GUI Based Model for Stroke Prediction

Jeena R S, Sukesh Kumar A

Abstract*: The innovations in the field of artificial intelligence have paved way to the development of tools for assisting physicians in disease diagnosis and prognosis. Stroke is a leading cause of disability in developing countries like India. Early diagnosis of stroke is required for reducing the mortality rate. Research shows that various physiological parameters carry vital information for the prediction of stroke. This research work focuses on the design of a graphical user interface (GUI) for the prediction of stroke using risk parameters. Data collected from International Stroke Trial database was successfully trained and tested using Support vector machine (SVM). The linear kernel of SVM gave an accuracy of 90 %. This work has been implemented in MATLAB which can be used to predict the probability of occurrence of stroke.

Index Terms: Stroke, Graphical User Interface (GUI), Support Vector machine (SVM)

I. INTRODUCTION

Stroke is a global health problem and one of the leading causes of adult disability. Stroke is the fourth major cause of death in India. In Trivandrum, the capital of the Indian state of Kerala, the incidence rate of stroke per year is 135.0 and 138.0 (per 1,00,000 inhabitants) in urban and rural community respectively [1]. Stroke or Cerebrovascular accident is a medical condition due to inadequate supply of blood to the brain cells, which damages them and results in their death. Blood flow may be interrupted either due to a clot in the blood vessel or a blood vessel rupture .Stroke caused due to a clot in the blood vessel is normally referred to as Ischemic stroke [2] and that due to a rupture of blood vessel is referred to as Hemorrhagic stroke. The inadequacy of oxygen and nutrients to the brain cells is referred to as ischemia which ultimately leads to their death. Ischemic stroke accounts for around 85 % of all strokes. Its incidence is accelerating in developing countries due to unhealthy lifestyles. The high metabolic rate of brain, sensitivity to changes in blood flow and dependence on continuous blood flow make strokes so dangerous. Prognosis of stroke is still a challenge and is very much appreciable in the field of medical research. Machine learning techniques are definitely worth exploring in predicting the possibility of stroke. Machine learning is a method of data analysis that automates logical model building. The iterative phase of machine learning is important because as models are exposed to new data samples, they are able to adapt independently. They learn from previous computations to produce consistent and repeatable decisions. Learning stops when the algorithm achieves a suitable level of performance. Support Vector Machines are extensively used in classification tasks for their ability to model the complex system.

II. RISK FACTORS

CVDs are chronic diseases that occur by long-term combined effects of risk factors. Stroke is enhanced by the increasing occurrence of modifiable risk factors. A risk factor is any characteristic of an individual that increases the probability of developing a disease.

A number of risk factors can increase the probability of occurrence of stroke. Lifestyle risk factors include diet, cigarette smoking habits [4], overweight and obesity, physical inactivity, alcohol consumption, family and genetic factors, age, sex, drug use , race, oral contraceptive use, geographic location, season, climate and socioeconomic factors whereas medical conditions include Atrial fibrillation[5], Blood pressure[3], Diabetes mellitus, Cholesterol, Mitral valve disease, elevated fibrinogen concentration, Sickle cell disease, Hyperlipidemia, Transient ischemic attack (TIA), migraine headaches and migraine equivalents. High blood pressure, heart disease and diabetes are the highest risk factors of stroke. But, they often do not cause symptoms in their earlier stages. Total count of risk factors is directly related to probability of stroke occurrence. There are several informative studies showing the evidences of using physiological parameters as risk factors for predicting near-term risk of stroke**.** The statistical association between a risk factor and the outcome can be tested using various regression models.

III. LITERATURE SURVEY

Saangyong Uhmn et al., [6] have presented the machine learning techniques, SVM, decision tree, and decision rule to predict the vulnerability of the liver disease, Chronic hepatitis from single nucleotide polymorphism data. The experimental results have shown that decision rule is able to distinguish chronic hepatitis from normal with the maximum accuracy of 73.20%, whereas SVM with 67.53% and decision tree with 72.68%.

Sumit Bhatia, Praveen Prakash, and G.N. Pillai [7] proposed a method for heart disease classification using Support Vector Machine (SVM) and integer-coded genetic algorithm (GA). Simple Support Vector Machine algorithm was used to find the support vectors in a fast, iterative and efficient way. Cleveland heart disease database was used in this work. For selecting the significant and relevant features and discarding the inconsistent and redundant ones, an integer-coded genetic algorithm was applied.

