A Glimpse to the Future: Identifying Stroke Risk Factors Using Data Visualization for Stroke Prediction

Manuscript received December 16, 2023; revised February 10, 2024

Marion Az-Zahra
Business and Management
Swiss German University
Jakarta, Indonesia
marion.azzahra@student.sgu.ac.id

***Abstract*— According to the World Health Organization (WHO), stroke stands as the second leading cause of death, accounting for 11 percent of total deaths in the world. This study identifies risk factors associated with stroke that could help predict the likelihood of a person getting a stroke by assessing input parameters such as gender, age, medical conditions related to stroke and cardiac disease, smoking status, body mass index, and average glucose level, among others. The results, presented with data visualization, showed that the above risk factors were significant in signifying stroke incidence.**

Keywords*— *data visualization, risk factors, stroke prediction

I. Introduction

Stroke is a global healthcare issue characterized by its familiar, severe, and disabling nature [1]. Across the majority of nations, stroke came out as the second leading cause of mortality and constitutes a primary factor contributing to acquired adult disability [2]. Stroke, which is caused by the abrupt death of some brain cells resulting from oxygen deprivation due to the loss of blood flow to the brain caused by arterial blockage or rupture, also proves to be a leading cause of both dementia and depression [3]. Notably, 70 percent of strokes and 87 percent of both stroke-related mortality and disability-adjusted life years transpire within low- and middle-income countries [4].

Given the increasing survival rates following the first incidence of stroke, the most significant health impact predominantly stems from the long-term consequences for patients and their families [5]. Despite significant advancements in the medical treatment of stroke, the absence of a universally applicable or efficient medical treatment leads to the continued reliance on rehabilitation for post-stroke care [6].

It is also worth noting that the risk factors associated with stroke are closely similar to those identified in coronary heart disease and various vascular conditions [7]. Effective preventive measures encompass targeting key modifiable factors, namely hypertension, high cholesterol levels, and diabetes [7]. Risks associated with lifestyle factors can also be identified, namely smoking status, lack of physical activity, poor eating habits, and abdominal obesity [8]. Knowing the fatality of stroke, it is essential to detect stroke likelihood in people starting as early as from their lifestyle factors. Thus, this study identifies risk factors associated with stroke to predict the possibility of people getting strokes using data visualization to present the data.

II. Literature Review

A developed prevention of stroke can be actualized once there is an increased awareness of stroke risk factors among the general populace [9]. Increasing knowledge regarding stroke warning symptoms among patients and their family members could lead to a faster duration between the first stroke incidence and the patient's arrival at the hospital, which results in improved treatment outcomes [9]. Further risk factors associated with stroke are provided in this section as follows:

A. Gender

Gender is among the essential identifiers for stroke incidence that could help predict which gender classification has the most cases of stroke. Prior research has conducted a study on stroke risk factors, with gender having exerted a significant influence on stroke risk and men being identified as having a more considerable relative risk [10]. However, a more recent study revealed that women are at a higher risk of getting a stroke in comparison to men, with women having a substantial number of risk factors for stroke [11], [12]. Women also have a higher likelihood of having a stroke compared to men, even at a young age [13]. To be more precise, women in the age range of 45 to 54 years old had a considerably higher likelihood of having had a stroke (odds ratio [OR] 2.39, 95% confidence interval [CI] 1.32—4.32) compared to men in the same age group [13].

B. Age

A study has proven that the likelihood of experiencing a stroke increases as one gets older, with the number of stroke incidences increasing every ten years after the age of 45, and more than 70 percent of all stroke incidences happen in people who are 65 years of age or older [14]. Another prior study also found that age is an independent risk factor for stroke, and the chance of developing stroke rises significantly between the ages of 55 and 59 [15]. In this age range, the risk for men increases from 5.9 percent to 11 percent and improves even further, reaching an incidence of 22.3 percent once they get 80 to 84 years old [16]. On the other hand, in the same age range, the risk for women increases from 3 percent to 7.2 percent but increases even further, reaching an incidence of 23.9 percent once they are older, in the age range of 80 to 84 years old [16].

