ECG classifier for task WA31

Marco Anteghini

Introduction

An electrocardiogram (ECG) is an electrogram of the heart that can be visualized as a graph of voltage versus time of the electrical activity of the heart [1]. Ectrodes are placed on the patient's skin to check the heart's rhythm and its electrical activity. It is essential to help diagnose and monitor conditions affecting the heart.
In recent studies Deep Learning (DL) approaches have been used to better classify different ECG signals, with the final goal to obtain reliable diagnosis to support the clinicians [2,3].

This study was inspired by "ECG Heartbeat Classification: A Deep Transferable Representation" [2] that applies ResNet, a CNN based architecture, to electrocardiogram (ECG) data. We reproduced a ResNet architecture but we also explored a newly generated LSTM architecture.

Part of the codes were adapted from:

1. https://www.kaggle.com/code/gregoiredc/arrhythmia-on-ecg-classification-using-cnn
2. https://github.com/spdrnl/ecg/blob/master/ECG.ipynb

Database

We used two dataset:

1. The MIT-BIH Arrhythmia dataset [4]
2. The PTB Diagnostic ECG database [5]

Both datasets contain standardized ECG signals. Each observation has 187 time-steps per heartbeat.
An example of a typical ECG signal is visible in Figure 1.

We investigated the dataset composition. The distribution of different classes in the MIT-BIH dataset is visible in Figure 2 and Figure 3. Figure 4 shows the distribution of two classes in the PTB database. The MIT-BIH dataset downloaded from kaggle at https://www.kaggle.com/shayanfazeli/heartbeat and present 5 lables (N, S, V, F, Q). Since the PTB dataset only has 2 labels (N,A) where N=Normal and A=Abnormal, we mapped the 5 labels from the MIT-BIH dataset in to a binary classification:

N -> N
S,V,F,Q -> A

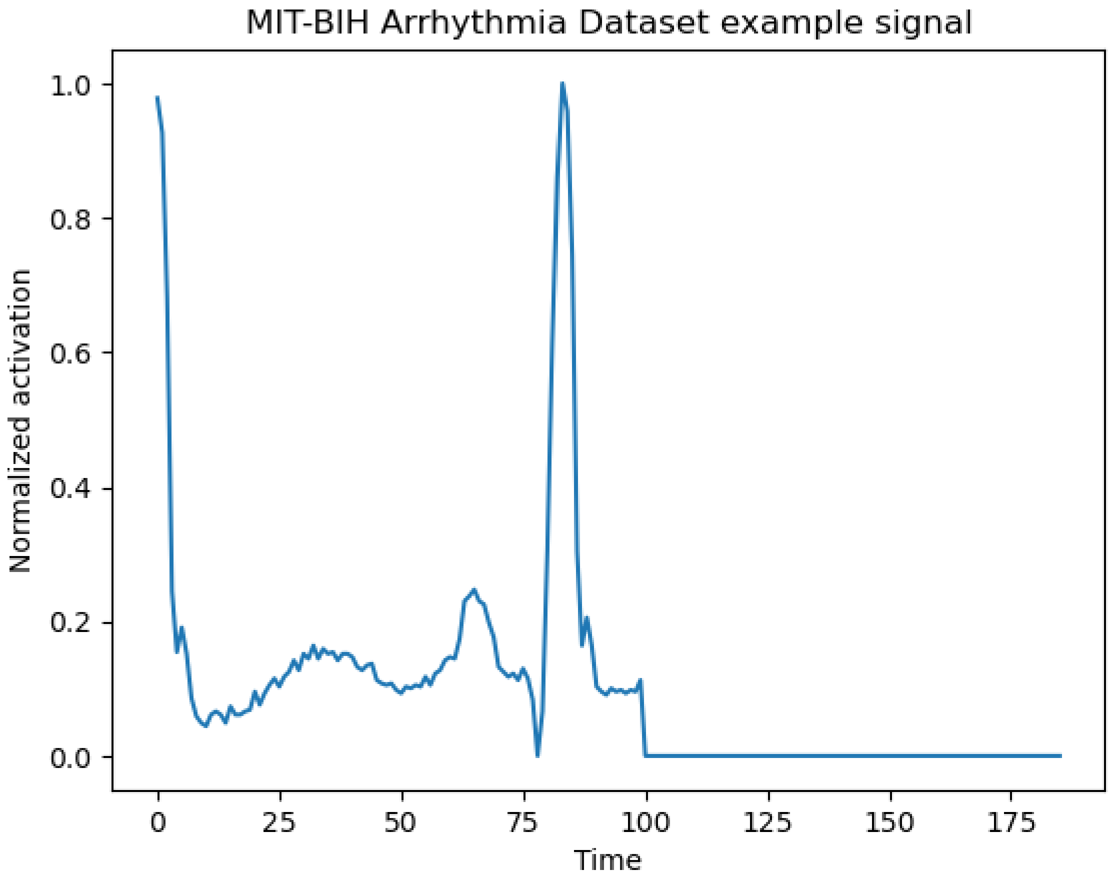

Figure 1. A sample ECG signal used in this study.

