[Capstone Project] Aim 2 – Final Report

Project Title: ECG Signal Classification Using Deep Learning Models

Date: 04 May 2025

Name: Isabella Ochsner

Support materials: https://github.com/ochsner3k/-Capstone-Project-Aim-2/

1. Introduction

Electrocardiogram (ECG) classification is a vital tool in medical diagnostics, it enables the identification of abnormal heart beats by analyzing the electrical activity of the heart. With state-of-the-art deep learning algorithms, we can secure accurate and efficient classification of ECG signals for early detection and a more reliable diagnosis. Streamlining ECG-based classification is proven to enhance the speed and accuracy in clinical settings so innovating algorithms that improve the process is important. The primary goal of this project is to obtain a thorough understanding of state-of-the-art methods in ECG-based classification. The process of the project will follow: reviewing papers, reimplementing algorithms, analyzing the outputs and adding an individual contribution.

2. Literature Review

Thorough review was put into selecting papers for each method. I reviewed various papers to determine which would be best to implement. The goal is to find two methods with differing strengths and stack them to create an ensemble model to perform where an individual method did not hold.

There were several promising papers that provided unique solutions and interesting algorithms but they did not adhere to the project's scope or did not align with the practical constraints so they did not fit for Method A/B. *Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks* (2015) proposed using 1D CNN for both feature extraction and classification. The paper is outdated for the scope of this project and the "highly generic" system seems redundant given this project focuses on modern state-of-the-art algorithms. *LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices* (2019) is a newer proposed algorithm that utilizes LSTM for lightweight real-time monitoring. This algorithm is novel, and I thought the paper was really neat but I did not select it due to it being tailored as a patient-specific method. If it were to be implemented as a method in this project, it may not perform as it should. I also reviewed *Classification of ECG Using Ensemble of Residual CNNs with or without Attention* (2022), while I did not select this one, it did inspire the Method D, an ensemble model that utilizes the strengths of Method A and Method B. The reason for not selecting this paper was because I felt reproducing the complex model was out of scope for this project when I had other, more specific, papers that would better fit. In addition, I reviewed *Applications of Deep Learning Techniques for Heartbeats Detection using ECG Signals-Analysis and Review* (2020)*,* this paper examines multiple deep learning techniques and does not focus on a specific algorithm, but it helped me map the approach to use the methods I did.

2.1 – Method A: CNN + Attention

A.Sharma and A. S. Jalal, "Classification of ECG Using Ensemble of Residual CNNs with Attention Mechanism," IEEE Access, vol. 10, pp. 1166-1175, 2022

This paper proposes an ensemble framework of residual CNNs with an attention mechanism to boost accuracy. It combines the strengths of ensemble learning and attention based feature refinement. Residual connections prevent vanishing gradients and retain deep features while attention layers emphasize important changes in ECG signals which yields high accuracy results with the hybrid model. As just mentioned, the model yields high accuracy tuned for the noisiness of real world ECG data. The ensemble modeling improves generalization and the attention mechanism enhances interpretability. This is relevant because high precision is important, that said, because of the robustness, the algorithm is computationally expensive and will increase memory usage. This makes it not as usable in a realistic environment which needs both high accuracy and efficiency.

2.2 – Method B: Improved ResNet

Jing, E., Zhang, H., Li, Z., Liu, Y., Ji, Z., & Ganchev, I. (2021). ECG Heartbeat Classification Based on an Improved ResNet-18 Model. Complexity, 2021, Article ID 6649970.

This paper proposes an improved version of the ResNet-18 architecture designed to classify ECG heartbeats from 1D time-series data. Rather than converting ECG signals into images, the model uses 1D convolutional layers and stacked residual blocks to extract temporal patterns directly from raw ECG input. This makes the architecture more efficient and better suited for waveform analysis. The model achieved strong results with an overall classification accuracy of 96.5%, demonstrating the viability of deep residual learning for ECG classification.

