POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Applied AI in Biomedicine

Beat Classifier for PPG Signals

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Authors: Edoardo Eula, Frederik Heinzel, Michele Masserini

Professor: Valentina Corino

Teaching assistants: Guadalupe García Isla, Francesca Lo Iacono

Academic year: 2023-2024

---

1. Introduction

Photoplethysmogram (PPG) signals generally offer a non-invasive and cost-effective method for the detection of blood volume changes in the microvascular bed of tissue in clinical and consumer devices [1]. Due to a growing use of PPG sensors in consumer wearables, the field of use cases for these signals is growing rapidly in recent years. The setting doesn't come without challenges though, as motion artifacts and other noise factors make signal interpretation complicated. A key step forward to provide important information about health in daily life is realised by the development of robust algorithms for PPG signal processing.

Premature Ventricular Complex (PVC) and Premature Atrial Complex (PAC) both represent benign cardiac arrhythmias and refer to early hearbeats from either the ventricles or the atria, respectively. Both can contribute to a false positive detection of atrial fibrillation (AF), an arrhythmia originating in the upper chambers of the heart. Methods have been proposed to detect AF using a wristband device [3] and other research has been conducted to detect PVC and PAC through PPG signals using a smartwatch [5].

The main task of this group project was to create a beat classifier for given PPG signals. The input of the beat classifier consisted of the PPG signal and the systolic peak position. The output consisted of two tasks:

- First, a detection and classification of **normal (N) and abnormal beats (AN)**.
- In the second task, a detection, classification and distinction should be made between the three beat types: **Normal (N), Supra-ventricular (S) and Ventricular (V) beats**.

The final classifier models are model ensemble structures, combining models trained in the time domain as well as the frequency domain, which in turn utilize complex model architectures with the integration of past labels, convolutional layers, residual blocks, identity blocks, squeeze and excitation modules and attention mechanisms. All pre-processing, data handling and model construction as well as training were performed using Python and Google Colab notebooks.

2. Materials and Methods

2.1. Data Exploration

Data from all patients were extracted and loaded. The data loading process involved it-

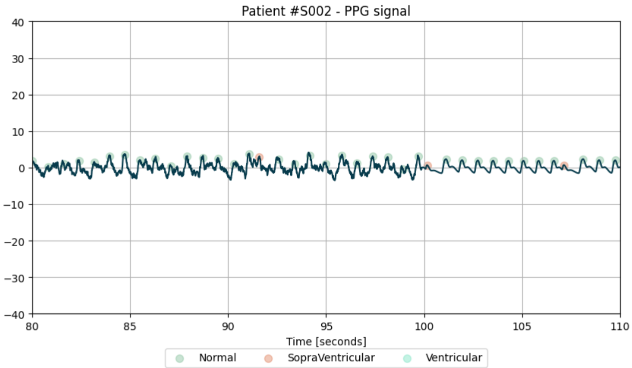

Figure 1: 30 seconds of the signal extracted from Patient S002 together with the peaks

erating through the provided files, where each set of three files corresponds to one patient, consisting of the PPG signal, annotation labels, and speech information. If successfully loaded, a patient instance is created, encapsulating relevant information such as the patient's ID, file path, sampling frequency, signal samples, PPG signal, speech annotations, and peak labels. During a first inspection, it was apparent that some signals were sampled at 128 Hz, others at 250 Hz. To ensure uniformity for subsequent analyses, all signals were downsampled to 128 Hz. The downsampling process involved interpolating the signal to match the desired frequency. Specifically, the patient class, implemented in the provided code, plays a crucial role in handling these adjustments, together with others used later in the pre-processing. After loading, a visual check was performed. In Figure 1 an example signal, together with the labels of the peaks is plotted. During the loading phase all the 105 patients were correctly loaded.

2.2. Data Pre-processing

The pre-processing pipeline used consisted on the following steps:

- Removal of patients with an high amplitude variation
- Removal of patients with just normal beats
- Butterworth bandpass filtering
- Data normalization

The initial step involved evaluating amplitude variations for each patient, aiming to identify and address motion artifacts' impact on photoplethysmogram (PPG) signals. Figure 2 displays a bar plot showcasing the ascending order of amplitude variations across all patients. Patients with elevated standard deviation values in their PPG signals, indicative of motion artifacts, were excluded from further analysis using a set threshold of 10.

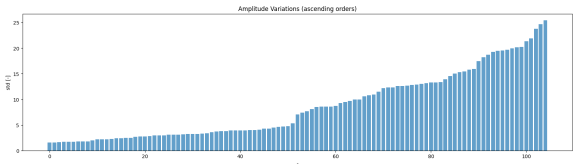

Figure 2: Amplitude Variations Barplot.

