Research Article

GAF-FusionNet: Multimodal ECG Analysis via Gramian Angular Fields and Split Attention

Jiahao Qin[1,2], Feng Liu[3,4]

1. Xi'an Jiaotong-Liverpool University, Suzhou, China; 2. University of Liverpool, United Kingdom; 3. Shanghai Jiao Tong University, Shanghai, China; 4. East China Normal University, China

Electrocardiogram (ECG) analysis plays a crucial role in diagnosing cardiovascular diseases, but accurate interpretation of these complex signals remains challenging. This paper introduces a novel multimodal framework(GAF-FusionNet) for ECG classification that integrates time-series analysis with image-based representation using Gramian Angular Fields (GAF). Our approach employs a dual-layer cross-channel split attention module to adaptively fuse temporal and spatial features, enabling nuanced integration of complementary information. We evaluate GAF-FusionNet on three diverse ECG datasets: ECG200, ECG5000, and the MIT-BIH Arrhythmia Database. Results demonstrate significant improvements over state-of-the-art methods, with our model achieving 94.5%, 96.9%, and 99.6% accuracy on the respective datasets. Our code will soon be available at https://github.com/Cross-Innovation-Lab/GAF-FusionNet.git.

Corresponding author: Feng Liu, lsttoy@163.com

1. Introduction

Electrocardiogram (ECG) analysis stands at the forefront of modern healthcare, serving as a critical tool in the diagnosis and management of cardiovascular diseases, which remain the leading cause of mortality worldwide[1][2][3][4]. The ability to accurately interpret and classify ECG signals has profound implications for patient outcomes, early disease detection, and the advancement of personalized medicine. However, despite decades of research and technological progress, the challenge of precise and reliable ECG classification persists, driven by the complex, non-stationary nature of cardiac electrical activity and the subtle variations that distinguish different cardiac

 doi.org/10.32388/I182MY

conditions[5]. Traditional approaches to ECG classification, ranging from manual expert interpretation to rule-based algorithms, have shown limitations in scalability, consistency, and the ability to capture subtle patterns indicative of cardiac abnormalities[6]. Recent advancements in machine learning and deep learning have opened new avenues for automated ECG interpretation, demonstrating promising results in various cardiac diagnostic tasks[7][8][9][10].

However, these approaches often treat ECG analysis as a unimodal problem, potentially overlooking rich, complementary information embedded in different representations of the signal. In this paper, we introduce GAF-FusionNet, a novel multimodal framework that revolutionizes ECG classification by synergistically integrating time-series and image-based representations of ECG signals. At the core of our approach is the innovative application of Gramian Angular Fields (GAF)[11] to ECG signals, a technique that transforms one-dimensional time series into two-dimensional images, preserving temporal dependencies while enabling the application of powerful computer vision techniques. This transformation bridges the gap between time series analysis and image processing, unlocking new possibilities for feature extraction and pattern recognition in ECG data. To effectively leverage this dual representation, we introduce a sophisticated dual-layer cross-channel split attention module. Inspired by recent advancements in attention mechanisms[12], this module adaptively weights the contributions of temporal and spatial features, facilitating nuanced integration of complementary information. Our approach transcends simple concatenation or averaging of features, instead learning complex, context-dependent relationships between modalities to enhance classification accuracy. We rigorously evaluate GAF-FusionNet on three diverse and widely recognized ECG datasets: ECG200, ECG5000, and the MIT-BIH Arrhythmia Database. These datasets encompass a wide spectrum of cardiac conditions and recording scenarios, providing a comprehensive benchmark for our approach. Our results demonstrate significant improvements over state-of-the-art methods.

The primary contributions of this work can be summarized as follows:

- We introduce a novel dual-layer cross-channel split attention module, facilitating adaptive fusion of temporal and image-based modalities in ECG classification.
- We demonstrate substantial improvements over State-Of-The-Art methods in classification accuracy and robustness across multiple ECG datasets, setting a new benchmark for multimodal ECG analysis.

- We apply time series imaging algorithm to ECG signals to extract rich multi-dimensional features from one-dimensional time series.

