IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2021, accepted February 22, 2021, date of publication February 26, 2021, date of current version March 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062709

Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges

OPEYEMI LATEEF USMAN[1,4], RAVIE CHANDREN MUNIYANDI[1], KHAIRUDDIN OMAR[2], AND MAZLYFARINA MOHAMAD[3]

[1]Research Centre for Cyber Security, Faculty of Information Science and Technology, University Kebangsaan Malaysia, Bangi 43600, Malaysia
[2]Research Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, University Kebangsaan Malaysia, Bangi 43600, Malaysia
[3]Center for Diagnostic, Therapeutic and Investigative Studies, Faculty of Health Sciences, University Kebangsaan Malaysia, Kuala Lumpur 50300, Malaysia
[4]Department of Computer Science, Tai Solarin University of Education, Ogun State, Nigeria

Corresponding author: Opeyemi Lateef Usman (p99943@siswa.ukm.edu.my)

This work was supported by the Universiti Kebangsaan Malaysia under Grant GGP-2019-023 and Grant GP-2019-K005539.

ABSTRACT Dyslexia is a neurological disorder that is characterized by imprecise comprehension of words and generally poor reading performance. It affects a significant population of school-age children, with more occurrences in males, thus, putting them at risk of poor academic performance and low self-esteem for a lifetime. The long-term hope is to have a dyslexia diagnostic method that is informed by neural-biomarkers. In this regard, large numbers of machine learning methods and, more recently, deep learning methods have been implemented across various types of dataset with the above-chance classification accuracy. However, attainment of clinical acceptability of these state-of-the-art methods is bedeviled by certain challenges including lack of biologically-interpretable biomarkers, privacy of dataset and classifiers, hyper-parameter selection/optimization, and overfitting problem among others. This review paper critically analyzes recent machine learning methods for detecting dyslexia and its biomarkers and discusses challenges that require proper attentions from the users of deep learning methods in order to enable them to attain clinically relevance and acceptable level. The review is conducted within the premise of implementation and experimental outcomes for each of the 22 selected articles using the Preferred Reporting Items for Systematic review and Meta-Analyses (PRISMA) protocol, with a view to outlining some critical challenges for achieving high accuracy and reliability of the state-of-the-art machine learning methods. As an evidence-based protocol for reporting in systematic reviews and meta-analyses, PRISMA helps to ensure clarity and transparency of this paper by showing a four-phase flow diagram of the selection process for articles used in this review. It is therefore, envisaged that higher classification performance of clinical relevance can be achieved using deep learning models for dyslexia and its biomarkers by addressing identified potential challenges.

INDEX TERMS Dyslexia detection, biomarkers, feature extraction, classification algorithms, machine learning, deep learning.

ACRONYMS AND DEFINITIONS

3D	Three-Dimensional
AD	Axial Diffusivity
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area under ROC curve
BET	Brain Extraction Tool
BOF	Bag of Feature
BOLD	Blood Oxygenation Level Dependent
CAT12	Computational Anatomy Toolbox version 12
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CTOPP	Comprehensive Test of Phonological Processing
CV	Cross-Validation
DARTEL	Diffeomorphic Anatomical Registration using Exponentiated Lie algebra
dMRI	Diffusion MRI
DNN	Deep Neural Network

The associate editor coordinating the review of this manuscript and approving it for publication was Rupak Kharel.

DT	Decision Tree
DTI	Diffusion Tensor Imaging
EEG	Electroencephalogram
EOG	Electro-Oculogram
FA	Fractional Anisotropy
FAST4	FMRIB's Automated Segmentation Tool v4
FDT	FMRIB Diffusion Toolbox
FEAT	FMRI Expert Analysis Tool
fMRI	Functional MRI
FMRIB	Functional Magnetic Resonance Imaging of the Brain
GM	Gray Matter
IEEE	Institute of Electrical and Electronic Engineers
KNN	K-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
Latent-DA	Latent Dirichlet Allocation
LDA	Linear Discriminant Analysis
LIBSVM	SVM Library of MATLAB
Linear-R	Linear Regression
LIPL	Left Inferior Parietal Lobule
LOFG	Left Occipital Fusiform Gyri
LOOCV	Leave-one-out Cross-Validation
LR	Logistic Regression
LSVM	Linear SVM
MATLAB	Matrix Laboratory
MD	Mean Diffusivity
MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
NPV	Negative Predicted Value
PANDA	Pipeline for Analyzing braiN Difussion imAges
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PPV	Positive Predicted Value
PPVT-III	Peabody Picture Vocabulary Test: Third Edition
PRISMA	Preferred Reporting Items for Systematic review and Meta-Analyses
RD	Radial Diffusivity
RF	Random Forest
ROC	Receiver Operating Characteristics
ROFG	Right Occipital Fusiform Gyri
SPM12	Statistical Parametric Mapping version 12
SURF	Speeded-Up Robust Feature
SVM	Support Vector Machine
SVM-RFE	Recursive Feature Elimination with SVM
T1-w/T2-w	T1 weighted image/T2 weighted image
TBSS	Tract-Based Spatial Statistics
TNR	True Negative Rate
TPR	True Positive Rate
VBM	Voxel-Based Morphometry
VOG	Video-Oculography
WART-R,3,4&5	Wide Range Achievement Test: Revision R, 3, 4, & 5
WASI, II,III&IV	Wechsler Abbreviated Scale of Intelligence, 2nd, 3rd, & 4th Editions
WJ-III	Woodcock-Johnson Third Edition
WM	White Matter
WoS	Web of Science

