# CHAPTER ONE

# INTRODUCTION

## 1.1 Background to study

A stroke is a medical emergency that occurs when the blood supply to the brain is blocked by a clot or when a blood vessel in the brain ruptures (Xiao et al., 2024). Strokes are the second leading cause of death worldwide and the fifth most common in the U.S. Also, A stroke is a medical emergency that happens when something prevents your brain from getting enough [blood](https://my.clevelandclinic.org/health/body/24836-blood" \t "https://my.clevelandclinic.org/health/diseases/_blank) flow (What Are Stroke Symptoms?, n.d.). The risk of stroke rises markedly with age, with 67% of strokes occurring in individuals over 70 years old (global regional, 2021). A stroke can cause different symptoms depending on which area of your brain it affects. Some of the most common symptoms include: [aphasia](https://my.clevelandclinic.org/health/diseases/5502-aphasia" \t "https://my.clevelandclinic.org/health/diseases/_blank) (trouble speaking or a complete loss of speech), b[lurry vision](https://my.clevelandclinic.org/health/symptoms/24262-blurred-vision" \t "https://my.clevelandclinic.org/health/diseases/_blank) or [double vision (diplopia)](https://my.clevelandclinic.org/health/diseases/22203-diplopia-double-vision" \t "https://my.clevelandclinic.org/health/diseases/_blank), n[ausea and vomiting](https://my.clevelandclinic.org/health/symptoms/8106-nausea--vomiting" \t "https://my.clevelandclinic.org/health/diseases/_blank), n[eck stiffness](https://my.clevelandclinic.org/health/symptoms/24477-stiff-neck" \t "https://my.clevelandclinic.org/health/diseases/_blank), [passing out or fainting](https://my.clevelandclinic.org/health/diseases/17536-syncope" \t "https://my.clevelandclinic.org/health/diseases/_blank), s[eizures](https://my.clevelandclinic.org/health/diseases/22789-seizure" \t "https://my.clevelandclinic.org/health/diseases/_blank), s[lurred or garbled speaking (dysarthria)](https://my.clevelandclinic.org/health/diseases/17653-dysarthria" \t "https://my.clevelandclinic.org/health/diseases/_blank),sudden worsening or loss of your senses (including [vision](https://my.clevelandclinic.org/health/articles/21204-vision" \t "https://my.clevelandclinic.org/health/diseases/_blank), [hearing](https://my.clevelandclinic.org/health/articles/17054-hearing" \t "https://my.clevelandclinic.org/health/diseases/_blank), [smell](https://my.clevelandclinic.org/health/diseases/21859-anosmia-loss-of-sense-of-smell" \t "https://my.clevelandclinic.org/health/diseases/_blank), [taste](https://my.clevelandclinic.org/health/diseases/21850-ageusia-loss-of-sense-of-taste" \t "https://my.clevelandclinic.org/health/diseases/_blank) and touch), Weakness or [paralysis](https://my.clevelandclinic.org/health/diseases/15345-paralysis" \t "https://my.clevelandclinic.org/health/diseases/_blank) on one side of your face and body. There are three main types of strokes (What Are the Different Types of Strokes?, 2016), Transient ischemic attack (TIA): a temporary blockage of blood flow to part of your brain Ischemic stroke: a blood clot blocks blood flow to part of your brainand Hemorrhagic stroke: a blood vessel in your brain ruptures or breaks, spilling blood into the surrounding tissues (Rachel Nall, 2024). Some of the main risk factors of stroke are, high blood pressure, diabetes, atrial fibrillation, high cholesterol, age, lifestyle factors, family history, ethnicity e.t.c.

A helpful tool to identify if someone is having a stroke is the FAST set of [warning signs](https://www.verywellhealth.com/stroke-symptoms-in-women-5214356) (Jividen, 2024) which include: “F” for face drooping: Does the person's face droop on one side, or does their smile look uneven? “A” for arm weakness: Can the person raise and hold both arms out equally? Does one drift down? “S” for speech difficulty: Is the person having difficulty speaking, including slurring their words, using nonsensical or garbled speech, or are they not able to speak at all? “T” for time to call 911: If the answer is yes to any of these questions, call 911 immediately. Normally when a patient has had a stroke, a computerized tomography (CT) scan immediately provides a diagnosis. In the case of ischemic stroke, magnetic resonance imaging (MRI) is efficient. Other ancillary diagnostic tests are carotid triplex and cardiac triplex. Strokes can be severe (extensive) or mild. In the vast majority of cases, the first 24 h are crucial. The diagnosis will highlight the treatment, which is usually pharmaceutical, and, in a few cases, surgical. Intubation and mechanical ventilation in the intensive care unit are necessary when the patient has fallen into a coma (Dritsas and Trigka, 2022).

