dHealth 2025
M. Baumgartner et al. (Eds.)
© 2025 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI250201

A Neural Network for Atrial Fibrillation Detection via PPG

Rishad HOWLADER[a,1], Monika JURGEC[a], Andrea SCHIMANKO[a], Christopher DOHNAL[a], Anne BUSCH[a], Martin BAUMGARTNER[b], Fabian WIESMÜLLER[b], Aaron LAUSCHENSKY[b], Dieter HAYN[b] and Günter SCHREIER[b]
[a] *University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria*
[b] *Austrian Institute of Technology, Vienna, Austria*

Abstract. Atrial fibrillation (AF) is a prevalent cardiac arrhythmia associated with severe complications such as ischemic stroke and heart failure. Early detection is essential for timely intervention; however, traditional diagnostic methods often lack scalability and accessibility. This project explores the use of photoplethysmography (PPG) signals recorded via smartphone applications to develop a predictive neural network model for AF detection. Data collection involved student participants using a smartphone app, supplemented with curated open-source datasets to enhance generalizability. A multilayer perceptron (MLP) was designed and trained using TensorFlow, with model performance evaluated based on accuracy, sensitivity, specificity, and F1-score. The initial model achieved an accuracy of approximately 75%, indicating the potential of PPG-based AF detection. Optimization efforts continue, focusing on improving predictive accuracy and minimizing false positives. Future research will refine the model architecture, validate it on larger datasets, and integrate findings into scalable digital health initiatives, such as the Austrian Digital Heart Program, supported by the Austrian Institute of Technology (AIT).

Keywords. Atrial Fibrillation, Photoplethysmography, Artificial Intelligence, Neural Networks, Mobile Health

1. Introduction

1.1. Atrial Fibrillation and Its Challenges

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, affecting approximately one-third of individuals aged 65 and older [1], [2]. It significantly contributes to morbidity and mortality, being associated with complications such as ischemic stroke, heart failure, and cardiovascular death [1], [3]. Early detection is critical, as anticoagulant therapy can reduce stroke risk by approximately 70% [4]. However, AF often remains undiagnosed due to its asymptomatic or paroxysmal nature, necessitating innovative diagnostic approaches beyond traditional methods [1].

The European Society of Cardiology (ESC) recommends opportunistic or systematic screening for individuals over 65 to mitigate AF-related risks [2]. Addressing this need requires solutions that are accessible, scalable, and cost-effective. Conventional diagnostic tools such as 12-lead electrocardiograms (ECG) and Holter monitoring are effective but often fail to capture transient AF episodes, especially in asymptomatic

[1] Corresponding Author: Rishad Howlader, Master Health Care Informatics, University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria, E-Mail: rishadarnab@outlook.com

patients [2], [6]. These methods are constrained by high costs, logistical challenges, and reliance on clinical infrastructure, limiting their suitability for large-scale population screening [2], [7].

1.2. Photoplethysmography and Artificial Intelligence

Smartphone-based photoplethysmography (PPG) has demonstrated high sensitivity and specificity in detecting AF, as shown in studies like the eBRAVE-AF trial [1], [5]. PPG provides a scalable, non-invasive diagnostic tool capable of detecting paroxysmal or asymptomatic AF episodes, making it particularly promising for systematic screening initiatives. Machine learning (ML) models, particularly multilayer perceptrons (MLPs), further enhance AF detection by identifying subtle patterns in PPG signals, enabling real-time, accurate, and automated AF classification without dependence on clinical infrastructure [6], [8].

1.3. Study Objectives

This study aims to develop an AI-based model for AF detection using PPG data collected via smartphones. The following key objectives were chosen to ensure a scientifically sound and clinically applicable approach:

Development and training of a neural network model to accurately detect AF: Given the complexity of AF detection, a robust neural network model is required to identify irregularities in PPG signals. A multilayer perceptron (MLP) was selected due to its ability to capture non-linear patterns in time-series data.

Cleaning and analysis of raw and open-source data collected during the measurement period: Data preprocessing is crucial for improving model performance, as raw PPG signals are often affected by motion artifacts, noise, and variability among participants. Proper data cleaning ensures that only high-quality, reliable inputs are used for model training.

Evaluation of the model's performance using accuracy, precision, recall, and F1-score: Instead of relying solely on accuracy, additional metrics such as precision, recall, and F1-score provide a comprehensive assessment of the model's reliability in distinguishing AF from non-AF cases. These benchmarks align with clinical standards for diagnostic model evaluation.

Summarizing findings to provide a foundation for nationwide AF screening programs supported by AIT: The ultimate goal is to assess whether AI-enhanced PPG analysis can contribute to nationwide AF screening efforts. While the Austrian Digital Heart Program, supported by AIT, aims to integrate AI into large-scale diagnostics, this study serves as a feasibility analysis to validate its practical application in non-clinical environments.

