International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Constructive Model for Dyslexia and ADHD Prediction using Feature Learning and Boosting Approaches

Pavan Kumar Varma Kothapalli*[1], V. Rathikarani[2], Gopala Krishna Murthy Nookala[3]

Submitted: 16/07/2023 **Revised**: 08/09/2023 **Accepted**: 21/09/2023

Abstract: Dyslexia and Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurological disorder characterized by speech and reading impairments. The disorder is commonly identified in school-aged people, most commonly in males and causes poor performance with low self-esteem. Based on the review, it is noted that there are various machine learning approaches are used for the prediction process, and the validation is done with the available dataset. However, the prediction process is complex in this cause due to the lack of a standard dataset, biologically-interpretable biomarkers, classifiers, under-fitting, over-fitting issues and so on. To successfully implement a better CDSS, some preliminary process needs to be done to enhance the prediction rate. It includes: data acquisition, pre-processing, and data augmentation process. Here, the available online dataset for dyslexia is occupied from the UCI Machine Learning source. Some appropriate features of dyslexia are acquired using Least Absolute Shrinkage and Selection Operator (LASSO) and Fisher-score Relief Model. A hybrid AdaBoosting is developed by integrating the conventional classifiers with the bagging and boosting model. The bagging and boosting process is considered during the training process. In this case, the model is simulated using the MATLAB 2020a simulation environment, and the performances are assessed to demonstrate the model's importance. Metrics including accuracy, error rate, precision, recall, and F1-score are examined together with other statistical measurements. To further compare these findings with currently used methods, these results are supplied. This investigation demonstrates that the expected model achieves a better trade-off and a higher level of prediction accuracy.

Keywords: *Dyslexia, ADHD, neurological disorder, Ada-boosting, fisher-score, least absolute shrinkage and selection operator.*

1. Introduction

The reading disorder called dyslexia makes the human to feel hard to read the words or letters effortlessly and more difficult to spell words. Our brain has phonological processing that associates the letters with the sound followed by words, statements, and passages. In mid-2000, the literature proves that dyslexic has issues in phonological processing that produce poor reading and understanding knowledge [1-3]. There is a challenge to align the letters with sounds that cause the individual to have poor reading skills. This reading disorder makes the individual slow readers, but they have innovative thinking to attain powerful reasoning abilities. The poor word decoding and reading issues continue while they grow up to affect their school performance and life. In the early and mid-2010, the literature proves that dyslexia is a neurological disease. People with dyslexia have the left anterior area of the brain found that cannot proceed with the speech patterns using neuroimaging in the year 2012, which caused the auditory processing illness [4]. Functional Magnetic Resonance Imaging (fMRI), Positron emission tomography (PET), Magnetoencephalography (MEG), Electroencephalogram (EEG), and Magnetic Resonance Imaging (MRI) are the techniques of neuroimaging that found there are variations in the design of the brain in people with dyslexia. This reading disorder has a deficiency in the left region of the brain, which is accountable for reading [5]. This deficiency is generally discovered by performing written and oral tests [6]. These test scores give clinical judgments. The important issue in this technique is that the clinical judgments are made based on expert analysis and the clinical decisions are not dependent on the objective [7].

Dyslexics' eye focus is varied from the non-dyslexics that are proven in recent studies. When the dyslexic kids look at the light spots, they can make a lot of eye movements that happen consecutively [8-11]. This deficiency has been found by observing the differences in eye movements when reading [12]. When the visual search is performed, and reading text, the eye movements of French dyslexics are followed. There is an observation that this problem has more obsessions, contrasted in visual tasks and reading by normal readers [13]. When reading Chinese, the dyslexic kids have the movement patterns that are examined for obsessions and looking for the time more for people with dyslexia than for normal individuals. The landing pose of obsession varies in both groups. To predict dyslexia from tracking eye movements, the statistical model has been constructed [14]—the achievement in obtaining the 80.18

[1] Research Scholar, Department of Computer Science and Engineering, Annamalai University,Chidambaram-608002,Tamilnadu, India
[2] Assistant Professor, Department of Computer Science and Engineering, Annamalai University,Chidambaram-608002,Tamilnadu,India
[3] Professor, SRKR Engineering College,Bhimavaram-534204,India
** Corresponding Author Email: kdvpkvarma@gmail.com*

percentile accuracy using the binary classifier SVM (Support Vector Machine). The neutral networks in people with dyslexia are examined to find the view pattern and obtain the 78-percentile accuracy [15]. Compared to neural networks, high accuracy is obtained in SVM. However, there are various flaws with the existing approaches like higher error rate, computational complexity and reduced prediction accuracy. To handle all these issues, a novel Ad boosting classifier model is imposed with the bagging and boosting concept to improve the drawbacks of the feeble classifier model. Similarly, most existing works does not discuss both dyslexia and ADHD; however, this work intends to perform prediction accuracy for both the disorders. The primary enhancements to research include:

1. To obtain a dataset after the available online dataset. Here, the Kaggle dataset for both ADHD and dyslexia is considered for the prediction process.

2. The features are learned using the LASSO, relief model and fisher score model. Here, most influencing features are analyzed with all these features where the features are considered for further classification purposes.

3. Here, an AdaBoost classifier is cast-off for cataloguing determinations where the drawback of the model is boosted by the bagging and boosting concept.

4. Finally, using the MATLAB 2020a environment, the simulation is completed, and parameters like accuracy, precision, recall, F1-score, and error rate are computed and evaluated against a number of other current methods.

