International Journal of Advanced Engineering and Technology Research
ISSN: 3106-8677 (Print), ISSN: 3106-8685 (Online) | Vol. 1, No. 1, 2026

FEENet: A Frequency-Enhanced EC Network for Cardiovascular Disease Detection

Tingfeng Liu *, Han Wen

College of Computer Science and Artificial Intelligence, Southwest Minzu University, Chengdu 610000, China
* **Corresponding author**

Abstract: With the development of society, the incidence of cardiovascular diseases has continued to rise. As an important non-invasive diagnostic tool, electrocardiography (ECG) has been widely used in clinical screening and auxiliary diagnosis. In recent years, deep learning-based automated ECG analysis methods have attracted considerable attention; however, most existing approaches rely on single time-domain features, making it difficult to fully exploit the discriminative information of ECG signals in both time and frequency domains. Therefore, this study proposes a Frequence-Enhance ECG Network (FEENet), which performs multi-scale feature extraction on both temporal ECG signals and their frequency-domain representations. A Transformer-based causal encoder (TCE) is then employed to model the relationships among features at different scales. Subsequently, a time–frequency cross-attention (TFCA) module is introduced to enable bidirectional interaction and deep fusion between temporal and frequency-domain features. Finally, a classification layer is used to produce accurate cardiovascular disease classification results. The proposed method achieves an accuracy of 89.6% and an F1-score of 77.4% on the PTB-XL dataset, demonstrating its effectiveness and strong classification capability in complex ECG classification tasks.

Keywords: Frequence-Enhance; Feature Pyramid; Transformer; Cardiovascular Disease; Electrocardiograph.

1. Introduction

With the acceleration of population aging and changes in lifestyle, the incidence of cardiovascular disease (CVD) has been steadily increasing, making it a major public health issue that seriously threatens human health. Electrocardiography (ECG), as an important non-invasive clinical tool for CVD diagnosis, is widely used for the detection of conditions such as arrhythmia and coronary heart disease [1]. However, long-term ECG monitoring generates a large volume of data, and manual interpretation is time-consuming and subject to inter-observer variability, making it difficult to meet clinical requirements for accuracy and real-time performance. Therefore, developing efficient and reliable automated ECG analysis algorithms is of significant research and practical importance.

In recent years, deep learning has been widely applied in the field of ECG analysis due to its end-to-end learning capability. Shunxiang Yang et al. [2] proposed a multi-view and multi-scale deep neural network to address the heterogeneity among the 12-lead ECG signals, integrating attention mechanisms and knowledge distillation, which significantly improved performance in multi-label classification tasks. Haozhen Zhang et al. [3] designed a multi-scale deep residual network combining 1D and 2D convolutions to effectively capture long-term temporal dependencies and inter-lead relationships, leading to notable performance gains. Apoorva et al. [4] incorporated Inception modules, residual connections, and channel attention mechanisms to enhance robustness and interpretability under different lead configurations and complex clinical scenarios. Nima et al. [5] focused on multi-label ECG classification under reduced-lead conditions and proposed a multi-branch convolutional neural network that integrates time-domain and frequency-domain features, along with a thresholding strategy to mitigate class imbalance, demonstrating strong performance across varying lead settings and noise conditions.

In summary, existing studies have achieved notable progress in automated ECG classification, multi-label modeling, and multi-lead feature extraction. However, there is still room for improvement in joint time–frequency modeling and multi-scale information fusion. To address these limitations, this paper proposes a Frequency-Enhanced ECG Network (FEENet) for multi-label cardiovascular disease classification. The proposed method takes both time-domain and frequency-domain features of ECG signals as dual-branch inputs, and integrates multi-scale feature extraction with a cross-attention-based feature interaction mechanism to achieve effective fusion of time–frequency information and enhance feature representation. As a result, the model enables accurate multi-label classification of cardiovascular diseases. Experimental results on the PTB-XL public dataset [6] demonstrate that the proposed method achieves competitive performance compared with several state-of-the-art approaches, attaining an accuracy of 89.6% and an F1 score of 77.4%.

2. Method

(1) Overall

The overall architecture of the proposed FEENet is illustrated in Fig. 1. The ECG time-domain signal and its corresponding frequency-domain representation are first fed into a five-layer feature pyramid to perform multi-scale feature extraction. The features at different scales are then processed by a Transformer Causal Encoder (TCE) and concatenated to form a unified multi-scale representation.