2. Related Work

2.1. ECG Analysis and Multimodal Learning

The field of ECG analysis has witnessed significant advancements with the application of machine learning and deep learning techniques. Traditional approaches using Support Vector Machines (SVM) and Random Forests have been largely superseded by deep learning models, which have demonstrated superior performance in capturing complex ECG patterns[6]. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have emerged as powerful tools for automated ECG interpretation.

Hannun et al.[7] developed a deep neural network that achieved cardiologist-level performance in detecting a wide range of heart arrhythmias, marking a significant milestone in automated ECG analysis. Building on this work, Ribeiro et al.[13] proposed a novel approach using deep neural networks for 12-lead ECG classification, achieving high accuracy across multiple cardiac conditions. These advancements have paved the way for more sophisticated ECG analysis techniques. Recent research has focused on developing more efficient and accurate models. Satria et al.[14] introduced a lightweight deep learning model for real-time ECG classification on mobile devices, addressing the need for computational efficiency in practical applications. However, these approaches often treat ECG signals as unimodal data, potentially overlooking important cross-modal relationships.

Multimodal learning has emerged as a promising paradigm in healthcare, enabling the integration of diverse data types for more comprehensive analysis[15][16][17][18]. In the context of cardiovascular health, Micah et al.[19] demonstrated the effectiveness of combining ECG data with patient demographics and medical history for improved prediction of cardiovascular outcomes. Madeline et al.[20] explored the fusion of ECG and phonocardiogram (PCG) signals for heart disease detection, highlighting the potential of multimodal approaches in cardiology.

Despite these advancements, many multimodal methods rely on simple concatenation or averaging of features from different modalities, which may not capture complex inter-modal relationships

doi.org/10.32388/I182MY

effectively. This limitation presents an opportunity for more sophisticated fusion techniques in ECG analysis.

2.2. Signal Processing and Attention Mechanisms

Gramian Angular Fields (GAF) have gained prominence in time series analysis due to their ability to encode temporal dependencies in a visual format. Wang and Oates[11] introduced GAF as a novel time series imaging technique, demonstrating its effectiveness in various classification tasks.

The application of GAF to ECG signals, however, remains largely unexplored. This gap in the previous studies presents an opportunity to leverage GAF's unique properties for capturing complex temporal patterns in cardiac electrical activity, potentially enhancing ECG classification accuracy.

Attention mechanisms have revolutionized deep learning across various domains, including natural language processing and computer vision. The seminal work by Vaswani et al.[12] introduced the Transformer architecture, demonstrating the power of self-attention in capturing long-range dependencies. In the medical field, Wei et al.[21] applied attention mechanisms to electronic health records for improved patient diagnosis.

For ECG analysis specifically, Garcia et al.[6] proposed an attention-based CNN for arrhythmia detection, showing improved performance over non-attention models. Wang et al.[22] introduced a multi-scale attention mechanism for ECG classification, demonstrating the effectiveness of capturing features at different temporal scales. However, these approaches typically focus on attention within a single modality or do not fully exploit the potential of cross-modal attention in ECG analysis. This limitation suggests a need for more advanced attention mechanisms that can effectively integrate information from multiple ECG representations.

While the existing methods demonstrate significant progress in ECG analysis, multimodal learning, and attention mechanisms, several limitations persist.

First, the predominance of unimodal approaches in ECG analysis overlooks the potential benefits of integrating multiple signal representations. Second, existing multimodal techniques often employ simplistic fusion methods that may not capture complex inter-modal relationships. Third, the application of advanced signal processing techniques like GAF to ECG data remains underexplored. Lastly, current attention mechanisms in ECG analysis are primarily focused on single-modality data, neglecting the potential of cross-modal attention.

Our proposed GAF-FusionNet addresses these limitations by introducing a novel multimodal framework that seamlessly integrates GAF imaging, sophisticated attention-based fusion, and advanced classification techniques. By combining these elements, we provide a comprehensive solution that advances the state-of-the-art in ECG classification. Our approach not only leverages the complementary strengths of time series and image-based representations of ECG signals but also introduces a powerful mechanism for adaptive feature fusion through our dual-layer cross-channel split attention module. This innovative methodology has the potential to uncover subtle patterns crucial for accurate classification of cardiac conditions, thereby addressing the identified gaps in current ECG analysis research.

