Report 1

Report 1

By
BI12-423 Nguyen Thi Thao

A report presented for the course
BI12— DS

ICT department
Ha Noi university of science and technology
March 6, 2024

Contents

1 Introduction

The heart is an organ which pumps blood throughout the body. As in the case of spiders and annelid worms, it can be a simple tube, but in the case of mollusks, it can be a slightly more complex structure with one or more receiving chambers (atria) and a primary pumping chamber (ventricle). The heart of a fish is a folded tube with three or four enlarged sections that resemble the chambers of the heart in mammals. The heartbeat is the result of the heart muscles contracting and relaxing regularly to pump blood throughout the body. The heart pumps approximately 8 pints of blood throughout the body every day at a rate of about 100,000 beats per day. This guarantees that waste items are removed from the body while oxygen- and nutrient-rich blood reaches all tissues and organs. Blood that has lost oxygen is sent by the heart to the lungs, where it picks up oxygen and expels carbon dioxide, a waste product of metabolism. The remainder of the body receives this oxygenated blood, which serves as vital fuel for cellular functions. Heart rate, as used in clinical diagnosis, is the regular contraction and expansion of the heart muscle, which produces a steady sound when the heart pumps blood. For evaluating the patient's general health as well as cardiovascular health, this is a crucial sign. Heart rate abnormalities or irregularities may be indicators of cardiovascular disease or other conditions, and they should be closely watched and assessed.

The World Health Organization (WHO) reports that heart disease claims the lives of over 17.9 million people year, with coronary artery disease and stroke accounting for 85% of these deaths. Heart disease claims the lives of over 200,000 Vietnamese citizens each year, surpassing the death toll from cancer. It is important to note that disorders including peripheral artery disease, coronary artery disease, and stroke are on the rise in younger people, whereas in the past, these conditions were usually associated with older adults.

Due to the fact that ECG is the most reliable source of information regarding the electrophysiological pattern of depolarization and repolarization of the heart muscles, heart-beat investigation is crucial for the early diagnosis of cardiovascular disorders such as arrhythmias and myocardial infarction (MI). An irregular heartbeat is a result of electrical signals not coordinating heartbeats, a condition known as arrhythmias. Heart attacks, also known as myocardial infarctions, are extremely threatening to human life because they obstruct the heart's flow of oxygen-rich blood, which can lead to severe cardiac arrest and even death [1].

In this report i was organized as follows: In Section 2, i present the background. Method part, which i put in the section 3. Section 4 and 5 are evaluation and conclusion repectively.

2 Background

Convolutional neural networks are able to extract complicated characteristics from two-dimensional data, such photographs, by using those data as input. The filters (kernels) are what allow this information to be extracted. The filter weights are changed during the training phase in order to create an accurate feature map for every class. A pooling layer needs to come after every convolutional layer in a CNN's design. To lighten the computational load, the pooling layers minimize overfitting and the quantity of network parameters. Max-pooling is the most popular kind of pooling; it works by choosing the highest value for each window. To perform the categorization of the characteristics that the kernels extracted, the convolutional network's last layers must be dense layers.

Electrocardiogram (ECG) signals are referred to as such in the context of Convolutional Neural Nets (CNNs). CNNs may be used to analyze ECG data for purposes including anomaly prediction, arrhythmia detection, and heartbeat categorization. The electrical activity of the heart at each time point is represented by a data point in a time-series data representation of an ECG signal. By doing convolutional operations throughout the temporal dimension of the data, CNNs are able to efficiently extract features from these signals, identifying patterns and correlations that are suggestive of various cardiac diseases. To enhance model performance, preprocessing methods like normalization and filtering can also be used on ECG data before to feeding them into CNNs. All things considered, CNNs provide a strong and adaptable method for using ECG data in a range of cardiac-related tasks, advancing medical monitoring and diagnosis.

3 Method

I used a time-tested technique called decision trees in this work. An input feature is assigned to each internal or intermediate node in a decision tree model. These are the intermediate nodes when a choice needs to be made amongst multiple options. The edges, which show the relationships between nodes, start with a node that has an input feature labeled with every conceivable value

for the target or output feature. On the other hand, the edge can lead to a decision node that is reliant on a different input characteristic. Leaf nodes hold the final class that was selected because they are labeled with a class or probability distribution among classes, indicating that the tree has classed the dataset into a particular class or probability distribution.