Subsequently, linear layers are applied to project the time-domain and frequency-domain features into the same dimensional space, which are then fed into a Time–Frequency Cross Attention (TFCA) module to capture the correlations between the two domains. The fused features generated by TFCA are concatenated with the original time-domain and frequency-domain features along the channel dimension, resulting in two feature maps with shape (Batch_size, 2C, L), where C denotes the number of channels and L represents the feature length.

These two sets of concatenated features are further modeled by TCE modules, and then jointly fed into a second TFCA module for deeper time–frequency feature fusion. Finally, the fused features are passed through a Softmax layer to produce the classification results.

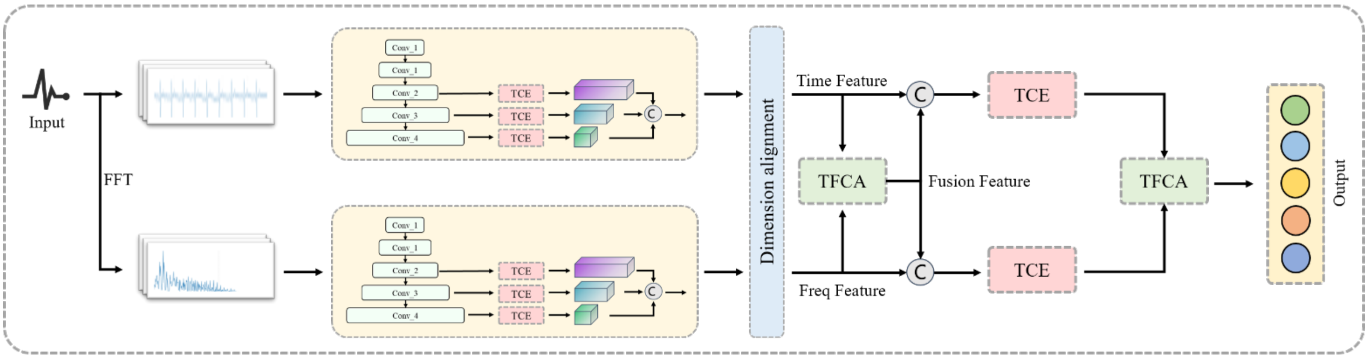

Fig.1 Overall Architecture of FEENet

(2) Preprocessing

The PTB-XL 12-lead ECG dataset is adopted as the experimental data in this study. During the data preprocessing stage, the raw data are first cleaned: samples with complete labels but missing values are padded with zeros, while samples with missing labels are directly removed. Subsequently, the cleaned ECG signals are transformed into the frequency domain using Fast Fourier Transform (FFT), and only the magnitude of the positive frequency components is retained as the frequency-domain representation. Finally, the processed time-domain signals and corresponding frequency-domain features are jointly fed into FEENet for model training.

(3) Feature pyramid

The feature pyramid consists of five layers, and the structure of each layer is illustrated in Fig. 2. Each layer is composed of three convolutional blocks, where the last block in each layer is embedded with a Squeeze-and-Excitation (SE) attention module [7] to enhance feature representation. A pooling layer is applied at each stage for down sampling. All convolutional layers adopt a unified configuration with kernel size 3, stride 1, and padding 1.

The SE module is composed of an average pooling layer, a multilayer perceptron (MLP), and a Sigmoid activation function. Specifically, given an input feature $X \in R^{C\times H\times W}$, global average pooling is first applied to obtain channel-wise statistics $z \in R^{C\times 1\times 1}$,. The vector z is then passed through a bottleneck structure consisting of a dimensionality reduction layer followed by a dimensionality expansion layer, as shown in Eq. (1):

$$W = \sigma\left(W2\delta\left(W1(z)\right)\right) \tag{1}$$

where $W_1 \in R^{\frac{C}{r}\times C}, W_2 \in R^{C\times\frac{C}{r}}$ denote the weight matrices for dimensionality reduction and expansion, respectively. The reduction ratio r is set to 16 in this study. $\delta(\cdot)$ represents the ReLU activation function, and $\sigma(\cdot)$ denotes the Sigmoid function. The resulting channel weights W are applied to the input feature X via channel-wise multiplication to achieve adaptive feature recalibration.

