



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: VIII Month of publication: Aug 2023

DOI: <https://doi.org/10.22214/ijraset.2023.55332>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Prediction of Learning Disability Using Machine Learning

Anjana V Ravindran¹, Anjana V J², Meenakshi P³

Department of Mathematics, Amrita Vishwa Vidyapeetham

Abstract: *This A learning disability is a neurological disorder. The children predicted with learning disability may find it difficult to spell, read, write, organize things and so on. Learning disabilities are not related to intelligence or motivation. People with learning disabilities have average or above-average intelligence but may need special accommodations and support to learn and succeed. Early identification, assessment, and intervention are critical for managing learning disabilities. With appropriate support and accommodations, people with learning disabilities can achieve their full potential and lead fulfilling lives. Machine Learning algorithms can be useful in predicting learning disabilities because they can analyse large amounts of data quickly and accurately, and they can identify patterns that may not be apparent to human observers. Deep learning models can lead to better and faster predictions and they are capable to work with unstructured data as well. While Machine Learning and Deep Learning algorithms have shown promise in predicting learning disabilities, it's important to use these tools in a responsible and ethical manner to ensure that individuals' privacy and autonomy are protected. By leveraging the power of these algorithms, we can help to ensure that children with learning disabilities receive the support they need to reach their full potential. In this study, six models, ANN and CNN, were assessed for their effectiveness in predicting specific learning disabilities. The performance measure used to evaluate the models was accuracy. Among the models, KNN is found to be the most accurate with 90.33% followed by Random Forest with 79.22%, CNN with 77.69% LSTM with 77.57%, ANN with 57.57% and SVM with 57.57%.*

Keywords: *Learning Disability, Machine Learning, Deep Learning, Ensemble Method, Neural Networks*

I. INTRODUCTION

A learning disability is a neurological condition which affects the brain's ability to send, receive, and process information. Learning disabilities include a group of disorders such as dyslexia, dyscalculia and dysgraphia. Each type of disorder may coexist with another. Early diagnosis and intervention are crucial for individuals with learning disabilities to receive the support they need to succeed in school and life.

Machine Learning and Deep Learning algorithms can help to identify patterns and predict outcomes based on large amounts of data. In the context of learning disabilities, these algorithms can be trained on data from assessments, medical records, and other sources to create models that can accurately diagnose learning disabilities in children at an early age. One child with learning disability may not have the same kind of learning problems as another child with learning disability. Some of the learning disabilities that specifically affects the academic performance are,

- Dyslexia - Reading impairment
- Dysgraphia - Writing and drawing impairments
- Dyscalculia - Mathematical problem-solving impairment

- 1) *Dyslexia:* A lack of precise and fluent word comprehension is characteristic of dyslexia. The children with dyslexia have difficulty with word identification, encoding and orthography. Reading comprehension is often hampered by the ability of reading very poor words. Individuals with dyslexia also lack phonemically and phonologically sensitive sensitivity, referring to the capacity to understand a spoken phrase, the phonemes, syllables, onsets, Rimos and other sound structures to be identified and manipulated. Children with dyslexia can also have impaired spelling which interferes with the correct and fluent combination of letters and letters with sounds.
- 2) *Dyscalculia:* Individuals with such a learning impairment show impaired mathematical calculation abilities, numbers and mathematical facts impossible to understand. Although features of LD in math vary from person to person, common characteristics include difficulty with counting and measurement, learning number facts, telling time, counting money and estimating number quantities.

- 3) *Dysgraphia*: The term is used to capture both the physical act of writing and the quality of written expression. Individuals with dysgraphia often have a tight, awkward pencil grip and body position. They will find difficulty in writing or drawing on a line or within margins, trouble organizing thoughts on paper, forming letter shapes and maintaining consistent spacing between letters or words are some of the challenges faced by these individuals.

Learning disabilities could arise for a number of reasons,

- a) *Heredity*: It is observed that a child, whose parents have had a learning disability, is likely to develop the same disorder.
- b) *Illness during and after birth*: An illness or injury during or after birth may cause learning disabilities. Other possible factors could be drug or alcohol consumption during pregnancy, physical trauma, poor growth in the uterus, low birth weight, and premature or prolonged labour.
- c) *Stress during infancy*: A stressful incident after birth such as high fever, head injury, or poor nutrition.
- d) *Environment*: Increased exposure to toxins such as lead (in paint, ceramics, toys, etc.)
- e) *Comorbidity*: Children with learning disabilities are at a higher than-average risk for attention problems or disruptive behaviour disorders.

