Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 2, pp. 261–271
DOI : 10.47974/JSMS-1252

Several implications for forecasting a brain attack utilizing mathematical techniques

Smita R. Pidurkar *
Department of Mathematics
G H Raisoni College of Engineering
Nagpur
Maharashtra
India

Seema Raut †
Department of Mathematics
G. H. Raisoni Institute of Engineering and Technology
Nagpur
Maharashtra
India

Abstract

The human brain is unquestionably the most complex organ in the body and is acknowledged as the body's control station. The wellness of this control station is depends on several human related factors. The uncontrolled behavior of this factors may cause brain stroke. In this study, survey data from 24,400 brain stroke patients are considered, along with important possible confounders such as age, gender, hypertension, heart disease, sugar, BMI, and smoking. Initially the association with this parameters with stroke were tested. Regression analysis was then performed to produce the projected model based on the data and its accuracy was tested.

Subject Classification: *68Q32.*

Keywords: *Logistics regression, True positive, False positive, Accuracy, Precision, Recall.*

* *E-mail:* **smita.pidurkar@raisoni.net (Corresponding Author)**
† *E-mail:* **seema.raut@raisoni.net**

1. Introduction

The WHO estimates that low- and middle-income nations experience eleven million brain strokes annually. This causes four million fatalities annually and leaves around 30% of survivors with significant disability. The 70% of survivors who fully recover from a stroke have a significantly increased risk of getting another one [1].

Strokes fall into two primary groups.1) Hemorrhagic stroke 2) Ischemic stroke. A ruptured blood vessel that results in bleeding into the brain is the cause of hemorrhagic stroke. An ischemic stroke occurs when a portion of the blood supply to the brain is restricted or obstructed, depriving the brain's tissue of oxygen and nutrients. Transient ischemic attacks, which are brief stops in blood supply to the brain, can happen to some patients (TIA).

Most typical signs of a stroke are vision loss or fading in one or both eyes, face Numbness / weakness in one side of the body's arm or leg ,an inability to hear or talk to other individuals, an unexpected, severe headache with no apparent cause , loss of balance or an unsteady gait.

Brain stroke is associated with a number of risk factors, including obesity and being overweight, using illicit drugs, binge drinking or drinking excessively [2], elevated blood pressure, smoking cigarettes, elevated lipids, Cardiovascular illness, cardiac defects, infections of the heart, or irregular heartbeat, COVID-19 infection, age, sex, familial or personal history of transient ischemic attack, heart attack [3].

2. Methodology

In this study's data from 24400 patients are used to prediction model, which takes into account the following key risk variables for stroke. Gender (z_1), Age (z_2), Hypertension (z_3), Heart diseases (z_4), Body Mass Index (z_5), Diabetics (z_6) and Smoking (z_7). These variables are expressed in numbers or nominal form. Age, body mass index, and diabetes are recorded in numerical form, while information on the parameters of gender, hypertension, heart disease, and smoking is obtained in nominal form. Here dependent variable consider as stroke which is dichotomous, if the patient has had a stroke, the dependent variable can be coded as y=1; otherwise, it can be written as y=0.i.e.

$$y = \begin{cases} 1, & \text{Patient have a stroke} \\ 0, & \text{Patient did not have a stroke} \end{cases}$$

This output is based on following parameters defined below:

i) Gender (z_1): Compared to men,women are more likely to die from stroke[4-7]. z_1 is specified as 1 for males and 0 for females in the prediction model.

ii) Age (z_2): Vasodilatation, or the narrowing of blood arteries that provide oxygen to the brain, has compounding consequences with age. Additionally, a lack of activity can compromise cardiovascular health and raise the risk of blood clots, which can result in a heart attack or stroke. Age, unfavorable lifestyle choices, and lack of exercise can cause blood vessels to thicken and stiffen, which can impede blood flow throughout the body. Atherosclerosis, a sort of arteriosclerosis, is caused by the fatty deposits substances and cellular debris. These could do so partially and result in a blood clot that also stops blood vessels, or they can completely or partially block blood vessels [8-10].