The work is divided into two sections: Section 2 offers a thorough analysis of the predicted model, and Section 3 offers a thorough explanation of the feature learning and classifier model. The predicted model's numerical results will be discussed with numerical results in section 4. Section 5 concludes with a summary of the research and a suggestion for further development.

2. Related Works

This section offers a complete examination of various existing approaches. Szucs et al. [16] state that the various wavelet transform is utilized in this scenario that uses the discrete wavelet transform. Few studies states that pre-processing is done manually, and feature extraction [17] is used when others use tools like Free Surfer, PANDA. The aim of pre-processing is to eliminate the null values and find suitable features. Subsequently pre-processing, the feature removal is done where suitable attributes are found and aligned with the scale of values. The values may be categorical or numerical. The range of attributes differs from various researches from 12 to 226 [18]. The group of dominant attributes is identified in the next step, which is essential to determine the object's class. Some researchers utilized manual selection to attain this [19]. A least absolute shrinkage, selection operator (LASSO), and SVM-RFE are used by the others. LASSO technique can be used to improve the interpretability and accuracy that performs the variable selection and regularization. These models are relevant for the regression models [20]. SVM-RFE chooses the attributes to consider the significance in portioning classes for SVM classifiers. The mentioned approach starts with the complete attributes groups and eliminates the number of attributes in the successive iterations. When the numbers of attributes are more, an essential task is to choose the suitable attributes due to computational complexity [21] – [23]. Moreover, the performance analysis of various attributes selection approaches is compared that are not depicted in the previous research.

Following the feature selection, the system training and categorization are carried out with the use of machine learning algorithms. The data set is divided into training and testing components. The data set for the current study was cross-validated ten times, however it is divided into ten similar sections.. Out of 10, 9 are required to train the algorithm when another set is required to test the performance [24] and other methods utilized the various split (for example, leave-one-out-cross-validation (LOOCV) [25], and fivefold. Supervised classification algorithms are utilized to test since the training data set has class information already, namely non-dyslexic or dyslexic. Naive Bayes, Support vector machine (SVM), Neural network, Logistic regression, Linear discriminant analysis (LDA) and K-Nearest Neighbor (K-NN) are considered as machine learning algorithms in the existing studies for classifying participants. Many studies used the common algorithm SVM. whenever the total amount of dimensions exceeds the number of specimens and there is a limited amount of attribute space, the SVM needs to give better performance even though the problematic is considered a binary organization issue, namely finding the non-dyslexic dyslexic users [26]. Moreover, the explanation of SVM is a complicated work. When the data set produces more noise, SVM does not execute well. Meanwhile, logistic regression is the easier method in understanding, implementing, and expecting to give better results for binary classification problems [27]. On the whole, a suitable classification approach is selected based on the data. Henceforth, the research gives the relative performance to indicate the different machine learning approaches results than announcing the chosen performance. The application of ensemble techniques is advantageous to attain good performances [28].

2.1 Reviews on performance measures

The tools utilized to assess the results in the current research are WEKA, Python, and MATLAB. The performance of the procedures used in this example to identify dyslexia using machine learning techniques is evaluated using a variety of measures. The ROC (Receiver Operating Characteristic) curve includes the sensitivity, accuracy, precision, specificity, mean square error (MSE), recall, negative predictive value (NPV), positive predictive value (PPV), and area. When specificity and sensitivity are employed to calculate the precise ratio of dyslexic and non-dyslexic recognized users, the number of categorized objects correlates to the total number of elements that relate to accuracy precisely[29]. A positive expected value or precision denotes the accurate fraction in identifying dyslexic users based on The overall amount of dyslexic people recognized during recall, which is the percentage of all dyslexic people properly identified. About 60% to 80% of accuracy is achieved in EEG-based techniques in various works [30] when 83.61% accuracy is achieved in the MRI scan-based method, and 80.24% accuracy is achieved in the game-based approach. Overall, it is really exciting to see how the techniques' change provides performance when data from various sources are merged altogether. Table 1 highlights the many specifics of the suggested methods to recognize dyslexia using machine learning techniques.

Table 1 Evaluation of numerous current methods.

Methods	**Age of User**	**Languages**	**Datatype**	**Test type**	**Technique of Machine learning**	**Performance Metric**
Machine Learning approaches	8-62	Hebrew	Text	Reading	Naïve Bayes and LDA	Perplexity, Accuracy
	12-54	Spanish	Eyetracking	Reading	SVM	Correctness
	9-10	Swedish	Eyetracking	Reading	SVM	Correctness
	10-14.7	Mandarin	MRI scans	Reading	Logistic regression, SVM	Sensitivity, Correctness, PPV, specificity, and NPV
	8.51-13.7	French, Polish, and German	MRI scans	Reading	Logistic regression, SVM, and random forest	The zone under the curve and accuracy
	7.1 (avg)	Malay	Text	Reading	KNN	Correctness
	7.1-60	English	Text	Online games	SVM	Precision, accuracy, recall
	Grades 6-7	Hebrew	EEG scans	Reading	Neural network, SVM	Confusion matrix
	>= 18	English	EEG scans	Typing and Writing	SVM	Sensitivity, accuracy, and specificity
	7-12	Malay	Image and Video	Reading	Naive Bayes, SVM, and K-means	Correctness
	8.5-12.5	Greek	Eyetracking	Reading	Naive Bayes, SVM, and Kmeans	MSE, Correctness
	Grade 3	Dutch	EEG scans	Reading	SVM and KNN	Sensitivity, accuracy, precision and specificity
	Grades k-6	English	Image	Handwriting	Convolutional neural network	Correctness

3. Methodology

Data acquisition, pre-processing including data normalization, feature learning, and classification model are the four initial stages of this study. In this case, the simulation is carried out in a MATLAB 2020b environment, and metrics like accuracy of predictions are assessed and contrasted with other current methods.

