OPEN ACCESS

International Journal of Innovations in Science & Technology

ECG Lead Selection for Disease Diagnostics Using CNN-Transformer

Muhammad Nabeel Raza[1], Sahar Waqar[2]
[1]School of Electrical Engineering and Computer Science (National University of Science and Technology).
[2]Department of Computer Engineering (University of Engineering and Technology).
*Correspondence: razanabeel512@gmail.com

Citation | Raza. M. N, Waqar. S, "ECG Lead Selection for Disease Diagnostics Using CNN-Transformer", IJIST, Vol. 07 Issue. 03 pp 1767-1778, August 2025

DOI | https://doi.org/10.33411/ijist/20257317671778

Received | June 20, 2025 **Revised |** Aug 01, 2025 **Accepted |** Aug 03, 2025 **Published |** Aug 04, 2025.

Electrocardiography (ECG) is crucial for diagnosing cardiovascular diseases (CVDs), which cause millions of deaths each year. This study addresses the challenge of CVD diagnosis in rural areas, where there is a shortage of skilled healthcare professionals and medical equipment. This study proposes a novel method to systematically compare different ECG leads using Deep Learning techniques, specifically a 1D CNN Transformer, to detect anomalies from minimal disturbances. The analysis was conducted using the PTB-XL dataset and further validated with Holter ECG-based records from the St. Petersburg INCART database. Minimal pre-processing was applied, limited to baseline wander removal, to maintain the intrinsic information of each lead. The results indicate that utilizing all leads significantly improves the F1 score, although lead II, V1, and V2 also provide comparable results in the INCART database. This study demonstrates that fewer leads can be effectively used to diagnose diseases, facilitating the creation of low-cost ECG machines suitable for deployment in rural areas. The code is publicly available at https://github.com/nabeelraza-7/ecg-lead-selection.

Keywords: Electrocardiography; Cardiovascular Diseases; Deep Learning; PTB-XL Dataset; ECG Lead Optimization

Introduction:

Electrocardiography (ECG):

Cardiovascular diseases (CVDs) can be diagnosed using various methods, including chest X-rays, echocardiograms, and computed tomography (CT) scans. However, electrocardiography (ECG) is widely used for diagnosing cardiovascular diseases (CVDs) because it provides valuable information about the heart's ability to circulate blood throughout the body. This is achieved by recording electrical activity from various anatomical positions using multiple leads. The heart's electrical impulses create potential differences, which are detected by sensitive electrodes placed on the body.

While ECG signals are unique to each individual, they contain recognizable patterns that medical professionals are trained to interpret. The absence of specific waves or the presence of minor artifacts in an ECG signal may indicate abnormal or compromised heart function. A standard ECG has 12 signals, and each signal is referred to as a lead. A sample of a 12-lead ECG is shown in Figure 1. Among the 12 ECG leads, leads I, II, III, aVF, aVL, and aVR are known as limb leads, while leads V1 to V6 are referred to as chest (precordial) leads. Each lead records the heart's electrical activity from a different angle, providing a comprehensive view of cardiac function. This allows professionals to understand which part of the heart is not functioning properly.

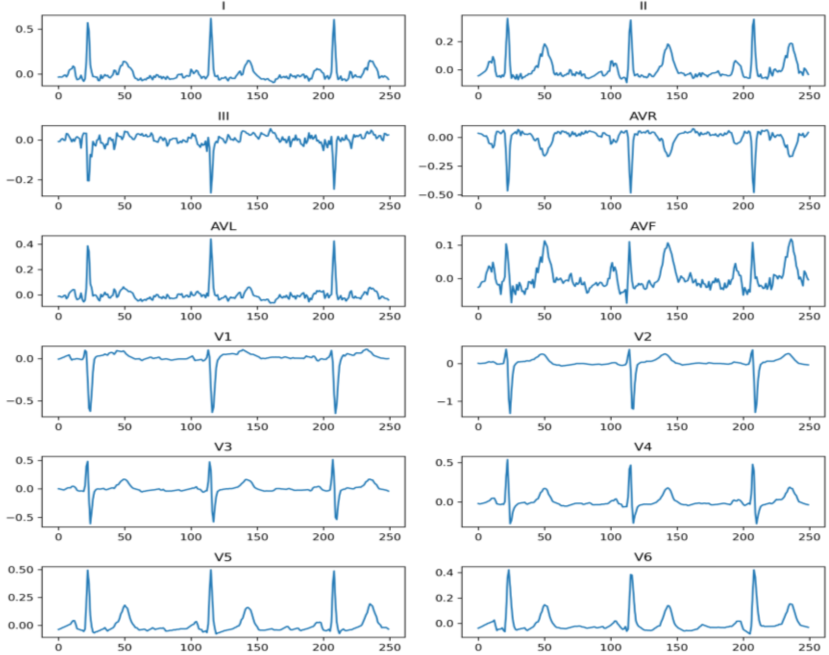


Figure 1. A standard 12-lead ECG sample from the PTB-XL dataset

Literature Review:

