**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 36, No. 2, November 2024, pp. 994~1001
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i2.pp994-1001 

# Performance of dyslexia dataset for machine learning algorithms

**J. Jincy[1,2], P. Subha Hency Jose[1]**
[1]Department of Biomedical Engineering, Faculty of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
[2]Department of Electronics and Communication Engineering, CSI College of Engineering, Ketti, India

**Article Info**

***Article history:***

Received Feb 22, 2024
Revised Jun 25, 2024
Accepted Jul 14, 2024

***Keywords:***

Dyslexia
EEG
K nearest neighbor
Machine learning
SVM

**ABSTRACT**

Learning disability is a condition usual amongst most populace due to poor phonological capability in humans making them impaired. One such neurological disorder is developmental dyslexia, a lack of reading and writing skills leading to difficulty in school education. The essential causes of developmental dyslexia are the consumption of more drug treatments during pregnancy, the over-the-counter purchase of medicines for minor ailments without the recommendation of physicians, and uncared-for head accidents during early life. The occurrence of this trouble is acute in India. Attempts were made by many to detect dyslexic children to reduce the intensity of this hassle. In this proposed effort, machine learning is used to locate significant styles characterizing people using EEG samples. A dataset is used for examination of developmental dyslexia, and classification is done using K nearest neighbor (KNN), decision tree, linear discriminant analysis (LDA), and support vector machine (SVM) to evaluate the performance. This piece of research work is done on MATLAB to provide results on simulation with classification accuracy of 90.76% for SVM, sensitivity of 89% for SVM, and LDA with 91.89% specificity for SVM providing optimum yield.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

P. Subha Hency Jose
Department of Biomedical Engineering, Faculty of Biomedical Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
Email: hency20002000@karunya.edu

## 1. INTRODUCTION

A neurological disorder termed dyslexia affects 5% to 15% of Indian kids, who are labeled as lazy because of their incapability in reading, writing, mathematic skills [1]. They yearn to be appreciated and accepted by society turning their life into social trauma. The learning disabilities like reading and writing curbs the life of young children's education and future endeavor. Genetic research shows that dyslexia is inherited from the family. Around 23-65% of children have taken over from the genes of their parents unnoticeably. Learning disabilities are of predominant type’s dyslexia, dyscalculia and dysgraphia. Dyslexia is a selected problem related to information sounds and phrases due to a lack of phonological processing [2], Dysgraphia is related to deficiency in writing words and scripts highly difficult to decode [3], Dyscalculia is an arithmetic disorder that causes poor mathematical and logical capacity [4].

There are many standardized tests for analysis of dyslexia with regard to reading, writing, spelling abilities, mental caliber, and working memory. Glancing at the tests the severity of the infirmity can be identified. The weak linguistic abilities can be assessed using word test and questionnaire connected to reading and writing based on clinical observation [5]. In addition to IQ tests neurobiological behavior in brain structure can be analyzed by using imaging tools and their behavior could be understood. Neural connectivity varies for dyslexic and normal children altering their brain pattern. Functional magnetic resonance imaging (FMRI) is a technique used to analyze word recognition based on changes in the blood flow in the frontal and occipital regions. The tests are based on images and words that are used regularly [6]. FMRI has been very beneficial it has drastic pitfalls that make the real neural pastime identification hard consequently leading to a maximum modified model this optimization set of rules complements the signal excellent within the vicinity of interest inside the mind [7]. Eye-tracking technology has proved exceptional strategies for characterizing dyslexia among regular and dyslexic readers. Eye movement is used to track and machine learning methods are used to assess the functions relating to fixations and saccades. The solution to dyslexia can be provided by early intervention with cerebral morphology which gains high accuracy with economic impact but quiet jerky [8]. Being a gender-oriented disorder functional and structural development and morphological study was done by EK Lambe [9] showed a different brain activation pattern between dyslexic males and females, which is influenced on analysis.