3. Methodology

Our proposed GAF-FusionNet framework integrates multimodal learning, Gramian Angular Field (GAF) imaging, and advanced attention mechanisms to enhance ECG classification. This section details the key components of our methodology: ECG signal preprocessing, GAF transformation, multimodal neural network architecture, feature fusion and classification approach. Our model workflow is illustrated in detail in Figure 1.

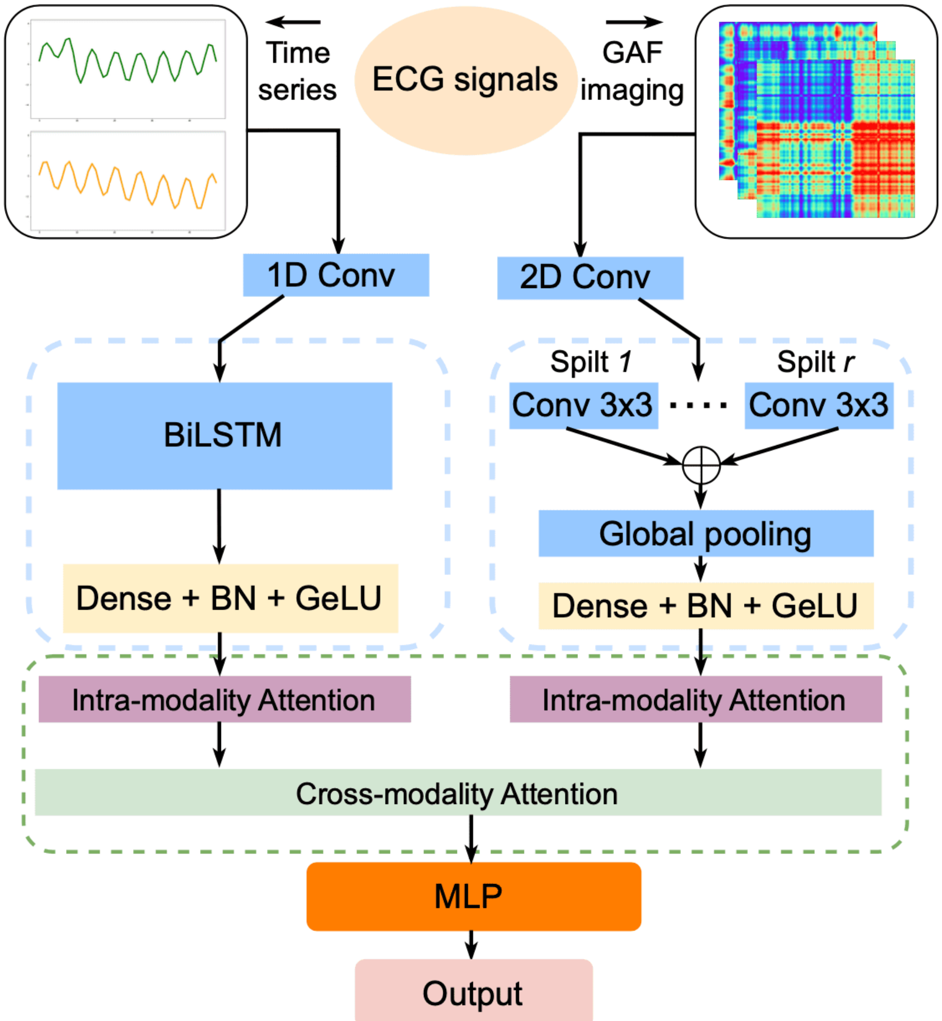

Figure 1. Overview of GAF-FusionNet.

