IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 6, June 2018

www.ijasrm.com

ISSN 2455-6378

Comparative Study of Different Machine Learning Algorithm for Dyslexia Children

Mrs. Selvi. H[1], Mr.Saravanan. M.S [2]

[1]Research Scholar, Research and Development Centre Science, Bharathiar University, Coimbatore, Tamil Nadu, India

[2]Professor, Department of Information Technology, Saveetha School of Engineering, India

Abstract

Over the globe, today there are many children's were affected by dyslexia problem. This paper is focusing on medical diagnostic problem – identifying and diagnosing dyslexia children is based on checklist containing the symptoms and signs of dyslexia using ANN techniques applied with WEKA . Many researchers research identifying or diagnosing dyslexia or non-dyslexia children in many ways. But the drawback in the existing system is they are research based on either Intellectual Intelligent (IQ) or Emotional Intelligent (EQ). They are not given accurate result for detecting the children who are affected by Dyslexia. The person's achievement in life is depends upon both knowledgeable and emotional intelligence. IQ is not only giving successful in life. We must need EQ also. The aim of the present research is to detect and diagnosis the dyslexia children based on both intellectual and emotional intelligent. So we expected that they give accurate result for identifying dyslexia students.

Keywords *- Dyslexia, Artificial Neural Network, Support Vector Machine, Naive Bayes, J48 Decision Tree*

1. Introduction

This paper is focusing on medical diagnostic problem - searching for a relationship between Intellectual intelligent, Emotional intelligent and dyslexia using artificial neural network with data mining. The problem has been described in several studies. But these studies do not provide any clear combined decisions. Artificial neural network is a powerful tool and very helpful to doctors on various areas of medical field. Data mining is the process of extraction of interesting patterns or information from large collection of data set. The present method available to determine dyslexia in children is based on check list containing the symptoms and signs of dyslexia. This traditional method is time consuming, not accurate and obsolete also.

1.1. Dyslexia

Oswald Berkhan was first identified dyslexia in 1881. In 1887, the word 'dyslexia' was coined by Rudolf Berlin. He was an ophthalmologist in Germany. He discussed about a young boy who had severely affected by dyslexia in spite of showing physical abilities and typical intellectual in all other respects. Dyslexia (dys - abnormal and lexis - language or words) comes from Greek word is one of the types of learning disability such as difficult to read, to write, to spell, to reason etc.

Dyslexia is not a disease but language based disability in which a person has difficult to read. They are not stupid or lazy. The children who have average or above-average intelligence, with right support however they can succeed not only in school life also succeed in their life. Parents and teachers can help dyslexia children by encouraging their powers, knowing their weaknesses, understanding the learning methods, working with specialists and learning about approaches for dealing with specific difficulties.

2. Literature survey

Athanasios S. Drigas and Rodi-Eleni Ioannidou in 2013 suggested AI methods to use different diagnosis of SEN (Special Education Needs) learners from dyslexia and autism, also to develop the excellence of life of SEN learners. Julie M. David, Kannan Balakrishnan in 2013 to improve a new procedure for assigning and

defining the importance of the missing value complaint method and dimensionality decrease method in the performance of fuzzy and neuro fuzzy classifiers with specific emphasis on prediction of learning disabilities in school age children.

Manghirmalani et al presented a soft computing method called Learning Vector Quantization to classify a child as learning ability or disability. Once examined with learning disability, rule based approach is used to classify them into types of learning disability. Kohli et al., (2010) presented a systematic method for identifying dyslexia at an early stage by using ANN. This study paper is the first dyslexia identification problems using ANN. Also, it cans covering the assessment results of dyslexia children between 2003 and 2007 based on test data. Using an error back-propagation algorithm the test data covers the input data of the system and the output result contains two categories such as dyslexic and non-dyslexic.

Anuradha et al., (2010) presented a paper for analysis of Attention Deficit Hyperactivity Disorder (ADHD). This research paper is more perfect and less time consuming. SVM algorithms are mainly suitable for classification and regression. A data-set and the results of a questionnaire conducted by doctors are used to analyze the disorder and implement with SVM module. The result of this supervised learning technique referred as percentage of 88,674% success in identifying between the ages six to eleven years old children.

