Combined Convolutional Neural Network-Transformer Network with Hand-Crafted Features for Imbalanced Multi-Label 12-Lead ECG Classification

Tianshi Xie[2], Poulomi Pal[1,2], Elaine Chew [1,2]

[1] Faculty of Life Sciences & Medicine, School of Biomedical Engineering & Imaging Sciences, King's College London, Lambeth Palace Rd, London SE1 7EU, UK
[2] Faculty of Natural, Mathematical & Engineering Sciences, Department of Engineering, King's College London, Strand, London WC2R 2LS, UK

Abstract

Cardiovascular diseases (CVDs) are the world's largest cause of mortality. There is consequently a great need for early and accurate detection of CVDs. To enhance the accuracy of CVD detection from 12-lead electrocardiograms (ECGs), we propose a novel Hand-Crafted Convolution Neural Network-Transformer Network (HC-CNN-TN) model. We used the PTB-XL, which is a large public dataset consisting of ECG data in 5 superclasses and 23 subclasses. The data consists of thousands of 10-second 12-lead ECGs uniformly sampled at 100 Hz. Solving the problem of detecting the category of an ECG from this dataset is a multi-label classification task with imbalanced data. We used a CNN-Transformer to extract high-dimensional features and time-dependent patterns. To enhance model performance, the hand-crafted features we extracted from the 12-lead ECGs are the QRS-complex, RR interval, heart rate, T wave and P wave. The weighted loss function is used to handle the imbalance. We achieved a precision of 0.75, recall of 0.76, F1-score of 0.76, and macro Area Under Curve (AUC) of 0.933. The same evaluation matrix is adopted for subclass classification, which produced similar results. Thus, the proposed model, which integrates hand-crafted and deep learning features, provides a promising way to classify multiple CVD categories with unbalanced samples.

1. Introduction

Cardiovascular diseases (CVDs) are some of the deadliest in the world, leading to approximately 17.9 million deaths, which represents 32% of all global deaths, in 2019 [1]. The most important non-invasive physiological signal used to diagnose CVDs at an earlier stage is the electrocardiogram (ECG). It is the electrical impulse of the atria and ventricles propagated to the skin and captured by machines when the heart alternates between contracting and relaxing [2]. The standard 12-lead ECG acquisition system uses 10 skin surface electrodes, namely six limb leads I, II, III, aVR, aVL, aVF, and six precordial (chest) leads V1-V6 [3]. Due to this comprehensive spatial coverage of cardiac electrical activity, the 12-lead ECG signal is considered very accurate in clinical diagnosis. After acquisition of the ECG signal, processing techniques are applied. Final results are derived via either statistical or artificial intelligence (AI) techniques. Presently, deep learning (DL) has become a powerful approach for automated ECG classification, such as Convolutional Neural Network (CNN) with BiLSTM, 1D ResNet34, and multi-branch CNN models [4–6]. However, existing DL models still face accuracy challenges, especially when dealing with complex or imbalanced ECG datasets.

To overcome this shortcoming, we develop a DL framework for multi-label classification of 12-lead ECG signals to increase classification accuracy. Here, we propose a hybrid model, a Hand-Crafted Convolutional Neural Networks–Transformer Networks (HC-CNN-TN), which integrates clinically meaningful hand-crafted features with CNN-Transformer features extracted from the 12-lead ECG signals. The PTB-XL database [7] from whence the ECG signals are obtained as input for evaluation, with model performance assessed on both superclass and subclass levels. By leveraging both domain knowledge and deep feature extraction, the proposed model aims to improve classification performance on complex and imbalanced ECG datasets.

2. Method

2.1. Dataset and Pre-Processing

The PTB-XL dataset [7] used in this paper is a large public dataset of 21799 clinical 12-lead ECGs each 10-seconds

