IT Project

Synoptic Assignment

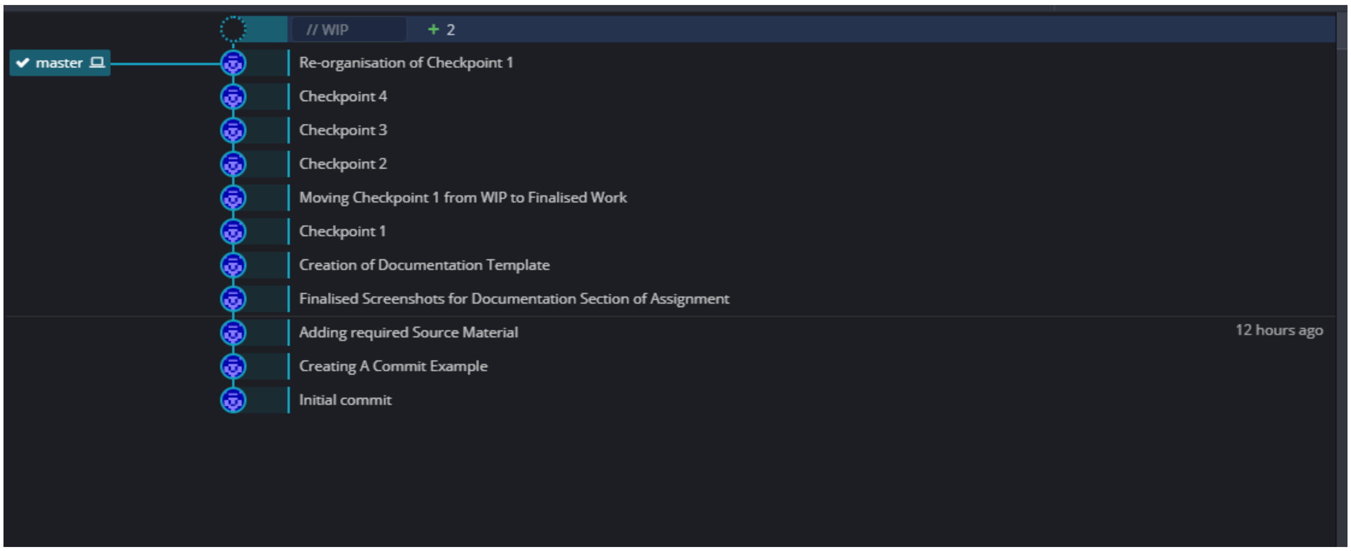
Sections A | B | C

By Matthew De Giorgio BAN 6.3A

Contents

Section A

Checkpoint	Code
Check Point 1	665502b7422f236eb49a0646e8f8813e64c3f4dc
Check Point 2	9c65cae960fa63ff4dba2dbe07f54832c24258e8
Check Point 3	28fa5788b0f116c90af7eb92288497eee7ce7e7f
Check Point 4	fa8761dcdd42d05a89c924875c17cb23d2e33263



Section B

Paper can be found here : https://arxiv.org/abs/1805.00794

Research Question

The authors of this paper had proposed to create a new method based on deep convolutional neural networks with aim of classifying five different types of heartbeats. Furthermore, they proposed an innovative way of knowledge transference to from this task to the Myocardial infarction classification task (MI) tackled by many other researchers. The MI classification task was included due to the fact it is classified using a cluster of heartbeats and other research papers had used the same dataset (MIT-BIH) for such a purpose.

Research Methodology

All research papers including this one that revolve around the classification of heartbeats and heat ailments are quantitative papers. This is only natural since to guarantee accuracy in the classification of these delicate phenomenon, the researcher would want to train and test their classifier on a much data as possible. The more distinct cases one can find in a dataset, the more accurate and precise the final classifier will be.

For this paper, *Kachuee et al* use the MIT-BIH database. The data had to be prepared before being used for neural network training. The beat extraction process involved the following steps:

- Splitting the ECG signal to ten seconds segments.
- Normalization of the amplitude values of to a range between 0 and 1
- Finding the set of all local maximums based on zero crossing of the first derivative.
- Finding the set of ECG R-peak candidates by applying a threshold to the normalized value.
- Find the median R-R intervals as the nominal heartbeat period of that window T
- For each R peak, selecting a part of the signal (1,2T long)
- Padding each part with 0 to equate all selections to a fixed length.