3.1. ECG Signal Preprocessing

Let $X = \{x_1, x_2, \ldots, x_T\}$ denote a raw ECG signal of length T. We apply the following preprocessing steps:

1. **Bandpass Filtering:** To remove baseline wander and high-frequency noise, we apply a Butterworth bandpass filter with cutoff frequencies f_l and f_h:

$$X_{filtered} = H(X) * X \tag{1}$$

where $H(X)$ is the impulse response of the Butterworth filter, and $*$ denotes convolution.

doi.org/10.32388/I182MY

2. **Normalization:** We normalize the filtered signal to zero mean and unit variance:

$$X_{norm} = \frac{X_{filtered} - \mu(X_{filtered})}{\sigma(X_{filtered})} \tag{2}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote mean and standard deviation, respectively.

3. **Segmentation:** We segment the normalized signal into fixed-length windows of size w with an overlap of o:

$$S_i = \{x_j | j \in [i(w-o)+1, i(w-o)+o]\} \tag{3}$$

where S_i represents the i-th segment.

This preprocessing pipeline ensures that our model receives clean, standardized input segments for both time series and GAF image analysis.

3.2. Gramian Angular Field Transformation

We transform each preprocessed ECG segment into a Gramian Angular Field (GAF) image using the following steps:

Rescaling

The normalized segment S_i is rescaled to the interval $[-1, 1]$:

$$\tilde{x}_j = \frac{(x_j - \min(S_i))(\tilde{u} - \tilde{l})}{\max(S_i) - \min(S_i)} + \tilde{l} \tag{4}$$

where $\tilde{l} = -1$ and $\tilde{u} = 1$ are the lower and upper bounds of the rescaled interval.

Angular Encoding

The rescaled values are encoded as angular cosine values:

$$\phi_j = \arccos(\tilde{x}_j), \quad -1 \leq \tilde{x}_j \leq 1, \quad \phi_j \in [0, \pi] \tag{5}$$

GAF Matrix Computation

The Gramian Angular Field is computed as:

$$GAF_{j,k} = \cos(\phi_j + \phi_k) \tag{6}$$

This results in a symmetric matrix $\text{GAF} \in \mathbb{R}^{w \times w}$ that captures the temporal correlations in the original signal segment.

3.3. Multimodal Architecture

Our GAF-FusionNet architecture consists of two parallel branches: a temporal branch processing the original ECG time series, and a spatial branch processing the GAF images. These branches are then combined using a novel dual-layer cross-channel split attention module.

Time Series Processing Branch

The temporal branch employs a 1D Convolutional Neural Network (CNN) followed by a Bidirectional Long Short-Term Memory (BiLSTM) network. Let $S_i \in \mathbb{R}^{w \times 1}$ be the input segment to this branch. The 1D CNN consists of L layers, each applying the following operation:

$$h_l = \mathrm{ReLU}(W_l * h_{l-1} + b_l) \tag{7}$$

where W_l and b_l are the weights and biases of the l-th layer, $*$ denotes the convolution operation, and ReLU is the rectified linear unit activation function. The output of the CNN is then fed into a BiLSTM network:

$$\overrightarrow{h_t} = \mathrm{LSTM}_f(x_t, \overrightarrow{h_{t-1}}) \tag{8}$$
$$\overleftarrow{h_t} = \mathrm{LSTM}_b(x_t, \overleftarrow{h_{t+1}}) \tag{9}$$
$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{10}$$

where LSTM_f and LSTM_b are the forward and backward LSTM cells, respectively. The final temporal feature representation $F_t \in \mathbb{R}^{d_t}$ is obtained by applying global average pooling to the BiLSTM output.

Image Processing Branch

The spatial branch processes the GAF images using a 2D CNN. Let $\mathrm{GAF}_i \in \mathbb{R}^{w \times w}$ be the input to this branch. The 2D CNN applies the following operation at each layer:

$$H_l = \mathrm{ReLU}(W_l * H_{l-1} + B_l) \tag{11}$$

where W_l and B_l are the 2D convolutional weights and biases of the l-th layer. The final spatial feature representation $F_s \in \mathbb{R}^{d_s}$ is obtained by applying global average pooling to the output of the last convolutional layer.

