Heartbeat Arrhythmia Detection using Deep Learning

Agustín Castillo Tosi
castosi@uw.edu
University of Washington
Washington, USA

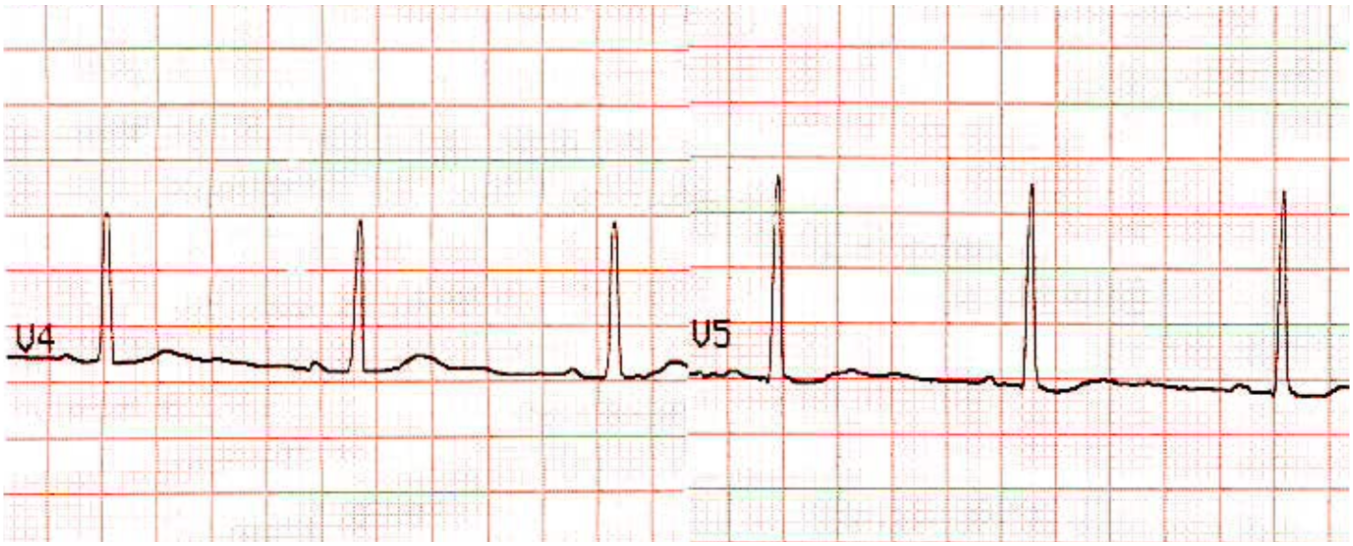

Figure 1: Example of a real ECG

ABSTRACT

The automatic detection of heart arrhythmias from electrocardiogram (ECG) measurements using Artificial Intelligence techniques has the potential to benefit health care by preventing some of the most common causes of death worldwide. This research has the objective of testing existing solutions as well as new proposed solutions for the task of detecting irregular heartbeat rhythms.

This project proposes two novel deep learning models for detecting the presence of arrhythmias on ECGs which leverage a combination of Convolutional Neural Networks (CNN) and Long Short Term Memory Networks (LSTM) working together.

1 INTRODUCTION

Over the last decade, innovative solutions using Artificial intelligence have been created for applications in many areas of Science and Medicine [4]. Advancements in medical technology, such as digital record systems and various medical devices, have contributed to the availability of vast amounts of data that doctors and researchers can easily access and utilize. AI applied to medicine can potentially improve modern healthcare, especially in disease diagnosis and treatment. Additionally, with the constant miniaturization of technology and the proliferation of wearable devices, patients can be monitored constantly while they perform their daily activities, which enables additional opportunities for disease prevention and medical prognosis.

According to the World Health Organization (WHO), the most common cause of death worldwide is cardiovascular diseases, which include conditions such as coronary artery disease, stroke, and heart failure—making it a high priority to tackle its diagnosis using AI tools. Fast methods that help detect cardiovascular diseases as early as possible are essential for patient survival.

