IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Diagnosis of Dyslexia Students Using Classification Mining Techniques

H. Selvi[1*], M.S. Saravanan[2]

[1]Dept. of Computer Science, Bharathiar University, India
[2]Dept. of Information Technology, Saveetha School of Engineering, India

Corresponding Author: selvibadrinarayanan@yahoo.com

DOI: https://doi.org/10.26438/ijcse/v7i5.2833 | Available online at: www.ijcseonline.org

Accepted: 08/May/2019, Published: 31/May/2019

Abstract - Now a day, all over the world **70-80%** of people with poor reading skills are likely dyslexic. One in five a student, or 15-20% of the population, has a language based learning disability. Dyslexia is the most common of the language based learning disabilities. Nearly the same percentage of males and females has dyslexia. Children suffering from a learning disability might face difficulties with reading, writing or mathematics but they excel in other areas of interests. It is in the interest of the society and especially the parents to identify the problem early in the development of the child and steer him/her towards a preferred field. They might lose their sense of self-worth and blame themselves for their situation. The model being proposed is a Web-based tool incorporating machine learning techniques (Decision trees) for predicting whether children (8-10 years) are at a risk of having Specific Learning Disability by showing the areas of learning disability on the basis of the clinical information and research.

Keywords: Dyslexia, Weka, SVM, Naïve Bayes, J48 Decision Tree, Neural Network

I. 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 in medical field. Because this tool is very helpful for doctors on various areas of medicine, such as diagnostic systems, biomedical analysis, image analysis, drug development etc. and also monitoring a lot of health problems. Data mining is the process of extracting hidden patterns and useful information from large set of data is now becoming part of current inventions.

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. Such dyslexia identification facilities are much less at schools or even in cities. Parents are either unaware or may not willing to take their children to undergo such as evaluation. Even if, teachers are advised, parents may hesitate to such evaluation process because of the unawareness of the society about dyslexia as they might think that the child may be mentally retarded. If the dyslexia determination facility is attached with schools and checkups are arranged as a routine process, dyslexia can be identified at an early stage
.

The main objectives of this work we used Waikato Environment for Knowledge Analysis (WEKA) version 3.7.5 as simulation tool which is an open source tool. The data set we used for diagnosis is real world data with 300 instances and 16 attributes. In the end part we check the performance comparison of different algorithms to propose the best algorithm for dyslexia diagnosis. So this will helps in early identification of dyslexia students and reduces the diagnosis time.

Section I contains the introduction of dyslexia detection, Section II contains the related work of detection of dyslexia, Section III explains methods of classifiers, Section IV and V contains proposed work and results & discussion and Section VI contains the conclusion.

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.

II. RELATED WORKS

Athanasios S. Drigas and Rodi-Eleni Ioannidou in 2013 suggested Artificial Intelligence 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 [2]. 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 chil-dren[3].

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 [4]. Kohli et al. 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 [5].

Anuradha et al. 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 su-pervised learning technique referred as percentage of 88,674% success in identifying between the ages six to eleven years old children [6].

Hernadez et al. introduced SEDA (‘Sistema Experto de Dificulta desparael prendizaje’ or ‘Expert System for Learning Difficulties’ in English) is a diagnostic tool for Learning Difficulties in elemen-tary education. Using the Expert Systems design methodology is developed which include an information base containing of a se-ries 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 [7].

Jain et al. introduced Perceptron based Learning Disability Detec-tor (PLEDDOR) model is used for identifying dyslexia, dyscal-culia, and dysgraphia using syllabus based test conducted by spe-cial 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 [8].

Arthi and Tamilarasi 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 neu-ral 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 dis-order [9].

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 signifi-cant difference between brainwaves of normal and learning Disa-bility children [10]. 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 [11].

Rahman urged that Increasing Intelligibility within the Speech of the Autistic Children by an Interactive Computer Game. There is no definite treatment for autism. Serving to autistic children by providing games and teaching facilities to improve their skills [12]. In the year 2013 Santos examines the first detection of Autism means that taking the symptoms of patient during childhood sup-ported by preverbal vocalization by using the classification tech-nique supervised learning SVM (support vector machine)[13]. Chaminade started a shot to use MRI study of young adults with autism interacting with a humanoid robot [14]. Prud’hommeaux et al. examines the difficulties for classification of non-standardized text of machine learning techniques [15]. Kathleen T Quach sug-gested that problem through the classification problem is that ASD may be a terribly heterogeneous disorder which will have sub-groups with totally different genetic expression signatures. To boost classification, it should be helpful to stratify the ASD class into subgroups and enrich the input set with clinical measures [16].