Machine Learning finds application in various fields and domains where the prediction of specific outcomes is required. It encompasses a class of algorithms that enable software applications to learn and improve their predictive accuracy without explicit programming. Classical machine learning is often classified based on the algorithm's learning approach to enhance its prediction accuracy. The fundamental principle of machine learning is to construct algorithms that can receive input data, employ statistical analysis to forecast an output, and continuously update their predictions as new data becomes available. The processes involved in machine learning bear resemblance to those of data mining and predictive modelling, as they involve the exploration of data to identify patterns and adjust program actions accordingly. This paper utilizes different machine learning techniques for the prediction.

II. LITERATURE SURVEY

In [1], A. Devi, Dr. G. Kavya, M. Julie Therese and R. Gayathri proposed a testing scale tool to diagnose and identify SLD. The tool allows the students, who are suspected to have SLD, to take up a quiz. Some specific test questions are repeated thrice based on the type of learning impairments. After completion of the test, resultant data is provided as input to the decision tree algorithm. Based on the marks obtained and time taken to complete the test, the decision tree algorithm predicts learning disabilities in children. The proposed approach is used to develop an integrated and user-friendly tool that is highly accurate in identifying reading, writing and mathematical disorders, and suggest the right way and most appropriate instructional activities to parents and teachers. A total number of 40 Children were participated in the online test with their parents' permission. The 12 of them were children without SLD problem and 28 of them were children with SLD problem. The result shows that the proposed web-based screening tool will definitely help the teachers in the school to identify those children with SLD's along with rehabilitation technique.

In [2], H. Atakan Varol, Subramani Mani, Donald L. Compton, Lynn S. Fuchs, and Douglas Fuchs presents application of machine learning methods on a 356-sample dataset for early prediction of reading disability among first graders. A wide array of classifiers consisting of Support Vector Machines, Decision Trees (CART and C4.5), Linear Discriminant Analysis, k Nearest Neighbour and Naive Bayes Classifiers were used in this study. As for the result SVM and Naive Bayes Classifiers are found to be the highest performing classifiers than the two decision tree algorithms, CART and C4.5.

In [3], K. Ambili and P. Afsar aims at analysing various data mining techniques for the prediction of learning disability. The paper highlights an improved machine learning approach for learning disability prediction in school children using a fusion of the conventional Naive Bayes and Neural Network Classifier. The data has been collected from a special school in Kerala. The collected data consists of 16 attributes that comprises the symptoms of children who were suffering from learning disabilities. The data was used as the training set for various algorithms. The testing data was collected from 30 school children through the questionnaire. The result shows that the fusion technique of Naive Bayes and Neural Network is found to be the best among classification and prediction algorithms in the diagnosis of learning disability when compared to other machine learning algorithms.

In [4], Julie M. David and Kannan Balakrishnan highlights the two machine learning approaches, viz. Rough Sets and Decision Trees (DT), for the prediction of Learning Disabilities (LD) in school-age children, with an emphasis on applications of data mining. In rough sets the attribute 4 reduction and classification are performed using Johnson's reduction algorithm and Naive Bayes algorithm respectively. For rule mining and in construction of decision trees, J48 algorithm is used. As for the data, decision table include 513 objects or cases of LD. For each case, 16 attributes were registered. The result of this study indicates that, the rules system represented by the decision trees may be significantly incorrect for inconsistent data with large number of variables.

It is found that, for selection of attributes, rough sets are very useful especially in the case of inconsistent data and it also gives the information about the attribute correlation which is very important in the case of learning disability.

The results obtained from this study is compared with that of other classifiers such as Naive Bayes, SVM and MLP and it is found that rough set is better in terms of classification and accuracy.

