

## Abstract

- Our project will be utilizing a dataset that helps predict if a person is at risk of having a stroke or not by looking at specific traits. Our overall motivation for choosing a brain stroke dataset is to understand the variables that contribute to brain strokes
- Finding clear patterns and differences between those who are and those who are not prone to strokes would greatly assist in discovering what specific features cause strokes.

## Background

- A stroke is a medical condition in which poor blood flow to the brain causes cell death. The main risk factor for stroke is high blood pressure. Other risk factors include high blood cholesterol, tobacco smoking, obesity, diabetes mellitus, a previous TIA, end-stage kidney disease, and atrial fibrillation.
- The variables in our specific dataset include gender, age, if the patient has hypertension or not, if the patient has heart disease or not, marriage status, type of work, residence type, average glucose level in the blood, body mass index, and smoking status.

## Methods

Insert methods here

- **Logistic Regression**
  - **a good model for classifying whether or not a person has a stroke or not.**
- **SVM**
  - **good to use due to its simplicity to implement and understand**
- **Random Forest**
  - **this model is versatile, mildly-complex, and uses bagging to generate its own datasets and produce a prediction.**
- **Perceptron Modeling**
  - **single layer neural network used for supervised binary classification.**

## Results



- The overall accuracy of each model was roughly the same with some minor variance, however our accuracy for classifying the positive class was significantly low due to imbalance in the dataset.

## Conclusion

- We were not able to garner a predictive model that worked correctly with our dataset to correctly predict variable weight and significance.

## Future Direction

- To further enhance our findings we would need to increase our sample size of patients who had strokes, because our sample size of stroke patients was only 5%.

## Acknowledgments

Bandi, Vamsi. *Prediction of Brain Stroke Severity Using Machine Learning* International Information and Engineering Technology Association, 2020.