POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Cardiac Arrhythmias' Classification using Photoplethysmographic Signals

AI in Biomedicine

Biomedical Engineering - Ingegneria Biomedica

Giovanni Veronelli, Pietro Monti, Alessandro Perricone
10579308, 10686153, 10661417
Academic year: 2023-2024

1. Introduction

Cardiac arrhythmias, particularly premature atrial contractions (PACs) and premature ventricular contractions (PVCs), are common irregular heartbeats that can serve as early indicators of more serious cardiac conditions. PACs are premature heartbeats originating from the atria, while PVCs arise from the ventricles. Although often benign, frequent PACs and PVCs can lead to more serious arrhythmias, such as atrial fibrillation or ventricular tachycardia, making their detection and monitoring critical [1, 2].

Traditionally, electrocardiography (ECG) has been the gold standard for detecting PACs and PVCs due to its ability to capture detailed electrical activity of the heart [3]. However, photoplethysmography (PPG) advent as a non-invasive, cost-effective alternative has opened new possibilities for continuous cardiac monitoring in various settings, thanks to the widespread distribution of wearable devices.

The goal of this study is to explore the efficacy of PPG in detecting PACs and PVCs, utilizing both traditional machine learning models and advanced deep learning techniques. The study first addresses a binary classification problem to distinguish between normal and abnormal beats. Following this, a three-class classification is conducted to differentiate among normal beats, PACs, and PVCs, aiming to enhance the precision of arrhythmia detection.

2. Materials and Methods

2.1. Dataset Overview and Characteristics

The dataset employed for this task comprises recordings from 105 subjects, each associated with a PPG signal, annotated beat peak positions, and corresponding labels. In particular, the labels are denoted as "N", "S", and "V", representing, respectively, normal beats, premature atrial contractions (PAC), and premature ventricular contractions (PVC).

The PPG recordings exhibited non-uniform sampling frequencies, with 62 recordings sampled at 128 Hz and 43 recordings at 256 Hz. All recordings were subject to high-frequency noise, likely due to motion artifacts from the subjects. Moreover, the dataset presented a significant class imbalance, with 92.8% of the beats labeled as "N", 3.9% as "S", and 3.3% as "V".

2.2. Data Preprocessing

In fig. 1, the processing pipeline of the signals is reported.

2.2.1 Exclusion of Subjects with Exclusively "N" Beats

To mitigate class imbalance and reduce redundancy, subjects exhibiting only "N" type beats were excluded from the dataset. This resulted in the removal of 14 signals, all of which were recorded at a sampling rate of 250 Hz.

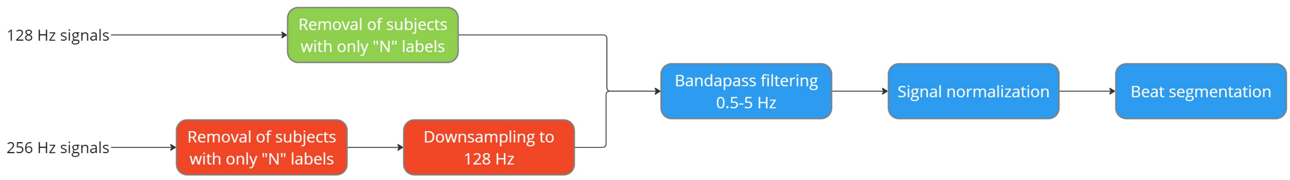

Figure 1: Data Preprocessing Pipeline.

2.2.2 Downsampling of 250 Hz signals

To ensure consistency in the analysis, it was necessary to standardize the sampling frequencies of the signals. The decision was made to downsample the 250 Hz signals to 128 Hz, based on several considerations:

- The majority of signals in the dataset were sampled at 128 Hz.
- A lower sampling rate reduces the volume of data, resulting in faster processing times and lower computational costs.
- Reduced data size contributes to more efficient storage utilization.
- A sampling frequency of 128 Hz is sufficient to capture all essential physiological information.

Before downsampling, the Fast Fourier Transform (FFT) of each 250 Hz signal was analyzed to confirm the absence of relevant frequency components above 64 Hz, in accordance with the Nyquist-Shannon sampling theorem.

2.2.3 Signal Filtering

Bandpass Butterworth filtering was then applied to all the signals. The analysis of the FFT of the signals previously performed, also helped to determine the cut-off frequencies of the filter. In order to remove the DC component of the signal, a high pass cut-off of 0.5 Hz was chosen. Instead, considering the low frequency content of the signals, a 5 Hz low pass cut-off was chosen.

