ISSN No:-2456-2165

Stroke prediction system using ANN (Artificial Neural Network)

Sai Chandan P Reddy
Electronics and instrumentation engineering
Ms ramaih Institute of technology
Bangalore,India

Abstract:- Stroke is the second leading cause of mortality worldwide, and it continues to be a huge health burden for both individuals and national healthcare systems. Hypertension, cardiac disease, diabetes, dysregulation of glucose metabolism, atrial fibrillation, and lifestyle factors are all potentially modifiable risk factors for stroke. Stroke is a life-threatening medical condition. When blood flow to a portion of your brain is halted or diminished, brain tissue is deprived of oxygen and nutrients, resulting in a stroke. Within minutes, brain cells begin to die. The authors aimed to derive a model equation for developing a stroke pre-diagnosis algorithm with the potentially modifiable risk factors. Ischemic embolic and haemorrhagic strokes account for the bulk of strokes. When a blood clot forms far away from the patient's brain, usually in the heart, it travels through the circulation and lodges in the patient's smaller brain arteries. Haemorrhagic stroke is another type of brain stroke that happens when a blood vessel in the brain ruptures or spills blood. Stroke is the world's second leading cause of death and one of the leading causes of death in persons over the age of 65[1]. By the method proposed, it would be able to mitigate the stroke by 99 percent, which is almost most of the time.

Keywords:- *Stroke prediction system, Artificial neural network, deep learning, prediction system, accuracy*.

I. INTRODUCTION

Causes of stroke mortality—Stroke mortality is caused by co-morbidities and/or complications. Stroke complications might manifest themselves at any time. The first week following the development of stroke symptoms and the first month following the commencement of the stroke is the most important time for survival, with the largest incidence of fatalities. Hyperglycemia, hypoglycemia, hypertension, hypotension, fever, infarct extension or rebreeding, cerebral edoema, herniation, coning, aspiration pneumonia, urinary tract infection, cardiac dysrhythmia and deep venous thrombosis are all complications of stroke.[2]Stroke's effect is predicted to expand, with five million strokes and heart Disease-related deaths expected in 2020, up from three million in 1998. This will be attributed to continuous health and demographic changes, which will result in a rise in vascular disease risk factors and the old population's age. As a result, the suggested method would be revolutionary in detecting a stroke in a patient at an early stage with much greater accuracy, which would be taken into consideration by all medical institutions for a better rate of preventing strokes at both an early and late age.

II. LITERATURE OVERVIEW

There has been a good work done in the field of stroke prediction system So far in [3] we can find that using Machine Learning models to Decision Tree, Naive Bayes and Artificial Neural Network it has become much easier for doctors to predict if a patient is having symptoms for stroke. During the prediction a system works on the Architecture flow which is (age, gender, BMI, smoking status, heart disease, previous medical history …etc) ML Techniques (Decision Tree, Artificial Neural Network, Naïve Bayes) Analysis (Stroke Prediction) Management (Suggestions and Improvements). [4] studies in which stroke patient's movement is taken as the input data in the process for the prediction by the use of expensive sensors, and they hope to be affordable in future. Therefore, [5] explains the study that ML algorithms, particularly the deep neural network, can better long-term outcome prediction for stroke patient. [6] developed a feature selection technique for classifying epilepsy episodes. The University of Bonn in Germany contributed the EEG dataset for this study. Subsets were obtained using a hybrid feature selection method. In this study, the mRMR, Fisher, Chi-Square, and Relief -F were used to identify features

III. METHODOLOGY

A. Data Preparation

Data was acquired from the Health Care dataset of strokes, which has been used in several studies. Because the data is difficult to come by, the only way to run the model and make a forecast was to use data from a reliable source. As a result, the dataset was derived from the Health Care dataset on strokes. This dataset is used to forecast if a patient is likely to suffer a stroke based on input parameters such as gender, age, and a variety of diseases and smoking status. The filtering strategy is used to pick a subset of the original train data for Machine Learning and Data Visualization applications.