1.1 Heart Arrhythmia

Irregular heartbeat rhythms, also now as arrhythmias [1], is a problem that occurs when the electrical signals that coordinate the heartbeats do not work correctly. This fault in the signal causes the heartbeat to be too fast, slow, or irregular. Arrhythmias can be grouped into two categories according to the speed of the heartbeat rate: Tachycardia and Bradycardia. Many heart diseases, including Myocardial Infarction, AV Block, Ventricular Tachycardia, and Atrial Fibrillation, can all be diagnosed by monitoring the heartbeat signals.

1.2 ECG

The simplest way to monitor and test the heartbeats is using an electrocardiogram (abbreviated as both ECG or EKG), which measures the heart's electrical conduction system. During an ECG, small electrical sensors called electrodes are attached to the skin, which detects and record the electrical signals produced by the heart as it beats. These signals are then displayed as a graph, which can be interpreted by a healthcare professional to diagnose various heart conditions [3]. Figure 2 shows the different sections that can be measured in the ECG waveform to detect irregularities in the heart rhythm.

AI can be used to analyze electrocardiogram (ECG) signals and detect abnormal patterns, such as arrhythmias and myocardial infarctions. This can help diagnose and monitor heart conditions more quickly and accurately, and also watch the effects of medications or treatments on the heart.

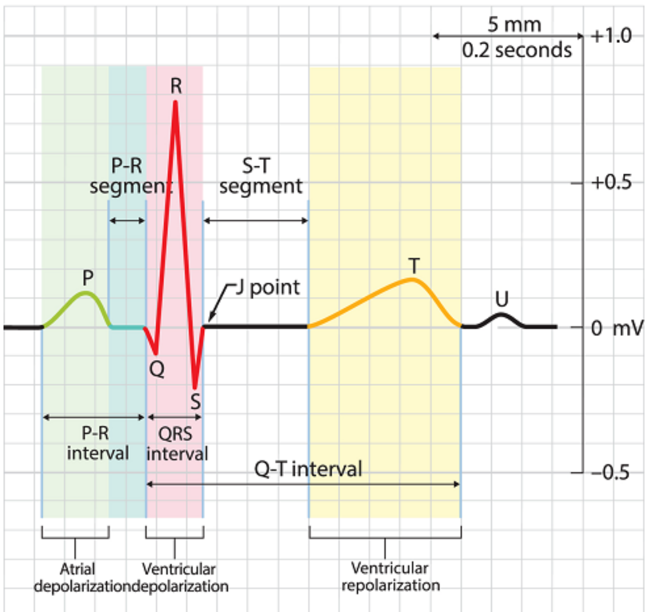

Figure 2: ECG waveform known as PQRST complex

1.3 Data

Several databases of ECG are publicly available. The data selected includes a collection of signals from the well-known heartbeat database MIT-BIH Arrhythmia Dataset [8]. This dataset in CSV format consists of 109,446 samples of ECG signals, each containing an example of 5 possible categories, including 3 types of **Irregular** heartbeats (Supraventricular premature beat, Premature ventricular contraction, and Fusion of ventricular and normal beat), **Normal** rhythm also known as **Sinus rhythm**, and **Unclassifiable** beats (that could be caused when the electrodes are being attached or detached from the patient or just noise). The subjects from whom the data was taken include men aged between 32 to 89 years and women aged between 23 to 89 years, but there is no mention of the racial composition of that group.

Detecting the presence of the mentioned rhythm categories in the ECG waveform is the first step in the diagnosis of heart problems. Table 1 show each category with its corresponding annotation symbol.

Table 1: Heartbeat Categories

Annotation	Meaning
N	Normal beat
S	Supraventricular premature beat
V	Premature ventricular contraction
F	Fusion of ventricular and normal beat
Q	Unclassifiable beat

Figure 3 shows the distribution of data across the five categories. It can be seen that it is very unbalanced, with the Normal category accumulating around 80 percent of the data. This difference in the number of samples for each label will present a challenge for training a model in the classification task because not all classes are equally represented. The solution for his problem will be discussed in section 2.