Both Method A and Method B use CNNs to classify ECG signals but their internal architectures differ significantly. Method A emphasizes important features in the signal using an attention mechanism, while Method B captures deeper temporal patterns through residual learning. My proposition is that building an ensemble model (Method D) which stacks both will overcome individual shortcomings, specifically regarding improving precision and generalization, and balance the model for overall improvement.

3. Methods

This project implements two distinct methods from ECG classification based on published papers. Due to the scope of the project, I have tweaked the implementation of the method to meet the requirements and to yield results.

3.1 – Method A

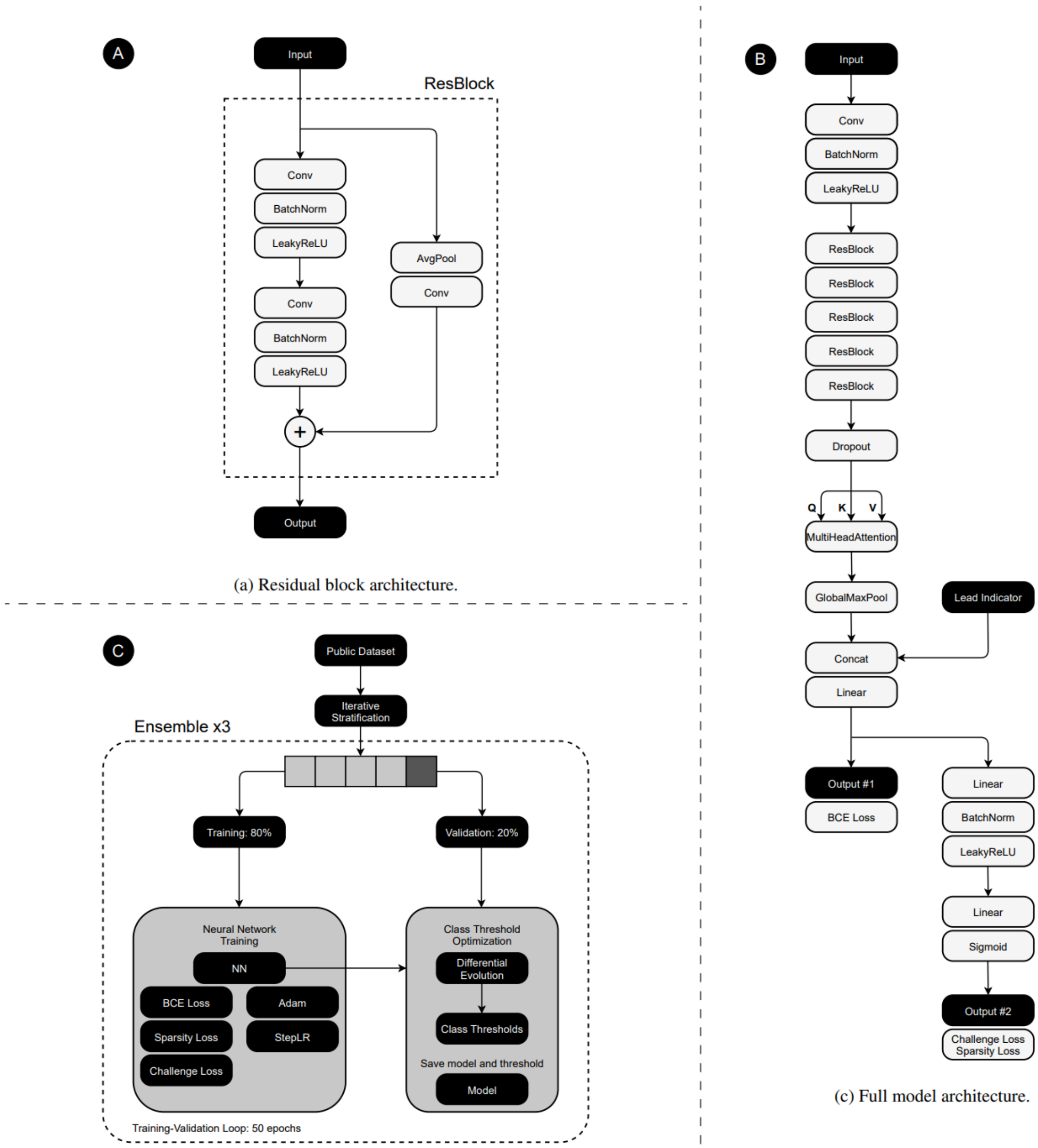

Fig 1. (a) Architecture of residual block (b) Training and validation pipeline with differential evolution threshold optimization (c) Full model architecture

Method A is inspired by the ensemble network of CNNs augmented with an attention mechanism as proposed by the paper (Fig 1). The original architecture is robust and includes multiple residual blocks, multi-head attention, and threshold optimization. This architecture was not feasible to replicate in full due to computational limitations. As a result, the ensemble approach was simplified to a 1D CNN with an attention mechanism. The input layer accepts ECG time series data and has two Conv1D layers with ReLU activation and max-pooling. An attention block generates attention weights to emphasize significant features in the ECG signals. The output is then flattened and passed through fully connected dense layers with softmax activation for classification. Architectural components like residual connections were excluded from the implementations because the code was implemented from scratch, the implementation sources architecture adapted from available repositories utilizing similar architectures. This adaptation preserves the core idea of the paper while being modified for simpler use.