C. Hypertension

Hypertension, the most pervasive risk factor for stroke, increases as people get older and has a 90 percent chance of

 EISSN: 2961-872X ● PISSN: 2961-8924 ● DOI: 10.56741/esl.v3i01.470

incidence during the lifetime of individuals who live 80 years and more [17]. Past research revealed that blood pressure raises the likelihood of getting a stroke 1.7 times more in men and 1.9 times more in women, per standard deviation increment beginning in middle age [18]. Additionally, prior research found that there was a stronger correlation between the morning blood pressure in older hypertensive patients and the risk of stroke compared to the blood pressure measured over 24 hours [19].

D. Cardiac Disease

Coronary heart disease and other vascular conditions are commonly caused by plaque accumulation in the arteries, blocking the oxygenated blood flow to the brain. The incidence of coronary heart disease and stroke has proven to be closely linked to hypertension and atherosclerosis, both of which are significantly influenced by risk factors, lifestyle choices, and environmental conditions [20]. Previous research also supported this statement that stroke risk factors are aligned with that of coronary heart disease and other vascular diseases [7]. It is precisely for this reason that the cardiac disease history of patients should be taken into account as the most prevalent risk factor for stroke.

E. Marital Status

Marital status serves as a significant social component that impacts the outcomes of acute myocardial infarction (AMI) and stroke [21]. Additionally, prior research has identified that being married or in a committed relationship has been linked to reduced mortality rates and increased survivability without complications despite having a history of heart disease or stroke [22].

F. Occupation

Stroke incidences in employed individuals often lead to early retirement due to stroke-related disability [23]. Although stroke incidences are less common in working populations in comparison to the general population, they still inflict a substantial socio-economic impact [23]. A recent study found that sitting occupations lead to an increased risk of experiencing a stroke [24]. This is because lack of physical activity has been identified as the second major risk factor for stroke, following hypertension [25]. It was also indicated that engaging in both occupational activity and free-time physical activity is correlated to a reduced risk of cardiovascular and all-cause death, regardless of the level of occupational activity levels [26].

G. Average Glucose Level

While high glucose level mainly affects individuals diagnosed with diabetes, there are still those who have no prior knowledge of their glucose levels, which can be fatal if not acknowledged [27]. However, there exists a suggestion that diabetes could increase the likelihood of getting a stroke, even in the absence of high blood pressure [28]. Moreover, impaired glucose tolerance caused by diabetes was proven to be a significant risk factor for cardiovascular diseases, including coronary heart disease and stroke, with a relative risk estimation of 1.9 (95% confidence interval [CI] of 1.2—3.2) [29].

H. Body Mass Index

Excess weight is scientifically proven to be associated with a high risk of several conditions that can potentially result in stroke, including hypertension [30]. Prior research found that women who have a higher body mass index (BMI) of 27 kg/m² or above had a considerably higher risk of contracting ischemic stroke; the same study also revealed that obesity and weight gain are significant risk factors for ischemic and total stroke in women, though this is not necessarily the case for hemorrhagic stroke [31]. Moreover, a similar study by other scholars but conducted on men revealed that men who have a body mass index of 30 kg/m² or above had a higher likelihood of getting strokes in comparison to those who had a body mass index of 23 kg/m² or lower [32].

I. Smoking Status

Various studies in the past have proved that smoking dramatically increases the likelihood of experiencing a stroke [33], [34], [35]. Smoking is also found to be a substantial factor contributing to the high risk of experiencing a stroke, with smokers facing a 50 percent greater risk compared to individuals who do not smoke [33], [36]. Furthermore, smoking is responsible for 12 percent to 14 percent of mortality resulting from strokes [14].