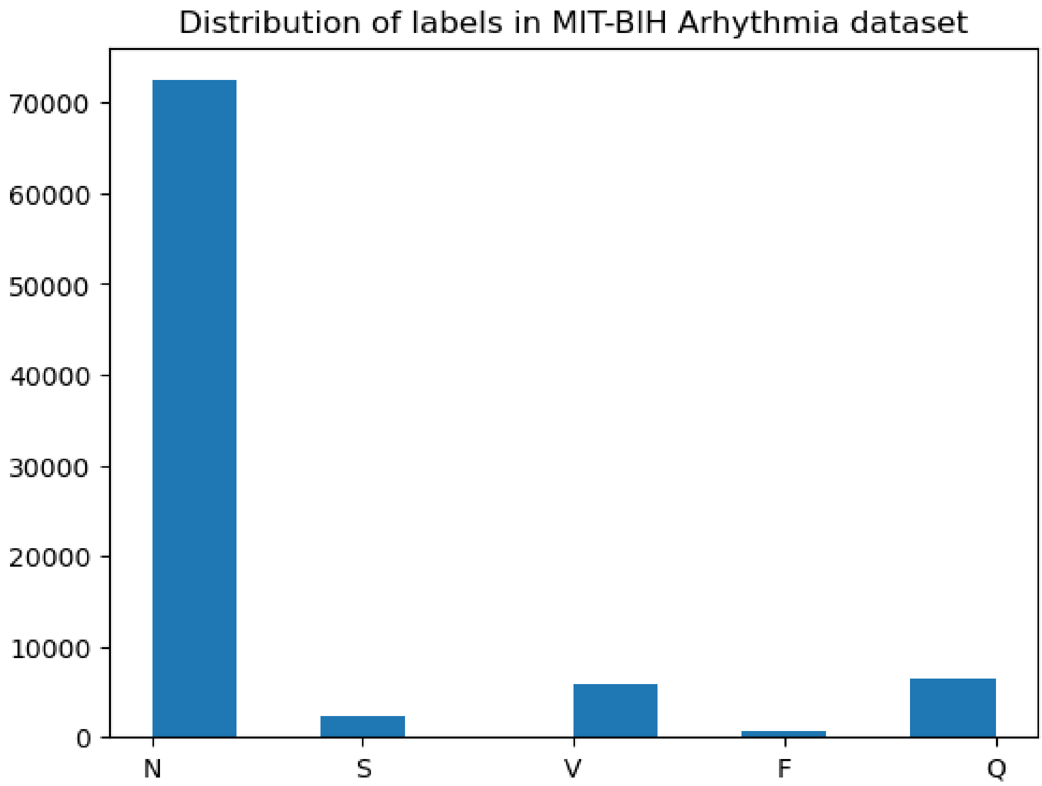

Figure 2. The classes distribution of the MIT-BIH dataset.