The existing studies use multiple leads for the classification and diagnosis of CVDs. However, usually one combination of any two leads, and 12 leads were found frequently for the said task. For better understanding, the literature has been categorized based on the number of ECG leads used: single-lead, two-lead, and 12-lead configurations. Additionally, with the broad spectrum of machine learning and deep learning algorithms used in the literature, the scope has been limited to the deep learning models due to their better results than the former.

To the best of our knowledge, existing literature does not adequately address intra-comparative analyses of different ECG leads in the context of deep learning-based models. This gives insights into the impact of different leads on diagnosing several types of heart diseases. The tabular comparison among the discussed studies has been illustrated in Tables 1, 2, and 3 for single, double, and 12 leads, respectively.

Single Lead:

In [1], Author performed a classification of Atrial Fibrillation (AF) using DenseNet CNN for a single lead. They achieved an F1 score of 0.82. Author, in their paper [2], created a realistic environment to create a robust model. They added random jitters so the model would be robust. Only lead II from the PTB Diagnostic ECG Database was utilized for the detection of Myocardial Infarction (MI). They trained on a single ECG lead with a CNN model. This model is trained on both noisy and noise-free datasets, achieving an accuracy of 93.53% with noise and 95.22% without noise.

Table 1. Research conducted using a single ECG Lead.

Ref.	ECG Lead	Model	Dataset	Disease	Results
[1]	I	DenseNet CNN	Physionet (2017)	AF	F1 = 0.82
[2]	II	CNN	PTB Diagnostic	MI	ACC-93.53%
[3]	III	Student-Teacher Model	Chapman ECG	Arrhythmia	ACC=97%
[4]	I	DNN, Xgboost	Physionet (2017)	NSR, AF, Alternative Rhythm, Noise	F1=0.84

In [3], Authors demonstrate that arrhythmia can be effectively detected using a single lead, rather than relying on all 12 leads. They proposed a student-teacher model with knowledge distillation. "The advanced teacher model is first developed using multi-lead ECG data, followed by the construction of the student model using a single lead from the 12-lead configuration. The experimental results demonstrated that the average accuracy of the student model with Lead-III was 97% in the classification of seven rhythms in the Chapman ECG.

In the paper [4], author used Deep neural networks as feature extractors and then used an ensemble of XGBoost models to achieve an F1 score of 0.84. They used a single-lead ECG from the Physionet dataset from 2017.

Double Lead:

In [5], a bidirectional LSTM is used for the diagnosis of arrhythmia. They used morphological features and RR interval. They trained both Bi-LSTM and Bi-LSTM with attention mechanisms, achieving 99% accuracy across all classes. The models are trained using lead II and V1 from the MIT-BIH Arrhythmia Dataset. Authors worked to improve the results from FusingTF and introduced a Low-Dimensional Denoising Embedded Transformer. It consisted of the transformer part as it is, but it introduced how low-dimensional denoising embeddings can improve performance. They used lead II and V1 from the MIT-BIH arrhythmia dataset and achieved 98.39% recall and 98.41% precision [6].

Authors used a multi-lead ECG to classify AF and Normal Sinus Rhythm (NSR). They used the spectrum and Temporal characteristics of the ECG signal to feed features to the classifier, which yielded an accuracy of about 98.3%. They used two ECG leads from a Holter ECG and tested them on the Physionet dataset [7].

Table 2: Research conducted using double ECG Leads.