iii) Hypertension (z_3): One of the most important things you can do to reduce your chance of having a stroke is to manage your hypertension. A second TIA or stroke can be avoided by decreasing blood pressure. [11-12].In the model, (z_3) is defined as 1 for patients with hypertension and 0 for those without.

iv) Heart Diseases (z_4): Arrhythmia may develop, depending on which part of the heart is damaged. This often stops blood flow to the brain and causes a stroke. Here, the parameter (z_4) is shown in nominal form, with values of 0 and 1 corresponding to having and not having heart disease, respectively.

v) Body Mass index (z_5): For both men and women, obesity and being overweight increased the risk of ischemic stroke and stroke. Additionally, hemorrhagic stroke risk was higher in men who were underweight, overweight, or obese. [13-17]

vi) Diabetes (z_6): Diabetes raises your chance of a stroke, which can damage your brain and leave you disabled or even dead. Individuals with diabetes are twice as likely to get a stroke compared to those who not. Additionally, diabetes might hinder your body's ability to recover from a stroke. [18].

vii) Smoking (z_7): It was discovered that smoking tobacco is directly responsible for 25 percent of all strokes. While utilizing nicotine replacement products (like as gum, patches, or e-cigarettes) might raise the risk of stroke, smoking conventional cigarettes still wreaks havoc on the respiratory system. As a vasoconstrictor, nicotine increases the risk of blood clots. For both smokers and non-smokers exposed to second hand smoke, smoking increases the risk of stroke. The risk of stroke is reduced by giving up smoking [19]-[20]. The stroke parameter is defined as follows: 0 for no smoking habit and 1 for a smoking habit.

The likelihood that the patient would experience a stroke was calculated using logistic regression by the supplied set of values for the selected independent factors.

The predicted result for multiple regression analysis is then E(y) given by (1), which stands for the probability that the subject would experience a brain attack.

$$E(y) = b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + b_4 z_4 + b_5 z_5 + b_6 z_6 + b_7 z_7 \tag{1}$$

$$E(y) = P(y = 1 z_1, z_2, z_3, z_4, z_5, z_6, z_7) = \frac{e^{b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + b_4 z_4 + b_5 z_5 + b_6 z_6 + b_7 z_7}}{1 + e^{b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + b_4 z_4 + b_5 z_5 + b_6 z_6 + b_7 z_7}} \tag{2}$$

The Chi test with a 0.05 significance level was used to examine the relationship between stroke and gender, age, hypertension, heart disease,

Table 1

Stroke with gender

Observed Cases:			
Stroke Count	**Column Labels**		**Grand Total**
Row Labels	**0**	**1**	
0	5419	8587	14006
1	5583	4811	10394
Total	11002	13398	24400
Expected cases:			
0	6315.33	7691	14006
1	4686.67	5707	10394
Total	11002	13398	24400
Chi square = 2.7131E-120 < alpha		Reject null hypothesis	

Table 2

Stroke with Hypertension

Observed Cases:			
Stroke Count	**Column Labels**		**Grand Total**
Row Labels	**0**	**1**	
0	13399	607	14006
1	4860	5534	10394
Total	18259	6141	24400
Expected cases:			
0	10481	3525.04	14006
1	7778.04	2615.97	10394
Total	18259	6141	24400
Chi square < alpha		Reject null hypothesis	

and smoking habit. The alternative hypothesis asserts that there is a relationship between stroke and the designated independent variable, while the null hypothesis states that there is no linkage between the two. And if the p-value is less than 0.05, the hypothesis is considered to be false. Otherwise, it will be approved. The following Table 1-4 shows the association of stroke with independent variables.

