Evaluating Factors that Influence Recurrent Ischemic Stroke

For Data 621 Final

By Group #3

Coco Donovan
Jean Jimenez
Marjete Vucinaj
Matthew Roland

Agenda

I. Abstract
II. Introduction
III. Methods
IV. Results
V. Discussion
VI. Conclusion
VII. Works Cited

Abstract

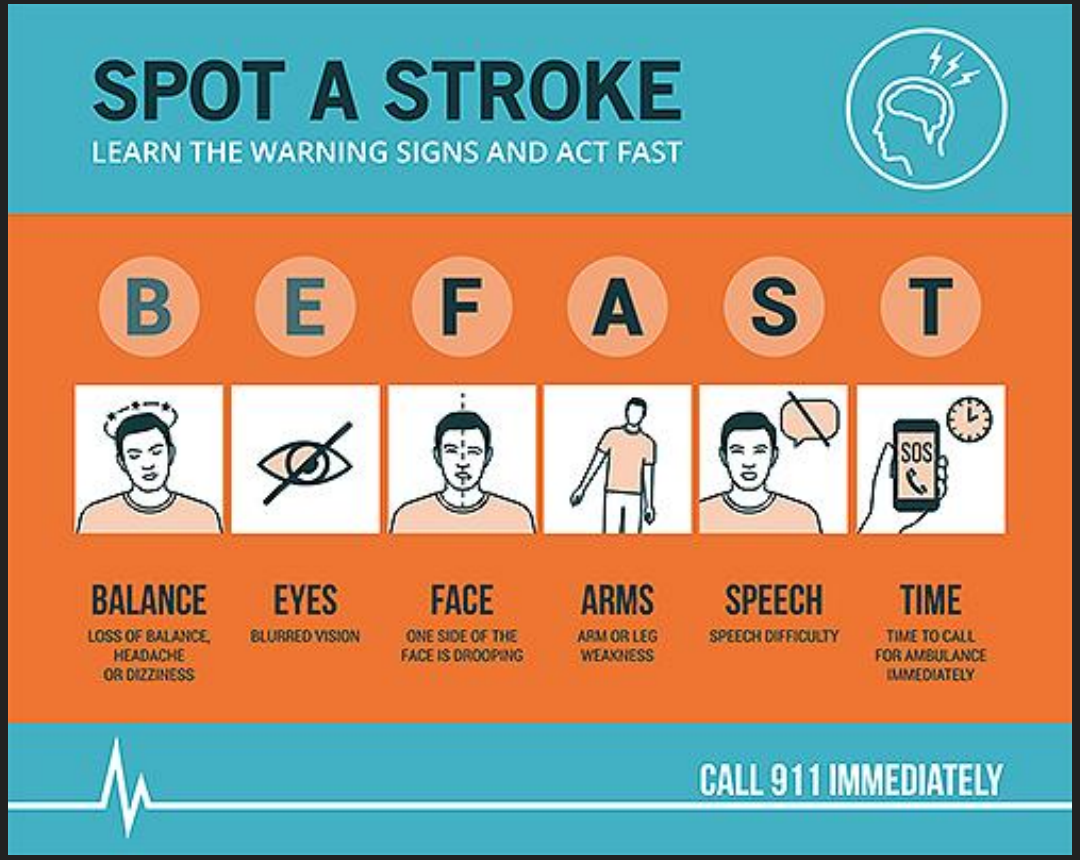
This study details our efforts to predict the likelihood of experiencing a stroke using a mix of demographic, physical health, and behavioral variables through a logistic regression model. We utilized demographic factors such as age, gender, and race/ethnicity in our model, alongside health indicators including heart health, family history of stroke, smoking status, BMI, and Diabetes status. By analyzing data from the StrokeHealth Outcomes (SHOUT) database, the logistic regression model will identify statistically significant predictors of stroke risk, accounting for potential confounders and interactions between variables. The findings, as they pertain to predictive success and relevant predictor variables, will inform strategic medical/public health interventions that aim to reduce stroke burden by identifying high-risk populations and modifiable risk factors. Ultimately, this research aims to contribute to our understanding of stroke determinants and guide preventive strategies to improve public health outcomes. Our study highlights meaningful relationships between patients' health and demographic info such as age and gender, in addition to the complexity that varied socio-demographics present when trying to understand medical outcomes.

