Stroke Prediction

An Exercise in Machine Learning
and Stroke Probability Predictions

Yakov Krasnikov

Charles Philpott

Troy Youngblood

Inspiration: Stroke Screening Tool

Cerebrovascular accidents (strokes) in 2020 were the 5th leading cause of death [1] in the United States.

A stroke occurs when the blood supply to a region of the brain is suddenly stopped by either a blockage or a rupture.

Blockage is an ischemic stroke and accounts for 87% [2] of all strokes.

Rupturing of a blood vessel is a hemorrhagic stroke and accounts for 13% [3] of all strokes.

Ischemic Stroke

Area at Risk

Lodged Blood Clot

Hemorrhagic Stroke

Cerebral Hemorrhage
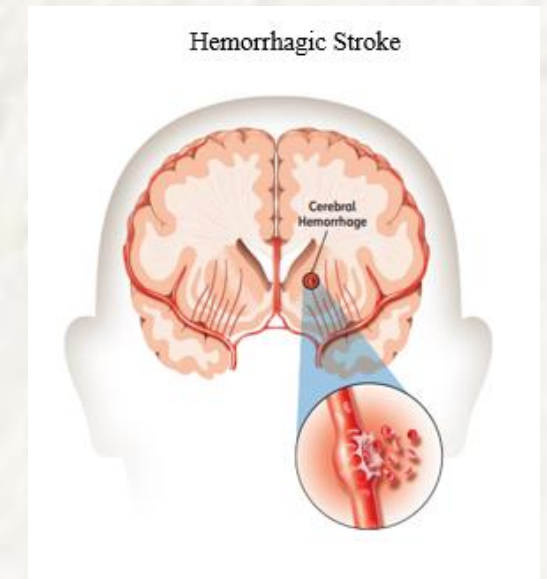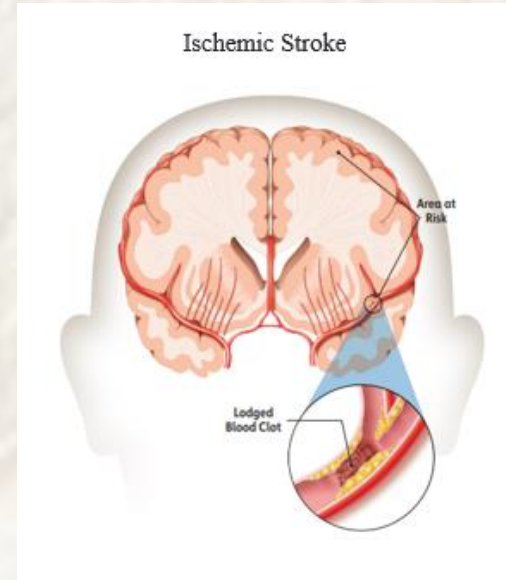
Source of Images [4]

80% of strokes are preventable [5] American Stroke Association

Data: 11 Clinical features for Predicting Stroke Events

- Data [4] from Kaggle

- Dataset used for model has 5109 unique rows

- 59% Female, 41% Male

- 5% have had a stroke

- 92% of strokes occur over age of 50
  - 38% of strokes between 70 – 79

- Obesity and Current/Former Smoker top 2 comorbidities

- Occur 86% of stroke victims had comorbidities

Data Attributes
- id unique identifier for each set of information
- gender: "Male, "Female"
- age: age of the patient
- hypertension: 0 no hypertension, 1 if hypertension
- heart_disease: 0 no heart disease, 1 if heart disease
- ever_married: "No" or "Yes"
- work_type: "children", "Govt_job", "Never_worked", "Private", or "Self-employed"
- Residence_type: "Rural" or "Urban"
- avg_glucose_level: average glucose level in blood
- bmi: body mass index
- smoking_status: "formerly smoked", "never smoked", "smokes", or "Unknown"
- stroke: 0 no stroke, 1 if stroke

80% of strokes are preventable [5] American Stroke Association

Hypothesis

A reliable predictive model can be developed if:

- Data and
- Key stroke attributes are correctly identified

Importance features of model will be compared against the risk factors from the American Stroke Association.

If the attributes are correctly identified by the model, the hypothesis will be considered validated.

Risk Factors from American Stroke Association

- Age (cannot be controlled)
- Heart Disease
- Diabetes
- High Blood Pressure
- Smoking
- Obesity
- Gender (cannot be controlled)

80% of strokes are preventable [5] American Stroke Association

Web Application

Conclusion

Hypothesis Criteria – model importance risk factors match American Stroke Association

- Age
- Heart Disease
- Diabetes
- High Blood Pressure
- Smoking
- Obesity
- Gender

Hypothesis Validated:

The top eleven in Feature Importance:

- Age
- Heart disease
- Diabetes (avg_glucose_level)
- High Blood Pressure (Hypertension)
- Married
- Smoking (smoking_status_former & smokes)
- work_type_self-employed & Private
- Obesity (bmi)
- Gender

(Grayed out attributes are dataset dependencies)

80% of strokes are preventable [5] American Stroke Association

Actionable Items

This model is one of many tools which are needed to increase awareness and help reduce stroke incidents.

- 80% of strokes are preventable.

Actionable item

- Support stroke prevention awareness programs
  - Exercise
  - Eating correctly
  - Programs to stop smoking

80% of strokes are preventable [5] American Stroke Association

Questions?

80% of strokes are preventable [5] , American Stroke Association

References

- [1] Ahmad FB, Cisewski JA, Miniño A, Anderson RN. Provisional Mortality Data — United States, 2020. MMWR Morb Mortal Wkly Rep 2021;70:519–522.
DOI: https://www.cdc.gov/mmwr/volumes/70/wr/mm7014e1.htm?s_cid=mm7014e1_w#F1_down

- [2] American Stroke Association, https://www.stroke.org/en/about-stroke/types-of-stroke/ischemic-stroke-clots

- [3] American Stroke Association, https://www.stroke.org/en/about-stroke/types-of-stroke/hemorrhagic-strokes-bleeds

- [4] American Stroke Association, *Explaining Stroke*, pages 1-20, https://www.stroke.org/-/media/stroke-files/stroke-resource-center/brochures/explaining_stroke_brochure_6_25_19.pdf?la=en

- [5] American Stroke Association, https://www.stroke.org/en/about-stroke

- [6] Stroke Prediction Dataset, *11 clinical features por predicting stroke events*, https://www.kaggle.com/fedesoriano/stroke-prediction-dataset

80% of strokes are preventable [5] American Stroke Association