ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202310.00239

Machine Learning in Early Screening for Developmental Dyslexia in Children: Postprint

Authors: Bu Xiao' ou, Wang Yao, Du Yawen, Pei Wang

Date: 2023-10-09T00:00:00+00:00

Abstract

Developmental dyslexia severely impacts children's academic achievement, mental health, and social adaptation capabilities. In recent years, machine learning has been gradually applied to early screening of children with dyslexia due to its powerful data processing and mining capabilities, accumulating substantial results across multiple domains including standardized psychoeducational testing, eye tracking, game-based testing, and brain imaging, and achieving more accurate, efficient, flexible, and reliable classification outcomes. However, machine learning still exhibits limitations in subject selection, data collection, translational potential, and security and privacy. Future research must prioritize the scientific rigor of early screening systems for preschool children with dyslexia, while actively constructing multimodal databases, identifying optimal algorithms among various alternatives to obtain optimal parameters, and ultimately achieving widespread implementation in clinical practice.

Full Text

Application of Machine Learning in Early Screening of Children with Developmental Dyslexia

Bu Xiaoou, WANG Yao, DU Yawen, WANG Pei

(Department of Special Education, Faculty of Education, East China Normal University, Shanghai 200062, China)

Abstract

Developmental dyslexia is a neurodevelopmental disorder with a neurobiological basis that not only severely restricts children' s academic achievement, but also negatively affects their psychological health and social adaptation. In recent years, machine learning has been gradually applied to the early screening of children with dyslexia due to its powerful data processing and mining capabilities, accumulating rich results in standardized psychoeducational testing, eye-tracking, game-based testing, and brain imaging, thereby achieving more accurate, efficient, flexible, and reliable classification outcomes. However, machine learning still has limitations regarding participant selection, data collection, translational potential, and security and privacy. Future research should focus on the scientific validity of early screening systems for preschool children with dyslexia, actively construct multimodal databases, identify optimal algorithms among multiple classifiers to obtain optimal parameters, and ultimately achieve widespread clinical application.

Keywords: developmental dyslexia, machine learning, early screening, children

Classification Number: R395

Developmental dyslexia (DD) is an extremely complex neurodevelopmental disorder characterized by persistent difficulties in reading, spelling, and writing despite normal intelligence and intact visual and auditory functions (Kaisar, 2020). The prevalence of dyslexia across different languages and cultures is approximately 5%–15% (Tamboer et al., 2016), and it shows intergenerational transmission (Zahia et al., 2020). Currently, children are typically not identified as having dyslexia until 2nd grade or higher during the process of acquiring reading skills (Sanfilippo et al., 2020). In economically underdeveloped countries, children from impoverished backgrounds are identified even later (Ballester et al., 2021). By this time, they have often missed the optimal intervention window —the early stage of enhanced brain plasticity from kindergarten to 1st grade (Fox et al., 2010). Numerous studies have found that children with dyslexia become trapped in a vicious cycle of poor academic performance, reduced self-efficacy, and insufficient learning motivation (Burns et al., 2022), and even exhibit extremely high dropout rates and mental health problems (Livingston et al., 2018). If these children do not receive timely identification and intervention, the negative effects of dyslexia may persist from early childhood through adulthood (Farah et al., 2021). Therefore, efficient early screening and effective early intervention are critical for the development of children with dyslexia.

To date, dyslexia screening has primarily relied on standardized psychoeducational testing (Lee et al., 2022), eye-tracking (Prabha & Bhargavi, 2019), web/mobile games (Borleffs et al., 2018), and brain imaging techniques (Usman et al., 2021). Standardized psychoeducational testing typically employs the IQ-achievement discrepancy model (Fletcher et al., 2019), response-to-intervention (RTI) model (Miciak et al., 2014), and pattern of strengths and weaknesses (PSW) model (Hale et al., 2010) to assess and quantify individuals' cognitive abilities such as intelligence, phonological processing, reading skills, and vocabulary development, thereby identifying individuals with dyslexia (Miciak & Fletcher, 2020). Regarding eye-tracking technology, researchers distinguish between children with and without dyslexia by recording eye movement characteristics during reading, including fixation/regression time and frequency, saccade amplitude and frequency, blink frequency and count, and binocular coordination (Hmimdi et al., 2021). Some researchers have also developed web-based e-learning systems and mobile games (e.g., Deslixate and GraphoGame) in a gamified format to generate specific phonological or cognitive tests, aiming to identify children with dyslexia through educational games (Larco et al., 2021; Ojanen et al., 2015). With the development of cognitive neuroscience technologies, an increasing number of studies have used brain imaging techniques to obtain brain structure, morphology, functional activation, and geometric properties, utilizing group mean differences to distinguish between children with dyslexia and typically developing children (Livingston et al., 2018; Sihvonen et al., 2021; Yang et al., 2021).