J. Stroke Recurrence

Stroke recurrence is not something to be taken lightly. It could lead to fatality, as past research has found that diabetes mellitus and atrial fibrillation are aligned with both the recurrence of stroke and recurrence-free survival [37]. This is also further supported by a recent study that revealed that patients with stroke who have hypertension, diabetes mellitus, atrial fibrillation, and coronary heart disease are statistically more likely to experience stroke recurrence [38].

III. Discussion And Analysis

This study uses a stroke prediction dataset from Kaggle [39] and presents the data analyzed with the Python 3 programming language by visualizing histograms and a heat map. The stroke prediction dataset contains data from a total of 5,110 patients.

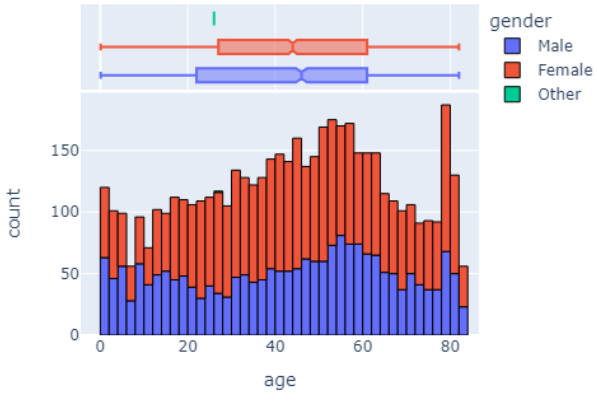

Fig. 1. Total Patients by Age

Fig. 1 counts the total of patients by age through classification of gender. Extracting one example from the data above, there are around 81 male patients aged 54 to 55.9 and about 82 female patients aged 60 to 61.9.

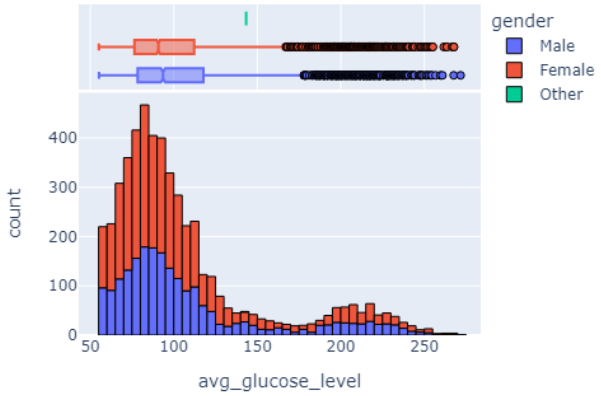

Fig. 2. Total Patients by Average Glucose Level

Fig. 2 counts patients' total by their average glucose level through gender classification. Extracting one example from the data above, around 228 female patients have an intermediate glucose level of 70 to 74.99 mg/dL. On the other hand, about 179 male patients have an average glucose level of 80 to 84.99 mg/dL. However, there is an alarming intermediate glucose level found in some patients. Some patients have an average glucose level of more than 250 mg/dL; thus, they are at high risk of getting strokes.

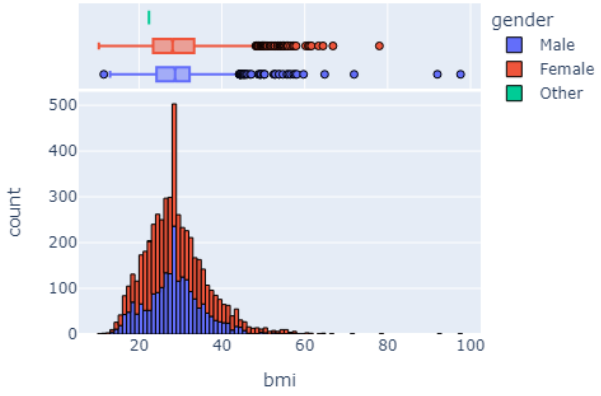

Fig. 3. Total Patients by Body Mass Index (BMI)

Fig. 3 counts patients' total by their body mass index (BMI) through gender classification. Extracting one example from the data above, most female patients (i.e., around 267 female patients) have a body mass index of about 28 kg/m² to 28.9 kg/m². They are also at risk of getting strokes because a body mass index between 25 kg/m² to 30 kg/m² is considered overweight and leads to obesity. On the other hand, most male patients (i.e., around 236 male patients) also have their body mass index at around 28 kg/m² to 28.9 kg/m².