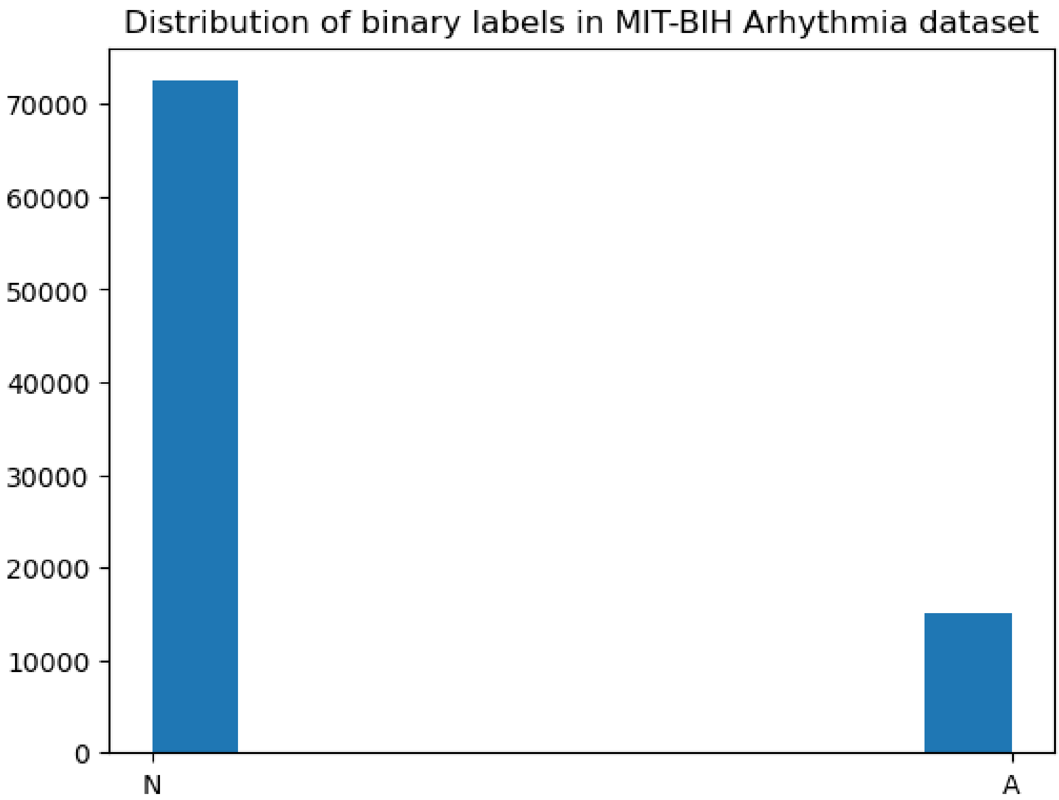

Figure 3. Distribution of the two classes in the MIT dataset.

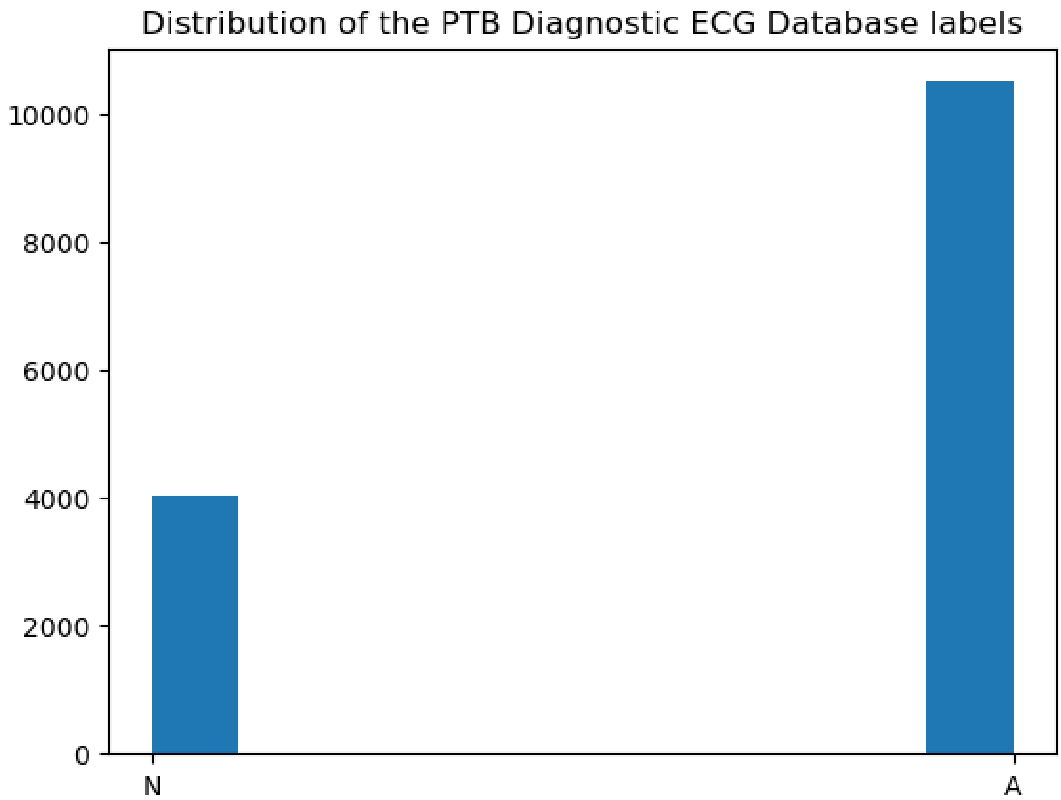

Figure 4. Distribution of the two classes in the MIT dataset.