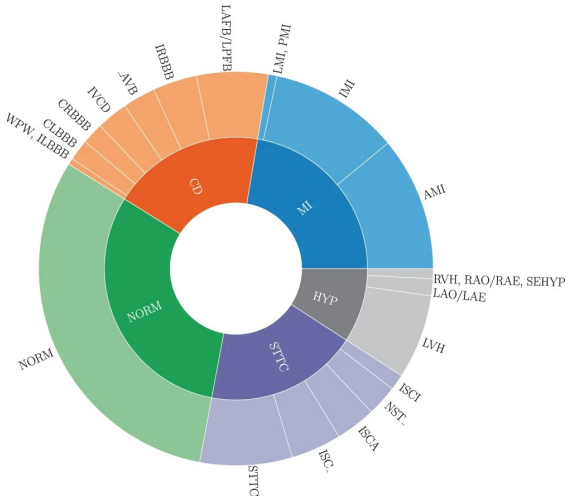

Figure 1: The distribution of PTB-XL: 5 superclasses in the inner circle and 23 Subclasses in the outer circle [7]

long obtained from 18869 patients. It has two sampling frequency versions: 100 Hz and 500 Hz. Since [8, 9] has shown that the two sampling frequencies do not lead to different results, 100 Hz is chosen. Moreover, 100 Hz data can significantly decrease the complexity of the model, speeding up the processing and training. As shown in Figure 1, five superclasses: Normal (NORM), Myocardial Infarction (MI), ST/T Change (STTC), Conduction Disturbance (CD), and Hypertrophy (HYP) are represented in the inner circle, each encompassing several of the 23 subclasses displayed in the outer circle.

In the pre-processing stage, we applied a fourth-order Butterworth bandpass filter to eliminate the baseline drift and artifacts of 12-lead ECG. The frequency response of the filter is in the range 0.5 Hz to 45 Hz. The publisher of the dataset divided the data into ten folds [7]. In this study, the training, validation and test sets are partitioned according to the ratios 8:1:1 for classification by the proposed DL. This project followed the recommended proportion, with the first eight folds serving as the training set, and the remaining parts being the test set and validation set, respectively.

2.2. Classification

This paper proposes a novel model named HC-CNN-TN, as illustrated in Figure 2. The model combines both hand-crafted feature extraction and DNN-based feature learning to leverage the complementary strengths of clinical knowledge and data representation. In the left branch of Figure 2, the hand-crafted features extracted from 12-lead ECG signals are passed through a feedforward module, which consists of a sequence of Linear, Batch Normalization, and ReLU layers. This structure enables the model to integrate features with different dimensions and physical units, and to automatically learn the complex

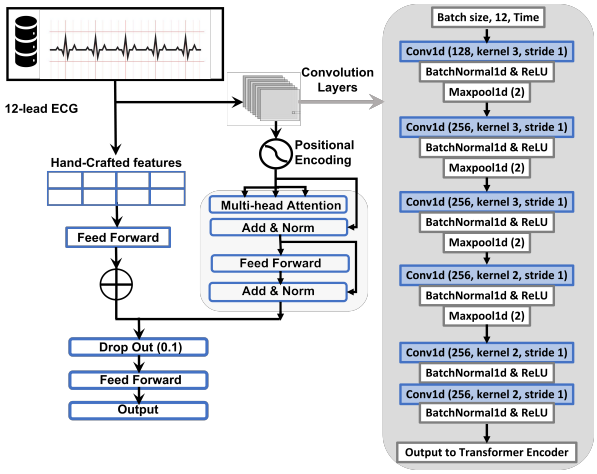

Figure 2: Model architecture of the HC-CNN-TN, the left branch shows the hand-crafted feature processing; and, the right branch is the CNN layers with positional encoding and Transformer encoder

inter-dependencies among them. The ReLU activation introduces non-linearity, enhancing the model's capacity to capture intricate decision boundaries from the hand-crafted inputs. In the DNN branch, shown on the right side of Figure 2, the 12-lead ECG signals are first reshaped into the format [Batch, 12, Time] and then passed through a series of 1D convolutional blocks. Each block consists of Conv1d, BatchNorm1d, ReLU, and MaxPool1d layers, which progressively extract local temporal and spatial fea-

Table 1: Hand-Crafted Features

ECG characteristic	Features
RR interval	mean, median, min RR difference, pNN60, Root Mean Square of Successive N-to-N Differences (RMSSD)
QRS complex	duration mean, duration standard deviation (SD), amplitude mean, amplitude SD, area mean, area SD, slope mean, slope SD
T wave	mean, T-wave alternans (TWA), Multi-scale permutation entropy (MPE)
QT interval	mean
Frequency	mean
Wavelet transform	level 4 entropy, level 3 entropy
Heart rate	min, max, mean, SD
P-wave	correlation coefficient, sample entropy, approximate entropy
Demographic information	age, gender

tures. The output of the convolutional layers is then fed into a Transformer Encoder equipped with positional encoding, enabling the model to capture long-range temporal dependencies across the ECG leads. Finally, the outputs from the Transformer Encoder and the hand-crafted feature feedforward module are concatenated, forming a unified feature representation. This concatenated tensor is then passed through a final feedforward layer with a dropout rate of 0.1 to obtain the output.