Chyl *et al.* [10] studied neuroimaging in the language areas of brain provides the greatest challenge and recommends future enhancement in the gray areas of research. Biscaldi *et al.* [11] explained saccades in five non-cognitive tasks. The criterion of the eye-movement data was composed of twelve persons who are considered to be dyslexic and were grouped into two groups (D1 and D2) based on various metrics. Comparing both groups, more details on their saccades and fixations were received. The standard tasks of the dyslexic subjects DysLexML is a machine learning tool used to classify dyslexia based on eye movement for a small feature set, which was found to be very accurate in the presence of noise using machine learning algorithms [12]. Small samples and small effects also provide efficient dyslexia treatment studies, according to researchers from Italy, allowing one to reach adequate power [13]. Electroencephalogram (EEG) is a non-invasive method that provides promising images of the cortical parts of the brain at low cost in response to various activities like reading and writing. On investigation done EEG-based classification framework provides a pattern that has meaningful data that can be arranged in a suitable order and classified using machine learning algorithms [14]. On systematic analysis of the existing research the following gaps are being identified as most of the diagnosis method are based on conventional IQ test and assistive technology, the entire cause of the condition is not understood [15]. Even though dyslexia can be predicted different ways the vast separation between the controls provide clear empathy towards the diseased. Smaller samples of the information cannot provide a generalization but rather large sample set can provide depth of the unfolding [16]. A novel method of dyslexia analysis is put forward to bring an end to blind belief on dyslexia with early intervention, possibility of detection using dataset and study the disorder in a neurological perspective with machine learning.

## 2. BACKGROUND STUDY

The EEG received from a dyslexic manager undergoes statistical analysis before classifying the output. The important steps taken to technique the uncooked signal are: i) preprocessing and artifact removal, ii) fact analysis using feature extraction, and iii) classification to get the preferred final results.

### 2.1. Preprocessing

EEG statistics collected from the use of 64 channels in the eyes closed notion for 2 minutes are sampled at 250 Hz. Data is imported into the mind-vision analyzer, spline interpolation is carried out, and immoderate artifacts are eliminated. The EEG signal undergoes extraction of the signal of interest using the fast fourier transform (FFT), decomposing it into four frequency bands: delta (0.5–4 Hz); theta (4–8 Hz); alpha (8–13 Hz); beta (13–30 Hz) [17]. The Biosemi device recorded EEG indicators sampled continuously at 2048 Hz, segmented off-line into 1.75-second epochs. The event-related potential (ERP) indicators that were obtained were baseline corrected with the aid of averaging the signals the cause of pre-processing is to put off unwanted noise and organize the signals based on applicable features [18]. Followed by pre-processing the values are assigned numerical expressions. The variety of capabilities varies from 12-256 [19].

### 2.2. Feature extraction

The mathematical tool for analyzing EEG is the FFT, where the power spectral density is calculated for all four frequency bands. The periodogram is generated from the correlated sequence. The discrete wavelet transform is a systematic method used for feature extraction. Out of the seven Daubechies (db), one proves to be an efficient one for dyslexia analysis [20]. The EEG signal is split into different frequencies using two sets of functions. The approximate coefficient is further decomposed into 5 levels, and the third detailed level of power gives 16–32 Hz, which is the beta band of the EEG signal that increases for dyslexics [21]. EEG based analysis can be done in time domain, frequency and time frequency domain innovative feature extraction can provide accuracy and reliable results Using assistive technologies like BCI provides enhanced results when channel selection is done using multi-channel BCI.

### 2.3. Feature reduction

Different data reduction methods comprise independent components analysis (ICA), principal components analysis (PCA), and discriminant analysis (LDA). High dimensional data is grouped into lower dimension a by forming subset such for training and testing, on which analysis can be done [22]. As the category of data is of low quality with redundant feature and noisy features data is cooked before analyzing [23]. The two vague categories, namely dyslexic and regular readers, are classified using reduced redundant features and less complexity with better accuracy.

### 2.4. Classification

Machine learning algorithms are proven classifiers used for clinical diagnostics. The outcome of the classifier may be skilled using supervised or unsupervised training and accuracy, sensitivity and specificity may be obtained. support vector machines (SVM), neural networks, decision trees, Bayesian classifiers, K means clustering, and logistic regression are varied classifiers that provide classification of metrics [24]. Usman *et al.* [25] did an analysis on the primary device mastering biomarkers and challenges primarily based on the output of twenty-two decided on articles the use of PRISMA. They concluded that SVM is proved to be the first-class classifier that offers nice outcomes regardless of extraordinary assets of records. In order to enable the open university for developing suitable courses for training low-engaged students a suitable model was developed and machine learning algorithms SVM, decision tree, gradient boosted, Navie bayes classifiers were used. On analyzing the statistical parameters kappa, recall, and accuracy was obtained. Logistic regression is an effective technique to clear up the category problem and to get the expected results. Thus, final results are continually based totally on the choice of the proper version to resolve a specific problem [26].