Once the dataset was finalized, the proposed neural network could be trained and evaluated. For each test conducted, a confusion matrix was created to extract key performance indicators that would be required for comparison and discussion of results.

Once the neural network's capabilities were defined and evaluated, the same networks and node weights were used to carry out the MI classification task.

Research Methodology of 2 other papers cited in research

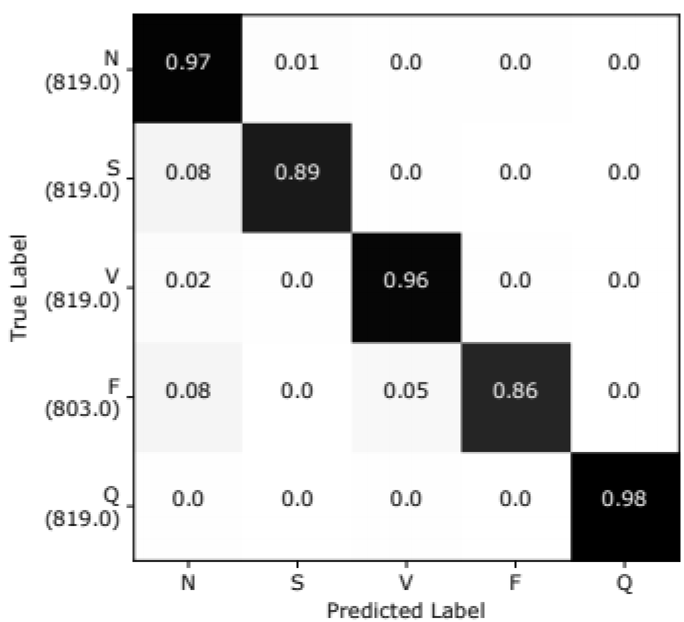
Comparing to two other papers by ***Archarya et al*** and ***Li and Zhou,*** One can see that similar approaches were taken.

For **Acharya's** research, the research methodology differed by the way the data was extracted and fed to the Convolutional Neural Network. After extraction, the dataset was up sampled and cleaned from noise to give the best possible chance for the more severe heartbeat types to be identified more often and correctly.

Li & Zhou used a similar extraction process but the random forest machine learning technique instead of a neural network. This technique has a stronger statistical prominence and is considered a step close to a white-box technique in comparison to neural network which are normally slated as black-box techniques since the users do not know the exact inner workings of the algorithm.

Presentation of Results

Kachuee and his team presented their results by including a sample confusion matrix to supplement the main result table.



The main results table basically compares the performance of their proposed neural network versus other papers mentioned in the literature. They also stated the type of approach used and the average accuracy (%)

TABLE II: Comparison of heartbeat classification results.

Work	Approach	Average Accuracy (%)
This Paper	**Deep residual CNN**	**93.4**
Acharya *et al.* [23]	Augmentation + CNN	93.5
Martis *et al.* [24]	DWT + SVM	93.8
Li *et al.* [25]	DWT + random forest	94.6

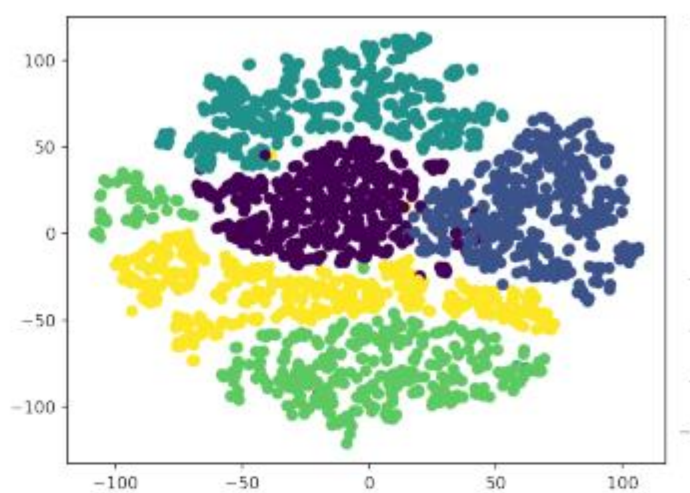
Finally, a table for the comparison of MI Classifiers was also presented. Accuracy, precision and recall were included to give more depth to the performance of each approach.

TABLE III: Comparison of MI classification results.

