2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340561

DWT-CNNTRN: a Convolutional Transformer for ECG Classification with Discrete Wavelet Transform

Congyu Zou[1], Mikhael Djajapermana[2], Eimo Martens[1], Alexander Müller[1], Daniel Rückert[2,4], Phillip Müller[2], Alexander Steger[1], Matthias Becker[3], and Utschick Wolfgang[2]

Abstract— **Cardiovascular diseases are the leading cause of death worldwide. The diagnoses of cardiovascular diseases are usually carried out by cardiologists utilizing Electrocardiograms (ECGs). To assist these physicians in making an accurate diagnosis, there is a growing need for reliable and automatic ECG classifiers.**

In this study, a new method is proposed to classify 12-lead ECG recordings. The proposed model is composed of four components: the CNN(Convolutional Neural Network) module, the transformer module, the global hybrid pooling layer, and a classification layer. To improve the classification performance, the model takes the discrete wavelet transform of ECG signals as the model inputs and utilizes a hybrid pooling layer to condense the most important features over each period.

The proposed model is evaluated using the test set of the China Physiological Signal Challenge 2018 dataset with 12-lead ECGs. It performs with an average accuracy of 0.86 and an average F1-scores of 0.83. The scores are particularly good for the block conditions (LBBB, RBBB, I-AVB). The main advantage of the proposed model is that, it obtains good results with a significantly smaller number of parameters compared to other individual and ensemble models.

Clinical relevance— **This work establishes a new ECG classifier model with high performance and low model size. It can make automatic ECG analysis more accessible, efficient, and accurate, especially in remote or underserved areas.**

Keywords— **ECG classification, Deep learning, Transformer, Discrete wavelet transform, CNN**

I. INTRODUCTION

Cardiovascular diseases are the leading cause of death worldwide [1]. The 12-lead electrocardiogram (ECG) is a classical clinical tool that helps cardiologists to diagnose cardiovascular diseases. However, the significant growth in the number of ECG examinations increases the cardiologists' workload, which eventually increases the chance of misclassifications. The diagnostic process can also be time-consuming due to doctors working with a large amount of heterogeneous data [2]. This has motivated researchers to develop a reliable and automatic system for diagnosing cardiovascular diseases based on ECG to assist cardiologists in making the correct diagnosis. An automatic ECG diagnostic system can be simplified to an ECG classification system or model, which predicts a class of cardiovascular diseases given ECG raw signals. such a system will likely gain in use as wearable health devices grow in popularity. Our goal is to build and optimize such a model.

Deep learning based classifier, in which useful features are learned automatically during the learning process, has proven to be an effective tool to do ECG classification in recent years [3], [4], [5]. The deep learning method was initially successful in classifying images. The Convolutional Neural Network (CNN) from [6] is the most often used method, due to its ability to efficiently extract morphology features from high-dimensional data. Many works applied CNN and achieved good results in ECG classification [3], [4], [7]. For example, the work in [7] used a combination of CNN and RNN, and has achieved the first rank on the China Physiological Signal Challenge 2018. In recent years, an attention-based network from [8], called transformer, has found great success in time-series classification, due to its ability to learn both short- and long-term dependencies. ECG classification, which can also be regarded as a time-series classification, is also applicable to this model architecture. The work in [9] used a deep learning model consisting of two parts: a deep network, which contains convolutional layers followed by a transformer encoder; a wide network, which processes the hand-crafted features. This work demonstrates how well the transformer can perform ECG classification tasks.

Previous works have several limits: models with good performances are always too big, meaning having too many parameters. Domain-specific hand-crafted features are still needed for rhythmical information on the ECG in some outstanding works. We employed discrete wavelet transform (DWT) coefficients to replace all other hand-crafted features. Some previous works focusing on heartbeat classification also use DWT [10], [11], but we apply it for long-term ECG segments instead of heartbeats. We employed Transformer architecture, like in [9], as base architecture, we then applied global hybrid pooling which combines global max pooling and global average pooling to compress the most important features over time [12]. With the innovative modification of the architecture and engagement of the DWT, we successfully reduce the model size significantly while maintaining high performance on ECG classification.