2.2.4 Signal Normalization

To standardize the PPG signals across different subjects, a Z-score normalization was applied. This method adjusts each signal to have a mean of zero and a standard deviation of one, ensuring that the data is centered and scaled uniformly.

2.2.5 Beat Segmentation

In order to analyze individual cardiac cycles within the PPG signals, a beat segmentation process was implemented. This approach involves dynamically extracting individual beats from the signal, centered around the identified systolic peaks. Given the variability in the distance between consecutive peaks, a dynamic windowing technique was employed. This method adapts the extraction window for each beat, based on the inter-peak intervals, ensuring that the relevant portions of the signal are captured effectively. For each identified peak, a segment of the signal is extracted by defining a window that spans a specified ratio before and after the peak location. This approach ensures that the complete waveform of the beat is captured, including both the ascending and descending phases of the pulse. By adjusting the window size dynamically according to the local inter-peak intervals, the method accounts for natural variations in heart rate and beat morphology across different subjects.

After completing these preprocessing steps, a total of 91 signals remained, all sampled at 128 Hz and filtered. The new label distribution was as follows: "N": 91.8%, "S": 4.5%, and "V": 3.7%. Moreover, a total of 215.350 beats were extracted.

2.3. Feature Extraction and Dataset Creation

The dataset used for classification was constructed by extracting various features from each beat. The selection of these features was informed by previous studies [4] [5] and a deep understanding of the characteristics of the PPG signal. The features were categorized into four primary types: temporal, morphological, statistical, and frequency-based.

- **Temporal Features**: These features describe the timing characteristics of the beat, such as peak amplitude, rise time, fall time, beat duration, and peak-to-peak interval.
- **Morphological Features**: These features capture the shape and structure of the beat, including maximum and minimum amplitude, mean and standard deviation of the amplitude, range, root mean square (RMS), area under the curve, full width at half maximum (FWHM), energy, systolic-to-diastolic ratio, and rise-to-fall ratio.
- **Statistical Features**: Statistical properties of the beat, such as skewness, kurtosis, and entropy, are included to quantify the distribution and complexity of the signal.
- **Frequency Features**: Frequency domain features were derived from the power spectral density (PSD) of the beat, such as total power, band power (within the 0.5-5 Hz range), dominant frequency, and specific FFT components. Additionally, the power ratio between the band power and the remaining power spectrum was computed.

In the end, for each beat, a total of 28 features were computed.

2.4. Outlier Removal

To enhance the robustness of the classification model, outliers within the dataset were identified and removed. The Z-score method was employed to detect outliers across multiple key features, including peak amplitude, various FFT components (fft1, fft2, fft3, fft4, fft5), and entropy.
For each feature, the Z-score was calculated to determine how many standard deviations a data point was from the mean. A threshold of 3 was set, meaning that any data point with a Z-score greater than 3 in any of the selected features was considered an outlier. This threshold effectively captures extreme values that may skew the analysis and impact the performance of the model.
After filtering out these outliers, a refined dataset was obtained, which excluded the anomalous beats. The distribution of the beat labels was analyzed both before and after the outlier removal to ensure that the balance of the dataset was not significantly disrupted.

- **Label distribution before outlier removal:**
  - **N:** 197,665 beats
  - **S:** 9,691 beats
  - **V:** 7,994 beats
- **Label distribution after outlier removal:**
  - **N:** 185,476 beats
  - **S:** 9,297 beats
  - **V:** 7,615 beats

Furthermore, after studying the correlation between features, several features were found to be highly correlated with others, which could introduce redundancy and reduce the effectiveness of the model. As a result, the following features were removed from the final dataset: *beat duration*, *band power*, *rms*, *range amplitude*, and *std amplitude*.

2.5. Binary Classification

The objective of the binary classification task was to distinguish between normal (N) and abnormal beats, where abnormal beats include both supraventricular (S) and ventricular (V) categories. To achieve this, two distinct approaches were employed: traditional machine learning models and a custom-designed feedforward neural network (FFNN).
The dataset was structured to facilitate the binary classification process. Initially, the features were separated from the labels, and the data was grouped by subject to maintain consistency in the class distribution across sets. This grouping allowed for a stratified split, ensuring that the proportion of normal and abnormal beats was maintained in each subset of the data.
For the machine learning models, the dataset was split into training and test sets, focusing on preserving the balance between the two classes. Meanwhile, the FFNN model required a more refined split, where the training data was further divided into training and validation subsets.