Alexander Genkin et al. have given an easy Bayesian logistic re-gression approach that uses a Laplace prior to avoid over fitting and produces sparse predictive models for text data. They applied this approach to a spread of document classification issues and show that it produces compact predictive models a minimum of as effective as those created by support vector machine classifiers or ridge logistic regression combined with feature selection [17].

Morris [18] has reviewed the conceptual and operational limita-tions of classic approaches in classification of learning disability. He found that through the new development of more reliable and valid classification method will remove many of the present prob-lems in clinical and research endeavors with learning disabled children. Suresh and Raja [19] have applied Functional Magnetic Resonance Imaging (firm) through image processing techniques to classify SLD (specific learning disability), to determine depth of severity, degree of recovery and therapy.

Cohen and Sedater [20] have used ANN technology in classifica-tion of autism among children. They compare ANN method with simultaneous and stepwise linear discriminate analysis. They found neural network methodology is superior to discriminate function analysis both in its ability to classify groups (92% vs. 85%) and to generalize to new cases that were not part of the training sample. It has been observed that several researchers have applied ANN for classification of disability in children with en-couraging results.

III METHODOLOGY

Many experiments are being carried out for evaluating the performance of many possible algorithms for the diagnosis of dyslexia which are:

A. Support Vector Machine

Support Vector Machine (SVM) has the high-performance capability to predict, analyses, regression and classify dataset. It is used to predict and analysis the dataset to regression and classification techniques. SVM is supervised learning algorithms which are mostly used data mining classification. It can perform classification by separating classes by discovering the optimal hyper plane. SVM gives the most accurate result by comparing other classification algorithms. Training data represented by support vector machine model. By maximizing the combined between the instances of two classes, it can minimize the error.

B.Naïve Bayes

Naïve Bayes has supervised learning algorithm classification technique which is a probabilistic classification learning algorithm based on applying Bayes theorem.

Naïve Bayes is used for diagnosis and prediction of the problem. The Naïve Bayes algorithm requires a small amount of training data during classification to predict to evaluate the parameter. The Naïve Bayes classification method used to predict, an associate of each class. For instance, the probability for the specified record for the target class. The class which has maximum probability is expected the most likelihood class. Bayes theorem is P (Y/X) = P (X/Y) ·P(Y) / P(X). Where P (X) is similar to whole classes and P (Y) = relative frequency of class Y

C.J48 Decision Tree

According to J48 decision tree is an open source Java implementation of C4.5 decision tree algorithm in weak platform. It is the extension of the earlier ID3 algorithm, which is developed by Ross Quinlan In J48 decision tree classification algorithm the branch of data distribution easily understandable and every leaf is pure gain evidence from the branch. J48 classification algorithm uses top-down greedy search methods for producing tree J48 decision tree produces sorting tree whose; the leaf denotes the ending class and the internal attributes represent a possible number of output of the branch features .

D. Artificial Neural Network (ANN)

Artificial neural network (ANN) also called "neural network', widely used in the real application based on biological neurons. ANN contains the interconnected nodes of artificial neurons and communicates each node by adjusting weights for each node and change its arrangement during message transfer [4]. ANN is learning algorithm, it can learn and adapt to change its structure during information received from the internal and external environment during learning [22]. ANN contains a set of the input layer, a hidden layer, and finally, contain output layers. These layers are interconnected to each other, in which weight is linked with each node. Neuron network is constructed multiplied the sum of all inputs by the sum of all weighted that are participating each processed.

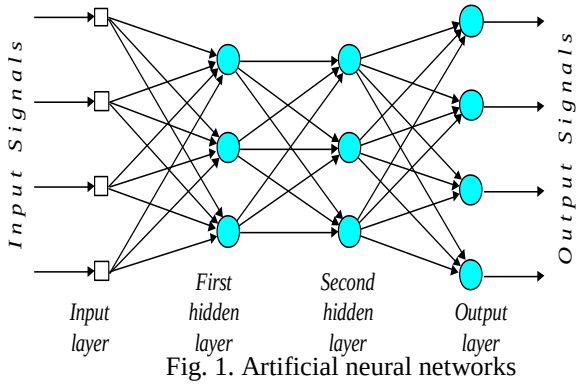

Fig. 1. Artificial neural networks

(Y) = X1W1+X2W2+.... +XnWn

Where W is the weight, X is input and f (Y) is activation function.

IV. PROPOSED WORK

The following are the steps involved in the proposed work:

5.1 Data Collection 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

5.2 Data Pre-processing Data pre-processing has been emerging as an important issue for mining the data due to the incompleteness and inconsistency of the real world data. The missing entries are filled up by using the average values.