*This work was supported by Klinikum Rechts der Isar der Technische Universität München, Technische Universität München and Fleischhacker GmbH & Co. KG. Fleischhacker GmbH & Co. KG also support the work financially

[1]Congyu Zou, Alexander Müller, Alexander Steger, and Eimo Martens are with Klinikum Rechts der Isar der Technische Universität München, Ismaninger Straße 22, München, 81675, Bayern, Germany . The corresponding author of this paper is Congyu Zou: congyu.zou@tum.de

[2]Mikhael Djajapermana and Utschick Wolfgang are with Technische Universität München, Arcisstraße 21, München, 80333, Bayern, Germany

[3]Matthias Becker is with Fleischhacker GmbH & Co. KG, An d. Silberkuhle 18, Schwerte, 58239, Germany

[4]Daniel Rückert is also with Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX, UK

In this study, we proposed a new ECG classification system called DWT-CNNTRN, which is a CNN transformer with a discrete wavelet transform branch. This system predicts the ECG classes for ECGs after preprocessing. This classifier achieved good classification performance with significantly fewer parameters than the state-of-the-art models.

To summarize, The main contributions of this research include: (1) The successful utilization of the discrete wavelet transform of several-heartbeats ECG signals to enhance classification performance. (2)The presentation of a lighter and better set of hyper-parameters for constructing the CNN and transformer modules for ECG classification. (3) The implementation of a global hybrid pooling layer in the field of ECG classification that combines the benefits of max pooling and average pooling layers, thereby enhancing classification performance.

II. Methods

The overall workflow of the method is illustrated in Figure 1a. The raw ECG signal will be fed to the model after data preprocessing, the model will output a 9-dimensional vector representing the probabilities of 9 different classes the ECG could be. The class with the highest probability will be regarded as the predicted class of the ECG from the model.

A. Data Pre-processing

The first pre-processing step is to remove the baseline wander of the ECG recordings. The baseline wander removal has two steps: 1) apply two consecutive median filters of 100 ms and 300 ms on the ECG recording to estimate the baseline. 2) subtract the baseline from the original ECG.

The next pre-processing step is to unify the length of the ECG recordings into $\ell = 20$ s(seconds). Longer ECG will be truncated, while it is less than 20 s, it will be padded with zeros at the end. Each ECG recording is also resampled to $f_R = 360$Hz to speed up the training process. In this way, the number of samples of each ECG recording is $n = \ell \cdot f_R = 7200$.

B. Apply discrete wavelet transform

Discrete Wavelet Transform (DWT) is a signal processing technique used for decomposing a signal into several subbands. After the selection of a wavelet function, DWT is performed by convolving a signal with a pair of scaling and the wavelet functions. Two sequences of approximation and detail coefficients are produced after the iterative convolution operations.

In this work, the db3 wavelet is used to make a 1-level discrete wavelet transform of the ECG. As illustrated in Figure 1a, each processed 12-lead ECG signal $\mathbf{e} \in \mathbb{R}^{n\times 12}$ is transformed into the approximate coefficients $\mathbf{c}_a \in \mathbb{R}^{\frac{n}{2}\times 12}$ and the detail coefficients $\mathbf{c}_d \in \mathbb{R}^{\frac{n}{2}\times 12}$. Both coefficients are then concatenated on the channel axis to create an output $\mathbf{o} \in \mathbb{R}^{\frac{n}{2}\times 24}$ as a single input data point.

Finally, a batch size $b = 16$ is chosen for training the model. Therefore, a 3-dimensional matrix $\mathbf{x} \in \mathbb{R}^{b\times \frac{n}{2}\times 24}$ will be used as the input for the learning model described in the next section.

