

Brain Stroke Prediction using Machine Learning

GROUP-11 :- DSBDA (MINI PROJECT)

Group Details

Group-11

Group members:

- 1) Deepak Abande (33101)
- 2) Sumit Chavan (33123)
- 3) Sakshi Dalvi (33117)
- 4) Divya Dhomase (33123)

Agenda

- u Introduction
- u Literature Survey
- u Proposed System Methodology
- u Demonstration
- u Results
- u Conclusion
- u References

Abstract

Tearing of the blood vessels present in the brain leads to serious health condition named as brain stroke. It can also occur when there is distraction in blood flow and other important nutrients to the brain parts. According to the report presented by World Health Organization (WHO), stroke is the main cause of death and disability in the world.

For heart stroke prediction, various work has been carried out but in order to predict brain stroke, very few work has been carried out. Various machine learning techniques and models are designed to detect the probability of stroke occurrence in the brain.

This mini project work has used machine learning algorithms like Logistic Regression, Decision Tree Classification, Support Vector Machine and has taken various physiological factors to train three different models for accurate prediction.

Keywords:

Stroke, machine learning, decision tree classification, support vector machine, Logistic regression.

Introduction

- u Stroke is the second leading cause of death in the world and it will remain as an important health burden for the individuals and for national healthcare systems. Potentially most identifiable risk factors for stroke occurrence include hypertension, cardiac disease, diabetes, atrial fibrillation, and also various lifestyle factors, etc.
- u A dataset used in the project work is referred from Kaggle with various physiological traits as its attributes to proceed with this task. These traits are then analyzed and used for the final detection/prediction. First of all, the dataset is cleaned and made ready for the machine learning model to understand which is the process of data- preprocessing. For this purpose, the dataset is initially checked for null values and fill them with not null values. Then to convert string values into integers Label encoding is performed followed by one-hot encoding.
- u After the step of Data Preprocessing, the mentioned dataset is used to split into two different parts i.e. train data and test data. Using this new data, a model is then built with the help of various Classification Algorithms. Using methods like confusion matrix, accuracy is calculated for all the algorithms and compared to get the best-trained model for prediction purpose. After proper analysis, the project work concludes which algorithm is most appropriate for the prediction of brain stroke.

Literature Survey

In order to get the required knowledge and information about various concepts and algorithms related to the present analysis of stroke, existing literature were studied. Some of the important conclusions were made from these surveys. Some of them are listed below.

2.1 “Prediction of Brain Stroke Severity Using Machine Learning. In: International Information and Engineering Technology Association (2020)”- Vamsi Bandi, Debnath Bhattacharyya, Divya Midhun Chakravarthy

The authors of this paper have performed the work of brain stroke prediction by using random forest algorithm which helped to analyze the levels of risks obtained from the strokes. Authors suggested that , this method will give better performance when compared to the existing algorithms. This research is limited to very less types of strokes and cannot be used for any new stroke type in the upcoming future.

2.2 **“Predicting stroke from electronic health records. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE (2019)”**- Nwosu, C.S., Dev, S., Bhardwaj, P., Veera Valli, B., John, D.