2.5.1 Traditional Machine Learning Models

The traditional machine learning models were trained using the extracted features, with the dataset split into training and test sets to evaluate model performance on unseen data. To address the issue of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training set, ensuring a balanced class distribution prior to training.
Several models were tested to determine the best approach. In particular, KNN, Decision Tree, Naive Bayes, Logistic Regression, and XGBoost were tested. For each model, hyperparameter tuning was conducted using Bayesian optimization.

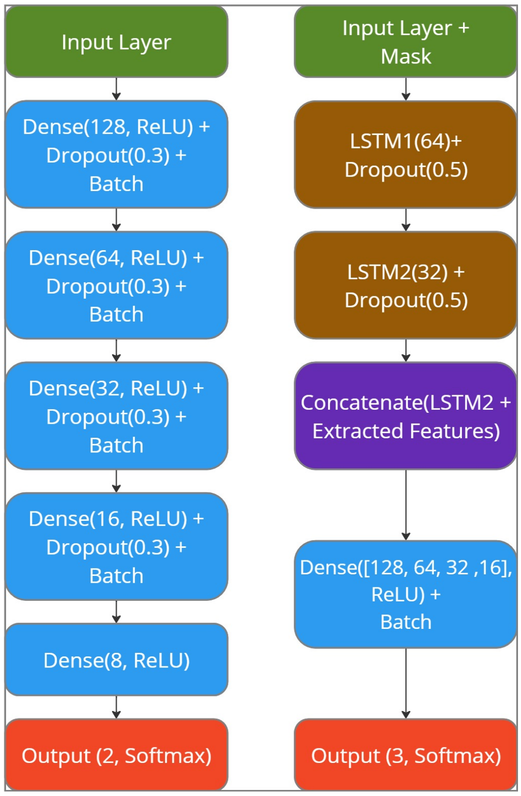

Figure 2: On the left, FFNN architecture for binary classification. On the right, custom deep learning model for three-class classification

2.5.2 Custom Feedforward Neural Network

A custom FFNN was designed and trained using the same feature set. The data was split into training, validation, and test sets, allowing for iterative tuning of the model based on the validation loss performance. To address the class imbalance, the class_weight parameter was used during training, which helped in effectively learning from the imbalanced data.
The architecture included multiple hidden layers, each employing ReLU activation functions. To enhance generalization and prevent overfitting, dropout and batch normalization were applied after most layers.

The network architecture (fig. 2) consisted of five hidden layers with gradually decreasing neuron counts (from 128 to 8 neurons), followed by a softmax output layer for binary classification. This design allowed the model to capture complex patterns in the data while maintaining robustness against overfitting, ensuring reliable performance across both training and validation sets.

2.6. Three-Class Classification

The approach for three-class classification was similar to the binary classification, involving the same data preprocessing steps and the use of both traditional machine learning models and a custom deep learning model. The goal was to classify the beats into three categories: normal (N), supraventricular (S), and ventricular (V).

For the machine learning approach, the same models (K-Nearest Neighbors, Decision Tree, Naive Bayes, Logistic Regression, and XGBoost) were employed, with hyperparameter tuning performed using Bayesian optimization. SMOTE was again used to balance the training data before fitting the models.

2.6.1 Custom Deep Learning Model

For the deep learning approach, a more sophisticated model was developed to handle the three-class classification task. This model combined an LSTM network, designed to capture temporal dependencies in the sequential beat data, with a feedforward neural network that processed the engineered features.

The architecture was as follows:

- LSTM Branch: The LSTM branch was responsible for extracting features from the sequential beat data. Two LSTM layers were used, with 64 and 32 units respectively, each followed by a dropout layer to prevent overfitting. Zero padding was applied to ensure all beats had the same length, and a masking layer was used to ignore the padded values during training.
- Engineered Features Input: In parallel with the LSTM branch, a separate input layer processed the pre-engineered features extracted from the beats.
- Concatenation and Dense Layers: The outputs of the LSTM branch and the engineered features input were concatenated and passed through several dense layers with ReLU activation functions and batch normalization. This structure allowed the model to integrate both sequential information and hand-crafted features effectively.
- Output Layer: A softmax activation function in the output layer produced the probability distribution across the three classes, facilitating the classification of each beat as normal, supraventricular, or ventricular.