C. Model Architecture

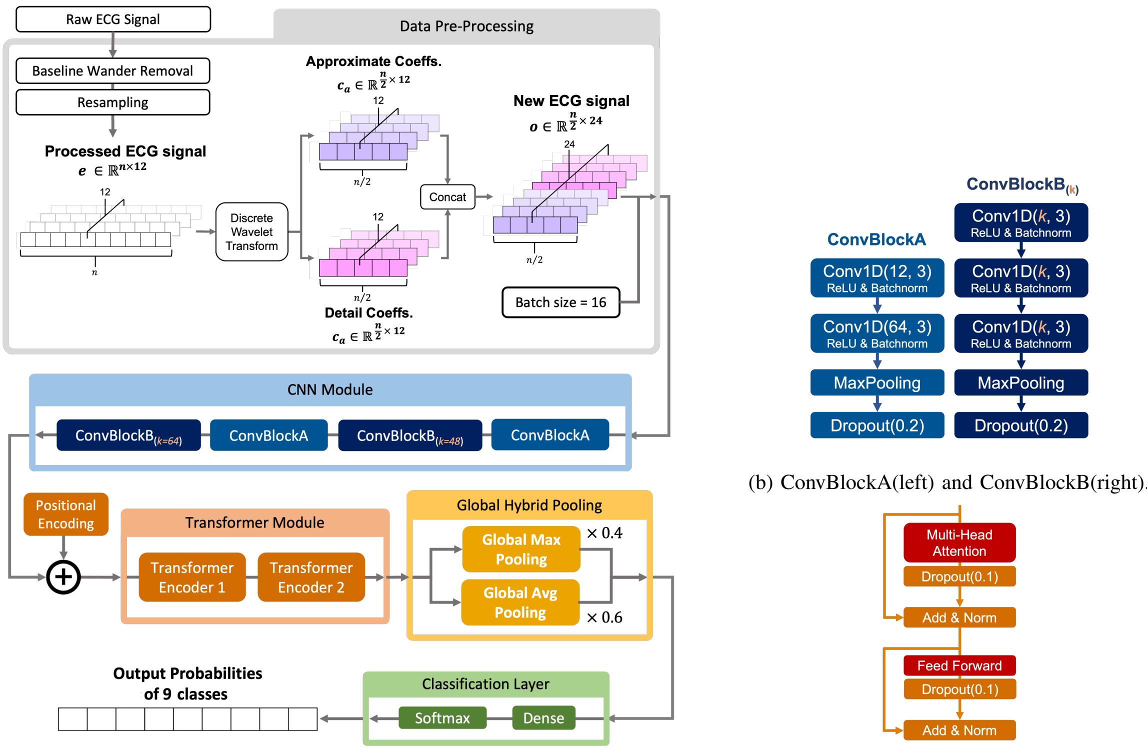
The proposed model is constructed by combining a Convolutional Neural Network(CNN), a stack of transformer Encoder layers, a hybrid pooling layer, and a final classification layer with softmax activation function.

1) CNN Module: The pre-processed ECG signals are taken as input into the CNN module. The CNN module consists of two types of convolutional blocks arranged in alternating turns. As illustrated in Figure 1b, one block, named ConvBlockA, has two convolutional layers, where the first convolutional layer has 12 kernels and the second layer has 64 kernels. The other block, named ConvBlockB, consists of three convolutional layers, each having the same number of kernels. The first ConvBlockB that appears in the CNN module has the number of kernels set to 64, while the second ConvBlockB has the number of kernels set to 48. Every convolution layer on both ConvBlockA and ConvBlockB is set to have a kernel size of 3 and a stride of 1. Padding with zeros is used to make sure the output of each convolution layer has the same dimension as the corresponding input. In addition, a ReLU function and a batch normalization layer follow each convolutional layer. At the end of every convolutional block, there is a max-pooling layer with a pool size of 3 and a dropout layer. The pooling layers summarize the most important features and compress the long ECG signals into a shorter sequence. Meanwhile, the dropout layer is set to drop 20% of the connection to prevent overfitting. By some convolutional and pooling layers, the CNN module learns the feature vectors of ECG signals. These feature vectors represent the extracted local structural information that is useful for classification.