It is worth noting that the feature sequences from the first and second convolutional blocks contain considerable redundancy and incur higher computational cost. Therefore, only the feature sequences from the third to the fifth convolutional blocks are retained for subsequent processing, achieving a balance between computational efficiency and representation capability.

(3) Time-Frequence Cross Attention

To fully capture the correlation between ECG time-domain signals and their frequency-domain representations, a Time–Frequency Cross Attention (TFCA) module is designed to model deeper interactions between the two domains. In this module, the time-domain features are treated as the query Q_t, while the frequency-domain features are used as the key K_f and value V_f. The attention weights are computed to adaptively enhance both time-domain and frequency-domain representations.

Specifically, three groups of one-dimensional causal convolutions with kernel size 1 are employed to project the time-domain and frequency-domain features into the corresponding representations, yielding Q_t, K_f, and V_f. Subsequently, a scaled dot-product attention mechanism is applied to compute the correlation between time-domain and frequency-domain features along the temporal dimension, as formulated in Eq. (2):

$$\mathrm{Attention}\left(Q_t, K_f, V_f\right) = \mathrm{Softmax}\left(\frac{Q_t K_f^{\top}}{\sqrt{d}}\right) V_f \tag{2}$$

The extracted features are mapped through a 1×1 convolution, followed by batch normalization and a nonlinear activation, and then combined with the original time-domain features via a residual connection to obtain the final output.

3. Experiment

(1) Dataset

The PTB-XL dataset consists of 12-lead clinical electrocardiogram (ECG) records acquired using Schiller AG devices between October 1989 and June 1996, and was publicly released in 2020. The dataset comprises 21,799 10-second recordings from 18,869 patients, available at sampling rates of both 500 Hz and 100 Hz. Diagnostic labels in PTB-XL are organized into a hierarchical structure of super classes and subclasses. The superclass labels categorize the ECGs into five major types: Normal (NORM), Myocardial Infarction (MI), ST/T Change (STTC), Conduction Disturbance (CD), and Hypertrophy (HYP). These superclass diagnostic labels have been extensively utilized in multi-label ECG classification research. In this study, 369 samples lacking labels were excluded, resulting in a final experimental set of 21,430 labeled samples with a sampling rate of 100 Hz.

(2) Metrics

This study employs Precision (PR), Recall (RE), F1-Score (F1), and Accuracy (ACC) to evaluate the models performance. The calculation methods for PR, RE, F1, and

ACC are defined as follows:

$$PR = \frac{TP}{TP + FP} \tag{3}$$

$$RE = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \times PR \times RE}{PR + RE} \tag{5}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Where TP, FP, TN, and FN represent True Positives, False Positives, True Negatives, and False Negatives, respectively. Specifically, PR indicates the proportion of actual positive labels among samples predicted as positive, RE reflects the ratio of correctly predicted positive samples to the total actual positive samples, and F1 serves as a harmonic mean to provide a comprehensive evaluation of PR and RE.

(3) Setting

This research is implemented based on the PyTorch 2.0.1 framework. The training process adopts the following configurations: Focal Loss is employed as the loss function, and AdamW is selected as the optimizer with parameters set to $\beta_1 = 0.9, \beta_2 = 0.999$ and a weight decay of 1×10^{-2}. The initial learning rate is set to 1×10^{-3}, which gradually decreases to 1×10^{-5} using a cosine annealing strategy. The dataset is partitioned into training, validation, and test sets with a ratio of 7:2:1. The training is conducted for 50 epochs with a fixed batch size of 64. All experiments are performed on a platform running Ubuntu 18.04, utilizing an NVIDIA GTX 4090 GPU and CUDA V11.8.89.

(4) Results

In this section, the proposed method is compared with existing state-of-the-art methods on the PTB-XL dataset across four metrics: ACC, PR, RE, and F1, to verify the models effectiveness. Subsequently, ablation studies are conducted to demonstrate the efficacy of the proposed approach in the multi-label classification of cardiovascular diseases.

