**Introduction**

A stroke is a medical condition caused by reduced blood flow to areas of the brain, resulting in cell death [(Sekerdag, Solaroglu, and Gursoy-Ozdemir 2018)](https://paperpile.com/c/TnBKZS/KZin). Symptoms depend on the anatomical location of the stroke and may include “sudden numbness or weakness of face, arm, or leg, especially on one side of the body; sudden confusion or trouble speaking or understanding speech; sudden trouble seeing in one or both eyes; sudden trouble walking, dizziness, or loss of balance or coordination; and sudden severe headache with no known cause.” [(Wall et al. 2008)](https://paperpile.com/c/TnBKZS/yNKp) A prolonged stroke may result in severe brain damage or death, but the outcome of the stroke is impacted by the timing and quality of subsequent medical care [(Maulden et al. 2005)](https://paperpile.com/c/TnBKZS/DK8S). Stroke can be diagnosed in several ways, including an examination of physical and neurological symptoms, and various imaging techniques such as computed tomography (CT) scans and magnetic resonance imaging (MRI) scans [(Yew and Cheng 2009)](https://paperpile.com/c/TnBKZS/7jRM). This helps in the evaluation of location, severity, and type of stroke.

Stroke has been linked to a variety of preconditions and diseases. Preconditions resulting in a heightened risk of stroke include hypertension and diabetes mellitus [(O’Donnell et al. 2010)](https://paperpile.com/c/TnBKZS/ZNGL). Given the large body of circumstantial evidence, the most effective methods of decreasing stroke risk include increased physical activity, a healthy diet, and decreased use of substances such as alcohol and tobacco [(Boehme, Esenwa, and Elkind 2017)](https://paperpile.com/c/TnBKZS/67IP).

Add paragraph

* As a leading cause of death worldwide (find citation), stroke is a pressing issue in the global health field.
* Add statistics about stroke around the world:
  + <https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(19)30034-1/fulltext>
  + <https://www.who.int/bulletin/volumes/94/9/16-181636/en/>

Similar pre-conditions for stroke are common in sub-Saharan Africa. 80 percent of the population in sub-Saharan Africa is rural, and on average, imaging is performed on a mere 10-13% of deserving patients in rural areas (Kawooya 2012). Medical attention is somewhat more accessible in urban areas, but the cost of treatment is still a limiting factor (citation). The vast majority of studies on the topic of stroke epidemiology use subjects based in high-income countries and are thus applicable to populations within those countries (rephrase or add citation); however, the World Stroke Organization estimates that 2 out of 3 strokes occur in low- or middle-income countries (citation), and the global scientific community has generally neglected these groups. Furthermore, people of African descent are nearly twice as likely to experience a stroke (citation).

**Methodology**

**Stroke Investigative and Research Education Network (SIREN)**

This article is based on data from the Stroke Investigative Research and Education Network (SIREN), a study that involves patients from 15 sites in Nigeria and Ghana. SIREN is one of the most comprehensive stroke studies performed using subjects in sub-Saharan Africa. This paper includes roughly 1133 stroke patients and 1709 control cases.

SIREN is a multicenter study, encompassing 15 studies across the northern and southern belts of Nigeria and Ghana. The protocol of SIREN has been detailed previously by Akpalu et al. [(Akpalu et al. 2015)](https://paperpile.com/c/TnBKZS/gMD7).

Stroke patients that were included in this study were consenting adults (age 18 or older) - in patients that were unconscious or aphasic (unable to communicate), consent was obtained from the next of kin. These patients first presented with clinical stroke within 8 days of known or suspected symptom onset. All cases had neuroimaging performed, with either CT or MRI, within 10 days of symptom onset. Although the patients were recruited from hospitals in order to prevent inaccurate stroke phenotyping, “a robust community engagement core incorporated community sensitization programs to enhance early presentation at SIREN hospitals and minimize referral bias.” [(F. S. Sarfo et al. 2018)](https://paperpile.com/c/TnBKZS/c8Jr) Control cases were stroke-free adults that consented to participate in SIREN - these cases were largely from the same communities where stroke patients were recruited. Their stroke-free status was verified with a questionnaire that has “98% negative predictive value.” [(F. Sarfo et al. 2016)](https://paperpile.com/c/TnBKZS/qi4h) To minimize potential confounding factors, controls were matched by age (with a variation of ±5 years), sex, and ethnicity. All study sites received approval from their institutional review board (IRB) and informed consent was given by all patients and controls.