2) Transformer Module: In the next step, the feature vectors produced by the CNN module are summed with a positional encoding. Since the attention mechanism on the transformer is order-agnostic as introduced in [13], the positional encoding is useful to embed positional information to the input. The sum of the CNN's feature vectors and the positional encoding is used as the input token for the transformer module.

The transformer module is formed by stacking two transformer encoders from [8], but with slight modifications. The modified transformer encoder is illustrated in Figure 1. Each encoder layer still consists of the multi-headed attention sub-layer and feed-forward sub-layer, and the sub-layers are connected in a residual structure. The modification comes from the additional dropout layer attached after each sub-layer. Table I summarizes all the used hyper-parameters on the transformer module. Those hyper-parameters are optimized using a small-scale grid search.

Through the attention mechanism, the transformer gives more weight to the parts which are important, and drowns out parts that are less relevant for the classification. The output of the transformer module is then forwarded to the global



(b) ConvBlockA(left) and ConvBlockB(right).

(a) Workflow of the proposed classification method.

(c) Modified Transformer Encoder.

Fig. 1: Method illustration.

TABLE I: Hyper-parameters for the transformer module.

	Transformer Encoder 1	Transformer Encoder 2
Num. attention heads	8	6
Embedding dimension	48	48
Feed forward dimension	48	48
Dropout rate	0.1	0.1

hybrid pooling layer, which will be described in the next subsection.

3) Global Hybrid Pooling Layer: The purpose of the global pooling layer is to compress the output of the transformer module into smaller dimensions by pressing only the most important information over the time frame defined by the CNN block. The implemented hybrid global pooling combines the global max pooling and the global average pooling. This is done by creating a convex combination of both pooling types:

$$\alpha \cdot \text{GlobalMaxPooling}(x) + (1 - \alpha) \cdot \text{GlobalAvgPooling}(x) \quad (1)$$

where $\cdot$ denotes the element-wise multiplication, x denotes the input, and α is set to 0.4, which is found by trial and error to produce the best performance.

4) Classification Layer: Finally, the classification layer learns to assign each ECG signal to one of the nine different classes. The classification layer consists of one dense layer with a softmax activation function. The dense layer has nine channels corresponding to the nine classes, and the softmax activation function outputs the predicted probabilities of the ECG signal associated with each of the nine classes of heart arrhythmia, and then assigns the most probable class to the corresponding ECG signal.

5) Training Setting: The training of the model is carried out using ADAM optimizer introduced in [14] with a learning rate of $\alpha = 0.001$ and parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$. The training objective is to minimize the cross-entropy between the output label of the model and the ground truth label, which is given by:

$$H(x,t) = -\sum_i p(x,i) \cdot \log(p(x,t)), \quad (2)$$

where $p(x,i)$ is the probability the model assigns the label i for the input x, and t is the ground truth label.

D. Performance Metrics

The classification performance of a learning model can be evaluated by Accuracy, Precision, and Recall. These performance metrics are calculated according to the following equations:

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \quad (5)$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative classification. While Precision measures the frequency of classification for the positive class that is actually true, Recall is a measure of how often the positive class is classified as such. The F1 score is the harmonic mean of Precision and Recall:

$$F_1 = \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2}\right)^{-1} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{6}$$

III. Results and Discussion

We use the dataset that comes from China Physiological Signal Challenge (CPSC) 2018 [15] to validate our methods. Details about this dataset are introduced in the Appendix. We choose this dataset because it is one of the most popular and rather abundant and balanced ECG datasets. 90% of the CPSC data is used as training data and the rest as test data. The training time for our model on a machine with Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz, 32GB RAM, and a GeForce RTX 2080Ti as the single GPU, is roughly 10 min with 120 training epochs.