However, symptoms of dyslexia in children show tremendous individual variability, as different etiological factors lead to different subtypes of dyslexia (Aaron et al., 1999). Additionally, traditional dyslexia detection techniques are inefficient and time-consuming, with unclear sensitivity and specificity indicators, making it difficult to meet the demand for large-scale and rapid screening of children with dyslexia (Usman et al., 2021). More importantly, dyslexia is associated with multiple neurological, behavioral, and environmental factors that interact in complex ways to cause the disorder (Catts et al., 2017; McGrath et al., 2020). Therefore, accurate diagnosis cannot be achieved by relying on single or a few factors alone (Catts & Petscher, 2022), and even traditional multifactorial methods cannot cover all possible factors and relationships (Walda et al., 2022). A relatively novel and effective approach for studying complex systems is machine learning (Kaisar, 2020). Machine learning (ML) uses computer algorithms to enable machines to learn patterns from large amounts of empirical data, automatically identifying patterns to make predictions or decisions (Gilvary et al., 2020). In recent years, because it can provide higher detection accuracy and better prediction results, some researchers have begun to apply machine learning to improve the precision and sensitivity of dyslexia screening. Therefore, this study aims to clarify potential development paths and directions for machine learning research in dyslexia by integrating the latest advances, main application areas, and future development directions of machine learning in dyslexia screening.

We conducted a literature search for studies using machine learning methods to classify and identify dyslexia since 2016, using databases including Web of Science, Elsevier Science Direct, EBSCO, and PubMed. The search keywords were "Dyslexia/Reading Disability" AND "Identification/Screening/Detection/Recognition/Prediction/Diagnosis" AND "Machine Learning/Deep Learning/Artificial Intelligence"AND "Child/Children/Preschool." Considering the rapid development and iteration of machine learning technology, and since the first Chinese machine learning study on dyslexia was published in 2016, the literature search date range was set from January 1, 2016, to October 1, 2022. The inclusion criteria were: (1) English empirical journal articles and conference papers with full text available, containing clear research questions, methods, and conclusions supported by detailed data; (2) Study participants were children under 18 years old, with both typically developing control groups and dyslexia groups, and children with dyslexia had no other comorbidities (such as dyscalculia, dysgraphia, autism, etc.); (3) Studies used or combined machine learning methods for dyslexia screening. We independently screened according to these criteria and finally included 25 articles in this systematic review (see Table 1). Figure 1 [Figure 1: see original paper] and Figure 2 [Figure 2: see original paper] show the literature screening process and the annual distribution of included literature, respectively.

Main Steps in Machine Learning-Based Early Screening for Dyslexia

Data Collection

The first step in machine learning-based dyslexia screening is obtaining data using appropriate technical means. Standardized psychoeducational testing provides the earliest and most extensive evidence for machine learning model construction. The data obtained show clear behavioral manifestations of individuals with dyslexia, mainly including reading, phonological processing, working memory, and auditory-visual discrimination. Chen et al. (2017) used the Dutch version of the McArthur-Bates Communicative Development Inventory (N-CDI) to measure early vocabulary development in 476 typically developing children aged 17-35 months and used machine learning algorithms to predict children at familial risk for dyslexia. Shamir et al. (2019) used a self-developed brief dyslexia screening tool (Zippy 6) to measure the cognitive and phonological abilities of 125 children aged 6–14 years (81 with dyslexia) and used machine learning algorithms to distinguish between children with dyslexia and typically developing children. Tolami et al. (2021) collected language samples from 54 children aged 8-11 years (29 with dyslexia), used computational linguistics methods to extract differential features of dyslexia such as spelling and grammatical errors, lexical diversity, grammatical complexity index, and readability, and used machine learning models to diagnose dyslexia. In Chinese dyslexia research, Wang and Bi (2022) collected cognitive test batteries from 399 children aged 7–13 years with dyslexia, measured reading fluency, reading accuracy, phonological awareness, morphological awareness, rapid naming, and orthographic awareness, and used deep learning models to predict symptoms of Chinese children with dyslexia. Lee et al. (2022) collected Chinese character datasets from 1,015 children aged 7-13 years (454 with dyslexia), used multiple algorithms to classify Chinese character response features (such as strokes, graphemes, tones, etc.), character structure, response features (such as orthography, phonetic roots, etc.), and personal characteristics, and ultimately constructed a Chinese dyslexia screening model based on core features such as Chinese character structure, writing accuracy, lexical status, strokes, tones, and grade level.