Stroke was verified in patients through clinical evaluation of symptoms, ECG, transthoracic echocardiography, and carotid Doppler ultrasound and neuroimaging techniques - more specifically, with computed tomography or magnetic resonance imaging. These were performed depending on the standardized protocol at each site [(F. S. Sarfo et al. 2018)](https://paperpile.com/c/TnBKZS/c8Jr). Ischemic stroke subtypes were identified using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) [(Kolominsky-Rabas et al. 2001)](https://paperpile.com/c/TnBKZS/IZ3m) and intracerebral hemorrhages were divided into Structural, Medication-Related, Amyloid Angiopathy, Systemic/Other Disease, Hypertension and Undetermined (SMASH-U) causes [(Meretoja et al. 2012)](https://paperpile.com/c/TnBKZS/39uG).

**Definition of Variables**

A subset of variables collected by SIREN were chosen based on their association with stroke. This information was collected from a variety of sources. These variables and their sources are listed in Appendix A.

**Pearson Correlation Coefficient**

The Pearson correlation coefficient (PCC), also known as the bivariate correlation, is a measure of the linear correlation between two variables. In other words, the PCC measures how “related” two variables are. For this project, the PCC represents the extent to which the presence of stroke varies with the selected attributes. It has a value between +1 and -1, where +1 is a completely linear positive relationship, -1 is a completely linear negative relationship, and 0 is no linear correlation. The PCC can be computed using Weka [(Witten, Frank, and Hall 2011)](https://paperpile.com/c/TnBKZS/sUmV).

(explain bootstrap confidence interval)

**Logistic Regression**

(further explanation - transcribe from powerpoint)

Logistic regression computes the probability of each class for a given subject in the test group. The probability threshold for designating a subject as positive or negative (stroke or control) varies; for example, with a default threshold of 0.5, a probability of less than 0.5 implies that a subject has no stroke, whereas a probability of greater than 0.5 implies that the subject is a stroke patient. The threshold can be calibrated to reduce error in the model.

**ROC curve**

The true positive rate (TPR) describes the likelihood that a given model will predict the stroke class for a subject who is a stroke patient. This figure is also referred to as the sensitivity. In the equation below, represents the number of true positive values and represents the number of false negative values.

TPR = Sensitivity =

**Eq. 1**

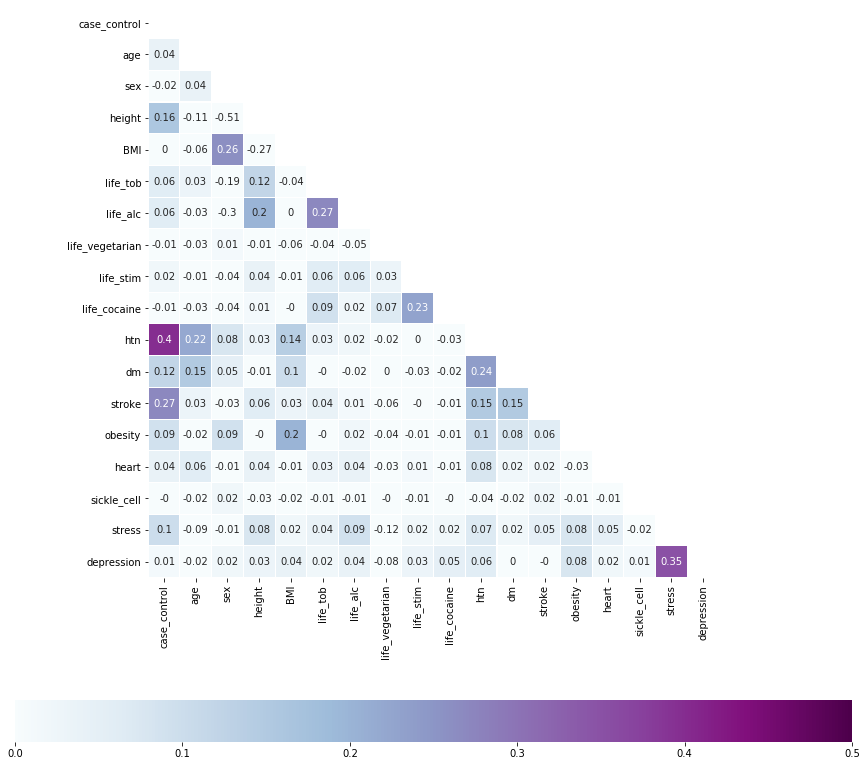
The false positive rate represents the probability that a model will predict the stroke class for a subject who is not a stroke patient. The false positive rate is also known as the inverted specificity. In the equation below, represents the number of true negative values and represents the number of false positive values.