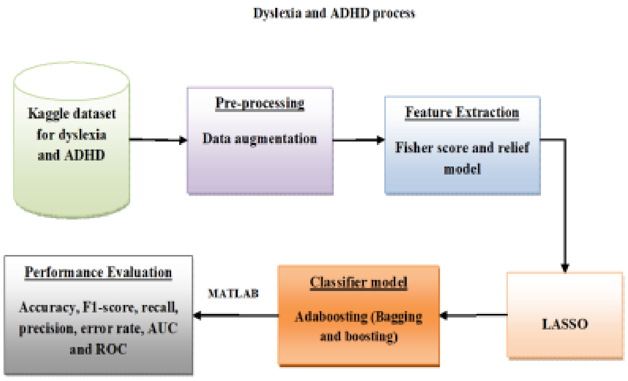

Fig 1 Block of the integrated fuzzy-SVM model

3.1 Data acquisitions:

In this study, two disorders—ADHD and dyslexia—are taken into consideration. For dyslexia, the dataset may be seen at https://www.kaggle.com/luzrello/dyslexia, and for ADHD, at https://www.kaggle.com/jerseyneo/reddit-adhd-dataset. Dyslexia is a unique learning condition that is linked to academic failure. The process of prediction is increasingly challenging and important. For 3600 participants, various techniques, such as online gamified assessments, are used; 80% of the participants are dyslexic, as predicted. Here, the samples' ages are tailored and the experiment was set in a different setting. The characteristics associated with this are listed in Table 1 below. While ADHD is a neurological issue that results in above-normal levels of behaviors, such as impulsive or hyperactive ones, analysis is taken into account for this condition. A person with ADHD has problems focusing on one task at a time or remains focused for a longer period of time. It has been reported that ADHD affects both children and adults.

3.2. Preprocessing with data extension

Increasing the amount of simulated samples and removing over-fitting problems are two benefits of the data augmentation approach known as image patching. Before an image is adjusted to a classifier model, image processing with augmentation is used to improve the image quality. It offers scaling, orientation, and color adjustments. Image augmentation involves modifying photos to produce different content iterations in order to expose the model to a wider variety of training samples. The training data is typically processed using image augmentation. The optimal pre-processing step is hence transformation, which is assessed as an augmentation. The photographs in the dataset that is being provided occasionally consist of low contrast pictures. Low contrast images are not ideal for use in the prediction process. As a result, images with contrast adjustments are widely sought for. It is less precise in situations where constant contrast adjustment is very appropriate when the training data provided lacks a constant contrast level representation. In fact, random visual contrast changes during training improve generalization. It's referred to as augmentation.

3.3 Fisher scores ReliefF model.

Fisher measure evaluates the gradation of data discernment of diverse classes. If the correlation between the class of dataset and feature is higher, then those features are considered for evaluation with higher quality and resolve be helpful for classification resolutions. If the quality of the characteristics set is less, features are superfluous. Fisher score for selecting features 'k' is provided in Eq. (1):

$$F_s(k) = \frac{\sum_{i=1}^{n} n_i\left(\mu_i^k - \mu^k\right)^2}{\sum_{i=1}^{n} n_i(\sigma_i^k)^2} \qquad (1)$$

3.4. Relief Model

The important aim of the relief algorithm is to determine the quality of the attributes by analyzing the capabilities of selected attributes randomly. This technique is given in the below algorithm. The chosen of each addressed attribute vector is the important property for nearby, similar class and different class vectors in this algorithm. Moreover, this step is managed for each chosen attribute, and these attributes will be located regarding the factor's positioning. Relief F-based Kernel Probability Estimation model is used as a theoretical technique to determine the corresponding functionalities from the obtained and feature vector that is trained. The SD's distance and the attribute set are determined where higher kernel probability estimation forms classified outcomes. This technique gives labeled data according to the corresponding features of the training set.

Algorithm 1 Relief model

Input: Training & Testing Feature, Label.

Output: Classified dyslexia and ADHD

1. Probability matrix array initialization;
2. For (*i*= 1 to size (*TS*))
3. For (j = 1 to size *(*TF*))*
4. Extract training set attributes;
5. Evaluate a distance of features and attributes;
6. Extract labels of training set;
7. Evaluate probability matrix array;
8. Compute SD of feature matrix and testing features;

9. Compute training feature probability;
10. Compute testing feature probability;
11. Feature verification probability;
12. Classified disease features;
13. End if;
14. End for;

End for;

The selected attribute is used for classification as input. The probability matrix is initialized in Eq. (2):

$$PM_{\pi} = P\ (q_1 = S_i) \tag{2}$$

Where, 's' is the training set state for i = 1, 2,... N, where 'N' denotes the number of elements of the training set., and 'q' is a state of testing sequence attribute. Features of the training set are obtained depending on the size of the testing attribute. The distance of training and testing attributes are determined in Eq. (3):

$$D_i = \sqrt{(s_i - V_i)^2 + \left(s_j - V_j\right)^2} \tag{3}$$

The distance value function is obtained from the associative labels. Based on the training and testing set labels, the probability array matrix is determined as shown in Eq. (14):

$$PM_{\pi} = \frac{\sum_{k=1}^{m} S_i\ (q_{i,j}\ (D_i))}{m} \tag{4}$$

The standard deviation of the feature matrix and the testing attributes are depicted below in Eq. (5): and Eq. (16):

$$\sigma_{Ts} = \sqrt{\frac{1}{m}\sum_{i=1}^{N}\left(S_i\left(q_{i,j}\ (d)\right) - \pi_i\right)^2} \tag{5}$$

$$\sigma_{TF} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left(V_i\left(q_{i,j}\ (d)\right) - \pi_i\right)^2} \tag{6}$$