### 2.5. Performance evaluation

Machine learning algorithms are proven classifiers used for clinical diagnostics. The outcome of the classifier may be skilled using supervised or unsupervised training, and accuracy, sensitivity, and specificity may be obtained. SVM, neural networks, decision trees, Bayesian classifiers, K-means clustering, and logistic regression are classifiers that provide the classification of facts. Figure 1 demonstrates the various level of analysis of EEG signals with artifacts and noises. On the other hand, the input dataset for the dyslexia classification in this hypothesis comprises interval values that are initially transformed into midpoints and then into an intuitionistic fuzzy domain representation. The selected instance's missing values are chosen, and the remaining characteristics that are compared to those of other examples in the full set. Data is divided into training and test sets using an appropriate machine learning technique, and the classification parameters are confirmed.

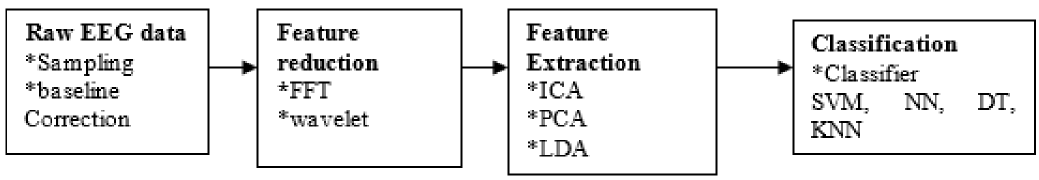


Figure 1. Various stages for EEG signal processing

## 3. PROPOSED SYSTEM

This framework, illustrated in Figure 2, provides a foundation for the interconnected approach, which renders it more likely to diagnose dyslexia accurately. The dataset is implementation uses MATLAB 2013a software with 4 GB RAM and 2.30 GHZ processor. The machine learning algorithm provided the finest solution in the dyslexia detection inside the KEEL datasets and identify dyslexia and normal controls. The dataset encompasses attributes count of 12 and no output labels is 2 with missing values (X/Y) which are replaced by determining KNN between their obtained mean value from the nearest neighbor. The acquired dataset had undergone up sampling and down sampling. The empirical study takes place by splitting the data into training set and test set with 10 fold cross validation to obtain the dyslexic and normal controls. The actual class obtained is compared with predicted class and accuracy, sensitivity and specificity are obtained.

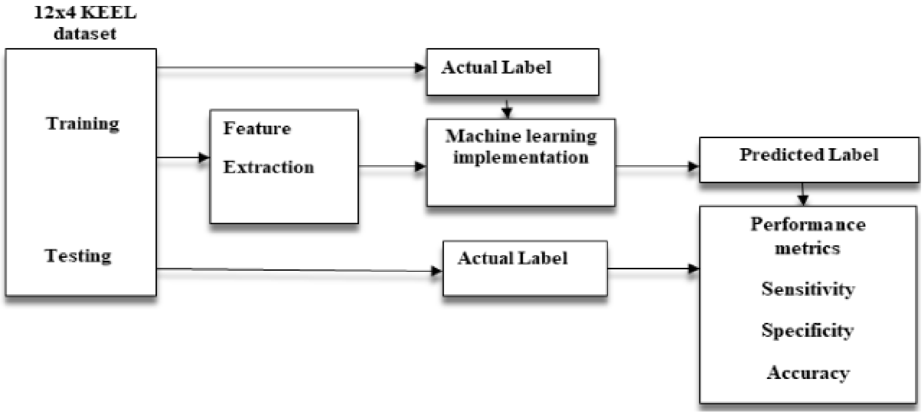


Figure 2. Proposed block diagram

For medical teams to gain social benefits, these combined dyslexia prediction model employing the KEEL data set offers an evaluation of a comparable experiment [27]. By combining numerous feature components, it can offer meaningful information on the underpinning trends and features of the data. Out of the scarcity of data this early diagnosis method provides a platform for giving confidence in journey of our study with limited resources.