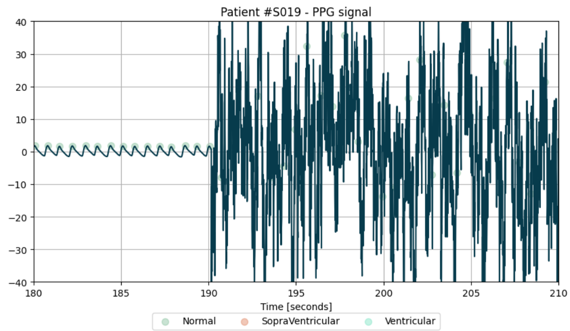

Figure 3: Patient S019's PPG signal affected by motion artifacts

To implement this exclusion, amplitude variations were computed for each patient, and those surpassing the threshold were subsequently removed from the dataset. This pre-processing step aimed to improve overall signal quality and diminish the influence of motion artifacts on subsequent analyses. Mitigating motion artifacts without accelerometer or comparable reference signals poses a significant challenge. Such artifacts can profoundly impact signal fidelity in physiological signals like PPG. Consequently, rather than employing filtration techniques, this project opted to exclude patients with notably high standard deviations in their PPG signals. This decision aligns with the understanding that motion artifacts embedded within the signal are complex to filter out effectively without additional reference signals. The elimination of such instances was chosen to preserve the dataset's integrity and reliability [7]. An example of a segment affected by motion artifacts in Patient S019's PPG signal is depicted in Figure 3.
In the second step, patients exhibiting solely normal beats were excluded from the analysis as their data lacked relevance for our study. The primary objective was to develop a beat classifier capable of identifying abnormal beats; con-

sequently, patients without any abnormal beats were deemed non-contributory to the specific goals of our investigation.
In the third and fourth steps of our pre-processing, we employed Butterworth bandpass filters with a frequency band of 0.5 Hz to 3 Hz and a filter order of 3. This choice ensures effective noise reduction while preserving relevant physiological information. The use of a third order filter strikes a balance between noise attenuation and signal fidelity.
Following the filtering, we applied normalization using the z-score technique. This step standardizes the signal amplitudes, ensuring a nearly zero mean and a standard deviation of one. Such normalization is essential for maintaining consistency and comparability across diverse signals with varying scales.
After completing all the pre-processing steps, the dataset was refined to include 59 patients, encompassing a total of 141,623 beats. The distribution of beats across different categories is illustrated in Table 1. The majority of beats were identified as normal, constituting 91.6% of the total beats, followed by supra-ventricular beats (4.8%) and ventricular beats (3.6%).

Class Counts

	Beats	Percentage
Normal	129,807	91.6%
Supra-ventricular	6,736	4.8%
Ventricular	5,080	3.6%

Table 1: Distribution of beats across different categories.

To facilitate the development of a robust beat classifier, our strategy involved segmenting the signals into windows, each encapsulating a single beat. This segmentation aimed to enable the classifier to focus on the morphology of individual beats. The windowing technique entailed extracting a beat and padding it to achieve a fixed length, ensuring uniformity across all windows. To determine the optimal fixed length, we examined the distribution of inter-beat distances, as depicted in Figure 4.

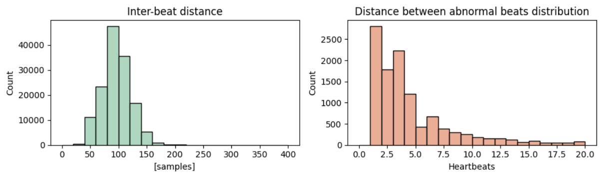

Figure 4: Left: Inter-beat distance distribution. Right: distance between abnormal beats distribution

Notably, the histogram revealed that the maximum length required for effective windowing was 200 samples.

2.3. Models

The primary objective of this project involved the development of two distinct models:

- The first model was designed to classify normal and abnormal beats.
- The second model aimed to classify normal, ventricular, and supra-ventricular beats.

Data preparation for both tasks followed a unified approach. The selected model architecture consisted of an ensemble model comprising two parallel models. The first model utilized time domain signals, while the second leveraged frequency domain signals, as suggested also in literature [4]. Additionally, both models incorporated a "history" aspect, by considering the labels of the past heartbeats.

2.3.1 Integration of Past Labels

Upon scrutinizing the data, a notable observation from the histogram in Figure 4 indicated that premature atrial contractions (PACs) and premature ventricular contractions (PVCs) often exhibited recurring patterns, as can be found in literature [2] [6]. Consequently, to enhance the precision of beat classification, we opted to include the label of the previous beat in our model, recognizing its potential to contribute to more accurate predictions overall for the current peak.

2.3.2 First Task: Detect Between Normal and Abnormal Beats

For the first task of detecting normal and abnormal beats, the abnormal beat classification encompassed both "S"-labeled and "V"-labeled beats. To address the class imbalance, we implemented under-sampling for the normal class

("N"). Before model training, we conducted a split of the dataset into training, validation, and test sets.

2.3.3 Second Task: Detect Among Normal, Supra-Ventricular, and Ventricular Beats

For the second task, our objective was to develop a model capable of distinguishing between normal ("N"), supra-ventricular ("S"), and ventricular ("V") beats. As seen in the Table 1, the dataset presented an inherent class imbalance, with the ventricular beats being the minority class. To address this imbalance and ensure robust model training, we implemented under-sampling for the normal and supra-ventricular classes, bringing their proportions closer to that of the ventricular class. This under-sampling strategy involved randomly selecting a subset of normal and supra-ventricular beats, aligning their count with the number of ventricular beats. This approach aimed to prevent the model from being biased towards the majority class, thereby promoting a more balanced learning process. Also here we applied a splitting of the dataset into training, validation and test sets.