A. Performance Evaluation

The proposed model achieves an average accuracy of 0.86, an average F1-score of 0.83, and an Area Under the receiver operating characteristic Curve(AUC) score of 0.97. The detailed performance of each class is shown in Table II and Figure 2. The model obtains a very good F1-score of more than 0.90 for three classes of arrhythmia, namely IAVB, AF, and RBBB. In particular, the highest F1-score is obtained for RBBB with 0.93. Four classes, namely STD, PVC, NSR, and LBBB, also have good F1-scores in the range of 0.80-0.90. However, the performance for the PAC and STE classes is not as favorable, with the lowest F1-score being recorded for the STE class at 0.63.

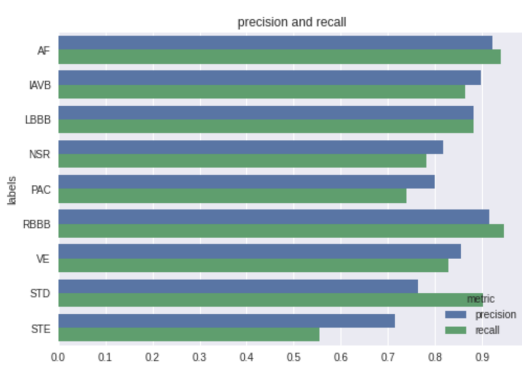


Fig. 2: Precision and recall of the purposed model for each class.

The proposed model has a very good performance for the classification of block conditions (IAVB, LBBB, and RBBB), while poor at classifying PAC and STE. The low F1-score of PAC could be caused by the cropping strategy, since the indicators of the PAC can appear after the cropping boundary. The lowest F1-score for STE is likely to be a consequence of data imbalance, since the number of STE samples in the dataset is 4 times less than STD and the model performs well on STD, whose indicators also lie in ST-segment. And STE is mostly falsely identified as NSR, which is indicated in the confusion matrix shown in Figure 3.

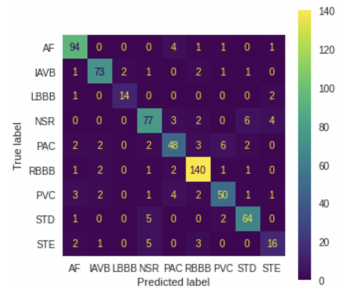


Fig. 3: Confusion matrix of the proposed model on test set.

B. Performance Comparison

The comparison of the F1-scores among recent similar works is shown in Table II. Overall, the proposed model is ranked second on the macro average F1-score. Although overall worse than Chen's model from [7], our model performs better on more classes, and significantly better on IAVB, RBBB, and STD.

When comparing the performance, metrics on label prediction are not the only factor we care about. The number of parameters becomes an important factor to consider when designing a learning model, as wearable health devices are becoming more popular. Having fewer parameters means reduced storage requirements and lower computational resources needed during inference. The comparison of the number of parameters between the most recent three models is shown in Table III. It is clear that our proposed model possesses significantly fewer parameters.

C. Ablation Analysis

To prove the effectiveness of applying discrete wavelet coefficients and the hybrid pooling layer, we also apply ablation analysis to our proposed model. In Figure II, we compared our proposed method to the model trained without discrete wavelet coefficients as additional features of ECG, and compare it further to the model which employs a normally-used average pooling layer instead of a hybrid pooling layer. It is clear to see that discrete wavelet coefficients and hybrid pooling layer indeed improved the model's macro-average F1-score. Within them, the hybrid pooling layer mainly increases model performance on STD and STE classes, and discrete wavelet coefficients mainly increase model performance on LBBB, NSR, PAC, and STE classes.