2. Methods

2.1. Study Design and Participant Recruitment

To develop an AI-based AF detection model using PPG data, a structured study was designed in collaboration with the Austrian Institute of Technology (AIT). The study

was conducted within the framework of the Austrian Digital Heart Program (ADHP), aiming to integrate machine learning techniques into large-scale digital health initiatives.

The data collection phase involved 16 student participants from the University of Applied Sciences Wiener Neustadt (FHWN). Over the course of four weeks (October 15 – November 15, 2024), participants recorded PPG signals using a smartphone application developed by AIT. This application utilized the built-in camera and flash of a smartphone to measure blood volume changes in peripheral circulation, which could then be analyzed for AF detection.

Before initiating data collection, participants were required to install the app on their personal Android or iOS devices. Upon first use, the app guided them through an initial calibration phase, ensuring that the correct finger placement was used to obtain high-quality PPG signals. Each measurement session lasted 60 seconds, during which the app recorded fluctuations in blood volume by analyzing changes in light absorption.

To complement the PPG data, participants provided demographic and health-related information, including age, gender, cardiovascular history, and lifestyle factors (e.g., smoking habits, exercise frequency). This information was crucial for later stages of analysis, ensuring a more comprehensive dataset for model training and validation.

A summary of the demographic characteristics of the study participants is provided in Table 1.

Table 1. Summary of the demographic characteristics of the study participants

Characteristic	Value
Age (mean ± SD)	23.4 ± 2.1 years
Gender (M/F)	9/7
Total Measurements	320
Average Duration per Measurement	60s

2.2. Data Processing

The raw data collected from the smartphone app contained inconsistencies due to motion artifacts, poor lighting conditions, and variations in device hardware. To ensure high-quality inputs for model training, a multi-step preprocessing pipeline was implemented.

Step 1: Data Cleaning. Each recorded PPG signal was examined for completeness and consistency. Measurements containing excessive noise, missing values, or abnormally low signal amplitudes were removed. Duplicate entries were also identified and excluded to prevent biases in model training.

Step 2: Data Augmentation. Since the majority of collected data represented normal sinus rhythm (non-AF cases), an external open-source dataset was integrated to balance the proportion of AF and non-AF samples. The external dataset consisted of validated recordings from AF-positive and AF-negative patients, improving the generalizability of the final model.

Step 3: Dataset Splitting. The final dataset, consisting of both internal and external data, was divided into training (70%) and testing (30%) subsets. This split ensured that the model was trained on a large proportion of data while maintaining an independent evaluation set to assess real-world performance.

2.3. Model Development and Training

To analyze PPG signals and classify AF cases, an MLP was implemented using TensorFlow. The model was designed to extract hidden patterns within the PPG signals, allowing for accurate AF detection.

Neural Network Architecture: The architecture of the MLP consisted of multiple layers, each playing a specific role in transforming the input data into meaningful predictions, listed in the following and visually represented in Figure 1:

- Input Layer: 128 features extracted from PPG signals.
- Hidden Layer 1: 64 neurons with ReLU activation to introduce non-linearity.
- Hidden Layer 2: 32 neurons with ReLU activation for deeper feature extraction.
- Hidden Layer 3: 16 neurons with ReLU activation for further refinement.
- Output Layer: A single neuron with Sigmoid activation, responsible for binary classification (AF vs. non-AF).

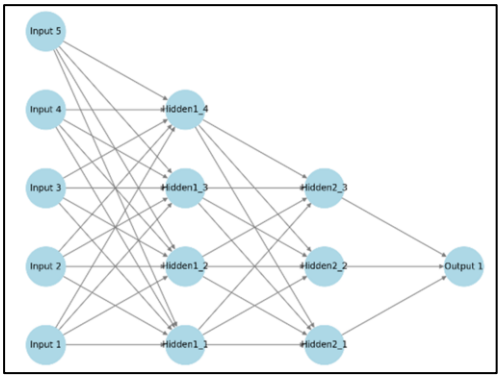

Figure 1. A visual representation of the used neural network structure

For training the network, binary cross-entropy was chosen as the loss function, given its suitability for binary classification tasks. The Adam optimizer was selected with a learning rate of 0.001, balancing convergence speed with model stability. Training occurred over 100 epochs with a batch size of 32, ensuring adequate exposure to data for meaningful learning.

Real-time validation was conducted during training to monitor the model's performance on unseen data, enabling early detection of overfitting. Early stopping techniques were applied whenever validation loss began to increase, ensuring the retention of the best-performing model.

Further optimization of model architecture, particularly targeting improvements in precision and recall, is planned. Future research may explore advanced architectures, including convolutional neural networks (CNNs) or recurrent neural networks (RNNs), to further enhance predictive performance and reliability in AF detection [9].