Notably, eye movement features have become commonly used indicators for machine learning-based dyslexia classification. Their combination with machine learning provides fine-grained information about cognitive processes (Raatikainen et al., 2021) and can serve as a high-precision screening tool for dyslexia. Bhargavi and Prabha (2020) collected eye movement data from 185 children aged 9–10 years and used multiple machine learning algorithms to improve prediction accuracy, finding that the optimal feature set with high accuracy included average fixation count, average fixation duration, average saccade time, total saccade count, and average fixation count. Ileri et al. (2022) recorded electrooculography (EOG) signals from 33 children aged 9–10 years (20 with dyslexia) during text reading and used machine learning to analyze different types of eye movement patterns to screen for dyslexia.

With the increasing popularity of smart mobile devices, data collection technologies based on web/mobile games have gained a broad user base. Researchers have developed various applications and games to support, detect, and treat dyslexia (Ahmad et al., 2022). Gamified designs mostly measure language ability, perceptual processing, working memory, executive function, and reading skills, using rich game elements to attract and motivate users. Rello et al. (2020) designed an online game to assess behavioral and cognitive deficits, collecting data from 3,644 users aged 7–17 years (including 392 individuals with dyslexia) to build a machine learning model for dyslexia screening. Rauschenberger et al. (2022) collected rhythm and frequency data from 313 children aged 7–12 years (including 116 with dyslexia) playing the web game "MusVis" and used machine learning for model training and prediction.

The essential characteristic of dyslexia is subtle spatially distributed changes in brain anatomy (Richlan et al., 2013; Tamboer et al., 2016; Vandermosten et al., 2012). Brain imaging data obtained through functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), electroencephalography (EEG), and positron emission tomography (PET) provide objective evidence for machine learning classification of dyslexia (Da Silva et al., 2021; Ortiz et al., 2020; Thiede et al., 2020). Most fMRI data focus on brain regions related to language and lexical decision-making, exploring functional differences in brain activation during reading tasks (Chimeno et al., 2014). Zahia et al. (2020) collected fMRI structural images from 55 Spanish children aged 9–12 years (including 18 with dyslexia) and used deep learning algorithms for automatic identification of children with dyslexia. Da Silva et al. (2021) collected high-resolution T1-w images from 32 Portuguese-speaking Brazilian children aged 8–12 years (16 with dyslexia) and used deep learning algorithms to classify important regions for visual representation. EEG can record high temporal resolution brain signals while preserving temporal and frequency domains, reflecting brain functional states during children' s cognitive processing and providing effective features for early diagnosis of dyslexia. Researchers mostly focus on five EEG frequency bands: delta, theta, alpha, beta, and gamma (Ortiz et al., 2020), exploring brain connectivity through phase synchronization between EEG channels and then extracting discriminative features for dyslexia identification. Zainuddin et al. (2019) collected EEG signals from 10 children with moderate dyslexia, 10 with severe dyslexia, and 10 control children aged 7–12 years, and used K-Nearest Neighbors (KNN) and Extreme Learning Machine (ELM) to screen for dyslexia through writing tasks. Formoso et al. (2021) collected EEG signals from 48 children aged 7–8 years (16 with dyslexia), measured phase synchronization between channels to reveal brain functional networks activated during auditory processing, and then used vector quantization unsupervised learning combined with Bayesian algorithms to extract discriminative features for dyslexia identification. In Chinese dyslexia research, Cui et al. (2016) used structural magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) to collect 3D T1-w images (MPRAGE) from 61 school-age children aged 10–14 years (28 with dyslexia) and used machine learning algorithms to distinguish children with dyslexia from typically developing children.

Nowadays, an increasing number of researchers are not limited to single-modality data collection. They integrate data from scales, behaviors, imaging, and other sources to improve the accuracy of dyslexia and biomarker detection. Among the 25 included articles, the proportions of data types used are as follows: standardized psychoeducational testing and eye movement features each account for 28%, followed by game testing data at 16%, MRI data at 12%, and fMRI and EEG data each at 8%.