Here, S_i length is denoted by 'm', and 'Vi 'length is denoted by 'n.' Based on the size of the training set (N), and size of the testing set (M), the probability of training or testing attribute is calculated as shown in Eq. (7): and Eq. (8):

$$P(TS) = (2 \prod \sigma_{Ts}^2)^\wedge \frac{-N}{2} * e^{\left\{\frac{-1}{2\sigma_{TS}}\right\}} \left|\left|TS - \pi_i\right|\right|^2 \tag{7}$$

$$P(TF) = (2 \prod \sigma_D^2)^\wedge \frac{-M}{2} * e^\wedge \left\{\frac{-1}{2\sigma_D}\right\} \left|\left|TF_i - \pi_i\right|\right|^2 \tag{8}$$

When the probability of the training set is greater than the probability of the testing set, the parameters are to classify the features depending on the attained testing attributes.

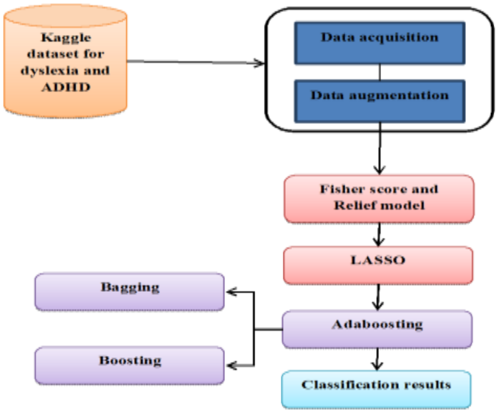

Fig 2 Flow diagram of the proposed model

3.5. Minimum Absolute Reduction and Collection Operator

LASSO has less assortment and shrinking functionalities based on the altering absolute value of the functions' coefficient. The attributes have few values of coefficient that are zero and attributes having negative coefficients are eliminated from the features' subset. This LASSO has a better performance for the values of the attribute with smaller coefficients. The large values of feature coefficient are there in the selected features' subsets. Unwanted attributes are identified in LASSO. However, the reliability of the attribute is improved by replicating the mentioned methods many times, ultimately considering the most often identified attributes as the essential one, which is known as the randomized LASSO feature. Since it utilizes parallel programming, this method has to be established in an effective computer. Fig 2 represents the movement of the anticipated model. It also illustrates the realization for the current application where q –I denote the vector of associated i^thsub-region keys.

3.6. Adaboosting

The disorders mentioned above, such as ADHD and dyslexia, are executed, and the task expects the effective AdaBoost technique. When the proposed approach is compared with the current neural networks or deep learning approaches, this technique anticipates that ADHD and dyslexia are efficient. In addition, the training time of data is modified depending on the requirements and minimizes computational complexity. The enhanced Adaboost algorithm can be classified into three phases. Here,

T_n=(〖(x〗 _1,y_1),(x_2,y_2), 〖…,x〗 _n,y_n) are the input training dataset where x_i is a feature set when y ∈{1,2,3} are the weak current classifier that failed in deployment.The existing approach has encountered the defects such as ID3 and CART being given to recognize the ability of Adaboost. The CART algorithm uses the recursive segmentation approaches. The discrete attributes are used in ID3 and planned to choose the features that have high values. Contrarily, the ID3 algorithm computes entropy, and CART calculates the GINI coefficient for each sample. A smaller coefficient of GINI with relevant division is required. CART algorithm usually divides the latest illustration usual into two sub-sample. Henceforth a non-leaf node at each of the decision tree (DT) possesses two types. CART produces the DT that has an easier structure and achieves higher precision.

1. The rules are generated that can be inferred.

2. Computation is less.

3. This model acquires capabilities to contract with discrete and continuous variables.

4. This model expresses the importance of definite features.

Primarily, the weight of training data should be initialized in Eq. (9) and Eq. (10):

$$T_1 = (w_{l1}, \dots w_{li}, \dots w_{1N}) \tag{9}$$

$$w_{Li} = \frac{1}{N}, i = 1, 2, \dots, N \tag{10}$$

Secondly, for m^th iteration, m=1,2,…,M. The weight distribution D_mhas the training dataset to achieve the fundamental classifier as mentioned in Eq. (11):

$$G_m(x): \gamma \rightarrow | \{1, 2, 3\} \tag{11}$$

Here, γ is data for the training set. Calculate the rate of error G_m (x) on actual results of classification with training data, and w_mi denotes the weight of i^th sample in m^th iteration as shown in Eq. (12) and Eq. (13):

$$e_m = \sum_{i=1}^{N} W_{mi} I(G_m x_i \neq y_i) \tag{12}$$

$$\sum_{i=1}^{N} W_{mi} \equiv 1 \tag{13}$$

Each step has the normalization, and the denominator is not divided by sample weight. Contrarily, the conventional AdaBoost algorithm has the classification error rate (CER) that hase_m ≤ 1/2. However, it is complicated to determine the CER, e_m ≤ 1/2 validate the features at each time. Considering the AdaBoost error rate as 1/2, error threshold limit as 2/3, and positive item X ensures a_m ≥0 when e_m≤ 2/3. Otherwise, am<0, sample weights need to be revised in the consecutive procedure. Then, determine the coefficient of classifier G_m (x) according to the rate of error e_m as shown in Eq. (14):