Ref.	ECG Lead	Model	Dataset	Disease	Results
[5]	II and V1	Bi-LSTM	MIT-BIH	Arrhythmia	ACC = 99%
[6]	II and V1	Transformer	MIT-BIH	Arrhythmia	REC=98.39%
[7]	I and II	LSTM	Physionet + proprietary dataset	AF and NSR	ACC=98.3%

12 Lead:

Some researchers utilize all 12 leads of a standard ECG. In [1], author employ an ensemble of Transformer models to predict 27 different diseases using 12-lead ECG recordings. They also incorporate convolutional neural networks (CNNs) to generate positional encodings. Their approach achieved the highest validation score of 0.533 in the PhysioNet Challenge. In [8], author developed a cascaded convolutional neural network (CCNN) that extracts features using a 1D CNN and then cascades the input to feed into a 2D ResNet block. They used all 12 ECG leads and achieved an 86.5% F1 Score.

In the paper [9], authors used a DNN to perform automatic diagnosis. They reported an F1 score of above 0 in the diagnosis of all diseases. Their solution, which utilized all 12 ECG leads, outperformed a medical doctor in diagnostic accuracy. Authors used 12-lead ECG signals to train a bi-directional LSTM. They achieved an F1 score of 74.15% [10].

Authors of [11] also used PTB-XL for different kinds of models. They reported the best results on different disease detection tasks. They achieved an F1 score of 0.891 for 2 disease classification, 0.68 for 5 disease classification, and 0.332 for 20 disease classification. They used entropy-based features and a 1D CNN for this model. Similarly, authors of [12] modelled different 12-lead datasets, including the INCART database. They achieved 0.996 accuracy in all arrhythmia classes.

Table 3. Research conducted using all 12 ECG Leads.

Ref.	ECG Lead	Model	Dataset	Disease	Results
[1]	all	Ensemble Transformer	Physionet Challenge	27 diseases	0.53 Validation Score
[8]	all	Cascaded CNN	China ECG Intelligence	Arrhythmias	F1=86.5%
[9]	all	DNN	Telehealth Network of Minas Gerais	6 diseases	F1=80%
[10]	all	Bi-LSTM	China Physiological Signal Challenge	9 diseases	F1=74.15%
[11]	all	Entropy-based CNN	PTB-XL	2, 5, 20 diseases	F1=0.891, 0.68, 0.332
[12]	all	Cascaded classifier	QT, MIT-BIH supraventricular database, INCART database	different depending on the dataset	ACC=0.996

Research Gap:

From the above-mentioned literature and many more [13][14][15], it is evident that even though 12 ECG leads are useful, many problems can be solved by just using one or two of them. However, the results from these studies are not comparable as the underlying task varies. Therefore, this study provides the formal means for comparison between different ECG leads.

Moreover, most existing research focuses on single-label multiclass classification rather than multilabel binary classification, possibly due to data sparsity. Additionally, it is notable that transformer models are still not used proactively, even though they are replacing RNNs in almost every benchmark. Compared to RNNs, Transformers enable better parallelization and global feature learning. The lack of research may be partly since Transformer models require significantly more parameters compared to recurrent neural networks (RNNs).

Problem Statement:

ECG devices can be expensive, and hospitals in remote areas might not have access to them. These devices have been designed to be used by humans and require technical training, making them more difficult to use in remote areas. The contributions of this study are as follows:

A CNN-Transformer-based neural network architecture to aid in diagnostics.

A detailed comparison of different ECG leads and their contributions in disease diagnostics.

The motivation behind this is to reduce the number of ECG leads used in diagnosis, enabling the design of more cost-effective ECG machines. This raises the question "Why are 12 leads traditionally used?", and the answer is that perturbations are clearer when viewed from different angles. Since the goal is to minimize human involvement, it is expected that the models will improve at detecting and predicting these subtle variations.

Objectives:

The primary objective of this research is to investigate the feasibility of using a reduced number of ECG leads for accurate cardiovascular disease diagnosis, aiming to facilitate the development of low-cost ECG machines suitable for deployment in rural areas. To achieve this, the study focuses on the following specific aims:

Develop a robust CNN-Transformer model to aid in the diagnosis of cardiac diseases.

Systematically compare the diagnostic value of individual ECG leads.

Novelty Statement:

The novelty of this work is this formal intra-lead comparison, which has not been adequately addressed in prior deep learning research.

Validate the model's performance and findings on the PTB-XL and St. Petersburg INCART datasets to ensure the results are generalizable.

