International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET207134

# A Survey on Detection of Stroke Using Various Machine Learning Approaches

**Priyanka S R[1], Deepa S S[2]**
[1]M. Tech CSSE Scholar, Government Engineering College, Idukki, India
[2]Associate Professor, Government Engineering College, Idukki, India

## ABSTRACT

Stroke is a sudden interruption of blood supply to brain. This is due to lack of oxygen caused by blockage of blood flow. Machine learning (ML) considered as a branch of artificial intelligence which is effective in spotting complex patterns in large medical data. ML is well suited in large medical applications especially those that depends on complex protomic and genomic measurement. There are several ML techniques that are used for various disease detection and predictions. This paper mainly focused on such techniques and feature selection mechanism that are useful for detecting stroke.

Keywords : Stroke, Machine Learning, Feature Selection

## I. INTRODUCTION

According to World Health Organization, stroke events are rapidly developing signs of focal disturbance of cerebral function. Stroke symptoms are paralysis, numbness or weakness in the arm, face, and leg, especially on one side of the body, trouble speaking, understanding speech, vision problems( trouble seeing in one or both eyes with vision blackened, vision blurred, or double vision), trouble walking, loss of balance or coordination, dizziness, sudden headache with an unknown cause. The growth of the high-throughput technologies has been outcome as exponential growth in the harvested data with regard to both dimensionality and sample size. Efficient and effective management of these data have become a challenging factor. Traditional manual management of these data sets appears impractical.

Thus techniques such as machine learning and data mining were developed to automatically discover knowledge and identify patterns from these data. Machine learning is provided with effective algorithms, applications, and frameworks to attain greater detection accuracy and value to enterprise data sets and contributing to diverse strategies. Various machine learning techniques and algorithms are discussed and finds the best approach that is suitable for the detection of stroke. Feature Selection is the process of selecting subsets from original features. Feature subset optimality is measured by means of evaluation criteria. Various feature selection techniques such as filter approach, wrapper approach, embedded approach and hybrid approach are also discussed.

## II. Machine Learning

**Machine learning** is a division of artificial intelligence that provides systems the ability to automatically learn and improve from experience. Machine learning focuses on the growth of computer programs that can acquire data and use it to learn for themselves. Steps to apply machine learning includes data collection, data exploration, training the model,

model performance evaluation, and performance improvisation.

### Machine Learning Algorithms

1. Supervised
2. Unsupervised
3. Semi Supervised
4. Reinforcement

**1. Supervised learning Algorithm:** As the name indicates the presence of a supervisor as a teacher. It is also known as predictive models. In this, we can use different attributes of a dataset to predict any other input value. It is a learning in which we teach or train the machine using data which is well labelled. That is data are already tagged with the correct label. After that, the machine is provided with a new set of data, so that, supervised learning algorithm analyses the training data and produces a correct outcome from labelled data. Decision tree and SVM are some of the examples of the supervised learning. The figure 1 shows the flow of supervised learning algorithm.

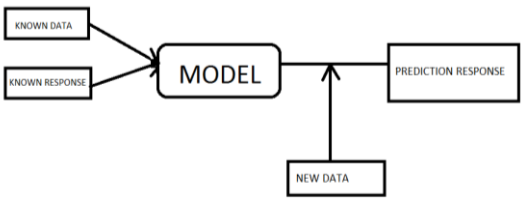


Fig. 1. Supervised learning algorithm

**2. Unsupervised learning Algorithm:** It is a machine learning technique, where you do not need to supervise the model. It helps you to find all kinds of unknown patterns in data. It is also known as descriptive models. The aim is to discern the data and get few structures within it. Clustering and Association are two types of unsupervised learning algorithms. It works better on transactional data.

**3. Semi Supervised learning Algorithm:** It falls between unsupervised learning and supervised learning. It is a class of machine learning tasks and techniques that also make use of unlabelled data for training. Classification, regression and predication are examples of semi supervised learning.

**4. Reinforcement learning algorithm:** In this type, machine is trained to take exact decisions based on the business requirement with the objective to maximize the efficiency. In the lack of training dataset, it is bound to learn from its experience. It allows machines and software agents to automatically discover the ideal behavior within a specific context, in order to increase its performance.

B. MACHINE LEARNING IN VARIOUS FIELDS

Machine learning allows machines to make decisions from massive data. Companies such as Google, Amazon, Accenture, Toyota, Hitachi, and Tesla & Johnson have embraced machine learning at immense scale and improved their products and services. Small companies also developed innovative applications using machine learning. Amazon launched machine learning platform in 2015 and exhibited more supportive reviews to customers, Google used the ML to translate text in 27 languages. Tesla adopted ML in Auto pilot. Machine learning is being used in a wide range of application domains and few foretastes are listed here.