2.3.4 Model Architectures

Time Model Architecture: The final time model (Figure 5 in the Appendix) consists of Long Short-Term Memory (LSTM), convolutional and batch normalization layers, as well as maxpooling and dropout layers. It starts with an initial LSTM layer with 256 units followed by two times the following layers: a convolutional 1D layers with 128 filters, a kernel size of 5 and a batch normalization layer. After that a maxpooling1d layer and dropout of 0.5 lead over to another LSTM with 256 units. In the last steps, the output of the LSTM goes through another dropout of 0.5 until it gets flattened and results in a dense layer with 512 units, which then after another dropout (0.5) finally reaches another dense layer and then concatenates with the historical beat labels through the past input layer until resulting in a dense output layer for classification with softmax activation.

Frequency Model Architecture: The final frequency model (Figure 6 part A until Figure 10 part E in the Appendix) combines convolutional layer, batch normalization layers, LSTMs, blocks in a ResNet-style with squeeze and excitation modules and attention blocks. It starts with an initial convolutional 1D layer with a filter of size 64 and a kernel size of 3, followed by batch normalization and a subsequent dropout layer of 0.5. Then the model increases the complexity and uses two residual blocks in a ResNet-style approach, which include convolutional layers and identity blocks, each enhanced with squeeze and excitation (SE) module combined with an attention module. In a next step, a cascade of four additional ResNet blocks are created using convolutional layers and identity blocks with increasing filter sizes. Lastly, the model includes two LSTMs (128 units each) and a dropout (0.5) before incorporating historical beat labels via an additional input layer and concatenation before resulting in the output layer with a softmax activation.

These architectures aimed to capture distinctive features and temporal dependencies within the beats, contributing to an accurate classification of normal and abnormal beats.

2.3.5 Ensemble Model

After both models were trained an ensemble model was built that combines predictions from two models with two mechanisms:

- **Argmax:** In this approach, the predictions from both models are combined by summing them element-wise. The final predictions are then obtained by selecting the class with the highest cumulative value. This mechanism essentially favors the model that assigns a higher confidence score to a particular class.
- **Averaging:** Here, the predictions from both models are averaged element-wise before determining the final class predictions. By taking the mean of the individual model predictions, this mechanism aims to create a more balanced and averaged decision, reducing the impact of extreme confidence values from either model.

After implementing both mechanisms, we observed that the results were relatively similar. However, we consistently chose the Argmax mechanism as it consistently yielded higher results, indicating its effectiveness in leveraging the strengths of each individual model to produce a more confident and accurate ensemble prediction.

3. Results

3.1. First Task Results: Normal vs Abnormal Beats

The results for Task 1 have been calculated using model precision, recall, and the F1-score of the 3 classes separately. The test set used for these evaluation consists of 30% of the available data set.

	Precision	Recall	F1-score
Normal	0.86	0.86	0.86
Abnormal	0.86	0.86	0.86
Per-Class Acc.	0.86	0.86	0.86
Weighted Acc.	0.86	0.86	0.86
Overall Acc.			**0.8598**

Table 2: Precision, Recall, and F1-score metrics for the argmax ensemble argmax combining the time model and frequency model predictions, as discussed in Section 2.3.2 for Task 1.

3.2. Second Task Results: Normal vs Supra-Ventricular vs Ventricular Beats

Similarly, in Task 2, the evaluation is extended to the three distinct categories, leveraging the same precision, recall, and F1-score metrics (see Table 3).

	Precision	Recall	F1-score
Normal	0.96	0.92	0.94
Sopraventr.	0.90	0.87	0.88
Ventricular	0.84	0.90	0.87
Per-Class Acc.	0.90	0.90	0.90
Weighted Acc.	0.90	0.90	0.90
Overall Acc.			**0.8958**

Table 3: Precision, Recall, and F1-score metrics for the argmax ensemble argmax combining the time model and frequency model predictions, as discussed in Section 2.3.2 for Task 2.

For graphical clarity, the confusion matrices of both time and frequency models for Task 1 (Figure 11 and Figure 12), and the time model, frequency model, and ensemble for Task 2 (Figure 13), are included in the Appendix.

4. Discussion

The evaluation results presented in Table 2 and Table 3 demonstrate the effectiveness of the ensemble model in classifying normal and abnormal beats, and distinguishing between Normal, Supraventricular, and Ventricular beats in the provided PPG data. Several key aspects contribute to the model's performance and merit detailed discussion.

4.1. Balancing Strategy and Training Efficiency

The balanced performance across the "Normal" and "Abnormal" classes in Task 1, and among "Normal," "Supraventricular," and "Ventricular" classes in Task 2, is a direct consequence of the implemented balancing strategy during the training phase. Under-sampling the majority class ("N") not only achieves class balance but also reduces training time. This approach mitigates the risk of overfitting, which ensures a robust model capable of generalizing well to unseen data, a critical aspect in both tasks.

4.2. Window Splitting Method and Signal Representation

The efficacy of the window splitting technique remains crucial for signal representation, significantly influencing overall model performance in both Task 1 and Task 2. By enforcing a fixed-length window and incorporating padding, the method mitigates the occurrence of multiple peaks within a single window. This is particularly essential, as the presence of multiple peaks in a short window often indicates irregular heart rhythms (arrhythmias) or conditions characterized by an abnormally rapid heart rate (tachycardia). The success of this strategy underscores the model's proficiency in efficiently capturing pertinent information from PPG signals across diverse beat categories.