Material and Methods:

Methodology:

An open-access PTB-XL dataset was used as the primary dataset in this study. This dataset contains a large number of annotated 12-lead ECG recordings and clear benchmark algorithms as opposed to the existing ECG datasets [16]. It covers a broad range of age groups and about an equal contribution of both male and female populations, and is available for 500 Hz and 100 Hz. Additionally, the St. Petersburg INCART database was used to compare the model with Holter ECG recordings. It also contains almost equal contributions of male and female records. It contains 30-minute-long recordings at 257 Hz. The dataset was downsampled and segmented to align with PTB-XL [17]. The proposed experiment is illustrated in Figure 2. The raw signals contained breathing and movement-based artefacts. For this, signals were pre-processed, and baseline wander was removed. The pre-processed signal was then passed to the residual neural network (ResNet), where training was performed. Finally, the trained model was tested using testing and validation datasets.

Datasets:

Physikalisch-Technische Bundesanstalt (PTB) PTB-XL dataset [16] and St. Petersburg INCART database of Holter ECG records [17] were used in this study. PTB-XL is a large dataset with a total of 21799 10-second-long 12-lead ECG recordings. The age group covered a broad range of 0−95 years with a median of 62. Each record was annotated and validated by cardiologists. This dataset was recently released in 2019 with public access.

The ECG signals are available at sampling frequencies of 500 Hz and 100 Hz. For this study, the 100 Hz signals are selected, as they contain five times fewer data points, reducing computational complexity. A higher frequency can confuse the model with unnecessary complexities. Moreover, the overall morphology of the signal remains intact with a lower frequency, and makes the training faster. The distribution of classes in the dataset is represented in Table 4.

PTB-XL provides train, validation, and test splits. For fair evaluation, the same splits were selected as recommended by the authors [16]. This dataset is annotated with multiple labels, indicating that a single patient can be diagnosed with multiple diseases.

As for the St. Petersburg INCART database, the 7 classes were treated as a multiclass single-label problem. The dataset comprises 75 records, each containing 30-minute-long Holter ECG recordings. Due to the use of Holter monitoring, the recordings exhibit relatively higher noise levels. The dataset includes recordings from 17 men and 15 women, with ages ranging from 18 to 80 years and a mean age of 58. It was downsampled from 257 Hz to 100 Hz and segmented into 10-second-long records to make it consistent. The distribution of classes is given in Table 5. A stratified random split of 60% training data, 20% validation data, and 20% testing data was selected for model training and evaluation.

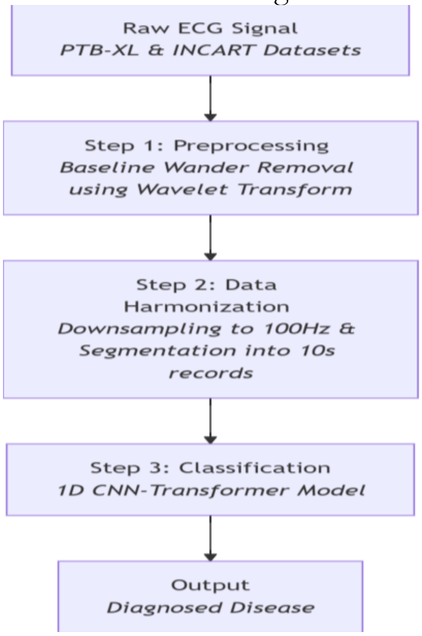


Figure 2. Methodology.

Table 4. Distribution of the diseases in the PTB-XL dataset.

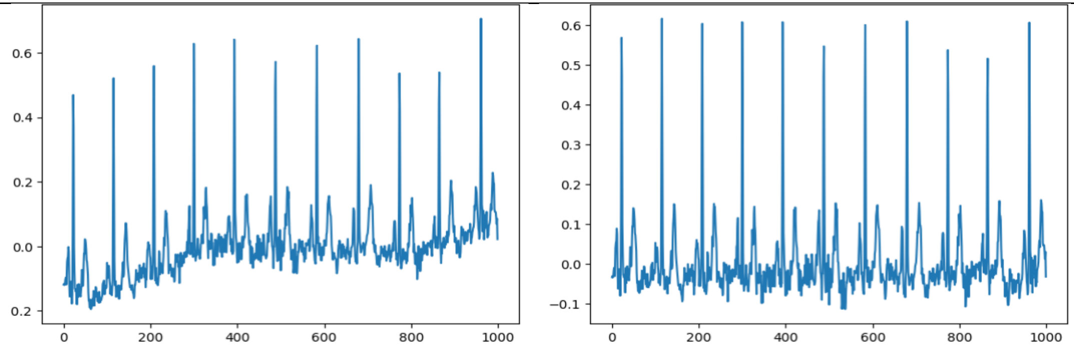
Records	Class
9514	Normal ECG(NORM)
5469	Myocardial Infarction (MI)
5235	ST/T Change (STTC)
4898	Conduction Disturbance (CD)
2649	Hypertrophy (HYP)

Table 5. Distribution of the diseases in the St. INCART database.