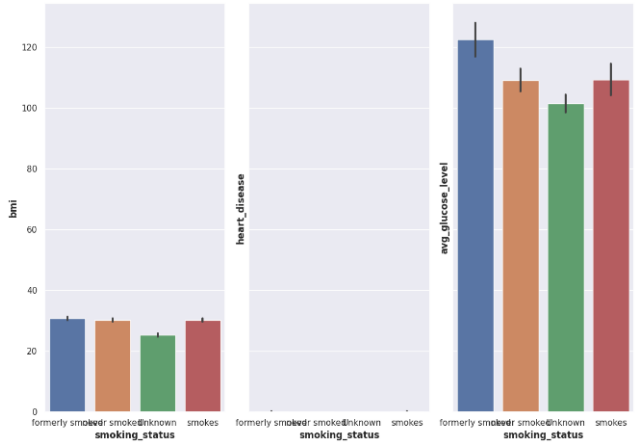

Fig. 4. FigSmoking Status Classification of Male Patients with Their Cardiac Disease History, Body Mass Index, and Average Glucose Level

Fig. 4 classifies male patients by their smoking status relating to their cardiac disease history, their body mass index, and their average glucose level. The blue-colored bar represents the smoking status of "formerly smoked," the orange-colored bar represents the smoking status of "never smoked," and the green-colored bar represents the smoking status of "unknown," meaning the information is unavailable for this patient. The red-colored bar represents the smoking status of "smokes," meaning they currently still smoke.

Extracting one example from the data above, those male patients who currently smoke have a body mass index of around 30 kg/m², meaning they are overweight and have a high risk of obesity and stroke. Meanwhile, those male patients who formerly smoked have the highest average glucose level of around 120 mg/dL. Additionally, male patients who either formerly smoked or currently smoke and either have or do not have heart disease are represented in the chart by a small number of either "0" (for not having heart disease) or "1" (for having heart disease).

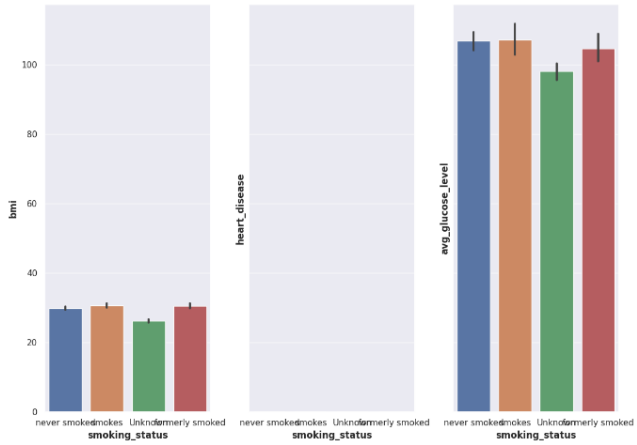

Fig. 5. Smoking Status Classification of Female Patients with Their Cardiac Disease History, Body Mass Index, and Average Glucose Level

Fig. 5 classifies female patients by their smoking status relating to their cardiac disease history, their body mass index, and their average glucose level. The blue-colored bar represents the smoking status of "never smoked," the orange-colored bar represents the smoking status of "smokes," meaning they still currently smoke; the green-colored bar represents the smoking status of "unknown," meaning the information is unavailable for this patient, and the red-colored bar represents the smoking status of "formerly smoked."