I. INTRODUCTION

Dyslexia, an extremely complicated neuro-developmental brain disorders, is attracting great attentions in recent time among researchers in modern Neuroscience [1]. It is a neurological disorder [1], [2] characterized by sluggish and inaccurate word comprehension and phonological impairment, affecting about 5-17% [3] of the general population in most languages and cultures. In most cases, the onset of this condition progresses to adolescence from childhood and can be counterproductive to academic performance [4]. Dyslexia may also have significant negative influence on children's self-esteem and self-perception development [4], [5]. Students with dyslexia experience high levels of bullying in academic settings as well as feelings of alienation and exclusion [1], [5], [6]. Adapting from the definition of International Dyslexia Association [7] and British Dyslexia Association [8], certain children with dyslexia exhibit deficits in higher-order processing or executive control systems. Apart from this, they also have visual attention span disorders that significantly affect their ability to read. They may also have memorizing issues as well as letter recognition difficulties [8]. Dyslexic children and adults, therefore, display major deficiencies in phonological processing, verbal working memory and speed of communication [8]. Dyslexia can be developmental or acquired depending on whether the victim's brain encountered deficit during developmental stages or sustained serious injuries such as stroke [9].

In Figure 1, brain is the principal neurological source of dyslexia which manifests as cognitive and behavioural deficits in the other three interacting elements. From this figure, it can be deduced that existing dyslexia interventions are approached from three domains namely: behavioural, cognitive, and biological domains respectively [10]. While phonological disorder otherwise, known as language disorder manifests when the affected individual is unable to develop association between letters (grapheme) and sound (phoneme), leading to reading and spelling impairments [8], visual deficit manifests when there is dysfunction in the magnocellular and parvocellular subsystems of the brain. Studies by Vidyasagar [11] and Danielli *et al.* [12] have shown that abnormality in the magnocellular pathways and cells is associated with various visual impairments, including dyslexia. Auditory deficit, otherwise called hearing impairment, manifests when there is impairment in central auditory

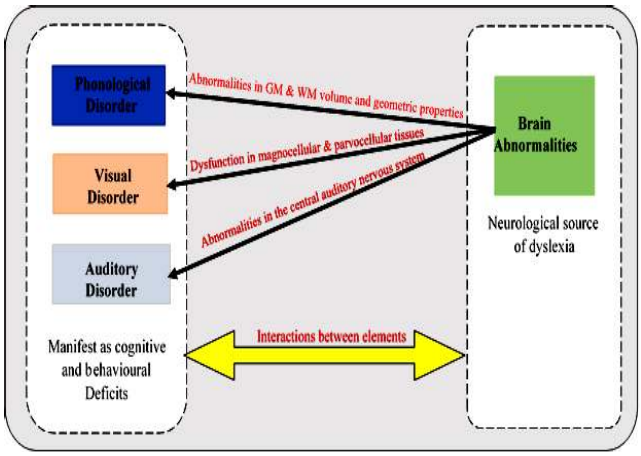


FIGURE 1. **Interaction between the elements of dyslexia: brain abnormalities, phonological disorder, visual deficit, and auditory dysfunction. Dyslexia occurs due to abnormalities and dysfunctionalities of certain brain subsystems such as phonological, visual and auditory as child develops right from birth. It manifests as both cognitive and behaviour deficits when a child attains school age.**

nervous system responsible for hearing leading to difficulty in sensing, processing, and understanding sound signals. Recent study by Kuhl *et al.* [13] has shown that dyslexia can be predicted through auditory cortex gyrification and abnormal neural connectivity within the speech processing system before a child begins to learn how to read and write. Summarily, these four elements constitute the basis upon which theories of developmental dyslexia are formulated.