Records	Class
6120	Normal ECG (NORM)
1980	Coronary artery disease, arterial hypertension, left ventricular hypertrophy (CAD, HTN, LVH)
1620	Earlier MI
1260	Transient ischemic attack (TIA)
1080	Coronary artery disease, arterial hypertension (CAD, HTN)
1080	Acute MI
360	Sinus node dysfunction

Preprocessing:

As previously discussed, dataset visualization revealed that the raw signals contained artifacts that caused deviations from the baseline. These artifacts are typically caused by respiration, improper electrode placement, and other patient body movements. Such types of artefacts are called Baseline Wander. The presence of these artefacts may cause difficulty in the detection of ST-segment changes. Additionally, this noise causes the model to learn superfluous information, causing hurdles in the accurate diagnosis of the desired deformities in the ECG signal. [18].



Before baseline wander removal (a)

After baseline wander removal (b)

Figure 3. Baseline Wander Removal.

Different types of features can be extracted from ECG leads, but to maintain a fair comparison between different ECG leads, it is important to keep them in their base form so that only the intrinsic information from each ECG lead is used for training. Therefore, only baseline wander was used as the preprocessing technique.

To eliminate baseline wander from the signal, wavelet transformation (WT) was employed. Unlike Fourier Transforms (FT), WT represented the signal in both time and frequency domains using wavelets, basis functions with finite support, and was shown to be highly effective in removing low-frequency artifacts [19]. Among different families of wavelets, Daubechies or dbN, i.e., N levels, has been proven to be most suitable for the ECG signals. N represents the number of vanishing moments. A higher level indicates the low-frequency band, i.e., the low-frequency component. If the coefficients of WT are zeroed above a certain level N, then the low-frequency component of the time domain signal can be removed, thus reducing the noise. The time domain signal can be attained using the inverse of wavelet traversal (IWT), afterwards using the equation below, where y[t] represents the original signal, y'[t] the reversed signal, IWT(.) the inverse wavelet transform, and WT(.) the wavelet transform. The de-trended signal y'[t] is reconstructed after making the wavelet coefficients above N equal to zero.

$$y'[t] = IWT(WT(y[t]), N = 6)$$

The Daubechies (db6) wavelet was used with N = 6 decomposition levels, as it is morphologically similar to the ECG signal. Some studies have also used db7 and db8. The selection of N = 6 levels was made experimentally by testing different levels and observing the baseline removal visually. This is a commonly used technique for removing baseline wander from the ECG signals [18]. Figure 3 (a) shows the normal signal, and Figure 3 (b) shows the same signal with its baseline removed.

Data imbalance is one of the major problems in medical datasets. Typically, negative cases vastly outnumber positive cases, and ECG-based datasets are no exception. In this regard, a lot of research has been done to tackle this problem, and a number of techniques have been introduced, like under-sampling, Synthetic Minority Oversampling Technique (SMOTE, etc. These techniques have been proven to help models generalize [20][21][22]. This study opts for down-sampling of the negative class for both datasets.

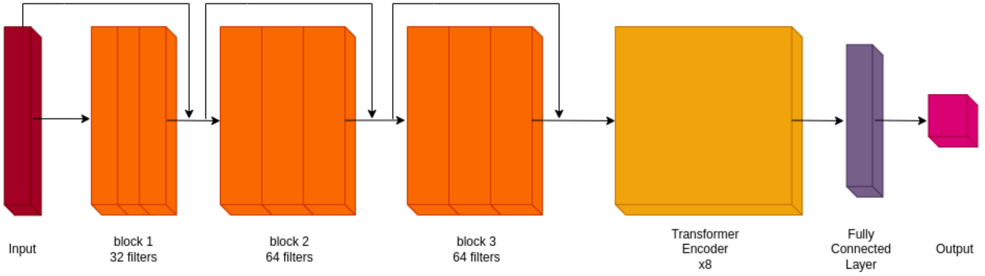


Figure 4. CNN-Transformer architecture

Model Architecture:

The model used for disease classification uses the residual neural network mentioned in [16] as a positional encoder for a transformer encoder head, followed by a fully connected layer for classification. The number of filters was reduced by half. It was shown in [23] that the residual neural network was the best-performing neural network when it comes to time series classification tasks [24]. It is tested on 44 benchmark datasets for time series classification tasks.