D. Transformer's Attention Mapping Analysis

Since we used the transformer in our model architecture, we could visualize the attention scores within the multi-head attention layer to get an insight into how the transformer pays attention to different parts of the ECG when it makes

TABLE II: F1-scores Comparison of the Proposed Methods and Other Similar Works

Work	Macro-Average	AF	IAVB	LBBB	RBBB	NSR	PAC	PVC	STD	STE
Ahmed et al. [16]	0.74	0.77	0.74	0.71	0.82	0.74	0.60	0.81	0.66	0.30
Yongchao et al. [17]	0.78	0.85	0.83	0.81	0.87	0.80	0.73	0.82	0.79	0.55
Chao et al. [18]	0.79	0.86	0.88	0.80	0.87	0.82	0.62	0.83	0.71	0.69
Yao et al.[19]	0.81	0.92	0.85	0.87	0.93	0.79	0.74	0.86	0.79	0.56
Chen et al.[7]	**0.84**	**0.93**	0.87	**0.88**	0.91	0.80	**0.83**	0.87	0.81	0.62
Proposed	0.83	0.91	**0.91**	0.85	**0.93**	**0.84**	0.76	0.80	**0.87**	**0.63**
(-) DWT	0.80	0.92	0.88	0.80	0.94	0.78	0.69	0.82	0.84	0.55
(-) HybridPooling	0.79	0.92	0.87	0.84	0.93	0.79	0.71	0.80	0.81	0.48

TABLE III: Comparison on the number of parameters between the proposed model and two other state-of-the-art models.

	Proposed	Chen et al.[7]	Yao et al. [19]
Number of Parameters	213.601	280.350	4.984.640

different arrhythmia classification decisions. As described in Section II-C, The output of the CNN module is regarded as input tokens for the transformer module. The token can represent the ECG within the receptive field of the particular neuro, The length of the fields is calculated to be 266 according to Eq.7:

$$\mathrm{rf}_\ell = \mathrm{rf}_{\ell-1} + \left(f_\ell - 1\right) \cdot \prod_{i=1}^{k-1} s_i \tag{7}$$

where rf_ℓ denotes the size of receptive field of ℓ-th layer, f_ℓ is the kernel size of ℓ-th layer, and s_ℓ is the stride of ℓ-th layer.

To plot the attention mappings in Figure 4: we selected the first head of the multi-head attention layer of the first transformer encoder, the attention scores are illustrated with the width and color density of the lines, where wider and darker lines mean higher attention scores. In the plots, the clips of ECG on the top and on the bottom are the same, while the top ones represent the deeper features processed through the transformer.

For some classes (especially PAC, PVC, STD, and STE), the transformer's attention is mainly on specific segments of the ECG signals where the irregularity appears. Figure 4 shows the attention map for PAC signals, where the transformer pays its attention to the segment with the premature beat. The transformer behaves similarly to PVC, but with stronger attention. For STD and STE conditions, the attention is mainly focused on the segment with ST depression and elevation, respectively. When the cardiac irregularities exhibit a pattern throughout the ECG, the transformer tends to distribute similar attention across the entire signal. The classes where such effect often occurs are the block conditions (IAVB, RBBB, LBBB), AF, and NSR conditions. Illustrative examples are displayed for the LBBB class in Figure 4, where the transformer focuses on the atypical QRS patterns, and for the AF class, where the attention is nearly uniformly distributed across the ECG due to the consistent absence of the P-wave.