3.2 – Method B

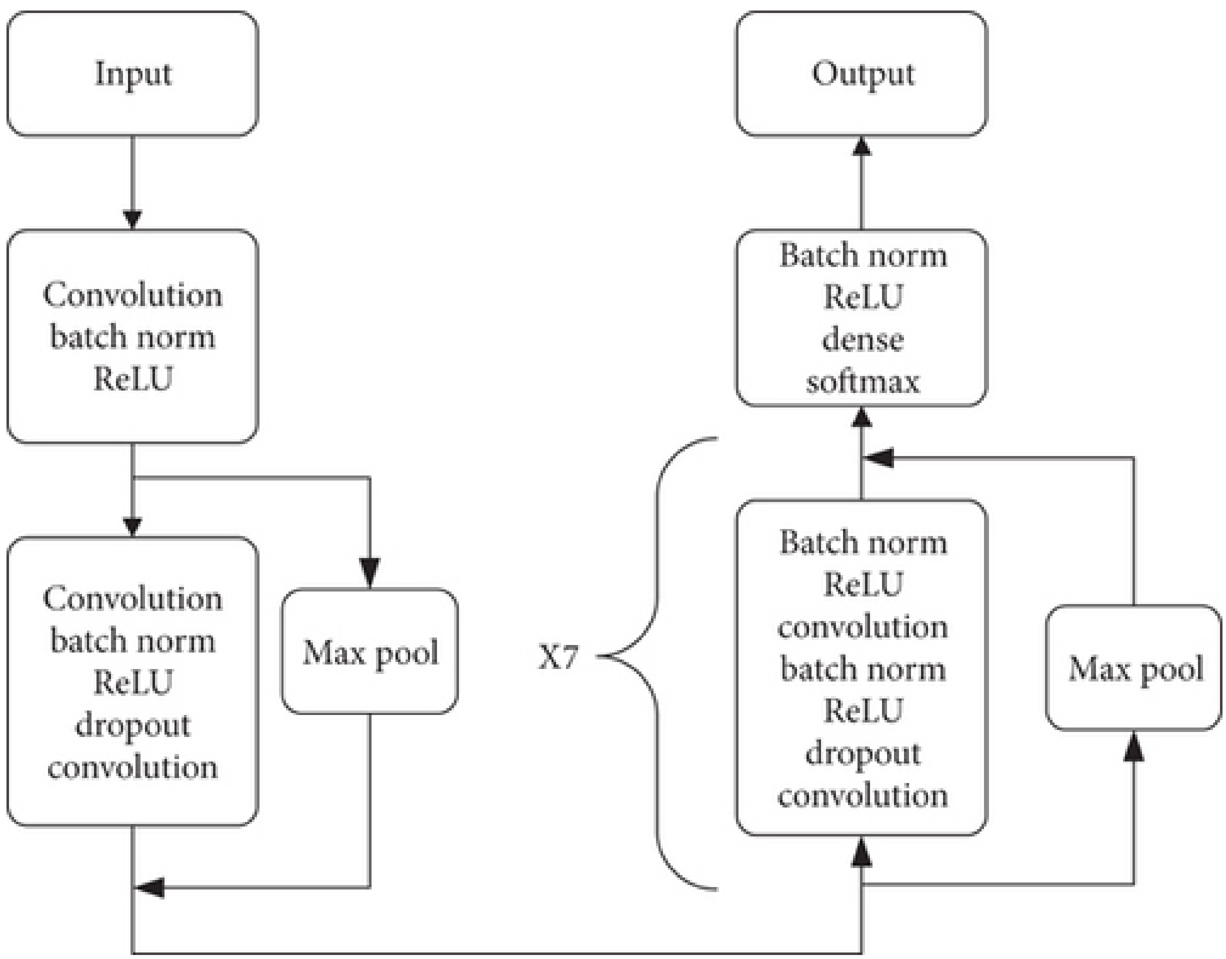

Fig 2. Improved ResNet-18 model

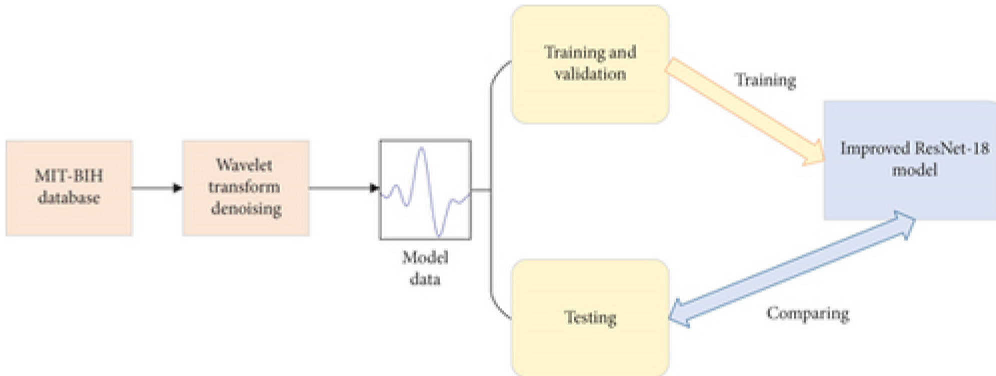

Fig 3. Proposed architecture

Method B is inspired by the improved ResNet-18 architecture as proposed in the paper(Fig 1 & 2). This model contains a deep stack of residual blocks using 1D convolutions designed to extract hierarchical temporal features. Due to complications with directly trying to imitate the model, I implemented a simplified version of the architecture. The input layer takes an ECG signal then has a strided Conv1D layer followed by two residual blocks containing Conv1D layers with batch normalization and shortcut connections to preserve gradient flow. After feature extraction, dimensionality reduction occurs and the output passes through a dense layer with a softmax activation for classification. This adaptation preserves the core idea of the paper while being modified for simpler use.