This hybrid model, leveraging both LSTM and traditional feature-based approaches, was designed to capture the complex nature of the PPG signals. By integrating temporal information from the LSTM with the engineered features, the model aimed to improve classification accuracy across the three classes.

2.7. Custom Metric

Given the diagnostic context in which these models will be applied, a custom metric was developed to prioritize the accurate classification of abnormal beats, which are critical for identifying potential arrhythmias. However, the metric also considers the importance of not misclassifying normal beats. This balanced approach ensures that the model not only detects abnormalities effectively but also maintains a high level of precision in identifying normal beats, making it reliable for clinical use.

The metric is implemented as follows, with slight changes between the binary and three class classification problems:

- **Precision for Normal Beats:** The function first calculates the precision for class 0. A low precision would indicate a high number of false negatives, meaning that many abnormal beats are incorrectly classified as normal.
- **Recall for Abnormal Beats:** Then, it calculates the recall for class 1 in binary classification, and for both class 1 and class 2 in the three-class scenario. In this context, recall measures the model's ability to correctly identify abnormal beats.
- **Threshold and Penalty Mechanism:** If the precision for normal beats falls below a specified threshold, the recall for abnormal beats is penalized. This penalty is set to the precision of class 0, but can be customized.
- **Output:** The function returns the recall for abnormal beats, adjusted by the penalty if necessary. In the three-class classification problem, the metric returns the average recall between class 1 and class 2.

2.8. Confidence Assessment

2.8.1 Machine Learning Models

For all the machine learning models tested, confidence in predictions was assessed using the *predict_proba* function. This function provides the predicted probabilities for each class. By examining these probabilities, we determined the level of confidence each model had in its prediction.

2.8.2 Custom Deep Learning Models

Instead, for the deep learning models, the confidence was derived directly from the output of the *softmax* layer. The *Softmax* activation function converts the raw model outputs into probabilities, reflecting the model's confidence across the possible classes.

3. Results

3.1. Binary Classification

For binary classification, both the machine learning models and the custom Feedforward Neural Network (FFNN), proved to be effective approaches for the task. However, the best performance was achieved by the custom FFNN, as shown in tab. 1.

Table 1: Binary Classification Results

Model	Precision	Recall	F1 score	Custom
FFNN	0.93	**0.93**	**0.93**	**0.88**
KNN	0.88	0.83	0.85	0.68
Dec Tree	0.87	0.82	0.84	0.67
Naive Bayes	0.77	0.86	0.80	0.81
Log Reg	0.90	0.86	0.88	0.73
XGBoost	**0.95**	0.80	0.85	0.60

The FFNN also demonstrated high confidence levels, as seen in tab. 2.

Table 2: FFNN Binary Classification Confidence

Class	Mean	Std	Min	Max
Class 0	0.97	0.09	0.50	1.00
Class 1	0.94	0.12	0.50	1.00

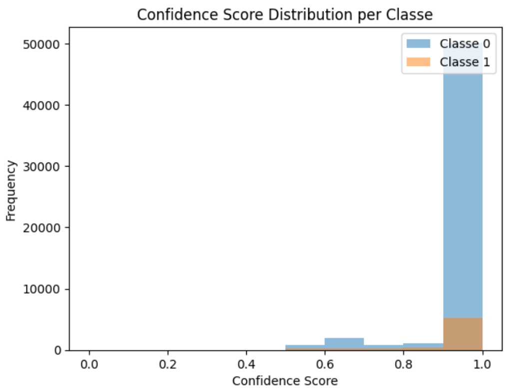

Figure 3: Confidence Score Distribution of FFNN

3.2. Three Class Classification

In the three-class classification task, several challenges were encountered, which will be thoroughly discussed in the following sections. The performance of all models was highly dependent on the ability to adjust model weights effectively. Despite these challenges, similar to the binary classification task, the best performance was ultimately achieved by the custom deep learning model, as shown in tab. 3.

Table 3: Three Class Classification Results

Model	Precision	Recall	F1 score	Custom
FFNN	**0.70**	**0.74**	**0.70**	**0.63**
KNN	0.61	0.64	0.62	0.47
Dec Tree	0.51	0.66	0.56	0.55
Naive Bayes	0.47	0.60	0.49	0.45
Log Reg	0.56	0.67	0.59	0.55
XGBoost	0.69	0.63	0.64	0.44

The deep learning model also demonstrated higher overall confidence levels, as shown in tab. ??. However, it exhibited lower confidence for the abnormal classes compared to the binary classification problem, reflecting the increased complexity of distinguishing between three classes.