Hernadez et al (2009) introduced SEDA ('Sistema Experto de Dificulta desparael prendizaje' or 'Expert System for Learning Difficulties' in English) is a diagnostic tool for Learning Difficulties in elementary education. Using the Expert Systems design methodology is developed which include a information base containing of a series of strategies for Psycopedagogy assessment. It was assessed by the scale of: Poor, Moderately Efficient and Efficient where 80% of the assessors rated the system as Well-organized.

Jain et al., (2009) introduced Perceptron based Learning Disability Detector (PLEDDOR) model is used for identifying dyslexia, dyscalculia, and dysgraphia using syllabus based test conducted by special educators using an ANN technique. Totally 240 children were subjected to this test and results gathered from various schools and hospitals in India. It was evaluated as simple and easy to replicate in huge volumes.

Arthi and Tamilarasi (2008) proposed a model is used to diagnosis the children with autism using ANN techniques. The original autistic data is changed into fuzzy value and given as an input to the neural network architecture with back propagation algorithm using pseudo algorithm. In future k-nearest neighbor algorithm for a comparative could be used in expecting research the autistic disorder.

Fonseca et al showed electroencephalograms (EEG) to notice abnormalities related to electrical activity of the brain by studying different brainwaves. He produced a result that there is a significant difference between brainwaves of normal and learning Disability children. Macas et al proposed a system for extracting the features of eye movements from frequency and time domain. They decided that back propagation method based classification gave better outcomes than that offered by Bayes and Kohonen network.

3. Dataset description

In this present work, the method of informal assessment is adopted for designing the tool for predicting the dyslexia children. Even though different types of checklists are generally available for assessing dyslexia characteristics, a check list containing the 16 most frequent and important characteristics (signs & symptoms) of dyslexia collected from the above assessment list, after eliminating the unwanted and redundant ones, is prepared suiting to the dyslexia conditions generally prevailing in Tamil Nadu. Table1 shows a brief description of the dataset that is being considered.

Table1: dataset descriptions

DATASET	NO. OF ATTRIBUTES	NO. OF INSTANCES
Collecting from Hospitals and schools	16	300

This general check list adopted in this research work is shown in given table below.

Table 2: set of attributes

S.NO	ATTRIBUTES	SIGNS & SYMPTOMS OF DYSLEXIA
1	DR	Difficulty with reading
2	DS	Difficulty with spelling
3	DWE	Difficulty with written expression
4	DSW	Difficulty with speeds in writing
5	DN	Difficulty with numbers
6	DC	Difficulty with concentration
7	ED	Easily distracted
8	DM	Difficulty with memory

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 6, June 2018

www.ijasrm.com

ISSN 2455-6378

9	LM	Lack of motivation
10	DNS	Does not like school
11	STL	Slow to learn
12	AVO	Avoidance
13	LC	Lost confidence
14	DH	Difficulty with handwriting
15	DLS	Difficulty in learning a subject
16	LE	Low effort

Pre-processing and transformation of the dataset are done using WEKA tools.

Transformation steps include:

- Replacing missing values
- Normalization of values

Figure 1 shows a description of dyslexia dataset. The dataset used by us contains 16 attributes and 300 instances for dyslexia students. We have applied different algorithms using WEKA data mining tool for our analysis purpose.