- Health care
- Finance
- Government
- Marketing
- Transport

## III. Feature Selection

Feature selection is the process of selecting a subset of original features based on the certain benchmark. It is an important technique and frequently used as a reduction technique for machine learning. Feature selection has been an active research area over the years in the fields such as machine learning and data mining.

Feature selection process basically consist of four steps:

- Subset Generation
- Subset Evaluation
- Stopping criterion
- Result Validation

Subset generation is a search procedure. It produces feature subsets for evaluation based on a certain search criteria. Each candidate subset is evaluated and then compared it with the previous best one according to a certain evaluation criterion. It replaces the previous one, if the new subset turns out to be better. Subset generation and evaluation process is repeated until a specified stopping criterion is achieved, then the selected best subset needs to be validated by preceding knowledge or through different evaluation. Figure 2 depicts the key steps of feature selection[2].

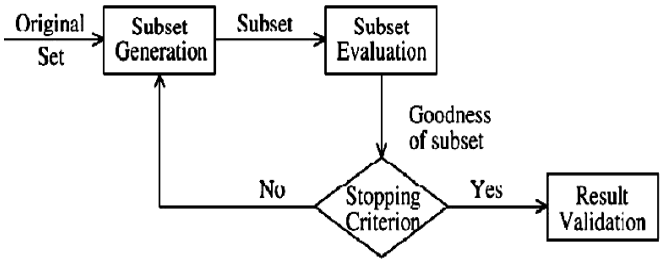


Fig. 2. Key steps of feature selection

A. OBJECTIVES OF FEATURE SELECTION

- To improve the performance of the model and avoid over fitting
- To provide faster and more cost-effective models
- To gain an intense insight into the processes that generated the data.

B. FEATURE SELECTION TECHNIQUES

Feature selection techniques can be organized into three categories, depending on how they combine the feature selection search with the construction of the classification model: filter methods, wrapper methods, and embedded methods and hybrid methods.

1. **Filter method**: It assess the relevance of features by looking only at the intrinsic properties of the data. In this method, a feature score is determined, and low-scoring features are removed. After that, this subset of features is given as input to the classification algorithm.

ADVANTAGES

a. they are easily scalable to very high-dimensional datasets,
b. Computationally simple and fast
c. Independent of the classification algorithm

Filter methods are divided into two

a. Univariate: It is fast, simple and independent of variable. Examples are Euclidean distance, information gain and gain ratio.
b. Multivariate: It models feature dependencies, independent of the classifier and have better computational complexity. Examples are correlation based feature selection and markov blanket filter.

2. **Wrapper method:** Wrapper methods are based on algorithms. A groundwork procedure within the area of all possible feature subsets is outlined. Numerous subsets of features are generated and evaluated. The analysis of a selected subset of features is obtained by training and testing a classification model, rendering this approach to a specific classification algorithm. A search algorithm is then 'wrapped' round the classification model.

ADVANTAGES

a. Embody the interaction between feature subset search and model selection
b. ability to take into account feature dependencies

Wrapper methods are divided into two

a. Deterministic: It is simple, interacts with the classifier, Models feature dependencies. Examples are ordered forward search, sequential backward elimination.

b. Randomized: It is less prone to native optima, interacts with the classifier, models feature dependencies. Examples are simulated annealing, genetic algorithm and randomized hill climbing.

3. **Embedded method:** In this the search for an optimal subset of features is built into the classifier construction, and may be seen as a search in the combined space of feature subsets and hypotheses. Like wrapper approaches, embedded approaches are specific to a given learning algorithm. The advantage is that they embody the interaction with the classification model, whereas at an equivalent time it's less computationally intensive than wrapper methods.

ADVANTAGES

a. Interacts with classifier
b. higher machine complexions
c. Models feature dependencies

4. **Hybrid method:** The hybrid model take advantage of the two models by exploiting their totally different analysis criteria in several search stage.

## IV. Literature Survey

We used 12 technical papers for literature review and inserted based on the importance and priority.

Dr. V. Ilango, et al. [1]: This paper depicts the study on various prediction techniques and tools for Machine Learning. The applications of Machine Learning in various domains are also mentioned here by highlighting on its necessary role in health care industry. It explains the process of machine learning in detail, machine learning algorithms like supervised learning, unsupervised learning, semi supervised and reinforcement learning algorithms. It explains about various steps to apply machine learning to data, then various machine learning tools (platform tools, library tools, Platforms versus Libraries, Graphical User Interfaces versus Command Line Interface versus Application Programming Interfaces, Local versus Remote) and machine learning in various health care provisions are also mentioned elaborately.