## 4. METHODS

Our suggested approach works by using a machine learning algorithm to identify dyslexia prediction hotspots and adapt research objectives accordingly. Using a trained algorithm, the unlabeled test dataset was mapped to identify similar classes [28]. Based on decision tree algorithm investigated by Turkish professional [29] using CART and Chi square model has low error rate and can be used for more generalized data. KNN variants showed that it can be used for any dataset owing to the fact of its versatility and open to change. It has less bias and high potential to create accurate classification for greater value of K [30]. The researchers from Pakistan proposed SVM approach furnishes that heterogeneous data can be classified with optimal sensitivity and specificity [31]. Hu *et al.* [32] practiced LDA to obtain dimensionality reduction, providing effective for sample data with low cost. Our classifier's overall performance is considered based on the confusion matrix and ROC (receiver's working curve).

In LDA is a classifier with [33], multi classes (dependent variables) are described based on target (independent variables). The different 12 attributes of dyslexia are classified based on the on the target classes. The gradient based LDA with local minima is used to reduce the cost function caused by the actual and predicted output iteratively. The learning rate is chosen such that convergence is obtained with least oscillations. It can be scalable for other larger dataset and flexible. Algorithm for LDA shown in Algorithm 1.

Algorithm 1. LDA
```
1:Pick A Random Starting Point X=Random(X)
2:Assign The Valuethreshold=0.000001
While Condition Is True;
Gradient = Compute_Gradient(X)
Next_X=Step(X, Gradient, Alpha=-0,001)
If Distance (Next_X, X)<Threshold ;
3: Assign Gradient Negative Step
4:When Converging Attained Process Stop
Break;
5: Continue If We Are Not Return X X=Next_X
```

SVM approach outperforms LDA based on decision boundary and error rate thus stands superior [34]. Thus, imported dataset undergoes train-test split and 10 fold cross validation and using Radial bias function the attributes are trained, classification is done to obtain optimal accuracy. Algorithm for SVM shown in Algorithm 2.

Algorithm 2. SVM
```
1: Start the partitioning
2: Check For j=1:k
3: determine the training set= k -1subsets;
4: Assume Testing set=remaining subsets;
```

```
5: Obtain Parameter_ optimization (k);
6: Again test on testing set & End the for statement;
7: Return accuracy of entire dataset
Decision Tree works as follows:
```

As we deal with low quality data decision tree classifiers are efficient in classifying missing data simplifying into simpler models with robust exceptions [35] from the dataset subset are formed as key 1 list and key 2 list based on specific attribute. Further splitting is done by measuring entropy value based on positive and negative classes maximum likelihood training is to reduce the complexity. Algorithm for decision tree shown in Algorithm 3.

Algorithm 3. Decision tree

```
1: Start keeping first attributes and the class attribute.
2: Compare the attribute name from the key1 list and key2 list, where key1 is the list to store attributes names based on the ascending order of the entropy value, and key2 is the list to store attributes names in original order.
3: Both are same then remove the attributes from the dataset and also remove the attribute from the key2 list and evaluate.
4: Do step until last attributes in the dataset.
```

K means clustering is unsupervised clustering algorithm in which optimal euclidean distance is calculated in dataset. It provides a preeminent recall rate for medical dataset [36]. Random K was initialized and mean was calculated. The mean coordinated is updated and average is calculated. Repeatedly iterating we get the cluster of dyslexic and normal readers. Algorithm for K means clustering shown in Algorithm 4.

Algorithm 4. K-means clustering

```
1: Calculate “d(x, xi)” i =1, 2… n; where d denotes the Euclidean distance between the points.
2: Arrange the calculated n Euclidean distances in non-decreasing order.
3: Let k be a positive integer, take the first k distances from this sorted list.
4: Find those k-points corresponding to these k-distances.
5: Let ki denotes the number of points belonging to the ith class among k points i.e. k ≥ 0
6: If ki >kj∀i ≠ j then put x in class i.
```

Summarizing KNN provides higher error prediction rate and with fast response. Dyslexia data set is large; SVM with higher dimensional space and non linear models can provide efficient methods of comparison. As machine is trying to replace humans in medical field less intervened setup towards prediction of dyslexia can uphold the society which in need of proper handling. Deploying the above models can provide a platform on which analysis can be built for a smarter society.