Table 3

Stroke with Heart Disease

Observed Cases:			
Stroke Count	**Column Labels**		**Grand Total**
Row Labels	**0**	**1**	
0	13709	297	14006
1	6867	3527	10394
Grand Total	20576	3824	24400
Expected cases			
0	11811	599	14006
1	8765	1629	10394
Grand Total	20576	3824	24400
Chi square < alpla		Reject null hypothesis	

Table 4

Stroke with Smoking

Observed Cases:			
Stroke Count	**Column Labels**		**Grand Total**
Row Labels	**0**	**1**	
0	7704	6302	14006
1	4546	5848	10394
Grand Total	12250	12150	24400
Expected Cases			
0	7032	6974	14006
1	5218	5176	10394
Grand Total	12250	12150	24400
Chi Square = 7.20569E-68 < alpha		Reject null hypothesis	

Since the null hypothesis is consistently rejected, the chi square value for each table shows that there is a relationship between the stroke and the specified independent variables. The strength of the correlation of the stroke with Glucose, BMI and Age can be also found by using correlation. The Table (5) shown the correlation between stroke and glucose.

Table 5

Correlation with Glucose

Stroke	No. of entries	Sum of entries	Mean	sum of all entries (y)	y^2	Standard Deviation	Correlation (R)
1	10394	1731633	166.5677	3101845	482570364	60.154039	0.565089
0	14006	1370378	97.84219				

With the rule for R values that if its value is in interval 0< 0.1, 0.1< 0.3, 0.3< 0.5, 0.5< 0.7, 0.7< 1 then Correlation's strength is Not correlated, Poorly Association, Average Correlation , Strong Correlation, Very strong association respectively ,it was found that the Stroke and Glucose is strongly correlated.

After finding association, the listed below terms are used to predict strokes:

i) Likelihood: The likelihood function shows the possibility that a random variable will be generated given a specific set of statistical parameter values.

ii) Odds: The odds ratio represents the likelihood that a patient will experience a stroke relative to their non-occurrence.

$$Odds = \frac{P(y=1z_1,z_2,z_3,z_4,z_5,z_6,z_7)}{P(y=0z_1,z_2,z_3,z_4,z_5,z_6,z_7)}$$

Probability that the person will have a stroke in terms of odd is given by

$$\text{Probability of stroke} = \frac{Odd}{1+Odd}$$

Using observations from 24400 patients and a cutoff value of 0.5, the projected mathematical model was determined to be

$$y = -5.79781 + 0.86257z_1 + 0.018353z_2 + 4.708578z_3 + 3.951188z_4 + 0.032222z_5 + 0.04262z_6 + 0.957562z_7 \quad (3)$$

2.1 *Testing of Model (3)*

The confusion matrix from (3) is obtained in Table (6)

Table 6

Confusion Matrix

		Predicted cases		
		0	1	Total
Actual cases	0	12925	1081	14006
	1	267	10127	10394
	Total	13192	11208	24400

Table 7

Accuracy, precision, and recall levels for various cutoff values

Cut off	Accuracy	Precision	Recall	Cut off	Accuracy	Precision	Recall
0	0.43	0.43	1	0.5	0.93	0.91	0.94
0.1	0.87	0.76	1	0.6	0.91	0.91	0.87
0.2	0.93	0.86	1	0.7	0.87	0.91	0.79
0.3	0.94	0.89	0.99	0.8	0.83	0.91	0.67
0.4	0.98	0.9	0.97	0.9	0.78	0.92	0.54

The model provided by equation (1) was determined to be 94.48% accurate. The Precision, which is the proportion of correctly predicted stroke cases to all anticipated stroke cases, was discovered to be 90.36%. The Recall measures the proportion of patients who were correctly predicted to have strokes to those who actually experienced them. This recall rating of 97.43% demonstrates how accurate the model is at identifying stroke patients.

It was discovered that cut off values affect the metrics Accuracy, Precision, and Recall. Below is a Table (7) that shows how Accuracy, Precision, and Recall relate to various cut-off levels.

According to a comparison shown in Figure 1 of these three metrics based on various cut off values, The accuracy rises with increase in cut off up to 0.4, at which point it decreases with increase in cut off, As the cut-off value increases, the precision also does while the recall value decreases.