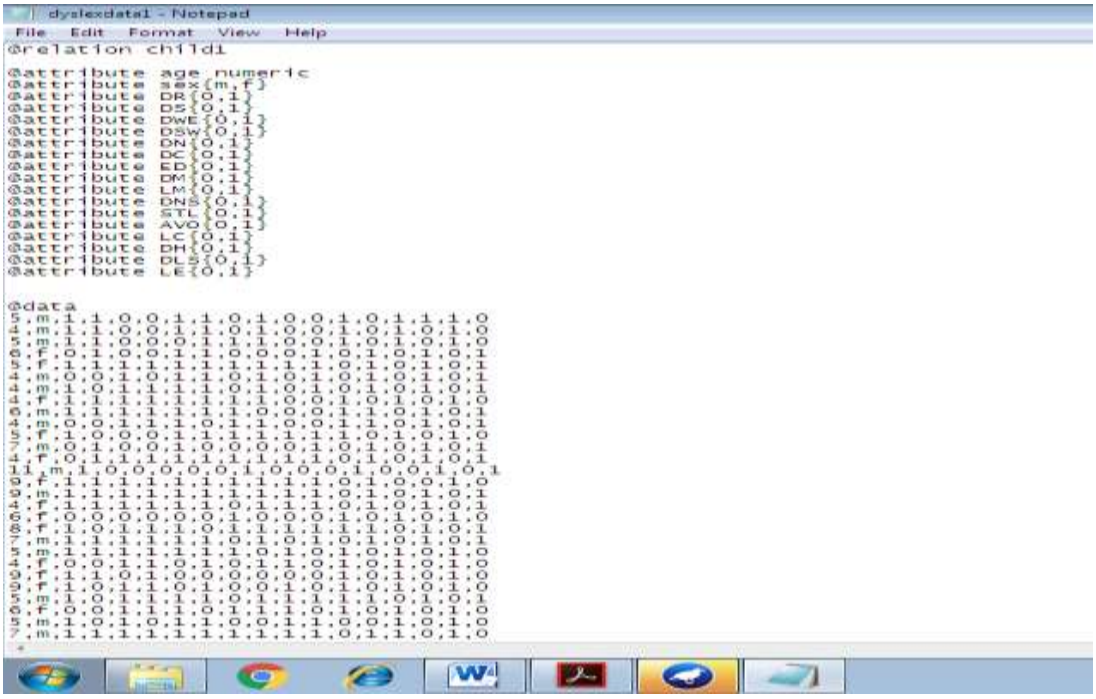

Fig.1 Screenshot view of Dyslexia Dataset

4. Algorithm used

Data mining is the process of mining information from bulky amount of data. It has very important to our real life data. Many tools are available for the purpose of prediction algorithm but they have some errors. Many of them cannot handle big data. There are many hospitals and health care institutions which collect huge amounts of patient data which becomes difficult to handle with currently existing systems. Machine learning algorithm helps to analyze and deriving hidden knowledge and information from the data sets. It improves accuracy and speed and also used in diagnosing several diseases. Now a days various efficient algorithms in data mining become popular such as SVM, ANN, Decision trees etc.

WEKA Tool is an open source data mining tool. It is developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. It is an advanced tool for developing machine learning (ML) algorithms for data mining tasks and their applications used to real-life data mining problems. These algorithms are applied directly to a dataset for data preprocessing, feature reduction, classification, regression, clustering, and association rules. It also includes visualization tools. The new machine learning algorithms can be used with it and existing algorithms can also be extended with this tool. We have applied following four commonly used classifiers for prediction on the basing on their performance. These classifiers are as follows:

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 6, June 2018

www.ijasrm.com

ISSN 2455-6378

4.1. Support vector machine (SOM)

In machine learning, SVM are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature.