4 Evaluation

4.1 Dataset

The dataset used in this study includes 109,446 samples that are divided into five different groups, providing a thorough depiction of ECG heartbeat signals. This dataset, which has a sampling frequency of 125 Hz, records fine details of heart activity. The information comes from the MIT-BIH Arrhythmia Dataset on Physionet, a well-known source that is often used in cardiology research.

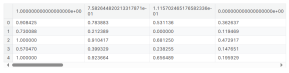

Figure 1: Dataset

The dataset is carefully prepared before analysis to guarantee consistency and dependability in the classification task. This includes preprocessing techniques including noise reduction, feature extraction, and normalization. This dataset facilitates improvements in medical diagnoses and monitoring by providing a useful resource for the development and assessment of algorithms targeted at precise heartbeat classification.

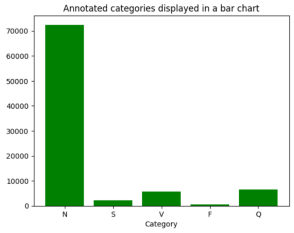

Figure 2: Imbalanced dataset

The data visualization that follows demonstrates the unequal distribution of the data. This may result in overfitting in labels with plenty of pictures and insufficient sample sizes for learning in labels with few data points, which might cause inaccuracies in the findings. By upsampling the minority categories to balance the dataset and correct the unbalanced data distribution, overfitting and other possible problems may be avoided.

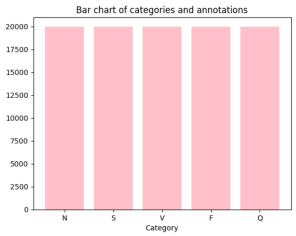

Figure 3: Balanced dataset

4.2 Metrics and results

I start by importing accuracy_score from sklearn.metrics and DecisionTreeClassifier from sklearn.tree. I then specify the original dataś characteristics and labels. The training dataś features and labels are denoted by X_train and y_train, whereas the test data's features and labels are represented by X_test and y_test. After that, I set up a Decision Tree classifier with the following parameters: random_state=42, max_depth=10. I use the f̈ittechnique to train the Decision Tree model using the training set. I use the äccuracy_scoreföunction to determine the modelś accuracy.

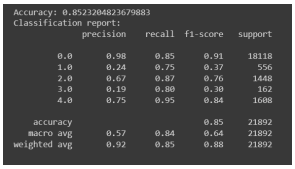

Figure 4: Classification results

ConfusionMatrixDisplay is a class that is imported from the sklearn.metrics library, enabling the confusion matrix to be shown in a visual chart. I then imported ConfusionMatrixDisplay from sklearn.metrics. I then made the following list of labels which the labels in this list correspond to the actual and expected classes in the confusion matrix. Here, the labels are included in the list. The ConfusionMatrixDisplay chart should then be created. With the labels given from the labels list and the confusion matrix calculated from y_test and y_pred, this line generates a ConfusionMatrixDisplay object. The ConfusionMatrixDisplay graphic should be shown.

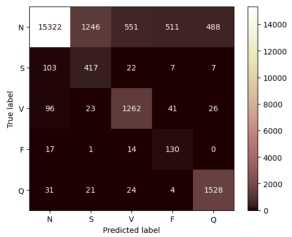

Figure 5: Enter Caption

The outcome that was seen above is dreadful. While labels "S" and "F" produced predictions with relatively low probability, label "N" demonstrated astonishingly effective outcomes. Due to the substantial influence this has had on the work as a whole, it has been determined that the Decision Tree model is not really appropriate for this particular challenge. I don't currently have a particular fix in mind.

5 Conclusion

We can plainly see from this task how important ECG is, as well as how CNN is used in their analysis. In order to show how adaptable the decision tree model is for precisely gauging the viability of machine learning, I presented it. I am not happy with the results of this article, though, as there were up to two incorrectly predicted labels, suggesting that there may not be enough precise answers. I'll tweak the settings and carry out further investigation in the future to ascertain the causes of the extremely low accuracy.

References

[1] U. R. Acharya, N. Kannathal, L. M. Hua, and L. M. Yi, "Study of heart rate variability signals at sitting and lying postures," *Journal of Bodywork and Movement Therapies*, vol. 9, pp. 134–141, 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:72594323