With a cut-off of 0.4, it was found that the false positive (FP) and true positive (TP) rates were, respectively, 0.077181 and 0.9743. The TP rate gauges how many patients are properly projected to have a stroke

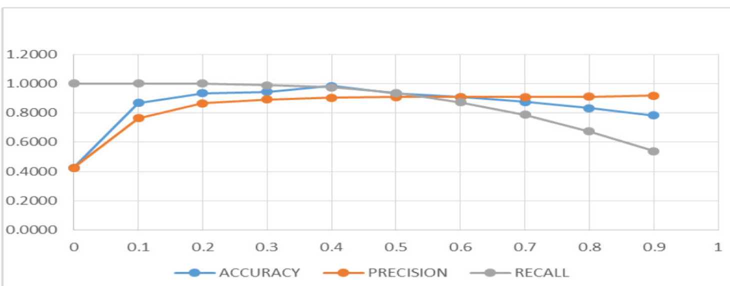

Figure 1

Relation between Accuracy, Precision, Recall for different cut off

compared to how many do, whereas the FP rate gauges how many patients are wrongly predicted to have a stroke compared to how many do not. Table (8) gives changes in TP and FP for various values of cut off and the graph in Figure (2) illustrates the relationship between true positive and false positive results.

Table 8

TP and FP values for different values of cut-off.

CUT OFF	FP	TP
	0.04	0.54
0	1	1
0.1	0.23	1
0.2	0.12	1
0.3	0.09	0.99
0.4	0.08	0.97
0.5	0.07	0.94
0.6	0.07	0.87
0.7	0.06	0.79
0.8	0.05	0.67
0.9	0.04	0.54

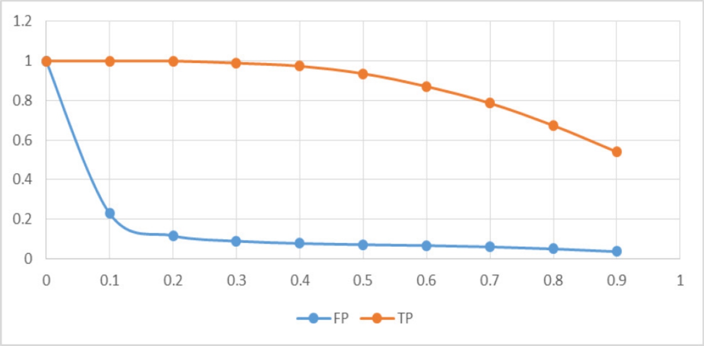

Figure 2

Region of Convergence

From the aforementioned factors of TP, FP, and accuracy, we determine the projected model's cut off value to be 0.4.

3. Results

The predicted model (1) of the stroke is created using logistic regression based on seven independent parameters that induce a stroke and whose data is in the form of nominal and numerical form. The model was evaluated using several cut-off values ranging from 0 to 1 for true positive and false positive rates, and it was ultimately determined the model has the maximum accuracy of 94.48% for a cut-off value of 0.4 for stroke prediction, according to the calculations. Moreover, the primary cause of stroke was shown to be hypertension, followed by heart disease.

4. Conclusion

On the basis of seven independent factors, including gender, age, hypertension, heart disease, body mass index, sugar, and smoking habit, it was shown that the aforementioned projected model provides close to 95% accurate prediction of stroke.

References

[1] https://www.who.int › Home › News › Detail.

[2] Daniel S, Bereczki D.Ideggyogy Sz.,Alcohol as a risk factor for hemorrhagic stroke. 57(7-8):247-56.PMID: 153304. (2004).

[3] Soliman RH, et al. Risk factors of acute ischemic stroke in patients presented to Beni-Suef University Hospital: prevalence and relation to stroke severity at presentation. Egypt doi: 10.1186/s41983-018-0012-4. Epub 2018 Apr 25. PMID: 29780228; PMCID: PMC5954772. (2018).

