ECG-KanFormer: Multi-label ECG Signal Classification Based on KAN and Transformer

Xuanshuo Fu, Lei Kang, and Javier Vazquez-Corral

Computer Vision Center, Barcelona, Spain
Universitat Autònoma de Barcelona, Barcelona, Spain
{xuanshuo, lkang, javier.vazquez}@cvc.uab.es

Abstract. Classification of electrocardiogram (ECG) signals poses substantial challenges due to their complexity and high dimensionality, necessitating sophisticated machine learning techniques for effective multi-class and multi-label classification. Kolmogorov Arnold Networks (KANs), renowned for their prowess in time-series analysis based on the Kolmogorov-Arnold representation theorem, assure that any continuous function can be expressed as a finite sum of functions of one variable. This capability positions KANs as a promising approach for approximating the complex functions characteristic of ECG data. In this paper, we introduce ECG-KanFormer, a novel integration of KANs and transformers, marking the first aggregation of these technologies in existing literature. By integrating KAN modules, which introduce non-linearity into the original linear transformer architecture, we enhance the transformer's ability to manage non-linear dependencies typical in ECG signals. We outline our proposed method where KAN modules replace the linear layers in standard transformers, thereby improving the model's learning capability from intricate ECG signal patterns. Our comprehensive experiments demonstrate that the ECG-KanFormer has achieved the state-of-the-art performance among existing approaches via various metrics, proving its effectiveness in the multi-class multi-label ECG classification task. This study paves a new way for end-to-end methods in ECG signal analysis and highlights the pioneering potential of KAN modules in medical diagnostics.

Keywords: Electrocardiogram · Kolmogorov–Arnold Networks · Transformers· Multi-label Classificatio· , PTB-XL.

1 Introduction

Heart-related ailments are the top cause of death worldwide, responsible for nearly one-third of all fatalities [5]. Electrocardiograms (ECGs), non-invasive tests that monitor the heart's electrical activity, are essential for diagnosing various cardiac issues such as heart attacks, ischemia, and arrhythmias. Despite their importance, manually reading ECG results is complicated, requires significant effort, and is error-prone. Thus, enhancing automatic ECG interpretation methods has become a critical field of medical research.

Deep learning has recently transformed how ECG analysis is conducted, enabling automated detection and classification of cardiovascular diseases. Traditionally, analysis has relied on single-lead ECGs [18, 23], which provide only a basic glimpse of heart activity and are often inadequate for diagnosing complex cardiac issues. In contrast, 12-lead ECGs [21, 4] offer a more detailed view of the heart's electrical activity from multiple angles, significantly improving the data available for diagnostic evaluations.

Convolutional neural networks (CNNs) [12], recurrent neural networks (RNNs) [17], and Transformers [25] are frequently used for multilabel classification of ECG signals, effectively extracting relevant features directly from the data. The uncertainty quantification for ECG classification has been evaluated in [7], showing ensemble-based approaches offer more robust estimates, particularly under external validation and out-of-distribution detection, and emphasizes the benefits of incorporating uncertainty into active learning for improved model performance. However, these models often face challenges in processing the lengthy sequences and intricate patterns characteristic of 12-lead ECGs, and may struggle to synthesize both local and global information across different leads effectively. Recently, the emergence of Kolmogorov Arnold Networks (KANs) modules shows promise in overcoming these issues. KANs improve performance in time-series tasks by more accurately capturing temporal dynamics, thus providing a notable advantage for managing the complex data from multi-lead ECGs. In this paper, we introduce the ECG-KanFormer, a novel model that combines KANs with transformers for enhanced multiclass multilabel ECG classification. This approach is validated using the PTB-XL dataset, a large-scale benchmark for ECG analysis. Our key contributions are as follows:

- We thoroughly investigate and optimize the use of the Short-Time Fourier Transform (STFT) to convert time-series ECG waveforms into the frequency domain, thereby enhancing the representation of signal characteristics and enabling more effective feature extraction for this task.
- We propose ECG-KanFormer, which, to the best of our knowledge, is the first model to integrate KAN into a Transformer architecture specifically designed for 12-lead ECG classification tasks. This model combines the strengths of KANs in approximating complex functions with the architectural advantages of transformers, tailored to address the unique challenges of ECG analysis.
- We conduct comprehensive experiments on the PTB-XL dataset using various metrics, demonstrating state-of-the-art performance in capturing both local variations and global trends within ECG signals. The results highlight the model's potential for practical deployment in clinical environments.