Data Preprocessing and Feature Extraction

The main purpose of data preprocessing is to enable algorithms to extract the most relevant and interpretable features from the dataset (Usman et al., 2021). For traditional machine learning methods, the first step of preprocessing is to convert data into quantitative (numerical) or qualitative (text categories) formats. Some scale or behavioral data use manual preprocessing methods, such as having experts label the data as non-dyslexia and dyslexia groups (Khan et al., 2018). In brain imaging research, data directly collected by researchers are usually high-dimensional multivariate data. Taking 64-channel EEG data as an example, even calculating one indicator on a single channel yields at least 64 feature values. When the number of feature values exceeds the sample size, machine learning can easily cause overfitting and reduce training and prediction speed. Therefore, it is necessary to reduce high-dimensional features to low-dimensional features to accelerate subsequent machine learning classification and training. For example, EEG signal preprocessing often uses Principal Component Analysis (PCA) to reduce data dimensions by removing secondary components, achieving dimensionality reduction (Ahire et al., 2022). Additionally, brain imaging data can be preprocessed using different software toolkits. For instance, fMRI images can use MATLAB' s SPM toolbox to automatically segment different tissue types, improving the comparability of pixels and voxels during data preprocessing (Zahia et al., 2020); or use the FreeSurfer image analysis suite to extract reliable cortical volume and thickness (Plonski et al., 2017).

The next step after preprocessing is feature selection and extraction, aiming to generate the most relevant and informative features from raw features (Abd Rahman et al., 2020) to form datasets required for classification. Features that can be selected from standardized psychoeducational tests generally include questionnaire/cognitive test scores, writing data, and phonological data. Eye movement data typically use statistical measures, discrete-based and velocity-based algorithms for feature selection, and Principal Component Analysis (PCA) for feature extraction. Feature extraction from fMRI data involves extracting brain cortical attribute features from brain tissue properties, with common features including cortical thickness, volume information, anisotropy scores, and activation patterns. In EEG data, Fourier transform and wavelet transform are generally used to extract temporal and frequency information from signals. Additionally, some new feature extraction methods have recently emerged, such as deep learning, which automatically extracts features from data by constructing different network structures, demonstrating good robustness and strong high-dimensional data processing capabilities. For example, in Ileri et al.' s (2022) study, Convolutional Neural Networks (CNN) provided automatic classification of segmented EOG signals without manual feature extraction. Table 2 summarizes the feature types in the 25 included articles.

Model Training and Classification

After feature extraction and selection, researchers can use machine learning for model training and classification. Machine learning is roughly divided into two types: unsupervised learning and supervised learning. Unsupervised learning is used to find patterns in input data without using any output data, while supervised learning is mainly used to predict future events (Russell & Norvig, 2010). In supervised learning, the purpose of training models is to learn ideal values for all weights and biases from labeled data. Recent studies generally use supervised learning algorithms to explore classification problems between individuals with dyslexia and typically developing populations. Common algorithms include Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Linear Regression (Linear-R), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naïve Bayes, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). Before training models, the entire dataset is usually divided into a testing set and a training set. Most existing machine learning studies on dyslexia use K-fold cross-validation, dividing the dataset into K equal parts, with K−1 parts used as the training set and 1 part as the testing set, using the average of K test results as the final performance evaluation. For example, Plonski et al. (2017) used 10-fold cross-validation, while AlGhamdi (2022) used 5-fold and 20-fold cross-validation. When sample sizes are small, some researchers also choose Leave-one-out cross-validation, a special form of K-fold cross-validation, to build models and evaluate classification results (Cui et al., 2016; Asvestopoulou et al., 2019).

Dyslexia identification is essentially a binary classification problem, i.e., distinguishing whether a user has dyslexia. The principle of SVM is to solve linear binary classification problems and can provide good performance for ultra-high-dimensional and sparse feature space data. Therefore, SVM has become the most widely used algorithm in dyslexia research. Shamir et al. (2019) used SVM to classify reading assessment data based on standardized tests and the Zippy 6 screening test, achieving 75% specificity and sensitivity. Prabha and Bhargavi (2019) proposed a Particle Swarm Optimization SVM (SVM-PSO) model to extract biomarkers of dyslexia from eye movement features. Compared with the Linear SVM model, this model achieved a prediction accuracy of 95%. Additionally, researchers combine SVM with other algorithms to identify children with dyslexia. For example, using RF to select the most important features as input for SVM achieved 89.7% accuracy and 84.8% recall (Raatikainen et al., 2021).