In [5], Peter Drotar and Marek Dobes employed a machine learning approach to identify handwriting deteriorated by dysgraphia. To achieve this goal, they collected a new handwriting dataset consisting of several handwriting tasks and extracted a broad range of features to capture different aspects of handwriting. These were fed to a machine learning algorithm to predict whether handwriting is affected by dysgraphia. Then they compared it with several machine learning algorithms and discovered that the best results were achieved by the adaptive boosting (AdaBoost) algorithm. The results show that machine learning can be used to detect dysgraphia with almost 80% accuracy, even when dealing with a heterogeneous set of subjects differing in age, sex and handedness. A total of 120 schoolchildren participated in data collection. Their age and sex distribution were recorded. The best performance was achieved by the AdaBoost classifier, with 79.5% prediction accuracy. Competitive performance was provided by SVM and random forest classifier, which lagged by only a few percentage points. The classification accuracy of other evaluated classifiers, such as Naive Bayes, decision trees, k-nearest neighbours and logistic regression, was notably lower. The accuracy scores of the other two models were also quite high, 72.5% for SVM and 72.3% for Random Forest classifier. In classification, Semi-supervised learning occurs when a large amount of unlabelled data is available. In such a situation, how to enhance predictability of classification through unlabelled data is the focus.

In [6], Pooja Manghirmalani Mishra and Dr. Sushil Kulkarni propose a methodology based on Support Vector Machine of semi-supervised 5 learning and implement it on the case samples of learning disability. It is observed that about 10% of children enrolled in school have a learning disability. A curriculum-based test was designed with respect to the syllabus of primary-level school going children. This test was conducted in schools for collecting LD datasets for testing. Historic data for LD cases were collected from LD Clinics of Government hospitals where the tests were conducted in real-time medical environments. The system was fed with 11 input units which correspond to 11 different sections of the curriculum-based test. Dataset consists of 340 cases of LD children. This case study has been carried out on more than 300 real data sets with the attributes, which represents the symptoms of LD, takes binary values and more work need to be carried out on quantitative data, as that is an important part of any data set. The accuracy of the algorithm was found to be 84.615% approx.

In [7], Rehman Ullah Khan, Julia Lee Ai Cheng and Yin Bee Oon proposed an automated diagnostic and classification system. The system is trained by pre-classified data of 857 school children scores in spelling and reading. The twenty-fifth percentile was applied on the scores to label the data. The scores of the twenty-fifth percentile and below were marked as indicators of children who were likely to have dyslexia while the scores above the twenty-fifth percentile were considered to be indicators of children who were non-dyslexic. The system has three components or modules: the first module is a diagnostic module which is a pre-screening application that can be used by experts, trained users and parents for detecting the symptoms of dyslexia. The second module is classification, which classifies the kids into two groups, non-dyslexics and suspicious for dyslexia in spelling and reading. And the third module is an analysis tool for researchers. The results show that 23% of children were at risk for dyslexia in the training data and 20.7% in the testing data with 98% of accuracy.

In [8], M. Mahalakshmi and Dr. K. Merrilance used images for screening individuals who have high risk to dyslexia. This work also motivates the application of machine learning in distributed environment. The proposed predictive model uses the machine-learning algorithms like Decision Tree(DT) and Random Forest (RF). The model is classified using Weka tool and Python implementation. The Naive Bayes is employed in this study to predict dyslexia using brain scans. This research project implements Apache SPARK, an in-memory framework that handles the 6 storage, processing, and execution issues associated with large data sets. There were 150 brain MRI pictures in the sample, with ages ranging from 24 to 35. 50 of them have been diagnosed with dyslexia. Adult brain scans were chosen for the study because they would have progressed through the developmental reading stage, having been exposed to a variety of study materials and methodologies. K-fold cross validation is used as a validation model. Images are first transformed to grayscale for Dyslexia prediction. The scans are analysed for three features: grey matter, white matter, and cortical thickness. The prediction accuracy of 81.5% is achieved using Decision Tree and 97.6% using Random Forest Algorithm.

In [9], G. Vanitha and M. Kasthuri view at various dimensions of research towards dyslexia. This review finds the research holes, challenges and opportunities in this field and also encourages to use Machine Learning (ML) algorithms in this research area.