5.3 Feature Selection

The feature is the process of eliminating some parts of the attributes from the data set. Feature selection used to improve the performance of the classification algorithm by removing unwanted attributes from the input. Some of the parameters which are used for Predictive data mining are

A. Sensitivity

It is also known as True Positive Rate. It is used for measuring the percentage of sick people from the dataset.
Sensitivity = Number of true positives/Number of true positives + Number of false negatives

B. Specificity

It is also known as True Negative Rate. It is used for measuring the percentage of healthy people who are correctly identified from the dataset.
Specificity = Number of true negatives/Number of true negatives + Number of false positives

C. Precision and recall

It is also known as positive predictive value. It is defined as the average probability of relevant retrieval.
Precision = Number of true positives/Number of true positives + False positives

D. Recall

It is defined as the average probability of complete retrieval.
Recall= True positives/True positives + False negative

E. Accuracy

A measure of a predictive model that reflects the proportionate number of times that the model is correct when applied to data. The formula for calculating the Accuracy,
Accuracy=Number of correctly classified samples/Total number of samples

E. Confusion Matrix

It is used for displaying the number of correct and incorrect predictions made by the model compared with the actual classifications in the test data. The matrix is represented in the form of n-by-n, where n is the number of classes. The accuracy of each classification algorithms can be calculated from that.

V. RESULT AND DISCUSSION

In data mining tools classification deals with identifying the problem by observing characteristics of problems amongst children and diagnose which algorithm shows best performance on the basis of WEKA's statistical output. Table 1 shows the WEKA data mining techniques that have been used in this paper along with other prerequisites like data set format etc. by using classification algorithms [21].

Table 1. Weka data mining technique by using different algorithms

Software	WEKA
Datasets	Dyslexia
Weka Data Mining Technique	Explorer
Classification Algorithms	SVM Naïve Bayes J48 Decision Tree Neural Network (Multi-layer perceptron)
Operating System	Windows 7
Dataset File Format	ARFF

In order to obtain better accuracy 10 fold cross validation was performed. For each classification we selected training and testing sample randomly from the base set to train the model and then test it in order to estimate the classification and accuracy measure for each classifier.

The system contains 16 attributes and 9 hidden neurons and 3 output nodes. The two output nodes are true and false. It contains learning rate is 0.3, epoch is 500, momentum is 0.2 and error of epoch is 0. The number of nodes is found through test and fault. Here back propagation method is used for altering the weights and biases so as to minimize a cost function. The activation function measured for each node is the binary sigmoid function. The parameters learning rate and momentum set values can be take priority over in the graphical interface. A decay parameters causes to the learning rate. Here one hidden layer is used. The number of hidden neuron is reduced in the training phase. The classifier

model of full training set and the stratified cross-validation summary is given below

Table 2: Performance comparison between different algorithms

S. No	Particulars	SVM	Naïve Bayes	J48 Decision Tree	Neural Network
1	Correctly Classified Instances	206	252	255	290
2	Incorrectly Classified Instances	94	48	45	10
3	Kappa statistic	0.318	0.594	0.6356	0.9134
4	Mean absolute error	0.313	0.231	0.246	0.036
5	Root mean squared error	0.560	0.366	0.3173	0.157
6	Relative absolute error	66.036	59.31	63.264	9.261
7	Root relative squared error	114.2	83.07	72.08	35.64
8	Total number of instances	300	300	300	300
9	Time taken to build model	0.14 sec	0.01 sec	0.11 sec	1.45 sec
10	Correct Classification Rate	84%	84%	93%	96.77 %
11.	Incorrect Classification Rate	16%	16%	7%	3.33 %

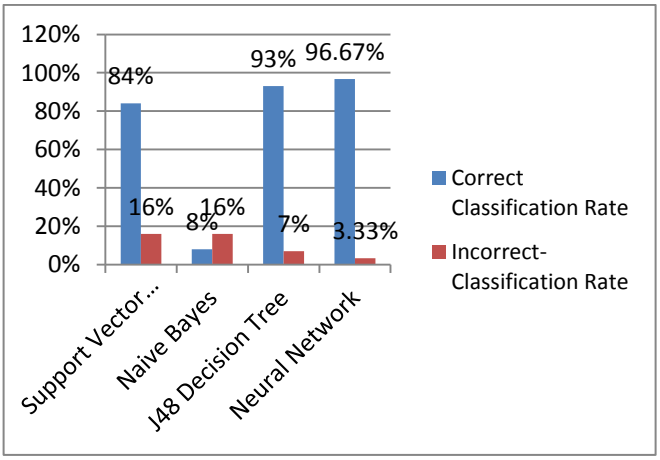

Fig. 2. Graph for Correct Classification VS. Misclassification rate

VI. CONCLUSION

From the results obtained, it is understood that how effectively the MLP with back propagation algorithm classifies the dyslexia dataset. The major issue studied from this study of prediction of dyslexia in children is failure of the classifier in handling the missing values in the datasets. The missing values contribution may be some times very important and significant. The second issue noticed is that some of the attributes in the check list have less contribution in dyslexia prediction. So we have to reduce the number of attributes for improving the performance of the classifier. Reducing the number of attributes is very effective and that will help to reduce the time taken for constructing the model. The results obtained shows that 96.67% accuracy with correctly classified instances and 3.33% accuracy in incorrectly classified instances. The findings show that there is no solution in the case of missing values present. Also some attributes are unwanted and hence have no contributions in predicting the dyslexia. But from the output of this classification method, it is understand how easily the learning disability can be predicted in the early stages itself.