B. Data Preprocessing

Records, points, vectors, patterns, occurrences, instances, samples, observations, or entities make up a dataset. A collection of characteristics that define data items capture the essential attributes of an item, such as the mass of a physical object or the time at which an event occurred. So, to begin increasing the dataset quality, the first step would be to Encode Categorical Features, which indicates that the dataset contains categorical values that must be encoded before the model can be trained and evaluated. The next step is to scale the variation in the features, since feature scaling is a means of standardizing the independent characteristics included in a data set in a specific range, in this instance BMI, age, and average glucose level (Figure 1). Finally, the ID feature should

ISSN No:-2456-2165

be removed since the number of entries isn't necessary, and detecting the null value as null poses an issue because it reduces the model's accuracy.

	id	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
count	5110.000000	5110.000000	5110.000000	5110.000000	5110.000000	4909.000000	5110.000000
mean	36517.829354	43.226614	0.097456	0.054012	106.147677	28.893237	0.048728
std	21161.721625	22.612647	0.296607	0.226063	45.283560	7.854067	0.215320
min	67.000000	0.080000	0.000000	0.000000	55.120000	10.300000	0.000000
25%	17741.250000	25.000000	0.000000	0.000000	77.245000	23.500000	0.000000
50%	36932.000000	45.000000	0.000000	0.000000	91.885000	28.100000	0.000000
75%	54682.000000	61.000000	0.000000	0.000000	114.090000	33.100000	0.000000
max	72940.000000	82.000000	1.000000	1.000000	271.740000	97.600000	1.000000

Figure 1: Data Preprocessing Overview

C. Feature Visualisation

Records, points, vectors, patterns, occurrences, instances, samples, observations, or entities make up a dataset. A collection of characteristics that define data items capture the essential attributes of an item, such as the mass of a physical object or the time at which an event occurred. So, to begin increasing the dataset quality, the first step would be to Encode Categorical Features, which indicates that the dataset contains categorical values that must be encoded before the model can be trained and evaluated. The next step is to scale the variation in the features, since feature scaling is a means of standardising the independent features in a data set in a specific range. A visualisation of urban and rural is shown to see how the dataset is getting related to each other. And finally a heat map correlation of all the gestures is shown which provides an idea of the features/attributes of the dataset.

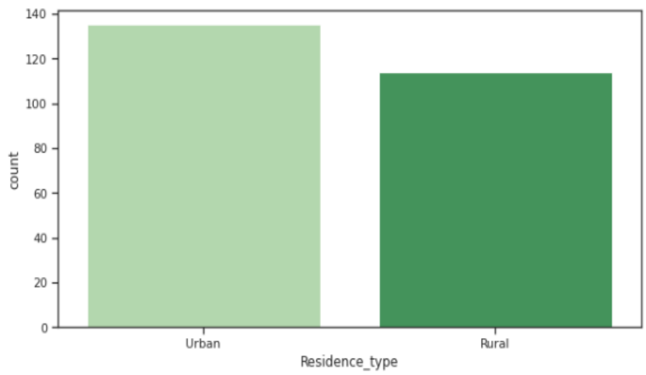

Figure 2: Residency type visualization

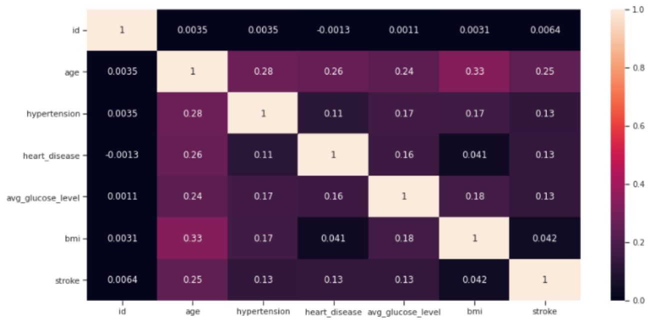

Figure 3: Heatmap Correlation

D. Model Architecture

The Artificial Neural Network (ANN) is made up of the structure and function of a biological neural network, which is the major reason for the creation of the Artificial Neural Network architecture. It is made up of neurons arranged in layers and neurons that are comparable to those found in the human brain. The input signal is absorbed by the processing unit, while the output layer provides network output. In these linked configurations, like in all network topologies, the input and output layers are always present. The Hidden layer, which keeps neurons out of both the input and output layers, is the third layer. These neurons are hidden from anyone engaging with the system, operating as a black box. By adding additional hidden layers containing neurons, the system's computing and processing capabilities may be improved, but the system's training phenomena grow more complex at the same time. The authors use TensorFlow to execute the ANN, and we'll start by utilizing the sequential function to initialize the ANN. Then we'd add levels to it, starting with the input layer, which would have the'relu' activation function. The activation function is used to determine if the output of a neural network is yes or no. It transforms values from -1 to 1 or 0 to 1, for example. We'll repeat the method for the four hidden layers, and then use the sigmoid as the activation function for the output layer. The ANN is then compiled, with Adam optimizer as the loss function, binary cross-entropy as the loss function, and accuracy as the metrics.