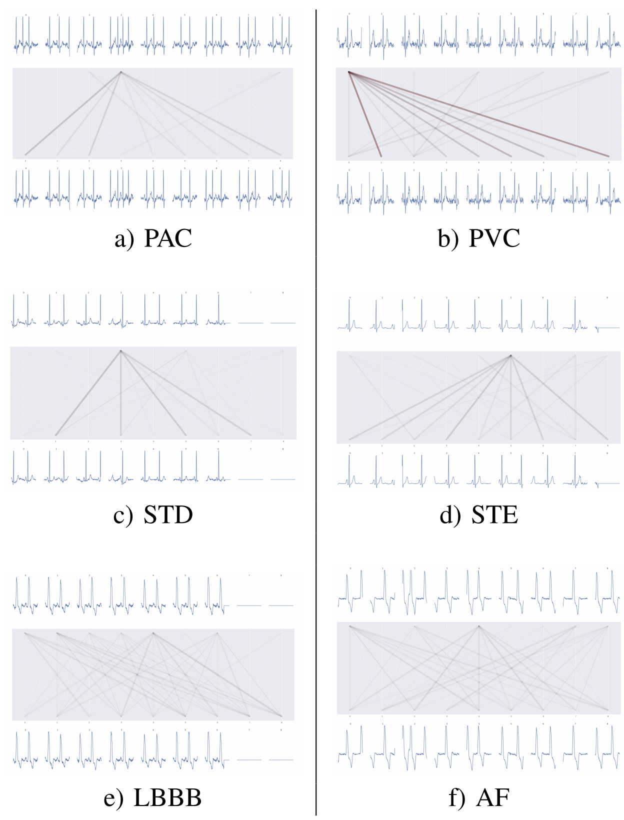


Fig. 4: Transformer's attention mapping in the case of correct classification for different ECG classes.

E. Limitation & Future Work

The biggest limitation of this work is that its performance on STE is still deficient, and far from practical usage. Its overall performance is not significantly better than previous work although it uses significantly fewer parameters and works faster and cheaper. This work is limited to classifying resting 12-leads ECG, a new classifier should be trained for other kinds of ECG.

In our work, the levels of transform and sets of coefficients were chosen using a narrow grid-search algorithm, so there could be some different coefficients selected by another heuristic optimization algorithm that outperforms our proposed method. Furthermore, the feature map offered by continuous wavelet transform might provide better results as suggested in [20]. Besides that, as discussed in Section III-A, a data imbalance could be the main factor undermining the model performance. Therefore, some techniques like SMOTE from [21] to offset data imbalances should be employed in future work to improve the performance.

IV. Conclusion

In this work, a new deep-learning model has been proposed for ECG classification. The model consists of four components: the CNN module, the transformer module with positional encoding, the global hybrid pooling layer, and a classification layer. The performance of the proposed model was evaluated using the CPSC2018 dataset, which achieved an average accuracy of 0.86 and an average F1 score of 0.83. The model showed exceptional performance for block conditions (I-AVB, LBBB, RBBB), however, it didn't perform as well on PAC, PVC, and STE categories. Nevertheless, the classification results still show that the proposed model can produce competitive F1-scores. The key advantage of the model is its ability to provide comparable categorization performance with significantly fewer parameters compared to other models.

Appendix

A. Dataset

China Physiological Signal Challenge (CPSC) 2018 has a training set of 6,877 12-lead ECG recordings and a test set with 2,954 ECGs. Each recording is sampled at 500 Hz, and lengths of the recordings ranges from 6 to 60 seconds. The dataset labels the 12-lead ECG recordings into nine classes of heart arrhythmias. The class distribution of CPSC 2018 can be seen in Table IV.

TABLE IV: ECG class distribution on the dataset of CPSC 2018 from [15].

ECG Class	**Number of Records**
NSR: Normal Sinus Rhythm	918
AF: Atrial Fibrillation	1221
I-AVB: 1st Degree AV Block	722
LBBB: Left Bundle Branch Block	199
RBBB: Right Bundle Branch Block	1675
PAC: Premature Atrial Contraction	544
PVC: Premature Ventricular Contraction	627
STD: ST Depression	786
STE: ST Elevation	185

References

[1] B. B. Kelly, V. Fuster, et al., Promoting cardiovascular health in the developing world: a critical challenge to achieve global health (2010).

[2] Z. Ebrahimi, M. Loni, M. Daneshtalab, A. Gharehbaghi, A review on deep learning methods for ecg arrhythmia classification, Expert Systems with Applications: X 7 (2020).