Table 1 shows the set of clinically significant hand-crafted features that capture key characteristics of the ECG signal to enhance model performance in multiclass cardiac classification tasks. Furthermore, to address the class imbalance in our multi-label ECG classification task, we employed the BCEWithLogitsLoss function with class-specific positive weights. The positive weight for each class i was computed based on the inverse square root of the class frequency ratio:

$$pos\_weight_i = \left(\frac{N - P_i}{P_i + \varepsilon} \right)^{0.4}, \quad (1)$$

where N is the total number of samples, P_i is the number of positive samples for class i, and $\varepsilon = 10^{-8}$ is a small constant to prevent division by zero. To avoid excessively large weights for rare classes, we applied clipping to the computed weights:

$$pos\_weight_i = \min\left(\max(pos\_weight_i, 0.8), 3.0\right). \quad (2)$$

This weighting scheme moderately emphasizes underrepresented classes while avoiding training instability caused by large weight values. The final class-wise weight vector was passed to the BCEWithLogitsLoss function through its pos_weight parameter.

3. Results

3.1. Experiment Setup

All experiments were conducted on an NVIDIA GeForce RTX 3050 GPU. The hyperparameters used for training are batch size 64, learning rate 3e-4 and 15 epochs. Binary Cross-Entropy loss function and Adam optimizer were chosen, and CosineAnnealingLR was applied as the learning rate scheduler to avoid premature convergence.

3.2. Evaluation and Benchmarks

The results in Table 2 demonstrate strong and balanced performance across NORM, CD, MI, STTC. The HYP classification result also suggests that weighted loss is effective in addressing the challenge of underrepresented classes. In Figure 3, based on the five confusion matrices for the 5 superclasses categories, the model shows a generally stable classification performance, with notable

Table 2: Results for Superclasses

class	precision	recall	F1-score
CD	0.80	0.74	0.77
HYP	0.62	0.62	0.62
MI	0.74	0.77	0.75
NORM	0.83	0.91	0.87
STTC	0.75	0.78	0.76

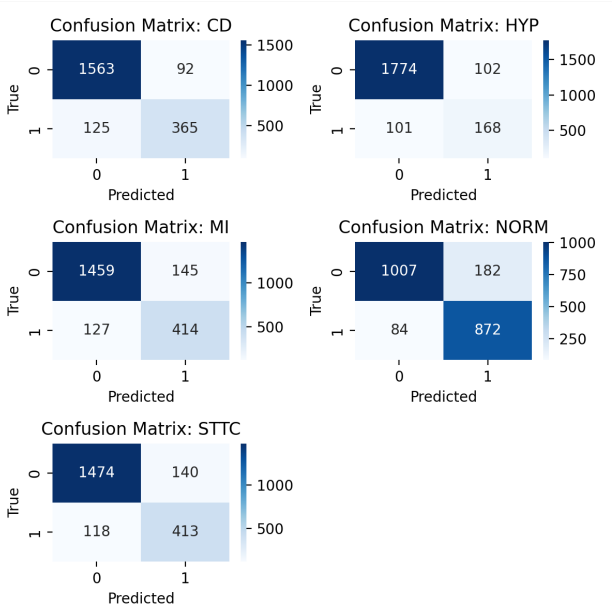

Figure 3: Confusion matrix for superclasses

strengths and some weaknesses. In the confusion matrix, the positive label and negative label are imbalanced: the number of positive labels is significantly less than negative labels. After assigning the weighted loss, the model makes more positive decisions.

In order to test this model on a more difficult task, this project experiments with 23 categories classification in the same dataset. The results include macro precision of 0.74, recall of 0.73, F1-score of 0.73 and AUC of 0.928. The performance is similar to the 5 superclass classification. This indicates that the HC-CNN-TN is also effective in classifying the subtypes of CVDs, and the performance can help guide diagnostic triage. In Table 3, the HC-CNN-TN results have macro precision of 0.75, recall 0.76, F1-score 0.76, and AUC 0.933. Compared to others, HC-CNN-TN has the best precision, recall and F1-score, and relatively high AUC.