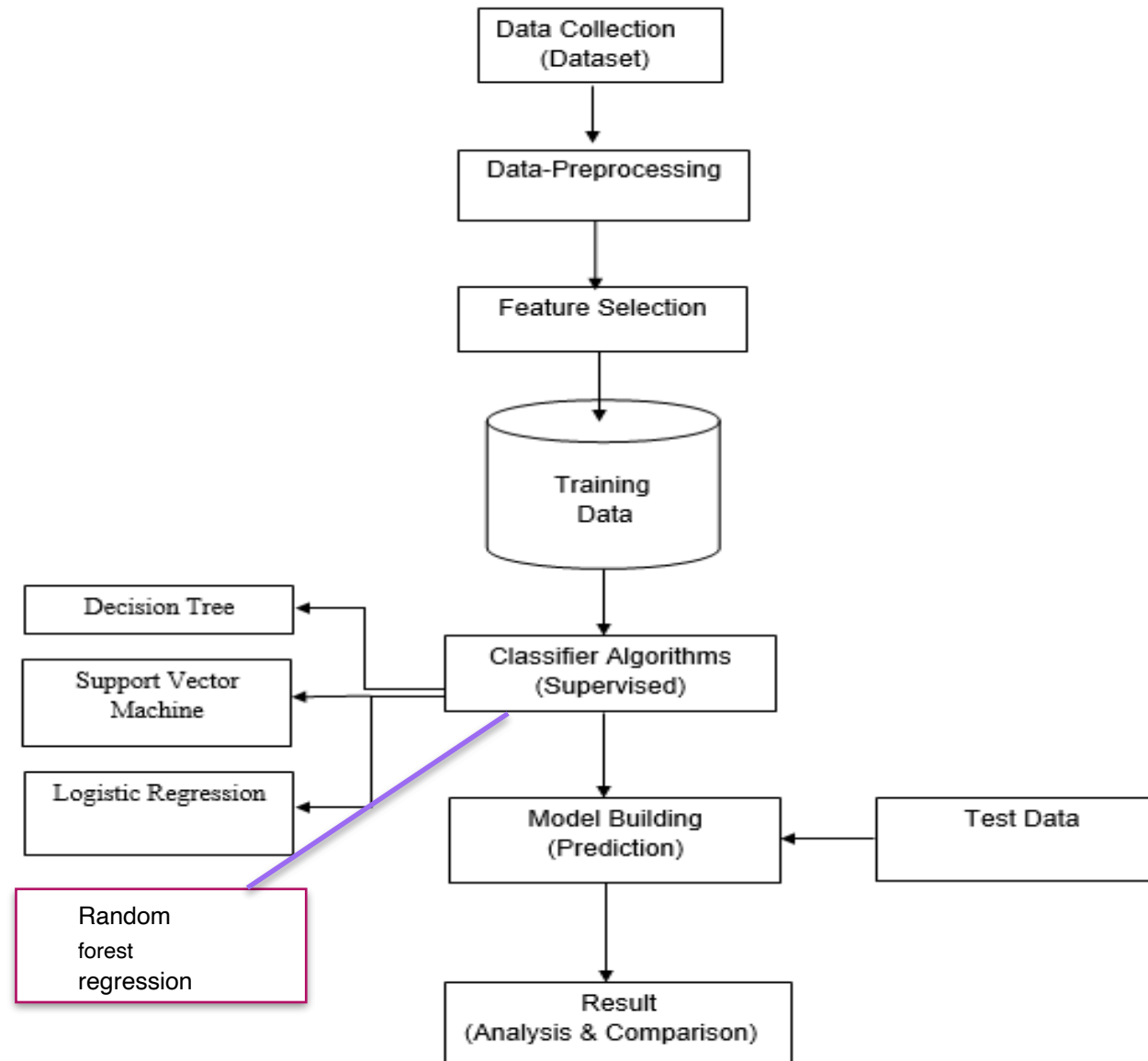
This paper shows that the model was trained using techniques like Decision Tree, Random Forest, and Multi-layer perceptron for prediction of stroke. With the slight differences, the obtained accuracies for the three methods were quite close. For Decision Tree, calculated accuracy was 74.31%, for Random Forest it was 74.53%, and for Multi-layer perceptron was 75.02%. According to this paper, Multi-layer perceptron is more accurate than the other two methods. Accuracy score was the only metric used for calculating the performance which might not always give favorable and accurate results.