EISSN: 2961-872X ● PISSN: 2961-8924 ● DOI: 10.56741/esl.v3i01.470

Extracting one example from the data above, those female patients who formerly smoked have a body mass index of around 30 kg/m², meaning they are overweight and have a high risk of obesity and stroke. Meanwhile, those female patients who currently smoke have the highest average glucose level of more than 100 mg/dL.

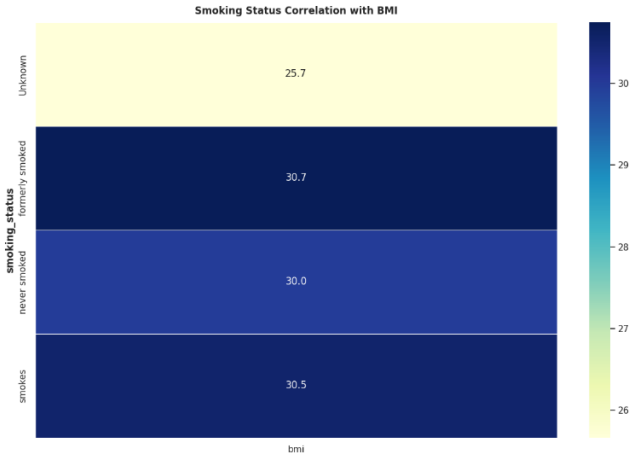

Fig. 6. Smoking Status Correlation with Body Mass Index Using A Heat Map Correlation

Fig. 6 shows the correlation between the patient's smoking status and their body mass index. Darker colors indicate stronger correlations between the two variables. Patients who had experience smoking, either formerly smoked or currently smoked, have a stronger correlation with their body mass index in a way that they have a high body mass index that indicates overweight or obesity. This is a potential danger, as they have a higher likelihood of getting strokes.

IV. Conclusion

Excluding age and gender, the majority of stroke risk factors identified in this paper are associated with lifestyle factors that can be mitigated, including smoking status, physical inactivity, dietary habits, and medical conditions related to stroke or cardiovascular diseases [8]. The results in this paper show that the majority of patients had a higher likelihood of getting strokes due to their increased body mass index and average glucose level, which are the identifiable major risk factors for stroke [28], [29], [30], effective preventive and mitigation measures that encompass the targeting of critical modifiable stroke risk factors [7] can be taken to prevent getting a stroke. This starts with changing lifestyle factors like more physical activity, healthy eating habits, and frequent exercise.

References

[1] C. P. Warlow, J. van Gijn, and M. Dennis, *Stroke: practical management*, 3rd ed. Oxford: Blackwell Publishing, 2008.

[2] World Health Organization, "The Top 10 Causes of Death," World Health Organization. Accessed: Dec. 14, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[3] M. O. Owolabi, S. Akarolo-Anthony, R. Akinyemi, D. Arnett, M. Gebregziabher, and C. Jenkins, "The burden of stroke in Africa: a glance at the present and a glimpse into the future," *Cardiovascular Journal Africa*, vol. 26, no. 2, pp. 27–38, 2015.

[4] K. Strong, C. Mathers, and R. Bonita, "Preventing stroke: saving lives around the world," *Lancet Neurol*, vol. 6, no. 2, 2007.

[5] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *The Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.

[6] L. Legg, A. Pollock, P. Langhorne, and C. Sellars, "A multidisciplinary research agenda for stroke rehabilitation," *British Journal of Therapy and Rehabilitation*, vol. 7, no. 7, pp. 319–324, 2000.

[7] W. Johnson, O. Onuma, M. Owolabi, and S. Sachdev, "Stroke: a global response is needed," *Bulletin of the World Health Organization*, vol. 94, no. 9, 2016.

[8] S. C. Johnston, S. Mendis, and C. D. Mathers, "Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling," *Lancet Neurology*, vol. 8, no. 4, 2009.