2 Related Work

Recent studies show that end-to-end deep learning methods for ECG classification significantly outperform traditional techniques using handcrafted features and conventional machine learning [11]. Hannun *et al.* [2] developed a deep neural

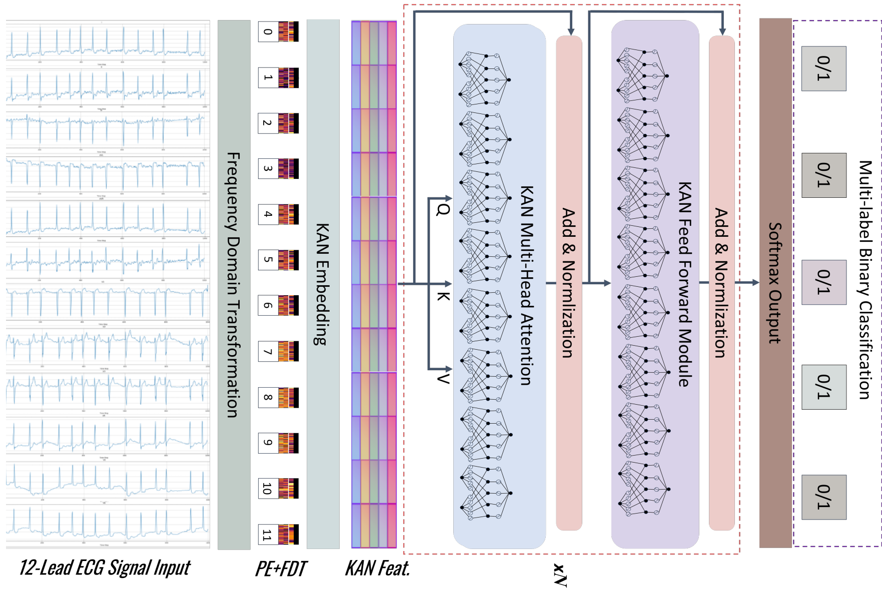

Fig. 1: ECG-KanFormer: The signal is initially processed using frequency domain analysis, after which the coded signal is input into ECG-KanFormer to extract relevant features. Subsequently, the features extracted by ECG-KanFormer are classified using various classification heads. In ECG-KanFormer, we employ KAN for deep integration with Vision Transformers (ViTs).

network (DNN) for classifying ECG rhythms from raw single-lead ECG data, achieving performance comparable to expert cardiologists. Mousavi *et al.* [19] introduced a sequence-to-sequence model for ECG heartbeat classification with strong inter- and intra-patient performance. Murugesan *et al.* [3] combined CNN and BLSTM to leverage ECG signals' hierarchical structure, learning patterns at beat, rhythm, and channel levels, and assessed model interpretability via attention visualizations. Martin *et al.* [17] proposed an LSTM network achieving state-of-the-art accuracy across various datasets and frequencies, suitable for real-time processing on portable devices for timely MI detection. Jing *et al.* [12] introduced a Beat-Level Fusion Network (BLF-Net) using an attention mechanism to weigh heartbeat contributions, enhancing diagnostic accuracy. Zhou *et al.* [25] adapted masked autoencoders for self-supervised representation learning from ECG time series, utilizing a lightweight Transformer as the encoder and a single-layer Transformer for the decoder.

In addition to developing neural network architectures, researchers are exploring additional techniques to enhance model effectiveness. Golany *et al.* [10] integrated physical insights into ECG classification by learning ordinary differential equations describing ECG signal dynamics. Rawi *et al.* [20] used the Tree of Parzen Estimator (TPE) [8] to optimize hyperparameters via a define-by-run approach. Qin *et al.* [6] employed a teacher-student model where the teacher, trained on multi-lead ECG, guides a student model using single-lead ECG, and introduced multi-label disease knowledge distillation (MKD) for multi-label disease diagnosis. Liu *et al.* [15] developed ECG-Text Pre-training (ETP), a framework leveraging cross-modal representations from ECG signals and textual reports for robust performance in linear evaluation and zero-shot classification tasks. Li *et al.* [14] proposed a Multimodal ECG-Text Self-supervised pre-training (METS) model that uses clinical reports to guide ECG self-supervised learning, achieving zero-shot classification by embedding ECG and text representations and calculating their similarity to classify ECG categories.