For mining complex patterns in big data, the emergence of deep learning has solved this challenge. Deep learning algorithms have more hierarchical structures, thus providing stronger modeling or abstract representation capabilities for phenomena and enabling simulation of more complex models. ANN is the foundation of deep learning, simulating brain neural network structure and function, and is particularly effective in uncertain recognition such as speech and image recognition (Lucchiari et al., 2014). Ahmad et al. (2022) used an ANN model to classify comprehensive game data, achieving 95% detection accuracy. With the development of neural networks, deep learning has evolved from shallow ANN. Among them, CNN is the most popular deep learning model for dyslexia classification (Usman et al., 2021). Da Silva et al. (2021) selected two network visualization techniques to learn high-level features in CNN input layers, accurately classifying brain states of children with dyslexia from brain imaging data (fMRI) alone, achieving 94.8% accuracy. Moreover, researchers have proposed a new CNN method based on EOG signals to identify dyslexia. Latifoglu et al. (2021) screened and tracked children with dyslexia through skipping lines and return eye movements during reading. They used a two-dimensional Convolutional Neural Network (2D-CNN) model to classify these spectrogram images, obtaining 99% accuracy, 100% sensitivity, 98.18% specificity, and 98.94% F-score. Ileri et al. (2022) recorded EOG signals from horizontal and vertical channels and applied a one-dimensional Convolutional Neural Network (1D-CNN) to classify signals from these two channels, with accuracies of 98.70% and 80.94%, respectively.

In fact, no single algorithm can be the best for all datasets. Algorithm selection is influenced by multiple factors including problem nature, dataset characteristics and quantity, data format, training and prediction time, and storage requirements. Therefore, researchers increasingly tend to find optimal algorithms among multiple classifiers to obtain optimal parameters. The overall trend in research is moving from single traditional machine learning algorithms to deep learning algorithms (Deep Neural Network, DNN) and comparing multiple different types of algorithms. Tolami et al. (2021) used linguistic features as classification indicators to build NB, KNN, SVM, LR, DT, and MLP models, with the deep learning MLP algorithm achieving the highest classification accuracy of

93.33%. In Chinese dyslexia research, Lee et al. (2022) used Chinese character features and personal characteristics as classification indicators, applying NB, KNN, SVM, DT, LR, and ANN to build models separately, finding that all six algorithms could distinguish children with dyslexia from typically developing children, with SVM achieving the highest accuracy of 80.0%. Based on the 25 included articles, the frequency of algorithm usage is as follows: SVM accounts for 27.3%, followed by KNN and LR each at 12.7%, CNN and RF each at 9.1%, NB and DT each at 5.5%, ANN at 3.6%, and BT, GA-BPNN, GB, ELM, LSTM, MLP, RUSBoosted, and ET each at 1.8%.

Performance Evaluation

The outcome variable for dyslexia is binary classification. For evaluating binary classification results, the first step is to categorize different sample classifications into four types: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Then, evaluation metrics are defined based on data classification. For binary classification problems, the most commonly used evaluation metric is overall accuracy. However, accuracy only reflects the proportion of correctly identified positive and negative samples out of the total sample. In practical applications, especially in clinical screening, the problem of disproportionately large positive-to-negative sample ratios often occurs. For these imbalanced data classification problems, multiple metrics are needed to evaluate classification model performance. Other commonly used evaluation metrics include Sensitivity, Specificity, Precision, Recall, F1 score, Kappa coefficient, AUROC curve and P-R curve, Positive Predictive Value (PV+), and Negative Predictive Value (PV−). Based on the 25 included articles, the performance of machine learning models (using accuracy as reference) is summarized as follows: standardized psychoeducational testing ranges from 68%-94.1%; eye-tracking testing ranges from 81.25%-99%; game testing ranges from 74%-99.9%; brain imaging data captured by EEG ranges from 89%-90%; brain imaging data captured by fMRI ranges from 65%-94.83%.