Keywords: Recurrent Stroke, Logistic Regression, Stroke Risk Factors, Predictive Modeling, Health Disparities.

Introduction

- Stroke is a condition that affects 795,000 Americans each year [Citation]
- Age, high blood pressure, high cholesterol, obesity, diabetes, and smoking are all risk factors of stroke
- Stroke affects people disproportionately, with Black adults being twice as likely to get a stroke than white adults [Citation]
- Acute Ischemic Stroke (AIS) is defined as an infarct (clot) in the brain that prevents blood flow
- Symptoms of stroke include but not limited to:
    - Lateral numbness (numb on one side of body)
    - Confusion
    - Speech deficits/ Aphasia
    - Sudden changes in Mobility (suddenly cannot move a body part or weakness)
    - Headaches
    - Loss of Consciousness
    - Death

Introduction



Introduction

- Since stroke is loss of blood flow to the brain, it is important to get to the hospital quick
- Good stroke outcomes are associated with quick time to treatment
- Recurrent stroke is a stroke that happens after an initial stroke
- Having a stroke increases your risk of having a recurrent stroke
- Recurrent strokes may have a 25.0 % mortality rate (although not that well researched)
- What are the risk factors associated with having a recurrent stroke (AIS)

Methods

- This study used records from the StrokeHealth Outcomes (SHOUT) database
- SHOUT is a database of patient with AIS who were treated in a comprehensive stroke center
- All adult patients with a diagnosis of AIS in 2023 was used in this study
- Recurrent stroke defined as patient having a stroke after their initial stroke hospital visit
- Data collected with IRB approval

Methods

- Data was de-identified, cleaned, and dummy variables were coded
- An exploratory data analysis was conducted
- Missing values were filled in using median imputation
- Logarithmic transformation was conducted on some factors to achieve a more normal distribution
- Data set was split into training and testing datasets
- A correlation matrix was created and factors with high correlation were eliminated
- A Logit, Lasso, Ridge regression and decision trees were all trained on the data
- MSE was used to evaluate the model we will ultimately use

Results

- The dataset contains a total of 29,632 records/ observations
- Of those, 4500 were recurrent stroke visits

Results

Table 1: Descriptive statistics of key numerical variables for patients in a recurrent stroke dataset, detailing means, standard deviations, medians, interquartile ranges, and value distributions along with counts of valid and missing observations.

Summary Statistics for Recurrent Stroke Numerical Variables

Stroke Dataset Analysis[1]

Variable	Mean	Standard Deviation	Median	Interquartile Range	Minimum	Maximum	Valid Observations	Missing Observations
age	71.36	14.45	73.00	21.00	18.00	121.00	23730	0
Length_of_stay_hours	167.63	214.77	118.00	127.00	-2,032.00	9,666.00	23728	2
MRS_discharge_score_cleaned	2.07	1.72	2.00	4.00	0.00	8.00	20345	3385
Arrival_NIHSS_score	6.55	10.68	3.00	8.00	-7.00	999.00	17788	5942
Arrival_NIHSS_score_cleaned	6.50	7.66	3.00	8.00	0.00	42.00	17786	5944
hasIVTPA	0.12	0.33	0.00	0.00	0.00	1.00	23730	0
BMI	27.79	7.67	26.78	7.40	2.03	259.18	19296	4434
TARGET	0.15	0.36	0.00	0.00	0.00	1.00	23730	0

[1] Analysis conducted on the stroke dataset.

Results - Model Selection

- A Logit, Lasso, Ridge regression and decision trees were all trained on the data
- MSE was used to evaluate the model we will ultimately use

Table 2: Comparison of mean squared errors (MSE) across different predictive models including logistic regression, lasso regression, ridge regression, and decision tree, evaluating their performance in a predictive analytics context.