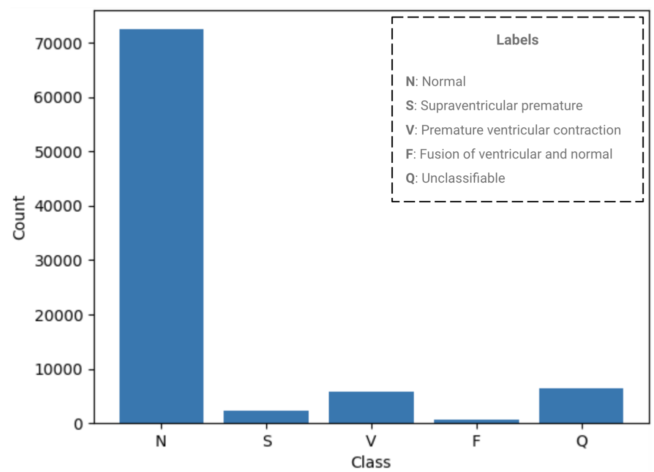

Figure 3: Data Distribution

This dataset of measurements taken over time can be described as Time Series of the univariate or 1-dimensional type. This data type is an ordered set of real values X of length equal to the number of real values T, where $X = [x_1, x_2, ..., x_T]$. This univariate Time Series Dataset is defined as $D = \{(X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)\}$ as a collection of pairs (X_i, Y_i) where X_i is a univariate time series with a Y_i possible class label [11] [5]. The task will consist in creating a model for dataset D to map a new input to a probability distribution of class labels.

1.4 Related work

Several solutions were proposed for this task using similar datasets to the one described in section 1.3. Still, the data on those studies differ in the number of cases and categories considered. In [10], they created their own database with 64.121 ECGs from 29,163 unique patients and 14 categories (including normal and noise). They developed a 34-layer Convolutional Neural Network that reaches an F1 score of 0.809, outperforming the average performance of the committee of cardiologists. The authors of [6] worked with a database of 12,186 ECG waves with three categories (Normal, Atrial Fibrillation, and Other) donated by a private company that manufactures ECG devices. The model they develop is a 13-layer Convolutional Neural Network that achieves an F1 score of 0.84 and 0.88 accuracy. Another related research is [9], which creates a Neural Network with 7 Convolutional Layers to classify ECG into four possible categories achieving an accuracy of 86 percent. It is also relevant to mention that non of those related articles mention details on the subjects' demographics of the ECG database used.

Other related researches are on the topic of Time Series Classification. A comprehensive review of deep learning methods for time series classification is done in [5], which proposes a taxonomy of models for time series classification and evaluates them using an archive of 97 univariate and multivariate time series datasets, reaching the conclusion that Fully Convolutional Networks (FCN) and Residual Networks (ResNet) are the best performant methods for this type of problems, outperforming state of the art techniques. In [11], a similar conclusion is achieved with FCN and ResNet using the same datasets. Finally, the article [7] explores using LSTM and FCN for time series classification.

2 METHOD

This research will focus on improving the results of previous work described in section 1.4 by creating models based on those solutions and tuning them to improve the accuracy using the **MIT-BIH** dataset. Additionally, a new neural network model architecture is proposed that uses a combination of LSTMs and CNNs differently from the previously mentioned research. Also, because this dataset differs slightly from the ones used in previous work, it is necessary to re-implement and test the performance of those models with this dataset to be used as a baseline benchmark in this research.

As mentioned in section 1.3, the dataset presents unbalanced classes, and this should be solved for the model to "see" the same amount of instances for each category during training. This is important to ensure that the model learns and generalizes well over all the classes in an even way. Several techniques exist to mitigate this problem of unbalanced data, like oversampling and undersampling. In this project, the latter will be used to balance the data.