$$\alpha_m = \log \frac{1 - e_m}{e_m} + \log_2 \tag{14}$$

The weight distributions of the training dataset are revised according to co-efficient a_mas shown in Eq. (15) & Eq. (16):

$$D_{m+1} = (w_{m+1_{,1}}, \dots, w_{m+1_{,i}}, \dots w_{m+1^{,N}}) \tag{15}$$

$$w_{m+1_{,i}} = \frac{w_{mi}}{Z_m} \exp\left(-a_m y_i G_m(x_i)\right) \tag{16}$$

Where Z_m as normalization factor produces 〖D_m〗 _(+1) that becomes PDF as shown in Eq. (17):

$$Z_m = \sum_{i=1}^{N} W_{mi} \exp(-a_m y_i G_m(x_i)) \tag{17}$$

Lastly, the weighted sample of misclassification by classifier G_m (x) is maximized after continuously training when classified weight samples are minimized. Hence, misclassified sample tasks in signified role in consecutive iteration. The design of linear combination of the classifier is achieved and the final classification as shown in Eq. (18) & Eq. (19):

$$f(x) = \sum_{m=1}^{M} a_m G_m(x) \tag{18}$$

$$G(x) = sign\big(f(x)\big) = sign\left(\sum_{m=1}^{M} a_m G_m(x)\right) \tag{19}$$

Trained Weak classifiers are combined with a strong classifier to achieve a risk prediction model. Linear combination of f(x) implements the weight-based 'M' classifiers. f(x) value denotes the instance categories 'x' and denotes the classification confidence as function sign gives three-segmented classifier results.

3.7. Bagging

Bagging has the aim to minimize the Decision Tree classifiers variance. Creating various subsets of data from the training set of samples [68] is the objective of this bagging technique. The group of subset data is selected randomly, which is required for training the decision tree. The ensemble of the various modes is obtained as the outcome. All the predictions from various trees are used to take the average is used later. This approach is very powerful compared to a single Decision Tree classifier. This technique requires minimizing the over-fitting issue and also managing the greater dimensionality data appropriately. This model evolves the problem of missing data, and the accuracy is maintained.

3.8. Boosting Techniques

Boosting is a recursive procedure based on the prediction of weight and alternates the weight. The weight is maximized when the instance is classified inaccurately. Generally, this technique establishes better predictive models. Boosting creates the various loss functions and tasks by joining the weak models to optimize the performances. The Boosting technique is applied to classify Ada boosting algorithms in the research in building the hybrid models.

Algorithm 2: Bagging model

1. BEGIN
2. Let $D = \{d_1, d_2, d_3, \dots \dots d_n\}$ be the given dataset;
3. $E = \{\quad\}$, the set of ensemble classifiers;
4. $C = \{c_1, c_2, c_3, \dots \dots c_n\}$, the set of classifiers;
5. $X =$ the training set, X D;
6. $L = n(D)$;
7. **For** $i = 1\ toL$**do**
8. $S(i) = \{BootstrapsampleIwithreplacement\}IX$;
9. $M_{(i)}$ = Model trained using $C_{(i)}$ on $S_{(i)}$;
10. $E = EC_{(i)}$;
11. Next I;
12. **for**$i = 1\ toL$;
13. $R_{(i)} = Y$classified by $E_{(i)}$;
14. next i;
15. Result = max $\left(R_{(i)}: i = 1,2, \dots \dots \dots, n\right)$;
16. **END;**

4. Mathematical outcomes and conversation

MATLAB 2020b's simulation environment is used to simulate the suggested ensemble model. Accuracy, Recall, Precise, and F-measure are used to validate the results. In the end, metrics are produced and evaluation is done using k-fold CV in Eq. (20) - Eq. (23):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

$$Recall = \frac{TP}{TP + FN} \tag{22}$$

$$F - measure = \frac{2\,(TP)}{2\,((TP) + (FN))} \tag{23}$$

Here, TP is True positive; TN- True Negative; FP- False Positive; FN – False Negative. To investigate the model's performance, AUC and Receiver Operating Curve (ROC) scalar curve value are employed. AUC is employed to assess how well the model can distinguish between class values. Here, Relief-based feature selection and the Fisher Score model are analysed to discuss the prevalent techniques. The simulation takes place in MATLAB circumstances. The relief related Kernel Probability Estimation relates to the classification process that uses the feature set's kernel functionality to determine the matching point from obtained feature vector in which variance, mean, and standard deviation of attributes are determined. After this, attribute extraction of the matching point is examined. The probability result of the maximum matching function is required to create the classification result. The labeled data is extracted from the training attributes determined and upgraded in the library based on the probabilities of matching points.