The study examined various research papers focused on the utilization of machine learning algorithms for predicting dyslexia and conducted a comprehensive analysis of the findings. This paper found out that generally, KNN, Random Forest and SVM are used for classification whose accuracy level is attained high. A combination of the above -mentioned methods is likely to provide better outcomes in detecting dyslexia.

In [10], Ms. Maitrei Kohli and Dr. T.V. Prasad proposes a systematic approach for identification of dyslexia and to classify or analyse potential cases more accurately and easily by use of ANN. This project is based on test data, designed to cover the evaluation results of potential dyslexic students, between the years 2003 – 2007. The results obtained suggested that out ANN model would perform better than the ones reported in the literature. The study takes into account student’s school assessment data for past five years. If the student has been continuously scoring very poor marks, then we presume that dyslexia identification tests are required. The next step is to fill up the questionnaires provided with our NN model. The questionnaires are automatically evaluated and results are split into one of the three ranges: (1) less than 33%, (2) between 33 and 45% or (3) greater than 45%. This project was implemented on a test and trial basis, and a maximum accuracy of 75% was obtained.

III. METHODOLOGY

The Python programming language in conjunction with various libraries served as the foundation for the effective data processing and analysis for this research. The libraries included are Matplotlib for data visualization, NumPy for numerical computations, TensorFlow for building and training machine learning models, Pandas for data manipulation and analysis, Keras for deep learning, Seaborn for advanced data visualization, and Scikit-learn for machine learning algorithms

A. Data Collection

The study utilized a secondary dataset that captures information about students’ performance on a computer-based test conducted by a special school in Kerala, aimed at predicting learning disability in their students. The dataset is in CSV format with 825 rows and 9 columns, where each column represents a different attribute of the test and each row provides specific information about a student’s performance. It contains 9 attributes, namely:

- 1) Type: This attribute represents the type of test conducted as part of the survey. It could include various types of assessments or evaluations administered to the students.
- 2) Test Name: This attribute indicates the specific name or identifier of the test taken by the students. It helps in identifying the particular examination or assessment carried out.
- 3) Round: The “Round” attribute signifies the round or session of the test. It can be used to distinguish between the multiple rounds conducted during the test.
- 4) Attend: This attribute captures the attendance information of the students during the test. It may indicate whether a student attended the test or if they were absent.
- 5) Question ID: The “Question ID” attribute denotes the identification number or code assigned to each question in the test. It helps in uniquely identifying individual questions within the dataset.
- 6) Answer ID: This attribute represents the identification number or code assigned to each answer option provided for the questions in the test. It aids in uniquely identifying the answer choices.
- 7) Marks: The “Marks” attribute records the marks or scores obtained by the students in the test. It ranges from 0 to 1, with 0 indicating a wrong answer and 1 indicating a correct answer.
- 8) Time: This attribute captures the time taken by students to complete the test or individual questions. It provides insights into the speed or efficiency of students in answering the questions.
- 9) Label: The “Label” is based on the attributes, marks and time. Table I illustrates its nature.

TABLE I

Time (seconds)	0-20		21-30		31-40		41-50		51-60	
Marks	0	1	0	1	0	1	0	1	0	1
Label	0	5	0	4	0	3	0	2	0	1

B. Data Preprocessing

Data preprocessing plays a vital role in preparing data for analysis, improving data quality, facilitating effective modelling, and ultimately leading to more accurate and reliable insights and predictions. The data was optimized by removing a few missing values, that happened to occur in the dataset, in order to ensure that the analysis is not biased or compromised due to incomplete data.

C. Model Training

Model training is a crucial step in the field of machine learning, where algorithms are trained to learn patterns and relationships from data in order to make accurate predictions or decisions. The training process involves optimizing the model's parameters or weights to minimize the difference between its predictions and the actual values or labels in the training dataset. It is essential to strike a balance between model complexity and overfitting to ensure the trained model can generalize well to unseen data and make accurate predictions in real-world scenarios. In order to assess the effectiveness of the machine learning models, we employed the train-test split function available in the sci-kit learn library. This function proved valuable in dividing the dataset into distinct training and test sets, with 80% of instances allocated for training purposes and the remaining 20% reserved for testing. Through random allocation, we ensured that both sets represented the overall dataset adequately, minimizing biases or systematic patterns in the splitting process. This partitioning technique allowed the machine learning models to be trained on a substantial portion of the dataset, enabling them to learn patterns and relationships within the training data. Following training, we evaluated the models' performance using the test set, which contained previously unseen data. The process was extremely helpful in providing insights into the models' ability to generalize to unfamiliar instances and make precise predictions on unseen data samples.