The architecture of the model consists of three residual blocks with {32,64,64} filters. The global averaging layer is replaced by an 8-layer transformer encoder with 5 attention heads. Compared to the residual neural network alone, faster convergence was observed. Transformers have been introduced as a replacement for recurrent neural networks, which makes them appropriate for the use case [25]. The architecture is illustrated in Figure 4.

Model Training:

For training, a random chunk of [N, 250] values is taken from [N, 1000] valued ECG signals, where N is the number of ECG leads in consideration. This prevents the model from overfitting and helps the model to generalize better. Moreover, a smaller number of values are propagated through the network, reducing computational load and training complexity. AdamW optimizer was used for training as it is more robust and better than the Adam optimizer. The parameters selected for training are mentioned in Table 6.

Table 6. Training parameters.

Parameter	Value
Learning Rate	1e-3
Betas	0.9, 0.999
Weight decay	1e-2
Batch size	128
Epochs	50
Early stopping patience	10 epochs over validation loss

These parameters were selected based on experiments conducted that led to faster and better convergence. More details about the parameters, model training, evaluation results, and other metrics can be found in the GitHub repository: https://github.com/nabeelraza-7/ecg-lead-selection.

Result and Discussion:

Results:

After a thorough evaluation, the results of testing are presented in Table 7 for the PTB-XL and INCART database. The F1 Score was chosen as the evaluation metric due to its ability to account for the imbalanced data classes.

For the PTB-XL dataset, the analysis demonstrates a clear performance advantage when using all 12 leads, which achieved a weighted F1 score of 0.7610. This score significantly surpasses the results of any single lead. Among the individual leads, V6 and AVR showed the strongest performance with F1 scores of 0.6634 and 0.6608, respectively. In contrast, lead III was the weakest performer on this dataset, with an F1 score of 0.5592. Confusion matrices for the 12-lead configuration are given in Figure 5.

Interestingly, the findings on the St. Petersburg INCART database present a different picture. While the 12-lead configuration still yielded the highest F1 score at 0.9967, several individual leads produced highly comparable results. Specifically, lead II, V2, and V1 achieved impressive F1 scores of 0.9818, 0.9863, and 0.9803, respectively. This suggests that for the types of anomalies present in the INCART database, a single lead can be nearly as effective as a full 12-lead ECG.

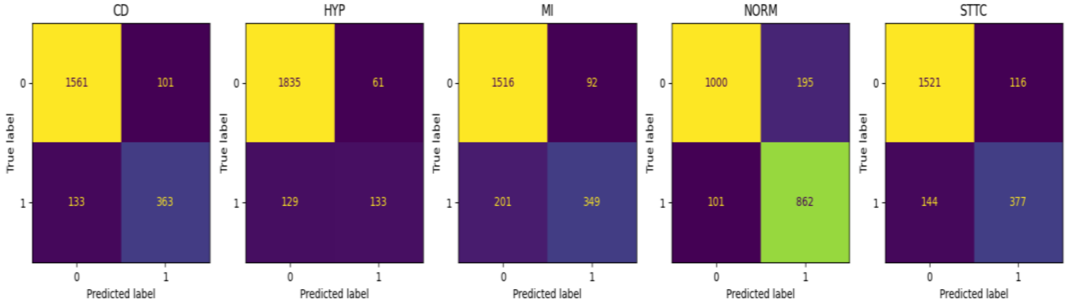


Figure 5. Confusion matrices for each class in PTB-XL with 12 leads

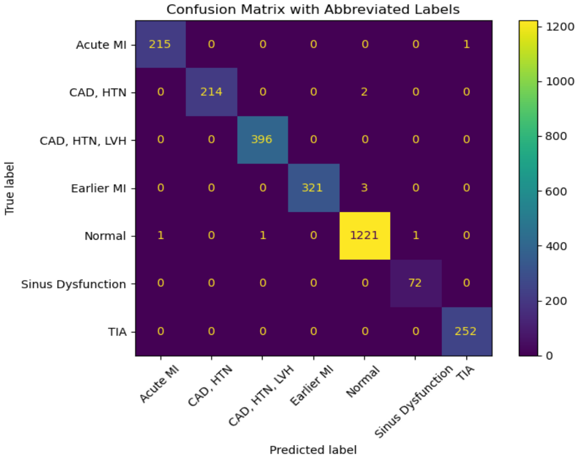


Figure 6. Confusion matrix for the INCART database with 12 leads

The model's high performance on this single-label classification task is further detailed in the confusion matrix in Figure 6, which shows minimal misclassification between the different cardiac conditions. The study notes that this difference in lead importance between the two datasets could be attributed to variations in data labelling methodologies.