Researchers have proposed several methods for detecting and classifying dyslexia and its markers using datasets obtained from several multiple sources. These sources include standardized psycho-educational tests, web-based/mobile-based games, eye movement tracking, MRI and EEG scans, MEG scans, PET scans, video and images captured during cognitive/phonological tasks. Standardized psycho-educational tests are used to measure study-participants' achievement, intelligence, phonological processing, reading skill and vocabulary development. In other words, they are used to assess and quantify students' cognitive abilities. They include Wechsler Abbreviated Scale of Intelligence (WASI, II,III&IV) [14], Wide Range Achievement Test: Revision R,3&4 (WRAT-R,3&4) [15], Woodcock-Johnson III (WJ-III) [16], Comprehensive Test of Phonological Processing (CTOPP) [17], [18], and Peabody Picture Vocabulary Test: Third Edition (PPVT-III) [19]. Web-based e-learning systems have been developed by [20]–[26] to assist schoolchildren suffering from dyslexia in reading and writing across various regions, languages and cultures. Similarly, mobile-based games such as Dyseggxia and GraphoGame have also been developed by researchers to assist dyslexic individuals in cognitive and phonological trainings [27]–[31]. For eye movement tracking, statistical measures such as dispersion threshold identification, velocity threshold identification algorithms and eye trackers have been utilized to detect learner's eye movement features during cognitive exercise. These features include fixations and saccades, used by recent studies to classify learners into dyslexic and non-dyslexic [32], [33]. Both MRI and EEG scans are used to generate and detect images showing alterations in learner's brain normal morphology with electrical impulses within it. For instance, different types of MRI scans have been utilized for finding neural-biomarkers of dyslexia in the structure, function, activation and geometric properties of different brain tissues and regions [1]. Such alterations were analyzed with the aid of voxel-based morphometry (VBM) analysis [34]. VBM study utilizes software packages such as BrainImage software and statistical parametric mapping such as Statistical Parametric Mapping and Computational Anatomy Toolboxes (e.g. SMP12 and CAT12) to normalize the stereotactic brain MRI slice to a common study-specific space, using voxel statistics to classify anatomical regions of brain tissue density [1], [34], [35]. By definition, a biomarker is a feature that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to therapeutic intervention. Meanwhile, neuroimaging biomarkers, often referred to as neural-biomarkers, use brain imaging techniques to obtain the morphology, function, micro-environment, metabolism or molecular content of the brain and its lesions. Apart from the fact that the above approach is very expensive, other drawbacks include low coverage (i.e., not suitable for large number of participants), undefined sensitivity and specificity metrics, and more focuses on neurobiological characteristics rather than biomarker definition. Other method for generating data for dyslexia detection are video-oculography (VOG) and electro-oculogram (EOG). Any deviation from normal reading and writing abilities detected during cognitive tasks with the aid of standardized psycho-educational tests is referred to as behavioural marker while alterations detected in brain morphology, eye movement and normal auditory system are called neural-biomarkers or simply, biomarkers.

The usage of traditional machine learning methods and deep learning algorithms for dyslexia and its biomarkers detection is on the increase recently. Models such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Decision Tree (DT), Random Forest (RF), Linear Regression (Linear-R), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naïve Bayes, K-Nearest Neighbor (KNN), and various state-of-the-art Convolutional Neural Networks (CNNs) have been exploited at various places and time to detect and classify dyslexia using data obtained with techniques mentioned above. Support vector machine (SVM) focuses on selecting critical points for a given classification task. Support vectors are the components of dataset related to the hyperplane separation of the two different classes. The Support vector algorithm finds decision function parameters which maximize the margin between the training samples and the class boundary. SVM's learning principle is based on systemic risk minimization [36], which addresses the problem of balancing the complexity of the model against the effectiveness of data fitting. ANNs are central nervous system-inspired machine learning algorithms that depict the

interconnection of well-established nodes, where each node is a computational unit that represents a biological neuron. A wide variety of ANNs architectures are available, with extensive usage areas for classification problems. DTs are a family of methods that use a branching decision model and its possible implications to improve decision-making. The expected outcome for the classification trees is a discrete category (the class). The main drawback of DTs is that they are prone to overfitting by creating complex models that are not properly generalized. RF is composed of a large number of individual de-correlated DTs that serve as an ensemble. Ensemble method is a supervised learning algorithm whose outcome of prediction is obtained by combining multiple basic classifiers into a single classification model. It aims to increase the accuracy and robustness of multiple classifiers over a single classifier [36]. As a meta-estimator that suits multiple DT classifiers on different dataset sub-samples, a RF uses average to boost predictive accuracy in order to control overfitting. Linear-R is the simplest type of regression analysis that estimates the relationship within a linear model between the independent and dependent variables. One method used where the dependent variable is categorical is LR, like group diagnosis. LDA is a classifier used to identify a linear combination of features in statistics, pattern recognition, and machine learning that characterizes or distinguishes two or more classes. It is closely related to analysis of variance (ANOVA), regression, PCA and factor analysis in a way in which linear combinations of variables are identified to better explain the data [37]. Naïve Bayes classifiers are a family of simple probabilistic classifiers or algorithms based on the application of Bayes' theorem. They are among the simplest Bayesian network models with kernel density estimation. Each pair of features classified by this model is independent of each other. The KNN classifier is a non-parametric approach used for classification and regression tasks. In both tasks, the input consists of the k-closest training samples in the feature space while the output depends on the given task: class in the case of the object's classification or property value in the case of regression. Deep neural networks (DNNs) are the most recent machine learning models used in dyslexia classification. Most popular among them is CNN which have undergone several stages of architectural developments over the last two decades. DNNs are derivatives of ANNs, particularly feed forward multilayer perceptron with more than two hidden layers. For detail discussion about derivation of CNN from ANN, see the attached supplementary file with the name **Supplementary1.pdf**. Supplementary1.pdf shows how CNN model uses filters to extract feature maps in each of the convolutional, activation, and pooling (sub-sampling) layers to predict patterns in the input images. Also, discussed in detail are the types of machine learning algorithms, the general use of deep learning in medical imaging, as well as the recent application of deep learning algorithms for neuroimaging analysis.

This study presents a critical review of recent advances in dyslexia detection using state-of-the-art machine learning methods. The review is conducted within the premise of implementation and experimental outcomes for each selected study with a view to highlighting some critical challenges mitigating against the clinical acceptability of these methods. The objective of this paper is therefore, to review the significance, contributions, performance, and limitations of recent studies which employ machine learning methods for dyslexia detection. Specific focus of this review includes data collection, data preparation and preprocessing, feature extraction and selection, training and classification, and performance evaluation. Machine learning is a growing field of research that can build predictive models from specific datasets. This review showed how traditional machine learning and deep learning methods have evolved over the last decade in the field of dyslexia and biomarker detection. By identifying potential challenges of state-of-the-art deep learning methods that are currently gaining attention in the diagnosis of dyslexia and suggesting approaches to addressing them, this will pave the way for them to reach a clinically-relevant and acceptable level, hence the need for this study.