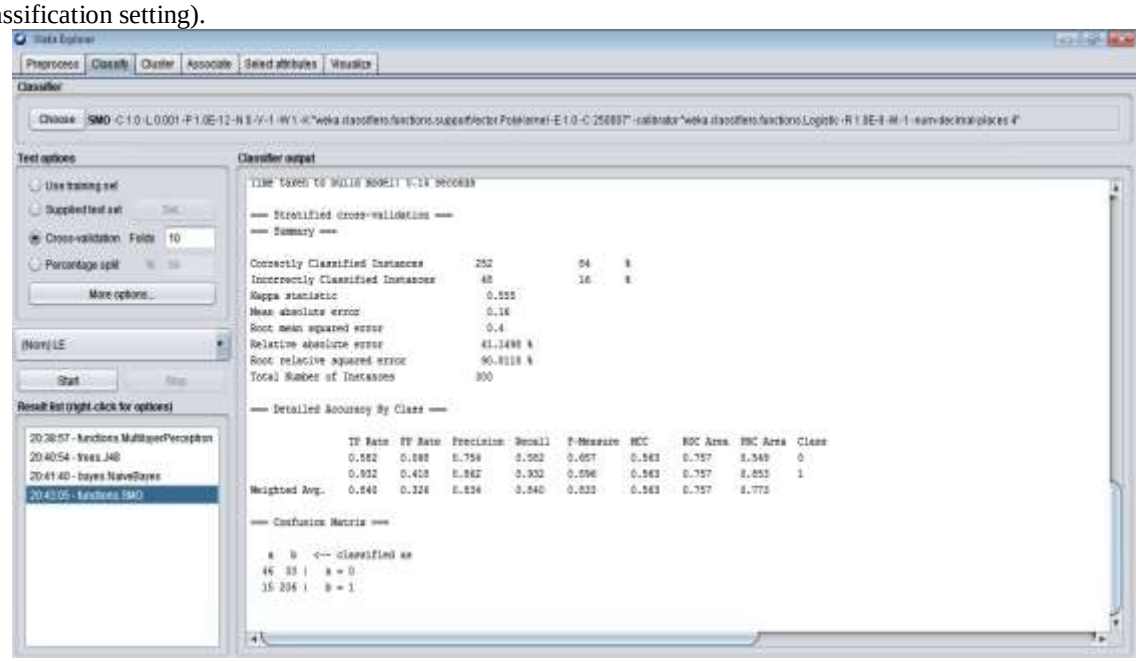

Figure 2: Machine Learning Algorithm Using SVM

4.2. Naive Bayes

In machine learning, Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s, and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. It also finds application in automatic medical diagnosis.

