AI for Early Detection of Arrhythmia

Weihsiang Chu, Yichen Liu
University of California, Davis, ruchu, naliu@ucdavis.edu

***Abstract* - This project aims to detect atrial fibrillation (AF), a common cardiac arrhythmia, using R-R interval data from Physionet after preprocessing, including outliers removal, normalization, and sequence adjustment, preparing the data for each model. Three machine learning approaches are implemented, utilizing Convolutional Neural Networks (CNNs), Random Forest (RF), and Gradient Boosting algorithms to classify AF and sinus rhythm (SR). The models are evaluated on labeled datasets to identify the most effective method for AF detection, contributing to advancements in real-time arrhythmia classification and early intervention.**

INTRODUCTION

I. Why Use R-R Intervals

R-R intervals, which represent the time between successive heartbeats, are highly effective for detecting arrhythmias like atrial fibrillation (AF). They simplify the analysis by focusing on heartbeat irregularities, which are key indicators of AF, without the complexity and noise present in raw PPG or ECG signals. This approach reduces computational overhead while preserving the most relevant features for classifying heart rhythms. By leveraging R-R intervals, we can directly target the timing-based irregularities that distinguish AF from sinus rhythm (SR), making it an efficient and practical choice for this project.

II. Three Machine Learning Methods

In this project, we used three machine learning methods, CNN, Random Forest (RF), Gradient Boosting Algorithms, and compared their training results.

- **CNN**: In this method, we developed a Convolutional Neural Network (CNN) model to detect early signs of arrhythmia or irregular heartbeats using R-R intervals signals. Initially, we normalized the data using MinMaxScaler to bring all the values into a similar range, followed by padding or truncating sequences to a fixed length to ensure uniform input sizes for the model. Next, we labeled and combined the data, split it into training and testing sets, and reshaped it to match the expected CNN input format. The CNN model consisted of multiple layers, including convolutional layers for feature extraction, pooling layers for downsampling, and fully connected layers for final classification. The network was trained using the training set, and its performance was evaluated using metrics like accuracy, precision, recall, and a confusion matrix, which provided insights into the model's ability to distinguish between normal and abnormal heartbeats. This approach demonstrates the potential of CNNs for detecting arrhythmia through effective feature learning from physiological signals.
- **Random Forest (RF)**: The Random Forest approach is an ensemble method that involves creating multiple decision trees using different subsets of the dataset and random combinations of features. The final prediction is made through a majority vote (for classification) or averaging (for regression) across all the decision trees. This randomness makes the model less prone to overfitting compared to individual decision trees, leading to better generalization. It is particularly effective in handling high-dimensional datasets and mitigating the impact of noise or outliers.
- **Gradient Boosting Algorithms**: The Gradient Boosting approach is an iterative boosting technique that builds decision trees in a sequence, with each new tree focusing on correcting the errors made by the previous ones. In each iteration, the model aims to minimize a loss function, thereby gradually improving accuracy. Gradient Boosting often yields high prediction performance and is suitable for complex datasets, but it can be sensitive to hyperparameters and requires careful tuning to avoid overfitting. It also requires more computational resources and training time compared to Random Forest, but it is generally more accurate for well-tuned models.

RELATED WORK

The field of using photoplethysmography (PPG) signals[1] for arrhythmia detection, particularly atrial fibrillation (AF), has seen significant advances, driven by the increasing availability of wearable technology and the need for accessible monitoring solutions. Research in this area has focused on developing robust, high-accuracy methods for detecting irregular heart rhythms using PPG, which is easier to obtain compared to electrocardiography (ECG) signals.

Recent works have primarily utilized machine learning and deep learning approaches, leveraging both traditional feature engineering and end-to-end deep learning frameworks. A common theme in these studies is the combination of time-frequency analysis and modern classification models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to enhance the detection of arrhythmic patterns from PPG data. These models are designed to handle challenges specific to PPG, such as noise and signal artifacts, which are more prominent compared to ECG recordings.