The entire article is broken down into the following structural components: Section 2 discusses the material and methods, on how the review was conducted, Section 3 presents discussion of findings and outline potential challenges of using state-of-the art machine learning models for dyslexia detection. Section 4 concludes the review and provide some future directions.

II. MATERIALS AND METHODS

A. IDENTIFICATION AND SELECTION OF STUDIES FOR THE REVIEW

A critical literature review was performed in this study to define a broad variety of machine learning methods for detecting and classifying dyslexia and its markers in the last one decade up to the year 2020 (2010-2020). The selection process follows the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [38] to discover and identify relevant articles. Suitable articles were first of all, discovered using electronic literature search method from the Google Scholar, Web of Science (WoS), Scopus, IEEE Xplore, PubMed and Science Direct using the combination of queries terms from the search keywords. Keyword searches for suitable articles include: ''classification methods'', ''classification problem'', ''biomarkers prediction'', ''intervention programmes'', ''assistive technologies'', ''machine learning methods'', ''deep learning algorithms'', ''neural networks'', ''deep neural networks'', ''neural-biomarkers'' and ''MRI dataset'' in the domain of dyslexia detection and analysis. Searched results were filtered to sieving out those articles that did not capture such discourse. Screening of the articles was based on the titles, abstracts, and search keyword. The total number of articles collected at the end of the search was 334. After the elimination of duplicate papers, the original total number of articles was reduced to 131. During the screening stage two reviewers were assigned to evaluate

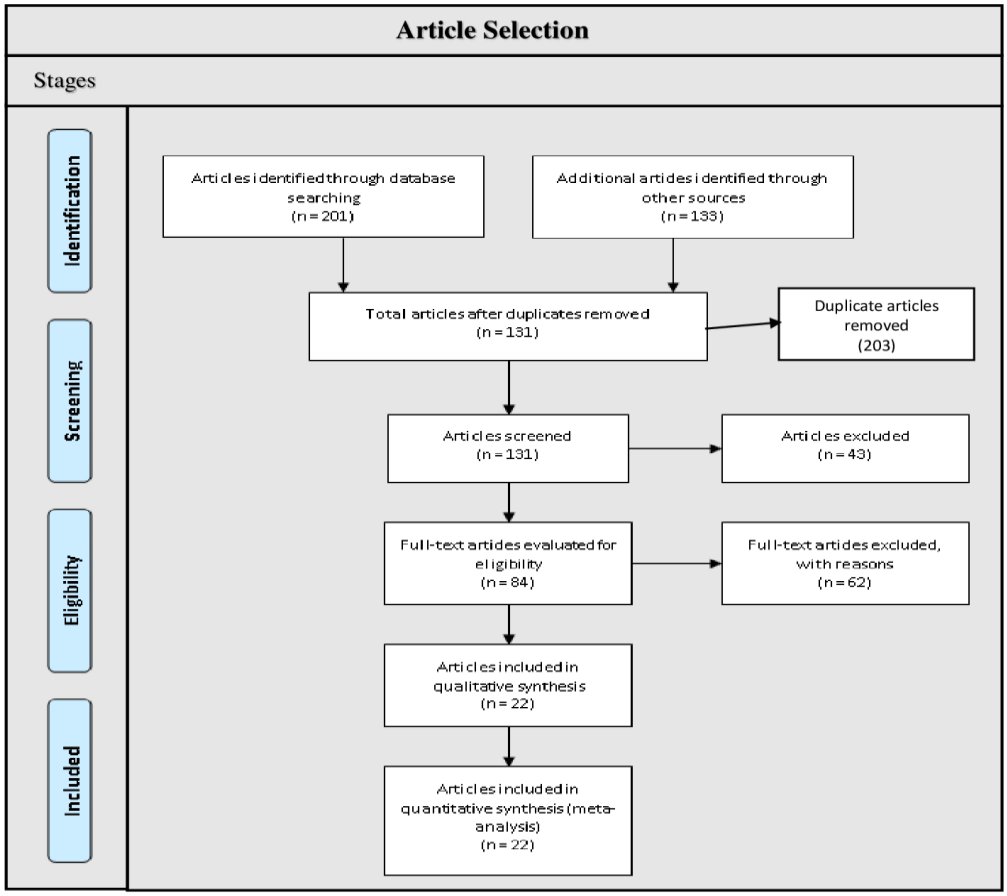


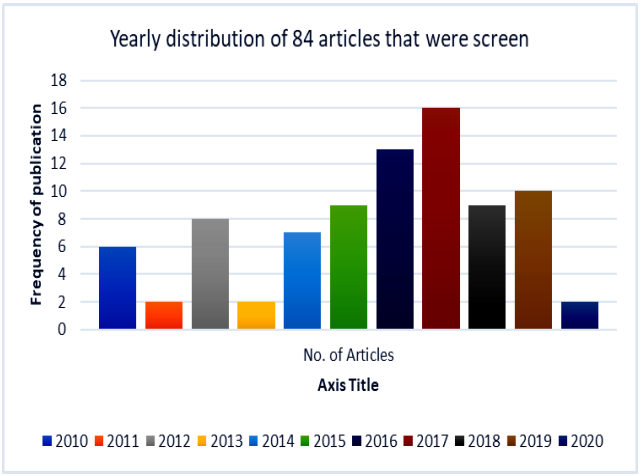
FIGURE 2. Flow diagram of publication selection process for selecting 22 final articles according to PRISMA protocol.