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 6, June 2018

www.ijasrm.com

ISSN 2455-6378

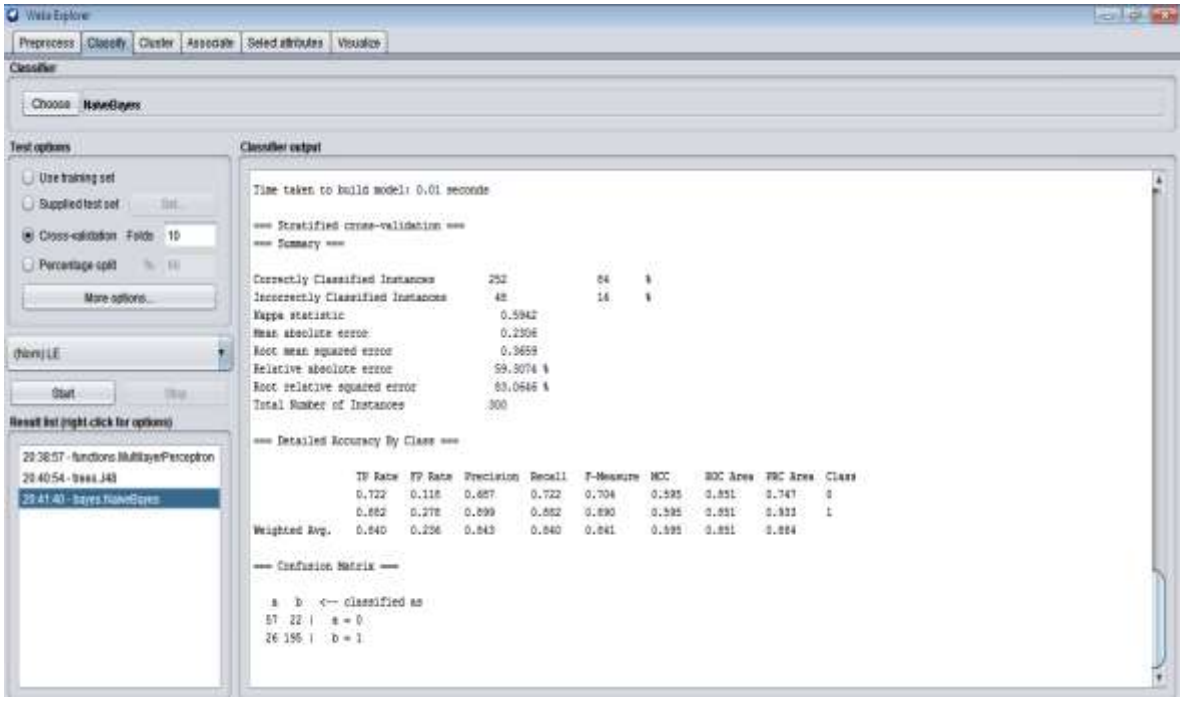

Figure 3: Machine Learning Algorithm Using Naïve Bayes

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

4.3. J48 Decision Tree

It's an algorithm used to generate a decision tree developed by Ross Quinlan. It is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by it can be used for classification, and for this reason, it is often referred to as a statistical classifier.

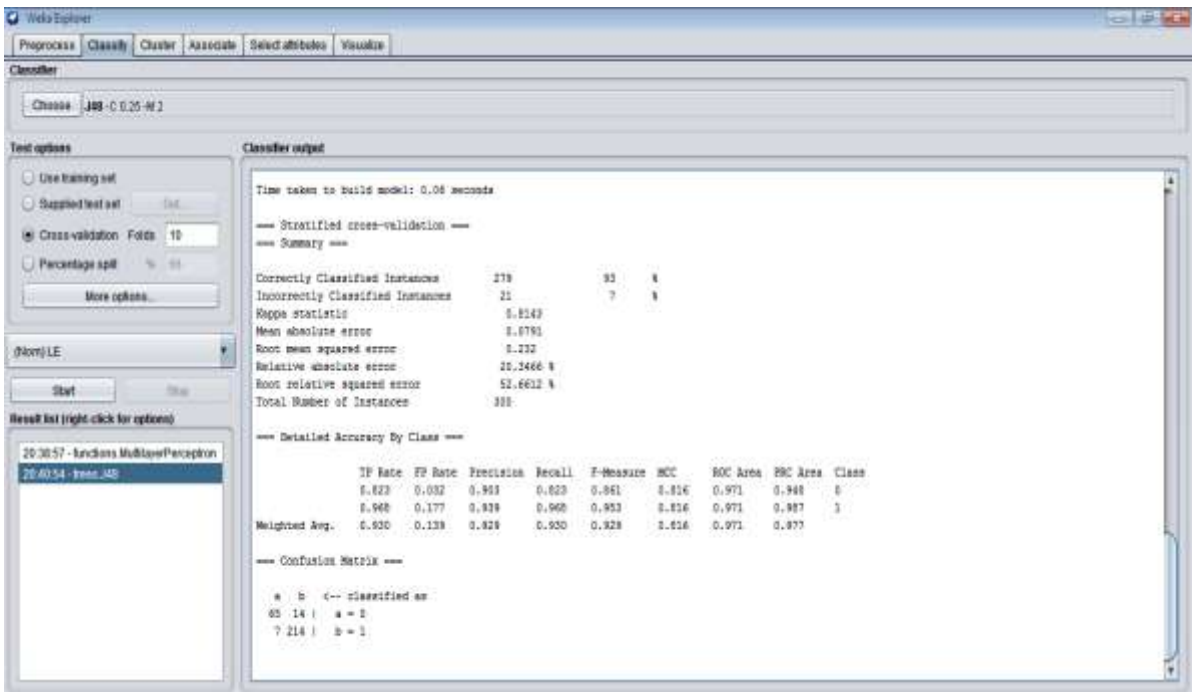

Figure 4: Machine Learning Algorithm Using J48 Decision Tree

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 6, June 2018

www.ijasrm.com

ISSN 2455-6378

4.4. Neural Network

They are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn (progressively improve performance) to do tasks by considering examples, generally without task-specific programming. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labeled as "cat" or "no cat" and using the analytic results to identify cats in other images. They have found most use in applications difficult to express in a traditional computer algorithm using rule-based programming.