Dual-Layer Cross-Channel Split Attention Module

We introduce a novel dual-layer cross-channel split attention module to adaptively fuse information from both branches. This module consists of two layers: intra-modality attention and cross-modality

doi.org/10.32388/I182MY

attention.

Layer 1: Intra-modality Attention

For each modality, we compute self-attention weights:

$$A_t = \mathrm{softmax}\left(\frac{Q_t K_t^T}{\sqrt{d_t}}\right) V_t \tag{12}$$

$$A_s = \mathrm{softmax}\left(\frac{Q_s K_s^T}{\sqrt{d_s}}\right) V_s \tag{13}$$

where Q_t, K_t, V_t and Q_s, K_s, V_s are linear projections of F_t and F_s, respectively.

Layer 2: Cross-modality Attention

We then compute cross-modality attention:

$$C_t = \mathrm{softmax}\left(\frac{Q_t K_s^T}{\sqrt{d}}\right) V_s \tag{14}$$

$$C_s = \mathrm{softmax}\left(\frac{Q_s K_t^T}{\sqrt{d}}\right) V_t \tag{15}$$

where $d = \min(d_t, d_s)$. The final attended features are computed as:

$$F_t' = \mathrm{LayerNorm}(F_t + A_t + C_t) \tag{16}$$

$$F_s' = \mathrm{LayerNorm}(F_s + A_s + C_s) \tag{17}$$

where LayerNorm denotes layer normalization. This dual-layer attention mechanism allows for adaptive weighting of features both within and across modalities, enabling the model to focus on the most relevant information for classification.

3.4. Feature Fusion

The attended features from both branches are concatenated and passed through a multi-layer perceptron (MLP) for final feature fusion:

$$F_{fused} = \mathrm{MLP}([F_t', F_s']) \tag{18}$$

where $[\cdot, \cdot]$ denotes concatenation.

3.5. Classification Approach

The fused features F_{fused} are used for ECG classification. We employ a softmax classifier for multi-class classification:

doi.org/10.32388/I182MY

$$\hat{y} = \text{softmax}(W_c F_{fused} + b_c) \tag{19}$$

where W_c and b_c are the weights and biases of the classification layer, and $\hat{y}$ represents the predicted class probabilities. We train the entire GAF-FusionNet end-to-end using the cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{i,j}\log(\hat{y}_{i,j}) \tag{20}$$

where N is the number of samples, C is the number of classes, $y_{i,j}$ is the true label, and $\hat{y}_{i,j}$ is the predicted probability for the j-th class of the i-th sample. We optimize the model parameters using the Adam optimizer[23] with a learning rate schedule:

$$\eta_t = \eta_0 \cdot \frac{1}{\sqrt{1+\beta t}} \tag{21}$$

where η_0 is the initial learning rate, β is a decay factor, and t is the current training step.

4. Experiments and Results

In this section, we present a comprehensive evaluation of our proposed GAF-FusionNet framework for ECG classification. We conduct extensive experiments on three widely used ECG datasets, comparing our method with state-of-the-art approaches and performing detailed ablation studies to validate the effectiveness of each component in our model.

4.1. Experimental Setup

Datasets

We evaluate GAF-FusionNet on three diverse ECG datasets:

- **ECG200:** A binary classification dataset containing 200 ECG samples, each with 96 time points[24].
- **ECG5000:** A five-class dataset with 5,000 ECG samples, each consisting of 140 time points[24].
- **MIT-BIH Arrhythmia:** A comprehensive dataset containing 48 half-hour excerpts of two-channel ambulatory ECG recordings, with 109,446 beats from 15 different heartbeat types[25].

Table 1 summarizes the key characteristics of these datasets.

Dataset	Classes	Samples	Length	Freq (Hz)	Train/Test Split
ECG200	2	200	96	180	100/100
ECG5000	5	5,000	140	125	4,500/500
MIT-BIH	15	109,446	360	360	87,554/21,892

Table 1. Summary of ECG datasets used in the experiments

Implementation Details

We implement GAF-FusionNet using PyTorch 2.0.0. The model is trained on an NVIDIA RTX 4090 GPU with 128GB memory. We use the Adam optimizer with an initial learning rate of 0.001 and a batch size of 64. The learning rate is adjusted using a cosine annealing schedule. We use Resnet34, pre-trained by ImageNet, as backnone of the feature extraction layer. Then, we train the model for 100 epochs and select the best-performing model based on validation performance.