## 5. RESULTS AND DISCUSSION

Our work utilizes Keel respository data set for dyslexia analysis using machine learning algorithms and 2×2 confusion matrix is gained using MATLAB tool. The dataset is low quality with 65 instances with 12 attributes crisp and vague values. The performance measurements produced for every machine learning method that is evaluated are presented in contingency tables. In the matrix, each row represents an actual class occurrence, and each column represents instances of a forecast class. The confusion matrix's row i and column j elements indicate the number of instances in which the predicted class is j and the actual class is i. Table 1 shows the results of the performance of the LDA algorithm in classifying dyslexia using the two labels targets or outcome. Our model recognized, 25 dyslexic individuals and grouped as dyslexic ($T_p$) while 32 skilled readers are identified as skilled reader ($T_n$). Nevertheless 5 dyslexic individuals are misclassified as skilled readers ($F_n$) and 3 skilled readers are misclassified as dyslexic ($F_p$) with a wide margin of separation.

In par with it Table 2 depicts the results of the SVM algorithm in classifying dyslexia with optimal line. Our model recognized, 24 dyslexic individuals and grouped as dyslexic ($T_p$) while 31 skilled readers are identified as skilled reader ($T_n$). Nevertheless 6 dyslexic individuals are misclassified as skilled readers ($F_n$) and 3 skilled readers are misclassified as dyslexic ($F_p$).

Table 1. Contingency matrix for linear discriminate analysis

| Actual vs predicted | Dyslexic | Skilled readers |
|---|---|---|
| Dyslexic | 25 ($T_p$) | 5 ($F_n$) |
| Skilled Readers | 3 ($F_p$) | 32 ($T_n$) |

Table 2. Contingency matrix for support vector machine

| Actual vs predicted | Dyslexic | Skilled readers |
|---|---|---|
| Dyslexic | 25 ($T_p$) | 6 ($F_n$) |
| Skilled readers | 3 ($F_p$) | 31 ($T_n$) |

A Decision tree classifier presented in Table 3 is used for handling nonlinear relationship for classifying dyslexia. Our model recognized, 25 dyslexic individuals and grouped as dyslexic ($T_p$) while 31 skilled readers are identified as skilled reader ($T_n$). Nevertheless 6 dyslexic individuals are misclassified as skilled readers ($F_n$) and 3 skilled readers are misclassified as dyslexic ($F_p$) with a good interpretability than SVM to overfit training data. Table 4 represent the performance of the KNN with, 31 skilled readers are identified as skilled reader (Tn). Nevertheless 6 dyslexic individuals are misclassified as skilled readers (Fn) and 3 skilled readers are misclassified as dyslexic (Fp).

Table 3. Contingency matrix for decision tree classifier

| Actual vs predicted | Dyslexic | Skilled readers |
|---|---|---|
| Dyslexic | 25($T_p$) | 6 ($F_n$) |
| Skilled readers | 3 ($F_p$) | 31 ($T_n$) |

Table 4. Contingency matrix for KNN classifier

| Actual vs predicted | Dyslexic | Skilled readers |
|---|---|---|
| Dyslexic | 24 ($T_p$) | 6 ($F_n$) |
| Skilled readers | 3 ($F_p$) | 31 ($T_n$) |

This kind of data visualization displays the different classifiers' performances. The performance metrics from the confusion matrix is be combined to get accuracy, sensitivity, and specificity:

$$\textit{Total Accuracy} = \sum_{i=1}^{n}(TP, Tn)/TotalNumberofTest\ data \quad (1)$$

$$Specificity = Tn/(FP + Tn) \quad (2)$$

$$Sensitivity = TP/(TP + Fn) \quad (3)$$

$T_p$=dyslexic classified as dyslexic
$F_p$=skilled readers misclassified as dyslexic
$T_n$=skilled reader classified as skilled readers
$F_n$=dyslexic misclassified as skilled readers
Comparing different performances metrics deriving from Table 5, SVM outperforms other algorithm with an accuracy rate of 90.76%, and specificity of 91.89%. On diving further the sensitivity of LDA is 89.28% near to SVM demonstrating a specific problem identification.

