

Effect of Signal Denoising on 1D-CNN Performance for ECG Beat Classification

Sumetee Jirapattarasakul¹, Rattawan Rattanakanpirom², Thaweesak Yingthawornsuk³

Abstract—Electrocardiogram beat classification plays an important role in computer-aided cardiac diagnosis. Although signal denoising is commonly applied before classification, its actual contribution to deep learning performance remains unclear because denoising can reduce noise while also modifying discriminative waveform morphology. This study investigates the effect of signal denoising on a fixed one-dimensional convolutional neural network for electrocardiogram beat classification. Three input conditions were compared: raw signal, bandpass-filtered signal, and wavelet-denoised signal. To improve evaluation reliability, the final experiment was conducted on four heartbeat classes using a record-wise protocol. The results showed that wavelet-denoised input achieved the highest overall accuracy and weighted F1-score, while raw ECG achieved the highest macro F1-score. These findings suggest that denoising can improve dominant-class recognition, but preserving original beat morphology remains important for more balanced arrhythmia classification.

Index Terms—ECG, arrhythmia classification, beat classification, denoising, 1D-CNN

I. INTRODUCTION

Electrocardiogram (ECG) analysis is one of the most widely used noninvasive techniques for monitoring cardiac activity and assisting the diagnosis of arrhythmia. Manual interpretation of long-term ECG recordings is time-consuming and subject to inter-observer variability, which has motivated extensive research on automatic heartbeat classification [1], [2], [4]. In recent years, deep learning approaches, especially convolutional neural networks, have shown strong performance in ECG classification tasks because they can learn relevant features directly from waveform data [6], [8], [11].

A common practice in ECG preprocessing is to apply denoising techniques before classification. Bandpass filtering is often used to suppress baseline wander and high-frequency noise, while wavelet denoising is frequently applied to reduce artifacts while preserving useful signal components [3], [5]. However, denoising is not always guaranteed to improve classification performance. For heartbeat classification, the classifier often depends on subtle morphological patterns around the P wave, QRS complex, and T wave. If denoising is too aggressive, important discriminative information may be attenuated or distorted.

Several recent studies have reported strong classification performance using different deep learning architectures, including CNN, CNN-LSTM, attention-based recurrent networks, and residual models [9], [10]. However, many of these works modify both preprocessing and model design at the same time, making it difficult to determine whether the improvement comes from denoising itself or from the classifier architecture.

Therefore, this study focuses on a controlled experimental question: *Does signal denoising improve the performance of a fixed 1D-CNN architecture for ECG beat classification?* To answer this question, the same 1D-CNN model was trained under three input conditions: raw ECG beats, bandpass-filtered ECG beats, and wavelet-denoised ECG beats. Only the preprocessing condition was changed, while the architecture, training setup, and evaluation protocol were kept constant.

The contributions of this paper are as follows:

- A controlled comparison of three ECG input conditions under the same 1D-CNN architecture.
- A record-wise experimental protocol to reduce data leakage between training and testing.
- A four-class evaluation setup that excludes the highly underrepresented Q class to improve metric reliability.
- A class-wise analysis showing how denoising affects dominant and minority heartbeat classes differently.

II. RELATED WORK

Recent non-survey studies have demonstrated that deep learning is highly effective for ECG heartbeat classification. Merdjanovska and Rashkovska proposed a patient-specific CNN-based approach for single-lead heartbeat classification [1]. Ullah *et al.* investigated deep learning methods for arrhythmia classification and showed that convolutional architectures can achieve competitive performance [2]. Darmawahyuni *et al.* combined rhythm- and beat-based deep representations for heart abnormality classification [4].

Hybrid architectures have also received attention. Yao *et al.* combined CNN and GRU for ECG heartbeat interpretation [3]. Ramachandran *et al.* proposed a hybrid CNN-LSTM

¹Department of Electrical and Information Engineering, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand

^{2,3}Department of Media Technology, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand

framework for ECG classification [5]. Islam *et al.* introduced HARDC, which combined dilated CNN with hierarchical attention-based recurrent modeling [9]. In addition, Khan *et al.* presented a one-dimensional deep residual network for ECG classification [6], while Issa *et al.* developed a deep model for single lead-II ECG heartbeat classification [8]. Zhou *et al.* addressed inter-patient heartbeat classification with a combination of multi-layer perceptron, capsule weighting, and sequence modeling [10]. More recently, Deng *et al.* proposed an instruction-driven one-dimensional CNN processor for ECG classification [11]. Pham *et al.* also presented heartbeat classification for arrhythmias and myocardial infarction using ECG signals [7].