4.3. Inclusion of Past Labels for Temporal Context

The incorporation of past labels into the model inputs, as elaborated in Section 2.3.1, is motivated by the recurring patterns observed in premature atrial contractions (PACs) and premature ventricular contractions (PVCs). This historical context significantly contributes to beat classification precision in both Task 1 and Task 2. Notably, the choice to utilize the two most recent labels aims to effectively manage the recall of the past and leverage information from the immediate temporal context. This consideration has proven suitable for accurate predictions in both time and frequency models.

4.4. Ensemble Model and Complementary Strengths

In Task 1, the time model outperforms the frequency model in both classes with correctly classified 86% "Normal" and 85% "Abnormal" labels (freq: 82% "N" and 84% "A", see Figure 11 and Figure 12 in the annex). In Task 2, both models excel in identifying "Supraventricular" beats with 86% correctly labeled beats for both models. Due to the ensemble framework the models complement each other as the ensemble model result is 90% correct classifications for this class, exceeding the individual model performances by a far. The time model also demonstrates superior performance in recognizing "Normal" beats with 96% (freq: 94%) correct labels, and the frequency model excels in identifying "Ventricular" beats with 83% (time: 81%) correct labels on the predictions, while the ensemble model achieved 84% correct classifications of "Ventricular" (see Figure 13). Summarized on can note that through the strategic combination in the ensemble, the respective strengths of these models are leveraged most effectively and the complementary strengths of both models are combined, resulting in enhanced overall classification accuracy across diverse beat categories.

4.5. Better model performance in Task 2

As noticeable in Table 2 and Table 3, the accuracy in the binary classification with 86% was lower than in the classification with three classes 90%. This was a somewhat surprising result, as one would intuitively expect the first task to be easier, especially as the "Abnormal" class consisted of exactly the "Supraventricular" and "Ventricular" class. The same amount of fine-tuning and effort on model architectures were spent on Task 1, with the same architectures working best in the end. A possible explanation seems to be that mixing both classes together make the assembled class actually harder to detect and predict than each class respectively. This could be due to the inherent complexity, with the underlying patterns and variations in the data being such that the task of differentiating between three distinct classes rather than merging two of them is inherently easier. Additionally, the information loss obtained by collapsing two classes into one reduces granularity and drops some of maybe the distinct information or features extractable per class and obscures the learnable common features of the new assembled class.

An approach to circumvent this issue would be to train the ensemble model on all three classes, such that it learns all class-inherent patterns and variations. Afterwards it would be predicting in a normal way on the test set but then at the end one had to go over the output again and remap a copy of the labels list (every "V" and "S" to "Abnormal").

5. Conclusion

In conclusion, our comprehensive approach, encompassing meticulous data pre-processing, the development of an advanced model in the time domain and another leveraging the FFT in the frequency domain, along with the strategic integration of an ensemble model, has resulted in robust beat classifiers for PPG signals. Achieving an overall accuracy of 86% in the first task and 90% in the second task highlights the effectiveness of our models in distinguishing normal and abnormal beats, as well as classifying among normal, supra-ventricular, and ventricular beats. Notably, the inclusion of past peak labels, providing a memory of the past, as well as the model ensemble of time and frequency domain contributes significantly to the models' precision, emphasizing their potential for practical applications in healthcare and wearable devices.

References

[1] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28:R1, 2 2007.

[2] B.-H. Chong, V. Pong, K.-F. Lam, S. Liu, M.-L. Zuo, Y.-F. Lau, C.-P. Lau, H.-F. Tse, and C.-W. Siu. Frequent premature atrial complexes predict new occurrence of atrial fibrillation and adverse cardiovascular events. *Europace*, 14:942–947, 7 2012.

[3] Valentina D.A. Corino, Rita Laureanti, Lorenzo Ferranti, Giorgio Scarpini, Federico Lombardi, and Luca T. Mainardi. Detection of atrial fibrillation episodes using a wristband device. *Physiological Measurement*, 38:787, 4 2017.

[4] Filipa Esgalhado, Beatriz Fernandes, Valentina Vassilenko, Arnaldo Batista, and Sara Russo. The application of deep learning algorithms for ppg signal processing and classification. *Computers*, 10(12):158, 2021.

[5] Dong Han, Syed Khairul Bashar, Fahimeh Mohagheghian, Eric Ding, Cody Whitcomb, David D. McManus, and Ki H. Chon. Premature atrial and ventricular contraction detection using photoplethysmographic data from a smartwatch. *Sensors*, 20:5683, 10 2020.

[6] Bao-tao Huang, Fang-yang Huang, Yong Peng, Yan-biao Liao, Fei Chen, Tian-li Xia, Xiao-bo Pu, and Mao Chen. Relation of premature atrial complexes with stroke and death: Systematic review and meta-analysis. *Clinical Cardiology*, 40:962–969, 11 2017.

[7] David Pollreisz and Nima TaheriNejad. Detection and removal of motion artifacts in ppg signals. *Mobile Networks and Applications*, 27:728–738, 4 2022.