Table 5. Algorithm models compression result

| Algorithm | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---|---|---|---|
| KNN | 85.7 | 86.48 | 86.15 |
| Decision tree | 78.57 | 83.78 | 81.53 |
| LDA | 89.28 | 86.48 | 87.69 |
| SVM | 89.28 | 91.89 | 90.76 |

The graphical representation chart in Figure 3 proves that despite of the low-quality dataset, boosted decision tree classifier is dealing with missing values providing higher sensitivity [37]. SVM provides wide separation between the classes of the input paving way for higher accuracy [38]. Being an introductory classifier KNN it has addressed the pattern recognition problem by considering the data in the neighbor data points for smaller sample with significant specificity.

As ample of machine learning algorithm has been reviewed our study uses few machine learning algorithms that provides significant specificity and sensitivity. While earlier studies have explored the impact of dyslexia in children in the different age groups and gender, they have not explicitly addressed its influence on early childhood between 5-10 years. Moreover, the publicly available dataset is having small samples comprimising of eye tracking, standardized test, FMRI and questionnaire. This study is tended to have a greater influence on utilization of machine learning algorithm to give accurate prediction without bias. We found that the performance metrics of various classifiers are correlated and successive training and testing exhibit their uniqueness. The proposed method may benefit the economic development of our country without adversely impacting illiteracy rate. However, hybridization can be provided to enhance the performance of detection technique with efficient feature reduction techniques.

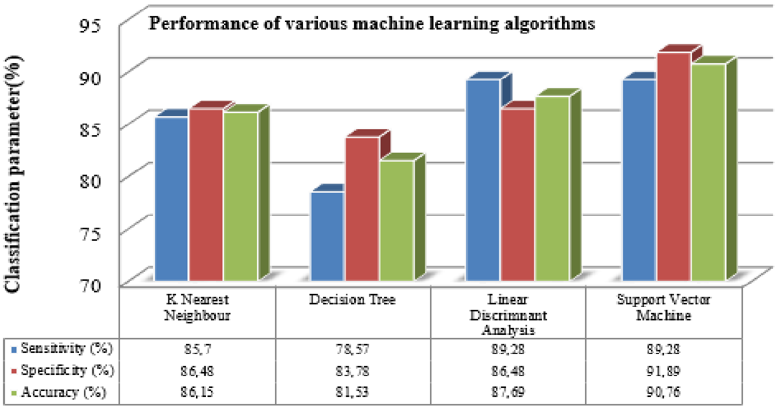


Figure 3. Graphical representation of various algorithms

## 5. CONCLUSION

Even though ample of testing methods of dyslexia is available our scope deals with dyslexia diagnosis using EEG in a pinch, the researchers and medical practitioners, can crack up on the traditional methods. Our study insist on learning disorder being a neurological condition can be studied using latest technology with accurate, specific result with high degree sensitivity. This various machine learning approach is an eye opener to the Special education teachers, doctors, and children. In future this condition is looked upon as a deficiency that is curable rather than causing setback in school education making our country doomed to darkness.

## ACKNOWLEGMENT

Authors thank their college management for providing their support and infrastructure towards the completion of the research work. We also express our sincere gratitude, to sincere hearts who shed their guiding light to accomplish this study with dedication and hard work.