The dataset is divided into two sets, with 80% for training and 20% for testing. For each proposed model architecture, a hyperparameter search is done using Grid Search over a defined set of hyperparameters and possible values they can take. Then the model with the best collection of hyperparameters found is trained again during more epochs with the training set and validated the results using the testing data to ensure a correct generalization of the learning. Finally, the resulting model was evaluated using the Accuracy and F1 metrics and the confusion matrix to see how many true positives and false negatives the model classifies from unseen data during training.

CNNs have already proved to perform well in this task, but other methods that can be used on time series data have not been applied to the ECG classification problem, mainly using LSTMs [2]. Previous work proposed using CNNs and LSTM networks working in parallel and then concatenating the result before the classification, achieving good performance for other time series classification tasks [7]. This research will apply this architecture for ECG classification to see how well it performs on this problem.

Additionally, this research proposes a different architectural combination of LSTM and CNN that was not used in the previous work to detect arrhythmias on ECGs. These models are all end-to-end approaches that can learn features without manual intervention, eliminating any human bias from the feature engineering process.

3 MODELS AND RESULTS

The following architectures will be compared in this research. These architectures will end in a classification layer of 5 output neurons corresponding with the number of expected output classes.

3.1 Convolutional Neural Network (CNN)

This neural network has been proven to be very effective in many problem domains, not only in Computer Vision but also in Time Series Classification [5]. In contrast with images that have two spatial dimensions, in this case, the data only has one dimension because it is a signal with values that changes over time. For this reason, the convolutional layers will be of 1 dimension [9].

This CNN model architecture consists of 4 blocks, each containing a 1D convolutional layer with 48 filters of size 7 with ReLU activation function and a max pooling layer of size 3. After that, it has a global pooling layer and one fully connected layer of 48 neurons, followed by the output layer.

This model achieves an accuracy of 89.8% and an F1 score of 0.914. Figure 4 shows the confusion matrix using the testing data, where it shows that the model correctly classifies almost all the cases with very few false positives. The class with fewer true positives is the *Supraventricular premature* beat with a Recall of 0.79.

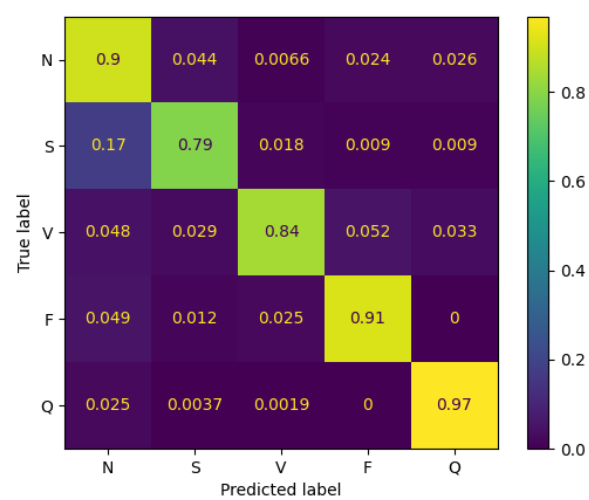

Figure 4: Confusion matrix of the CNN model

3.2 LSTM

Because the type of data has temporal dimension characteristics, Recurrent Neural Networks can be used. This type of neural network possesses several problems when applied to large datasets with many dimensions because of the "vanishing gradient problem" and connecting together information separated by a long time frame [2]. This problem of "long-term dependencies" makes it problematic to be applied in large time series data. Long Short Term Memory Networks (LSTM) were created to solve these problems by adding a mechanism to save a context state. An LSTM model can be used before a classification layer.

This model has an LSTM layer with 64 units, followed by a global pooling and a fully connected layer of 16 units. With an accuracy of 36.4 % and an F1 score of 0.436, this model does not achieve good performance. Although this model does not perform very well, looking at Figure 5, we see that the class *Unclassifiable* beat has a high recall, probably meaning not generalizing well enough on other classes. It is worth mentioning that this type of model is rarely used in Time Series Classification because they are mainly designed to predict the output for each element in the time series, and the vanishing gradient problems make it very difficult to train [5].