FPR = Specificity =

**Eq. 2**

The Receiver Operating Characteristic (ROC) curve, is useful when comparing the accuracy of machine learning algorithms. The ROC curve is a plot of the false positive rate versus the true positive rate for a range of different probability thresholds between 0 and 1. The area under the ROC curve can be used to characterize the model skill.

**Results**

**Pearson correlation coefficient**

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Correlation** | **95% Confidence Interval** |
| Hypertension | 0.396 | 0.362 to 0.430 |
| Past stroke | 0.270 | 0.245 to 0.295 |
| Height | 0.163 | 0.131 to 0.198 |
| Diabetes | 0.121 | 0.084 to 0.159 |
| Stress | 0.100 | 0.063 to 0.139 |

**Table 1. Correlation between occurrence of stroke and hypertension, past stroke, height, diabetes, and stress. (add description of dataset - table must be independent from text) Included with each correlation is a bootstrap confidence interval.**

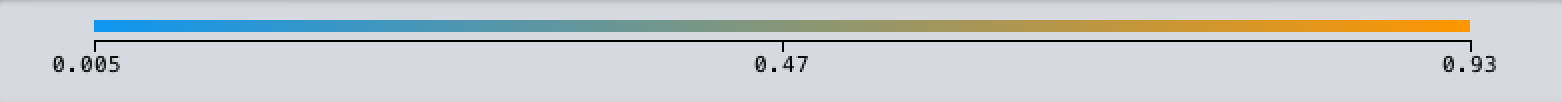
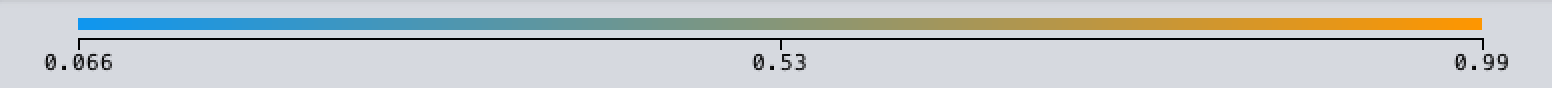
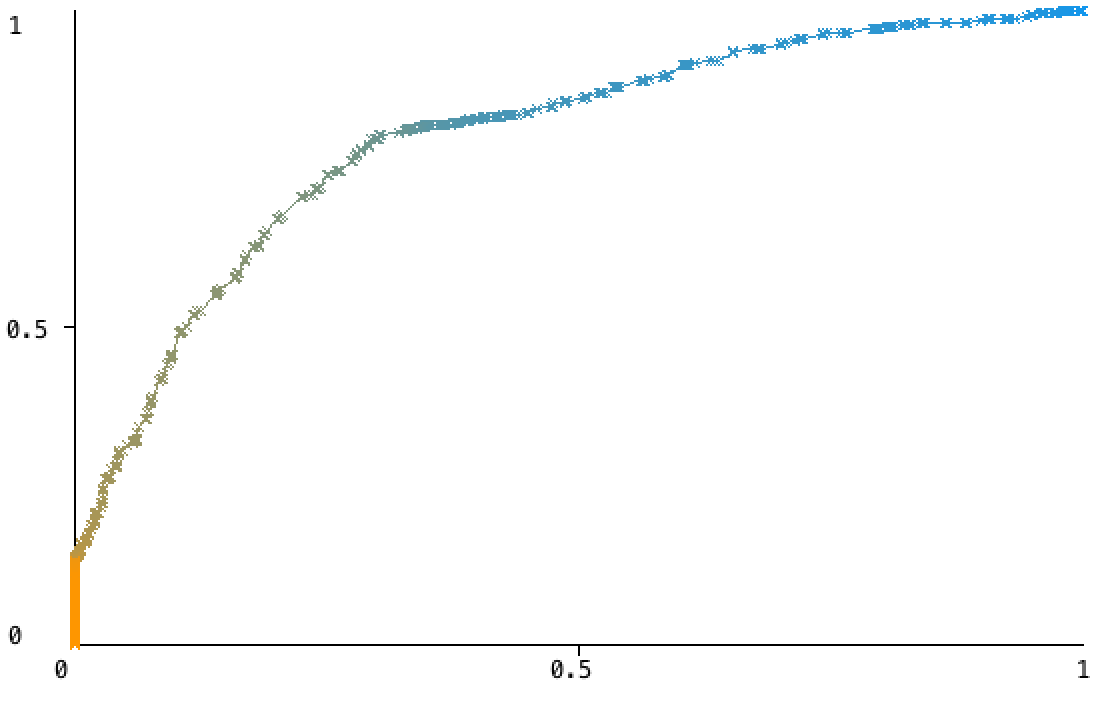
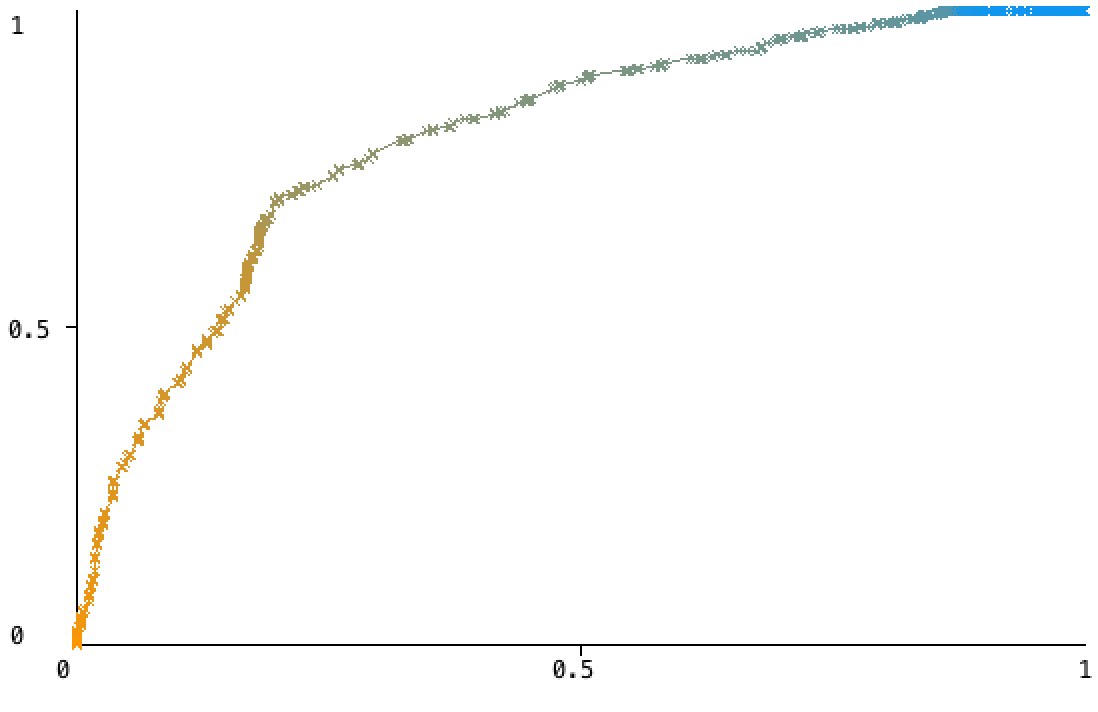
Above is a table listing the PCC for each attribute selected. Hypertension exhibits the highest correlation with stroke, with a value of 0.396. Past stoke exhibits the next highest correlation with stroke, with a value of 0.27. Height, diabetes, and stress are similarly correlated with stroke, at 0.163, 0.121, and 0.1 (respectively).

**Logistic Regression**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **TP Rate** | **FP Rate** | **ROC Area** | **Class** |
|  | 0.773 | 0.295 | 0.800 | 0 |
|  | 0.705 | 0.227 | 0.800 | 1 |
| **Weighted Average** | 0.746 | 0.268 | 0.800 |  |

**Table 2. True positive rate, false positive rate, and ROC area of class 0 (control) and class 1 (stroke case), in addition to the weighted average of these values - as computed by WEKA (add citation).**

The ROC curves look different. Why are the ROC areas identical?



**Table 3. Logistic regression ROC curves of class 0 (control, left) and class 1 (stroke case, right).**

**Discussion**

**Acknowledgements**

**Appendix A**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable code** | **Survey definition** | **Value** | **Source** |
| htn | Have you been diagnosed by your doctor with hypertension? | 1 - Yes  0 - No |  |
| stroke | Have you been diagnosed by your doctor with stroke? | 1 - Yes  0 - No |  |
| height | Height of subject (cm) | Number, min: 120, max: 200 |  |
| dm | Have you been diagnosed by your doctor with diabetes mellitus? | 1 - Yes  0 - No |  |
| life\_vegetarian | Are you a vegetarian? | 1 - Yes  0 - No |  |
| stress | Have you been diagnosed by your doctor with stress in the last 2 weeks? | 1 - Yes  0 - No |  |
| present\_weight | Present weight of subject (kg) | Number, min: 30, max: 170 |  |
| obesity | Have you been diagnosed by your doctor with obesity? | 1 - Yes  0 - No |  |
| heart | Have you been diagnosed by your doctor with heart disease? | 1 - Yes  0 - No |  |
| life\_stim | Have you ever used stimulants? | 1 - Yes  0 - No |  |
| depression | Have you been diagnosed by your doctor with depression in the last 4 weeks? | 1 - Yes  0 - No |  |
| age | Age of subject | Integer, min: 0 |  |
| life\_alc | Which best describes the subject’s history of alcohol use? | 0 - Never used alcohol  1 - Formerly used alcohol  2 - Past 12 months  3 - Past 30 days  4 - Currently uses alcohol |  |
| life\_tob | Which best describes the subject’s history of tobacco use? | 0 - Never used tobacco products  1 - Formerly used tobacco products  2 - Currently uses tobacco products |  |
| life\_cocaine | Have you ever used cocaine? | 1 - Yes  0 - No |  |
| sickle\_cell | Have you been diagnosed by your doctor with sickle cell disease? | 1 - Yes  0 - No |  |
| sex | Sex of subject | 1 - Male  2 - Female |  |

**References**

[Akpalu, Albert, Fred Stephen Sarfo, Bruce Ovbiagele, Rufus Akinyemi, Mulugeta Gebregziabher, Reginald Obiako, Lukman Owolabi, et al. 2015. “Phenotyping Stroke in Sub-Saharan Africa: Stroke Investigative Research and Education Network (SIREN) Phenomics Protocol.” *Neuroepidemiology* 45 (2): 73–82.](http://paperpile.com/b/TnBKZS/gMD7)

[Boehme, Amelia K., Charles Esenwa, and Mitchell S. V. Elkind. 2017. “Stroke Risk Factors, Genetics, and Prevention.” *Circulation Research*. https://doi.org/](http://paperpile.com/b/TnBKZS/67IP)[10.1161/circresaha.116.308398](http://dx.doi.org/10.1161/circresaha.116.308398)[.](http://paperpile.com/b/TnBKZS/67IP)

[Kawooya, Michael G. 2012. “Training for Rural Radiology and Imaging in Sub-Saharan Africa: Addressing the Mismatch between Services and Population.” *Journal of Clinical Imaging Science* 2 (June): 37.](http://paperpile.com/b/TnBKZS/Sgp3)