Although these studies achieved promising results, direct comparison remains difficult because many works adopt different data splits, augmentation strategies, denoising procedures, and architectures. In contrast, the present work intentionally limits the scope to one classifier and varies only the signal preprocessing condition. This design allows a more direct assessment of the impact of denoising.

III. MATERIALS AND METHODS

A. Dataset and Class Grouping

This study used the MIT-BIH Arrhythmia Database, which is widely employed in recent ECG classification studies [4], [6]–[8]. The original heartbeat annotations were initially considered under the commonly used AAMI-style grouping. However, the Q class contained only 8 training beats and 7 testing beats, which was too limited for reliable evaluation. Therefore, the Q class was excluded from the final experiment, and the classification task was conducted on four classes:

- **N**: normal and related beats
- **S**: supraventricular ectopic beats
- **V**: ventricular ectopic beats
- **F**: fusion beats

A record-wise split was applied to reduce data leakage between the training and testing sets. In this experiment, 22 records were used for training, 22 records were used for testing, and 4 records were excluded. After beat extraction and class mapping, the final dataset contained 50993 training beats and 49684 testing beats.

B. Train/Test Split

Table I summarizes the number of beats in the training and test sets for each class. The test set distribution was highly imbalanced, with class N dominating the dataset, while classes S and F contained far fewer samples.

C. Beat Extraction and Normalization

Each ECG beat was extracted using a fixed window centered around the annotated beat position. The extracted beat length was 259 samples, corresponding to 99 samples before the reference point and 160 samples after it. Each beat was then normalized using z-score normalization before being used as input to the classification model.

TABLE I
NUMBER OF BEATS IN THE TRAINING AND TEST SETS

Class	Train	Test
N	45847	44239
S	944	1837
V	3788	3220
F	414	388
Total	50993	49684

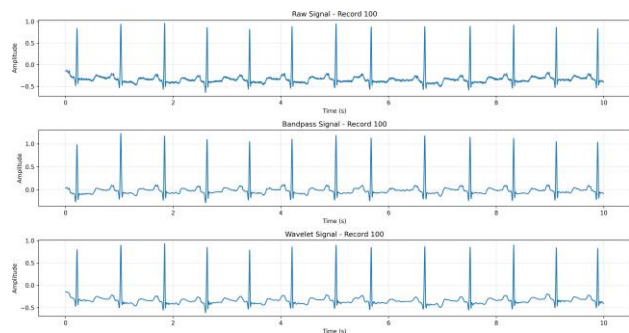


Fig. 1. Comparison of raw, bandpass-filtered, and wavelet-denoised ECG signals from the same record segment.

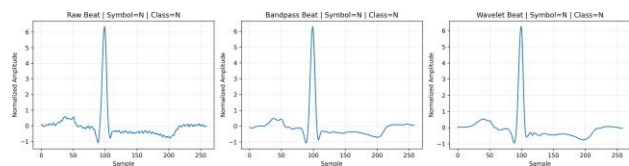


Fig. 2. Comparison of the extracted beat under raw, bandpass-filtered, and wavelet-denoised conditions.

D. Preprocessing Conditions

Three input conditions were compared in this work.

1) *Raw*: In the first condition, the ECG beat was used directly after segmentation and normalization, without additional denoising.

2) *Bandpass*: In the second condition, a third-order Butterworth bandpass filter with a passband of 0.5 to 40 Hz was applied before beat extraction.

3) *Wavelet*: In the third condition, wavelet denoising was applied using the db6 wavelet with level-4 decomposition and soft-thresholding before beat extraction.