Appendix A

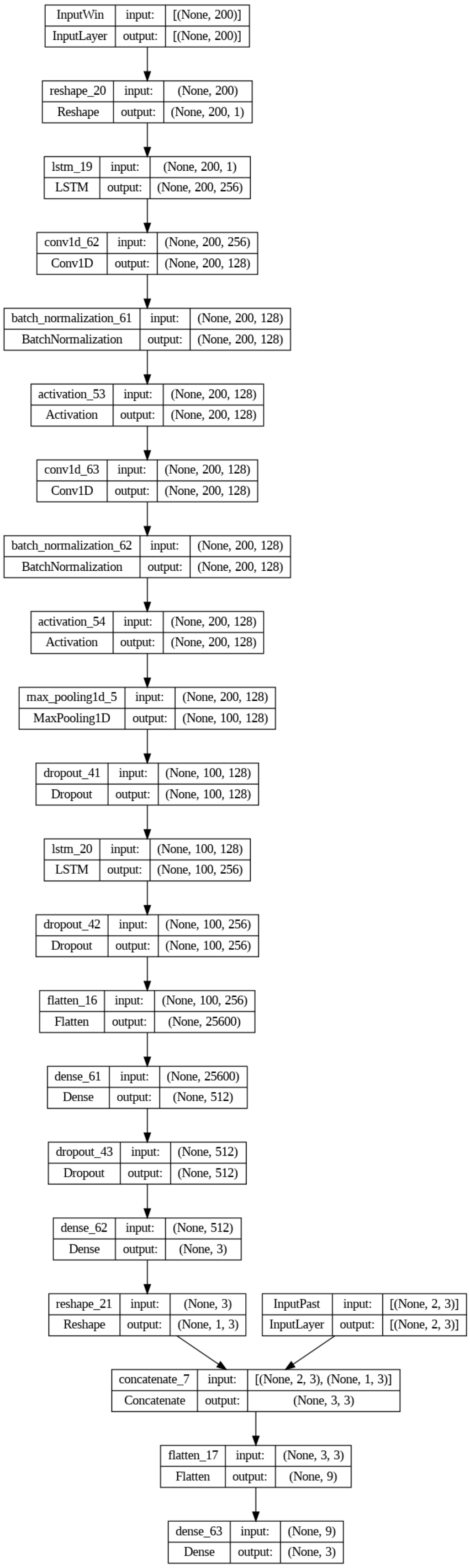

Figure 5: Time model architecture

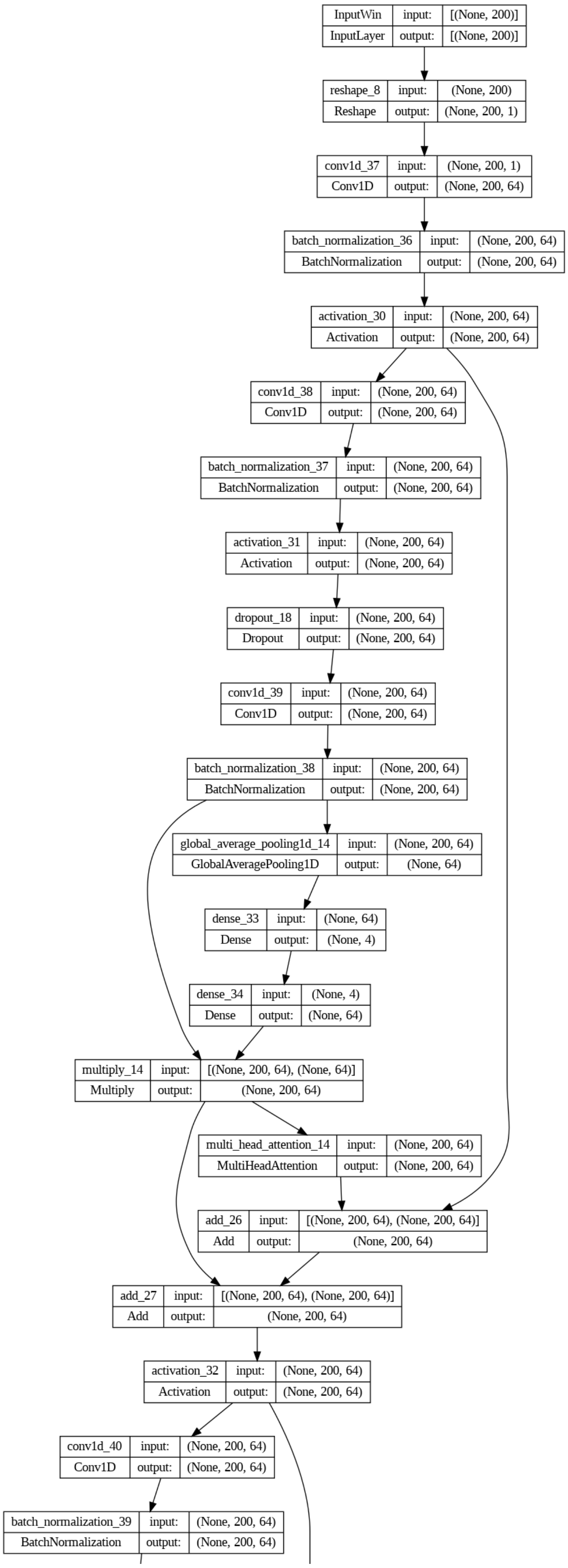

Figure 6: Frequency model architecture (part A)