Model Comparison

Mean Squared Errors of Predictive Models

Model	Mean Squared Error
Logistic Regression	0.1094
Lasso Regression	0.1094
Ridge Regression	0.1095
Decision Tree	0.1277

Results - Visualizing MSE

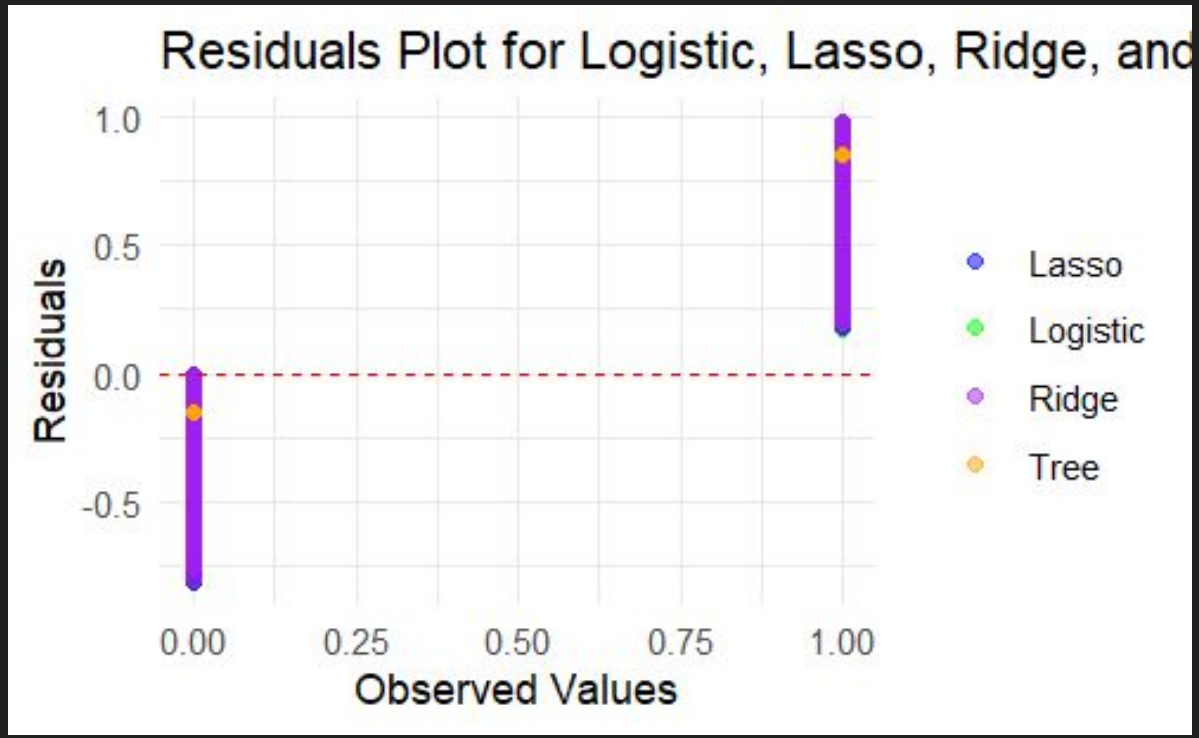


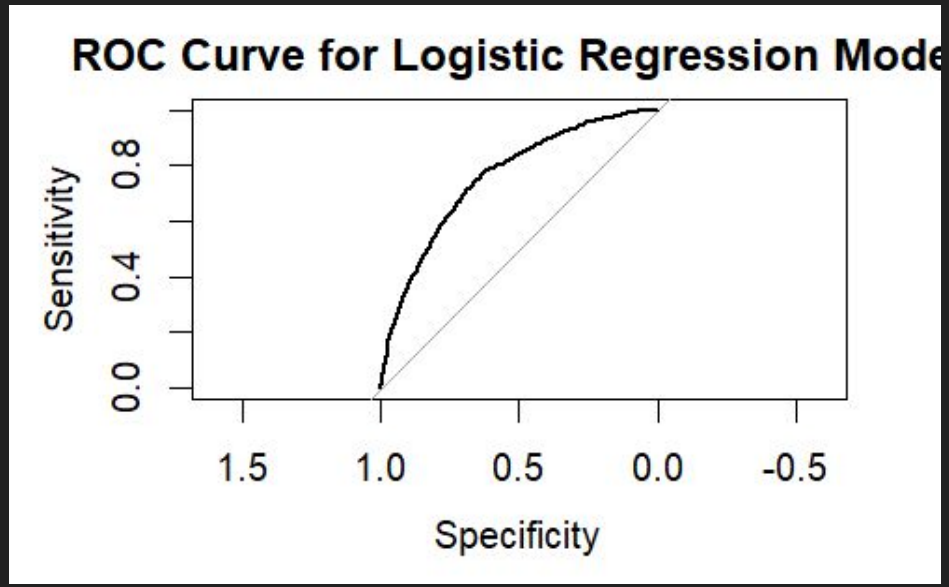
Figure 1: Residuals plot comparing model fits for logistic regression, lasso regression, ridge regression, and decision tree models based on observed binary outcomes.