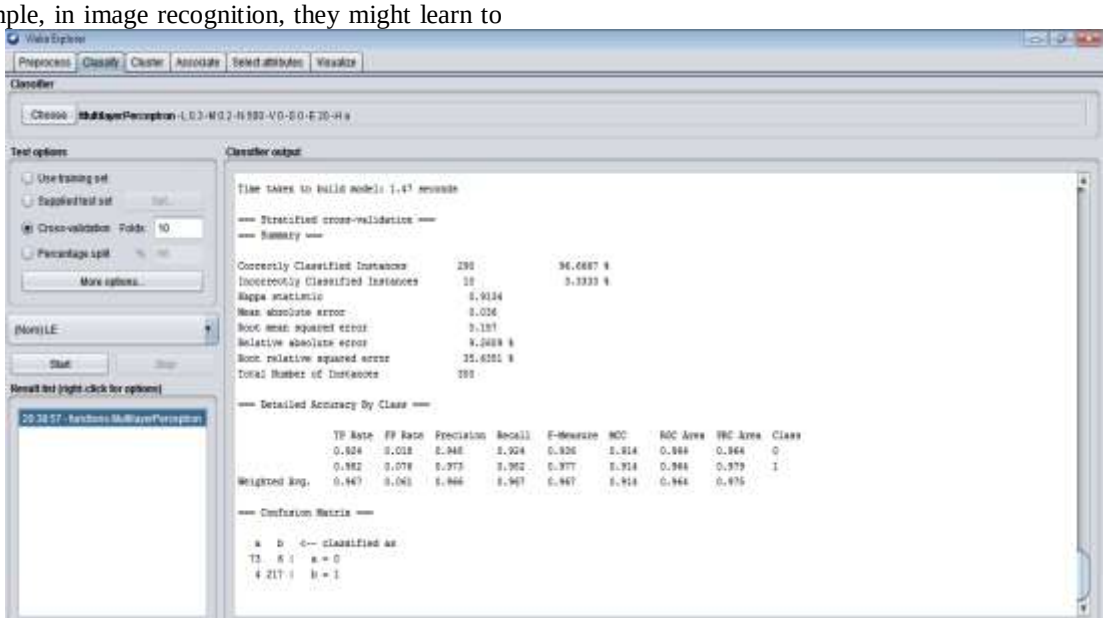

Figure 5: Machine Learning Algorithm Using Neural Network

5. Performance comparison

The aim of this paper is to find the best among the machine learning algorithms discussed in previous section for dyslexia detection. Using these algorithms we have built four different classifiers using with WEKA tool. Because this tool, well-suited for developing new machine learning schemes. We have applied 10-fold cross validation in order to check the results of every classifier for unknown instances. This result of whole experimentation is shown in Table 2. While the above figures shows the pictorial representation of efficiency of these algorithms.

Table 3. Comparison of different machine learning algorithms

Algorithm classification	Correct Classification Rate	Incorrect-Classification Rate
Support Vector Machine	84 %	16 %
Naive Bayes	8 %	16 %
J48 Decision Tree	93 %	7 %
Neural Network	96.6667 %	3.3333 %

