Stroke Prediction Project

Problem: Predicting Stroke Occurrence

Team Members:
- Walaa Youssef
- Rocio Sampayo
- Perla Lampinstein
- Lorena Olalde Rios

Introduction:
According to the Center for Disease Control and Prevention, in 2021, strokes accounted for 1 in 6 cardiovascular disease-related deaths. Within the United States, a stroke occurs every 40 seconds, resulting in a stroke-related death every 3 minutes and 14 seconds.

- Every year, over 795,000 Americans experience a stroke, with 610,000 being first or new occurrences.
- Notably, nearly 1 in 4 strokes (about 185,000) affect individuals with a history of a previous stroke.
- Approximately 87% of all strokes are ischemic, involving the blockage of blood flow to the brain.

The financial impact of strokes in the U.S. reached nearly $56.5 billion between 2018 and 2019, encompassing healthcare costs, stroke treatment medications, and lost workdays.

Long-term Disability and Stroke: Stroke stands as a leading cause of serious long-term disability, notably reducing mobility in over half of stroke survivors aged 65 and older.2

Stroke Statistics by Race and Ethnicity: While stroke is a leading cause of death for Americans, the risk varies among different racial and ethnic groups. Non-Hispanic Black adults face nearly double the risk of a first stroke compared to White adults, and both non-Hispanic Black adults and Pacific Islander adults exhibit the highest stroke death rates.1 The stroke death rate increased from 38.8 per 100,000 in 2020 to 41.1 per 100,000 in 2021.1

Age-Related Stroke Risk: Although stroke risk generally increases with age, strokes can occur at any age. In 2014, 38% of individuals hospitalized for strokes were under the age of 65.3

Importance of Early Action: Recognizing warning signs and symptoms of stroke is crucial for prompt action. In one survey, 93% of respondents identified sudden numbness on one side as a symptom, but only 38% were aware of all major symptoms and knew to call 9-1-1

during a stroke.Patients reaching the emergency room within 3 hours of initial symptoms often experience less disability three months post-stroke compared to those with delayed care.

Dataset:

The dataset we will be using can be found on the following link:

https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

Project Objective:

The primary objective of the Stroke Predictor project is to develop an accurate and reliable predictive model that can assess an individual's risk of experiencing a stroke. This model will be designed to analyze diverse datasets, including demographic information, lifestyle factors, medical history, and genetic predispositions, to identify patterns and correlations associated with stroke occurrence.

Tools:

- Python Pandas
- Tableau
- Jupyter Notebook

Libraries:

- Scikit-learn
    - ❖ sklearn.model_selection
    - ❖ sklearn.ensemble
    - ❖ sklearn.preprocessing
    - ❖ sklearn.metrics
- TensorFlow

Database:

- PostgreSQL

Visualizations: Below are examples and are subject to change.

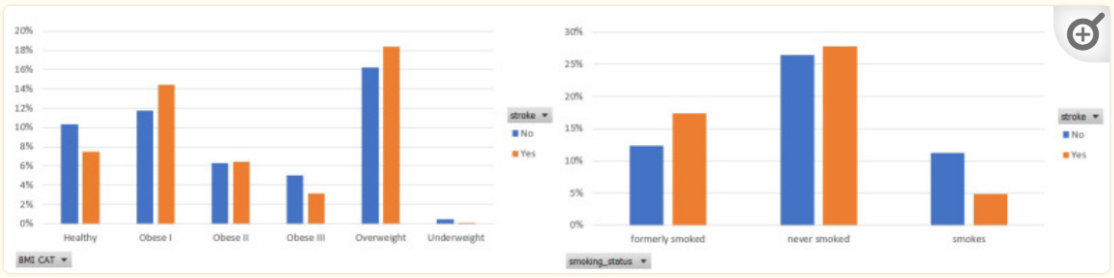

Figure 3

Participants distribution per BMI category and smoke status in the balanced dataset.

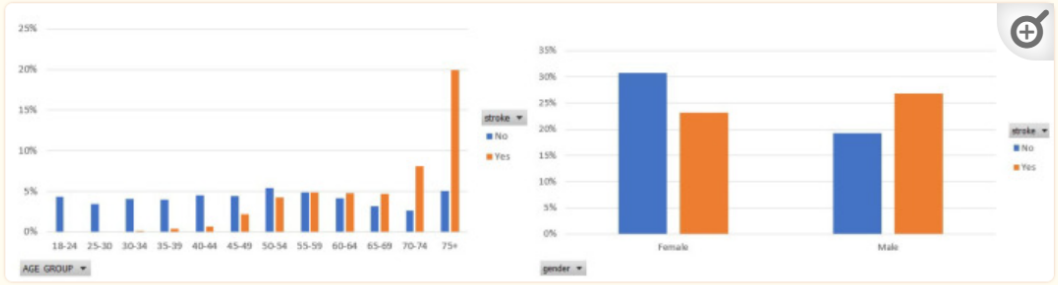

Figure 1

Participants distribution per age group and gender type in the balanced dataset.