Recently, machine learning algorithms have shown great promise in accurately predicting stroke risk based on various The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi . clinical risk factors. By leveraging these algorithms, clinicians can identify high-risk patients and intervene early, potentially reducing the number of stroke-related complications and improving patient outcomes (Mridha et al., 2023). Different machine learning classifiers such as Random Forest, Support Vector Machine, Decision Tree, and Logistic Regression are used in medical fields for many years making correct analyses and predicting accurate results relying on patterns with the big imbalanced dataset (Biswas et al., 2022).

The [trouble with accuracy](https://towardsdatascience.com/accuracy-paradox-897a69e2dd9b" \t "https://encord.com/blog/an-introduction-to-balanced-and-imbalanced-datasets-in-machine-learning/_blank) is that it’s not necessarily a good metric for determining how well a model will predict outcomes. It’s counterintuitive (hence the paradox), but depending on the data that it encounters during training, a model can become biased towards certain predictions that result in a high percentage of accurate predictions but poor overall performance. A model might report having very accurate predictions, but, in reality, that accuracy might only be a reflection of the way it learned to predict when trained on an imbalanced dataset (Nikolaj Buhl, 2022). In the article,”A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach” Ten machine learning classifiers after being verified showed more than 90% accurate results before balancing the data and four classifiers displayed more than 96% accurate results after data-balancing using the oversampling method.

This research will use the SASMOTE (Self-inspected Adaptive Synthetic Minority Over-sampling Technique) data balancing technique, to rule out the issue of imbalanced data, in other words prevent the model from becoming biased towards one class. In other words, the model will no longer favor the majority class just because it contains more data.

## 1.2 Motivation of study

With machine learning and artificial intelligence in the picture, we are gradually moving toward a highly automated and intelligent system of life where a wide range of tasks—from healthcare to agriculture, finance, and even daily decision-making—can be performed with minimal human intervention. The potential of these technologies to transform our world is immense, particularly in critical sectors such as medicine, where early diagnosis and prevention can save lives.

However, despite the impressive capabilities of machine learning models, their effectiveness heavily depends on the quality and balance of the data used to train them. One of the major challenges in predictive modeling is dealing with imbalanced datasets, where one class significantly outweighs the other. For example, in stroke prediction, the number of people who have never experienced a stroke may far exceed those who have. If not properly addressed, this imbalance can cause the model to become biased toward the majority class, leading it to make seemingly accurate predictions that are, in fact, misleading.

Therefore, it's not enough for a model to be accurate in general; it must be precise in identifying minority classes that often represent the most crucial outcomes.

Recognizing and addressing data imbalance through proper preprocessing techniques such as resampling, synthetic data generation, or using cost-sensitive algorithms is essential to ensure fair and effective predictions.

## 1.3 Statement of the problem

With machine learning and artificial intelligene in the picture we are getting closer to a system of life where the stroke disease can be detected early in a persons life, this is definitely beneficial to the population of the world at large. but following up various research in line with machine learning and medicine it has been noticed that an imbalanced data can lead to misguided answers by a predictive machine learning model, if this is not properly looked into some errors might go unnoticed as the model might appear to be accurate in it’s prediction.

## 1.4 Aim and Objectives of Study

This project aims to develop a stroke predictive system that uses the SASMOTE data balancing technique to oversample genetic data of stroke patients, with the following objectives:

1. Gathering of stroke patient genetic data from opensource sites
2. Training a machine learning model combined with the sasmote data balancing method
3. Developing a simple user interface

## 1.5 Scope and Limitation of Study

While this study aims to support stroke prediction using machine learning, it is important to understand its current limitations. Although the goal is to improve early detection and assist healthcare systems, this model is not yet ready to be used as a real medical tool. In real-life situations, especially in hospitals, any system used to make health decisions must be carefully tested and approved by medical professionals.

This research is being developed without the involvement of a full medical team, which means it hasn’t been reviewed by doctors or tested in a clinical environment. Also, the model was trained using open-source data, which might not include enough examples from all age groups, regions, or health backgrounds. This could affect how well the model performs in different situations.

Therefore, this study should be seen as a step toward developing better tools in the future. It highlights the potential of AI in healthcare but should only be used for research and learning purposes for now, not for making real medical decisions.