Table 4: Custom Classification Confidence

Class	Mean Conf	Std Conf	Min Conf	Max Conf
N	0.97	0.09	0.35	1.00
S	0.64	0.14	0.34	0.98
V	0.68	0.11	0.35	1.00

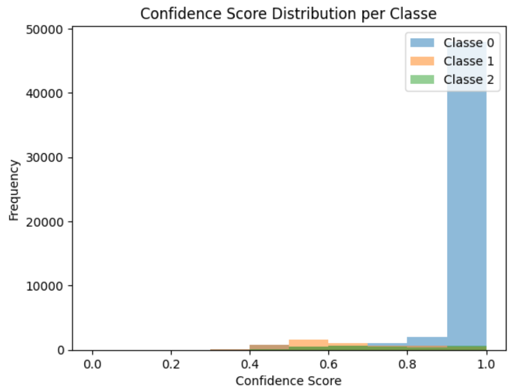

Figure 4: Confidence Score Distribution of custom model

4. Discussion

In the binary classification task, no particular issues were encountered during testing. The custom FFNN consistently outperformed traditional machine learning models, demonstrating superior precision, recall, and F1 scores. Additionally, the confidence assessment showed that the FFNN provided high levels of certainty in its predictions. The custom metric implemented, which penalizes the model when precision for normal beats drops below a certain threshold, also helped ensure a balance between detecting abnormal beats and avoiding false positives.
The three-class classification task, however, presented more challenges. Although the custom deep learning model once again delivered the best overall performance, the results underscored the complexities involved in distinguishing between normal, supraventric-

ular, and ventricular beats. In particular, the accurate classification of supraventricular beats (class S) was notably difficult. After multiple attempts to address this issue, the most effective solution involved manually adjusting the class weights. In the deep learning model, the weight for class S was manually set to 39.0, significantly higher than the original value computed by the *class_weight* tool. This manual adjustment proved crucial for improving the model's ability to correctly classify supraventricular beats. Similarly, better results were also achieved with the decision tree, logistic regression, and XGBoost classifiers by manually adjusting the weight for this class.

In conclusion, the custom Feedforward Neural Network (FFNN) demonstrated high efficiency in addressing the binary classification problem, although the machine learning models also proved to be effective. For the three-class classification problem, while baseline results were achieved, there remains significant room for improvement, particularly in effectively distinguishing between supraventricular and ventricular beats. Future enhancements could focus on exploring the non-linear characteristics of the signals or investigating time-frequency representations, such as wavelet transforms, which may reveal additional distinguishing features. Moreover, further hyperparameters search and the use of more complex machine learning models, might enhance the obtained results.

References

[1] Han KD. Choi JI. et al. Kim, Y.G. Premature ventricular contraction is associated with increased risk of atrial fibrillation: a nationwide population-based study. *Sci Rep*, 11, 2021.

[2] Jose Gomez-Arroyo MD PhD Daniel G. Jovin BS Narayan G. Kowlgi MD Karoly Kaszala MD PhD Alex Y. Tan MD Jayanthi N. Koneru MD Jordana Kron MD Kenneth A. Ellenbogen MD Jose F. Huizar MD Sampath Gunda MD, Daniel Akyeampong BS. Consequences of chronic frequent premature atrial contractions: Association with cardiac arrhythmias and cardiac structural changes. *The Official Journal of the World Society of Arrhythmias*, 30, 2019.

[3] Jianyuan Hong, Hua-Jung Li, Chung chi Yang, Chih-Lu Han, and Jui chien Hsieh. A clinical study on atrial fibrillation, premature ventricular contraction, and premature atrial contraction screening based on an ecg deep learning model. *Applied Soft Computing*, 126:109213, 2022.

[4] N.A.; Lopez D.M. Millán, C.A.; Girón. Analysis of relevant features from photoplethysmographic signals for atrial fibrillation classification. *Int. J. Environ. Res. Public Health*, 17, 2020.

[5] Elyas Sabeti, Narathip Reamaroon, Michael Mathis, Jonathan Gryak, Michael Sjoding, and Kayvan Najarian. Signal quality measure for pulsatile physiological signals using morphological features: Applications in reliability measure for pulse oximetry. *Informatics in Medicine Unlocked*, 16:100222, 2019.