In summary, current approaches for ECG classification can be categorized into three main types: 1. CNN-based methods. These approaches use convolutional kernels to extract meaningful feature representations from the ECG signals [12]. 2. RNN-based methods. These methods leverage RNNs to address the sequential nature of ECG signals, effectively capturing temporal dependencies [17]. 3. Transformer-based methods. Transformer-based has proved to be effective in sequential data such as handwriting [13], speech [9], and ECG [25]. These models are advantageous due to their ability to capture long-range dependencies and parallelize computations, leading to improved efficiency and accuracy. Vision Transformer (ViT) [1] is a highly successful architecture for various tasks, as it processes images by dividing them into a sequence of patches. For ECG data, waveforms can be transformed into visual-like features using the STFT, making it possible to apply ViT to ECG classification tasks. Additionally, KAN [16] have recently been proposed as an alternative to multilayer perceptrons (MLPs), demonstrating advantages in scalability and in avoiding catastrophic forgetting. Therefore, integrating KAN with ViT for ECG tasks represents a promising research direction with significant potential.

3 Methodology

As shown in Figure 1, the signal is first analyzed in the frequency domain, then fed into ECG-KanFormer for feature extraction. These features are subsequently classified using different classification heads. ECG-KanFormer integrates KAN with ViT for enhanced performance.

3.1 Frequency Domain Transformation and Analysis

As shown in Figure 2, we first analyze the data in the frequency domain. Frequency domain analysis examines a signal's spectrum to identify key frequency components, enhance data compression, and support pattern recognition and diagnostics. In ECG analysis, it effectively assesses heart rate variability, detects arrhythmias, identifies cardiac pathologies, removes noise and artifacts, and improves signal processing, making it a crucial tool for monitoring heart health and diagnosing diseases.

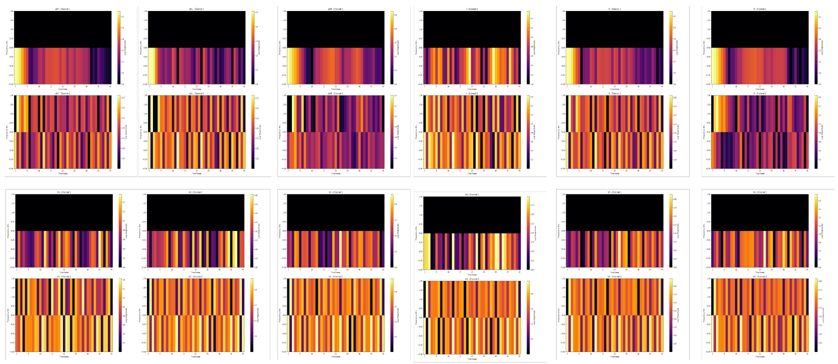

Fig. 2: Frequency Domain Transformation and Analysis: After performing frequency domain analysis, the signals from each lead are displayed in the figure. In this representation, the signal from each lead is decomposed, making the features easier to extract and analyze.

A wide range of signal waveforms can be generated through the modulation of sinusoidal functions. This process is mathematically represented by the following equation:

$$f(t) = Asin(\omega t + \varphi) + k \tag{1}$$

The complete wave signal can be analyzed by examining its amplitude (A), angular velocity (ω)—where angular velocity corresponds to the frequency (f), thus frequency is also commonly used to describe it in signal processing—phase (φ), and offset (k). The sinusoidal functions sine and cosine represent different phases.

According to the Delicacy condition, within each cycle, the signal may contain only consecutive or finitely many Type-I interruptions, including removable discontinuities and jump discontinuities; the number of local maxima and minima must be finite; and the signal must be absolutely integrable over one cycle. Any periodic signal that satisfies the above conditions can be represented by a series of sinusoidal functions in a harmonic relationship, as described by the following equation:

$$f(t) = c_0 + \sum_{n=1}^{\infty} c_n cos(n\omega t + \varphi) + k \tag{2}$$

It is known that integrating a sine wave over the entire real line, from negative infinity to positive infinity, results in zero. This is obtained by considering the principal value and accounting for the periodic nature of the sine wave.

A trigonometric transformation of Equation 2 yields:

$$f(t) = c_0 + \sum_{n=1}^{\infty} [c_n cos\varphi cos(n\omega t) - c_n sin\varphi sin(n\omega t)] \tag{3}$$

Let $a_n = c_n cos\varphi$ and $b_n = -c_n sin\varphi$. Then Equation 3 can be expressed as:

$$f(t) = c_0 + \sum_{n=1}^{\infty}[a_n cos(n\omega t) + b_n sin(n\omega t)] \tag{4}$$

Assuming $f(t)$ is detected using a $sin(k\omega t)$ function, we derive:

$$\int_0^T f(t) sin(k\omega t) dt \tag{5}$$

Assuming $f(t)$ contains sine wave coefficients with $n\omega$ corner frequencies as b_n, according to orthogonality and Equation 5, we have:

$$\int_0^T f(t) sin(k\omega t) dt = b_n \int_0 T sin(n\omega t)^2 dt \tag{6}$$

Furthermore, we can derive:

$$\int_0^T f(t) sin(n\omega t) dt = b_n \frac{T}{2} \tag{7}$$

$$b_n = \frac{2}{T} \int_0^T f(t) sin(n\omega t) dt \tag{8}$$

Similarly, we can obtain a_n.