A.K.Shafreen Banu et al. [2]: This survey projected with several feature selection approaches. Basic method, merits and demerits of feature selection available at recent innovation are explained. Feature selection is a noteworthy analysis topic of near future and active field to be extended owing to size of datasets offered. To boost the performance of the classifier the feature selection optimizes that with the assistance of chosen subsets. No individual algorithm performs well. It has been surveyed that hybrid approaches performs well for numerous datasets.

N. Leibowitz et al. [3]: Purpose of this paper is to guage a novel quantitative approach to the assessment of proprioception deficits in stroke patients. Here an automated protocol is designed and implemented where a magnetic motion tracking system and a sensor attached to each of the patient's hands. It permits registration of trajectories in 3D coordinates. During this the patient's affected and healthy hands are placed below and above a square board respectively. With vision blocked, the patient's affected hand is stirred to at least one of four locations, and so the patient is instructed to calmly position the healthy hand directly top of his/her perceived location of the affected hand. The positional difference between the two hands is recorded by the system mechanically. Several times this procedure is repeated and therefore the magnitude and direction of errors are used to quantify the proprioception deficit. The data for this study was collected in a sample of 22 stroke patients and an age-matched cluster of neurologically intact subjects. Stroke patients had considerably higher mean distance error compared with the control group, and showed higher instability in repeated performance. The system provides a reliable quantitative measure of upper limb proprioception, and considerable advantage over the traditional means applied in the clinic.

G.E. Wang et al. [4]: The comparison of hybrid methods based on SVM is presented in this paper to address feature selection. It specialize in the foremost relevant features for use in representing information so as to delete those features considered as irrelevant. It not only reduced the number of variables but also eliminates the noise inputs. The results of the study finds that the hit rates of hybrid feature selection methods are higher than single methods, especially when the instances equal to both parts. The proposed approaches selects the foremost relevant variables, however not essentially the optimal ones. The four methods employed in the paper achieved similar classification accuracy.

Guixiong Liu et al. [5]: This paper proposes a multi-class classification method of support vector machines based on double binary tree (DBT-SVM) to solve the problems of 'irreversibility', 'error accumulation' and randomicity of classification order in multi-class classification of support vector machines based on binary tree (BT-SVM). According to the method, each Sub-classifier of BT-SVM is modified. After unknown samples are classified by the modified BT-SVM, the negative output of its final sub-classifier is again classified by an Auxiliary BT-SVM so that the misclassified samples mixed in the negative output can be classified correctly. Experimental results shows that the classification accuracy can be improved using DBT-SVM method, while the general classification accuracy does not decrease.

Sancho Salcedo-Sanz et al. [6]: This paper proposes two approaches for feature selection and ranking based on Simulated Annealing (SA) and Walsh analysis, which use a Support Vector Machine (SVM) as classifier. These approaches are inspired by one of the key problems in the insurance sector: predicting the bankruptcy of a non-life insurance company. The prediction is based on accounting ratios that measure the health of the companies. The two approaches, decides about the financial state of each studied company. The proposed methods are applied for predicting the insolvency of non-life insurance companies. Good results are obtained, consisting of low probability of error given by the SVM and discarding noisy and redundant ratios.

Raid Alzubi et al. [7]: Machine learning approach for detecting patients with hypertension disease using SNPs data is presented in this paper. The proposed approach consists of using the CMIM filter feature selection method for selecting a subset of the most informative SNPs. A supervised classification step distinguish between affected samples and healthy samples of SNP data. Supervised classification experiments utilizing five classification algorithms(k-NN, LDA, NB, ANN, SVM) are been used, and showed that the ensemble approach using the SVM, 5-NN, and NB classifiers achieved the highest classification accuracy.

Yvan Saeys et al. [8]: This paper reviewed the main contributions of feature selection research in a set of well-known bioinformatics applications. Two main issues in the bioinformatics domain are: the large input dimensionality, and the small sample sizes. To overcome these problems, feature Selection techniques have been designed by researchers from bioinformatics, machine learning and data mining. The main aim of this review is to make aware of the benefits, and even the necessity of applying feature selection techniques.

T. Kalaiselvi et al. [9]: This paper describes the algorithm of GSO and reviewed GSO based on the several field like clustering, optimization problem, multicast routing problem (MQMR) problem and multi-robot based problems. The result confirms that the outcomes of GSO are better when compared to the other optimization methods namely, PSO, ACO and GA.

Debasish Ghose et al. [10]: This paper presents multimodal function optimization, using a GSO algorithm, with the applications to collective robotics.

