Cardiovascular Disease Prediction using Multi-layer Perceptron classifier and Logistic Regression

Jobb Q. Rodriguez
Ateneo de Naga University
Maryville Subdiv., Brgy. San Felipe
Naga City, Camarines Sur,
+639619116147
jobb.rodriguez22@gmail.com

ABSTRACT

In this paper, the author used Multi-layer Perceptron classifier (MPC) and Logistic Regression (LR) to identify which configuration of MPC performs best and identify if LR outperforms MPC in predicting the cardiovascular disease given the dataset, *Cardiovascular Disease Dataset* [1]. Moreover, the author aimed to identify if the *Body Mass Index (BMI)* can improve the performance of the models. The MPC model, without BMI, with the default configuration performed the best with an average test accuracy score of 73.39%. The MPC model, with BMI, with the default configuration performed the best with an average test accuracy score of 73.36%. Without BMI, the MPC's average test accuracy score was 1.36% higher than LR. With BMI, the MPC's average test accuracy score was 1.33% higher than LR. The BMI does not improve the models' performances.

CCS Concepts

• Computing methodologies→Machine learning and Modeling and simulation

Keywords

Cardiovascular Disease; Multi-layer Perceptron classifier; Logistic Regression; Classification; Machine Learning

1. INTRODUCTION

The cardiovascular disease status must be predicted given the set of features from the *Cardiovascular Disease Dataset* [1]. The author's tasks were to identify which configuration of MPC performs best, identify if LR outperforms MPC performs best, and identify if the BMI can improve the performance of the models. Accuracy, the percentage of correct predictions, was the metric used for measuring the models' performances.

2. RESULTS

2.1 Performance of the MPC Configurations without the BMI

Without the BMI, the MPC had three (3) configurations. The first configuration was the MPC's default configuration. The second configuration modified the MPC's hidden layer sizes to 75. The third configuration modified the MPC's hidden layer sizes to 75, the maximum number of iterations to 300, and the tolerance for the optimization to 0.00001.

Table 1. The Average Accuracy Scores of the MPC Configurations without the BMI

Configuration	Average Accuracy Scores (%)	
	Train	Test
First	73.92	73.39
Second	73.85	73.40
Third	73.89	73.43

For the average accuracy score for the training dataset, the first configuration performed best with an accuracy score of 73.92%. For the average accuracy score for the test dataset, the third configuration performed best with an average accuracy score of 73.43%.

2.2 Performance of LR without the BMI

Without the BMI, the LR used its default configuration. The average accuracy score for the training dataset was 72.06%. The average accuracy score for the test dataset was 72.03%.

2.3 Performance of the MPC Configurations with the BMI

With the BMI, the MPC had the same three (3) configurations with the MPC configurations without the BMI.

Table 2. The Average Accuracy Scores of the MPC Configurations with the BMI

Configuration	Average Accuracy Scores (%)	
	Train	Test
First	73.88	73.36
Second	73.86	73.42
Third	73.83	73.49

For the average accuracy score for the training dataset, the first configuration performed best with an accuracy score of 73.88%. For the average accuracy score for the test dataset, the third configuration performed best with an average accuracy score of 73.49%.

2.4 Performance of LR with the BMI

With the BMI, the LR used its default configuration. The average accuracy score for the training dataset was 72.06%. The average accuracy score for the test dataset was 72.03%.

3. ANALYSIS

3.1 The Top Performing MPC Configuration

Using the data from Table 1, the third configuration was the top performing MPC configuration without the BMI. It had an average test accuracy score of 73.43% which was 0.07% higher than the second configuration and 0.13% higher than the first configuration. Its average test accuracy score was 0.46% lower than its average training accuracy score which showed the configuration's reliability to predict given an unknown dataset. However, given that the third configuration had a lower number of hidden layers, a higher maximum number of iterations, and a lower tolerance for the optimization, the first configuration, with an average test accuracy score of 73.39%, was the top performing MPC configuration without the BMI. Its average test accuracy score was 0.53% lower than its average training accuracy score which shows the configuration's reliability to predict given an unknown dataset.

Using the data from Table 2, the third configuration was the top performing MPC configuration with the BMI. It had an average test accuracy score of 73.49% which was 0.03% higher than the second configuration and 0.04% higher than the first configuration. Its average test accuracy score was 0.34% lower than its average training accuracy score which showed the configuration's reliability to predict given an unknown dataset. However, given that the third configuration had a lower number of hidden layers, a higher maximum number of iterations, and a lower tolerance for the optimization, the first configuration, with an average test accuracy score of 73.36%, was the top performing MPC configuration with the BMI. Its average test accuracy score was 0.52% lower than its average training accuracy score which shows the configuration's reliability to predict given an unknown dataset.

3.2 MPC's Results vs LR's Results

Given the analysis from 3.1 and the results in 2.2, without the BMI, the MPC outperforms LR in the given dataset. The MPC's top performing configuration had an average training accuracy score of 73.92% which was 1.86% higher than LR's average accuracy training score and an average accuracy test score of 73.39% which was 1.36% higher than LR's average accuracy test score.

With the BMI, the MPC outperforms LR in the given dataset. The MPC's top performing configuration had an average training accuracy score of 73.88% which was 1.82% higher than LR's average accuracy training score and an average accuracy test score of 73.36% which was 1.33% higher than LR's average accuracy test score.

3.3 Models Without BMI's results vs Models With BMI's results

Table 3. The Average Test Accuracy Scores of the Top Performing MPC and LR with the BMI and without the BMI

BMI	Average Test Accuracy Scores (%)		
	Present	Absent	Difference
MPC	73.36	73.39	- 0.06
LR	72.03	72.03	0.00

For LR, there was no difference between the presence and the absence of the BMI. For the top performing MPC configuration, the average test accuracy score with the BMI was 0.06% lower thant the average test accuracy score with the BMI. However, given that the difference is not below -2.00% and is not above 2.00%, the BMI does not improve the models' performances.

4. CONCLUSION

With or without the BMI, the first configuration was the top performing MPC configuration. Moreover, with or without BMI, the MPC outperforms LR in the given dataset. Also, the BMI does not improve the models' performances.

5. REFERENCE

[1] Svetlana Ulianova. 2019. Cardiovascular Disease dataset. Retrieved from https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset