Robust Ensemble Feature Selection Technique with Explainable AI for Dyslexia Prediction Using Eye Movements

Resma Madhu P.K[*,1], Karishma Kaine T [2], Priyanka Panda [2]*
[1]Department of Electronics and Communication Engineering,
Amrita School of Engineering – Coimbatore, India

[2]Department of Electronics and Communication Engineering,
PSG Institute of Technology and Applied Research, India
E-mail:pk_resmamadhu@cb.amrita.edu
** - Corresponding author*

Abstract:
Early and easy detection of dyslexia is a crucial prerequisite for effective treatment. Machine Learning (ML) based classifiers are widely employed to detect dyslexia using eye movement recorded during reading activity. Abundance of available features, diversity of feature selection techniques, and variety of ML classifiers necessitate extensive investigation to determine the optimal dyslexia prediction model. Hence, there is a need for optimal feature selection methods to select the best features, which has been addressed in this work. The proposed work aims to develop a robust ensemble feature selection technique which can provide a holistic view of various feature ordering methods to bring the best out of them. Features extracted using various methods like statistical, dispersion-based and velocity-based techniques are considered holistically. Explainable AI (XAI) principle can provide insight on features efficiency in ML based dyslexia prediction. SHapley Additive exPlanations (SHAP) principle of XAI is combined with traditional feature selection methods like principal component analysis (PCA) and random forest (RF) to order the features. Among the four different rank aggregation methods, Robust Rank Aggregation (RRA) can aggregate the ordered features to improve the classification accuracy. Performance improvement with the inclusion of features is analyzed using accumulation effect. Experimental results illustrate that ~58% of feature reduction is observed and stacking classifier provide a reliable classification accuracy of 96.4%.

Keywords: Dyslexia Prediction, Machine Learning Classifier, Explainable AI, Ensemble Feature Selection, Robust Rank Aggregation, Accumulation Effect

1. **Introduction**

Globally, ~15% of the population is affected by dyslexia, making it is one of the most common neurodevelopmental disorders (Wagner et al., 2020). About 10-12% of school students are affected with dyslexia, which significantly impacted their studies (Yang et al., 2022). Dyslexia primarily affect reading and writing abilities of individuals (Ziegler et al., 2020). Students with dyslexia have challenges in decoding the words and retaining information during reading tasks (Mukhtar et al., 2024). These challenges make them deficits in phonological processing and cognitive information handling. Early detection of dyslexia is one of the requisites for specialized intervention to improve their academic and emotional outcomes. Proactive approaches such as individualized reading programs and adaptive teaching strategies can enhance academic performance. Hence, there is a need for an efficient and easy-of-use dyslexia prediction system to screen for dyslexia in larger groups of individuals.

Individuals with dyslexia have difficulty in language processing capability, which makes their eye movement disrupted during reading (Kirkby et al., 2008). This signature eye movement pattern can be a marker to detect dyslexia. Eye movements are fundamentally classified as saccades, fixations, smooth pursuit, vergence, and vestibulo-ocular movements. Among these, saccades and fixations are most relevant to reading. Saccades are characteristics by rapid jumps of the eye focus, enabling the gaze to shift across text during reading. Fixations are the pauses in the eye movement to process visual information. Dyslexic readers exhibit typical eye movement patterns, such as increased fixation duration, higher fixation count, and irregular saccades including regressions and shorter amplitudes. Hence, a ML technique with visual processing capability can efficiently predict dyslexia from the eye movement pattern, which has been investigated in this work.

The rest of the paper is organized into sections as follows; Section 2 describes the work related to dyslexia prediction using ML and its associated challenges. The methodology of the proposed rank aggregation method for optimal feature selection is elaborated in Section 3. Section 4 presents the experimental results of the proposed work. And Section 5 concludes the work with the future direction of research.

2. Related Work

Eye-tracking techniques augmented with recent development in ML emerge as an effective non-invasive tool for screening dyslexia, especially in school children.(Coenen et al., 2024).

ML based detection of dyslexia is broadly classified into two types: feature-based and feature-free.

In feature-based methods, statistical and domain-specific features are extracted from eye-tracking data. It includes fixation duration, saccade length, regression count, and gaze path patterns, which provides inferences on dyslexia. The extracted features are used by the ML classifiers to detect dyslexia accurately. ML techniques such as Support Vector Machines (SVM), Random Forest (RF), and XGBoost are widely used in dyslexia prediction. These ML classifiers provide interpretable outputs and transparency in inferencing. (Appadurai & Bhargavi, 2021). It also demands less computational power and can perform effectively with smaller datasets. However, their performance greatly relies on the quality and selection of features, which demands extensive domain knowledge and ML expertise.

On the other hand, feature-free methods such as Convolutional Neural Networks (CNNs) (Svaricek et al., 2025), Recurrent Neural Networks (RNNs), and Transformers eliminate the need for feature extraction and generate inference directly from the eye-tracking data like gaze maps or scan paths (Haller et al., 2022)(Alqahtani et al., 2023) Hybridization of multiple deep learning models are also proposed to improve the prediction performance. (Dewanjee & Muntaha, 2024)Large language models (LLM) powered active inferencing approach is employed to simulate the reading pattern of normal and dyslexia person(Donnarumma et al., 2023).This eliminates the need for expert knowledge and enables us to observe the complex nonlinear relationship between eye movement and dyslexia. Lack of interpretability, demand for large, labeled datasets, and higher computational resource requirement are some of the downfalls of this method.

Automated feature selection is one of the promising alternative approaches to eliminate the need for extensive investigation and domain knowledge to design a feature-based ML classifier. It also leverages the low computational requirement of these classifiers and improves its classification performance (Rello & Ballesteros, 2015). Several techniques have been proposed in literature to order and select the features for improving dyslexia prediction. Dimensionality reduction techniques like Linear Discriminative Analysis (LDA) and PCA are used widely to select features based on their statistical characteristics (Jothi Prabha & Bhargavi, 2022) Filter-based approaches using correlation coefficients and mutual information, are used to rank features based on their statistical relevance (Vajs et al., 2022) Wrapping techniques such as Recursive Feature Elimination (RFE) are used to eliminate irrelevant features.(Prabha et al., 2021) Embedding classifiers like RF and XGBoost can provide interpretable results and can be used to extract importance, which is used to rank the features (Liyakathunisa et al., 2023). These feature selection techniques utilize various modalities to identify features that demonstrate strong characteristics in one or more specific aspects (Li et al., 2017). Hence, there is a need for a fusion technique that integrates multiple feature selection methods to harness the strengths of each and achieve more robust performance.

Motivated by these challenges, the proposed work aims to develop an ensemble feature selection technique, which can include features ordered with various techniques. It enables the proposed work to select optimal features after aggregation and to design efficient ML classifier with reliable classification accuracy.

Major contribution of this work is listed as follows,

1. Feature ordering using PCA, RF-FI, and XAI-SHAP
2. Development of rank aggregation method to fuse the ordered features
3. Use of accumulation effect to analyze the improvement in performance with inclusion of ordered features
4. Selection of optimal features and best ML classifier to detect dyslexia accurately.

3. Methodology

The block diagram of the proposed work on dyslexia prediction model using eye-tracking data using robust feature selection technique with XAI is illustrated in Figure 1. Like other ML based classification problems, the proposed work also involves data acquisition, feature extraction and prediction. However, Ensemble feature selection technique is proposed to determine the features that need to be extracted to optimally predict dyslexia. To achieve this objective, the proposed work mainly comprises of two phases: the design phase and the deployment phase. The design phase is a one-time action executed to determine the optimal features and the best ML classifier for prediction. The deployment phase uses these features and ML classifier to detect dyslexia, and it is executed for every individual during screening process.

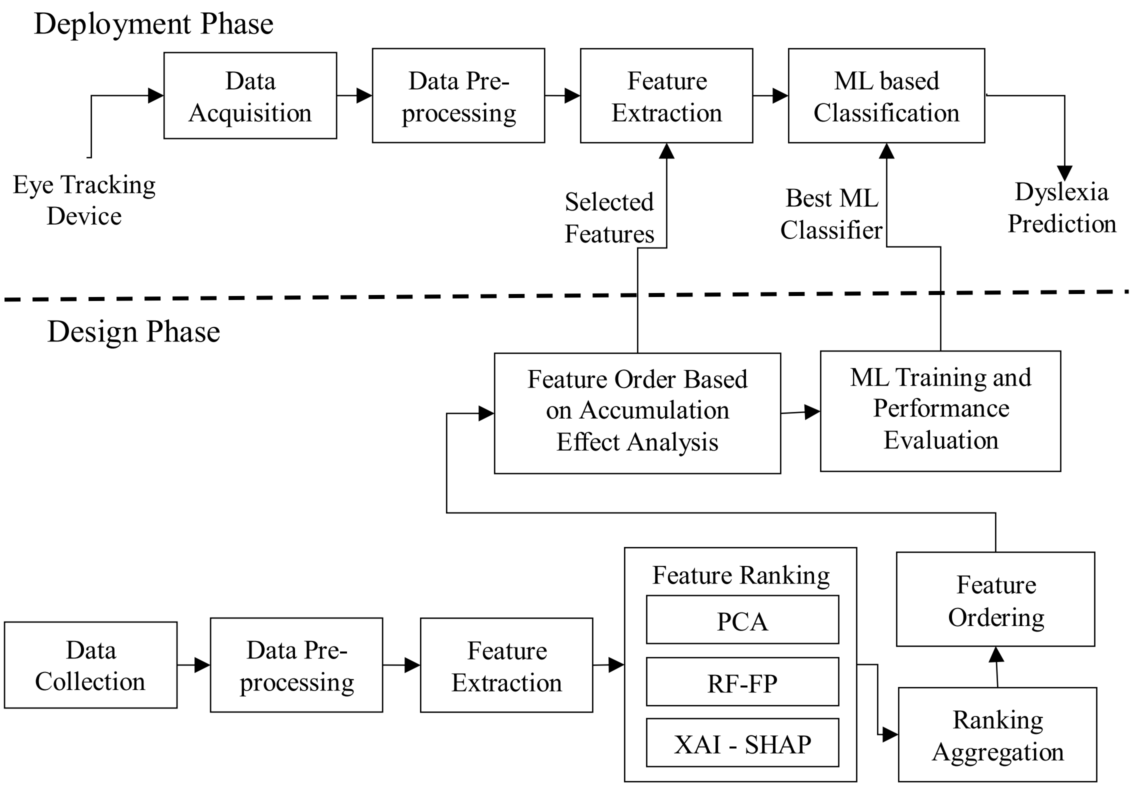


Figure 1. Proposed robust ensemble feature selection technique with XAI

In the design phase, raw eye-tracking data is first collected and pre-processed to remove noise and to handle inconsistencies. Various features describing cognitive or visual behaviors are extracted to provide inference on dyslexia. Subsequently, multiple feature ranking techniques, such as principal component analysis (PCA), Random Forest based feature importance (RF-FP), and explainable AI using SHAP (SHapley Additive exPlanations) are applied to evaluate the relevance of each extracted feature.

The identified three feature ordering techniques (PCA, RF-FI, and XAI-SHAP) have different principles of operations, which can be effectively fused to get the best out of these methods as in Figure 2. PCA is an unsupervised technique that enables to capture the feature variances effectively and order based on unique contribution of the features. RF-FP, leveraging ensemble predictions capability of RF. It can effectively detect feature importance and handles non-linear relationships. Meanwhile, XAI-SHAP provides model-agnostic interpretability by highlighting feature interactions and its contribution to model performances. The fusion of these methods can provide a comprehensive ranking mechanism that reduces biases, ensures a balance between model performance and transparency, and enhances the robustness of the dyslexia prediction model.

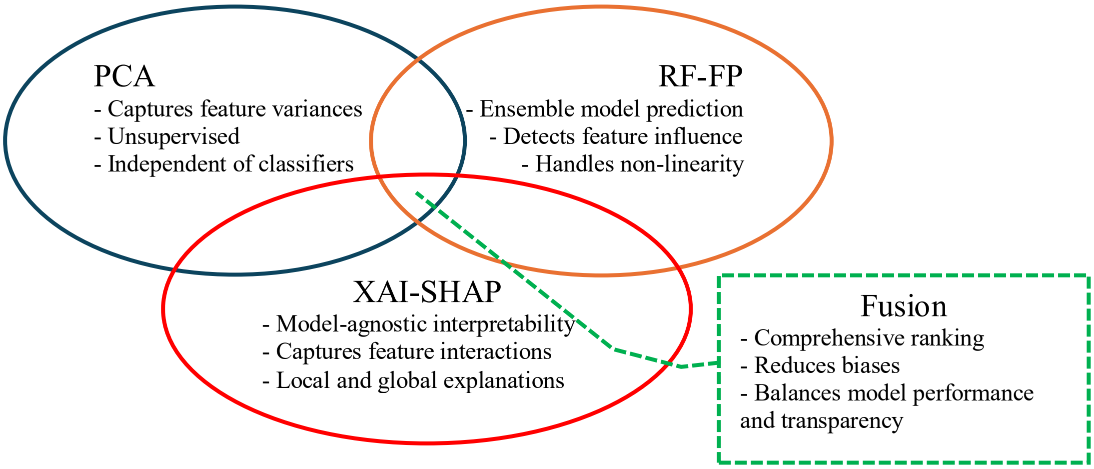


Figure 2. Characteristics of the proposed ensemble feature selection

The resulting ranked lists are merged using various ranking aggregation strategies to produce a robust and consensus-based feature ordering. The resultant feature order is used to analyze the impact of accumulated features on the classification performance. The features are added until further accumulation fails to provide significant improvement in classification performance. These selected features are used to provide inferences to extract only those features for dyslexia prediction during deployment. Various ML classifiers are investigated with the selected features, and the best ML classifier is used to detect dyslexia of the person. During deployment phase, real-time eye-tracking data is acquired via an eye tracking sensor by subjecting the person to perform a reading activity. The acquired data is pre-processed, and the selected features are only extracted to classify the dyslexia using the trained ML-based classifier.

3.1. Data Set Description

Eye movements were recorded using the Ober-2, an infrared-based eye-tracking device, while the person under test performs instructed activities. The eye movements are recorded in two dimensions namely X and Y axis with a sample time in milliseconds for both eyes. This raw eye-tracking data considered in this study is obtained from (Nilsson Benfatto et al., 2016). It consists of recordings of 185 children aged between 9 and 10 years, of which 97 were diagnosed with high-risk dyslexia (HR-D) and 88 are with low-risk dyslexia (LR-D). The collected dataset is segregated into training and testing dataset in the ratio of 70:30 enabling it to train and evaluate the ML based classifier.

3.2. Feature Extraction

The features described in (Prabha & Bhargavi, 2020) have been extracted from the acquired eye tracking dataset. Fixations and saccades are key components of eye movement that play a crucial role in reading, which are significantly affected by dyslexia. Hence, features related to fixations and saccades are extracted using three different methods namely, statistical, dispersion-based and velocity-based features as described in Table 1. Statistical measurements such as means, standard deviations, and range of eye movements for both eyes are extracted. Dispersion of eye movement during fixation are extracted using dispersion-based algorithms. Temporal features such as speed and frequency of saccades and fixation are extracted using velocity-based techniques. Hence, statistical features are more and less complex to extract, whereas dispersion and velocity features need customized algorithm to extract the features from the acquired eye movement dataset. Thus, the proposed technique aims to select the optimal features from these features set to provide a reliable dyslexia prediction.

Table 1. Description of features and its types

Type of Features	No. of Features
Statistical based Features	30
Dispersion based Features	16
Velocity based Features	11
Total	**57**

3.3. Feature Ordering Methods

3.3.1. PCA based Feature Ordering

PCA is a widely used dimensionality reduction technique. It transforms the original feature space into a set of orthogonal components, known as principal components (PC) that capture the variance available among the feature set (Jayabal et al., 2018). PCA transformation is based on the Eigen values and Eigen vectors of the feature covariance matrix. The Eigen values provide inferences on the strength of the PC to capture the data variances as illustrated in Figure 3. It provides inferences on the number of PCs required to maximize the data inclusion (Ahmad et al., 2022). In the proposed work, PCs are determined for the eye tracking feature sets, and it is observed that about 15 PCs can be included to accommodate a higher volume of data. Thus, PCA effectively reduces the dataset dimensionality while retaining the most critical information. Based on the explainable variance of the selected PCs, PCA can be applied to order the features based on their contribution to the most informative components. Thus, PCA-based feature ordering provides an efficient method to identify dominant features in high-dimensional dyslexia feature sets.

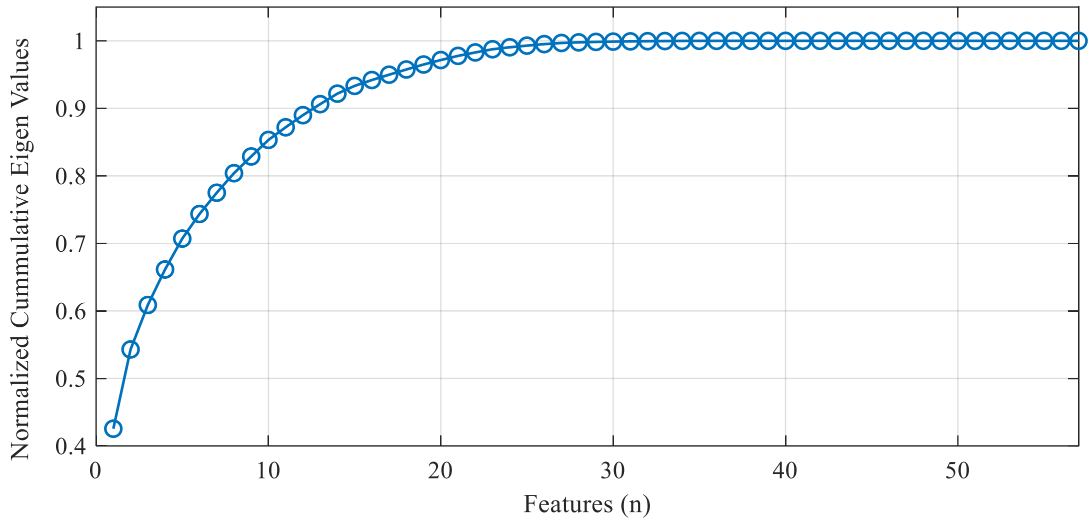


Figure 3. Variations in PC strength (Eigen values) across features.

3.3.2. RF-FI based Feature Ordering

RF is a robust ensemble ML classifier that provides a reliable predictive performance and provides an inherent mechanism for estimating FI. It operates by aggregating the results of multiple decision trees, each tress is trained on different subsets of the data, to form a collective consensus for classification. RF-FI is leveraged to identify influential features to improve the classification accuracy in dyslexia prediction. The importance score is determined using the Gini index by analyzing the contribution of the feature to reduce impurity (Madhu PK et al., 2021). It is evaluated across all the decision trees and aggregated to provide the overall FI. By ranking features based on these scores, researchers can prioritize those eye-tracking features that have the greatest impact the decision of the ML model.

In the proposed work, RF-FI is evaluated for the available features extracted from the eye-movement tracking dataset as described in Section 3.2. Figure 4 shows the distribution of RF-FI across the considered features. It is observed that the dispersion and velocity-based features (from feature no. 30) have more importance in influencing the dyslexia classification.

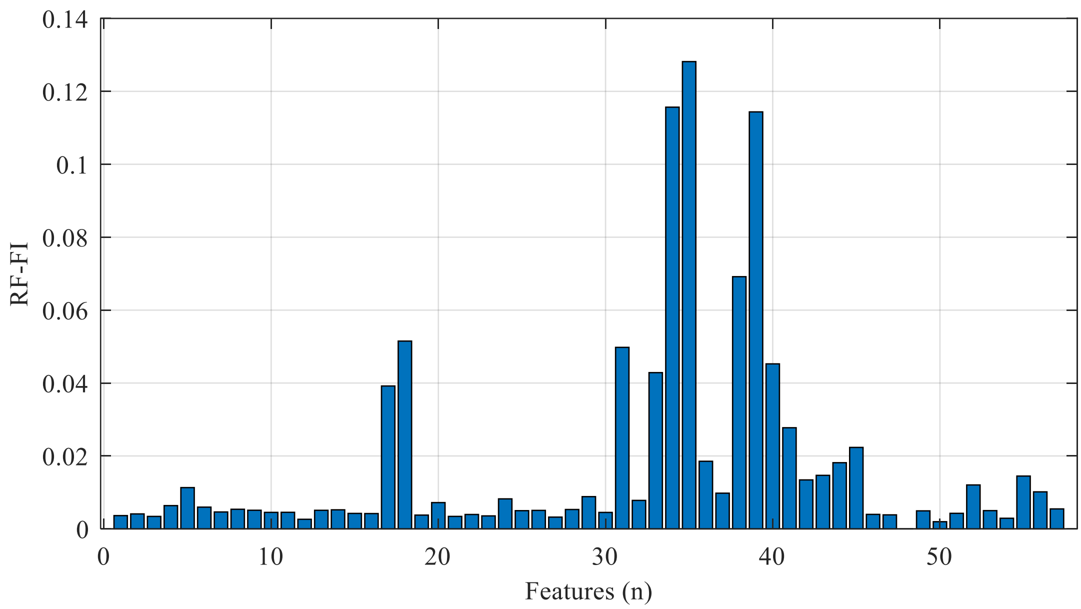


Figure 4. Distribution of RF-FI across features.

3.3.3. XAI-SHAP based Feature Ordering

SHAP is one of the well-established XAI principals which is based on Shapley values from cooperative game theory (Robaa et al., 2024). It provides transparency, accountability, and interpretability of the features impact on the ML classification problems. Hence, powerful tool for feature selection in ML classification, especially in sensitive domains like dyslexia prediction (Ding et al., 2025). SHAP assigns an importance-value for each feature based on its (Robaa et al., 2024)contribution to the model's output. It also provides a consistent and interpretable way to understand feature impact (Lundberg & Lee, 2017). In the proposed work SHAP is used to order the dyslexia features extracted from the recorded eye movements.

RF is used as base classifier to analyze the feature impact on the classification as performance illustrated in Figure 5. SHAP values of top 5 features are shown for illustration, and it is observed that negative value indicates a LR-D class and positive indicates HR-D class. Eye-tracking velocity features such as scan path and fixation duration exhibited higher positive SHAP values. It indicates a higher likelihood of dyslexic (HR-D) with higher duration. On the other hand, average reading speed demonstrated an inverse trend, i.e., lower speeds are associated with higher positive SHAP values, suggesting a higher predictive contribution toward HR-D class. Thus, it is evident that SHAP can select features, which serves as a reliable discriminative indicator for ML based dyslexia prediction.(Robaa et al., 2024)

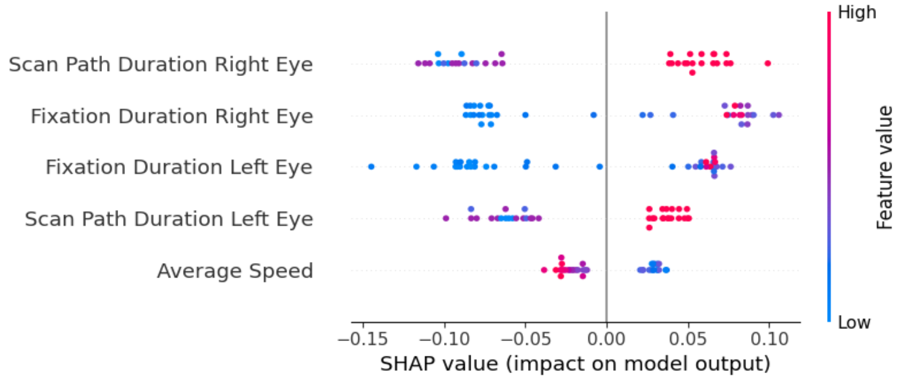


Figure 5. Summary of SHAP value distribution for top 5 features

3.4. Ensemble Rank Aggregation Methods

Features ordered using PCA, RF-FI, and XAI-SHAP are aggregated to provide ensemble feature selection technique. Spearman rank correlation coefficients are evaluated to analyze the variations brought in by these feature ordering techniques (Qiu et al., 2024)(Hauke & Kossowski, 2011)as in Figure 6. The heat map shows the correlation or similarity across the feature ranking. It is observed that PCA based ranking exhibits a low correlation with RF-FI (ρ = 0.18) and SHAP (ρ = 0.10), reflecting its unsupervised technique and focus on data variance rather than predictive contribution (independent of ML classifier). It enables the proposed ensemble technique to have a complementary perspective provided by different ranking strategies. Supervised methods like RF-FI and SHAP are highly correlated in ordering the feature for dyslexia prediction.

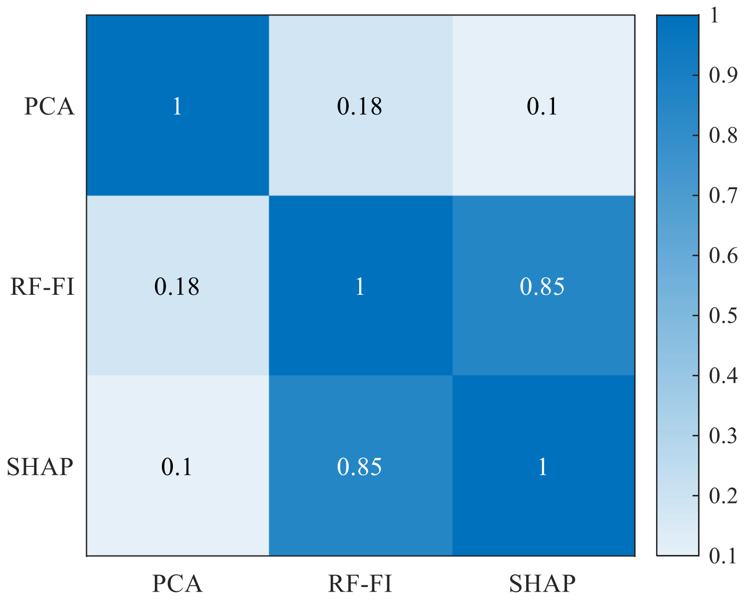


Figure 6. Heat map for spearman coefficient of feature ordering methods

In the proposed work, about four different feature rank aggregation techniques namely, Borda Count, Robust Rank Aggregation (RRA), Markov-Chain based Ranking, Bayesian Ranking techniques are considered. These techniques are widely used and enable us to analyze the classification performance across different modality of rank aggregation.

3.4.1. Borda Count Aggregation

Borda Rank Aggregation is a widely used method for combining multiple ranked lists into a single consensus ranking

Boehmer, N.,]. It enables to aggregate by synthesizing diverse perspectives on feature relevance and to identify the important features across multiple ranking systems. Each feature receives a score based on its position in individual rankings, and the final score determines its overall rank. For the given feature (i), the Borda score (BS_i) is determined as in (1) across the available feature ordering methods ($N_r = 3$) i.e. PCA, RF-FI and XAI-SHAP. The consensus ranks of the features are determined using Borda score as in (2).

$$BS_i = \sum_{j=1}^{N_r} \left(N_f - r_{ij}\right) | i \in (1, N_f] \tag{1}$$

$$\boldsymbol{r_{bc}} = argsort\ (\boldsymbol{BS}) \mid \boldsymbol{BS} = \{BS_i\}, i \in [1, N_f] \tag{2}$$

3.4.2. Robust Rank Aggregation (RRA)

Robust Rank Aggregation (RRA) is a statistical technique which combines multiple feature ranks and it is efficient to minimize the influence of noise or outliers (Cucuringu, 2016). Unlike simple averaging, RRA evaluates the statistical significance of a feature ranking using a probabilistic model. It determines the features consistently ranked highly across different ranking methods. The rank matrix ($\boldsymbol{r} \in \mathbb{R}^{N_f \times N_r}$) is formulated using the ranks obtained for the considered features (N_f) with the various ranking methods (N_r). Each rank is normalized as in (3) enabling to determine probability distribution. For each feature (j), the minimal probability across the ranking techniques is determined as in (4).

$$\tilde{r}_{ij} = \frac{r_{ij}}{N_f} \tag{3}$$

$$r_i^{min} = min_j \left(\tilde{r}_{ij}\right) \mid j \in [1, N_r] \tag{4}$$

The p-value of the features (p_i) are evaluated using the cumulative distribution function (CDF) of the Beta distribution (β_{CDF}) as in (5). The shape of the Beta CDF is determined by two shape parameters namely α and β. The α parameter is responsible for concentrating the distribution near 0, whereas beta controls the concentration near 1. These shape parameters determine the skewness and overall shape of the Beta distribution.

$$p_i = 1 - \beta_{CDF}(r_i^{min}, \alpha, \beta) \tag{5}$$

In the proposed work, the shape parameter α is set to unity and β is set to the number of ranking system (N_r) used. It makes the CDF encounter negatively skewed distribution improving the robustness of the rank aggregation as in Figure 7. A low p-value is received for the features that have a very low minimum ranking, which makes the feature statistically significant and consistently important. Hence, RRA is capable of detecting robust features that repeatedly rank higher enabling to include the best performances across various ranking methods with different modalities. Finally, the features with lower p-values are assigned a higher ranking as in (6) and the aggregated feature rank is generated.

$$\boldsymbol{r_{rra}} = argsort\ (\boldsymbol{p}) \mid \boldsymbol{p} = \{p_i\}, i \in [1, N_f] \tag{6}$$

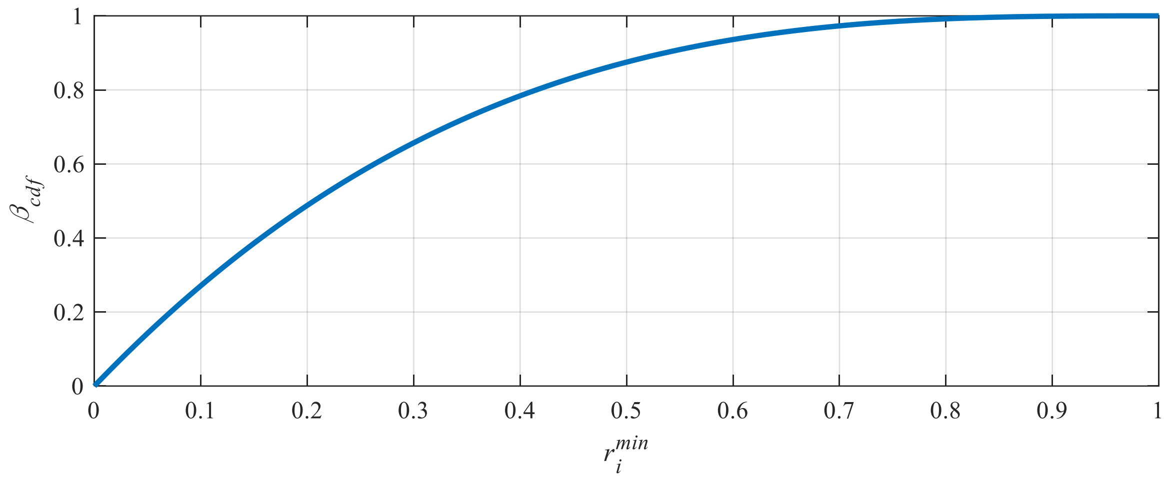


Figure 7. Beta CDF distribution used in RRA.

3.4.3. Markov Chain-Based Rank Aggregation

Markov Chain-based rank aggregation is a probabilistic technique that models the process of selecting a consensus ranking (Lin, 2010). It is inspired from PageRank algorithm and enables a robust method for combining multiple features ranking available. This approach is robust to noise and inconsistencies among individual ranking systems, which motivates to investigate it for the proposed work on dyslexia feature ordering. In this method, each feature rank is treated as a state in a Markov chain, and transition probabilities are derived from pairwise preference comparisons across multiple ranking methods.

In Markov Chain-based rank aggregation, a transition matrix ($\boldsymbol{T} \in \mathbb{R}^{N_f \times N_f}$) is constructed to indicate the preference of a particular feature over the other features in each ranking system. For instance, each element in the translation matrix ($T_{ij} \in \boldsymbol{T}$) quantitatively indicates the preference of i^{th} feature over the j^{th} feature. The translation matrix is normalized as in (7) using the teleportation factor (τ) as similar to PageRank algorithm (Sharma et al., 2020). The stationary probability distribution function ($\boldsymbol{P_m}$) calculated using the normalized translation matrix ($\widetilde{\boldsymbol{T}}$). Power iteration method was adopted to find the dominant Eigen vector of the normalized translation matrix to result in a stationary probability distribution. The state vector of $\boldsymbol{P_m}$ for the given t^{th} iteration is noted as P_m^t and determined using the past state and the normalized translation matrix ($\widetilde{\boldsymbol{T}}$) as in (8). The iteration continues until convergences ($\boldsymbol{P_m^{t+1}} \approx \boldsymbol{P_m^t}$) and the stationary probability is determined. Finally, the resultant stationary probability is sorted to determine the aggregated ranking for dyslexia features as in (9).

$$\widetilde{\boldsymbol{T}} = \tau \boldsymbol{T} + (1-\tau)\frac{1}{N_f}\boldsymbol{I}^{N_f \times N_f} \tag{7}$$

$$\boldsymbol{P_m^t} = \boldsymbol{P_m^{t-1}}\widetilde{\boldsymbol{T}} \tag{8}$$

$$\boldsymbol{r_{mc}} = argsort\,(\boldsymbol{P_m^t}) \mid \boldsymbol{P_m^{t+1}} \approx \boldsymbol{P_m^t}, \boldsymbol{P_m^t} = \{p_i^m\}, i \in [1, N_f] \tag{9}$$

3.4.4. Bayes Rank Aggregation

Bayesian rank aggregation is a probabilistic framework that enables to model the uncertainty and latent quality of each feature rank. It employs Bayes' theorem to combine the probable ranking from each feature ordering (Lin, 2010). Each feature ordering method is treated as independent evidence and uses Bayesian update to compute a fused probability on the importance of each feature. Initially, the ranks from various feature ordering are converted to the evidence probability (P_e^{ij}) as in (10). Bayes theorem is applied using product rule to fuse this evidence probability across the feature ordering methods to obtain the resultant probability (P_b^i) as in (11). Finally, the resultant ranks are obtained using the calculated probability as in (12).

$$P_e^{ij} = \frac{1/r_{ij}}{\sum_{i=1}^{N_f} 1/r_{ij}} \tag{10}$$

$$P_b^i = \frac{\prod_{j=1}^{N_r} P_e^{ij}}{\sum_{i=1}^{N_f} \prod_{j=1}^{N_r} P_e^{ij}} \tag{11}$$

$$\boldsymbol{r_{br}} = argsort\ (\boldsymbol{P_b}) \mid \boldsymbol{P_b} = \{P_b^i\}, i \in [1, N_f] \tag{12}$$

The correlation across these rank aggregation methods (Borda Count, RRA, Markov Chain, and Bayesian Rank Aggregation) is evaluated using the Spearman rank correlation heat map as illustrated in Figure 8. More correlation between RRA and Markov Chain is observed, indicating that both methods have highly similar feature rankings. This is mainly due to the consistency between these probabilistic models-based approaches in capturing underlying feature importance. Simpler scoring mechanism of Borda Count shows significant volition with RRA and Markov Chain, which may overlook the complex relationships between features and dyslexia prediction. Bayesian Rank, being a probabilistic product-based approach exhibits a moderate correlation with Borda Count and remains weakly correlated with the others. Hence, there is a need for further analysis on the downstream process of dyslexia prediction using ML classifier.

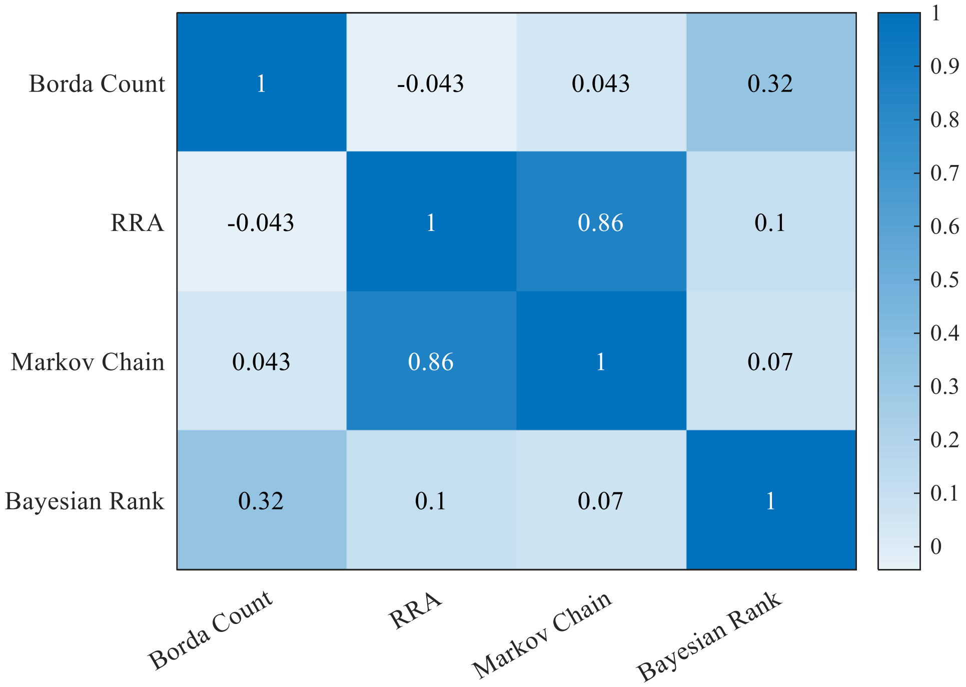


Figure 8. Heat map of Spearman rank correlation for rank aggregation methods.

3.5. Feature Order Based Accumulation Effect Analysis

The effect on classification performance with the inclusion of features (Jayabal et al., 2018) is analyzed as in Figure 9. The extracted features and various rank aggregation order obtained using various methods are considered. RF classifier is considered as a baseline ML classifier to ensure uniformity. An Incrementor module uses the feature order to iteratively accumulate feature index. The feature indexes are used by the feature selector to select and form a new feature set, which is a subset of the overall feature set. The selected feature subset is used to train a ML classifier, whose performance is evaluated using standard metrics like test data accuracy. A trigger mechanism is then employed to signal the incrementor to expand the feature subset by including additional features in the next iteration and continues until all the features get accumulated. Performance improvement resulted with the inclusion of the feature is analyzed. It enables us to identify the optimal feature set to detect dyslexia with reliable classification accuracy.

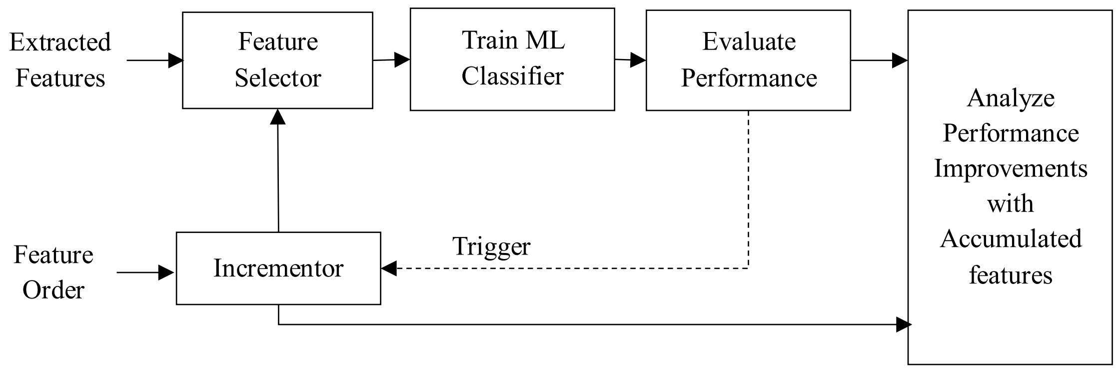
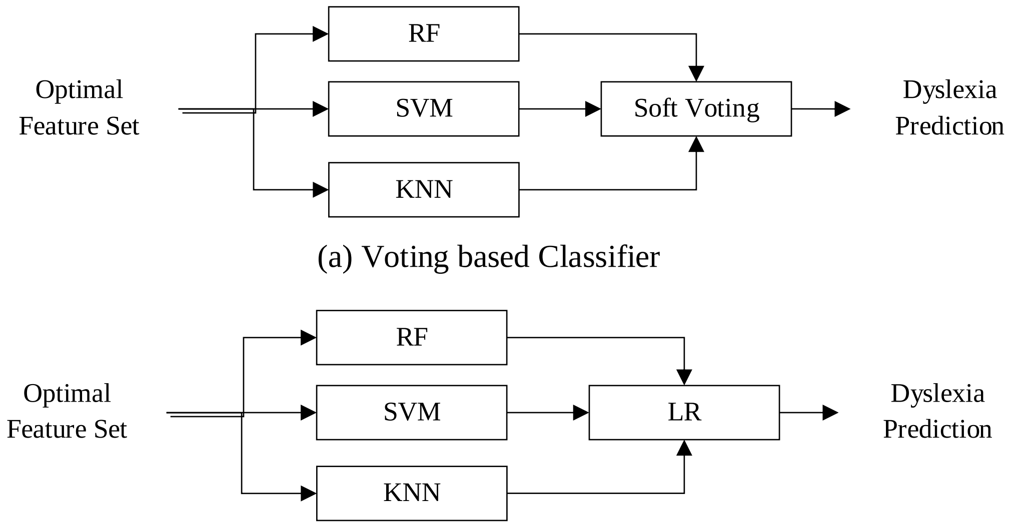


Figure 9. Feature order-based accumulation effect analysis

3.6. Training Machine Learning Model

The selected features are used to train the various ML classifiers to evaluate the prediction performance. Four base classifiers namely, RF, Support vector machines (SVM) and K-nearest neighborhood (KNN) are considered in this work. This enables us to understand the capability of leveraging the performance of simple base classifiers with optimal feature selection. Also, these base classifiers are less computationally intensive and can be deployed in edge devices, which enables faster and on-field prediction of dyslexia.

Additionally, two variants of ensemble classifiers constructed using the base classifiers is also evaluated (AlGhamdi, 2022). Voting and stacking based classifiers are constructed as illustrated in Figure 10. The voting classifier is obtained by using predictions from RF, SVM, and KNN. These predictions are combined using soft voting. Unlike hard voting, final prediction is determined using the predicted probabilities rather than class labels. This enables a smother decision and efficient in handling uncertainties. Similar to voting classifier, stacking classifier also uses the predictions from RF, SVM, and KNN. However, it uses a meta-classifier to fuse the prediction probabilities. In this work, linear regression (LR) is used as meta-classifier, which is pre-trained to understand the sampling strategy to combine the predictions from different base classifiers. It enables combining the performance of multiple classifiers and improves overall accuracy.



(b) Stacking based Classifier

Figure 10. Block diagram of ensemble ML classifier

4. Results and Discussion

The accumulation effect is analyzed for various rank aggregation methods by inclusion of all the 57 features using test data classification accuracy as shown in Figure 11. It is observed that the Borda and Bayesian based ranking aggregation exhibits similar behavior and inclusion of features from rank 10 to 45 does not provide much improvement in classification accuracy. As more features are included, there is a significant improvement in accuracy. It makes both rank aggregation methods include predominate of the features to achieve reliable classification accuracy. On the other hand, Markov-chain based rank aggregation results in a higher accuracy with a smaller number of features. However, accuracy does not show significant improvement amidst inclusion of more features. RRA results in improvement in accuracy with inclusion of features beyond rank 20. It reaches the highest classification accuracy (0.9455) for the selected 24 features. Hence, 24 features ordered using RRA are considered as optimal features.

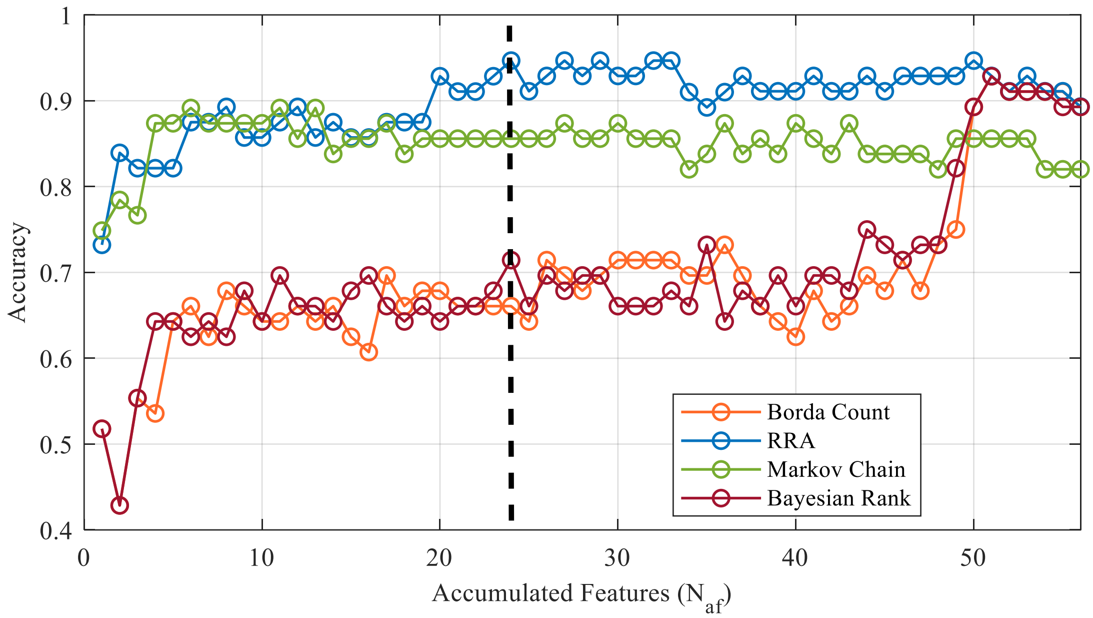


Figure 11. Accumulation effect analysis of rank aggregation methods.

Further, the performance of the RRA based rank aggregation is compared with the individual feature ordering techniques. The correlation of the feature ordered with RRA is compared with the individual techniques like PCA, RF-FI, and XAI-SHAP as shown in Figure 12. It is observed that the RRA predominantly orders feature as like RF-FI and XAI-SHAP. However, it also provides significant importance to the PCA with improved correlation. This indicates the proposed RRA can combine the best of the feature orders inferred by various modalities.

Figure 12. Heat map comparison of RRA with individual feature ordering techniques.

Performance improvement observed using accumulation effect with RRA is also compared with the individual feature ordering techniques as illustrated in Figure 13. A surge in accuracy is observed with the inclusion of features (above 20) ordered by PCA and XAI-SHAP techniques, whereas RF-FI is insensitive to feature inclusion until more than 45 features are considered. The peak performance along with the corresponding selected features across all the considered methods are evaluated as in Table 2. The PCA is capable of reducing the features to some extent but unable to provide higher classification accuracy. Other techniques like RF-FI and XAI-SHAP provide higher classification accuracy with more features. However, RRA remains superior with its robust aggregation characteristics and able to provide higher classification accuracy.

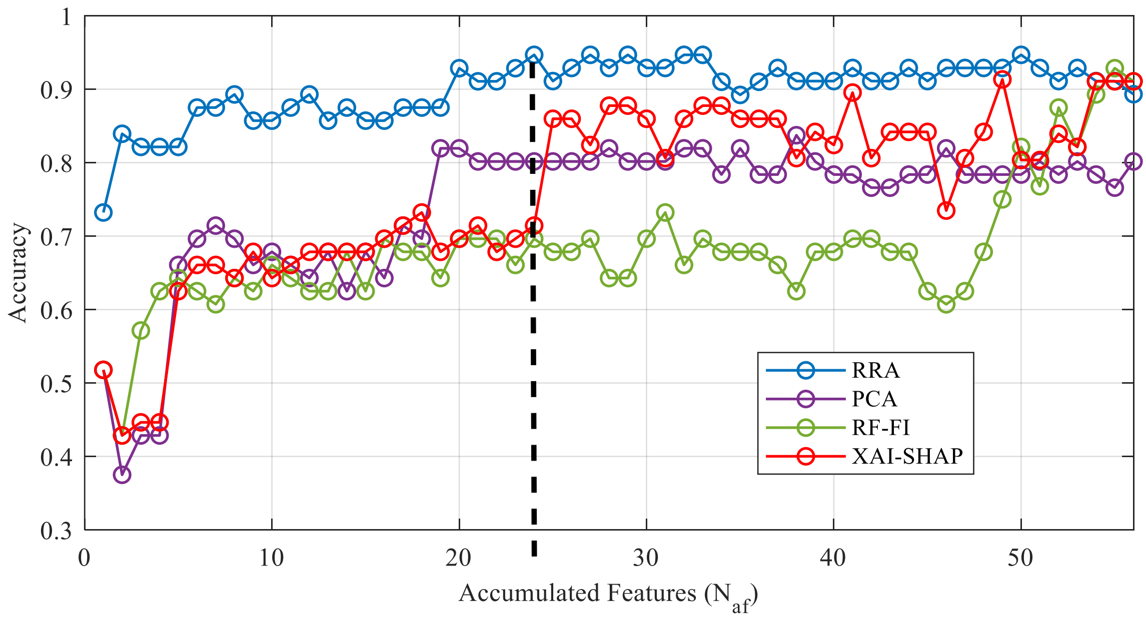


Figure 13. Accumulation effect analysis of RRA with individual feature ordering methods.

Table 2. Performance comparison of feature ordering techniques

Feature Selection Method	No. of Selected Features	%Reduction in Features	Classification Accuracy
PCA	38	33.33	0.8373
RF-FI	55	3.51	0.9286
XAI-SHAP	49	14.04	0.9133
RRA (Aggregation)	**24**	**57.90**	**0.9455**

** Total no. of features: 57*

The reduction in features is also analysed based on its types namely, Statistical, Dispersion and Velocity-based as in Figure 14. The selected feature from each type is evaluated and it is observed that velocity-based features are selected predominantly. It clearly indicates the reliability of the proposed method as velocity of eye movement is a trustable indicator for dyslexia. The statistical based features are least selected followed by dispersion features.

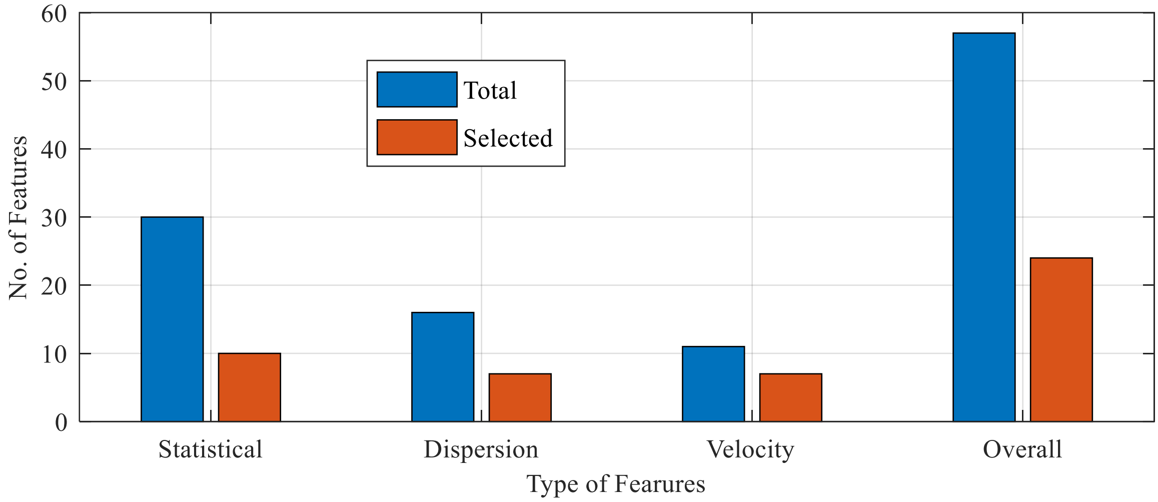


Figure 14. Analysis of feature selection based on its type

With the selected features, the performance is evaluated for various classifier as described in Section 3.6. The classification accuracy is evaluated for training and testing data to analyse the potential overfitting as in Figure 15 and to generalize the ML classifier for deployment. Among the base classifiers, RF exhibits a better performance, and NBC fails to provide reliable accuracy for training and testing dataset. The experimental results also suggest that ensemble methods, particularly stacking, is more effective in capturing the complex patterns in the features of eye movement data for accurate dyslexia prediction. Thus, the proposed work results in 24 features ordered using RRA and stacking classifier to predict dyslexia accurately.

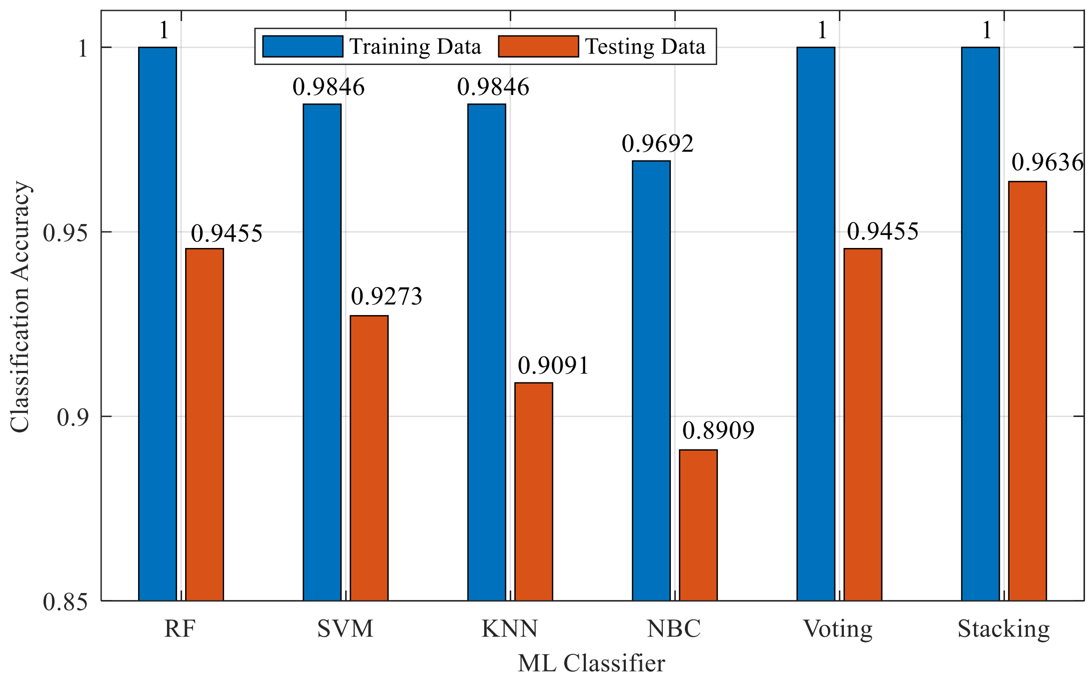


Figure 15. Performance analysis of various ML classifiers.