Overview of the analysis

In the notebook available at
https://github.com/MarcoAnteghini/ECG_classifier/blob/main/notebooks/models.ipynb we show all the step to obtain: 1) A base model; 2) a RESNET model; 3) an LSTM model. The models are available at
https://github.com/MarcoAnteghini/ECG_classifier/tree/main/models.
These models were trained on the MIT-BIH dataset.

In the notebook available at
https://github.com/MarcoAnteghini/ECG_classifier/blob/main/notebooks/validation_and_transfer_learning.ipynb we show a validation process where we validate the model trained on the MIT-BIH dataset on the PTB dataset.
Given the low performances, we further applied a transfer learning approach. First, we divided the PTB dataset in training set and test set. Secondly, we re-trained both the RESNET model and the LSTM model on the new training set. Finally we tesed the re-trained models on the test set.

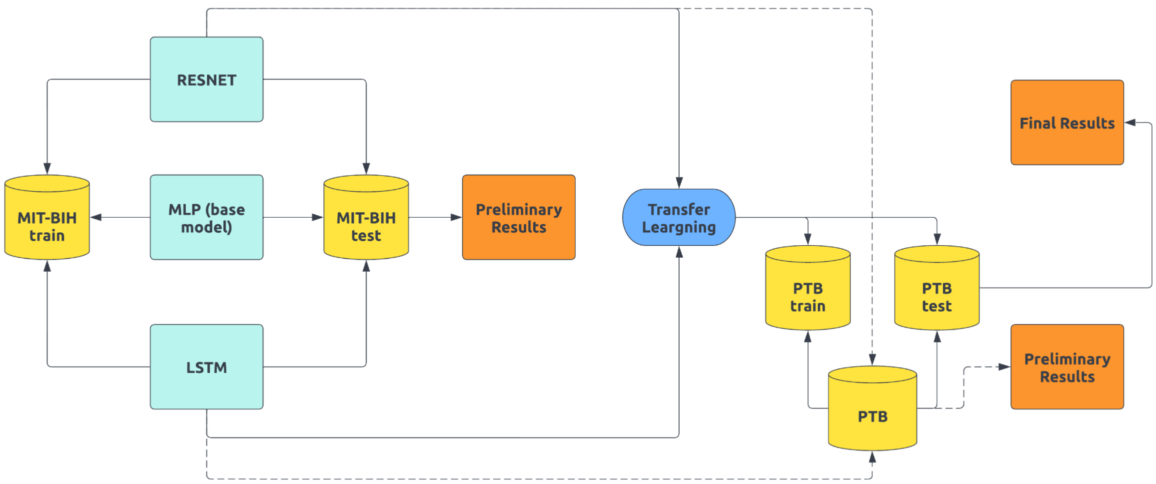

Figure 5. Overview of the full analysis. After training and testing the models on the MIT-BIH dataset we first validated the models against the PTB dataset (dashed arrows), then we applied a transfer learning approach after splitting the PTB dataset in training and test set.

Results

In the notebooks is shown that in order to reliably classify the PTB dataset we need to apply a transfer learning strategy with both the RESNET model and the LSTM model. Despite better performances of the RESNET model in classifying the MIT-BIH dataset (see notebook models for the Tsne visualization), the LSTM reached an higher accuracy on the PTB test set after re-training (LSTM - 97.70% vs RESNET - 92.96%).

REFERENCES:

[1] Lilly, Leonard S. Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty, 6th Edition. Lippincott Williams & Wilkins. pp. 70-78. ISBN 978-1-4698-9758-5. OCLC 1229852550. (2016).

[2] Kachuee, M., Fazeli, S., Sarrafzadeh, M.: Ecg heartbeat classification: A deep trans-ferable representation. In: 2018 IEEE International Conference on Healthcare In-formatics (ICHI). pp. 443-444. IEEE (2018).

[3] Darmawahyuni A, Nurmaini S, Rachmatullah MN, Tutuko B, Sapitri AI, Firdaus F, Fansyuri A, Predyansyah A. Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification. PeerJ Comput Sci. pp. 8:e825. (2022).

[4] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45–50. (2001).

[5] Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 S 317. (1995).