[9] N. Stroebele, F. Mueller-Riemenschneider, C. H. Nolte, J. Mueller-Nordhorn, A. Bockelbrink, and S. N. Willich, "Knowledge of risk factors, and warning signs of stroke: a systematic review from a gender perspective," *International Journal of Stroke*, vol. 6, no. 1, pp. 60–66, 2011.

[10] G. Boysen *et al.*, "Stroke Incidence and Risk Factors for Stroke in Copenhagen, Denmark," *Stroke*, vol. 19, no. 11, pp. 1345–1353, 1988.

[11] N. Dagres *et al.*, "Gender-Related Differences in Presentation, Treatment, and Outcome of Patients With Atrial Fibrillation in Europe. A Report From the Euro Heart Survey on Atrial Fibrillation," *J Am Coll Cardiol*, vol. 49, no. 5, pp. 572–577, Feb. 2007, doi: 10.1016/j.jacc.2006.10.047.

[12] G. Y. H. Lip, T. Watson, and E. Shantsila, "Anticoagulation for stroke prevention in atrial fibrillation: Is gender important?," *European Heart Journal*, vol. 27, no. 16. pp. 1893–1894, Aug. 2006. doi: 10.1093/eurheartj/ehl140.

[13] A. Towfighi, J. L. Saver, R. Engelhardt, and B. Ovbiagele, "A midlife stroke surge among women in the United States," *Neurology*, vol. 69, no. 20, pp. 1898–1904, 2007.

[14] M. Kelly-Hayes, "Influence of age and health behaviors on stroke risk: Lessons from longitudinal studies," *Journal of the American Geriatrics Society*, vol. 58, no. SUPPL. 2. Oct. 2010. doi: 10.1111/j.1532-5415.2010.02915.x.

[15] R. Marinigh, G. Y. H. Lip, N. Fiotti, C. Giansante, and D. A. Lane, "Age as a risk factor for stroke in atrial fibrillation patients: Implications for thromboprophylaxis," *Journal of the American College of Cardiology*, vol. 56, no. 11. Elsevier USA, pp. 827–837, Sep. 07, 2010. doi: 10.1016/j.jacc.2010.05.028.

[16] P. A. Wolf, R. D. Abbott, and W. B. Kannel, "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study," *Stroke*, vol. 22, no. 8, pp. 983–988, 1991.

[17] R. S. Vasan, A. Beiser, and S. et al Seshadri, "Residual Lifetime Risk for Developing Hypertension in Middle-aged Women and Men The Framingham Heart Study," *JAMA*, vol. 287, no. 8, pp. 1003–1010, 2002.

[18] S. Seshadri, P. A. Wolf, and A. et al. Beiser, "Elevated Midlife Blood Pressure Increases Stroke Risk in Elderly Persons: The Framingham Study," *Arch Intern Med*, vol. 161, no. 19, pp. 2343–2350, 2001.

[19] K. Kario *et al.*, "Morning Hypertension: The Strongest Independent Risk Factor for Stroke in Elderly Hypertensive Patients," 2006.

[20] Y. Komachi *et al.*, "Geographic and Occupational of Risk Factors in Cardiovascular Diseases in Japan," *Jpn Circ J*, vol. 35, no. 2, pp. 189–207, 1971.

[21] C. Zhu *et al.*, "The association of marital/partner status with patient-reported health outcomes following acute myocardial infarction or stroke: Protocol for a systematic review and meta-analysis," *PLoS One*, vol. 17, no. 11, 2022.

[22] K. K. Andersen and T. S. Olsen, "Stroke case-fatality and marital status," *Acta Neurol Scand*, vol. 138, no. 4, pp. 377–383, 2018.

[23] K. Berger, H. Schulte, M. D. Stuegbauer, and G. Assmann, "Incidence and Risk Factors for Stroke in an Occupational Cohort: The PROCAM Study," *Stroke*, vol. 29, no. 8, pp. 1562–1566, 1998.