Table 7. Model evaluation results, F1 Score (weighted).

ECG Lead	PTB-XL	INCART
I	0.6209	0.9057
II	0.6456	0.9818
III	0.5592	0.9579
AVR	0.6608	0.9588
AVL	0.5793	0.9202
AVF	0.5976	0.9722
V1	0.5688	0.9803
V2	0.5818	0.9863
V3	0.5892	0.9519
V4	0.6080	0.9564
V5	0.6513	0.9466
V6	0.6634	0.9267
All 12 Leads	0.7610	0.9967

Discussion:

At present, it is challenging to directly compare our results with the majority of state-of-the-art (SOTA) models due to the variation in metrics and the number of classes utilized by different researchers in model optimization. Nevertheless, as outlined in the related work section, it is noteworthy that the study referenced as [14] obtained an F1 score of 0.68 for a five-class classification task. In contrast, our approach has achieved an F1 score of 0.761 under similar conditions, demonstrating a notable performance improvement.

Regarding the INCART database, it was not feasible to conduct a comparative analysis based on the F1 score, as no prior research using this database has reported its results using this specific metric. It should be noted that the INCART database is a single-label, multi-class classification, whereas PTB-XL is a multi-label, multi-class classification dataset. The results in the literature and our study also highlight that the complexity of these tasks is considerably different. But, in either case, our introduced CNN-Transformer model performs exceptionally well.

Conclusion:

The results indicate that the use of all 12 leads is not always necessary, as seen in the INCART database, where certain leads provide insights into the detected disease. Reducing the number of leads can decrease the resources required for manufacturing ECG machines. However, this approach may not apply to all cases of disease detection. Moreover, the proposed CNN-Transformer architecture is more robust because of its CNN-based feature extraction and Transformer-based classifier part, which is well-suited for temporal modelling. The simplicity and speed of the model suggest its potential for use in wearable devices and sensors.

A primary challenge is that this deep learning approach does not inherently provide a clear, interpretable reason for a particular diagnosis, which remains a significant hurdle for clinical adoption.

Future Recommendations:

Future work should therefore focus on integrating explainable AI (XAI) techniques to illuminate the model's decision-making process. Further investigation could also explore dynamic lead selection, where the model could identify the most diagnostically relevant leads for a specific patient or suspected condition, optimizing the balance between performance and resource use.

Author's Contribution: Each author equally contributed to this study in its methodology design, architecture design, implementation, and testing.

Conflict of interest: The authors declared that they have no conflicts of interest regarding the publication of this paper.

References:

[1] J. Natarajan, A., Chang, Y., Mariani, S., Rahman, A., Boverman, G., Vij, S., & Rubin, "A wide and deep transformer neural network for 12-lead ECG classification," *IEEE Access*, pp. 1–4, 2020.

[2] M. A. U. Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci. (Ny).*, vol. 415–416, pp. 190–198, 2017, doi: https://doi.org/10.1016/j.ins.2017.06.027.

[3] F. A.-M. Majid Sepahvand, "A novel method for reducing arrhythmia classification from 12-lead ECG signals to single-lead ECG with minimal loss of accuracy through teacher-student knowledge distillation," *Inf. Sci. (Ny).*, vol. 593, pp. 64–77, 2022, doi: https://doi.org/10.1016/j.ins.2022.01.030.

[4] J. X. Shenda Hong, Meng Wu, Yuxi Zhou,Qingyun Wang, Junyuan Shang, Hongyan Li1, "ENCASE: an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks," *Comput. Cardiol. (2010).*, vol. 44, 2017, [Online]. Available: https://www.cinc.org/archives/2017/pdf/178-245.pdf

[5] B. Z. Runchuan Li, Xingjin Zhang, "Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2933473.

[6] W. W. Jian Guan, Wenbo Wang, Pengming Feng, Xinxin Wang, "Low-dimensional

Denoising Embedding Transformer for ECG Classification," *arXiv:2103.17099*, 2021, doi: https://doi.org/10.48550/arXiv.2103.17099.