Results - Logit Model

- Age showed a negative coefficient (-0.0144, $p < 0.001$), indicating that the likelihood of recurrent stroke decreases with increasing age
- The length of hospital stay also had a slight negative impact on recurrence rates (-0.0003, $p = 0.047$)
- Clinical severity scores such as the MRS discharge score (0.1053, $p < 0.001$) and the Arrival NIHSS score (0.0113, $p < 0.001$) were positively associated with stroke recurrence, suggesting that more severe initial symptoms are linked to higher risks of recurrence

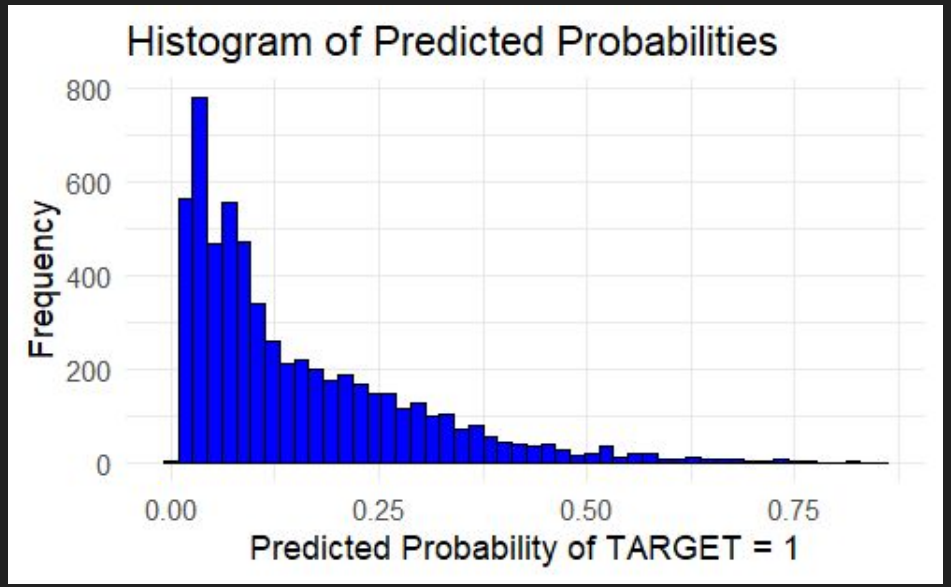
Results - Logit Model

Figure 3: Receiver Operating Characteristic (ROC) curve for evaluating the performance of a logistic regression model, plotting sensitivity against 1-specificity



Results - Model Performance

- Probabilities mainly cluster at lower values, indicating low risk of stroke recurrence for most patients.
- Most frequent probability range is 0 to 0.25, with a peak near 0.1.
- Distribution is right-skewed with fewer high probability cases.
- Small tail extends towards higher risk, up to a probability of 1.
- Right skewness suggests that most individuals are at lower risk or reflects conservative model predictions.
- Possible influence of model training on an imbalanced dataset with fewer high-risk cases.



