1. **Abstract**

Stroke, a fatal non-communicable disease of any age, kills more people than AIDS, Tuberculosis and Malaria put together in each year. WHO estimated around 6.2 million deaths because of stroke in 2008. As the incidence, prevalence, mortality and disability rates are increasing, overall stroke burden has increased globally. Almost 70% of patients are unaware of their mild stroke, 30% seek medical attention lately and another 30% suffer from recurrent stroke, before seeking attention. Data mining, with its several techniques for classification and regression, plays a leading role in developing an effective model of risk prediction in the context of healthcare. Even though stroke prevention is a complex medical issue, primary prevention could be feasible by using data mining classification techniques that will assess risk factors to predict the likelihood of the disease among mass people. This work is aimed at providing an analysis of different data mining classification algorithms like Naïve Bayes (NB), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), etc. on a newly created dataset of 435 patient’s risk factors to find the algorithm with the best accuracy to propose a tool for the end users to check risk prediction.

**General Term :** Data Mining and techniques.

**Keywords:** Stroke risk factor, Data mining, KDD, Naïve Bayes, Random forest, Random Tree, Support Vector Machine, ANN, J48, Risk Prediction.

1. **Introduction**

A stroke happens when blood flow is occluded in a part of the brain. The lack of oxygen damages the brain cells that can have potentially disabling effects on the patient. (Services, 2018) When it happens the brain cells divested of oxygen and brain cells begin to die. Then the abilities controlled by that area of the brain such as memory and muscle control are lost. Some strokes affect the muscles used to urinate. There are two types of stroke. They are:

1. Ischemic Stroke
2. Haemorrhagic Stroke

**1. Ischemic Stroke**

An Ischemic stroke is the most common type. It occurs when blood supply is cut off to part of the brain. It accounts for the majority of all strokes. Approximately 85% of strokes are ischemic caused by vascular occlusion.

An ischemic can occur because of lesions caused by atherosclerosis. These lesions may form in the small arteries of the brain and they can block blood flow to the brain. (Services, 2018)

**2. Hemorrhagic Stroke**

A hemorrhagic stroke is a different kind of stroke caused by bleeding in the brain. It happens when a blood vessel breaks and bleeds into the brain. Blood spills into or around the brain and creates swelling and pressure, damages cells and tissue in the brain.

Hemorrhagic stroke is caused by a rupture in a weakened blood vessel in the brain. Hemorrhagic stroke account for about 20% of all strokes. (Services, 2018)

In this thesis paper, we will predict the risk of stroke. A man who has not yet stroked will talk about the possibility of having a stroke in the future. It is a more challenging task in healthcare sectors to predict the diseases from the voluminous medical databases. At present, data mining techniques will help us a lot to predict risk. Data mining techniques which include classifications, clustering, association rule mining for finding risk prediction. In this research work, Naïve Bayes Support Vector Machine (SVM) classifier algorithm is used for stroke risk prediction.

The brain is one of the largest and most complex organs in the human body. It is made up of more than 100 billion nerves that communicate in trillions of connections called synapses. [1] It is the central organ of the human nervous system. It controls most of the activities of the body, processing, integrating and coordinating the information it receives from the sense organs and making decisions as to the instructions sent to the rest of the body. Because of an ischemic stroke, brain cells start to die or damage. So, symptoms occur in the body parts (face, eyes, arms, legs etc.) that these brain cells control. There is the number of factors which increase the risk of having an ischemic stroke. Some of them are listed below:

* Age and gender
* High blood pressure
* Diabetes
* Ischemic heart diseases
* Smoking
* The family history of stroke
* Stress and depression
* Overweight and obesity
* Abnormal cholesterol levels