E. Visualization of Signal Output

To qualitatively compare preprocessing behavior, signal-level and beat-level outputs were visualized. Fig. 1 shows the raw, bandpass-filtered, and wavelet-denoised ECG signals from the same record segment. Fig. 2 shows the corresponding extracted beat under the three preprocessing conditions. These visualizations help illustrate how each denoising method changes waveform morphology before classification.

TABLE II
1D-CNN ARCHITECTURE USED IN ALL EXPERIMENTS

Layer	Configuration
Input	Beat segment of length 259
Conv Block 1	Conv1D(32, kernel=7), BatchNorm, Max-Pool
Conv Block 2	Conv1D(64, kernel=5), BatchNorm, Max-Pool
Conv Block 3	Conv1D(128, kernel=3), BatchNorm
Pooling	GlobalAveragePooling1D
Dense	Dense(128), Dropout(0.30)
Output	Dense(4), Softmax

F. 1D-CNN Architecture

The same 1D-CNN architecture was used in all experiments. The network consisted of three convolutional blocks followed by a dense classifier. Batch normalization was applied after each convolutional layer, and max pooling was used in the first two convolutional blocks. Global average pooling was then used before the dense and output layers. Keeping the architecture fixed ensured that any performance difference was primarily caused by the preprocessing condition rather than the model design. The output layer contained four neurons with softmax activation, corresponding to the four heartbeat classes.

G. Training Setup

The model was trained using the Adam optimizer with a learning rate of 0.001, batch size of 256, and up to 12 epochs. Early stopping and learning-rate reduction on plateau were used during training. Class weights were also applied to alleviate the effect of severe class imbalance.

H. Evaluation Metrics

The experimental results were evaluated using accuracy, macro precision, macro recall, macro F1-score, weighted F1-score, per-class F1-score, and confusion matrices. Because of the strong class imbalance, macro F1-score was treated as an important performance indicator in addition to overall accuracy.

I. Overall Experimental Workflow

Fig. 3 illustrates the complete pipeline of the study, including dataset preparation, record-wise splitting, beat extraction, preprocessing under three conditions, classification using the same 1D-CNN model, and final performance evaluation.

IV. RESULTS

A. Overall Performance

Table III presents the overall classification performance under the three input conditions. Among the three conditions, the wavelet-denoised ECG signal achieved the highest overall accuracy at 0.8867 and the highest weighted F1-score at 0.8898. In contrast, the raw ECG signal achieved the highest macro precision at 0.4662 and the highest macro F1-score at 0.4302, indicating a better balance across classes. The bandpass-filtered condition produced lower overall performance than both raw and wavelet conditions.

B. Class-wise Performance

Table IV shows the F1-score for each class under the three preprocessing conditions. The wavelet condition achieved the highest F1-score for the dominant N class and the S class, while the raw condition achieved the best F1-scores for the V and F classes. Bandpass remained competitive on class V, but its performance on class N was clearly lower than that of the other two conditions. These results suggest that wavelet denoising can improve dominant-class recognition, whereas raw ECG better preserves abnormal-beat morphology for some minority classes.

C. Confusion Matrix Analysis

Fig. 4 shows the confusion matrices for all three conditions. Under the wavelet condition, the model achieved the strongest classification performance for the dominant N class, which explains the highest overall accuracy. However, the raw condition preserved stronger classification performance for class V and also slightly better recognition of class F. The bandpass condition showed intermediate behavior, with competitive V performance but a larger drop in N-class recognition.

D. Training Curves

Fig. 5 shows the training and validation curves for all three preprocessing conditions. The wavelet condition converged to the best overall test accuracy, whereas the raw condition achieved the best macro F1-score. The bandpass condition showed lower overall performance and less favorable convergence behavior than the other two conditions.

V. DISCUSSION

The updated four-class experiment produced a clearer and more reliable evaluation than the earlier five-class setup because the highly underrepresented Q class was removed. After excluding Q, the results showed that wavelet denoising achieved the highest overall accuracy and weighted F1-score, while raw ECG achieved the highest macro F1-score. This indicates that the effect of denoising depends on which evaluation criterion is emphasized.