Discussion

- Hypertension, hyperlipidemia, high BMI, increased length of stay.
- Findings align with known risk factors of recurrent stroke.
- Black individuals have higher instances of recurrent stroke, indicating significant racial disparities.
- Type 2 diabetes mellitus (not type 1) linked with increased recurrent stroke risk.
- Contrary to the extant literature, younger individuals (<50 years) showed higher risk of recurrent stroke.
    - Potential characteristic of our dataset; perhaps younger individuals who have had a stroke are at risk for recurrence
    - Potential limitation in our model

Conclusion

- Engaging in healthy lifestyle approaches post-stroke is crucial to reduce recurrence risk
- Poor diets, stress, and lack of activity are linked with higher recurrence rates
- Encourage lifestyle modifications and appropriate medication use, especially important for younger individuals
- Black individuals face higher risks of recurrent stroke
    - Possible influences include discrimination impacting health and cerebrovascular disease risk

Conclusion - Limitations and Future Directions

- Potential for a more suitable model to better represent our data
- Possibility that key predictive features are missing from the dataset
- Incorporate additional variables and larger datasets to improve model accuracy
- Add time series component to track individual cases of recurrent stroke over time
- Based on findings, develop targeted interventions to mitigate risk factors identified, particularly in demographics showing higher risk

Works Cited

1. Albright, K. C., Huang, L., Blackburn, J., Howard, G., Mullen, M., Bittner, V., ... & Howard, V. (2018). Racial differences in recurrent ischemic stroke risk and recurrent stroke case fatality. Neurology, 91(19), e1741-e1750.
2. Beatty Moody, D. L., Taylor, A. D., Leibel, D. K., Al-Najjar, E., Katzel, L. I., Davatzikos, C., ... & Waldstein, S. R. (2019). Lifetime discrimination burden, racial discrimination, and subclinical cerebrovascular disease among African Americans. Health Psychology, 38(1), 63.
3. Davis, B. A. (2010). Discrimination: A Social Determinant Of Health Inequities. Health Affairs Forefront.
4. Engel-Nitz, N. M., Sander, S. D., Harley, C., Rey, G. G., & Shah, H. (2010). Costs and outcomes of noncardioembolic ischemic stroke in a managed care population. Vascular Health and Risk Management, 6, 905-913.
5. Faraway, J. J. (2006). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman and Hall/CRC.
6. Faraway, J. J. (2014). Linear Models with R - 2nd Edition. Chapman and Hall/CRC.
7. Ferrone, S. R., Boltyenkov, A. T., Lodato, Z., O'Hara, J., Vialet, J., Malhotra, A., ... & Sanelli, P. C. (2022). Clinical outcomes and costs of recurrent ischemic stroke: A systematic review. Journal of Stroke and Cerebrovascular Diseases, 31(6), 106438.
8. ICD-10-CM Diagnosis Code I63.9: Cerebral infarction. (2024). Retrieved from https://www.icd10data.com/ICD10CM/Codes/I00-I99/I60-I69/I63-/I63.9
9. Kolmos, M., Christoffersen, L., & Kruuse, C. (2021). Recurrent ischemic stroke–a systematic review and meta-analysis. Journal of Stroke and Cerebrovascular Diseases, 30(8), 105935.
10. Sanelli, P., Yang, B., O'Hara, J., Jimenez, J., Gribko, M., Martinez, G., Feizullayeva, C., Adinarayan, K., Hoang, A., Qui, M., Wang, J., Boltyenkov, A., Sangha, K., Sanmartin, M., Elhabr, A., & Katz, J. (2023, April 11). Stroke health outcomes database. Retrieved April 11, 2023.
11. Shah, S., Liang, L., Kosinski, A., Hernandez, A. F., Schwamm, L. H., Smith, E. E., ... & Xian, Y. (2020). Safety and outcomes of intravenous TPA in acute ischemic stroke patients with prior stroke within 3 months: Findings from get with the guidelines–stroke. Circulation: Cardiovascular Quality and Outcomes, 13(1), e006031.
12. Singh, R. J., Chen, S., Ganesh, A., & Hill, M. D. (2018). Long-term neurological, vascular, and mortality outcomes after stroke. International Journal of Stroke, 13(8), 787-796.
13. Tsao, C. W., Aday, A. W., Almarzooq, Z. I., et al. (2022). Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation, 145, e153–e639. https://doi.org/10.1161/CIR.0000000000001052
14. WHO EMRO | Stroke, Cerebrovascular accident | Health topics. (2023). World Health Organization - Regional Office for the Eastern Mediterranean. Retrieved from https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html