D. Algorithms Used

- 1) *Random Forest*: A popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. Random Forest is a collection of Decision Trees. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
- 2) *Support vector machines (SVM)*: A set of supervised learning methods used for classification, regression, and outliers' detection. SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes.
- 3) *K-Nearest Neighbour (KNN)*: KNN assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using KNN algorithm.
- 4) *Artificial Neural Network (ANN)*: ANNs are usually simply called neural networks (NNs) or neural nets. Commonly, ANN has an input layer, an output layer as well as hidden layers. The input layer receives data from the outside world which the neural network needs to analyse or learn 16 about. Then this data passes through one or multiple hidden layers that transform the input into data that is valuable for the output layer. Finally, the output layer provides an output in the form of a response of the Artificial Neural Networks to input data provided.
- 5) *Convolutional Neural Network (CNN)*: A type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. CNNs are trained using a large dataset of labelled images, where the network learns to recognize patterns and features that are associated with specific objects or classes. Once trained, a CNN can be used to classify new images, or extract features
- 6) *Long short-term memory (LSTM)*: A kind of recurrent neural network (RNN) used in the field of deep learning. A traditional RNN has a single hidden state that is passed through time, which can make it difficult for the network to learn long-term dependencies. LSTMs address this problem by introducing a memory cell, which is a container that can hold information for an extended period of time. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate.

IV. RESULT ANALYSIS

In this study, accuracy is the metric employed to assess the performance of classification models. The mentioned models are applied on the data to determine their accuracy. Based on our outcome, KNN is found to be the most accurate with 90.33% of accuracy followed by Random Forest with 79.22% of accuracy.

LSTM and CNN has an accuracy around 77% whereas SVM and ANN were found to have an accuracy of 57.57%.

TABLE II
ACCURACY TABLE

Proposed Model	Accuracy in Percentage
Random Forest	79.22
SVM	57.57
KNN	90.33
ANN	57.57
CNN	77.69
LSTM	77.57

The specific accuracy of each model will depend on a range of factors such as the size of the dataset, the complexity of the problem being solved, and the particular parameters of the model being used. Nonetheless, the algorithms used here represent powerful tools for extracting patterns and making predictions from large, complex datasets.

