Our proposed system architecture has been depicted in fig.1 initially, a vast amount of data is collected from 635 peoples. The pre-processing stage was associated following the missing tuple handling method. After cleaning data, risk factors of 606 peoples have been used for the experiment. The original dataset including the risk factors of 606 people has been used for selecting the best prediction algorithm using WEKA. The dataset was feed to the classification algorithms for simulation. After evaluating the performance accuracy using 10-Fold Cross-Validation and 80:20 Percentage Split techniques we got Decision Tree as best algorithm. We have enabled an end-user tool for risk prediction which generates result from the Decision Tree algorithm as it gives us best accuracy. Then the generated output will be shown to the end-user and the processed dataset will feed to the database (which will be used as a trained dataset for the end-user tool) which will redouble the dataset.

Early stage risk prediction of stroke using data mining techniques.

Healthcare Leadership