A likely explanation is that wavelet denoising improved recognition of the dominant N class, which strongly influenced both accuracy and weighted F1-score due to the severe class imbalance. In contrast, the raw signal preserved more discriminative morphological details for abnormal beats, especially V and F, which improved macro-level balance across classes. This behavior suggests that denoising can be beneficial for majority-class recognition but may also suppress waveform characteristics needed for minority-class discrimination.

The bandpass-filtered condition produced intermediate results. Although it remained competitive on class V, its performance on class N was notably lower than that of the raw and wavelet conditions. This implies that the selected bandpass setting may not preserve discriminative information as effectively as the other representations.

Another key challenge is that classes S and F remain highly difficult under all conditions. Their F1-scores are still

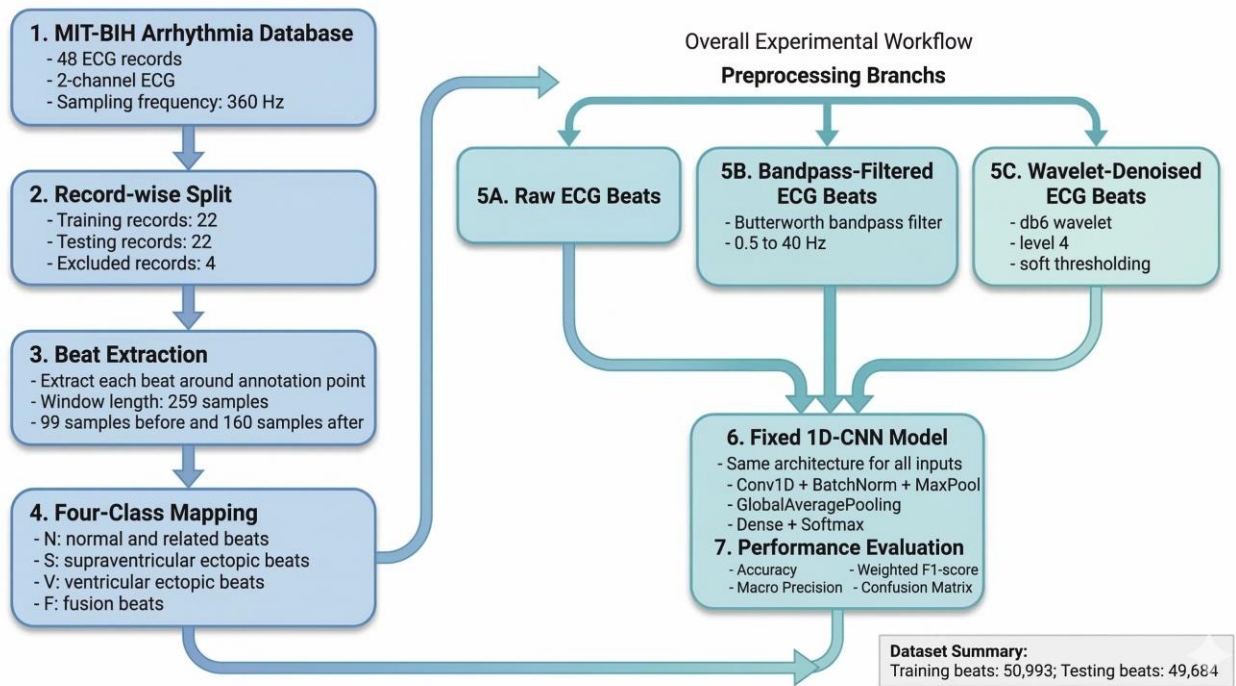


Fig. 3. Overall experimental workflow of the proposed study. The MIT-BIH Arrhythmia Database was divided using a record-wise split, followed by beat extraction, four-class heartbeat mapping, preprocessing under three conditions, 1D-CNN classification, and performance evaluation.