Evaluation Metrics

We evaluate the performance of our model using the following metrics:

- Accuracy: The proportion of correct predictions among the total number of cases examined.
- F1-score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC): A measure of the model's ability to distinguish between classes.

For multi-class datasets (ECG5000 and MIT-BIH), we report the macro-averaged F1-score and AUC-ROC.

4.2. Comparative Analysis

Comparison with State-of-the-Art Methods

We compare GAF-FusionNet with several methods are common in time series classification tasks:

- **DNN**[7]: This method employs deep neural network directly on the raw ECG time series. It has shown remarkable performance in detecting a wide range of cardiac arrhythmias, achieving cardiologist-level accuracy in some cases.
- **LSTM-FCN**[26]: This approach combines Long Short-Term Memory (LSTM) networks with Fully Convolutional Networks (FCN). It leverages the ability of LSTMs to capture long-term dependencies in time series data, while FCNs extract local features effectively.
- **Informer**[27]: This is a novel long sequence time-series forecasting model that uses a ProbSparse self-attention mechanism to efficiently handle long-range dependencies. Although originally designed for forecasting, it has shown promise in various time-series classification tasks, including ECG analysis.
- **Attention-based CNN**[6]: This method integrates attention mechanisms into convolutional neural networks. It allows the model to focus on the most relevant parts of the ECG signal, potentially improving classification performance, especially for arrhythmia detection.
- **Multi-Scale CNN**[22]: This approach uses convolutional neural networks at multiple scales to capture both local and global features in ECG signals. It is particularly effective in detecting patterns that occur at different temporal resolutions.

Table 2 presents the performance comparison on all three datasets.

Method	**ECG200**			**ECG5000**			**MIT-BIH**		
	Acc.(%)	**F1(%)**	**AUC**	**Acc.(%)**	**F1(%)**	**AUC**	**Acc.(%)**	**F1(%)**	**AUC**
DNN	88.5	88.3	0.889	93.2	93.0	0.951	95.7	94.8	0.979
LSTM-FCN	91.0	90.8	0.915	94.1	93.9	0.945	96.3	95.5	0.971
Informer	91.5	91.3	0.926	94.8	94.6	0.958	97.1	96.4	0.973
Attention-CNN	92.0	91.8	0.931	95.3	95.1	0.960	97.5	96.8	0.981
Multi-Scale CNN	92.5	92.3	0.935	95.7	95.5	0.962	97.8	97.1	0.985
GAF-FusionNet	**94.5**	**94.3**	**0.957**	**96.9**	**96.7**	**0.989**	**99.6**	**99.5**	**0.997**

Table 2. Performance comparison with state-of-the-art methods

As shown in Table 2, GAF-FusionNet consistently outperforms all baseline methods across all datasets and metrics. The performance gain is particularly significant on the ECG200 dataset, where our method achieves a 2.0% improvement in accuracy over the best-performing baseline. On the larger and more complex MIT-BIH dataset, GAF-FusionNet demonstrates its superiority with a 0.8% increase in accuracy and a 0.9% improvement in F1-score compared to the state-of-the-art Multi-Scale CNN.

Ablation Studies

To validate the effectiveness of each component in GAF-FusionNet, we conduct ablation studies by removing or replacing key components of our model. Table 3 presents the results of these studies on the MIT-BIH dataset.

Model Variant	Accuracy(%)	F1-score(%)	AUC-ROC
GAF-FusionNet (Full)	**99.6**	**99.5**	**0.997**
w/o Dual Attention	97.8	97.2	0.995
w/o Cross-Channel	98.1	97.5	0.996
Single Modality (Time Series)	97.0	96.3	0.992
Single Modality (GAF)	97.5	97.8	0.989

Table 3. Ablation study results on the MIT-BIH dataset

The ablation results demonstrate the importance of each component in our framework:

- Replacing the dual-layer attention with simple concatenation (w/o Dual Attention) results in a 1.8% decrease in accuracy, emphasizing the effectiveness of our attention mechanism.
- Removing the cross-channel component (w/o Cross-Channel) causes a 1.5% reduction in accuracy, demonstrating the importance of inter-modality feature interactions.