the papers independently. After the exclusion of 43 papers, 84 qualified articles were then distributed to three reviewers for thorough screening exercise. These were then progressed to another screening stage after meeting the criteria. The inclusion criteria were as follows: (1) articles published in English between 2010 and 2020; (2) articles used a particular machine learning method or combinations thereof for the identification of dyslexia; (3) articles used the dyslexia dataset category referred to in the introductory section. Following the screening process, the articles were then sent to three reviewers for further consideration of the eligibility of the articles in accordance with Table 1. The reviewers then compared their analyses until they reached a consensus and agreement. Finally, only 22 articles were selected for critical review. Figure 2 displays a flowchart of the selection process for articles from the early stage of the search to the final number of selected articles.

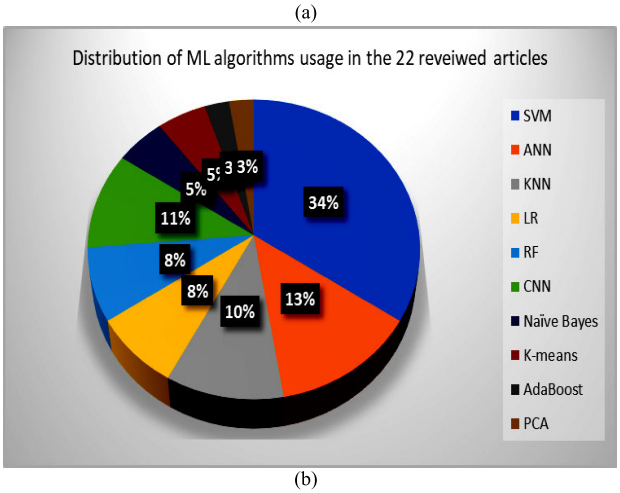
In-line with recent dyslexia detection reviews based on machine learning [9],[39] the specific focus of this review includes data collection, data preparation and pre-processing, extraction and selection of features, training and classification, and performance evaluation. Existing reviews, however, are limited in scope to a wide range of classical machine learning methods and/or specific dataset types (e.g. EEG scans) as in the case of Perera *et al.* [39] without consideration for deep learning methods. Figure 3(a) shows the annual distribution of 84 articles that were thoroughly screened while the pie chart of Figure 3(b) shows the frequency of use of machine learning algorithms in the final 22 reviewed articles. Figure 3(a) shows that most of the studies were published between 2014 and 2019, with 64 articles accounting for 76.2% of the total. Between 2010 and 2013, a total of 18 articles (21.4%) were published with the highest number of articles in the year 2012. Only 2.4% of the total articles were published in the year 2020. In Figure 3(b), 34% of the 22 articles reviewed used the SVM algorithm, followed by

TABLE 1. Criteria of inclusion and exclusion.

Criteria of Inclusion	Criteria of Exclusion
Article in English language (AND) Articles that used a machine learning method (AND) Articles that used combination of machine learning methods	Articles that unrelated to dyslexia detection (AND) Articles that do not fulfil any inclusion criterion



(a)



(b)

FIGURE 3. **(a). Annual distribution of 84 articles thoroughly screened. (b). Distribution of machine learning usage by researchers of final 22 reviewed articles.**

the ANN algorithm (13%). The percentage of studies that used the KNN algorithm is 10.5%, while the LR and RF algorithms are 8% each. CNN algorithm usage accounts for 10.5%, Naïve Bayes and K-mean algorithms account for 5% each, while AdaBoost and PCA account for 5% each.

B. DATA COLLECTION

The first step in dyslexia biomarker detection is acquisition of dyslexia-related datasets using techniques mentioned in the introductory remarks. Accordingly, three categories of datasets have been collected by early studies for machine learning methods analysis. The first category of dataset shows the behavioural symptoms noticeable in dyslexic individuals. They are related to reading, writing, phonological awareness, working memory and facial engagement during learning process. Psychologists have collected this category of datasets using standardized tests such as WASI, WRAT, WJ, CTOP, and PPVT [14]–[17], [19] to identify dyslexic individuals based on their poor performance. Conducting standardize testing is very tedious, time-consuming, and not suitable for larger number of participants due to wider variations in the exhibited behavioural patterns. Studies based on machine learning methods obtain this category of datasets from students' handwritings, typing, reading, test scores and facial engagements across varying languages and cultures. Spoon *et al.* [40] and Spoon *et al.* [41] collected handwriting dataset from 88 subjects (comprises 11 dyslexics, 77 controls) and 100 subjects (comprises 22 dyslexics, 78 controls) respectively for machine learning method analysis. The handwriting samples which were collected with the assistants of participants' parents and class teachers from grades K1-6 schoolchildren of both Spanish and Latin-speaking origins. Both Kohli and Prasad [42] and Khan *et al.* [43] collected students' test scores for training machine learning models during dyslexia identification analysis. The test scores datasets were collected from English-speaking and Malay-speaking students respectively. Images of students' facial engagements have been collected by Abdul Hamid *et al.* [44] for machine learning dyslexia detection. Participants comprises 30 dyslexic individuals (20 males, 10 females) between the age 7-12 years old drawn from pool of students at Dyslexia Association of Malaysia (DAM) Ampang and Bangi areas in Malaysia. Reading test-errors dataset have been collected by Lakretz *et al.* [45] for machine learning analysis from 313 Hebrew-speaking dyslexic subjects, aged 7 to 62 years, at the Language and Brain Laboratory in Tel-Aviv University, Israel.