Research Applications of Machine Learning-Based Early Screening for Dyslexia

Identifying Predictive Factors of Dyslexia

The primary advantage of machine learning lies in its model flexibility, i.e., its ability to fit quite complex multiplicative interactions or nonlinear relationships, thereby producing remarkable predictive accuracy. This statistical effect is particularly prominent in predictive problems such as predicting suicide risk among Weibo users and depression-susceptible populations. Machine learning helps identify predictive factors of dyslexia, enabling effective detection of children at risk for dyslexia for timely intervention, thus reducing the possibility of reading failure after children become literate and even into adulthood. For example, Tamboer et al. (2016) used MRI technology to build an SVM prediction model, finding that the most reliable classification voxels were located in the Left Occipital Fusiform Gyrus (LOFG), Right Occipital Fusiform Gyrus (ROFG), and Left Inferior Parietal Lobule (LIPL), with sensitivity reaching 82% and specificity reaching 78%. Therefore, these brain regions are potential biomarkers for dyslexia types related to reading, spelling, and phonology. Prabha and Bhargavi (2019) built a prediction model based on SVM-PSO, showing that eye movement features such as average fixation count, average fixation duration, average saccade time, total saccade count, and average fixation count can serve as risk prediction indicators for dyslexia in children. The model achieved a prediction accuracy of up to 95%. Formoso et al. (2021) collected children' s EEG signals and used phase synchronization between EEG channels to represent connectivity between brain regions. The results showed that discrimination ability was strongest in alpha and beta bands under 16 Hz stimulation, with AUC values reaching 0.95. In Chinese dyslexia research, Wang and Bi (2022) built a Chinese dyslexia prediction model based on a Genetic Algorithm-Back Propagation Neural Network (GA-BPNN), finding that reading accuracy was the strongest factor in predicting Chinese reading difficulties, with phonological awareness, pseudo-character accuracy, morphological awareness, reading fluency, rapid digit naming, and non-character reaction time also making important contributions to prediction. Previous studies suggested that dyslexia was related to performance on virtual Hebb-Williams maze tasks, but classification of children with dyslexia through real-time observation of task performance was not feasible (Gabel et al., 2021). Yu et al. (2022) pioneered the prediction of reading ability based on maze task performance using machine learning algorithms, achieving real-time feedback on reading risk in percentage form. They used virtual maze data, reading level, and personal characteristics as indicators, and the model built using the RUSBoosted Trees algorithm achieved classification accuracy above 70%. With the development of computer networks and electronic products, online assessment platforms or applications for preliminary dyslexia screening have gradually become popular. Researchers developed an application called Dystech that analyzes audio signals in real-world environments, finding that machine learning algorithms related to appropriate audio signal processing can extract patterns inaccessible to human experts, with screening accuracy reaching 80% (Radford et al., 2021). Rauschenberger et al. (2022) designed a language-independent game called MusVis, using RF and Extra Trees (ET) to train on collected game data, achieving classification accuracies of 74% and 69% for German and Spanish dyslexia, respectively.

Assisting in Screening Children with Dyslexia

Traditional screening for dyslexia in children is mainly conducted through professional medical institutions and research organizations, primarily using standardized psychoeducational testing combined with children' s external behaviors and parent reports. Although eye-tracking and brain imaging technologies have provided more objective technical support for dyslexia screening in recent years, it is almost impossible to use these complex measurement tools for large-scale identification of every individual with dyslexia. Moreover, these measurement tools have drawbacks such as high cost, time consumption, poor accessibility, and narrow access channels. Therefore, machine learning is used to assist clinical screening and automated identification, which can not only incorporate a large number of objective classification indicators to improve accuracy but also be convenient, fast, and reduce waiting costs. Asvestopoulou et al. (2019) developed a dyslexia screening tool called DysLexML, which recorded children' s fixation points during silent reading through eye-tracking, applied LSVM to build a screening model, and achieved 97% accuracy. Notably, DysLexML maintains good robustness and accuracy even in the presence of noise, enabling it to cover larger and more diverse populations and laying the foundation for developing inexpensive eye-tracking screening tools in less controlled, large-scale environments (such as kindergartens).

Predicting High-Risk Children with Dyslexia

Dyslexia is widely believed to have a genetic basis (approximately 70%) (Su et al., 2012; Galaburda et al., 2006). Several candidate genes for dyslexia, such as ROBO1, DCDC2, DYX1C1, and KIAA0319, have been confirmed to play important roles in children' s brain development (Galaburda et al., 2006). Even if some children at familial risk do not develop dyslexia, their performance on tasks such as spelling, non-word reading, and reading comprehension is still worse than that of typically developing children (Lyytinen et al., 2005). Early prediction of high-risk children with dyslexia makes early prevention and intervention possible. This early prediction function can be achieved by training machine learning prediction models. Skeide et al. (2016) believed that the neuroplasticity of brain regions important for literacy might be regulated by genetic variants, thereby pre-limiting children' s reading and writing abilities. They collected gray/white matter volumes and literacy-related gene information from children in grades 4–8 and kindergarten–1st grade, used LSVM to build dyslexia prediction models, achieving accuracies of 73% and 75%, respectively. Chen et al. (2017) analyzed group-level differences in total vocabulary and vocabulary categories in children aged 17–35 months based on the vocabulary development inventory, using SVM to classify children at familial risk for dyslexia and typically developing children. The results showed that the risk prediction model achieved 68% accuracy, 70% sensitivity, and 67% specificity, indicating that machine learning can distinguish between children at familial risk for dyslexia and typically developing children at pre-literacy early stages. At this stage, children at familial risk for dyslexia already show functional and structural changes in temporoparietal and occipitotemporal regions, similar to changes observed in individuals with dyslexia (Hosseini et al., 2013; Kraft et al., 2015).