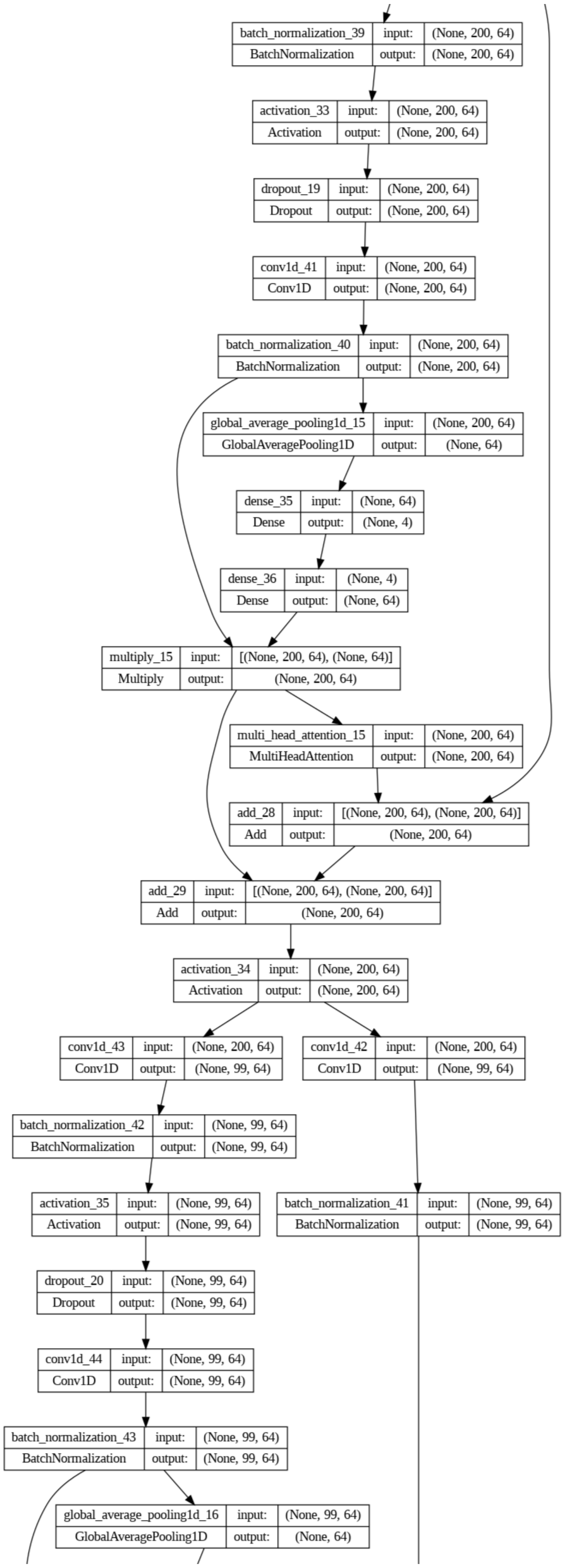

Figure 7: Frequency model architecture (part B)

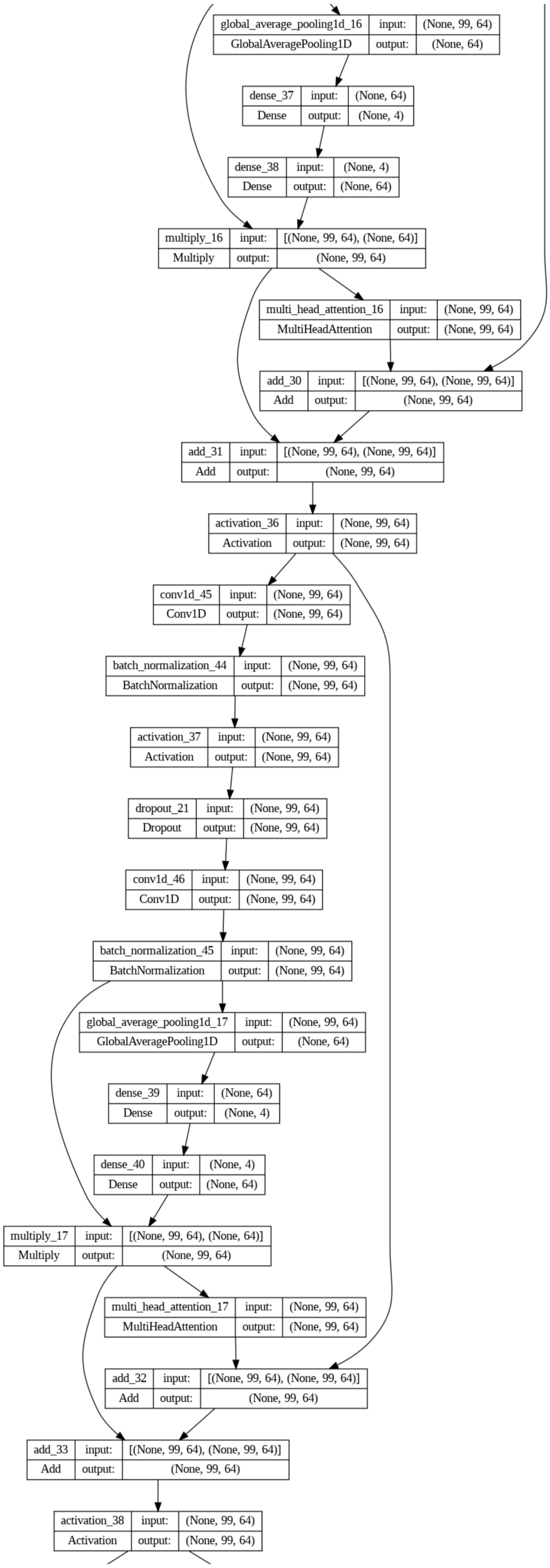

Figure 8: Frequency model architecture (part C)