[3] H. Song, D. Rajan, J. J. Thiagarajan, A. Spanias, Attend and diagnose: Clinical time series analysis using attention models, in: Thirty-second AAAI conference on artificial intelligence, 2018.

[4] P. Nejedly, A. Ivora, I. Viscor, J. Halamek, P. Jurak, F. Plesinger, Utilization of residual cnn-gru with attention mechanism for classification of 12-lead ecg, 2020.

[5] S. W. Chen, S. L. Wang, X. Z. Qi, S. M. Samuri, C. Yang, Review of ECG detection and classification based on deep learning: Coherent taxonomy, motivation, open challenges and recommendations, Biomedical Signal Processing and Control 74 (October 2021) (2022) 103493. doi:10.1016/j.bspc.2022.103493.
URL `https://doi.org/10.1016/j.bspc.2022.103493`

[6] Y. LeCun, P. Haffner, L. Bottou, Y. Bengio, Object recognition with gradient-based learning, in: Shape, contour and grouping in computer vision, Springer, 1999.

[7] T.-M. Chen, C.-H. Huang, E. S. Shih, Y.-F. Hu, M.-J. Hwang, Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model, Iscience 23 (3) (2020) 100886.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.

[9] A. Natarajan, Y. Chang, S. Mariani, A. Rahman, G. Boverman, S. Vij, J. Rubin, A wide and deep transformer neural network for 12-lead ecg classification, in: 2020 Computing in Cardiology, 2020, pp. 1–4.

[10] M. Llamedo, J. P. Martínez, Heartbeat classification using feature selection driven by database generalization criteria, IEEE Transactions on Biomedical Engineering 58 (3 PART 1) (2011) 616–625. doi:10.1109/TBME.2010.2068048.

[11] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, C. Ou, Automatic ECG classification using continuous wavelet transform and convolutional neural network, Entropy 23 (1) (2021) 1–13. doi:10.3390/e23010119.

[12] Z. Tong, G. Tanaka, Hybrid pooling for enhancement of generalization ability in deep convolutional neural networks, Neurocomputing 333 (2019) 76–85. doi:10.1016/j.neucom.2018.12.036.

[13] Y. H. Tsai, S. Bai, M. Yamada, L. Morency, R. Salakhutdinov, Transformer dissection: An unified understanding for transformer's attention via the lens of kernel, CoRR abs/1908.11775 (2019).

[14] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[15] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, E. Ng, An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection, Journal of Medical Imaging and Health Informatics 8 (2018) 1368–1373.

[16] A. Mostayed, J. Luo, X. Shu, W. Wee, Classification of 12-Lead ECG Signals with Bi-directional LSTM Network (2018) 1–16arXiv:1811.02090.
URL `http://arxiv.org/abs/1811.02090`

[17] Y. Wang, B. Xiao, X. Bi, W. Li, J. Zhang, X. Ma, Pay Attention and Watch Temporal Correlation: A Novel 1-D Convolutional Neural Network for ECG Record Classification, 2019 Computing in Cardiology Conference (CinC) 45 (2019) 2019–2022. doi:10.22489/cinc.2019.185.

[18] C. Che, P. Zhang, M. Zhu, Y. Qu, B. Jin, Constrained transformer network for ECG signal processing and arrhythmia classification, BMC Medical Informatics and Decision Making 21 (1) (2021) 1–13. doi:10.1186/s12911-021-01546-2.
URL `https://doi.org/10.1186/s12911-021-01546-2`

[19] Q. Yao, R. Wang, X. Fan, J. Liu, Y. Li, Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network, Information Fusion 53 (2020) 174–182.

[20] P. S. Addison, Wavelet transforms and the ECG: A review, Physiological Measurement 26 (5) (2005). doi:10.1088/0967-3334/26/5/R01.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique Nitesh, Journal of Artificial Intelligence Research 16 16 (2) (2002) 321–357. doi:10.1002/eap.2043.