Therefore, using a_n and b_n, we can determine the amplitude and phase of this wave:

$$c_n = \sqrt{a_n^2 + b_n^2} \tag{9}$$

$$\varphi = arctan(-\frac{b_n}{a_n}) \tag{10}$$

According to Euler's formula:

$$e^{i\theta} = cos\theta + isin\theta \tag{11}$$

We can obtain:

$$F(f(t)) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{12}$$

The ECG signal typically exhibits a smooth and continuous waveform, whose Fourier representation is dominated by low-frequency components due to the rapid decay of high-frequency terms. This spectral concentration prevents large abrupt variations and supports the signal's "delicacy." Moreover, ECG signals are effectively band-limited, with their spectra decaying beyond a certain frequency threshold, thereby avoiding unpredictable high-frequency behavior and ensuring stable and predictable signal characteristics. Since ECGs consist of periodic cardiac cycles, including P waves, QRS complexes, and T waves, they can be decomposed into fundamental sine and cosine components, whose regularity further reinforces structural consistency. In addition, Fourier analysis facilitates the separation and filtering of high-frequency noise while preserving essential signal components. Therefore, the ECG waveform satisfies the Delicacy condition and can be decomposed accordingly, enabling more effective extraction of frequency-domain information and detailed ECG features.

3.2 Kolmogorov–Arnold–Transformer Networks

We proceed with the above frequency domain analysis to deconstruct the ECG signal into different waveforms, considering each waveform as a distinct token.

Different features are extracted by various feature extraction heads of the Transformer, and the extracted features are fused accordingly, as shown in the following equation:

$$Attention(Q, K, V) = SoftMax(\frac{Qk^T}{\sqrt{d}} + B)V \tag{13}$$

within each window during the Multi-Head Self-Attention (MSA) computation, every pixel (or token/patch) undergoes the generation of the corresponding query (Q), key (K), and value (V). Subsequently, the relative position bias is introduced after matching Q and K, followed by division by $\sqrt{d}$ as outlined in Equation 13. At this stage, the process accomplishes both the extraction of global features based on image location and the modeling of dependencies.

Based on the above feature calculations, the feature extraction head is responsible for capturing features across different regions within the ECG signal. Different frequencies of the ECG signal contain critical information about the heartbeat. The feature extraction head aims to simultaneously model global relationships and inter-region features. The specific methodology is outlined as follows:

$$C_{out} = 0.5T(G(X))(1 + \frac{e^{T(G(X))} - e^{-T(G(X))}}{e^{T(G(X))} + e^{-G(X))}}) \tag{14}$$

$$T(X) = \sqrt{\frac{2}{\pi}}(X + 0.044715X^3) \tag{15}$$

$$G(X) = \frac{X - E(X)}{\sqrt{Var(X) + \varepsilon}} * \gamma + \beta \tag{16}$$

where ε denotes a sufficiently small constant, while γ and β represent two trainable hyperparameters. X denotes the input image, $E(\cdot)$ signifies the computed expectation, and $Var(\cdot)$ signifies the computed variance. This phase concludes the more granular modeling of the ECG signal achieved through the use of tokens.

Meanwhile, we further improve the Transformer's extraction and modeling of global temporal features through the use of Kolmogorov–Arnold Networks (KAN). To some extent, this approach fundamentally eliminates the reliance on linear weight matrices by using learnable functions instead of fixed activation functions.

KAN are based on the Kolmogorov-Arnold representation theorem and differ from traditional Convolutional Neural Networks (CNNs) by using learnable activation functions at the edges of the network. This allows each weight parameter in KAN to be replaced by a univariate function, which is typically parameterized as a spline function. This approach provides a high degree of flexibility and the ability to simulate complex functions with fewer parameters, thereby enhancing the interpretability of the model. This is represented in the following equation:

$$f(x) = f(x_1, x_2, ..., x_n) = \sum_{q=0}^{2n} \Phi_q(\sum_{p=1}^{n} \phi_{q,p}(x_p)) \tag{17}$$

The above equation represents the concept of multivariate continuous functions based on the theory that any multivariate continuous function f can be expressed as a combination of a finite number of univariate continuous functions.

