**RISK PREDICTION OF STROKE USING DATA MINING**

**Introduction:**

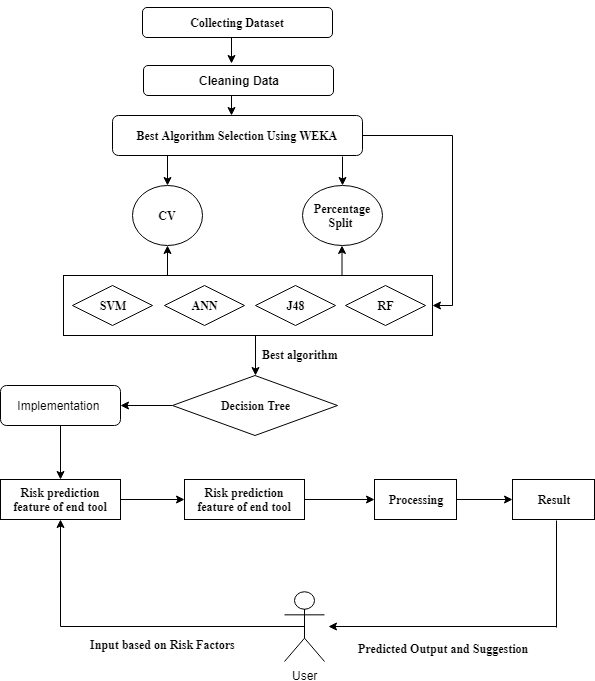
In healthcare industry, data mining plays an important role in predicting diseases. For detecting a disease number of tests should be required from the patient. But using data mining technique the number of tests can be reduced. This reduced test plays an important role in time and performance [1]. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis [2].

Knowledge Discovery is the nontrivial process of extracting implicit, novel, and useful information from large volume of data. Major applications of Knowledge Discovery and Data Mining in healthcare fall into four categories: (a) Clinical Medicine: Modern hospitals and clinical centers surpassed their traditional role as a place for diseases’ diagnosis and treatment and now acting as a mass database and a source of complex clinical, laboratory, equipment use, and drug management data which can be analyzed for disease diagnosis and decision making; (b) Public Health: including early outbreak detection, healthcare and syndromic surveillance; (c) Healthcare Text mining: including mining medical literature, as well as mining clinical data such as patients’ clinical records; and (d) Healthcare Policy and Planning: including detecting expensive clinical profiles among patients diagnosed with a specific chronic illness which has a high disease’s burden such as diabetes [3].

Data mining plays an important role in predicting diseases. Through different types of data mining technique heart disease, Diabetes and Breast cancer and so on can be predicted easily [4].

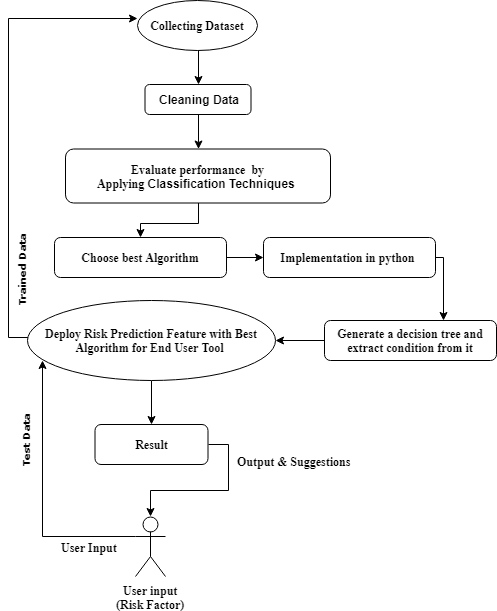
A stroke is a sudden interruption in the blood supply of the brain. Most strokes are caused by an abrupt blockage of arteries leading to the brain ([ischemic stroke](http://www.strokecenter.org/patients/about-stroke/ischemic-stroke/)).  Other strokes are caused by bleeding into brain tissue when a blood vessel bursts ([hemorrhagic stroke](http://www.strokecenter.org/patients/about-stroke/intracerebral-hemorrhage/)). Because stroke occurs rapidly and requires immediate treatment, stroke is also called a brain attack. When the symptoms of a stroke last only a short time (less than an hour), this is called a transient ischemic attack (TIA) or mini-stroke. The effects of a stroke depend on which part of the brain is injured, and how severely it is injured. Strokes may cause sudden weakness, loss of sensation, or difficulty with speaking, seeing, or walking [5].

If the prediction of risk factors is possible you may be able to lower your risk factors and prevent or delay a stroke[6].

****

**Fig 1:** Proposed System Architecture

**System Architecture:** Our system architecture has been delineated in fig.1. Initially, an original dataset including the risk factors of 435 people has been used for selecting the best prediction algorithm. Dataset was cleaned then as most of the people don’t have any idea of their Cholesterol level and all of them are in same races. Then the processed dataset was feed to the database (which will be used as a trained dataset for the end-user tool) and to the classification algorithms for simulation. The performance accuracy has been evaluated using 10-Fold Cross Validation and Percentage Split techniques. Finally, according to the best accuracy, the best algorithm will be chosen for enabling the risk prediction feature of the tool.

****

**Fig 2:** Flow Chart of Methodology:

**Flow Chart:** A flow chart of our methodology has been portrayed in fig 2. Our methodology works according to this flow chart. Here in the first step dataset will be collected and Cleaned. After that Dataset will be sent to the preprocessing stage to be as binary data from binary nominal data. Then seven most popular classification techniques will apply on the dataset. The performance of all algorithms will be evaluated by 10-Fold Cross Validation and 80:20 percentage split techniques. With the best performed algorithm will be selected and also be used for developing a reliable tool with risk prediction feature for the end user. Then end user will give the answer of all questions from questionnaire form as input. Here the train dataset will be updated every time.

**References:**

1. Online.

Available: <https://www.researchgate.net/publication/319851535_Multi_Disease_Prediction_using_Data_Mining_Techniques>

1. Online

Available:

(Siri Krishan Wasan, 2006)

1. Online

Available:

(AICN, 2014)

1. Online

Available: <https://www.researchgate.net/publication/319851535_Multi_Disease_Prediction_using_Data_Mining_Techniques>

1. Online

Available:

(Patient, n.d.)

1. Online

Available:

(National Heart, n.d.)

(Patient, n.d.)

# References

*AICN*. (2014, April 30). Retrieved from AICN: http://www.ieeeottawa.ca/aicn/data-mining-and-knowledge-discovery-in-healthcare-and-medicine/

Kamaraj, K. (2016). Multi Disease Prediction using Data Mining Techniques. *Research Gatee*, 1-2.

National Heart, L. a. (n.d.). *patient, patient and family*. Retrieved from National Heart, Lung and Blood Inistitution: https://www.nhlbi.nih.gov/health-topics/stroke

*Patient*. (n.d.). Retrieved from The internet stroke Center: http://www.strokecenter.org/patients/about-stroke/what-is-a-stroke/

Siri Krishan Wasan, V. B. (2006). The Impact of Data Mining Techniques On Medical Diagnostics. *Data Science Journal*, 119-126.