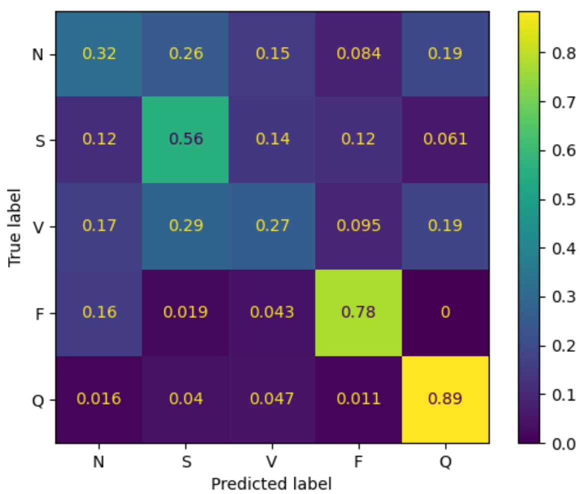

Figure 5: Confusion matrix of the LSTM model

3.3 LSTM and CNN in parallel

This architecture combines two types of models into one, running them in parallel. In this model, an LSTM and a CNN run simultaneously, and then the outputs are concatenated and passed onto a classification layer. This architecture has proven to be effective in Time Series Classification tasks [7], and in this research, it is applied to the problem of ECG classification.

The model has five convolutional layers with 64 filters of size 10 each and an LSTM with 80 units. Figure 6 summarizes the model's architecture. For the testing set, it reaches an accuracy of 84,7% and an F1 score of 0.878. Figure 7 shows the confusion matrix for this model.