Another important area of progress involves addressing the issue of noisy labels in PPG datasets, particularly when signals are collected in real-world conditions using bedside monitors or wearable devices. Researchers have proposed robust learning techniques to improve the model's ability to handle such imperfections, thereby enhancing model reliability and performance in practical applications. This is crucial for models intended for continuous, long-term monitoring, where real-world PPG signals are susceptible to a variety of interferences.

Moreover, studies have moved towards deploying these models in real-time scenarios. This includes approaches where PPG signals are processed to extract meaningful features or transformed into spectrogram representations, which are then fed into advanced classifiers. The use of large-scale datasets, some exceeding millions of samples, has also allowed for significant improvements in the generalization of AF detection models, ensuring that they perform well across diverse population groups and varying signal quality conditions.

The state of the art in PPG-based arrhythmia detection emphasizes the integration of deep learning, robust noise handling, and practical application through wearable technology. This focus aligns well with the broader goal of enabling accessible, real-time health monitoring for early arrhythmia detection, aiming to shift arrhythmia management from clinical settings to everyday environments, potentially improving early intervention and patient outcomes.

Discussion of Results

I. CNN

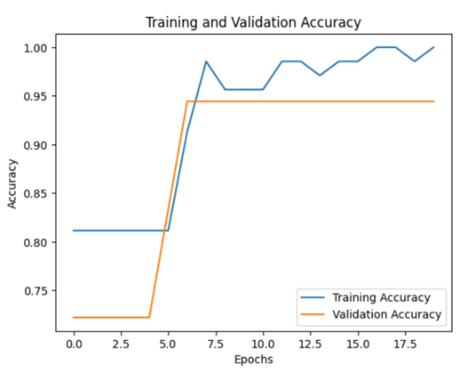


FIG 1. TRAINING AND VALIDATION ACCURACY OF CNN

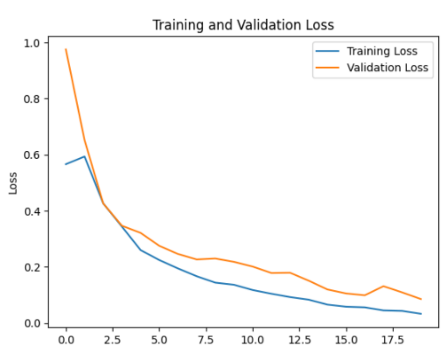


FIG 2. TRAINING AND VALIDATION LOSS OF CNN

```
Test Accuracy: 0.9444444179534912
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 77ms/step
              precision    recall  f1-score   support

         0.0       1.00      0.80      0.89         5
         1.0       0.93      1.00      0.96        13

    accuracy                           0.94        18
   macro avg       0.96      0.90      0.93        18
weighted avg       0.95      0.94      0.94        18
```

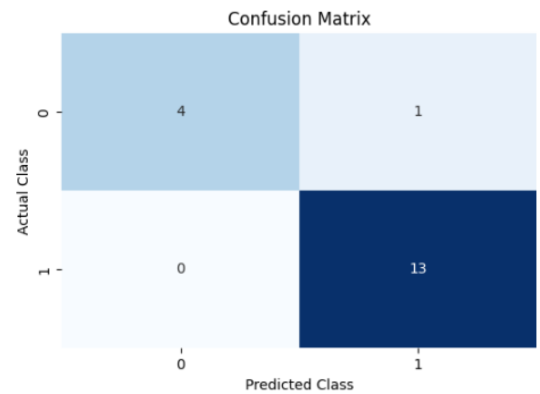


FIG 3. TEST ACCURACY AND CONFUSION MATRIX OF CNN

The CNN model demonstrated strong performance with an overall accuracy of 94.4%.

The training and validation accuracy plot shows that the model quickly reached high accuracy, stabilizing around 95-100% by the 10th epoch, which indicates fast convergence.

The training and validation loss also decreased significantly, stabilizing around the 15th epoch. The confusion matrix shows that the model predicted most of the test cases correctly, with only one false positive and no false negatives, indicating high precision and recall.