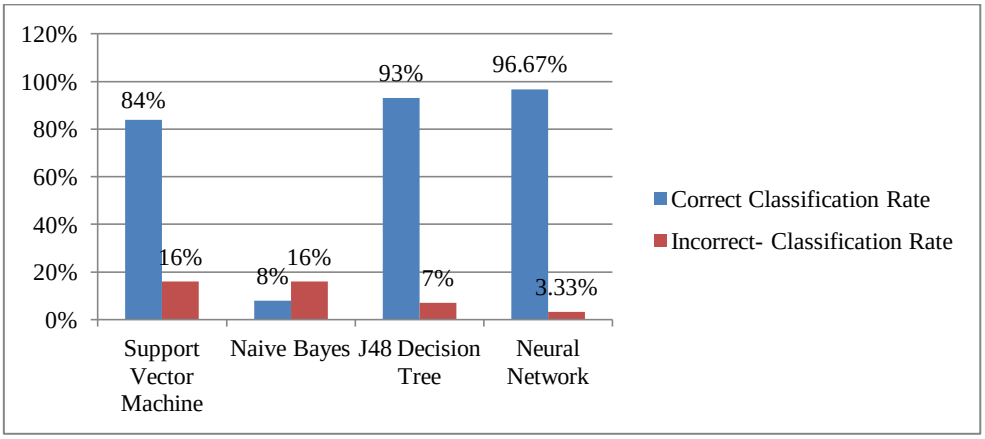

Fig. 6. Graph for Correct Classification VS. Misclassification rate

6. Conclusion

In this research paper, we tried to find the predictive performance of different classifiers. We select four common machine learning algorithms classifiers are implemented with same data and outcomes are compared on the basis of correct classification and incorrect classification rate according to results. It can be concluded that Neural Network classifier is the best as compared to Support Vector Machine, Decision Tree and Naive Bayes. After examining the quantitative data generated from the computer models, Moreover their performance is closely competitive showing slight difference. So, many experiments on several other datasets need to be measured to draw a more general conclusion on the comparative performance of the classifiers.

References

Part of this research work, we discuss the field of dyslexia with these references

[1] Anuradha, J et al, "Diagnosis ofADHD using SVM algorithm", Proceedings of the Third Annual ACM Bangalore Conference (2010)

[2] Arthi. K and Tamilarasi, A, "Prediction of autistic disorder using neuro fuzzy system by applying ANN technique", International Journal of Developmental Neuroscience 26, 699–704 (2008)

[3] Athanasios S. Drigas and Rodi-Eleni Ioannidou,"A Review on Artificial Intelligence in Special Education", Ag.Paraskevi, 15310, Athens, Greece, 2013.

[4] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, Vol. 2, 1998, pp. 121-167. doi:10.1023/A:1009715923555

[5] Disability World, 2003,"UNICEF and Disabled Children and Youths", Disability World, No. 19, (online at www.disabil ityworld.org/ 06 − 08 03/index.shtm).

[6] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[7] Fabian, H. P., Chan, K. S., Ho, K. Y., & Leong, S. K. (2004). A Study on Decision Tree. 2nd Engineering & Technology Student's Congress. Kota Kinabalu.: SKTM.

[8] Hernadez, J et al, "Learning Difficulties Diagnosis for Children's Basic Education using Expert Systems", WSEAS Transactions on Information Science and Applications (2009)

[9] J. Anuradha, Tisha, V, Ramachandran, K. V. Arulalan, B.K. Tripathy, Diagnosis of ADHD using SVM algorithm, In Proceedings of the Third Annual ACM Bangalore Conference, 2010.

[10] Jain et al, "Computational Diagnosis of Learning Disability", International Journal of Recent Trends in Engineering (2009)

[11] James Freeman and David Skapura, "Neural networks: Algorithms, applications and Programming Techniques", Pearn Education, 2007.

[12] Julie M. David, Kannan Balakrishnan: Machine Learning Approach for Prediction of Learning Disabilities in School Age Children, Int. J. of Computer Applications, ISSN-0975-8887, 9(10), Nov. 2010, pp 7-14. http://www.ijcaonline.org/archives/volume9/ number 1/1432-1931

[13] Kohli, M., Prasad, T.V, "Identifying Dyslexic Students by Using Artificial Neural Networks", Proceedings of the World Congress on Engineering, London, U.K, vol. 1(2010)

[14] Lineu C. Fonseca et al, "Quantitative EEG in children with learning disabilities", Analysis of band power, 64(2-B):376-381, 2006

[15] Manghirmalani et al, "Learning Disability Diagnosis and Classification-a Soft Computing Approach", IEEE World Congress on Information and Communication Technologies (WICT); DIO: 10.1109/WICT.2011.6141292, 2011

[16] Martin Macas et al, "Bioinspired methods for analysis and classification of reading eye movements of dyslexic children", Department of Cybernetics, Czech Technical University in Prague, Czech Republic NiSls Symposium 2005

[17] Michael W.Berry et.al," Lecture notes in data mining", *World Scientific*(2006)