Xing et al. [8] conducted a survey of 1000 patients, the results show that SVM gives an accuracy of 92.1%, artificial neural networks gives an accuracy of 91.0% and decision trees with 89.6% accuracy. TNF, IL6, IL8, HICRP, MPO1, TNI2, sex, age, smoke, hypertension, diabetes, and survival are the attributes used in this work.

Revised Version Manuscript Received on February 20, 2017.

Jeena R S, Department of Electronics & Communication, College of Engineering, Trivandrum, (Kerala). India. E-mail: jeena_rs@yahoo.com

Dr. Sukesh Kumar A, Associate Professor, Rajiv Gandhi Institute of Development Studies, Vellayambalam, Trivandrum (Kerala). India. E-mail: drsukeshkumar@yahoo.in

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Similarly, Chen et al. [9] compared the accuracy of SVM, neural networks, Bayesian classification, decision tree and logistic regression. Considering 102 input samples, SVM had the highest accuracy of 90.5%, neural networks 88.9%, Bayesian 82.2%, decision tree 77.9%, and logistic regression 73.9%. The work by Alexopoulos, Dounias, and Vemmos [10] was focused on the application of inductive ML methods in medical diagnosis of stroke.

IV. PROPOSED WORK

This research work focuses on the development of a graphical user interface (GUI) model for the prediction of stroke using Support vector machine (SVM) with 12 input parameters. An overview of SVM and GUI design in MATLAB is given in the following sections.

A. Support Vector Machine

SVM is a widely used supervised machine learning algorithm for classification developed by Vladimir N. Vapnik and the current standard incarnation (soft margin)was proposed by Vapnik and Corinna Cortes in 1995 [11]. In pattern classification, given a set of input samples and the corresponding class labels, the aim is to confine the implicit relation among the patters of the same class, so that when a test sample is given, the corresponding output class label is retrieved. It merges linear algorithms with linear or non-linear kernel functions that make it a dominant tool in data mining and medical imaging applications. It outperforms other classifiers even with small numbers of training samples.
Dataset for this work is taken from International Stroke trial Database. [12] Database includes patient information, patient history, hospital details, Country, risk factors and symptoms. After preprocessing, 350 samples are taken in this work. Polynomial, quadratic, radial basis function and linear functions are applied and all give different accuracy. A comparison has been made between classification accuracy of various kernel functions.

Table 1 Shows the Symptoms and Risk Factors Given as Input to SVM.

1	Age
2	Sex
3	Walking symptoms
4	Atrial Fibrillation
5	Face deficit
6	Arm / Hand deficit
7	Leg/ Foot deficit
8	Blood Pressure
9	Dyphasia
10	Hemianopia
11	Visuospatial disorder
12	Cerebellar signs

Table 1. Input Parameters

B. Design of GUI

A graphical user interface (GUI) is a pictorial interface to a program. A good GUI can make programs easier to use by providing them with a consistent appearance and with intuitive controls like pushbuttons, list boxes, sliders, menus, and so forth. In this work, a MATLAB GUI is created using a tool called guide, the GUI Development Environment. This tool allows a programmer to layout the GUI, selecting and aligning the GUI components to be placed in it. Common users can indicate symptoms they are experiencing and get a prediction from the system.

V. PERFORMANCE ANAYSIS

SVM has been implemented with different kernel functions and the appropriate choice of kernel function for detection of stroke has been investigated by the same authors [13]. Table 2 shows the performance metrics for various kernel functions of an SVM classifier.

Table 2. Experimental Results

KERNEL	Accuracy	Precision	Sensitivity	Specificity	F1 Score
Linear	91 %	84.7%	100 %	78.75 %	91.7
Quadratic	81 %	79.6%	87 %	73.4%	83.3%
RBF	59%	89%	27%	96%	41%
Polynomial	87.9%	84.8%	94.7%	80%	89 %

Linear kernel function is a better choice due to better accuracy level. A MATLAB GUI has been developed for the common man to enter the risk parameters and symptoms to predict the probability of occurrence of stroke.

The prediction has been done with the help of LIBSVM. LIBSVM [29] is integrated software for support vector classification, regression and distribution estimation. It supports multi class classification. Also it is possible to evaluate likelihood measures, i.e., either scores or posterior probabilities.