Advantages and Limitations of Applying Machine Learning to Early Screening of Dyslexia in Children

Advantages

In recent years, the application of machine learning in dyslexia and biomarker detection has increasingly gained researchers' favor, with advantages mainly reflected in three aspects. First, machine learning can identify complex nonlinear relationships between variables, providing more accurate screening and developmental prediction for dyslexia. Dyslexia is the result of interactions among multiple factors (Morris et al., 1998). Traditional statistical methods (such as logistic regression) identify single or multiple predictors that have weak predictive power or cannot reflect interactions between factors, failing to fully mine data. Machine learning is more suitable for analyzing complex problems like dyslexia. Taking Back Propagation Neural Networks (BPNN) as an example, BPNN, as a simulation of human brain working mechanisms, can not only handle fuzzy mapping relationships but also identify complex nonlinear relationships between variables (Lyu & Zhang, 2019). Whether in alphabetic languages or Chinese, BPNN models can effectively screen children with dyslexia by collecting cognitive test or phonological test data related to reading. Second, compared with human identification methods, machine learning avoids the influence of subjective understanding biases on the one hand, and can automate repetitive tasks, analyze more data in less time, and achieve higher accuracy and repeatability than manual algorithms on the other hand. Third, machine learning has powerful high-dimensional data processing capabilities, can extract additional, key discriminative information from brain imaging data, and detect subtle imaging abnormalities that may reflect important pathophysiological mechanisms but are unobservable to the naked eye. Differences in brain function and development are early signs of dyslexia risk. As age increases, rapid synapse formation changes children' s brain activation patterns, but brain structure remains unchanged from childhood to adulthood unless severely injured or critically ill. Therefore, high-dimensional brain imaging data can provide more accurate results for dyslexia identification. For example, Da Silva et al. (2021) achieved 94.8% accurate classification of children with developmental dyslexia from brain imaging, while using feature visualization techniques (CAM) and gradient-based feature visualization techniques (Grad-CAM) in convolutional neural network layers responsible for learning high-level features, providing visualization images of brain regions related to strategic control and attention processes during reading in children with dyslexia and typically developing children. This prediction of brain states at the slice level and subsequent generation of finer-grained feature information related to classification can improve model interpretability.

Limitations and Implications

First, there is a lack of research on participant groups during the optimal intervention period. Dyslexia is heritable, with 68% of monozygotic twins and up to 40%-60% of first-degree relatives both having dyslexia (Vogler et al., 1985).

Several candidate genes for dyslexia, such as ROBO1, DCDC2, DYX1C1, and KIAA0319, have been confirmed to play important roles in children' s brain development (Galaburda et al., 2006). The early stage of enhanced brain plasticity in children occurs during kindergarten to 1st grade, which is the optimal period for early intervention in dyslexia (Fox et al., 2010). Studies have found that effective intervention for high-risk children with dyslexia in kindergarten and 1st grade (average effect size of 0.31–0.84) is far more effective than for high-risk children in 2nd and 3rd grade (average effect size of 0.23–0.27) (Wanzek & Vaughn, 2007). Therefore, accurate early identification of children with dyslexia, especially those at familial risk, before the optimal intervention period is crucial. Unfortunately, based on the reviewed literature, only one machine learning study on dyslexia had participants in a younger age range (17–35 months), while participants in other studies were aged 6–17 years. Research on the pre-literacy stage (ages 3–7) is almost blank. Future research needs to widely collect data on children before they become literate, focusing on genetic and environmental risk factors, establishing multimodal databases, and using machine learning' s powerful classification functions to screen children with dyslexia and establish relatively stable behavioral/biomarkers, ultimately building a convenient, fast, accurate, and scientific early screening system.

Second, data quality in machine learning research is uneven, collection standards are not unified, and sample sizes are insufficient. Database collection for dyslexia based on machine learning shows a trend from single databases to multiple databases and from single-modality to multimodal heterogeneous data. Since databases come from different laboratories and populations, collection standards have not been unified, data distribution characteristics differ, and large amounts of data are incompatible and structurally complex. Therefore, it is necessary to establish standardized heterogeneous databases to improve computational power required by models and avoid resource waste. The phenomenon of non-unified collection standards is particularly prevalent in brain imaging databases for children with dyslexia. On the one hand, imaging equipment models and parameters that are not unified can affect data quality to some extent. Due to the lack of authoritative and fixed standards, the reproducibility of brain imaging cannot be consistently recognized. On the other hand, classification accuracy largely depends on sample size. Compared with questionnaire and behavioral data, brain imaging data in publicly/non-publicly available dyslexia-related databases are relatively small. Models trained with small samples can easily fall into overfitting of small samples and underfitting of target tasks. To address these issues, first, international cooperation can establish a unified standard platform for data collection and sharing to achieve reproducible application of brain imaging data. Second, remedies can be made by increasing training data, reducing the space models need to search, and optimizing the process of searching for optimal models.