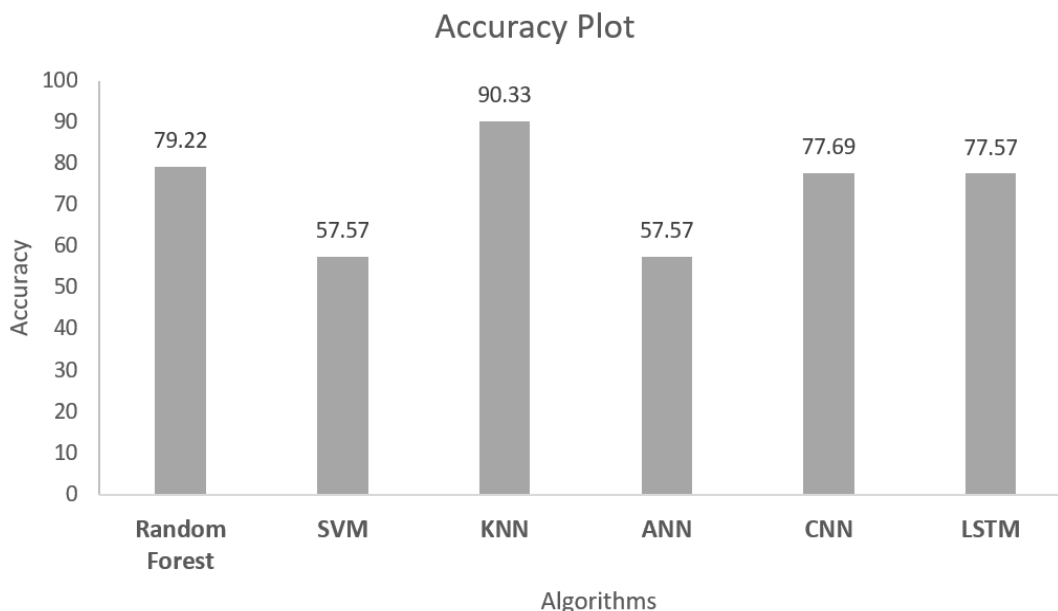


Figure 1: Comparison of accuracy for different models

V. CONCLUSION

A learning disability is a neurological condition which affects the brain’s ability to send, receive, and process information. A child with a learning disability may have difficulties in reading, writing, speaking, listening, understanding mathematical concepts, and with general comprehension. Numerous Machine Learning and Deep Learning techniques are utilized to predict learning disabilities in children. Early diagnosis and intervention are crucial for individuals with learning disabilities to receive the support they need to succeed in school and life. To determine the most accurate algorithm for a given dataset, a comparative analysis of existing machine learning models can be conducted. Adequate preprocessing of the data before prediction is a crucial step that cannot be overlooked. Innovative strategies can be developed to enhance the accuracy of predictions. This study mainly focuses on predicting learning disability on a collected secondary dataset using machine learning and deep learning algorithms thereby comparing the accuracy between them. The algorithms used here are K-Nearest Neighbour, Random Forest, Long Short-Term Memory, Support Vector Machine, Artificial Neural Network and Convolutional Neural Network. It is found that between the algorithms used KNN has highest accuracy with 90.33% followed by Random Forest whose accuracy is 79.22%. Thus, with the help of machine learning and deep learning models early diagnosis of learning disability can be achieved to a certain level.



REFERENCES

- [1] A.Devi, Dr. G. Kavya, M. Julie Therese and R. Gayathri. Early Diagnosing and Identifying Tool for Specific Learning Disability using Decision Tree algorithm. Proceedings of the Third International Conference on Inventive Research in Computing Applications (ICIRCA-2021).
- [2] H. Atakan Varol, Subramani Mani, Donald L. Compton, Lynn S. Fuchs, and Douglas Fuchs. Early Prediction of Reading Disability using Machine Learning. AMIA Annu Symp Proc. 2009; 2009: 667–671.
- [3] K. Ambili and P. Afsar. A Framework for Learning Disability Prediction in School Children using Naive Bayes - Neural Network Fusion Technique. ISSN: 0975 – 6760— NOV 15 TO OCT 16— VOLUME – 04, ISSUE – 01.
- [4] Julie M. David and Kannan Balakrishnan. Machine Learning Approach for Prediction of Learning Disabilities in School-Age Children. International Journal of Computer Applications. 9. 10.5120/1432-1931
- [5] Drotar, P., Dobes, M. Dysgraphia detection through machine learning. Sci Rep 10, 21541 (2020). <https://doi.org/10.1038/s41598-020-78611-9>
- [6] Pooja Manghirmalani Mishra and Dr. Sushil Kulkarni. Classification of Data using Semi-Supervised Learning (A Learning Disability Case Study). Pooja Manghirmalani Mishra IJCET Volume 4, Issue 4, July-August (2013), pp. 432-440
- [7] Khan, Rehman and Lee, Julia and Oon, Yin Bee. (2018). Machine Learning and Dyslexia: Diagnostic and Classification System (DCS) for Kids with Learning Disabilities. January 2018.
- [8] M. Mahalakshmi and Dr. K. Merrilance. Prediction of Dyslexia using Machine Learning Algorithms. IJCRT -Volume 10, Issue 5 May 2022 -ISSN: 2320-2882.
- [9] J G. Vanitha and M. Kasthuri. Dyslexia Prediction Using Machine Learning Algorithms – A Review. International Journal of Aquatic Science, ISSN: 2008-8019, Vol 12, Issue 02, 2021.
- [10] Ms. Maitrei Kohli and Dr. T.V. Prasad. Identifying Dyslexic Students by Using Artificial Neural Networks. Proceedings of the World Congress on Engineering 2010 Vol I -WCE 2010, June 30 - July 2, 2010, London, U.K.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)