The learnable spline function used in KAN is illustrated below:

$$spline(x) = \sum_i c_i B_i(x) \tag{18}$$

which is dominated by the B-spline function. $spline(x)$ denotes the spline function, c_i are the coefficients optimized during training, and $B_i(x)$ represents the B-spline basis functions defined on the grid.

By defining intervals in which each basis function $B_i(x)$ is active and significantly influences the shape and smoothness of the function, the network gains more control. This allows the model to achieve higher accuracy.

The substitutions are represented as KAN layer and MLP layer:

$$KAN(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ ...... \circ \Phi_1 \circ \Phi_0) \tag{19}$$

$$MLP(x) = (W_{L-1} \circ \sigma \circ W_{L-2} \circ \sigma ...... \circ W_1 \circ \sigma \circ W_0 \circ \sigma) \tag{20}$$

In this context, W in MLP represents the weights, and σ denotes the activation function. The counterparts of KAN and MLP are described in the equation above. We use KAN to replace MLP in the overall model.

In summary, we decompose the data, model and features into three distinct stages. First, we analyze ECG signals in the frequency domain to extract waveform signal information at different frequencies, representing the decomposition of the signals. Second, we extract various features from these waveform signals using different feature extraction heads of the Transformer, focusing on specific information for each frequency band; this represents the decomposition of features. Lastly, we utilize KAN to replace the Multilayer Perceptron (MLP) in the model. KAN facilitates the representation of high-dimensional complex functions as combinations of simpler one-dimensional functions, concentrating on optimizing these individual functions rather than the entire multivariate space. This approach reduces the complexity and the number of parameters required for the Transformer, enhancing accuracy and interpretability.

Through triple decomposition, we obtain increasingly detailed information about the model, which enhances the Transformer's ability to capture global features and focus on more precise information points.

3.3 Loss Function

The task addressed in this paper is a multi-categorization problem. We first encode multiple categories using binary codes: for N categories, we use N binary

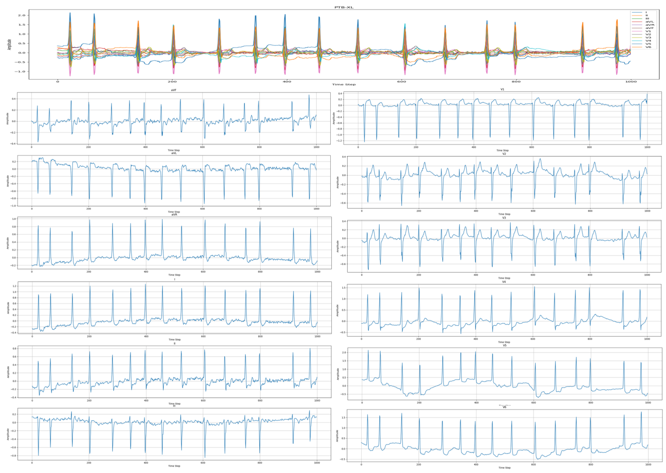

Fig. 3: PTB-XL dataset:Signal data is recorded and stored in a proprietary compressed format. For all signals, we offer 12 sets of standard leads (I, II, III, AVL, AVR, AVF, V1, V2, V3, V4, V5, V6).

bits, where each bit represents a category. A bit value of 0 indicates that the data does not belong to that category, while a bit value of 1 indicates that it does. Each data point contains N subsets of these binary category codes.

For each binary bit, we use Binary Cross-Entropy Loss (BCELoss) to compute the loss, as given by the following formula:

$$BitLoss = -\omega_n[y_n \cdot logx_n + (1 - y_n) \cdot log(1 - x_n)] \tag{21}$$

The total loss is:

$$Loss = \frac{1}{D}\sum_{i=0}^{D}\sum_{n=0}^{N} BitLoss_n \tag{22}$$

where D represents the number of data points, and N represents the number of categories. For the data to be correctly classified, all categories contained within each data point must be accurately categorized, making our classification requirements more stringent.

4 Experimentation

4.1 Data Description

ECG is a crucial non-invasive diagnostic tool for cardiovascular diseases, increasingly augmented by machine learning-based algorithms. The PTB-XL dataset(As shown in figure 3) includes 21,837 clinical 12-lead ECG recordings from 18,885 patients, each with a duration of 10 seconds. The patient cohort consists of 52%

Table 1: Statistics of PTB-XL dataset on five superclasses.

Number of Records	Class	Description
7185	NORM	Normal ECG
3232	CD	Myocardial Infarction
3064	STTC	ST/T Change
2936	MI	Conduction Disturbance
815	HYP	Hypertrophy

Table 2: Performance comparison among the state-of-the-art approaches (best results in bold).