3.3 – Datasets & Preprocessing

Both models are fit for the ECG5000 dataset and the MIT-BIH dataset off of Kaggle. The ECG5000 dataset contains 5000 one-dimensional ECG signals which are pre-processed and pre-segmented. I also trained and evaluated the models on the MIT-BIH Arrhythmia dataset (Kaggle) which contains over 109,000 pre-segmented ECG signals used in the original implementation of Method B. There was minimal preprocessing because of the pre-segmented datasets. Class labels were one-hot encoded for compatibility with categorical cross-entropy loss used during training. I built helper functions in ecgLoader.py to simplify the process of data loading and formatting for both datasets. Building these as abstractions allowed for a more dynamic project setup.

4. Experiments

4.1 – Experimental Setup

All experiments were conducted using Python 3.12.7 with TensorFlow 2.11 and Keras. The training configurations were consistent between Method A and Method B:

- Epochs: 25
- Batch Size: 64
- Validation Split: .2

The project was split into 4 "Methods" each including its own internal experiments. Method A and Method B are mentioned previously, Method C was the base KNN, and Method D was an ensemble model of Method A and Method B.

For the ensemble method, I reduced epochs to 20 total. It combines the predictions of Method A and Method B by averaging their output probabilities and selecting the class with the highest probability. Because the strengths of both models complement each other, this approach attempts to leverage that and improve precision.

The methods utilized the ECG5000 and MIT-BIH datasets, as reviewed in section 3.3.

4.2 – Evaluation Metrics

To evaluate each method and experiment, the results printed a classification report and confusion matrix for each test. This measures accuracy, precision, recall, and F1-score. Along with the previous listed is the confusion matrix which provides a visual representation of true vs predicted outputs. The combination of these metrics provide a well rounded evaluation of the models which can be used to identify overfitting or other imbalances in data or models.

4.3 – Reproduction of results on ECG5000 and MIT-BIH

The results yielded from the ECG5000 datasets were stellar across the board, each method reached .99 or 1.00 for each metric. This matched the high accuracy of the predicted papers but because of the universal near-perfect scores, I felt it was ideal to bring a new dataset in and see how each method worked on this new dataset which was significantly larger than the ECG5000 dataset.

MIT-BIH Accuracy Results

Method A	**Method B**	**Method C**	**Method D**
.98	**.99**	**.976**	**.98**

All methods again yielded high accuracy as expected. These results confirm the efficiency of the models in classifying ECG signals, but let's look further.