TABLE III
OVERALL PERFORMANCE COMPARISON OF THE THREE INPUT CONDITIONS

Condition	Accuracy	Macro Precision	Macro Recall	Macro F1-score	Weighted F1-score
Raw	0.8648	0.4662	0.4063	0.4302	0.8758
Wavelet	0.8867	0.4136	0.4298	0.4189	0.8898
Bandpass	0.7452	0.4141	0.4229	0.4091	0.8064

TABLE IV
PER-CLASS F1-SCORE UNDER DIFFERENT PREPROCESSING CONDITIONS

Class	Support	Raw	Wavelet	Bandpass
N	44239	0.9273	0.9476	0.8498
S	1837	0.0219	0.0388	0.0380
V	3220	0.7589	0.6881	0.7444
F	388	0.0126	0.0011	0.0040

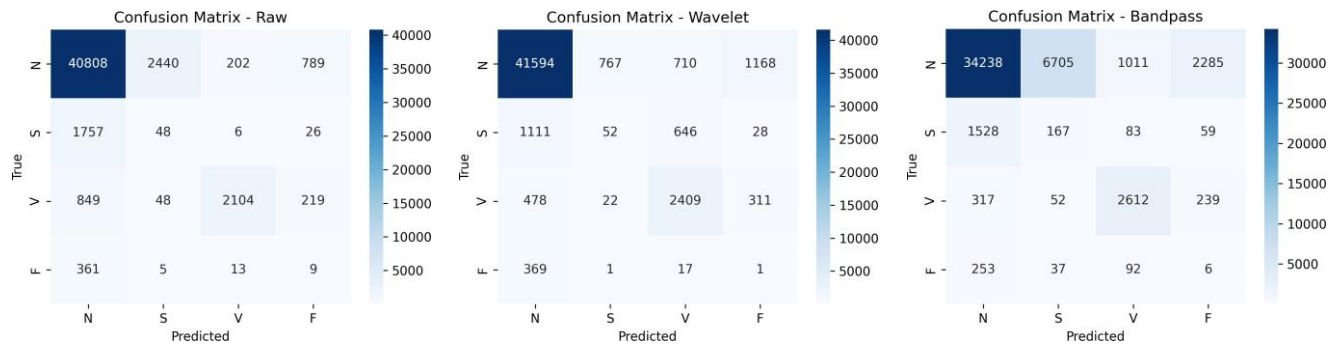


Fig. 4. Confusion matrices of the 1D-CNN model under the raw, wavelet-denoised, and bandpass-filtered conditions for the four-class experiment.

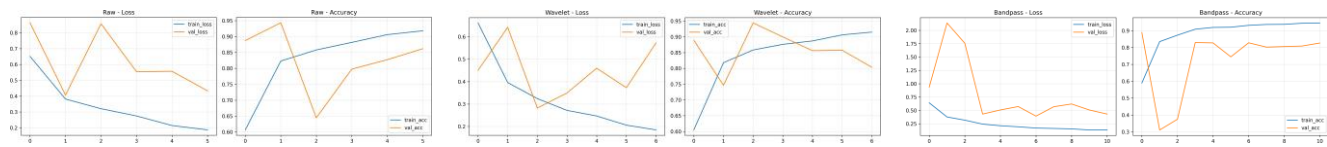


Fig. 5. Training and validation loss and accuracy under the raw, wavelet-denoised, and bandpass-filtered conditions for the four-class experiment.

very low, reflecting both class imbalance and the difficulty of separating these arrhythmic patterns from other classes. Therefore, future work should explore stronger imbalance-aware learning methods, such as focal loss, oversampling, class-balanced sampling, or additional lightweight baselines under the same record-wise protocol.

Overall, the results suggest that there is no single universally best preprocessing method for every metric. Wavelet denoising is advantageous for overall accuracy, whereas raw ECG remains more favorable when balanced class performance is considered.

VI. CONCLUSION

This paper presented a controlled comparison of three ECG input conditions for heartbeat classification using a fixed 1D-CNN architecture: raw ECG, bandpass-filtered ECG, and wavelet-denoised ECG. After removing the highly underrepresented Q class, the final experiment was conducted on four heartbeat classes: N, S, V, and F. The results showed that wavelet denoising achieved the best overall accuracy and weighted F1-score, while raw ECG achieved the best macro F1-score and the strongest performance on the V and F classes. These findings indicate that denoising can improve dominant-class recognition, but preserving original beat morphology remains important for more balanced arrhythmia classification. Future work will investigate parameter-tuned denoising methods, stronger imbalance mitigation, and additional baseline models under the same record-wise evaluation setup.