Third, it cannot yet achieve high translational power in clinical practice and gain wider use. Although numerous studies have found that changes detected in brain morphology, eye movements, and normal auditory systems can serve as neurobiomarkers for dyslexia identification, traditional outpatient screening for dyslexia still relies primarily on standardized psychoeducational testing (behavioral markers). This is because standardized psychoeducational testing has advantages such as representative test content, high standardization, high reliability and validity, and convenient and economical use. Machine learning currently does not yet possess the necessary conditions for clinical translation. First, training data lack representativeness. Experimental data usually seek estimation of typical samples under the premise of controlling experimental irrelevant variables, but if our goal is to create generalizable prediction algorithms, samples need to include a large number of individualized cases in real life. Second, machine learning models have low interpretability and transparency. There exists an "algorithm black box" where there is an unobservable space between input data and output answers, and even developers cannot fully understand the specific details of algorithm operation. Third, machine learning performance metrics lack clinical applicability; for example, F1 scores and recall may not be applicable in clinical environments and are difficult for clinicians and researchers to understand. Finally, there is insufficient validation research on intervention methods. The ultimate goal of early screening for dyslexia is to provide effective early intervention for children. However, only two studies currently link machine learning with dyslexia intervention (Atkar & Jayaraju, 2021; Oliaee et al., 2022). Previous EEG research on dyslexia based on machine learning mainly distinguished children with dyslexia from typically developing children through group differences in EEG (especially power in individual frequency bands). Oliaee et al. (2022) innovatively classified children with dyslexia before and after specific treatment plans, providing a new method for evaluating the effectiveness of dyslexia treatment programs. They used PCA and Sequential Floating Forward Selection (SFFS) algorithms to extract optimal feature subsets from recorded EEG signals and found that EEG signals of children with dyslexia showed significant changes in spectral and phase correlation features in different regions before and after receiving Transcranial Direct Current Stimulation (tDCS) treatment and cognitive training, with classification accuracy of the most discriminative feature subset reaching 92%. Atkar and Jayaraju (2021) used a deep learning-unsupervised learning Generative Adversarial Network (GAN) model to generate raw audio data of two- or three-letter Hindi words, establishing a MelGAN system using generated data. The system accelerates recovery by having children with dyslexia repeat correct pronunciations of words, aiming to provide teachers with an effective auxiliary tool. Although using machine learning to evaluate intervention effects and assist in creating intervention tools has begun to enter researchers' 视野, their practicality and verifiability still need further improvement.

Finally, participant data security and privacy protection are threatened. Machine learning model training requires large amounts of data, but databases often contain large amounts of private data such as personal identity information and family information. How to prevent privacy leakage at low cost and high efficiency becomes extremely important. Usman and Muniyandi (2020) constructed a method for secure classification of dyslexia based on CNN models and Residue Number Systems (RNS). They developed a pixel-bitstream encoder using special modules of RNS, encrypted the 7-bit binary values of each pixel in MRI data in training and test sets before using cascaded CNN for classification, and then used encrypted test datasets to predict children with dyslexia. Additionally, establishing informed consent and ethical review before data sharing also helps prevent potential data misuse.

In summary, machine learning has been gradually applied to early screening of dyslexia. Data collection methods have shifted from single-modality to multimodal heterogeneous data, and multiple models are used to verify optimal classification effects, with classification performance ranging from 67%–100%. Currently, the most used machine learning algorithm is SVM, and future deep learning is expected to achieve higher classification performance for dyslexia. In application, machine learning research on dyslexia still has problems such as small sample sizes, low clinical practice rates, insufficient multimodal data integration, and classification performance needing improvement. Moreover, there is a lack of research on child groups during the optimal intervention period, and true early screening of children with dyslexia has not been achieved. Future research should first focus on risk identification in preschool children, concentrating on marker research for early screening of dyslexia. Second, since dyslexia is not specific to any region, language, or culture, language-independent data collection methods need to be developed to establish unified standard dyslexia databases. Finally, future research needs to collect data from multiple sources (such as scales, behaviors, brain imaging, etc.), hybridize multiple models, and consider multimodal deep learning frameworks to improve machine learning predictive power, continuously optimize constructed dyslexia screening models, and ultimately achieve widespread use in clinical practice.

Note: References marked with were included in the systematic analysis.*

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.