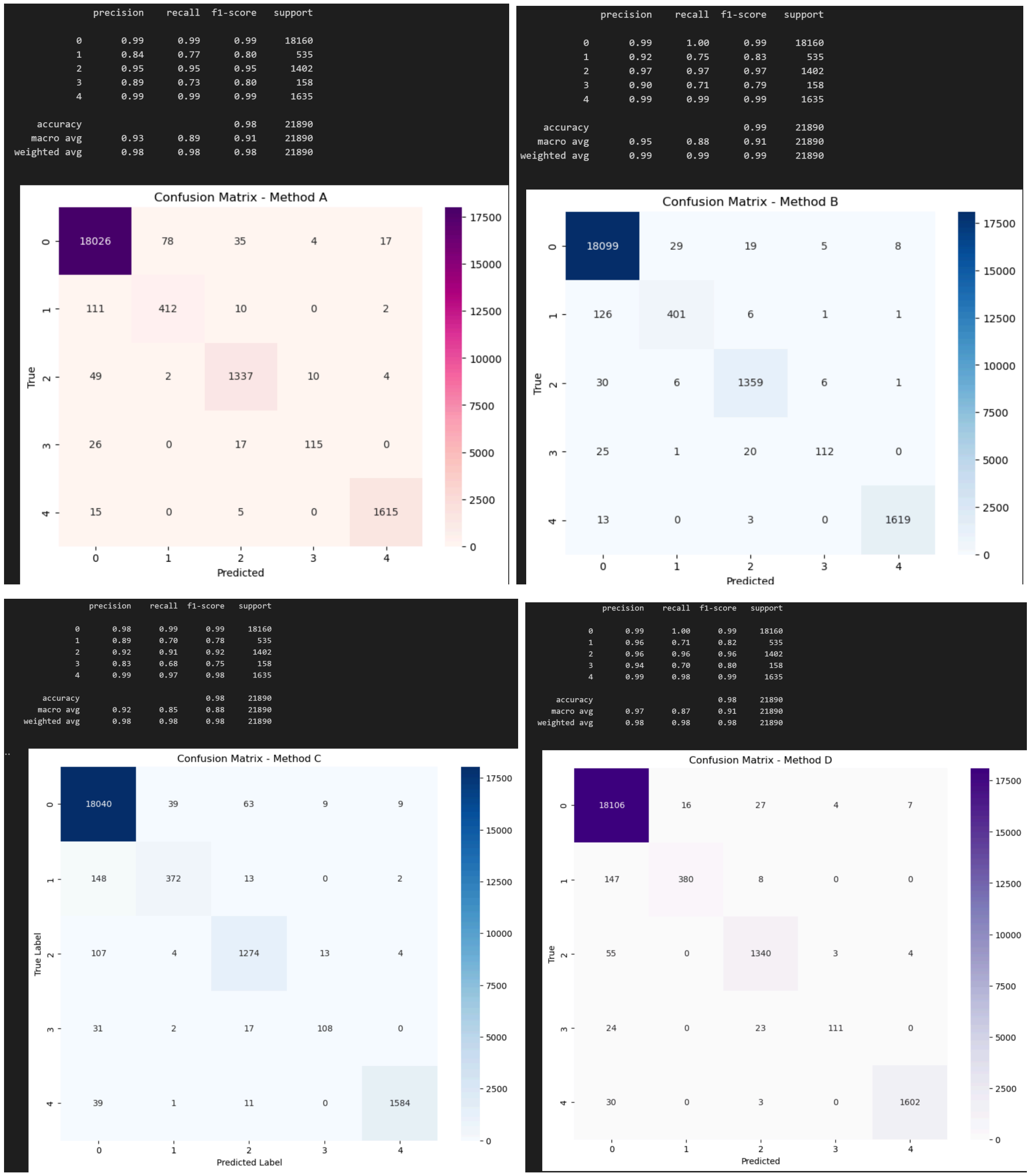

Fig 4. Results for Methods A, B, C, and D

All four methods produce viable results across the board. See Section 4.5 for more details.

4.5 – Analysis

While I would not say I was able to reproduce the results of the original papers, I was able to implement simplified architectures with the base process of both models. Due to the lack of public source code and the complexity of the original designs, exact replication was outside the scope of this project.

The implementations both received 98% accuracy which outperforms the original output of Method B. That said, the results did not stand out. When compared to Method C, the KNN baseline, both models scored higher across the board but by limited margins. I don't think this speaks to Method A/B being weak- they could be much stronger if identically implemented- but more so to KNN being a reliable general model. That said, all models performed similarly in class 0, 2, and 4. While all methods have a visible dip in scores in class 1 and 3, KNN stands out with its weaker scores. This could be due to KNN being a model susceptible to class imbalance, because Methods A, B, D are tailored for complex temporal dependencies because of their CNN-based architectures.

Model A performed as expected with strong interpretability due to the attention mechanism. It struggled with rare classes where localized attention may not capture global context. Method B also performed as expected, it was consistent across all classes because of deeper feature representation enabled by residual connections. With Method D, the ensemble model received the highest precision scores across classes, it improved overall performance meeting the hoped and expected outcome. It underperformed in recall in class 1 and 3. This means the model made fewer false positives at the cost of missing some true positives. This tradeoff makes sense because it balances generalization and sensitivity- in a literal sense, because the model averages the outputs. Overall, the experiment yielded encouraging results.

