Deep Learning for Automated ECG Anomaly Detection: A Comprehensive Evaluation

Jane Researcher, John Smith, Emily Chen
Department of Biomedical Engineering, University of Example

Abstract

We present a deep learning approach for automated detection of cardiac anomalies in electrocardiogram (ECG) recordings. Our 1D convolutional neural network achieves 97.3% accuracy on the MIT-BIH Arrhythmia Database using 5-fold cross-validation. We employ data augmentation techniques including signal warping, Gaussian noise injection, and time-shift perturbations to improve model robustness. The proposed method outperforms prior approaches by 2.1 percentage points while using a simpler architecture.

1. Introduction

Electrocardiogram (ECG) analysis is fundamental to the diagnosis of cardiac arrhythmias. Manual interpretation of ECG recordings is time-consuming and subject to inter-observer variability. Automated classification systems can assist clinicians in rapid, consistent diagnosis.

Recent advances in deep learning have shown promise for ECG classification tasks. Convolutional neural networks (CNNs) can learn discriminative features directly from raw signals, eliminating the need for hand-crafted feature engineering. However, challenges remain in achieving robust performance across the diverse morphologies present in real-world ECG data.

In this work, we propose a 1D CNN architecture specifically designed for single-lead ECG anomaly detection. Our key contributions include: (1) a lightweight architecture requiring fewer parameters than existing approaches, (2) a comprehensive data augmentation strategy that improves generalization, and (3) rigorous evaluation using 5-fold cross-validation on a standard benchmark.

2. Methods

2.1 Dataset

We use the MIT-BIH Arrhythmia Database (PhysioNet, DOI: 10.13026/C2F305), which contains 48 half-hour excerpts of two-channel ambulatory ECG recordings from 47 subjects studied at Boston's Beth Israel Hospital. Following the AAMI standard, heartbeats are classified into five categories: Normal (N), Supraventricular (S), Ventricular (V), Fusion (F), and Unknown (Q).

2.2 Preprocessing

Raw ECG signals are bandpass filtered at 0.5-40 Hz using a 4th-order Butterworth filter to remove baseline wander and high-frequency noise. Individual heartbeats are segmented by detecting R-peaks using a threshold-based method and extracting 360-sample windows centered on each peak.

2.3 Model Architecture

Our model consists of three 1D convolutional blocks, each comprising a Conv1d layer, batch normalization, ReLU activation, and max pooling. The three blocks use 32, 64, and 128 filters respectively with kernel size 5. A global average pooling layer reduces the feature maps to a fixed-size vector, followed by a fully

connected classifier with a hidden layer of 64 units and dropout (p=0.5).

2.4 Training

We train with the Adam optimizer using a learning rate of 0.001 for 100 epochs with batch size 32. We employ 5-fold cross-validation to ensure robust evaluation and report the mean performance across all folds. Data augmentation includes three techniques applied during training:

1. Signal warping: Random time-domain warping with magnitude factor 0.1
2. Gaussian noise injection: Additive noise with standard deviation 0.01
3. Time-shift perturbation: Random circular shifts of up to 10 samples

Class imbalance is handled using weighted cross-entropy loss, with weights inversely proportional to class frequency.

2.5 Hardware

All experiments were conducted on a single NVIDIA V100 GPU with 16GB memory. Training a single fold takes approximately 15 minutes.

3. Results

Our model achieves 97.3% accuracy (F1: 0.947) on the MIT-BIH Arrhythmia Database, outperforming prior CNN-based methods by 2.1 percentage points. Per-class results are summarized below:

Class	Precision	Recall	F1
N	0.991	0.986	0.988
S	0.878	0.862	0.870
V	0.952	0.941	0.946
F	0.891	0.837	0.863
Q	0.947	0.962	0.954

The model shows strongest performance on the Normal class and faces the greatest challenge with the underrepresented Fusion class.

4. Discussion

The lightweight architecture achieves competitive performance while requiring significantly fewer parameters than recent transformer-based approaches. The data augmentation strategy proved critical: without augmentation, accuracy drops to 94.1%, demonstrating its importance for generalization.

The 5-fold cross-validation results show low variance (std: 0.3%), indicating that the model generalizes well across different patient subsets. This is particularly important for clinical deployment where robustness across patients is essential.

5. Conclusion

We demonstrate that a relatively simple 1D CNN architecture, combined with effective data augmentation and proper evaluation methodology, can achieve state-of-the-art performance on ECG anomaly detection. Future work will explore multi-lead ECG analysis and transfer learning from larger datasets.

References

[1] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50, 2001.

[2] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220, 2000.

[3] Kachuee M, Fazeli S, Sarrafzadeh M. ECG Heartbeat Classification: A Deep Transferable Representation. IEEE ICHI, 2018.