Table 1 Presentation metrics contrast

Metrics	**Correctness**	**Precision**	**Recall**	**F1 score**	**AUC**	**ROC**	**Error rate**
Fisher+ ReliefF and boosting model (Dyslexia)	95.3	92.2	97.5	96	0.98	0.92	0.065
Fisher+ ReliefF and boosting model (ADHD)	95.6	94.97	97.6	96.56	0.98	0.926	0.54
Ensemble model (Dyslexia)	92.21	91.98	96.61	94.21	0.986	0.81	0.078
Ensemble model (ADHD)	95	93.80	98	94.65	0.99	0.92	0.076
Linear	78.2	79.21	77.	89.	0.8	0.8	0.0

SVM			8	15	6	45	99
Hybrid SVM-PSO	72.2	61	76	68	0.85	0.83	1.01
Standard RF	84.7	83.5	77.5	71	0.75	0.81	1.57
Standard NB	83.61	81.1	76	78.9	0.85	0.79	1.62

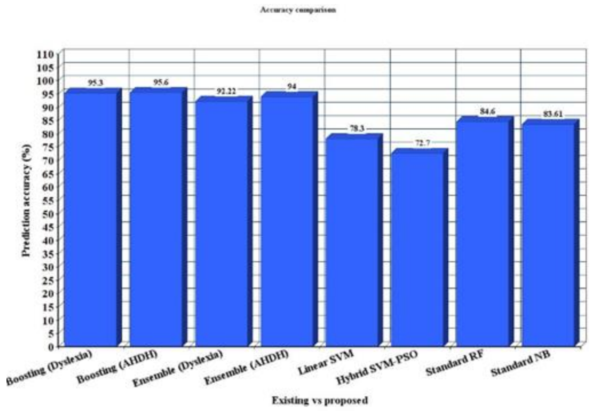

Fig 3 Accuracy contrast

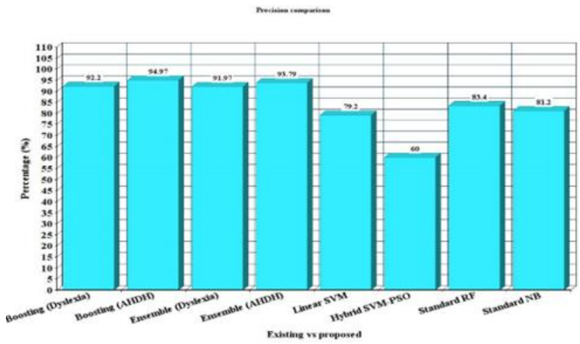

Fig 4 Precision assessment

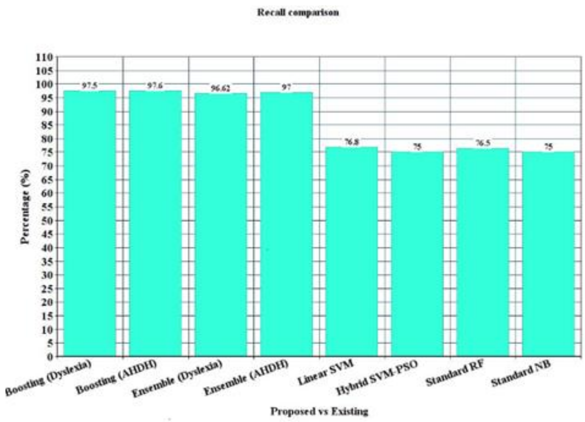

Fig 5 Recall assessment

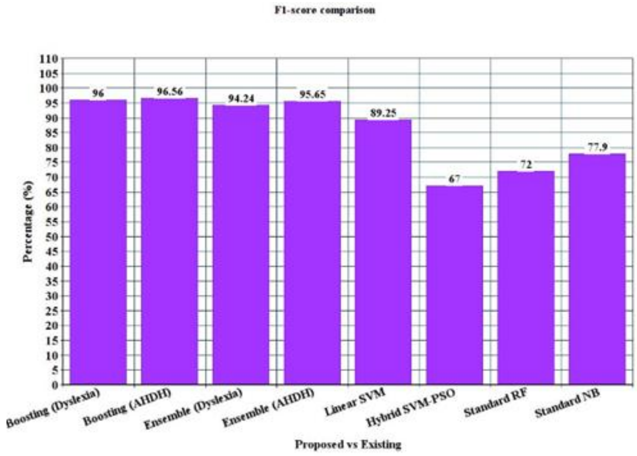

Fig 6 F1-score **assessment**

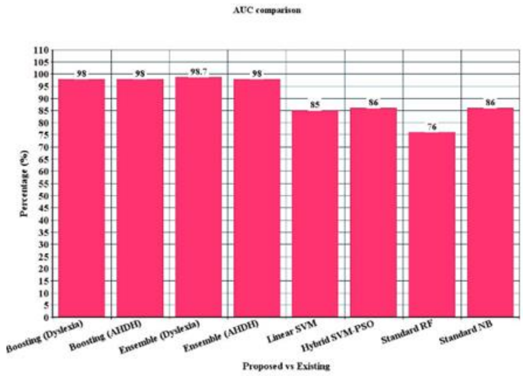

Fig7AUC assessment

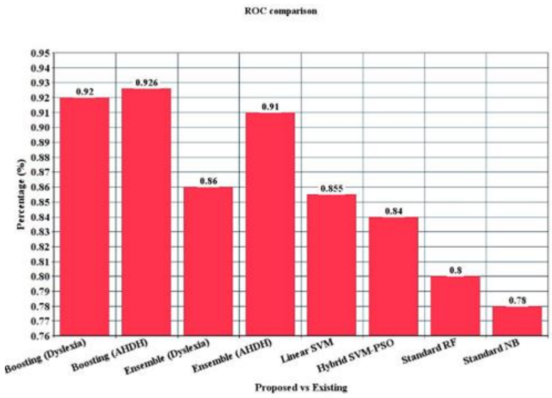

Fig 8 ROC assessment

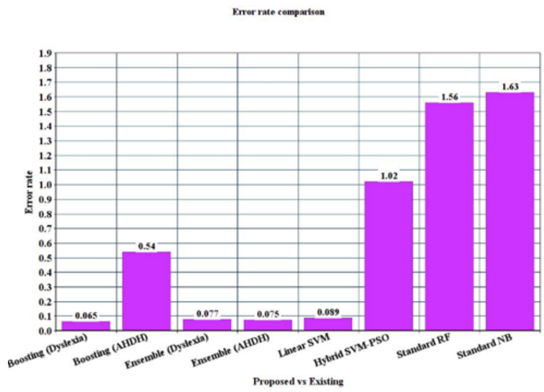

Fig 9 Error rate assessment

Table 1 contrasts the suggested ensemble approach with the current methods (for the prediction of dyslexia) based on a variety of performance measures. The ensemble model is compared to the widely used linear SVM, hybrid SVM-PSO, standard RF, and standard NB classifier models The accuracy of the predicted model is 95.3%, which is greater than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, standard RF, and NB classifier model in 3.08%, 1.3%, 17%, 22.6%, 10.7%, and 11.69% (See Fig 3). The expected model's precision is 92.2%, which is greater than the ensemble model (which accounts for dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier models (see Fig. 4). The expected model's recall is 97.5%, which is 0.8%, 0.5%, 20.7%, 22.5%, 21%, and 22.5% higher than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier models (See Fig. 5). The expected model's F1-score is 96%, which is greater than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, standard RF and NB classifier model, by 1.76%, 0.35%, 6.75%, 29%, 24%, and 18.1%. (See Fig 6). The ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier model all have AUCs of 98%, which is 0.007%, 0.007%, 13.7%, 12.7%, 22.7%, and 12.7% lower than the predicted model (see Fig 7). The ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier model all have ROCs lower than 92%, which is 6%, 1%, 6.5%, 8%, 12%, and 14% higher than the predicted model (see Fig 8). The expected error rate is 0.065, which is considerably lower than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, standard RF and NB classifier model (See Fig 8), and is 0.475, 0.012, 0.01, 0.024, 0.955, 1.495, and 1.565. Similar to this, Table 1 contrasts the proposed ensemble model's performance metrics with those of the current techniques (for the prediction of ADHD). The expected model's accuracy is 95.6%, which is greater than the ensemble model (which accounts for dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier models by 3.38%, 1.6%, 17.3%, 22.9%, 11%, and 11.99%. The expected model's precision is 94.97%, which is greater than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, standard RF, and NB classifier model by 3%, 3.18%, 15.77%, 34.97%, 11.83%, and 13.77%. The recall of the expected model is 97.6%, which is greater than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, standard