IV. EXPERIMENTAL RESULTS

Now, as we progressed from data collection through preprocessing to feature selection to model construction and application, the author approach was applied in the algorithm proposed and the steps included to get the results. The author used 100 epochs for training the dataset on the model, if the epochs are less, then it means the model has not been trained well, and we need to train the model properly to get the best accuracy than the other algorithms. When the epochs are 25 we are getting an accuracy of 75 percent, gradually increasing it we are getting the accuracy of 89 percent at 50 epochs and finally increasing it to 100 we are getting an accuracy of 99 percent which is a tremendous accuracy as it predicts cancer 99 out of 100 times, which is a great boost to the health business.

V. CONCLUSION

The stroke prediction decision support system supports and guides clinicians in making the finest, most accurate, and quickest judgments possible while also minimizing total treatment costs. The proposed strategy reduces treatment costs and improves quality of life by predicting strokes at an early stage. We were able to achieve a whopping accuracy of 99 percent on a particular dataset by employing Artificial Neural Networks, which is a fantastic result in terms of science and creativity and will help us have fewer patients die from strokes and necessary help would be provided at an early stage.

REFERENCES

[1.] Akash, Kunder & Shashank, H & .S, Srikanth & A.M, Thejas. (2020). Prediction of Stroke Using Machine Learning.

[2.] “What is Alzheimer's Disease? Symptoms & Causes |alz.org,” Alzheimer’s Association.[Online].

[3.] S. Levy-Tzedek, "A Machine-Learning Model for Automatic Detection".

[4.] M. K. AKASH, S. H. N, S. S and T. A. M, “Stroke Prediction Using Machine Learning,” vol. 1, 2020.

[5.] M. JoonNyung Heo, M. Jihoon G. Yoon, M. Hyungjong Park and M. P. Young Dae Kim, "Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke".

[6.] Farrikh Alzami, Juan tang et al.,”Adaptive hybrid feature selection-based classifier ensemble for epileptic seizure classification,” IEEE Access, vol. 6, pp. 29132-29145, 2018.

[7.] T. Kansadub, S. Thammaboosadee, S. Kiattisin and C. Jalayondeja, "Stroke risk prediction model based on demographic data," 2015 8th Biomedical Engineering International Conference (BMEiCON), 2015, pp. 1-3, doi: 10.1109/BMEiCON.2015.7399556.

[8.] Indarto, E. Utami and S. Raharjo, "Mortality Prediction Using Data Mining Classification Techniques in Patients With Hemorrhagic Stroke," 2020 8th International Conference on Cyber and IT Service Management (CITSM), 2020, pp. 1-5, doi: 10.1109/CITSM50537.2020.9268802.

[9.] C. -H. Lien, F. -H. Wu, P. -C. Chan, C. -M. Tseng, H. -H. Lin and Y. -F. Chen, "Readmission Prediction for Patients with Ischemic Stroke after Discharge," 2020 International Symposium on Computer, Consumer and Control (IS3C), 2020, pp. 45-48, doi: 10.1109/IS3C50286.2020.00019.

[10.] J. Chen, Y. Chen, J. Li, J. Wang, Z. Lin and A. K. Nandi, "Stroke Risk Prediction with Hybrid Deep Transfer Learning Framework," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2021.3088750.

[11.] J. Cho, Z. Hu and M. Sartipi, "Post-stroke discharge disposition prediction using deep learning," SoutheastCon 2017, 2017, pp. 1-2, doi: 10.1109/SECON.2017.7925299.

[12.] E. Zamsa, "Medical software user interfaces, stroke MD application design," 2015 E-Health and Bioengineering Conference (EHB), 2015, pp. 1-4, doi: 10.1109/EHB.2015.7391403.

[13.] Le Zheng et al., "Risk prediction of stroke: A prospective statewide study on patients in Maine," 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015, pp. 853-855, doi: 10.1109/BIBM.2015.7359796.