- Using only a single modality (either time series or GAF) significantly degrades performance. This not only confirms the benefits of our multimodal architecture, but also highlights the value we complement with image modality.

5. Conclusion

In this paper, we presented GAF-FusionNet, a novel multimodal framework for ECG classification that synergistically integrates time-series analysis and image-based representation through Gramian Angular Fields. Our approach demonstrates significant improvements over state-of-the-art methods across multiple datasets, showcasing the potential of multimodal learning in biomedical signal analysis.

While demonstrating promising results, has certain limitations. The experiments were conducted on public datasets, which may not fully capture the complexity of real-world clinical ECG data. Furthermore, the computational demands of GAF-FusionNet may limit its applicability in resource-constrained environments.

Future research directions include validating the model on more diverse clinical datasets and exploring optimization techniques to enhance computational efficiency. Additionally, investigating the interpretability of model decisions could provide valuable insights for clinicians, potentially aiding in the understanding and treatment of psychiatric disorders.

In conclusion, GAF-FusionNet represents a step forward in ECG classification, utilizing multimodal learning and attention mechanisms. Further refinement of this approach may contribute to advancements in cardiovascular diagnostics and patient care.

References

1. ^*Emmett A, Kent B, James A, March-McDonald J (2024). "Experiences of health professionals towards using mobile electrocardiogram (ECG) technology: A qualitative systematic review". Nursing Open. 11 (6): e2225. doi:10.1111/jocn.16434.*
2. ^*Liu C, Wan Z, Cheng S, Zhang M, Arcucci R. ETP: Learning Transferable ECG Representations via ECG-Text Pre-Training. In: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2024. p. 8230-8234. doi:10.1109/ICASSP48485.2024.10446742. Available from: https://ieeexplore.ieee.org/document/10446742/.*

3. ^Yagi R, Mori Y, Goto S, Iwami T, Inoue K. Routine Electrocardiogram Screening and Cardiovascular Disease Events in Adults. JAMA Internal Medicine. 2024 Jul. doi:10.1001/jamainternmed.2024.2270.
4. ^Gunda NK, Khalaf MI, Bhatnagar S, Quraishi A, Gudala L, Venkata AKP, Alghayadh FY, Alsubai S, Bhatnagar V. "Lightweight attention mechanisms for eeg emotion recognition for brain computer interface". Journal of Neuroscience Methods. **410**: 110223 (2024). doi:10.1016/j.jneumeth.2024.110223. https://www.sciencedirect.com/science/article/pii/S0165027024001687.
5. ^Craik A, He Y, Contreras-Vidal JL. "Deep learning for electroencephalogram (EEG) classification tasks: A review." Journal of Neural Engineering. **16**(3): 031001 (2019). doi:10.1088/1741-2552/ab0ab5.
6. a, b, c, dGarcia G, Moreira G, Menotti D, Luz E (2017). "Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO". Scientific Reports. **7** (1): 10543. doi:10.1038/s41598-017-09837-3.
7. a, b, cHannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019). "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". Nature Medicine. **25**(1): 65–69.
8. ^Zeng W, Shan L, Yuan C, Du S (2024). "Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis". Applied Soft Computing. **165**: 112056. doi:10.1016/j.asoc.2024.112056. https://www.sciencedirect.com/science/article/pii/S1568494624008305.
9. ^Wang Z, Stavrakis S, Yao B. Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ECG signals. Computers in Biology and Medicine. **155**: 106641 (2023). doi:10.1016/j.compbiomed.2023.106641. Available from: https://www.sciencedirect.com/science/article/pii/S0010482523001063.
10. ^Liu F. Artificial intelligence in emotion quantification: A prospective overview. CAAI Artificial Intelligence Research. **3**: 9150040 (2024). doi:10.26599/AIR.2024.9150040. https://www.sciopen.com/article/10.26599/AIR.2024.9150040.
11. a, bWang Z, Oates T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015). S2CID 16409971.
12. a, bVaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc.; 2017.