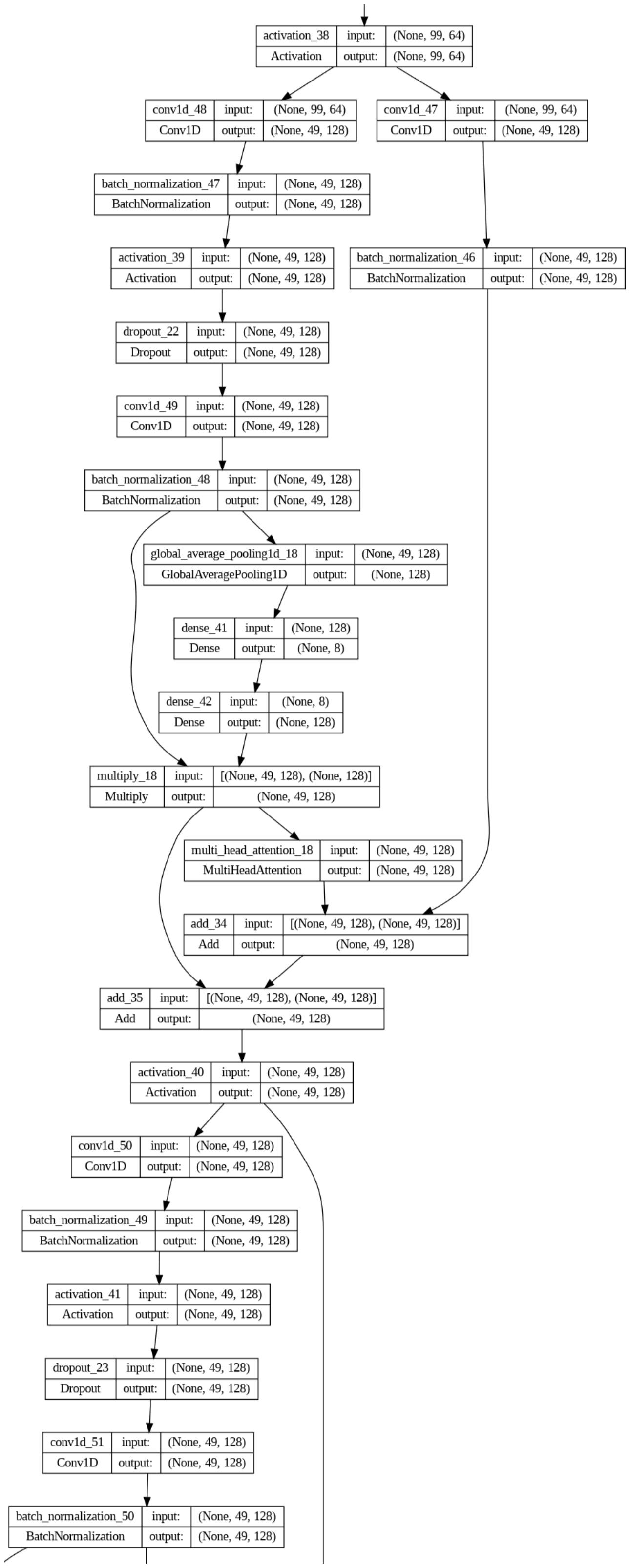

Figure 9: Frequency model architecture (part D)