ACKNOWLEDGMENT

The authors would like to thank the Department of Media Technology and the Department of Electrical and Information Engineering, King Mongkut's University of Technology Thonburi, as well as the developers of the open-source tools and publicly available datasets used in this study.

REFERENCES

- [1] E. Merdjanovska and A. Rashkovska, "Patient-specific heartbeat classification in single-lead ECG using convolutional neural network," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2021, pp. 932–936. <https://doi.org/10.1109/EMBC46164.2021.9630366>
- [2] W. Ullah, I. Siddique, R. M. Zulqarnain, M. M. Alam, I. Ahmad, and U. A. Raza, "Classification of arrhythmia in heartbeat detection using deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, Art. no. 2195922, 2021. <https://doi.org/10.1155/2021/2195922>
- [3] G. Yao, X. Mao, N. Li, H. Xu, X. Xu, Y. Jiao, and J. Ni, "Interpretation of electrocardiogram heartbeat by CNN and GRU," *Computational and Mathematical Methods in Medicine*, vol. 2021, Art. no. 6534942, 2021. <https://doi.org/10.1155/2021/6534942>
- [4] A. Darmawahyuni, S. Nurmaini, M. N. Rachmatullah, B. Tutuko, A. I. Sapitri, F. Firdaus, A. Fanyuri, and A. Predyansyah, "Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification," *PeerJ Computer Science*, vol. 8, Art. no. e825, 2022. <https://doi.org/10.7717/peerj-cs.825>
- [5] D. Ramachandran, R. Suresh Kumar, A. Alkhayyat, R. Q. Malik, P. Srinivasan, G. G. Priya, and A. G. Adigo, "Classification of electrocardiography hybrid convolutional neural network-long short term memory with fully connected layer," *Computational Intelligence and Neuroscience*, vol. 2022, Art. no. 6348424, 2022. <https://doi.org/10.1155/2022/6348424>
- [6] F. Khan, X. Yu, Z. Yuan, and A. U. Rehman, "ECG classification using 1-D convolutional deep residual neural network," *PLoS One*, vol. 18, no. 4, Art. no. e0284791, 2023. <https://doi.org/10.1371/journal.pone.0284791>
- [7] B.-T. Pham, P. T. Le, T.-C. Tai, Y.-C. Hsu, Y.-H. Li, and J.-C. Wang, "Electrocardiogram heartbeat classification for arrhythmias and myocardial infarction," *Sensors*, vol. 23, no. 6, Art. no. 2993, 2023. <https://doi.org/10.3390/s23062993>
- [8] M. F. Issa, A. Yousry, G. Tuboly, Z. Juhasz, A. H. AbuEl-Atta, and M. M. Selim, "Heartbeat classification based on single lead-II ECG using deep learning," *Heliyon*, vol. 9, no. 7, Art. no. e17974, 2023. <https://doi.org/10.1016/j.heliyon.2023.e17974>
- [9] M. S. Islam, K. F. Hasan, S. Sultana, S. Uddin, P. Lio', J. M. W. Quinn, and M. A. Moni, "HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN," *Neural Networks*, vol. 162, pp. 271–287, 2023. <https://doi.org/10.1016/j.neunet.2023.03.004>
- [10] C. Zhou, X. Li, F. Feng, J. Zhang, H. Lyu, W. Wu, X. Tang, B. Luo, D. Li, W. Xiang, and D. Yao, "Inter-patient ECG heartbeat classification for arrhythmia classification: a new approach of multi-layer perceptron with weight capsule and sequence-to-sequence combination," *Frontiers in Physiology*, vol. 14, Art. no. 1247587, 2023. <https://doi.org/10.3389/fphys.2023.1247587>
- [11] J. Deng, J. Yang, X. Wang, and X. Zhang, "A novel instruction driven 1-D CNN processor for ECG classification," *Sensors*, vol. 24, no. 13, Art. no. 4376, 2024. <https://doi.org/10.3390/s24134376>