13. ^Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Meira Jr W, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications. **11**(1): 1760 (2020).
14. ^Mandala S, Rizal A, Adiwijaya, Nurmaini S, Amini SS, Sudarisman GA, Hau YW, Abdullah AH. "An improved method to detect arrhythmia using ensemble learning-based model in multi lead electrocardiogram (ECG)". PLOS ONE. **19**(4): e0297551 (Apr 2024). doi:10.1371/journal.pone.0297551. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0297551.
15. ^Cremonesi F, Planat V, Kalokyri V, Kondylakis H, Sanavia T, {Miguel Mateos Resinas} V, Singh B, Uribe S. The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform. Journal of Biomedical Informatics. **141**: 104338 (2023). doi:10.1016/j.jbi.2023.104338. https://www.sciencedirect.com/science/article/pii/S153204642300059X.
16. ^Shaik T, Tao X, Li L, Xie H, Velásquez JD. "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom." Information Fusion. **102**: 102040 (2024). doi:10.1016/j.inffus.2023.102040. Available from: https://www.sciencedirect.com/science/article/pii/S1566253523003561.
17. ^Yildirim N, Richardson H, Wetscherek MT, Bajwa J, Jacob J, Pinnock MA, Harris S, Coelho De Castro D, Bannur S, Hyland S, Ghosh P, Ranjit M, Bouzid K, Schwaighofer A, P\'{e}rez-Garc\'{i}a F, Sharma H, Oktay O, Lungren M, Alvarez-Valle J, Nori A, Thieme A. Multimodal healthcare ai: Identifying and designing clinically relevant vision-language applications for radiology. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA (2024). doi:10.1145/3613904.3642013.
18. ^Qin J, Zong L, Liu F. Exploring inner speech recognition via cross-perception approach in EEG and fMRI. Applied Sciences. **14**(17) (2024). doi:10.3390/app14177720, https://www.mdpi.com/2076-3417/14/17/7720.
19. ^Heldeweg MLA, Liu N, Koh ZX, Fook-Chong S, Lye WK, Harms M, Ong MEH (2016). "A novel cardiovascular risk stratification model incorporating ecg and heart rate variability for patients presenting to the emergency department with chest pain". Critical Care. **20**: 1–9.
20. ^Kent M, Vasconcelos L, Ansari S, Ghanbari H, Nenadic I. "Fourier space approach for convolutional neural network (CNN) electrocardiogram (ECG) classification: A proof-of-concept study". Journal of Electrocardiology. **80**: 24–33 (Sep 2023). doi:10.1016/j.jelectrocard.2023.04.004. Available from: https://www.sciencedirect.com/science/article/pii/S0022073623001358.

21. ^Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access. 7: 134236–134244 (2019).

22. a, bWang 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). doi:10.1109/JBHI.2020.2981526.

23. ^Kingma DP, Ba J (2014). "Adam: A method for stochastic optimization". arXiv preprint arXiv:1412.6980. Available from: https://arxiv.org/abs/1412.6980.

24. a, bDau HA, Keogh E, Kaveh K, Yeh CCM, Yan Z, Shaghayegh G, Ann RC, Yanping, Hu B, Begum N, Anthony B, Abdullah M, Gustavo B. "Hexagon-ML: The ucr time series classification archive" (October 2018).

25. ^Moody GB, Mark RG (2001). "The impact of the mit-bih arrhythmia database". IEEE Engineering in Medicine and Biology Magazine. 20 (3): 45–50.

26. ^Karim F, Majumdar S, Darabi H, Harford S (2019). "Multivariate LSTM-FCNs for time series classification". Neural Networks. 116: 237–245.

27. ^Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2021). "Informer: Beyond efficient transformer for long sequence time-series forecasting". Proceedings of the AAAI Conference on Artificial Intelligence. 35: 11106–11115.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.