[Kolominsky-Rabas, P. L., M. Weber, O. Gefeller, B. Neundoerfer, and P. U. Heuschmann. 2001. “Epidemiology of Ischemic Stroke Subtypes according to TOAST Criteria: Incidence, Recurrence, and Long-Term Survival in Ischemic Stroke Subtypes: A Population-Based Study.” *Stroke; a Journal of Cerebral Circulation* 32 (12): 2735–40.](http://paperpile.com/b/TnBKZS/IZ3m)

[Maulden, Sarah A., Julie Gassaway, Susan D. Horn, Randall J. Smout, and Gerben DeJong. 2005. “Timing of Initiation of Rehabilitation after Stroke.” *Archives of Physical Medicine and Rehabilitation* 86 (12 Suppl 2): S34–40.](http://paperpile.com/b/TnBKZS/DK8S)

[Meretoja, Atte, Daniel Strbian, Jukka Putaala, Sami Curtze, Elena Haapaniemi, Satu Mustanoja, Tiina Sairanen, et al. 2012. “SMASH-U: A Proposal for Etiologic Classification of Intracerebral Hemorrhage.” *Stroke; a Journal of Cerebral Circulation* 43 (10): 2592–97.](http://paperpile.com/b/TnBKZS/39uG)

[O’Donnell, Martin J., Denis Xavier, Lisheng Liu, Hongye Zhang, Siu Lim Chin, Purnima Rao-Melacini, Sumathy Rangarajan, et al. 2010. “Risk Factors for Ischaemic and Intracerebral Haemorrhagic Stroke in 22 Countries (the INTERSTROKE Study): A Case-Control Study.” *The Lancet* 376 (9735): 112–23.](http://paperpile.com/b/TnBKZS/ZNGL)

[Sarfo, Fred, Mulugeta Gebregziabher, Bruce Ovbiagele, Rufus Akinyemi, Lukman Owolabi, Reginald Obiako, Onoja Akpa, et al. 2016. “Multilingual Validation of the Questionnaire for Verifying Stroke-Free Status in West Africa.” *Stroke; a Journal of Cerebral Circulation* 47 (1): 167–72.](http://paperpile.com/b/TnBKZS/qi4h)

[Sarfo, Fred Stephen, Bruce Ovbiagele, Mulugeta Gebregziabher, Kolawole Wahab, Rufus Akinyemi, Albert Akpalu, Onoja Akpa, et al. 2018. “Stroke Among Young West Africans: Evidence From the SIREN (Stroke Investigative Research and Educational Network) Large Multisite Case-Control Study.” *Stroke; a Journal of Cerebral Circulation* 49 (5): 1116–22.](http://paperpile.com/b/TnBKZS/c8Jr)

[Sekerdag, Emine, Ihsan Solaroglu, and Yasemin Gursoy-Ozdemir. 2018. “Cell Death Mechanisms in Stroke and Novel Molecular and Cellular Treatment Options.” *Current Neuropharmacology* 16 (9): 1396–1415.](http://paperpile.com/b/TnBKZS/KZin)

[Wall, Hilary K., Brianne M. Beagan, June O’Neill, Kathleen M. Foell, and Cynthia L. Boddie-Willis. 2008. “Addressing Stroke Signs and Symptoms through Public Education: The Stroke Heroes Act FAST Campaign.” *Preventing Chronic Disease* 5 (2): A49.](http://paperpile.com/b/TnBKZS/yNKp)

[Witten, Ian H., Eibe Frank, and Mark A. Hall. 2011. “Introduction to Weka.” *Data Mining: Practical Machine Learning Tools and Techniques*. https://doi.org/](http://paperpile.com/b/TnBKZS/sUmV)[10.1016/b978-0-12-374856-0.00010-9](http://dx.doi.org/10.1016/b978-0-12-374856-0.00010-9)[.](http://paperpile.com/b/TnBKZS/sUmV)

[Yew, Kenneth S., and Eric Cheng. 2009. “Acute Stroke Diagnosis.” *American Family Physician* 80 (1): 33–40.](http://paperpile.com/b/TnBKZS/7jRM)