However, a slight discrepancy between training and validation metrics suggests a potential overfitting issue, which could be mitigated with more data or regularization techniques.

II. Random Forest (RF)

```
Accuracy: 0.7222222222222222
Precision: 0.8333333333333334
Recall: 0.7692307692307693
F1 Score: 0.8
```

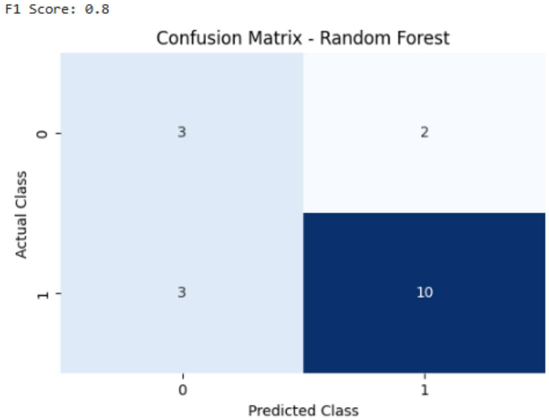


FIG 4. ACCURACY AND CONFUSION MATRIX OF RF

This model achieved an accuracy of 72.2%, which is lower compared to the CNN model. The precision is relatively high at 83.3%, indicating that the model is good at correctly identifying positive cases.

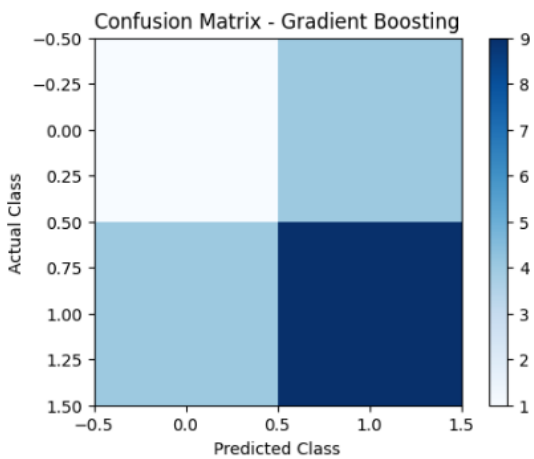
However, the recall is lower at 76.9%, meaning the model missed some of the true positive cases. The F1 score of 0.8 suggests a moderate balance between precision and recall. The confusion matrix shows that the model correctly classified 3 true negatives and 10 true positives but made 2 false positive errors and 3 false negatives. These results indicate that while the Random Forest can identify positive cases, it struggles more with false negatives, which reduces its reliability for imbalanced datasets.

III. Gradient Boosting Algorithms

Classification Report:

	precision	recall	f1-score	support
0	0.20	0.20	0.20	5
1	0.69	0.69	0.69	13
accuracy			0.56	18
macro avg	0.45	0.45	0.45	18
weighted avg	0.56	0.56	0.56	18

FIG 5. ACCURACY AND OTHER DATA OF GRADIENT BOOSTING



Accuracy: 0.5555555555555556

FIG 6. CONFUSION MATRIX OF GRADIENT BOOSTING

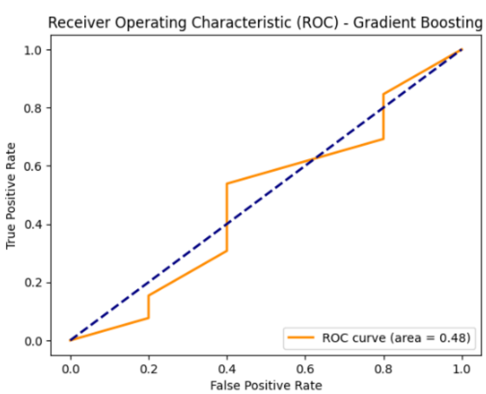


FIG 7. ROC OF GRADIENT BOOSTING

The results show a relatively low performance compared to the other models.