Work	Accuracy (%)	Precision (%)	Recall (%)
This Paper[1]	**95.9**	**95.2**	**95.1**
Acharya *et al.* [27][1]	93.5	92.8	93.7
Safdarian *et al.* [28][1]	94.7	–	–
Kojuri *et al.* [29][2]	95.6	97.9	93.3
Sun *et al.* [30][3]	–	82.4	92.6
Liu *et al.* [31][3]	94.4	–	–
Sharma *et al.* [26][3]	96	99	93

[1]: PTB dataset, ECG lead II
[2]: dataset collected by authors, 12-lead ECG
[3]: PTB dataset, 12-lead ECG

Visualization of the MI Classifier results were also included to give the reader an indication of the clustering of the network results.



(a)

Discussion of Results

The researcher simply describes how the results show that the proposed network is quite accurate at both Classification activities even though very minimal retooling of the network was required. They also note that the network was the second most accurate of the mentioned works.

Future Implications of the Study

A small comment on the possibility of using 12 lead ECGs to increase the accuracy of the network was made. This came about because the leading approach used for MI classification used 12 lead ECGs to train their classifier and therefore comparison is not as clear then the other approaches who all used dual lead ECGs.

Section C

Where to find the Data

The MIT-BIH Arrythmia database is maintained by the MIT Laboratory for Computational Physiology and is Hosted on **Physionet.org**. Once a user account is created, one can download and of the datasets hosted on the website and may even participate in any of the events. Each dataset included the MIT-BIH db are also hosted on GitHub incase the main Physionet website is down.

URL : https://physionet.org/content/mitdb/1.0.0/

How Was the Data Collected?

The MIT-BIH Arrythmia Database is over 40 years old. Data collection started in 1975 at the **Beth Israel Deconess Medical Center** located in Boston. The dataset is comprised of ECGs taken from various medical patients. The current data set contains 48 half-hour excepts made up of dual channel ECG Recordings from 47 different subjects studied by BIH Arrhythmia laboratory between 1975 and 1979. 23 of these recordings were chosen at random from a sample of 4000 ECG recordings collected from a population of Inpatients and outpatients. The remainder recordings (25) were selected from the same set of 4000 recordings but they were picked with a focus to include less common but still clinically significant arrhythmias that would have not been represented well enough in a random selection process.

The current dataset is given in a digitized format and contain annotations made by 2 independent cardiologists. A total of 110,000 Annotations are currently included within the database.

Cite : Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

An Alternative way of representing the source data

Any time this dataset was use in research; it would always be presented in a table like the following:

Class	0	1	2	3	4
Beat Type	Normal Beats (N)	Supraventricular Ectopic Beats (S)	Ventricular Ectopic Beats (V)	Fusion Beats (F)	Unknown Beats (Q)
Beats	72471	2223	5788	641	6431

Table 3.1.: Summary of Training Data

For my dissertation, apart from using a table, I also utilized a donut chart to show how disproportionate the dataset truly was. The donut chart easily transmits to the reader the proportionality of each class within the dataset.

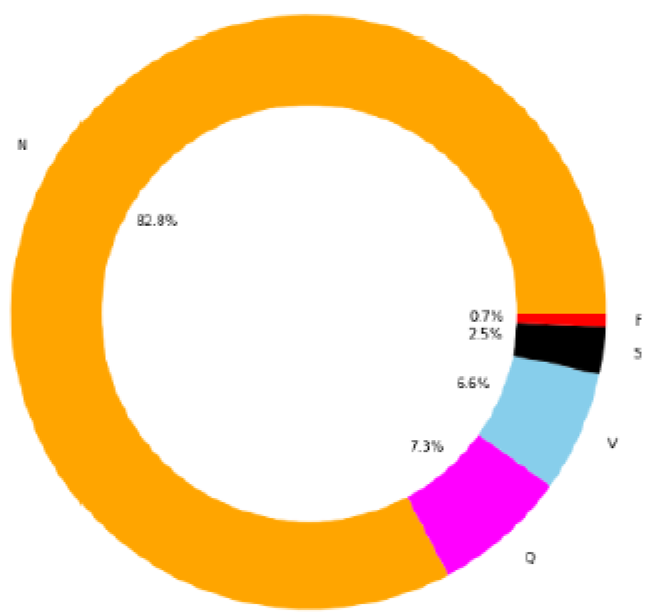


Figure 3.7: A donut chart visualising the class distribution in the default training data.