4. Discussion

The present work achieves state-of-the-art performance on both the five diagnostic superclasses and the 23 diagnostic subclasses in the PTB-XL dataset, showing its effective-

Table 3: Benchmarks

Model	precision	recall	F1-score	AUC
X-ECGNet [9]	0.74	0.76	0.75	0.93
Lightweight Multi-receptive Field CNN [10]	0.73	0.71	0.72	0.93
CNN with Entropy Features [11]	0.71	0.66	0.68	0.91
HC-CNN-TN	**0.75**	**0.76**	**0.76**	**0.93**

ness in capturing diverse and clinically relevant patterns. To address the label imbalance problem in the dataset, where negative samples are significantly outnumbered by positive ones, a weighted loss strategy was employed during training. Compared with previous studies that mainly relied on CNNs, ResNet-based models, or recurrent networks, our approach consistently achieved high precision, recall, and F1-score in handling the imbalanced nature of the PTB-XL dataset. However, several challenges remain. The proposed model lacks interpretability, and the results are limited to the PTB-XL dataset.

5. Conclusion

To conclude, the proposed HC-CNN-TN model demonstrates strong capability in performing multi-label classification of 12-lead ECG signals. This approach assigns higher importance to underrepresented classes, effectively mitigating bias toward dominant categories and improving the recall and F1-score of minority diagnoses such as HYP. Overall, the results indicate that the HC-CNN-TN framework has state-of-the-art performance for the 12-lead ECG classification task. In future work, we plan to enhance the interpretability of the HC-CNN-TN model and release the attribution results. The attribution will show visually the parts of the ECG that contribute more to the results, thus enhancing confidence in the model's predictions. Moreover, transfer learning will be adopted to adapt the model to specific diagnostic tasks and external datasets. These improvements will broaden the scope of the application, making the model suitable for more specific scenarios.

Acknowledgments

EC and PP are funded by the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (Grant agreement No. 788960).

References

[1] Organization WH. Cardiovascular diseases (cvds) fact sheet, 2025. URL https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] Rahman Protik MN, Khatun F, Islam MM. Analyzing qrs complex, st segment and qt interval of ecg signal to determine the effect of having energy drinks on hypercalcaemia. In 16th Int'l Conf. Computer and Information Technology. 2014; 109–114.

[3] Kaewfoongrungsi P, Hormdee D. The comparison between linear regression derivings of 12-lead ecg signals from 5-lead system and easi-lead system. In 2017 17th International Symposium on Communications and Information Technologies (ISCIT). 2017; 1–6.

[4] Pandey C, Choudhury AD, Khandelwal S. qxai: Quantifiable xai for cardiac diseases. In 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). 2024; 233–238.

[5] Khambampati S, Dondapati S, Kattamuri TV, Pathinarupothi RK. Curenet: Improving explainability of ai diagnosis using custom large language models. In 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). 2023; 1–7.

[6] Ayano YM, Schwenker F, Dufera BD, Debelee TG, Ejegu YG. Interpretable hybrid multichannel deep learning model for heart disease classification using 12-lead ecg signal. IEEE Access 2024;12:94055–94080.

[7] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. Ptb-xl: A large publicly available ecg dataset. Scientific Data 2020;URL https://doi.org/10.1038/s41597-020-0495-6.

[8] van de Leur RR, Blom LJ, Gavves E, Hof IE, van der Heijden JF, Clappers N, Doevendans PAM, Hassink RJ, van Es R. Automatic triage of 12-lead ecgs using deep convolutional neural networks. Journal of the American Heart Association Cardiovascular and Cerebrovascular Disease 2020;9.

[9] Azzem YCH, Harrag F. Explainable deep learning basedsystem for multilabel classification of 12-lead ecg. In 2023 International Conference on Networking and Advanced Systems (ICNAS). 2023; 1–6.

[10] Feyisa DW, Debelee TG, Ayano YM, Kebede SR, Assore TF. Lightweight multireceptive field cnn for 12-lead ecg signal classification. Computational Intelligence and Neuroscience 2022;2022:8413294.

[11] Śmigiel S, Pałczyński K, Ledziński D. ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset. Entropy 2021;23(9):1121. ISSN 1099-4300. URL https://doi.org/10.3390/e23091121.

Address for correspondence:

Tianshi Xie
King's College London, Strand, London WC2R 2LS, UK
tianshi.xie@kcl.ac.uk