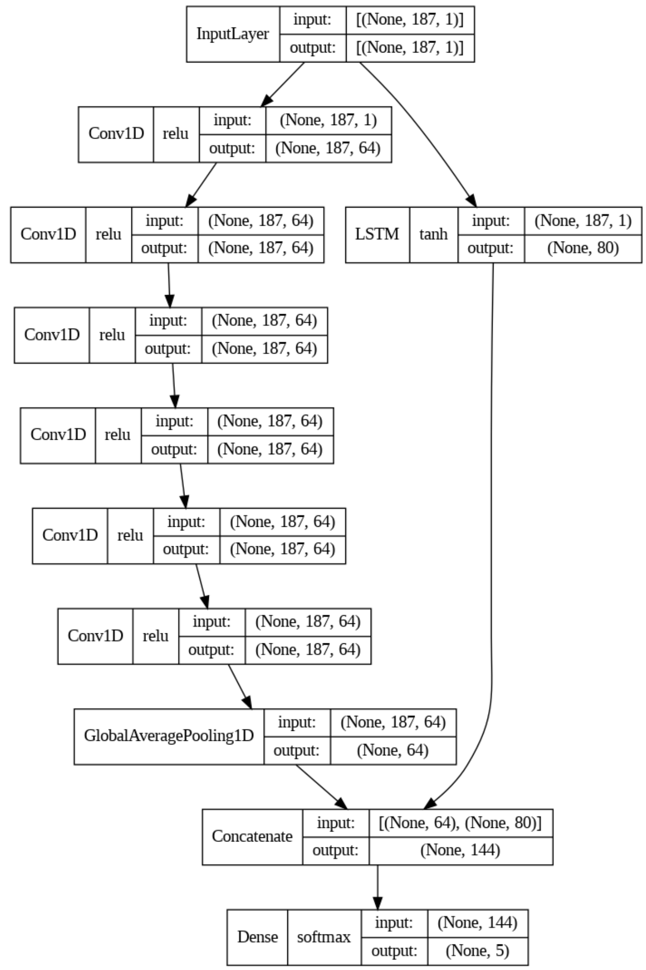

Figure 6: LSTM and CNN in parallel model architecture

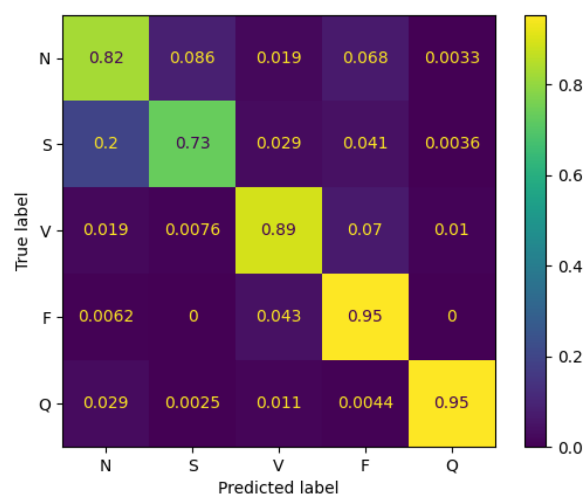

Figure 7: Confusion matrix of the LSTM and CNN in parallel model

3.4 LSTM and CNN sequential

After testing models already used in other research, either for the ECG classification or other time series classification tasks, a new model is proposed that was not used in previous research mentioned in the related work section 1.4 in this article.

In this new architecture, the input data is first feed to a CNN, and the result will be sent to the LSTM; after that, the result is handed to the classification layer. The idea behind this design is to use a CNN as a feature extractor for the LSTM model. The CNN has five blocks, each containing a 1D convolutional layer with 64 filters of size 10 with ReLU activation function and a max pooling layer of size 3. The following LSTM layer has 48 units. A summary of this architecture can be seen in Figure 8.

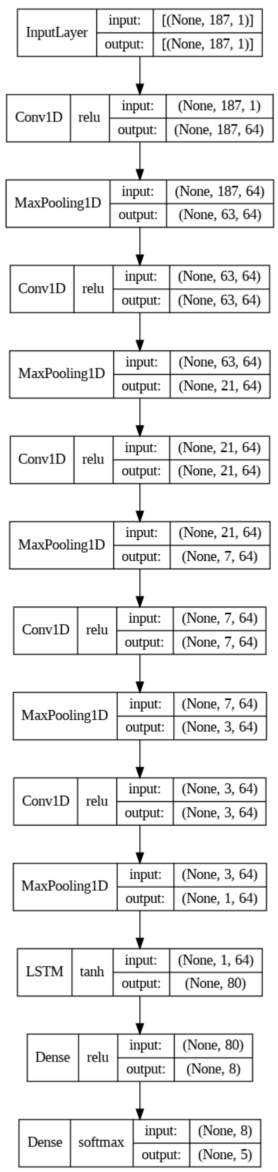

Figure 8: LSTM and CNN sequential architecture

After testing the model predictions using the testing set, it yields an accuracy of 91.3% and an F1 score of 0.927. The confusion matrix of this model, seen in Figure 9, also shows very high performance, and it can be seen that the true positives for all the classes are very high. However, the Recall for *Supraventricular premature* beat class is lower than the rest with a value of 0.77.

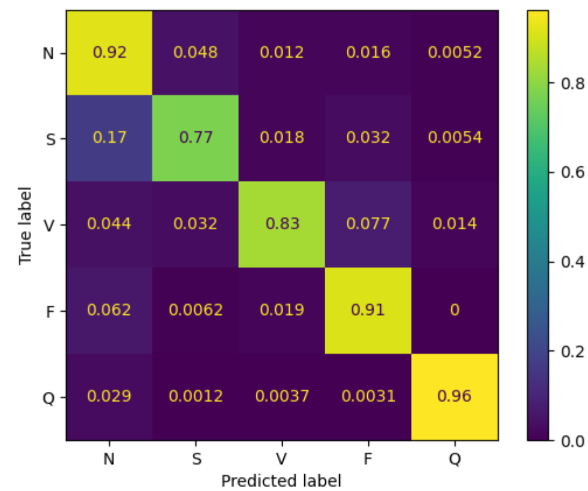

Figure 9: Confusion matrix of the LSTM and CNN sequential model

3.5 Models comparison

The models featured here, except for LSTM, perform well when evaluated with "unseen" data during training, achieving high accuracy and F1 score. Additionally, based on the confusion matrices, nearly all samples are correctly classified for each class.