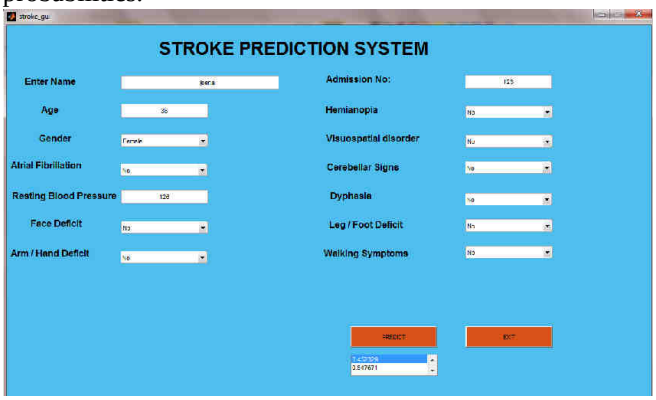

Figure 1. GUI for Stroke Prediction

VI. CONCLUSION

A GUI model for the prediction of stroke has been developed in MATLAB using Support Vector Machine as the classifier. Diagnosis of stroke during initial stages is crucial for timely prevention and cure. This model aids in predicting the probability of stroke based on symptoms and risk factors. Performance of the system can be improved by incorporating more risk factors and symptoms which requires training from a much larger database.

REFERENCES

1. Subha PP,Pillai Geethakumari SM, Athira M, Nujum ZT, Pattern and risk factors of stroke in the young among stroke parients admitted in medical college hospital, Thiruvananthapuram., Ann indian Acad Neurol 2015;18:20-3.
2. Barry L. Zaret, M.D., Marvin Moser, M.D., Lawrence S. Cohen, Chapter 18 *Stroke - Lawrence M. Brass*, M.D.

(pgs 215-234)
3. MacMahon S, Rodgers A. The epidemiological association between blood pressure and stroke: implications for primary and secondary prevention. *Hypertens Res.* 1994;17(suppl 1):S23-S32.
4. Shinton R, Beevers G. Meta-analysis of relation between cigarette smoking and stroke. BMJ
5. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort: the Framingham Heart Study. JAMA. 1994;271:840-844
6. Saangyong Uhmn, Dong-Hoi Kim, Jin Kim, Sung Won Cho, Jae Youn Cheong, "Chronic Hepatitis Classification Using SNP Data and Data Mining Techniques", Frontiers in the Convergence of Bioscience and Information Technologies, FBIT 2007,pp.81 - 86 , 11-13 Oct. 2007B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
7. S. Bhatia, P. Prakash and G.N. Pillai, SVM based Decision Support System for Heart Disease Classification with Integer-coded Genetic Algorithm to select critical features, Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, pp.34-38, 2008.
8. Yanwei Xing, Jie Wang and Zhihong Zhao Yonghong Gao 2007 "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease" Convergence Information Technology, 2007. International Conference November 2007, pp 868-872.
9. Jianxin Chen, Guangcheng Xi, Yanwei Xing, Jing Chen, and Jie Wang 2007 "Predicting Syndrome by NEI Specifications: A Comparison of Five Data Mining Algorithms in Coronary Heart Disease" Life System Modeling and Simulation Lecture Notes in Computer Science, pp 129-135
10. Alexopoulos, E., Dounias, G.D., and Vemmos, K. (1999). "Medical diagnosis of stroke using inductive machine learning". In Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course in Artificial Intelligence-ACAI99, Chania, Greece, 20-23.
11. C. Cortes and V. Vapnik, "Support-vector networks," Machine learning,vol. 20, no. 3, pp. 273–297, 1995..
12. Sandercock, Peter; Niewada, Maciej; Czlonkowska, Anna. (2011). International Stroke Trial database (version 2), University of Edinburgh. Department of Clinical Neurosciences.
13. Jeena R S, Dr Sukesh Kumar A, 'Stroke Prediction using SVM', Proceedings on International Conf. on Control, Instrumentation, Communication and Computational Technologies, (ICCICCT-2016),Tamil Nadu

Jeena R S is currently working as Assistant Professor in the Department of Electronics and Communication Engineering, at College of Engineering Trivandrum, Kerala. She received her M.tech degree from University of Kerala in the year 2004. She is currently doing her research in the area of biomedical imaging.

Dr. Sukesh Kumar A is currently working as Associate Professor, Rajiv Gandhi Institute of development Studies, Vellayambalam. His research interest is mainly in the field of biomedical image processing. He has produced a number of PhDs and have numerous national and international journal publications.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.