Method	Year	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
VGG16 [22]	2014	62.14	–	–	**90.93**
Seq2Seq [19]	2019	84.19	–	–	86.54
DMSFNet [24]	2020	62.51	–	–	90.72
LSTM [17]	2021	84.17	78.54	**78.37**	90.3
MVKT-ECG [6]	2023	82.2	–	–	84.3
METS [14]	2024	84.2	69.4	62.6	–
ETP [15]	2024	–	–	–	83.5
ECG-KanFormer	–	**87.32**	**78.86**	64.07	90.47

male and 48% female, with an age range from 0 to 95 years (median age 62, interquartile range 22). Up to two cardiologists annotated the ECG waveform data, resulting in a multilabeled dataset where diagnostic labels were aggregated into superordinate and subcategories. The dataset encompasses a broad spectrum of diagnostic categories, including a significant proportion of healthy records. Additionally, the dataset can be augmented with metadata on demographics, other diagnostic statements, diagnostic likelihoods, manually annotated signal attributes, and suggested methods for partitioning training and test sets.

As shown in Table 1, the dataset is organized into five broad disease categories, with each ECG recording potentially associated with multiple categories. We use ECG-KanFormer to predict diagnoses for each ECG based on the extracted diagnostic labels. Performance is evaluated using accuracy, precision, recall, and macro Area Under the ROC Curve (AUC).

4.2 Metrics

In this paper, we employ four metrics to assess the effectiveness of our method: Accuracy, Precision, Recall and macro averaging AUC. The definition of Accuracy, Precision and Recall are given by the following equations:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{23}$$

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

Table 3: Performance across various hyperparametric configuration.

Layer	N_FFT	Parameter(hop, win, aug)	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
3	512	16 64 (0.0, 1.0)	86.50	78.67	59.71	89.39
3	512	16 64 (0.0, 0.8)	**86.88**	**82.09**	58.23	89.69
6	1024	16 64 (0.2, 0.2)	**87.32**	**78.86**	**64.07**	**90.47**
6	1024	16 64 (0.5, 0.5)	86.87	77.90	62.58	90.38
8	100	25 100 (0.0, 0.0)	86.15	77.07	60.27	89.08
8	512	25 64 (0.0, 0.8)	86.38	79.49	58.56	88.96
8	1024	25 64 (0.0, 0.8)	86.50	77.25	61.23	89.30

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

where TP (True Positives), FP (False Positives), TN (True Negatives), and FN (False Negatives) represent the counts of each respective outcome in the classification tasks. This metric quantifies the overall proportion of correct predictions (both true positives and true negatives) out of the total number of cases examined. All the Accuracy, Precision, Recall and AUC are metrics where higher values indicate better performance.

4.3 Implementation Details

In this work, we set the hop size P_h to 16, window size P_w to 64, number of layers P_l to 6, and number of heads P_h to 8. Data augmentation uses probabilities of 0.2 for lead shuffling and 0.2 for masking one-tenth of the signals. We train the model on an NVIDIA A40 using Adam with learning rate 2×10^{-4}, betas $(0.9, 0.98)$, and $\epsilon = 10^{-9}$. The learning rate decays every 100 iterations with $\gamma = 0.8$.

4.4 Experimental Analysis

As shown in Table 2, the proposed ECG-KanFormer outperforms many existing methods, mainly owing to the deep integration of the Transformer architecture with the distinctive modeling capability of KAN. Unlike RNN-based models such as LSTM and Seq2Seq, which process sequences step by step and are therefore limited in efficiency, scalability, and long-range dependency modeling, Transformers enable parallel sequence processing and use self-attention to capture global dependencies without suffering from severe vanishing-gradient issues. Meanwhile, KAN differs from conventional MLPs by replacing fixed linear weight matrices and fixed activation functions with learnable univariate functions, typically parameterized by splines. Incorporating KAN into the Transformer framework therefore enhances the model's flexibility in representing complex local and global relationships in long ECG sequences.

Compared with LSTM and Seq2Seq, ECG-KanFormer achieves superior overall performance. Although LSTM can model temporal dependencies through forget, input, and output gates and obtains an accuracy of 84.17% with balanced precision and recall, its sequential computation limits training and inference efficiency, especially for long signals. Seq2Seq achieves a similar accuracy of 84.19%, but its RNN-based structure still restricts its ability to capture long-range dependencies. In contrast, ECG-KanFormer benefits from multi-head self-attention, which allows the model to attend to different parts of the ECG signal simultaneously, and from learnable spline functions, which further improve its ability to model diverse dependencies. This is particularly important for ECG analysis, where clinically meaningful components such as P waves, QRS complexes, and T waves may contribute differently to diagnosis. By assigning adaptive importance to these waveform segments, ECG-KanFormer can extract more discriminative features, especially from long and noisy ECG sequences.