Table 2 provides a summary of the evaluation metrics that were compared. The model that combines a CNN and an LSTM in sequence performs the best, with an accuracy of 91.3%, an F1 score of 0.927, and high Recall values for all labels. This model outperforms the others. Upon examining the confusion matrices, we observe that all models, except for the LSTM, exhibit some confusion between *Normal* and *Supraventricular premature* rhythm classes.

A summary of all the evaluation metrics compared can be seen in Table 2. The model with a CNN and an LSTM sequentially connected appears to be the best-performing one with an accuracy of 91.3%, an F1 score of 0.927, and very high Recall values for all labels, outperforming the other models. Looking at the confusion matrices we see that all the models present a comparable amount of confusion between *Normal* and *Supraventricular premature* rhythm classes.

Table 2: Summary comparing all the model's metrics

Model	Accuracy	F1
CNN	89.8%	0.914
LSTM	36.4%	0.436
CNN + LSTM in parallel	83.7%	0.878
CNN + LSTM sequential	91.3%	0.927

4 CONCLUSION

This research explored several Deep Learning models to solve the classification of heartbeat rhythms into *Normal beat*, *Supraventricular premature beat*, *Premature ventricular contraction*, *Fusion of ventricular and normal rhythm*, or *Unclassifiable beat*, using a dataset of ECG waveforms. 2 new models were introduced for this task, one already proved for time series classification problems, and a novel one. This new model combines CNN and LSTM neural networks and proved to be the best among the other models tested here, giving an accuracy of 91.3% and an F1 score of 0.927 when predicting samples from the testing set.

It is worth noticing some of the limitations of this research. First, the demographics of the dataset is very limited; more samples from subjects of different ages, sex, and races should be included, as these are important factors that could impact the probability of heart disease, minimize bias and increase the fairness of the models. Second, the class imbalance is an issue that ideally should be resolved by adding more cases of unrepresented labels. Finally, more types of arrhythmias should be considered. Future work should tackle these issues.

This technology can potentially help cardiology experts save time during ECG analysis and, as a result, improve healthcare services.

REFERENCES

[1] [n. d.]. Heart arrhythmia - Symptoms and causes. https://www.mayoclinic.org/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668

[2] 2015. Understanding LSTM Networks – colah's blog. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[3] 2022. ECG/EKG Interpretation Basics Review. https://www.registerednursern.com/ecg-ekg-interpretation-basics-review/

[4] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25, 1 (Jan. 2019), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[5] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33, 4 (July 2019), 917–963. https://doi.org/10.1007/s10618-019-00619-1 arXiv:1809.04356 [cs, stat].

[6] Sebastian D. Goodfellow, Andrew Goodwin, Robert Greer, Peter C. Laussen, Mjaye Mazwi, and Danny Eytan. 2018. Towards Understanding ECG Rhythm Classification Using Convolutional Neural Networks and Attention Mappings. In *Proceedings of the 3rd Machine Learning for Healthcare Conference*. PMLR, 83–101. https://proceedings.mlr.press/v85/goodfellow18a.html ISSN: 2640-3498.

[7] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2018. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* 6 (2018), 1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939 arXiv:1709.05206 [cs, stat].

[8] George B Moody and Roger G Mark. 1992. MIT-BIH Arrhythmia Database. https://doi.org/10.13026/C2F305

[9] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky. 2017. Deep Learning for ECG Classification. *Journal of Physics: Conference Series* 913, 1 (Oct. 2017), 012004. https://doi.org/10.1088/1742-6596/913/1/012004 Publisher: IOP Publishing.

[10] Pranav Rajpurkar, Awni Y. Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y. Ng. 2017. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1707.01836 arXiv:1707.01836 [cs] version: 1.

[11] Zhiguang Wang, Weizhong Yan, and Tim Oates. 2016. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. https://doi.org/10.48550/arXiv.1611.06455 arXiv:1611.06455 [cs, stat] version: 4.