The second category of dataset used by machine learning studies for dyslexia detection is by using brain imaging modalities to capture distinctive behaviours and activations of the study-participants' brains. These imaging modalities include fMRI, MEG, EEG, PET etc. EEG dataset have been collected by Al-Barhamtoshy and Motaweh [46] from 80 schoolchildren around Makkah and Jeddah in Saudi Arabia for early dyslexia detection. The schoolchildren whose age ranged between 7-18 years old were made to undergo Gibson test of cognitive skills prior the collection of EEG dataset. Perera *et al.* [47], [48] collected the brain activation signal patterns of 32 English-speaking participants (17 dyslexics, 15 controls, 15 males, 17 females) for dyslexia detection study using EEG headset. All the study-participants are right-handed, aged 18+ years with normal/corrected-to-normal vision and normal hearing. Other researchers who have used EEG datasets for dyslexia detection using machine learning methods include Chen Wan Fadzal *et al.* [49], Karim *et al.* [50], and Fried and Breznitz [51], [52]. For MRI datasets, Chimeno *et al.* [37] collected 3D structural images for both fMRI and DTI scans from 52 schoolchildren aged between 9 and 12 years old. The two types of MRI dataset were obtained at different acquisition times (i.e., 5 sec for fMRI and 4.2 sec for DTI). In three independent studies using classical machine learning methods, Plonski *et al.* [53], Jednorog *et al.* [54], and Plonski *et al.* [55] collected a high-resolution T1-weighted (T1-w) images from 236 study participants (130 dyslexics, 106 controls) who are mainly schoolchildren. The samples were drawn from among schoolchildren in three different

countries: French, German and Polish. French samples comprise of 84 subjects (39 dyslexics, 45 controls), German samples comprise of 71 subjects (45 dyslexics, 26 controls), while Polish samples comprise of 81 subjects (46 dyslexics, 35 controls). Cui *et al.* [56] obtained a high-resolution 3D T1-w images (MPRAGE) based on the protocol of the Imaging Centre for Brain Research at Beijing Normal University. The total number of subjects that participated in the study was 61 (28 dyslexics, 33 controls). All subjects were right-handed Mandarin native speakers, who regularly attend schools. Tamboer *et al.* [3] followed a standard acquisition protocol of the Spinoza Centre for Neuroimaging in Amsterdam to collect 3DTI full-brain scans of 49 first-year psychology students who were native Dutch speakers (22 dyslexics, 27 controls, 45 women, 8 left-handed). Specifically, 3D T1-w images Turbo Field Echo Sequences (TFES) were collected by the study for machine learning dyslexia detection. BOLD functional images MRI dataset were collected by Zahia *et al.* [57] from 66 schoolchildren, aged between 9-12 years old for dyslexia biomarker analysis. These children who were native Spanish-speakers, were recruited from schools and the University Hospital of Cruces Paediatric Ophthalmology and Neurology Department in Bilbao, Spain. Usman and Muniyandi [58] collected 45 T1-w images of schoolchildren (19 dyslexics, 26 controls) from Kaggle Database for dyslexia biomarker detection.

The third category of dataset is associated with eye-movement patterns during the cognitive tasks. This category of datasets has been collected by Rello and Ballesteros [59], Rello *et al.* [60], Benfatto *et al.* [61], Rello and Ballestero [59] and Rezvani *et al.* [62] using machines such as eye movement tracker and EEG scans. Other recent machine learning studies combine the method of collecting first category of dataset with method of collecting either category two or category three dataset for optimal dyslexia detection and analysis.

Although collecting second and third categories of dataset can be beneficial in achieving higher accuracy for dyslexia and its biomarkers detection, they are relatively very expensive to collect. Also, their coverage is usually limited to a smaller set of participants [9]. In addition, participants, most especially, kids may behave in an unusual way during the scanning process.

While there are several online platforms for collecting first category dyslexia datasets (e.g. TestingMon.com and Pearson), second and third categories, particularly, for some aspects of neural dyslexia biomarkers are publicly available in major databases and repositories such as Kaggle, OpenNeuro, BishopBlog and The Eckert Laboratory, etc. Table 2 presents various sources and types of datasets for dyslexia study.

