Journal of Mechanics Engineering and Automation 11 (2021) 180-182
doi: 10.17265/2159-5275/2021.06.004

Stroke Prediction

Sydney Caulfeild[1,2], Sarah Pak[1,2], Nathanael Yao[1,2] and Hoz Rashid[1,2]

1. QMIND-Queen's AI Hub

2. Queen's University, Kingston, Ontario K7L 3N6, Canada

Abstract: Stroke prevention methods were explored because strokes are the leading cause of death in Canada and the number of deaths is expected to grow over the next decade. As a solution to this problem, a model was created to predict a person's risk of having a stroke based on their physical traits and lifestyles. Current models that use facial recognition to detect strokes provide certain benefits but are less desirable than the proposed stroke prediction model as they can not detect strokes before they happen. The team utilized a multilayer perceptron model on data containing attributes such as gender, blood pressure, and BMI (body mass index) to predict a person's likelihood of stroke. The model was able to achieve 96% accuracy on testing data.

Key words: Stroke, artificial intelligence, multilayer perceptron.

1. Introduction

1.1 Motivation

Stroke is the third leading cause of death in Canada. Strokes are caused by the sudden loss of brain function that follows a brain blood vessel blockage or rupture. Its symptoms include loss of sensation, difficulty speaking, vision difficulties, headache, and loss of coordination [1].

The effects of a stroke can range from mild to severe, and among those who survive their stroke, many never fully recover. According to Statistics Canada, approximately 36% of these survivors are left with significant disabilities five years after their stroke and more than 40% require assistance with activities of daily living [1].

In a study conducted in 2015 by the American Heart Association, the number of individuals experiencing the effects of stroke in Canada is projected to increase from 405,000 in 2013 to between 654,000 and 726,000 in 2038 [2].

With the high rates of stroke in Canada and the expected growth in the next decade, the team was motivated to explore methods for stroke prediction and prevention.

More specifically, different machine learning models for stroke prediction were compared and optimized.

1.2 Related Works

There have been various studies on stroke prediction, including research conducted in 2018 that used images of facial features to determine whether a person has had a stroke [3].

They focused on the expressional asymmetry and mouth askew to make predictions. To classify facial stroke, the SVM (Support Vector Machine), RF (Random Forest), and Bayesian Classifier were adopted as classifiers. The research found that the recognition accuracy of SVM, RF, and Bayes is 100%, 95.45%, and 100%, respectively [3].

1.3 Problem Definition

Models that use facial expressions require pictures to inform a patient if they are having a stroke. Rather than identifying a case of stroke, the desired solution will use data on people's lifestyles to state their likelihood of having a stroke. Not only are those data more accessible than pictures, but they can also quantify a patient's susceptibility to a stroke and allow them to change their unhealthy habits to prevent one from

Corresponding author: Sydney Caulfeild, BASc.

occurring.

2. Methodology

2.1 Data Generation

Many factors affect one's susceptibility to stroke, including gender, blood pressure, BMI (body mass index), and physical activity [4]. A dataset was sourced from Kaggle as it contains all the major stroke risk factors [5]. The dataset has 11 attributes and a label indicating whether each patient has had a stroke.

A heatmap was used to analyze the correlation between all attributes and stroke. As seen in Fig. 1, BMI has the weakest relationship to stroke, however, the performance decreased when the attribute was removed. BMI is also the only attribute containing null values which were all replaced with the BMI attribute's mean.

2.2 Model Creation

The selected model for this solution is a multilayer perceptron. Multilayer perceptrons are a feedforward neural network that has an input layer, an output layer, and hidden layers. Neural networks such as multilayer perceptrons are useful for predicting results based on input data because the networks are structured in different layers.

After some testing with other types of models such as an SVC model, the team chose to use a multilayer perceptron because it is effective at handling a wide range of different features. Since the stroke prediction dataset has 11 attributes, the results that the team got from the multilayer perceptron model were better than other models that were tested.

2.3 Additional Analysis

Further modifications were done to the model to optimize the multilayer perceptron model's performance. The number of epochs, nodes, and layers was varied to analyze their impact on the model's behavior. All modifications were evaluated by comparing the model's maximum validation and training scores.