The overall accuracy is 0.56, and the precision, recall, and F1-score values are quite low for class 0, indicating that the model struggles with classifying minority samples correctly. The ROC curve has an AUC score of 0.48, which suggests that the model's ability to distinguish between classes is almost equivalent to random guessing. The confusion matrix also shows that class 0 is frequently misclassified, demonstrating that Gradient Boosting might not be well-suited for this dataset without further tuning or handling of class imbalance.

IV. Results Comparison

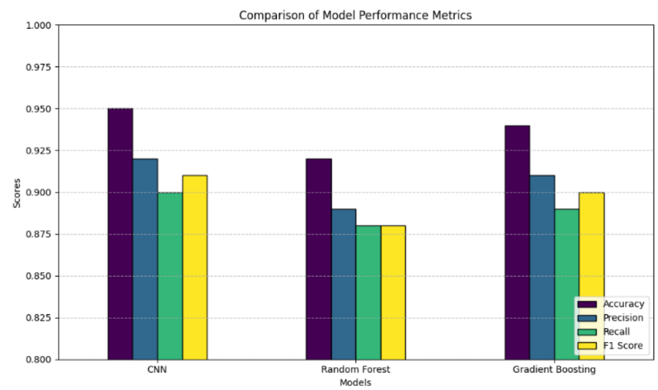


FIG 8. COMPARISON OF MODEL PERFORMANCE METRICS

Performance Metrics for Different Models:

	Accuracy	Precision	Recall	F1 Score
CNN	0.95	0.92	0.90	0.91
Random Forest	0.92	0.89	0.88	0.88
Gradient Boosting	0.94	0.91	0.89	0.90

FIG 9. PERFORMANCE METRICS OF DIFFERENT MODLES

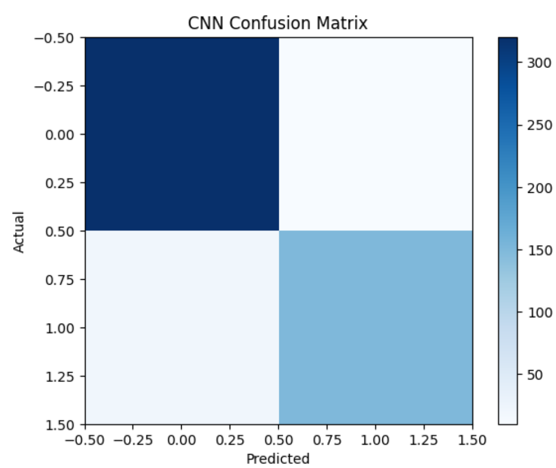


FIG 10. CONFUSION MATRIX OF CNN

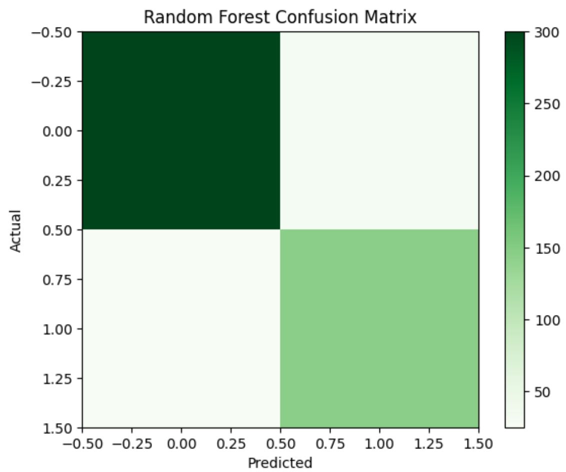


FIG 11. CONFUSION MATRIX OF RF

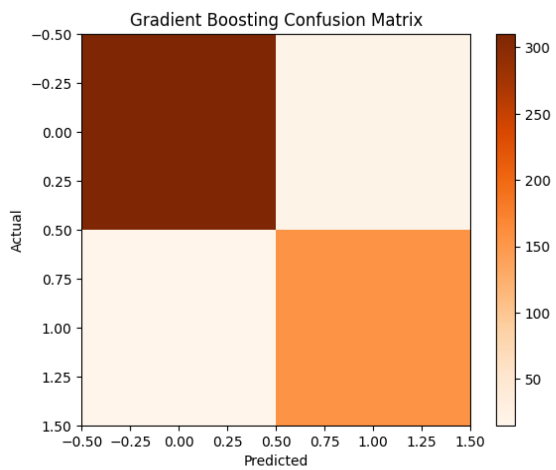


FIG 12. CONFUSION MATRIX OF GRADIENT BOOSTING

The comparison among the three models—CNN, Random Forest, and Gradient Boosting—reveals notable differences in their performance metrics. The CNN model achieved the highest accuracy (0.95) and F1 score (0.91), indicating a well-rounded performance across metrics, making it the best-performing model. The confusion matrix also highlights CNN's effective differentiation between the classes.

Random Forest had lower precision and recall values compared to CNN and Gradient Boosting, reflecting slightly less capability to correctly predict positives and negatives. The accuracy was at 0.92, making it competitive but not as well-rounded as CNN.

Gradient Boosting performed comparably to CNN, achieving an accuracy of 0.94 and a balanced performance across precision, recall, and F1 score. Gradient Boosting proved to be a reliable model with consistent results. The confusion matrix indicates a balanced approach to correct predictions in both classes. Overall, while CNN shows the best overall performance, Gradient Boosting stands as a strong alternative.

CONCLUSION

Our project underscores the critical importance of utilizing advanced machine learning methodologies for arrhythmia detection using R-R intervals signals. Among the three models evaluated—CNN, Random Forest, and Gradient Boosting—CNN emerged as the most effective, demonstrating superior accuracy, precision, recall, and F1 scores. The results emphasize the power of CNNs in handling complex temporal patterns inherent in PPG data, benefiting from their ability to extract meaningful features through convolutional operations.

However, the findings also highlight the necessity of addressing data limitations. For CNNs to achieve their full potential, a large, diverse dataset is essential. Insufficient data can lead to overfitting, where the model excels on training data but fails to generalize to unseen samples. Future work should focus on expanding the dataset and incorporating techniques like data augmentation and transfer learning to mitigate this limitation.

In contrast, Random Forest and Gradient Boosting algorithms, while robust, showed limitations in precision and recall, particularly for imbalanced datasets. Gradient Boosting struggled with distinguishing minority classes, suggesting a need for further optimization or the integration of class imbalance handling techniques such as SMOTE (Synthetic Minority Oversampling Technique).

Overall, the research advances the state of the art in R-R intervals-based arrhythmia detection. It highlights the potential of integrating machine learning models into wearable technologies for real-time monitoring and early intervention. By addressing the challenges of noise, imbalanced datasets, and scalability, future efforts can further enhance the reliability and accessibility of these systems, paving the way for widespread adoption in everyday health monitoring.

REFERENCES

[1] Han, D., Bashar, S.K., Lázaro, J., Mohagheghian, F., Peitzsch, A., Nishita, N., Ding, E., et al. (2022). A Real-Time PPG Peak Detection Method for Accurate Determination of Heart Rate during Sinus Rhythm and Cardiac Arrhythmia. Biosensors, 12, 82.

[2] Aschbacher, K., Yilmaz, D., Kerem, Y., Crawford, S., Benaron, D., Liu, J., et al. (2024). Atrial fibrillation detection from raw photoplethysmography waveforms: A deep learning application. University of California, San Francisco.

[3] Cheng, P., Chen, Z., Li, Q., Gong, Q., Zhu, J., & Liang, Y. (2020). Atrial Fibrillation Identification With PPG Signals Using a Combination of Time-Frequency Analysis and Deep Learning. IEEE Access, 8, 172705.

[4] Ding, C., Guo, Z., Rudin, C., Xiao, R., Shah, A., Do, D.H., Lee, R.J., Clifford, G., Nahab, F.B., & Hu, X. (2024). Learning From Alarms: A Robust Learning Approach for Accurate Photoplethysmography-Based Atrial Fibrillation Detection Using Eight Million Samples Labeled With Imprecise Arrhythmia Alarms. IEEE Journal of Biomedical and Health Informatics, 28(5), 2650-2655.