



Stroke Predication Using Machine Learning

Dinesh C. Dobhal Associate Professor, Department of CS&E, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.,
dineshdobhal@gmail.com

Kumar Kartik Research Scholar, Department of CS&E, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.
kkartik2097@gmail.com

Prajwal Gupta Research Scholar Department of CS&E, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.
gssshivam216@gmail.com

Shweta Chauhan Assistant Professor, School of Management, Graphic Era Hill University, Dehradun, India.

Abstract : When the brain's blood flow is disrupted or decreased, a stroke happens. Because of a shortage of oxygen and nutrients in the brain, a stroke can result in the death of brain cells. By evaluating the effectiveness of predictive data mining approaches, a lot of effort has been done to predict different diseases. In this work, we assess various methods for predicting stroke in a dataset from a cardiovascular health (CHS) study. Hemorrhagic and ischemic strokes make up the majority of cases. An ischemic embolism stroke typically occurs when a clot in bloodstream travels from the heart through the bloodstream to a smaller artery in the brain. Because hemorrhagic stroke happens when an artery in the brain leaks or ruptures, it is regarded as a different type of stroke. Stroke is one of the most serious diseases for persons over 65 and the second leading cause of mortality in the globe. It harms the brain in the same way that a "heart attack" harms the heart. A stroke can result in death as well as costly medical care and long-term impairment. A stroke claims the life of one person every four minutes, however up to 80% of strokes can be avoided with early detection or prediction.

Here, different machine learning classification approaches are used to create a classification model after dimensionality reduction using a principal component analysis algorithm and feature selection using a decision tree algorithm.

An ideal stroke predictions model with an accuracy of

95.27 was produced after studying and contrasting the effectiveness of categorization using various methods and the accuracy of the variant model.

Keywords: Stroke, Cardiovascular health, Oxygen, Dataset, Model, Blood, Dataset.

1. Introduction : An extremely serious medical emergency is a stroke. It occurs when a brain blood artery rupture or, more frequently, when there is a blockage [1]. If untreated, brain cells swiftly start to degrade. Stroke can cause fatalities or severe disabilities. stroke signs, a high intensity headache with no known reason, such as dizziness, walking or balance issues, confusion, speech difficulties, or difficulty understanding others. Weakness or numbness in the body, especially on one side. Changes in vision in one or both eyes, difficulty in eating etc. The brain damage increases in direct proportion to time. This is because brain cells begin to degenerate due to oxygen deprivation within minutes [2]. The portions of the body that the brain regulates stop functioning once the brain tissue has died. As a result, the major cause of organ failure is stroke. Drugs that can stop blood clots and reduce brain damage typically need to be administered for a brief period of time, usually within three hours of the onset of symptoms. Diagnose, you can begin the examination in the ambulance. Imaging tests like a CT scan, an MRI, or an ultrasound will be performed on you once you enter the emergency room. There may be further tests, such as an ECG and an EEG (test of heart electrical activity) (test of electrical activity of the brain) [3].

Strokes are the second-leading causes of death and the third-leading cause of morbidity worldwide. An artery leading to the brain can get blocked or burst, cutting off blood supply and causing certain brain cells to suddenly die from a lack of oxygen. This disease is known as a stroke. It also contributes significantly to depression and dementia [4].

Globally, low- and middle-income nations account for 87% of stroke-related fatalities and disability-adjusted life expectancy, as well as 70% of all strokes [5]. Over the past 40 years, the number of stroke cases has doubled in low- and middle-income nations. Over the previous ten years, stroke rates have fallen by 42% in high-income nations. Persons who live in low- and middle-income nations have stroke 15 years earlier on average and die from it more frequently than people who live in high-income countries [6].

Strokes typically affect persons who are at the height of their productivity. The country's developing socioeconomic issue has received little attention thus far, despite having a significant impact on it. The risk factors for coronary heart disease, stroke, and other vascular illnesses are fairly similar. Addressing important modifiable factors including diabetes, high blood

pressure, and excessive cholesterol levels are all part of effective prevention measures. Additionally, it can reduce or even eliminate the hazards brought on by bad eating habits, smoking, inactivity, and abdominal obesity. Even in certain low-income nations, this combination of preventative measures has been successful in lowering stroke mortality [7].