Table 1. Overall performance comparison between FEENet and existing methods

Model	ACC(%)	PR(%)	RE(%)	F1(%)
[8]	82.2	/	/	62.6
[9]	62.7	/	/	73.6
[10]	/	75.4	77.8	76.4
FEENet	89.6	80.0	75.1	77.4

Table 1 presents the overall performance comparison of various models for the 5-class multi-label classification task on the PTB-XL dataset. Overall, the proposed FEENet achieves state-of-the-art results across all metrics except for Recall, reaching 89.6%, 80.0%, and 77.4% in ACC, PR, and F1, respectively. Compared to existing methods, FEENet shows a significant improvement in ACC, indicating a stronger discriminative capability in overall prediction accuracy. The enhancement in PR further demonstrates that the model maintains a lower false-positive rate when identifying positive categories. Although the Recall of FEENet (75.1%) is slightly lower than that of the method proposed by Xiao Q et al.[10] (77.8%), the marginal gap does not undermine FEENets comprehensive performance advantage in terms of F1-score.

Considering the characteristics of multi-label cardiovascular disease classification, the PTB-XL dataset presents significant challenges, including label co-occurrence and class imbalance, which impose stringent requirements on the models feature representation and generalization capabilities. The superior performance of FEENet across multiple metrics demonstrates its enhanced capacity for feature extraction and label correlation modeling. This indicates that FEENet can more effectively capture discriminative information within ECG signals and uncover latent associations between multiple co-existing diseases.

Table 2. Category-specific performance comparison on the PTB-XL dataset

Model	NORM	MI	STTC	CD	HYP
[11]	84.4	71.3	75.0	75.1	45.6
[12]	85.5	74.0	75.7	76.4	50.9
[9]	86.2	74.5	77.4	76.5	53.3
[10]	86.8	76.8	80.0	77.8	60.5
FEENet	86.9	76.6	76.2	80.9	66.3

Table 2 further illustrates the classification performance of various models across the five categories in the PTB-XL dataset. FEENet achieves the best results in the NORM, CD, and HYP categories. Notably, its advantages in the CD and HYP categories are more pronounced, reaching 80.9% and 66.3% respectively—an improvement of 3.1% and 5.8% over the method by Xiao Q et al.[10], reflecting superior category-specific discriminative power.

It is worth noting that the HYP category is a relatively minority class in the PTB-XL dataset, accounting for only 12% of the total samples, which highlights a significant class imbalance issue. The fact that FEENet achieves the highest performance (66.3%) on the HYP category, outperforming all other methods, suggests that FEENet possesses greater robustness in handling class imbalance. It can effectively mine key discriminative features from minority samples, thereby reducing the risk of missed detections for rare conditions.

(5) Ablation

To verify the effectiveness of the frequency-domain branch and the TFCA module within FEENet, systematic ablation studies were conducted on the PTB-XL dataset, with the results summarized in Table 3.

From the perspective of overall metrics, the baseline model (without the frequency-domain signal or TFCA) achieved an ACC of 88.2% and an F1-score of 72.9%. Upon introducing the frequency-domain signal, the ACC increased by 1.0% and the F1 improved by 2.2%, suggesting that frequency-domain features effectively supplement temporal information and enhance the models discriminative power. When only the TFCA module was incorporated, the ACC rose to 89.2% and the F1 reached 74.5%, demonstrating that TFCA improves feature representation quality by strengthening feature interactions. The model reached its peak performance when both the frequency-domain branch and the TFCA module were integrated, yielding an ACC of 89.6% and an F1 of 77.4%. This outperforms all single-module configurations, illustrating a superior synergistic enhancement effect between frequency-domain modeling and the cross-feature fusion mechanism.

From the perspective of category-specific F1-scores, the full model demonstrates consistent improvements across most categories, with the most prominent performance gain observed in the minority class HYP. Specifically, the F1-score for the HYP category gradually increased from 56.1% in the baseline model to 66.3%, marking the most significant improvement. Due to the low sample proportion of the HYP category in the dataset, the class imbalance issue is particularly acute, often making it difficult for traditional models to fully learn its discriminative features. However, by incorporating the frequency-domain branch and the TFCA module, the model more effectively captures the waveform variations induced by HYP and reinforces key diagnostic information through the attention weighting mechanism, thereby enhancing its discriminative capability for the HYP category.