2.3 **“ Prediction of Stroke using Data Mining Classification Techniques: International Journal of Advanced Computer Science and Applications (IJACSA) (2018)”**- Ohoud Almadani, Riyadh Alshammari

In this paper, the authors have used different data mining classification algorithms and techniques for prediction of the possibility of a stroke. The dataset in the paper was taken from the Ministry of National Guards Health Affairs Hospitals, Kingdom of Saudi Arabia. The three classification algorithms used were JRIP, C.5, and Multi layers perceptron. With the help of these algorithms, the model obtained an accuracy of around 95%. Though the paper claims to obtain an accuracy of 95%, the time taken for training and predicting is much high as the authors have used complex algorithms for implementation purpose.

Proposed system methodology

Proposed System Flow Diagram



Stroke Dataset from Kaggle

Attribute Name	Type (Values)	Description
1. id	Integer	A unique integer value for patients
2. gender	String literal (Male, Female, Other)	Tells the gender of the patient
3. age	Integer	Age of the Patient
4. hypertension	Integer (1, 0)	Tells whether the patient has hypertension or not
5. heart_disease	Integer (1, 0)	Tells whether the patient has heart disease or not
6. ever_married	String literal (Yes, No)	It tells whether the patient is married or not
7. work_type	String literal (children, Govt_job, Never_worked, Private, Self-employed)	It gives different categories for work
8. Residence_type	String literal (Urban, Rural)	The patient's residence type is stored
9. avg_glucose_level	Floating point number	Gives the value of average glucose level in blood
10. bmi	Floating point number	Gives the value of the patient's Body Mass Index
11. smoking_status	String literal (formerly smoked, never smoked, smokes, unknown)	It gives the smoking status of the patient
12. stroke	Integer (1, 0)	Output column that gives the stroke status

Algorithms

- u Logistic Regression
- u Support Vector Machine
- u Decision tree classifier
- u Random Forest

Demonstration

[https://colab.research.google.com/drive/
13vFUqgVrzUUjtHWLj0aRbPy5YI99HuGE?usp=sharing#scrollTo=connected-teens](https://colab.research.google.com/drive/13vFUqgVrzUUjtHWLj0aRbPy5YI99HuGE?usp=sharing#scrollTo=connected-teens)

Results

	Accuracy	Precision score	Recall score	F1 score
Logistic regression	78%	0.15%	0.72%	0.25%
Support vector machine	72%	0.13%	0.76%	0.22%
Decision tree classifier	89%	0.16%	0.22%	0.19%
Random Forest Regression	94%	0.29%	0.04%	0.07%

Conclusion

- u Stroke is a critical medical condition that should be treated before it becomes critical. Building an effective machine learning model can definitely help in the early prediction of stroke and reduce the severe impact on the future.
- u In this project, we showed the performance of various machine learning algorithms for successfully predicting the stroke based on multiple physiological attributes. Out of all the algorithms chosen, performance of Decision tree classification was best with an accuracy of 89%. Among all the precision, recall and F1 scores obtained, Decision tree has performed better.

References

List all the material used from various sources for making this seminar.

- u Prediction of Brain Stroke Severity Using Machine Learning- Bandi V, Bhattacharyya D, Midhunchakkravarthy D. In: International Information and Engineering Technology Association, (2020).
- u Predicting Stroke from Electronic Health Records- Nwosu, C.S., Dev S., Bhardwaj P., Veeravalli B., John D. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE (2019).
- u Prediction of Stroke using Data Mining Classification Techniques- Almadani O, Alshammari R. In: International Journal of Advanced Computer Science and Applications (IJACSA) (2018).
- u Probability of stroke: a risk profile from the Framingham Study- Wolf P, Ralph B, Belanger A, Kannel W. Journal of American Heart Association(2017)
- u Development of an Algorithm for Stroke Prediction: A National Health Insurance Database Study in Korea- Seung Nam Min a, Se Jin Park b,Dong Joon Kim c, Murali Subramaniyam d , Kyung-Sun Lee. Eur Neural (2018).
- u Dataset named 'Stroke Prediction Dataset' from Kaggle:
- u <https://www.kaggle.com/ahmtcnbs/stroke-prediction-xgboost-97/data>



Thank You!