GSO is much same to ACO and PSO but with important differences. An important feature of the algorithm is the use of an adaptive local-decision domain, which is used effectively to detect the multiple optimum locations of the multimodal function. Agents in the GSO algorithm have a finite sensor range and carries a luciferin amount with them. The GSO algorithm is memory less and the glow-worms do not maintain any information in their memory. Some theoretical results related to the luciferin update mechanism of luciferin levels of the glow-worms are provided. Simulation result shows that the GSO algorithm succeeds in capturing multiple optima of several multimodal test functions.

Divya Jain et al [11]: A hybrid machine learning framework is used for the diagnosis of breast cancer and diabetes using efficient feature selection and classification technique. It identifies significant risk factors related to the chronic disease datasets by applying different feature selection techniques. Hybridization of ReliefF Feature Ranking with Principal Component Analysis (PCA) method is used in this paper. The effectiveness of the presented feature selection method is evaluated using k-nearest neighbour method. It is used for Classification of the diabetics and breast cancer patients. The hybridization improves the accuracy of the classifier with the proposed feature selection technique for the chronic disease datasets. The performance of the hybrid framework is found to be best in comparison to five other techniques such as Correlation Based feature Selection (CBS), Fast Correlation Based Feature Selection (FCBF), Mutual Information Based Feature Selection (MIFS), MOD Tree Filtering Approach and ReliefF Feature Selection

Yonglai zhang et al. [12]: This paper proposed a feature selection model for detecting the risk of stroke. It is based on a dataset that was obtained from biomedical tests and electronic archives on 792 patients at a community hospital in Beijing. This model uses standard deviation as a parameter to rank the features, a filter-based variable, and SVM for classification, which was effective in feature selection of stroke. The first 6 features: CK, AGE, LDH, α-HBD, ALP and SCr, the most important risk factors of stroke. The proposed model achieved an accuracy of 82.58% by means of 18 features from the original data set. The new map thus represents an effective detection that can help to identify patients with an increased risk of stroke events

## V. Conclusion

In this paper, we reviewed about the machine learning and feature selection techniques which is used for the detection of stroke. It has been noticed that accuracy may vary with the difference in datasets, feature selection techniques and optimization techniques. In most of the papers, it has been realized that SVM achieved high accuracy among classification algorithms. Ranking of features based on some parameters will improve the accuracy of detection of stroke.

## VI. REFERENCES

[1]. Dr. V. Ilango et al." Predictive Analytics in Health Care Using Machine Learning Tools and Techniques" International Conference on Intelligent Computing and Control Systems'17.

[2]. A.K.Shafreen Banu et. al. "A Study 0f Feature Selection Approaches for Classification" IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems'15.

[3]. N. Leibowitz et al." Automated measurement of proprioception following stroke" Research gate article on Disability and Rehabilitation'14.

[4]. G.E. Wang et al. "Comparison of Feature Selection Approaches based on the SVM Classification" IEEE conference on Industrial Engineering and Engineering Management'2008.

[5]. Guixiong Liu et al. “Multi-Class Classification of Support Vector Machines Based on Double Binary Tree” IEEE International Conference on natural computation'08

[6]. Sancho Salcedo-Sanz et al. “Feature Selection Methods Involving SVMs for Prediction of Insolvency in Non-life Insurance Companies” International conference on Financial Economy

[7]. Raid Alzubi et al. “ SNPs-based Hypertension Disease Detection via Machine Learning Techniques” IEEE conference on neural networks and learning systems

[8]. Yvan Saeys “A review of feature selection techniques in bioinformatics” journal on bioinformatics'07

[9]. Kalaiselvi Thiruvenkadam “A Review on Glow worm Swarm Optimization” International Journal of Information Technology'17

[10]. K.N. Krishnanand et al. “Glow-worm swarm based optimization algorithm for multimodal functions with collective robotics applications” International Journal on Multiagent and Grid Systems'06

[11]. Vijendra Singh et al. “Diagnosis of Breast Cancer and Diabetes using Hybrid Feature Selection Method” 5th IEEE International Conference on Parallel, Distributed and Grid Computing'18

[12]. Yonglai zhang et al. “Risk Detection of Stroke Using a Feature Selection and Classification Method” IEEE Translations and content mining'18

**Cite this article as :**

Priyanka S R, Deepa S S, "A Survey on Detection of Stroke Using Various Machine Learning Approaches", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 1, pp. 305-311, January-February 2020. Available at doi : https://doi.org/10.32628/IJSRSET207134
Journal URL : http://ijsrset.com/IJSRSET207134