Table 3. Ablation studies of different modules in FEENet

Frequence	TFCA	ACC	F1	F1 of per classes				
				NORM	MI	STTC	CD	HYP
/	/	88.2	72.9	83.6	75.4	72.8	77.0	56.1
√	/	89.2	75.1	86.5	75.9	74.8	77.5	60.9
/	√	89.2	74.5	86.6	77.4	72.2	77.0	59.0
√	√	89.6	77.4	86.9	76.6	76.2	80.9	66.3

In summary, the frequency-domain branch in FEENet provides the model with complementary spectral information distinct from temporal waveform features, enabling the capture of disease-related rhythmic structures and energy distribution characteristics. Simultaneously, the TFCA module enhances the expressive power of critical information through adaptive feature selection and cross-feature interaction mechanisms. The synergistic modeling of time and frequency domains, combined with the dynamic feature fusion mechanism, effectively strengthens the model's ability to distinguish complex and minority categories, achieving a synchronized improvement in both overall performance and class balance.

4. Conclusion

In the automated diagnosis of cardiovascular diseases, multi-label classification requires the simultaneous characterization of inter-disease correlations and complex signal structures, which places higher demands on a models feature representation capabilities. Inspired by the paradigm of joint time-frequency modeling, this paper proposes FEENet, a multi-label classification model that integrates frequency-domain features with a cross-modal feature attention mechanism. By constructing a dual-branch architecture (time-frequency domains) and introducing the TFCA module to facilitate interaction and fusion between heterogeneous features, the models ability to represent complex pathological patterns is significantly enhanced. Experimental results on the 5-class multi-label task of the PTB-XL dataset demonstrate that FEENet achieves superior performance in terms of overall ACC, Precision, and F1-score. Notably, the substantial improvement in the minority class HYP validates the effectiveness and stability of the proposed method. This approach aims to improve the accuracy and class balance of automated cardiovascular disease identification, providing more reliable decision support for clinical computer-aided diagnosis.

Despite the promising results in overall performance and minority class recognition, there remains room for further improvement in specific categories, such as STTC. Future work will focus on integrating more refined imbalance learning strategies and exploring robust multi-modal joint modeling methods to further enhance the models generalization capability and practical application value in complex clinical scenarios.

References

[1] Alberdi A , Aztiria A , Basarab A .Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review[J].Journal of Biomedical Informatics, 2016, 59:49-75.

[2] Yang S , Lian C , Zeng Z ,et al. A Multi-view Multi-scale Neural Network for Multi-label Ecg Classification [J].IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(3):648-660.

[3] Zhang H , Zhao W , Liu S .SE-ECGNet: A Multi-scale Deep Residual Network with Squeeze-and-Excitation Module for ECG Signal Classification[J]. 2020, 2685-2691.

[4] Srivastava A, Pratiher S, Alam S, et al. A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG[J]. Physiological Measurement, 2022, 43(6): 064005.

[5] Wickramasinghe N L , Athif M .Multi-label classification of reduced-lead ECGs using an interpretable deep convolutional neural network[J].Physiological Measurement, 2022, 43(6):14.

[6] Wagner P , Strodthoff N , Bousseljot R D ,et al.PTB-XL, a large publicly available electrocardiography dataset [J].Scientific Data, 2020, 7(1):154.

[7] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[8] Qin Y Z, Sun L, Chen H, et al. MVKT-ECG: Efficient single-lead ECG classification for multi-label arrhythmia by multi-view knowledge transferring[J]. Computers in Biology and Medicine, 2023, 166: 107503.

[9] Jyotishi D, Dandapat S. An attentive spatio-temporal learning-based network for cardiovascular disease diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(8): 4661-4671.

[10] Xiao Q, Li Y, Lee K, et al. Integrating learnable expert knowledge into deep learning-based multi-label ECG classification[J]. Measurement Science and Technology, 2025, 36(5): 056113.

[11] Yao Q H, Wang R X, Fan X M, et al. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network[J]. Information Fusion, 2020, 53: 174-182.

[12] Jyotishi D, Dandapat S. An attention based hierarchical LSTM model for detection of myocardial infarction[C]//2020 IEEE 17th India Council International Conference (INDICON). New Delhi, India: IEEE, 2020: 1-5.