RF, and NB classifier model by 0.98%, 0.6%, 20.8%, 22.6%, 21.1%, and 22.6%. The expected model's F1-score is 96.56%, which is greater than the ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier model by 2.32%, 0.91%, 7.31%, 29.56%, 24.56%, and 18.66%. The expected model's AUC is 98%, which is identical to the ensemble model (dyslexia and ADHD), and 13%, 12%, 22%, and 13% higher than the linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier models, respectively. The ensemble model (dyslexia and ADHD), linear SVM, hybrid SVM-PSO, conventional RF, and NB classifier model have ROCs that are 6%, 1%, 6.5%, 8%, 12%, and 14% lower than the expected model, which has a ROC of 926%. The expected error rate is 0.054, which is lower than the ensemble model's (dyslexia and ADHD), linear SVM's, hybrid SVM-PSO's, conventional RF and NB classifier model's, and their respective error rates of 0.023, 0.021, 0.035, 1.0396, 0.966, 1.506, and 1.576.Based on the investigation, it is understood that the predicted model outperforms the numerous current approaches.

5. Conclusion

The problems of ADHD and dyslexic are identified reasonably with high correctness that creates severe impact possible on the long-term disease influence of a human without concerning the cultural and social background. The earlier diagnosis of the disease is the important stage to achieve the goal. Many kinds of research are tried in the forecast of ADHD and dyslexia already by means of machine learning techniques. The research prefers the same route and an enhanced theoretical approach with a large dataset to train the model. The study establishes the algorithm of relief feature selection that can give a strongly associated attribute set is utilized. It is utilized along with different machine learning approaches. This research determines that the relief feature model works well, especially with highly influenced features that are achieved using the feature selection algorithm and generate accuracy significantly greater than the associated works. About 95.3% and 95.6% of accuracy are achieved for dyslexia and ADHD predication. A goal in the future is to globalize the technique more and more. This model can be used with other algorithms for feature selection and is strong in opposition to the dataset, where the missing data level is huge. Another future technology is the application of deep learning approaches. The first important goal of this study is to enhance the existing system and a new and theoretical approach to building the model and establish a useful model that is implemented easily and practically.

References

[1] Altoè, G., Bertoldo, G., ZandonellaCallegher, C., Toffalini, E., Calcagnì, A., Finos, L., &Pastore, M. (2020) Enhancing statistical inference in psychological research via prospective and retrospective design analysis. Frontiers in Psychology, 10:2893.

[2] Bates, D., Maechler, M., Bolker, B., & Walker, S.

(2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1-48

[3] Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E., &Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14, 365-376.

[4] Cirino, P. T., Rashid, F. L., Sevcik, R. A., Lovett, M. W., Frijters, J. C., Wolf, M., & Morris, R. D. (2002). Psychometric stability of nationally normed and experimental decoding and related measures in children with reading disability. Journal of Learning Disabilities, 35(6), 526-539.