For medical diagnosis tasks such as ECG classification, balancing precision and recall is crucial. High recall helps reduce missed diagnoses, while high precision reduces false positives, unnecessary examinations, and patient anxiety. ECG-KanFormer maintains a relatively high precision, indicating fewer false-positive predictions than previous methods. Although its recall is slightly lower than that of LSTM, it still identifies a substantial proportion of true positives, suggesting a more cautious and reliable prediction behavior. Furthermore, the AUC of 90.47% demonstrates its strong class-discrimination ability, which is particularly valuable for imbalanced medical datasets, as it indicates stable performance across different decision thresholds.

As shown in Table 2, we have compared our experiments with popular backbone networks from recent years as well as the primary temporal signal analysis networks. The results in Table 2 demonstrate that ECG-KanFormer ranks first in terms of both accuracy and precision. We outperform the second-place METS [14] by 3.12% in accuracy and the second-place LSTM [17] by 0.32% in precision.

Our AUC is 90.47%, ranking third. This is significantly higher than the 84.3% achieved by the main ECG timing signal analysis algorithm MVKT-ECG [6] and the 83.5% achieved by ETP [15] in the last two years, representing improvements of 6.71% and 6.97%, respectively. Just 0.46% from the highest value.

Our model's high accuracy and precision, combined with its low AUC, and recall, may indicate a performance bias due to data imbalance. Specifically, the model appears to perform well on the majority classes (those with higher frequencies), resulting in high overall accuracy and precision, but struggles with the minority classes (those with lower frequencies), leading to low recall. A low AUC suggests that the model does not discriminate effectively across all classes, while low recall implies that the model fails to correctly identify all samples from the positive classes.

In the next step of the research to further improve our model, we can consider resampling the data (e.g., oversampling a few classes or undersampling a majority of classes), adjusting the decision threshold of the model, using a weighted loss function to balance the importance of each class, or integrating methods to better improve the overall performance and robustness of the model.

Table 4: Ablation study comparing ViT performance w/ and w/o the KAN module replacing the MLP.

KAN	Acc. (%)	Prec. (%)	Rec. (%)	AUC (%)
–	86.09	73.78	58.06	87.81
✓	**87.32**	**78.86**	**64.07**	**90.47**
Improvement	1.23	5.08	6.01	2.66

As shown in Table 3, we have investigated and analyzed ECG-KanFormer with various parameters and layer configurations. This research explores the impact of different model layers, N_FFT (number of FFT points), hop size, window size, and augmentation techniques on the network's performance. The aim is to understand how these parameters and network depths affect the model's effectiveness.

We find that varying network depths affect the fitting of categorized ECG data, with deeper networks tending to overfit more quickly. The choice of N_FFT influences the richness of the features extracted by the model, while the interplay between hop size and window size affects the coding length of the initial input to the network.

In our exploration of model depth and parameters, we discovered a dynamic equilibrium between model depth, N_FFT, and other factors when analyzing ECG signals. By homomorphic adjustment, we can balance the complexity between the model and the features. Reducing model complexity necessitates increasing feature complexity, and vice versa.

4.5 Ablation Study

As shown in Table 4, we perform an ablation study comparing KAN-enhanced ViT models at different depths with the original ViT. Our model consistently outperforms ViT across accuracy, precision, recall, and AUC, improving precision by 5.08%, AUC by 2.66%, accuracy by 1.23%, and recall by 6.01%.

The effectiveness of ECG-KanFormer for time-series data analysis, particularly for ECG data classification, is demonstrated. The deep integration of KAN with ViT proves to be highly effective, highlighting that this approach is a promising avenue with considerable potential for further exploration.

5 Conclusion

In this paper, we propose a novel ECG-KanFormer network architecture, which, to our knowledge, is the first instance where KAN are deeply integrated with ViT for ECG signal analysis. This approach utilizes frequency domain analysis to transform waveforms into visual-like features, allowing the ViT structure to extract long-range global features. Furthermore, the deep integration of KAN with ViT demonstrates advantages in scalability and interpretability. We believe that ECG-KanFormer has significant potential for application across a wide range of sequential data tasks. In the future, we plan to incorporate expert cardiologists' knowledge as rules within the model to enhance its performance and interpretability in medical applications.