C. DATA PREPARATION AND PREPROCESSING

The main aim of data preparation and preprocessing is to enable classifier extract most relevant interpretable features from the used dataset. Effective classification of dyslexia biomarkers using conventional machine learning classifiers is largely dependent on the type of preprocessing tasks performed on the input dataset before they are fed into the classification models. Preprocessing operations are carried out for a variety of purposes, ranging from data normalization, features extraction, tissue segmentation, smoothing (also known as de-noise), alignment with a specific image template and modulation to mention just a few. For traditional machine learning methods such as ANN, SVM, KNN etc., the first preparation and preprocessing step is to convert the dataset into a quantitative (numerical) or qualitative (textual) format. A typical example of this is the conversion of EEG scans dataset into high pass and low pass filters using different wavelet transformation approaches [52]. Others used standard algorithms to filter noise and irrelevant records from dyslexia biomarker dataset as in the case of Al-Barhamtoshy and Motaweh [46] who employed rule-based Fourier transform algorithm to eliminate noise from EEG scans.

Studies such as Khan *et al.* [43] and Kohli and Prasad [42] employed manual data preprocessing methods while vast number of image-based dyslexia biomarker studies used different type of software packages to preprocess their datasets. Following the concept of Bag of Feature (BOF) image classification concept, Abdul Hamid *et al.* [44] implemented Speeded-Up Robust Feature (SURF) descriptor in Computer Vision Toolbox of MATLAB to extract facial engagement features from image dataset. Jednorog *et al.* [54] automatically segmented different tissue types from the T1-w brain images using SPM8 software running on MATLAB 7.11 software. FAST4 software was used by Tamboer *et al.* [3] to segment gray matter (GM) tissue from DTI generated brain image dataset. Recent version of SPM package (SPM12, CAT12) was utilized by Zahia *et al.* [57] and Usman and Muniyadi [58] to improve comparability of pixels and voxels during data preprocessing. This software package was implemented in MATLAB (R2017b) and (R2018b) working environments respectively. Cui *et al.* [56] created a map of white matter (WM) volume in the MNI space using VBM8 toolbox in SPM8 software. In another scenario, Cui *et al.* [63] developed a Pipeline for Analyzing braiN Diffusion imAges (PANDA) MATLAB toolbox for preprocessing dyslexia biomarker-related diffusion (dMRI) dataset. FreeSurfer image analysis software was used by Plonski *et al.* [53], [55] to derive accurate cortical volume and thickness features from the T1-w brain image dataset. The Brain Extraction Tool (BET), the FMRIB Diffusion Toolbox (FDT) and the Tract-Based Spatial Statistics (TBSS) function software [64] were used to obtain fractional anisotropy (FA) values from DTI dataset collected by Chimeno *et al.* [37]. Statistical measures were implemented on a computer system by Prabha and Bhargavi [32], [33] to extract different fixation events and saccades eye movements from the raw eye tracking dataset. Following the concept of DeepWriter proposed by Xing and Qiao [65], Spoon *et al.* [40], [41] generated 50 randomly patches of handwriting features from each image handwriting

TABLE 2. Sources of dyslexia datasets.

Data Source	Description	Website	Data Type
BishopBlog	Sharing of MRI dyslexia datasets	http://deevybee.blogspot.com/2012/05/sharing-of-mri-dyslexia-datasets.html	MRI scans
BishopBlog	Neuronal migration in language learning impairment	http://deevybee.blogspot.com/2012/05/neuronal-migration-in-language-learning.html	MRI scans
Kaggle	Dyslexia Handwriting Datasets	https://www.kaggle.com/izasazanita/dyslexia-handwriting-dataset	Handwriting images
NEMAR	EEG/ERP dataset for public download	https://sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html	EEG Scans
OpenNeuro	EEG/EPR dataset for Go-nogo categorization and detection task	https://openneuro.org/datasets/ds002680/versions/1.0.0	EEG scans
OpenNeuro	Auditory-visual shift study	https://openneuro.org/datasets/ds002893/versions/1.0.0	EEG scans
OpenNeuro	Auditory Cortex Mapping datasets	https://openneuro.org/datasets/ds003082/versions/1.0.0	MEG and MRI scans
OpenNeuro	Multimodal neuroimaging dataset to study spatiotemporal dynamics of visual processing in human	https://openneuro.org/datasets/ds002814/versions/1.2.1	T1w, T2w and BOLD fMRI scans
OpenNeuro	Visual Oddball Task (256 channels)	https://openneuro.org/datasets/ds002578/versions/1.0.1	sMRI scans
Pearson	Online platform for WRAT intelligence test	https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Brief/Wide-Range-Achievement-Test-%7C-Fifth-Edition/p/100001954.html	WRAT-5
TestingMon	Online platform for Woodcock-Johnson test of cognitive abilities.	https://www.testingmom.com/tests/woodcock-johnson/	WJ achievement test
The Eckert Laboratory	Neuroimaging dataset for developmental and age-related communication problems	http://eckertlab.org/	Brain fMRI and DTI scans

dataset used. Other data preparation and preprocessing techniques used in the dyslexia detection related machine learning studies include Least Absolute Shrinkage and Selection Operator (LASSO) and Recursive Feature Elimination with SVM (SVM-RFE) used by Benfatto *et al.* [61] and Rezvani *et al.* [62] to remove redundant and noisy features from the eye tracking datasets. This study only appraises the data preparation and processing methods; however, comparative analysis of these methods is not covered in this work.

D. FEATURE EXTRACTION AND SELECTION

Appropriate feature extraction and selection process is an important task in the detection of biomarkers of dyslexia because the number of features to be predicted is largely attributable to computational complexity. The purpose of feature extraction and selection is to generate most relevant and highly informative features from the original features [67]. This is done in order to eliminate redundant data or noise in the training dataset before they are fed into the machine learning predictive models, often called classifier. Features could be numerical or categorical (text/an aspect of an image). As presented in Table 3, studies have shown that relevant and discriminative features have been selected from heterogeneous datasets for machine learning dyslexia detection.

