, M. Barandas, A. Campagner, L. Ronzio, F. Cabitza, H. Gamboa, Interpretable heartbeat classification using local model-agnostic explanations on ECGs, Comput. Biol. Med. 133 (2021), https://doi.org/10.1016/j.compbiomed.2021.104393.

[30] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, C. Rakovski, A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients, Sci. Data. (2020), https://doi.org/10.1038/s41597-020-0386-x.

[31] S.A. Hicks, J.L. Isaksen, V. Thambawita, J. Ghouse, G. Ahlberg, A. Linneberg, N. Grarup, I. Strümke, C. Ellervik, M.S. Olesen, T. Hansen, C. Graff, N.H. Holstein-Rathlou, P. Halvorsen, M.M. Maleckar, M.A. Riegler, J.K. Kanters, Explaining deep neural networks for knowledge discovery in electrocardiogram analysis, Sci. Rep. 11 (2021) 1–11, https://doi.org/10.1038/s41598-021-90285-5.

[32] H. Lee, S. Yune, M. Mansouri, M. Kim, S.H. Tajmir, C.E. Guerrier, S.A. Ebert, S. R. Pomerantz, J.M. Romero, S. Kamalian, R.G. Gonzalez, M.H. Lev, S. Do, An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets, Nat. Biomed. Eng. 3 (2019) 173–182, https://doi.org/10.1038/s41551-018-0324-9.

[33] J.K. Kim, S. Jung, J. Park, S.W. Han, Arrhythmia detection model using modified DenseNet for comprehensible grad-CAM visualization, Biomed. Signal Process. Control. 73 (2022) 103408, https://doi.org/10.1016/j.bspc.2021.103408.

[34] Y. Elul, A.A. Rosenberg, A. Schuster, A.M. Bronstein, Y. Yaniv, Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning–based ECG analysis, Proc. Natl. Acad. Sci. U. S. A. 118 (2021) 1–12, https://doi.org/10.1073/pnas.2020620118.

[35] V. Jahmunah, E.Y.K. Ng, R.S. Tan, S.L. Oh, U.R. Acharya, Explainable detection of myocardial infarction using deep learning models with Grad-CAM technique on ECG signals, Comput. Biol. Med. 146 (2022), https://doi.org/10.1016/j.compbiomed.2022.105550.

[36] M.I. Alhusseini, F. Abuzaid, A.J. Rogers, J.A.B. Zaman, T. Baykaner, P. Clopton, P. Bailis, M. Zaharia, P.J. Wang, W.J. Rappel, S.M. Narayan, Machine learning to classify intracardiac electrical patterns during atrial fibrillation: Machine learning of atrial fibrillation, Circ. Arrhythmia Electrophysiol. 13 (2020) E008160, https://doi.org/10.1161/CIRCEP.119.008160.

[37] J. Chang, J. Lee, A. Ha, Y.S. Han, E. Bak, S. Choi, J.M. Yun, U. Kang, I.H. Shin, J. Y. Shin, T. Ko, Y.S. Bae, B.L. Oh, K.H. Park, S.M. Park, Explaining the rationale of deep learning glaucoma decisions with adversarial examples, Ophthalmology 128 (2021) 78–88, https://doi.org/10.1016/j.ophtha.2020.06.036.

[38] B.I. Kuo, W.Y. Chang, T.S. Liao, F.Y. Liu, H.Y. Liu, H.S. Chu, W.L. Chen, F.R. Hu, J. Y. Yen, I.J. Wang, Keratoconus screening based on deep learning approach of corneal topography, Transl. Vis. Sci. Technol. 9 (2020) 1–11, https://doi.org/10.1167/tvst.9.2.53.

[39] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays, Comput. Methods Programs Biomed. 196 (2020) 105608, https://doi.org/10.1016/j.cmpb.2020.105608.

[40] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792, https://doi.org/10.1016/j.compbiomed.2020.103792.

[41] W. Shi, L. Tong, Y. Zhu, M.D. Wang, COVID-19 automatic diagnosis with radiographic imaging: explainable attention transfer deep neural networks, IEEE J. Biomed. Heal. Informatics. 25 (2021) 2376–2387, https://doi.org/10.1109/JBHI.2021.3074893.