[4] Antonio González-Hermosillo J, et al. Committees and Investigators. Gender differences and management of stroke risk of nonvalvular atrial fibrillation in an upper middle-income country doi: 10.1016/j.ijcha.2018.12.017. PMID : 30705937; PMCID: PMC6349010. (2019).

[5] Vuković V,et al,Women and stroke: how much do women and men differ? A review--diagnostics, clinical differences, therapy and outcome, 2009 Sep; 33(3) : 977-84.PMID: 19860135 (2009).

[6] Bushnell CD,et al. Advances in Stroke: Stroke in Women. Stroke. 2022 Feb;53(2):605-607. doi: 10.1161/STROKEAHA.121.036975. Epub 2022 Jan 24. PMID: 35073191.https://doi.org/10.1155/2012/136398 (2022).

[7] Rexrode KM, et al. The Impact of Sex and Gender on Stroke. doi: 10.1161/CIRCRESAHA.121.319915. Epub 2022 Feb 17. PMID: 35175851; PMCID : PMC8890686. (2022).

[8] Yao XY, et al. Age- and gender-specific prevalence of risk factors in patients with first-ever ischemic stroke in china. Stroke. doi: 10.1155/2012/136398. Epub 2012 Jun 17. PMID: 22762013; PMCID: PMC3385649. (2012).

[9] Gu H et al.Age- and Sex-Associated Impacts of Body Mass Index on Stroke Type Risk: A 27-Year Prospective Cohort Study in a Low-Income Population in China. doi: 10.3389/fneur.2019.00456. PMID: 31118920; PMCID: PMC6504695. (2019).

[10] Xiaoshuang Xia, et al. Prevalence and risk factors of stroke in the elderly in Northern China: data from the *National Stroke Screening Survey Journal of Neurology* volume 266, pages 1449–1458 (2019).

[11] Lattanzi S, et al. Elevated blood pressure in the acute phase of stroke and the role of Angiotensin receptor blockers. PMID: 23431423; PMCID: PMC3574652.https://doi.org/10.1155/2013/941783. (2013).

[12] Blood pressure medications can lead to increased risk of stroke, May 29, 2015, Source:University of Alabama at Birmingham, https://www.sciencedaily.com/releases/2015/05/150529193554.htm (2015).

[13] Shiozawa M, et al. Association of Body Mass Index with Ischemic and Hemorrhagic Stroke. Nutrients. doi: 10.3390/nu13072343. PMID: 34371853; PMCID: PMC8308685. (2021).

[14] Kurth T, et al. Prospective study of body mass index and risk of stroke in apparently healthy women. doi: 10.1161/01.CIR.0000161822.83163.B6. PMID: 15837954. https://pubmed.ncbi.nlm.nih.gov/15837954/ (2005).

[15] Hu G, et al. Body mass index, waist circumference, and waist-hip ratio on the risk of total and type-specific stroke. doi: 10.1001/archinte.167.13.1420. PMID: 17620537. (2007).

[16] Yonemoto K, et al. Body mass index and stroke incidence in a Japanese community: the Hisayama study. doi: 10.1038/hr.2010.220. Epub 2010 Nov 25. PMID: 21107333. (2011).

[17] Shiozawa M, et al.. Association of Body Mass Index with Ischemic and Hemorrhagic Stroke. Nutrients. doi: 10.3390/nu13072343. PMID: 34371853; PMCID: PMC8308685. (2021).

[18] Chen R, et al. Diabetes and Stroke: Epidemiology, Pathophysiology, Pharmaceuticals and Outcomes. doi: 10.1016/j.amjms.2016.01.011. PMID: 27079344; PMCID: PMC5298897. (2016).

[19] Shah RS, Cole JW. Smoking and stroke: the more you smoke the more you stroke. doi: 10.1586/erc.10.56. PMID: 20602553; PMCID: PMC2928253. (2010)

[20] Pan B, et al. The relationship between smoking and stroke: A meta-analysis. Medicine (Baltimore).doi: 10.1097/MD.0000000000014872. (2019).

[21] https://www.kaggle.com/datasets.