[24] A. Kumar, M. Prasad, and P. Kathuria, "Sitting occupations are an independent risk factor for Ischemic stroke in North Indian population," *International Journal of Neuroscience*, vol. 124, no. 10, pp. 748–754, 2014.

[25] M. O'Donnell, D. Xavier, L. Liu, and H. et al. Zhang, "Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study," *The Lancet*, vol. 376, no. 9735, pp. 112–123, 2010.

[26] A. Holtermann, J. L. Marott, F. Gyntelberg, and K. et al. Sogaard, "Does the benefit on survival from leisure time physical activity depend on physical activity at work? A prospective cohort study," *PLoS One*, vol. 8, no. 1, 2013.

[27] V. R. Bhatt *et al.*, "Risk factors of stroke," *Journal of Institute of Medicine*, vol. 30, no. 3, pp. 37–41, 2008.

[28] J. Tuomilehto and D. Rastenyte, "Diabetes and glucose intolerance as risk factors for stroke," *J Cardiovasc Risk*, vol. 6, no. 4, pp. 241–249, 1999.

[29] J. Tuomilehto, A. Schranz, D. Aldana, and J. Pitkäniemi, "The Effect of Diabetes and Impaired Glucose Tolerance on Mortality in Malta," *Diabetic Medicine*, vol. 11, no. 2, pp. 170–176, 1994, doi: 10.1111/j.1464-5491.1994.tb02015.x.
[30] J. Stamler, "Epidemiologic findings on body mass and blood pressure in adults," *Ann Epidemiol*, vol. 1, no. 4, pp. 347–362, 1991.
[31] K. M. Rexrode, C. H. Hennekens, W. C. Willett, and G. A. et al. Colditz, "A Prospective Study of Body Mass Index, Weight Change, and Risk of Stroke in Women," *JAMA*, vol. 277, no. 19, 1997.
[32] T. Kurth, J. M. Gaziano, and K. et al. Berger, "Body Mass Index and the Risk of Stroke in Men," *Arch Intern Med*, vol. 162, no. 22, pp. 2557–2562, 2002.
[33] P. A. Wolf, R. B. D'Agostino, and W. B. Kannel, "Cigarette smoking as a risk factor for stroke: The Framingham Study," *JAMA*, vol. 259, 1988.
[34] T. A. Manolio, R. A. Kronmal, and G. L. et al. Burke, "Short-term predictors of incident stroke in older adults," *The Cardiovascular Health Study*, vol. 27, pp. 1479–1486, 1996.
[35] B. L. Rodriguez, R. D'Agostino, and R. D. et al. Abbott, "Risk of hospitalized stroke in men enrolled in the Honolulu Heart Program and the Framingham Study: A comparison of incidence and risk factor effects," *Stroke*, vol. 33, pp. 230–236, 2002.
[36] R. Shinton and G. Beevers, "Meta-analysis of relation between cigarette smoking and stroke," *BMJ*, vol. 298, pp. 789–794, 1989.
[37] T. Hillen, C. Coshall, K. Tilling, A. G. Rudd, R. McGovern, and C. D. A. Wolfe, "Cause of Stroke Recurrence Is Multifactorial: Patterns, Risk Factors, and Outcomes of Stroke Recurrence in the South London Stroke Register," *Stroke*, vol. 34, no. 6, pp. 1457–1463, 2003.
[38] S. Zheng and B. Yao, "Impact of risk factors for recurrence after the first ischemic stroke in adults: A systematic review and meta-analysis," *Journal of Clinical Neuroscience*, vol. 60, pp. 24–30, 2018.
[39] Fedesoriano, "Stroke Prediction Dataset." Kaggle, 2020. Accessed: Dec. 14, 2023. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/data

 EISSN: 2961-872X ● PISSN: 2961-8924 ● DOI: 10.56741/esl.v3i01.470