[7] S. H. W. Yen Chun Chang, "AF Detection by Exploiting the Spectral and Temporal Characteristics of ECG Signals with the LSTM Model," *Comput. Cardiol. (2010).*, 2018, [Online]. Available: https://scholar.nycu.edu.tw/en/publications/af-detection-by-exploiting-the-spectral-and-temporal-characterist

[8] L. Z. Xiuzhu Yang, Xinyue Zhang, Mengyao Yang, "12-Lead ECG arrhythmia classification using cascaded convolutional neural network and expert feature," *J. Electrocardiol.*, vol. 67, pp. 56–62, 2021, doi: https://doi.org/10.1016/j.jelectrocard.2021.04.016.

[9] T. B. S. & A. L. P. R. Antônio H. Ribeiro, Manoel Horta Ribeiro, Gabriela M. M. Paixão, Derick M. Oliveira, Paulo R. Gomes, Jéssica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr., "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nat. Commun.*, 2020, doi: https://doi.org/10.1038/s41467-020-15432-4.

[10] W. W. Ahmed Mostayed, Junye Luo, Xingliang Shu, "Classification of 12-Lead ECG Signals with Bi-directional LSTM Network," *arXiv:1811.02090*, 2018, doi: https://doi.org/10.48550/arXiv.1811.02090.

[11] K. P. Sandra Śmigiel, "ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset," *Entropy*, vol. 23, no. 9, p. 1121, 2021, doi: https://doi.org/10.3390/e23091121.

[12] Z. H. Genlang Chen, "A cascaded classifier for multi-lead ECG based on feature fusion," *Comput Methods Programs Biomed*, 2019, doi: 10.1016/j.cmpb.2019.06.021.

[13] T. I. & M. G. Serkan Kiranyaz, "Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias," *Sci. Reports Vol.*, 2017, doi: https://doi.org/10.1038/s41598-017-09544-z.

[14] A. P. Neha Shukla, "[Retracted] ECG-ViT: A Transformer-Based ECG Classifier for Energy-Constraint Wearable Devices," *J. Sensors*, 2022, doi: https://doi.org/10.1155/2022/2449956.

[15] S. Y. Yunan Wu, Feng Yang, Ying Liu, Xuefan Zha, "A Comparison of 1-D and 2-D Deep Convolutional Neural Networks in ECG Classification," *arXiv:1810.07088*, 2018, doi: https://doi.org/10.48550/arXiv.1810.07088.

[16] N. S. Patrick Wagner, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, 2020, [Online]. Available: https://physionet.org/content/ptb-xl/1.0.3/

[17] L. A. A. A. L. Goldberger, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, 2000, doi: 10.1161/01.cir.101.23.e215.

[18] R. S. T. Shubhojeet Chatterjee, "Review of noise removal techniques in ECG signals," *IET Signal Process.*, vol. 14, no. 9, p. 12, 2020, doi: 10.1049/iet-spr.2020.0104.

[19] C.-C. C. & F. R. Tsui, "Comparing different wavelet transforms on removing electrocardiogram baseline wanders and special trends," *BMC Med. Inform. Decis. Mak.*, vol. 20, 2020, doi: https://doi.org/10.1186/s12911-020-01349-x.

[20] T. M. A. Matloob Khushi, Kamran Shaukat, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3102399.

[21] T. M. Alam *et al.*, "A Machine Learning Approach for Identification of Malignant Mesothelioma Etiological Factors in an Imbalanced Dataset," *Comput. J.*, vol. 65, no. 7, pp. 1740–1751, Jul. 2022, doi: 10.1093/COMJNL/BXAB015.

[22] M. K. Xiaoyan Yang, "Biomarker CA125 Feature Engineering and Class Imbalance

Learning Improves Ovarian Cancer Prediction," *IEEE Asia-Pacific Conf. Comput. Sci. Data Eng.*, 2020, [Online]. Available: https://www.computer.org/csdl/proceedings-article/csde/2020/09411607/1taFbRdyYUw

[23] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," *IEEE Int. Jt. Conf. Neural Netw.*, vol. 2017-May, pp. 1578–1585, Jun. 2016, doi: 10.1109/IJCNN.2017.7966039.

[24] L. I. & P.-A. M. Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.*, vol. 33, pp. 917–963, 2019, doi: https://doi.org/10.1007/s10618-019-00619-1.

[25] I. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, p. 30, 2017.