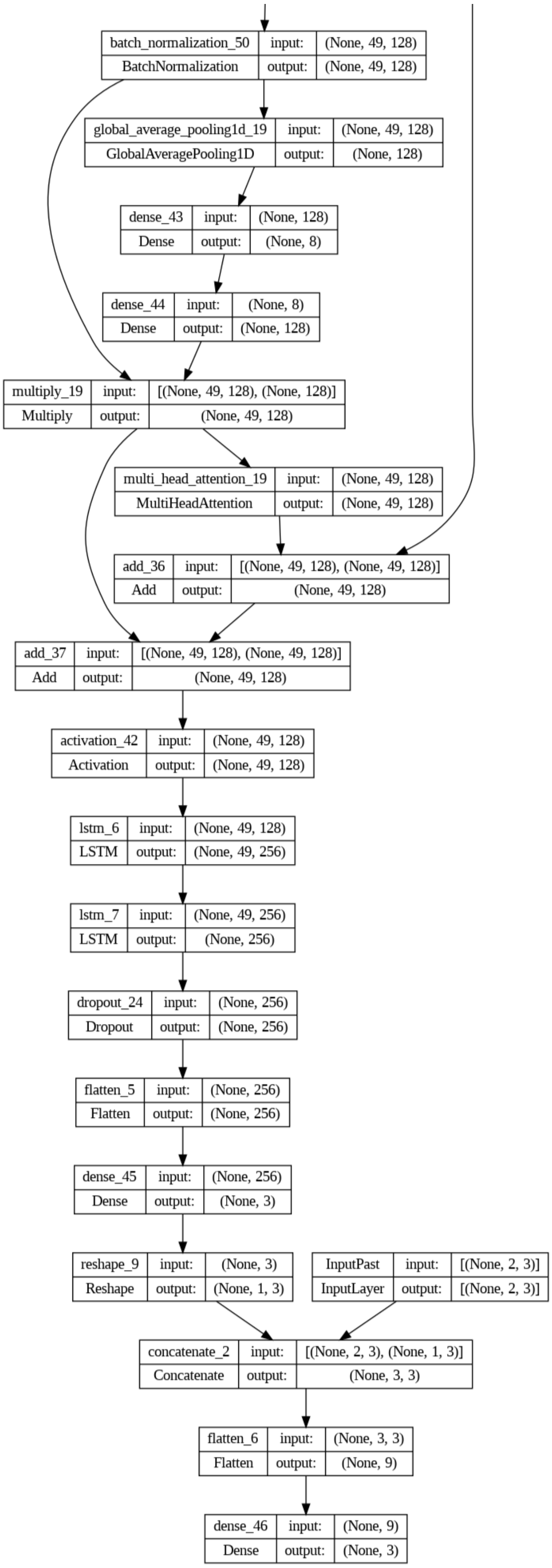

Figure 10: Frequency model architecture (part E)

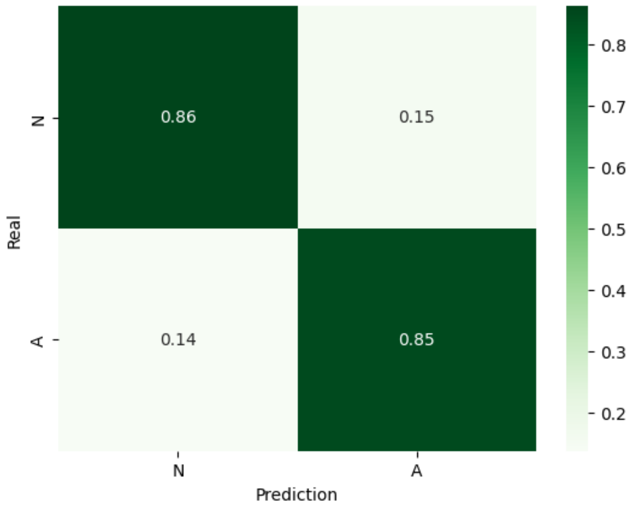

Figure 11: Task 1: Time model confusion matrix, discriminating between Normal and Abnormal beats

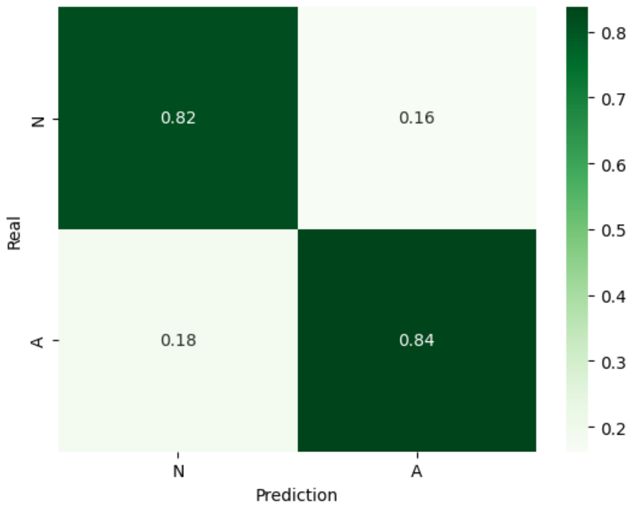

Figure 12: Task 1: Frequency model confusion matrix, discriminating between Normal and Abnormal beats

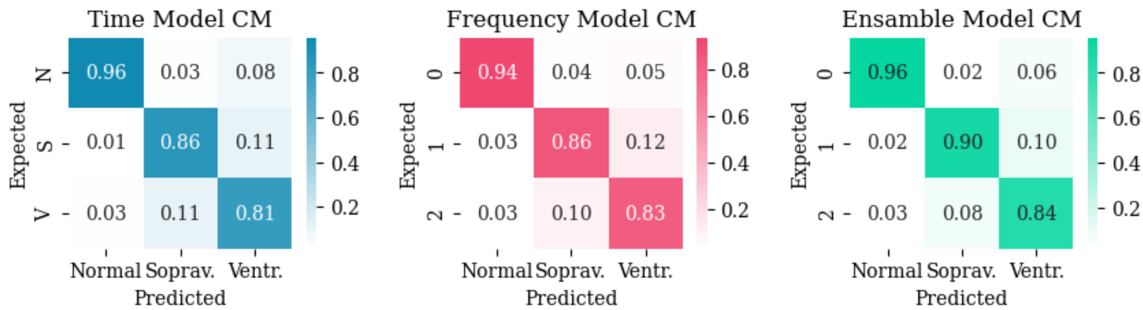

Figure 13: Task 2: Time/Frequency/Ensemble model confusion matrix, discriminating between Normal, Sopraventricular and Ventricular beats