4.6 – Challenges & Setbacks

I faced multiple limitations when developing this project. Ideally I would have been able to replicate the architecture detailed in the papers of Method A and Method B but varying factors prevented a successful replication. I struggled a lot trying to implement them from scratch, citing various Github repositories using similar deep learning architectures to base my methods on. Ensemble stacking was omitted due to compute constraints and limited implementation guidance in Method A. And for Method B, the full depth of the residual stack was reduced for tractability.

Despite limitations, the core goals of implementing and comparing two deep learning models as well as an ensemble between those methods were met with satisfactory results.

5. Conclusions and Discussion

This capstone project thoroughly analyzes two state-of-the-art methods in ECG classifications by reviewing recent papers, selecting algorithms of interest, attempting to implement them and replicant results, and analyzing their performance.

5.1 – My Contributions

While I wasn't able to successfully replicate the proposed architecture in my selected methods, I wanted to further my analysis and understanding of deep learning by constructing an ensemble model to prove my hypothesis that the complementary strengths of the individual models ensemble could yield a more robust performance. The approach outperformed the other individual models in accuracy, precision, and class balance. I additionally adapted the models for the MIT-BIH dataset, expanding on the ECG5000 baseline data.

5.2 – Future Work

While reproducing the models in the papers did not go according to plan, the project and results inspire future tuning and updates to these experiments. The next step would be to complete a full architectural application, not just for the individual models, but to enhance the ensemble model. The ensemble model averaged the outputs of Method A and B but after reviewing the results and seeing the tradeoff and where the model underperformed, it inspires further research into tweaking the model, introducing weighted averages which could potentially address the recall drop observed in the results. With more computational resources, implementing a faithful reproduction could better individual performance which could improve ensemble performance.

This project demonstrates the strengths of different models for ECG-based classification. Although the full reproduction of the original research was constrained, the simplified implementations of Method A and Method B yield expected results, Method C demonstrated the baseline KNN model, and Method D validated the hypothesized idea that combining models with complementary strengths can improve results. The outcome of these experiments propose further research in improving the ensemble model and satisfy the in-depth understanding of state-of-the-art methods in ECG-based classification.

6. References

"ECG Heartbeat Categorization Dataset." *Www.kaggle.com*, www.kaggle.com/datasets/shayanfazeli/heartbeat?resource=download.

Jing, Enbiao, et al. "ECG Heartbeat Classification Based on an Improved ResNet-18 Model." *Computational and Mathematical Methods in Medicine*, vol. 2021, 30 Apr. 2021, pp. 1–13, https://doi.org/10.1155/2021/6649970.

Kiranyaz, Serkan, et al. "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks." *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, Mar. 2016, pp. 664–675, https://doi.org/10.1109/tbme.2015.2468589.

Murat, Fatma, et al. "Application of Deep Learning Techniques for Heartbeats Detection Using ECG Signals-Analysis and Review." *Computers in Biology and Medicine*, vol. 120, May 2020, p. 103726, https://doi.org/10.1016/j.compbiomed.2020.103726.

Nejedly, Petr, et al. "Classification of ECG Using Ensemble of Residual CNNs with Attention Mechanism." *2021 Computing in Cardiology (CinC)*, 13 Sept. 2021, pp. 1–4, ieeexplore.ieee.org/abstract/document/9662723/, https://doi.org/10.23919/cinc53138.2021.9662723. Accessed 27 Dec. 2024.

---. "Classification of ECG Using Ensemble of Residual CNNs with or without Attention Mechanism." *Physiological Measurement*, vol. 43, no. 4, 2022, pp. 10.1088/1361-6579/ac647c, pubmed.ncbi.nlm.nih.gov/35381586/, https://doi.org/10.1088/1361-6579/ac647c.

Saadatnejad, Saeed, et al. "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices." *IEEE Journal of Biomedical and Health Informatics*, 2019, pp. 1–1, https://doi.org/10.1109/jbhi.2019.2911367.

7. Sharing Agreement

Do you agree to share your work as an example for next semester? *Yes.*

Do you want to hide your name/team if you agree? *No.*