6 Acknowledgements

This work has been partially supported by the predoctoral program AGAUR-FI ajuts (2026 FI-3 00470) Joan Oró, which are backed by the Secretariat of Universities and Research of the Department of Research and Universities of the Generalitat of Catalonia, as well as the European Social Plus Fund; the Beatriu de Pinós del Departament de Recerca i Universitats de la Generalitat de Catalunya (2022 BP 00256), and the Grant PID2024-162555OB-I00 funded by MCIN/AEI/10.13039/501100011033, ERDF "A way of making Europe". JVC also acknowledges the 2025 Leonardo Grant for Scientific Research and Cultural Creation from the BBVA Foundation.

References

1. Dosovitskiy et al., A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
2. Hannun et al., A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine **25**(1), 65–69 (2019)
3. Murugesan et al., B.: Ecgnet: Deep network for arrhythmia classification. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA). pp. 1–6. IEEE (2018)
4. Park et al., J.: Study on the use of standard 12-lead ecg data for rhythm-type ecg classification problems. Computer Methods and Programs in Biomedicine **214**, 106521 (2022)
5. Lindstrom et al., M.: Global burden of cardiovascular diseases and risks collaboration, 1990-2021. Journal of the American College of Cardiology **80**(25), 2372–2425 (2022)
6. Qin et al., Y.: Mvkt-ecg: Efficient single-lead ecg classification for multi-label arrhythmia by multi-view knowledge transferring. Computers in Biology and Medicine **166**, 107503 (2023)
7. Barandas, M., Famiglini, L., Campagner, A., Folgado, D., Simão, R., Cabitza, F., Gamboa, H.: Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram. Information Fusion **101**, 101978 (2024)
8. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. Advances in neural information processing systems **24** (2011)
9. Chang, O., Liao, H., Serdyuk, D., Shahy, A., Siohan, O.: Conformer is all you need for visual speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 10136–10140. IEEE (2024)
10. Golany, T., Freedman, D., Radinsky, K.: Ecg ode-gan: Learning ordinary differential equations of ecg dynamics via generative adversarial learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 134–141 (2021)

11. Hong, S., Zhou, Y., Shang, J., Xiao, C., Sun, J.: Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Computers in biology and medicine **122**, 103801 (2020)
12. Jing, J., Zhang, J., Liu, A., Gao, M., Qian, R., Chen, X.: Ecg-based multiclass arrhythmia classification using beat-level fusion network. Journal of Healthcare Engineering **2023**(1), 1755121 (2023)
13. Kang, L., Riba, P., Rusiñol, M., Fornés, A., Villegas, M.: Pay attention to what you read: non-recurrent handwritten text-line recognition. Pattern Recognition **129**, 108766 (2022)
14. Li, J., Liu, C., Cheng, S., Arcucci, R., Hong, S.: Frozen language model helps ecg zero-shot learning. In: Medical Imaging with Deep Learning. pp. 402–415. PMLR (2024)
15. Liu, C., Wan, Z., Cheng, S., Zhang, M., Arcucci, R.: Etp: Learning transferable ecg representations via ecg-text pre-training. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8230–8234. IEEE (2024)
16. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
17. Martin, H., Morar, U., Izquierdo, W., Cabrerizo, M., Cabrera, A., Adjouadi, M.: Real-time frequency-independent single-lead and single-beat myocardial infarction detection. Artificial intelligence in medicine **121**, 102179 (2021)
18. Mathews, S.M., Kambhamettu, C., Barner, K.E.: A novel application of deep learning for single-lead ecg classification. Computers in biology and medicine **99**, 53–62 (2018)
19. Mousavi, S., Afghah, F.: Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach. In: IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 1308–1312. IEEE (2019)
20. Rawi, A.A., Elbashir, M.K., Ahmed, A.M.: Deep learning models for multilabel ecg abnormalities classification: A comparative study using tpe optimization. Journal of Intelligent Systems **32**(1), 20230002 (2023)
21. Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P., Andersson, C.R., Macfarlane, P.W., Meira Jr, W., et al.: Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications **11**(1), 1760 (2020)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
23. Varon, C., Morales, J., Lázaro, J., Orini, M., Deviaene, M., Kontaxis, S., Testelmans, D., Buyse, B., Borzée, P., Sörnmo, L., et al.: A comparative study of ecg-derived respiration in ambulatory monitoring using the single-lead ecg. Scientific reports **10**(1), 5704 (2020)
24. Wang, R., Fan, J., Li, Y.: Deep multi-scale fusion neural network for multi-class arrhythmia detection. IEEE journal of biomedical and health informatics **24**(9), 2461–2472 (2020)
25. Zhou, Y., Diao, X., Huo, Y., Liu, Y., Fan, X., Zhao, W.: Masked transformer for electrocardiogram classification. arXiv preprint arXiv:2309.07136 (2023)