ACKNOWLEDGEMENT

The authors would like to thank Almighty God for making every-thing possible according to His divine will and providence, especially in giving them strength, knowledge, and wisdom towards the completion of this document. They would also like to thank their family and friends.

REFERENCES

[1] Disability World, 2003,"UNICEF and Disabled Children and Youths", Disability World, No. 19, (online at www.disabil ity-world.org/ 06 − 08 03/index.htm).

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

[3] 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.

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

[5] 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).

[6] Anuradha, J et al, "Diagnosis of ADHD using SVM algorithm", Proceedings of the Third Annual ACM Bangalore Conference (2010) https://doi.org/10.1145/1754288.1754317.

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

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

[9] 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).
[10] Lineu C. Fonseca et al, "Quantitative EEG in children with learning disabilities", Analysis of band power, 64(2-B):376-381, 2006.
[11] 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.
[12] Md. Mustafizur Rahman, S. M. Ferdous, Syed Ishtiaque, "Increasing Intelligibility in the Speech of the Autistic Children by an Interactive Computer Game", Multimedia(ISM), pp 383 – 387, 2010.
[13] Joan F. Santos, NiritBrosh, Tiago H. Falk, Lonnie Zwaigenbaum, Susan E. Bryson, Wendy Roberts, Isabel M. Smith, Peter Szatmari and Jessica A. Brian, "Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers", International Conference on Acoustics, Speech and Signal Processing, pp 7567 – 7571, 2013.
[14] Suresh P. and Raja, B. K, 2011, "A Review on Analysis and Quantification of Specific Learning Disability (SLD) with fMRI using Image Processing Techniques", IJCA Proceedings on International Conference on VLSI, Communications and Instrumentation (ICVCI) , vol. 5, pp. 24-29, Foundation of Computer Science.
[15] Prud'hommeaux et al., "Classification of atypical language inautism", in Proceedings of the 2nd Workshop on Cognitive Modelingand Computational Linguistics, pp: 88-96, 2011.
[16] Kathleen T Quach et al., "Application of Artificial Neural Networksin Classification of Autism Diagnosis Based on Gene ExpressionSignatures".
[17] Alexander Genkin et al., "Large-scale Bayesian logistic regression fortext categorization", Technometrics, pp: 291-304, 2007. RachnaAhuja et al, / (IJCSIT) International Journal of Computer Science and Information Technologies https://doi.org/10.1198/004017007000000245.
[18] Morris, R. D., 1988, "Classification of learning disabilities: Old problems and new approaches", Journal of Consulting and Clinical Psychology, vol. 56, no. 6, pp.789-794. https://doi.org/10.1037/0022-006X.56.6.789.
[19] Folorunsho, Olaiya. "Comparative Study of Different Data Mining Techniques Performance in knowledge Discovery from Medical Database." International Journal 3, no. 3 (2013). *International Journal of Engineering & Technology* 3411
[20] Cohen, I.L., Sudhalter, V., Landong-Jimenez, D. and Keogh, M., 1993, "A Neural Network Approach to the Classification of Autism", Journal of Autism and Developmental Disorders, vol. 23, no. 3, pp. 443-466. https://doi.org/10.1007/BF01046050.
[21] James Freeman and David Skapura, "Neural networks: Algorithms, applications and Programming Techniques", Pearson Education, 2007.

Author's Profile

H. SELVI is pursuing Ph.D. in the stream of computer science from Bharathiar University, Coimbatore. She is working as an Assistant Professor in the Department of Computer Application, Bhaktavatsalam Memorial College, Chennai. Her research area is based on detection of dyslexia using artificial neural network techniques.

M.S.Saravanan has obtained Ph.D. from the Bharathiar University in the year 2013 in India. He is currently working as Professor in the Department of Information Technology in Saveetha University, Chennai, India. He is also a member of IEEE, CSI (Computer Society of India), ISTE (Indian Society of Technical Education) and the registered member of IAENG (International Association of Engineers). He has published eighty-six international publications and presented twenty-four research papers in international and national conferences, having 19 plus years of teaching experience in various institutions in India and Abroad.