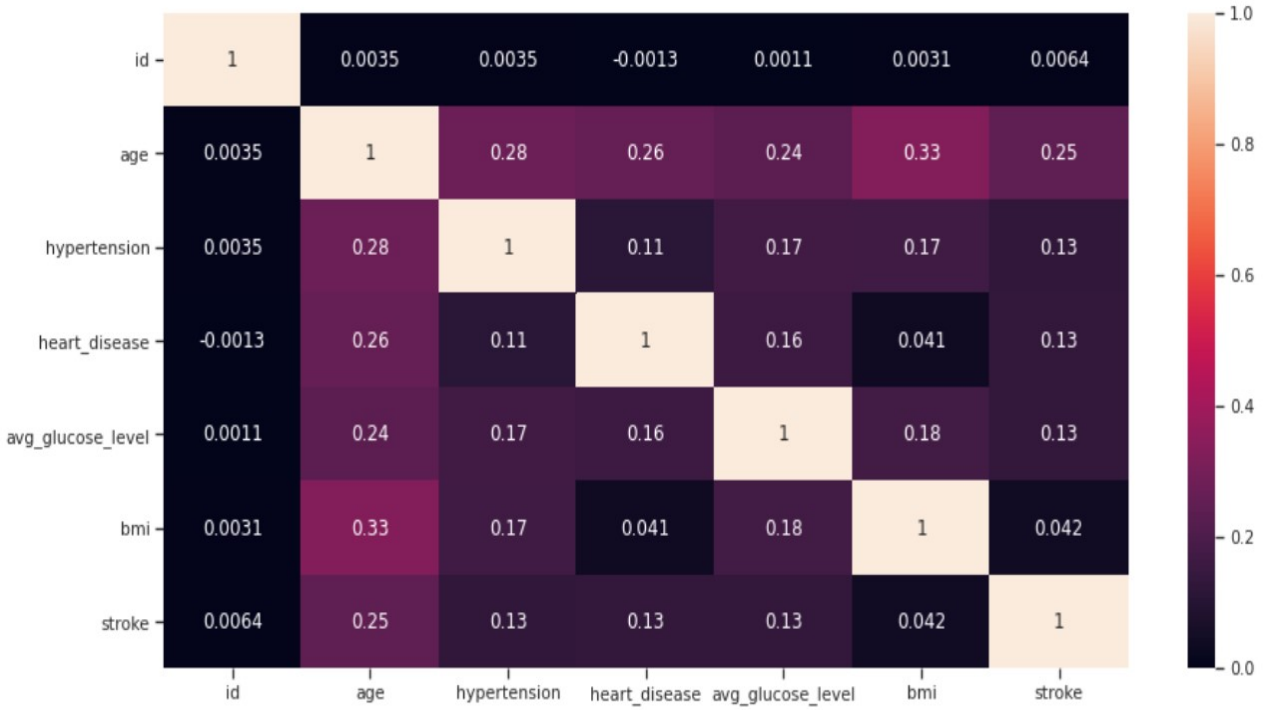

Fig. 1 A heatmap used to visualize the relationship strength among attributes and stroke.

Table 1 Results of the different model variations. The training values are the model's accuracy during its training while the testing values are the model's accuracy using the test data.

Model	Training	Testing
V1	0.960	0.952
V2	0.960	0.9508
V3	0.9544	0.9615
V4	0.9579	0.9532

3. Results and Discussion

In the first version of our model, V1, values were normalized using a mapping function that takes in data and outputs an array of unique strings. The multilayer perceptron had a top layer with 10 nodes, a middle layer with 20 nodes, and an output layer with 3 nodes. The model was tested with 150 epochs and a batch size of 30.

The first version of the model demonstrated overfitting after 16 epochs. Therefore, the second version, V2, had only 30 epochs and another middle layer with 10 nodes. However, as seen in Table 1, the testing accuracy for the second version decreased.

For the third version of our model, V3, the number of nodes in each dense layer of the multilayer perceptron was modified. The number of nodes in the top layer was equal to the number of features, two middle layers were added with 30 nodes and 10 nodes respectively, and an output layer with 3 nodes was used.

In the final version of the model, V4, the number of nodes was changed again for the four dense layers. In the top layer the number of nodes was again equal to the number of features. There were two middle layers with 40 nodes and 30 nodes, and an output layer with 2 nodes.

The third version produced the highest percent accuracy during testing and as such was selected as the final model.

4. Conclusions and Future Work

The team built a model that predicts a person's chance of having a stroke given certain inputs. Using data to assess a person's risk of having a stroke may help assist doctors in identifying people who are at high risk and give them appropriate advice and treatment.

The next steps for the model include testing different learning methods such as RF to determine which learning method is optimal for this dataset. Although a multilayer perceptron model was selected because of its efficiency at handling multiple features, it is important to test other models.

Furthermore, the model may be improved by testing out different methods of data preprocessing to know which data are most useful for predicting a stroke. For instance, other inputs can be removed or modified to see which inputs are important to getting an accurate prediction.

Finally, more data should be collected to construct our own dataset to put our model to the test against current data.

References

[1] Krueger, H., Koot, J., Hall, R. E., O'Callaghan, C., Bayley, M., and Corbett, D. 2015. "Prevalence of Individuals Experiencing the Effects of Stroke in Canada: Trends and Projections." NIH.

[2] Government of Canada. 2019. "Stroke in Canada: Highlights from the Canadian Chronic Disease Surveillance System." Accessed March 2, 2021. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/stroke-canada-fact-sheet.htm.

[3] Chang, C.-Y., Cheng, M.-J., and Ma, M. H.-M. 2018. "Application of Machine Learning for Facial Stroke Detection." IEEE 2018.

[4] American Stroke Association. 2021. "Stroke Risk Factors." Accessed Feb. 2, 2021. https://www.stroke.org/en/about-stroke/stroke-risk-factors.

[5] Soriano, F. 2021. "Stroke Prediction Dataset." Accessed Jan. 8, 2021. https://www.kaggle.com/fedesoriano/stroke-prediction-dataset?fbclid=IwAR0JNuSXufCZgf16g4Hoiq8mfoULN28r6wtxZ-zXEWx-qWJQttIhDOieC2k.