## REFERENCES

[1] T. N. I. Express, “Dyslexia not just a mental health problem, it’s a social problem, says expert,” 2020.
[2] J. Jincy and P. S. H. Jose, “Survey on intervention and diagnosis of dyslexia,” *International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Feb. 2021, doi: 10.1109/ICAECT49130.2021.9392606.
[3] P. Drotár and M. Dobeš, “Dysgraphia detection through machine learning,” *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-78611-9.
[4] S. Haberstroh and G. Schulte-Körne, “Diagnostik und Behandlung der rechenstörung,” *Deutsches Arzteblatt International*, vol. 116, no. 7, pp. 107–114, Feb. 2019, doi: 10.3238/arztebl.2019.0107.
[5] M. Poulsen, H. Juul, and C. Elbro, “A national test of dyslexia,” *Annals of Dyslexia*, vol. 73, no. 3, pp. 337–355, Jul. 2023, doi: 10.1007/s11881-023-00285-5.
[6] G. H. Glover, “Overview of functional magnetic resonance imaging,” *Neurosurgery Clinics of North America*, vol. 22, no. 2, pp. 133–139, Apr. 2011, doi: 10.1016/j.nec.2010.11.001.
[7] S. Mohamad, W. Mansor, and K. Y. Lee, “Review of neurological techniques of diagnosing dyslexia in children,” in *Proceedings - 2013 IEEE 3rd International Conference on System Engineering and Technology, ICSET 2013*, Aug. 2013, pp. 389–393, doi: 10.1109/ICSEngT.2013.6650206.
[8] A. Jothi Prabha and R. Bhargavi, “Predictive model for dyslexia from fixations and saccadic eye movement events,” *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105538, Oct. 2020, doi: 10.1016/j.cmpb.2020.105538.
[9] E. K. Lambe, “Dyslexia, gender, and brain imaging,” *Neuropsychologia*, vol. 37, no. 5, pp. 521–536, Feb. 1999, doi: 10.1016/S0028-3932(98)00146-8.
[10] K. Chyl, G. Fraga-González, S. Brem, and K. Jednoróg, “Brain dynamics of (a)typical reading development—a review of longitudinal studies,” *npj Science of Learning*, vol. 6, no. 1, Feb. 2021, doi: 10.1038/s41539-020-00081-5.
[11] M. Biscaldi, S. Gezeck, and V. Stuhr, “Poor saccadic control correlates with dyslexia,” *Neuropsychologia*, vol. 36, no. 11, pp. 1189–1202, Nov. 1998, doi: 10.1016/S0028-3932(97)00170-X.
[12] T. Asvestopoulou *et al.*, “Dyslexml: Screening tool for dyslexia using machine learning,” *arXiv preprint arXiv:1903.06274*, 2019.
[13] E. Toffalini, D. Giofrè, M. Pastore, B. Carretti, F. Fraccadori, and D. Szűcs, “Dyslexia treatment studies: a systematic review and suggestions on testing treatment efficacy with small effects and small samples,” *Behavior Research Methods*, vol. 53, no. 5, pp. 1954–1972, Mar. 2021, doi: 10.3758/s13428-021-01549-x.
[14] H. Perera, M. F. Shiratuddin, and K. W. Wong, “Review of EEG-based pattern classification frameworks for dyslexia,” *Brain Informatics*, vol. 5, no. 2, Jun. 2018, doi: 10.1186/s40708-018-0079-9.
[15] M. J. Snowling, C. Hulme, and K. Nation, “Defining and understanding dyslexia: past, present and future,” *Oxford Review of Education*, vol. 46, no. 4, pp. 501–513, Jul. 2020, doi: 10.1080/03054985.2020.1765756.
[16] K. Vasileiou, J. Barnett, S. Thorpe, and T. Young, “Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period,” *BMC Medical Research Methodology*, vol. 18, no. 1, Nov. 2018, doi: 10.1186/s12874-018-0594-7.