3. Results

The effectiveness of the developed neural network model for detecting AF was assessed using multiple clinically relevant performance metrics. While accuracy is a common measure for overall model correctness, the metrics precision, recall (sensitivity), and the F1-score provide deeper insight into the model's diagnostic reliability and its suitability for real-world AF screening applications. These metrics were used, the achieved values are listed in Table 2:

- **Accuracy**: The percentage of correctly classified cases (AF and non-AF).
- **Precision**: The proportion of predicted AF cases that were correctly identified.
- **Recall (Sensitivity)**: The proportion of actual AF cases that were successfully detected.
- **F1-Score**: A balanced measure that considers both precision and recall.

Table 2. List of performance metrics as achieved by the neural network in classifying the PPG signals

Metric	Value
Accuracy	75%
Precision	72%
Recall	68%
F1-Score	70%

A confusion matrix (see Figure 2) provides a detailed breakdown of the model's classification performance. This matrix visualizes the number of true positives (correctly identified AF cases), false positives (normal cases misclassified as AF), false negatives (AF cases missed by the model), and true negatives (correctly identified non-AF cases).

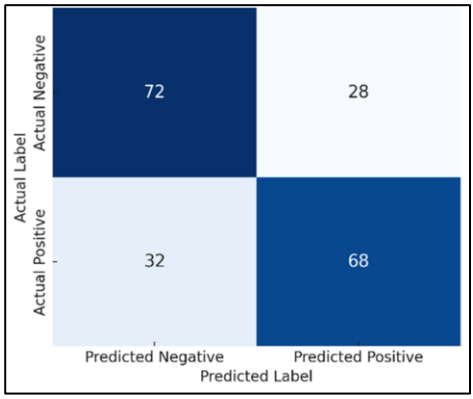

Figure 2. Visualization of classification results as a confusion matrix

4. Discussion

This study demonstrates the feasibility of employing smartphone-based PPG signals and an AI-driven MLP for AF detection. The developed model achieved an accuracy of 75%, precision of 72%, recall of 68%, and an F1-score of 70%, confirming its potential as a scalable, accessible, and cost-effective screening tool. These advantages position PPG-

based screening as a viable alternative to traditional ECG-based diagnostics, especially within large-scale initiatives such as the Austrian Digital Heart Program (ADHP).

However, despite these promising results, there remains considerable potential for optimization. Specifically, improving recall (to reduce false negatives) and precision (to minimize false positives) is critical for clinical implementation. Missed AF cases could lead to delayed therapeutic interventions, while false positives may cause unnecessary healthcare interventions and patient anxiety.

Several factors currently limit model performance. The relatively small and homogeneous dataset, primarily consisting of young, healthy student participants, restricts generalizability. Moreover, inconsistent signal quality arising from variations in smartphone hardware, lighting conditions, and motion artifacts during measurements emphasizes the need for rigorous preprocessing and data normalization procedures.

Future work will address these limitations by incorporating larger and more diverse datasets, including clinical patient populations, to enhance the robustness and real-world applicability of the developed AI algorithms. Further optimization strategies include detailed hyperparameter tuning, experimenting with advanced architectures (such as CNNs or RNNs), and employing advanced preprocessing methods (e.g., synthetic data generation or noise-filtering algorithms). Additionally, subsequent studies will prioritize clinical validation, regulatory approval processes, and seamless integration into national digital health frameworks, particularly within programs like the ADHP. While results demonstrate promising potential for AI-driven AF detection, further optimization and validation are crucial to transitioning this technology into a scalable, clinically viable screening tool.

References

[1] F. Brandt, M. Kreiser, and A. Bauer, “Smartphonebasiertes Screening auf Vorhofflimmern: Erkenntnisse der eBRAVE-AF-Studie,” Monitor Versorgungsforschung, vol. 16, no. 3, pp. 49–53, 2023.

[2] R. Schleberger et al., “Update Vorhofflimmern: Die ESC-Leitlinien 2020 sowie aktuelle Daten zur frühen antiarrhythmischen Therapie,” Herzschrittmachertherapie & Elektrophysiologie, vol. 32, pp. 257–263, 2021.

[3] Y. Guo et al., “Photoplethysmography-Based Machine Learning Approaches for Atrial Fibrillation Prediction,” Journal of the American College of Cardiology: Asia, vol. 1, no. 3, pp. 399–408, 2021.

[4] A. Bauer et al., “Detection of atrial fibrillation with a smartphone camera: Results from the DETECT AF PRO trial,” Europace, vol. 21, no. 1, pp. 41–47, 2019.

[5] Austrian Institute of Technology (AIT), “Nationwide Screening Program for Atrial Fibrillation,” internal documentation, 2024.

[6] L. Rizas et al., “Smartphone-based screening for atrial fibrillation: Results from a pragmatic randomized trial,” Nature Medicine, vol. 28, no. 9, pp. 1823–1830, 2022.

[7] M. Freyer et al., “Rationale and Design of the eBRAVE-AF Trial,” American Heart Journal, vol. 241, pp. 26–34, 2021.

[8] Preventicus GmbH, “Validation of smartphone-based PPG technology for AF detection,” internal study report, 2023.