The most effective treatment for unidentified forms of stroke in low- and middle-income countries is still unknown because the majority of guidelines are based on evidence from high-income nations. For instance, 34% of strokes in low- and middle-income nations (vs. 9% in high-income countries) are hemorrhagic subtypes, as are up to 84% of stroke patients in these regions (vs. 9% in high-income regions). Within three years of diagnosis, 16% of people in high-income nations pass away. A course of treatment based on an ischemic stroke (vs hemorrhagic stroke) diagnosis determined by computed tomography is recommended by current guidelines for the management of acute stroke (CT) [8].

In environments with limited resources, doctors must make tough clinical decisions such as whether to administer anticoagulants to patients and how closely to monitor blood pressure without the ability to distinguish between ischemia and ischemic events. The disproportionate burden of stroke in low- and middle-income countries may be explained by these treatment difficulties, insufficient rehabilitation services, a lack of preventative efforts, and a lack of knowledge of the particular risk factors linked with stroke in these nations. It is unknown what causes the early beginning of infancy, the prevalence of hemorrhagic subtypes, and the high mortality [9].

In low- and middle-income nations, there is a need to better understand the potential intrinsic risk factors for this epidemic. Important risk factors for stroke incidence, subtypes, and outcomes in people of African heritage are examined by the Stroke Research and Education Network [10]. Through a four-step process of monitoring, prevention, acute treatment, and rehabilitation, it is possible to direct targeted prevention efforts by understanding the genetic basis for the interconnections between risk variables [11].

With the help of this integrated approach for Type, low- and middle-income nations may create guidelines for the prevention, diagnosis, and rehabilitation of stroke. Considerations for formulating guidelines for treating patients with acute stroke of unknown aetiology in the presence of a CT scanner are provided for resource-constrained settings [12]. In his talk, Berkowitz emphasised the effectiveness of supportive care, including euthanasia and normoglycemia for stroke patients, early activation, and quick seizure therapy. He offered blood pressure guidelines for stroke patients in such circumstances and advised against excessive aspirin use [13].

It was also stressed how important secondary prevention is. The “World Health

Organization (WHO)", the "World

Stroke Organization", and the "World Federation of Neurosurgeons" must work together to relaunch the original WHO global stroke initiative in order to manage acute stroke in resource-constrained settings, improve surveillance data, and direct effective prevention and treatment. Protocols for lowering cardiovascular risk and preventing stroke are included in the WHO package for core interventions in non-communicable illnesses for primary care in resource-constrained settings [14].

In addition to expanding educational initiatives for stroke prevention, treatment, and rehabilitation through partnerships, WHO intends to create recommendations for the management of acute stroke in low- and middle- income countries.

2. Research Methodology and Experimental setup : The steps for making a prediction using machine learning algorithms are as follows. .

Step 1: Importing the necessary Libraries and loading the dataset.

Step 2: Checking for Null Values during exploratory data analysis.

Step 3: Checking missing values using "Bar Plot" (Ref Fig 1.1).

Step 4: Using the KNNImputer algorithm to address the missing values .

Step 5: Visualizing the distribution of stroke using countplot and pie plot (Ref Fig 1.2).

Step 6: Analyzing normal distribution of the data (Ref Fig 1.3).

Step 7: visualizing countplot for hypertension and heartdisease w.r.t stroke (Ref Fig 1.4).

Step 8: Handling categorical Data, converting categorical features into numerical using dummy.

Step 9: Dropping unnecessary features which will not affect the efficiency of the prediction model.

Step 10: Plotting Correlation (Ref Fig 1.5).

Step 11: Using StandardScaler to scale the train and test sets into scaled versions.

Step 12: Splitting the dataset for Training and Testing datasets

Step 13: Modeling and prediction using Logistic Regression (Ref Fig 1.6).

Step 15: Modeling and prediction using Logistic Regression with SMOTE (Synthetic Minority Oversampling Technique), (Ref fig 1.7).

Step 16: Modeling and prediction using Support Vector Classifier, (Ref Fig 1.8).

Step 17: Using GridSearchCV to find the best parameter for SVC and modeling and prediction using SVC (Ref Fig 2.1; 2.2).

3. Results analysis and Future Scope:

The results obtained during study are presented in following table

Table1 : Accuracy of ML algorithms

S.No	Model	Accuracy of the model
1	Linear Regression	94.4227
2	SMOTE	79.3911
3	Support Vector Classifier	86.5076
4	GridSearchCV	95.2694