5. Conclusion

Eye movement is one of feature rich marker to detect dyslexia among students. It is one of the easily acquirable data and does not require specialized test process. Hence, it is one of the preferred techniques for dyslexia prediction, which can be used to screen large number of people. However, challenges and complexity associated in understanding the eye movement is addressed using ML techniques. ML uses features extracted from eye movement signals to predict the dyslexia. Three different feature extraction techniques namely statistical, dispersion-based and velocity-based methods are used. These methods provide about 57 features in combination. Availability of large no. of features demands feature ordering and selection techniques to improve the prediction performance with optimal feature set. Three different feature ordering techniques including traditional PCA and RF-FI with XAI-SHAP are considered. These techniques use different modalities to order the feature based certain principles like feature variances (PCA), feature importance (RF-FI), and feature contribution (XAI-SHAP), respectively. Spearman correlation analysis illustrates lower correlation of PCA with RF-FI (0.18) and XAI-SHAP (0.1).

Ensemble technique has been proposed to aggregate these feature order to fuse these modalities and improve the prediction accuracy. Four widely used rank aggregation techniques namely, Borda Count, RRA, Markov-Chain based Ranking, Bayesian Ranking techniques are investigated to aggregate the feature order. It is observed that RRA can aggregate the feature order optimally, which has been ascertained using accumulation analysis. Accumulation effect tends to analyse the improvement in classification accuracy with inclusion of feature in the order suggested by RRA. About 24 features are selected as optimal set providing a classification accuracy of 94.6% using the baseline RF classifier. It makes the proposed work to eliminate ~58% of features and improve the classification accuracy as compared with the other feature selection techniques. Spearman correlation analysis demonstrates that RRA achieves stronger alignment with PCA (0.57) and other feature ranking methods such as RF-FI (0.85) and XAI-SHAP (0.88). This indicates that RRA effectively integrates the individual feature orderings to extract the best from each technique and contributing to optimal model performance. Performance analysis with other ML classifiers like SVM, RF, KNN and ensemble classifiers like voting and stacking classifiers are also carried out. Results indicate that stacking classifier constructed using RF, SVM, and KNN with LR as meta-classifier can provide a classification accuracy of 96.36% with the selected 24 features.

The proposed work can be extended to include features extracted from other medium like EEG, neuro-imaging and behavioural datasets for holistic analysis of diversified characteristics. Adaptive feature selection with SHAP and confidence-driven approach to dynamically select the features to improve the model prediction accuracy will be investigated in future.

Reference

Ahmad, N., Rehman, M. B., El Hassan, H. M., Ahmad, I., & Rashid, M. (2022). An Efficient Machine Learning-Based Feature Optimization Model for the Detection of Dyslexia. *Computational Intelligence and Neuroscience*, *2022*(1), 8491753.

AlGhamdi, A. S. (2022). Novel ensemble model recommendation approach for the detection of dyslexia. *Children*, *9*(9), 1337.

Alqahtani, N. D., Alzahrani, B., & Ramzan, M. S. (2023). Deep learning applications for dyslexia prediction. *Applied Sciences*, *13*(5), 2804.

Appadurai, J. P., & Bhargavi, R. (2021). Eye movement feature set and predictive model for dyslexia: Feature set and predictive model for dyslexia. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *15*(4), 1–22.

Coenen, L., Grünke, M., Becker-Genschow, S., Schlüter, K., Schulden, M., & Barwasser, A. (2024). A Systematic Review of Eye-Tracking Technology in Dyslexia Diagnosis. *Insights into Learning Disabilities*, *21*(1), 45–65.

Cucuringu, M. (2016). Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and SDP synchronization. *IEEE Transactions on Network Science and Engineering*, *3*(1), 58–79.

Dewanjee, S., & Muntaha, S. (2024). Hybrid Deep Learning for Dyslexia Identification through Heterogeneous Cognitive and Behavioral Data Analysis. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7.

Ding, N., Peng, P., Tang, J., Ding, Y., & Zhao, J. (2025). An investigation of phonological predictors in Chinese developmental dyslexia using a machine learning approach. *Reading and Writing*, 1–25.

Donnarumma, F., Frosolone, M., & Pezzulo, G. (2023). Integrating large language models and active inference to understand eye movements in reading and dyslexia. *ArXiv Preprint ArXiv:2308.04941*.

Haller, P., Säuberli, A., Kiener, S. E., Pan, J., Yan, M., & Jäger, L. (2022). Eye-tracking based classification of Mandarin Chinese readers with and without dyslexia using neural sequence models. *ArXiv Preprint ArXiv:2210.09819*.

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, *30*(2), 87–93.

Jayabal, R., Vijayarekha, K., & Rakesh Kumar, S. (2018). Design of ANFIS for hydrophobicity classification of polymeric insulators with two-stage feature reduction technique and its field deployment. *Energies*, *11*(12), 3391.

Jothi Prabha, A., & Bhargavi, R. (2022). Prediction of dyslexia from eye movements using machine learning. *IETE Journal of Research*, *68*(2), 814–823.

Kirkby, J. A., Webster, L. A. D., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin*, *134*(5), 742.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 1–45.

Lin, S. (2010). Rank aggregation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(5), 555–570.

Liyakathunisa, Alhawas, N., & Alsaeedi, A. (2023). Early prediction of dyslexia risk factors in kids through machine learning techniques. In *Kids Cybersecurity Using Computational Intelligence Techniques* (pp. 225–242). Springer.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Madhu PK, R., Subbaiah, J., & Krithivasan, K. (2021). RF-LSTM-based method for prediction and diagnosis of fouling in heat exchanger. *Asia-Pacific Journal of Chemical Engineering*, *16*(5), e2684.

Mukhtar, I. S., Ezinne, N. E., Mohamad Shahimin, M., Mohd-Ali, B., Oghre, E., Zeried, F. M., & Osuagwu, U. L. (2024). Age-Matched Comparative Analysis of Binocular Vision Anomalies among Children with Dyslexia in Northern Nigeria. *Pediatric Reports*, *16*(3), 566–578.

Nilsson Benfatto, M., Öqvist Seimyr, G., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for dyslexia using eye tracking during reading. *PloS One*, *11*(12), e0165508.

Prabha, A. J., & Bhargavi, R. (2020). Predictive model for dyslexia from fixations and saccadic eye movement events. *Computer Methods and Programs in Biomedicine*, *195*, 105538.

Prabha, A. J., Bhargavi, R., & Harish, B. (2021). An efficient machine learning model for prediction of dyslexia from eye fixation events. *New Approaches Eng. Res*, *10*(6), 171–179.

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. *Proceedings of the 12th International Web for All Conference*, 1–8.

Robaa, M., Balat, M., Awaad, R., Omar, E., & Aly, S. A. (2024). Explainable AI in Handwriting Detection for Dyslexia Using Transfer Learning. *ArXiv Preprint ArXiv:2410.19821*.

Sharma, P. S., Yadav, D., & Garg, P. (2020). A systematic review on page ranking algorithms. *International Journal of Information Technology*, *12*(2), 329–337.

Svaricek, R., Dostalova, N., Sedmidubsky, J., & Cernek, A. (2025). INSIGHT: Combining Fixation Visualisations and Residual Neural Networks for Dyslexia Classification From Eye-Tracking Data. *Dyslexia*, *31*(1), e1801.

Vajs, I., Ković, V., Papić, T., Savić, A. M., & Janković, M. M. (2022). Spatiotemporal eye-tracking feature set for improved recognition of dyslexic reading patterns in children. *Sensors*, *22*(13), 4900.

Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., Liu, G., & Beal, B. (2020). The prevalence of dyslexia: A new approach to its estimation. *Journal of Learning Disabilities*, *53*(5), 354–365.

Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., Zhao, J., & Weng, X. (2022). Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. *Brain Sciences*, *12*(2), 240.

Ziegler, J. C., Perry, C., & Zorzi, M. (2020). Learning to read and dyslexia: From theory to intervention through personalized computational models. *Current Directions in Psychological Science*, *29*(3), 293–300.