[5] Dai, L., Zhang, C., & Liu, X. (2016). A special Chinese reading acceleration training paradigm: To enhance the reading fluency and comprehension of Chinese children with reading disabilities. Frontiers in Psychology, 7:1937.

[6] Ebrahimi, L., Pouretemad, H., Khatibi, A., & Stein, J. (2019). Magnocellular based visual motion training improves reading in Persian. Scientific Reports, 9:1142

[7] Giofrè, D., Cumming, G., Fresc, L., Boedker, I., &Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. PLOS ONE, 12(4), e0175583.

[8] Franceschini, S., Bertoni, S., Gianesini, T., Gori, S., &Facoetti, A. (2017). A different vision of dyslexia: Local precedence on global perception. Scientific Reports, 7:17462.

[9] Horowitz-Kraus, T., Vannest, J. J., Kadis, D., Cicchino, N., Wang, Y. Y., & Holland, S. K. (2014). Reading acceleration training changes brain circuitry in children with reading difficulties. Brain and Behavior, 4(6), 886-902.

[10] Koen, B. J., Hawkins, J., Zhu, X., Jansen, B., Fan, W., & Johnson, S. (2018). The location and effects of visual hemisphere-specific stimulation on reading fluency in children with the characteristics of dyslexia. Journal of Learning Disabilities, 51(4), 399-415.

[11] Lofti, S., Rostami, R., Shokoohi-Yekta, M., Ward, R. T., MotamedYeganeh, N., Mathew, A. S., & Lee, H. J. (2020). Effects of computerized cognitive training for children with dyslexia: An ERP study. Journal of Neurolinguistics, 55:100904

[12] *Luniewska, M., Chyl, K., Debska, A., Kacprzak, A., &Plewko, J. (2018). Neither action nor phonological video games make dyslexic children read better. Scientific reports, 8:549.

[13] Meng, X., Lin, O., Wang, F., Jiang, Y., & Song, Y. (2014). Reading performance is enhanced by visual texture discrimination training in Chinese-speaking children with developmental dyslexia. PlosONE, 9(9):e108274. H

[14] Wang, L. C., Liu, D., &Xu, Z. (2019). Distinct effects of visual and auditory temporal processing training on reading and readingrelated abilities in Chinese children with dyslexia. Annals of Dyslexia, 69, 166-185.

[15] Wang, L. C. (2017). Effects of phonological training on the reading and reading-related abilities of Hong Kong children with dyslexia. Frontiers in Psychology, 8:1904.

[16] Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature.Plos Biology, 15(3), e2000797.

[17] Andersson, U. &Lyxell, B. Working memory defcit in children with mathematical difculties: A general or specifcdefcit? Journal of Experimental Child Psychology 96, 197–228 (2007).

[18] Staikova, E., Gomes, H., Tartter, V., McCabe, A. &Halperin, J. Pragmatic defcits and social impairment in children with ADHD. Journal of Child Psychology and Psychiatry 54, 1275–1283 (2013).

[19] Helland, W., Posserud, M., Helland, T., Heimann, M. &Lundervold, A. Lanugage Impairments in Children with AD(H)D and in Children with Reading Disorder. Journal of Attention Disorders 20, 581–589 (2016).

[20] Melby-Lervåg, M. & Hulme, C. Is working memory training efective? A meta-analytic review.Developmental Psychology 49, 270–291 (2013).

[21] Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Prim. 2015;1:15020.

[22] Chang Z, Lichtenstein P, Asherson PJ, Larsson H. Developmental twin study of attention problems high heritabilities throughout development. JAMA Psychiatry. 2013;70:311–8.

[23] Pettersson E, Larsson H, Lichtenstein P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. Mol Psychiatry. 2016;21: 717–21

[24] Rydell M, Taylor MJ, Larsson H Genetic and environmental contributions to the association

between ADHD and affective problems in early childhood-A Swedish population-based twin study. Am J Med Genet B Neuropsychiatr Genet. 2017;174: 538–46.

[25] Faraone SV, Ghirardi L, Kuja-Halkola R, Lichtenstein P, Larsson H. The familial co-aggregation of attention-deficit/hyperactivity disorder and intellectual disability: a register-based family study. J Am Acad Child Adolesc Psychiatry. 2017;56:167–74 e1.

[26] Brikell I, Ghirardi L, D'Onofrio BM, Dunn DW, Almqvist C, Dalsgaard S, et al. Familial liability to epilepsy and attentiondeficit/hyperactivity disorder: a nationwide cohort study. Biol Psychiatry. 2018;83:173–80

[27] Zayats T, Athanasiu L, Sonderby I, Djurovic S, Westlye LT, Tamnes CK, et al. Genome-wide analysis of attention deficit hyperactivity disorder in Norway. PLoS ONE. 2015;10: e0122501.

[28] Mooney MA, McWeeney SK, Faraone SV, Hinney A, Hebebrand J, Consortium I, et al. Pathway analysis in attention deficit hyperactivity disorder: an ensemble approach. Am J Med Genet B Neuropsychiatr Genet. 2016;171:815–26.

[29] Akutagava-Martins GC, Salatino-Oliveira A, Genro JP, Contini V, Polanczyk G, Zeni C, et al. Glutamatergic copy number variants and their role in attention-deficit/hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. 2014;165:502–9

[30] Carrey NJ, MacMaster FP, Gaudet L, Schmidt MH. Striatal creatine and glutamate/ glutamine in attention-deficit/ hyperactivity disorder.J Child AdolescPsychopharmacol.2007; 17:11–7.