[17] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A review on mental stress assessment methods using eeg signals," *Sensors*, vol. 21, no. 15, p. 5043, Jul. 2021, doi: 10.3390/s21155043.
[18] A. Frid and Larry M. Manevitz, "Features and machine learning for correlating and classifying between brain areas and dyslexia," *arXiv preprint arXiv:1812.10622*, 2018.
[19] A. L. Golande and T. Pavankumar, "Early prediction of heart abnormalities using robust pre-processing and features extraction algorithm," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 7, pp. 380–394, Jul. 2020, doi: 10.5373/JARDCS/V12I7/20202019.
[20] S. Parmar and C. Paunwala, "A novel and efficient wavelet scattering transform approach for primitive-stage dyslexia-detection using electroencephalogram signals," *Healthcare Analytics*, vol. 3, p. 100194, Nov. 2023, doi: 10.1016/j.health.2023.100194.
[21] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography signal processing: a comprehensive review and analysis of methods and techniques," *Sensors*, vol. 23, no. 14, p. 6434, Jul. 2023, doi: 10.3390/s23146434.
[22] W. Chen, Y. Cai, A. Li, Y. Su, and K. Jiang, "EEG feature selection method based on maximum information coefficient and quantum particle swarm," *Scientific Reports*, vol. 13, no. 1, Sep. 2023, doi: 10.1038/s41598-023-41682-5.
[23] R. Nisbet, G. Miner, and K. Yale, "A data preparation cookbook," in *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier, 2017, pp. 727–740.
[24] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010, doi: 10.1016/j.eswa.2010.06.065.
[25] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, "Advance machine learning methods for dyslexia biomarker detection: a review of implementation details and challenges," *IEEE Access*, vol. 9, pp. 36879–36897, 2021, doi: 10.1109/ACCESS.2021.3062709.
[26] S. Sperandei, "Understanding logistic regression analysis," *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.
[27] N. D. Alqahtani, B. Alzahrani, and M. S. Ramzan, "Deep learning applications for dyslexia prediction," *Applied Sciences (Switzerland)*, vol. 13, no. 5, p. 2804, Feb. 2023, doi: 10.3390/app13052804.
[28] R. Alanazi, "Identification and prediction of chronic diseases using machine learning approach," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–9, Feb. 2022, doi: 10.1155/2022/2826127.
[29] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, p. 100130, Nov. 2023, doi: 10.1016/j.health.2022.100130.
[30] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, Apr. 2022, doi: 10.1038/s41598-022-10358-x.
[31] S. M. S. Shah, F. A. Shah, S. A. Hussain, and S. Batool, "Support vector machines-based heart disease diagnosis using feature subset, wrapping selection and extraction methods," *Computers and Electrical Engineering*, vol. 84, p. 106628, Jun. 2020, doi: 10.1016/j.compeleceng.2020.106628.
[32] L. Hu, J. Xu, L. Tian, and W. Zhang, "Self-centralized jointly sparse maximum margin criterion for robust dimensionality reduction," *Knowledge-Based Systems*, vol. 206, p. 106343, Oct. 2020, doi: 10.1016/j.knosys.2020.106343.
[33] A. Sharma and K. K. Paliwal, "Cancer classification by gradient LDA technique using microarray gene expression data," *Data and Knowledge Engineering*, vol. 66, no. 2, pp. 338–347, Aug. 2008, doi: 10.1016/j.datak.2008.04.004.
[34] R. Madhanagopal, R. C. Avinaash, and K. Karthick, "Comparison of support vector machines and linear discriminant analysis for indian industries," *Statistika Mathematika*, vol. 4, no. 3, pp. 74–80, 2012.
[35] Y. Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, p. 130, 2015.
[36] V. Sheth, U. Tripathi, and A. Sharma, "A comparative analysis of machine learning algorithms for classification purpose," *Procedia Computer Science*, vol. 215, pp. 422–431, 2022, doi: 10.1016/j.procs.2022.12.044.
[37] J. Loveline Zeema and D. Francis Xavier Christopher, "Evolving optimized neutrosophic C means clustering using behavioral inspiration of artificial bacterial foraging (ONCMC-ABF) in the prediction of dyslexia," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1748–1754, May 2022, doi: 10.1016/j.jksuci.2019.09.008.
[38] P. J. Rose, "Uncertainty based chaotic pigeon inspired optimized feature selection for effective Dyslexia prediction using density peak clustering," *International Journal of Computing and Digital Systems*, vol. 14, no. 1, 2023.

## BIOGRAPHIES OF AUTHORS

**J. Jincy** received her UG in Department of Electronics and Instrumentation Engineering at Karunya University, India. She did her PG in Embedded system in the same institute in the year 2008. Being an active Academician working in CSI college of Engineering, Ketti, India for 14 years is working for EEG based dyslexia analysis using machine learning algorithms. Her areas of interest are biomedical signal processing and machine learning. She published paper in international conference with the title survey on intervention and diagnosis of dyslexia, exploration of learning disorder using machine learning and preliminary study of dyslexia using machine learning. She can be contacted at email: jincydhas22@gmail.com.

**P. Subha Hency Jose** is associate professor at Karunya Institute of Technology and Science, Karunya University, India. She holds Ph.D. degree in process control engineering with specialization in biomedical instruments and artificial intelligence in health care. Being an eminent researcher has published paper in 43 international journals, 38 international conferences, with 28 Scopus citations. Her research project include fetal ECG monitoring in rural areas, health monitoring for pregnant women in rural areas with emergency trigger for visually impaired and so on. As a successful mentor she has completed guiding 38 M Tech scholars and 6 research scholar on for pursuing thesis work. She can be contacted at email: hency20002000@karunya.edu.