Various machine learning dyslexia detection studies have selected and used phonological features obtained manually from standardized tests, learners' handwritings, questionnaires and test scores [40]–[43]. Many image-based machine learning studies made use of MRI scans to extract and select features from brain tissue properties. Such properties include volumetric information, geometric (shape-based) measures, diffusion parameters, fractional anisotropy (FA), and activation patterns [3], [54]–[58], [68]. For EEG based machine learning studies, features relating to brain electrical signals or eye movement patterns have been extracted using various techniques, for example discrete wavelet transform [51], [52].

Al-Barhamtoshy and Motaweh [46] extracted feature relating to the brain electrical signals from the study samples using statistical functions and Fourier transform algorithms. The algorithms employed rule-based paradigm to filter noise and irrelevant features from the records of electrical signals dataset collected. Also, Artefact Subspace Reconstruction (ASR) based on sliding-window PCA was exploited by Perera *et al.* [47] to filter unwanted signals and noise from EEG brain electrical activity signals collected for dyslexia biomarker detection. Such unwanted signals include eye blinks and body movements.

In an MRI-based studies, Destrieux Atlas was used to select 742 geometric brain properties features at the superior and middle temporal gyri, subparietal sulcus, and prefrontal areas of the brain in an independent study conducted by Plonski *et al.* [53], [55] and Jednorog *et al.*. These areas are located at both left and right hemisphere of the brain cortex.

TABLE 3. Feature extraction methods for dyslexia biomarkers detection using machine learning.

Authors /year	Feature extraction method	Selected features
Kohli and Prasad (2010) [42]	This study used cognitive and behavioural evaluation test results.	Test of rapid naming, short-term memory evaluation and sequencing skills, non-word reading and IQ test.
Khan et al. (2018) [43]	Scikit-learn machine learning libraries of Python was used to extract and select features from test dataset conducted in Malay language.	Behavioural markers i.e., spelling and reading features.
Abdul Hamid et al. (2018) [44]	SURF descriptor was implemented in MATLAB to extract interest points from images of frontal faces collected.	Behavioural markers: Eye region, whole face area.
Lakretz et al. (2015) [45]	Reding errors features were manually extracted by expert from recorded audio responses obtained from dyslexic individuals during Hebrew reading screening test.	Phonological and cognitive marker: Encoded reading error features.
Al-Barhamtoshy and Motaweh (2017) [46]	This study used statistical functions and Fourier transform based on rule-based algorithm.	Brain electrical signals features
Perera et al. (2018) [47]	Unwanted and noisy signals were filtered using ASR implemented based on a sliding-window PCA in order to extract most relevant features.	Brain electrical activity signal features.
Prabha and Bhargavi (2019) [32]	PCA feature selection algorithms were used to select most relevant features from fixation and saccades eye movements dataset tracked with the aid of Ober-2.	Eye fixation and saccades features
Karim et al. (2013) [50]	KDE method was used to extract 353 most relevant brain activities signal features. This was implemented using *ksdensity( )* function in MATLAB.	Brain activities signal features.
Frid and Breznitz (2018) [52]	Most informative ERP signals were extracted using a Discrete Wavelet Transform method.	ERP brain activities signals features.
Chimeno et al. (2014) [37]	BET, FDT and TBSS were used to extract FA features from DTI scans while FEAT and BET were utilized to extract activation patterns from fMRI scans. These two features are associated with language, speech, and lexical decisions.	Phonological and Cognitive biomarkers: activation patterns, FA features.
Plonski et al. (2017) [55]	FreeSurfer image analysis suit was employed to implement algorithms used to extract 740 brain cortical volume and thickness features from 74 segmented brain hemisphere regions.	Fold index, GM volume, cortical thickness, surface area, mean curvature
Cui et al. (2016) [56]	VBM8 implemented in SPM8 (PANDA) was used to extract WM volume features. Other diffusivity features were selected directly from the WM density images. Most discriminative features were selected using nested-LOOCV: inner and outer.	WM neuroimaging features: FA, MD, AD, and RD.
Tamboer et al. (2016) [3]	VBM method was used to extract GM volumes in the three brain phonological areas on DTI scans of the study participants. This activity was preceded by tissue-type segmentation and affine registration. Most relevant voxels were extracted from these areas.	Phonological biomarkers: GM volume voxels features.
Rello et al. (2018) [60]	Kolmogorov-Smirnov test was performed to extract most informative computer-game features.	Phonological and cognitive markers: Computer-game features i.e., hits and misses.
Rezvani et al. (2019) [62]	Brain Connectivity Toolbox (BCT) of MATLAB was employed to extract and select the most relevant local network features from electrical activity brain signals.	Weighted connectivity graph features.
Spoon et al. (2019) [40], [41]	Keras and TensorFlow libraries of Python was used to implement CNN classifier which extract high-level features from patches of handwriting images.	Cognitive and behavioural biomarkers: image patches of handwriting features.
Zahia et al. (2020) [57]	SMP12 implemented in MATLAB 2018b was used to extract 3D voxel volumes of activated areas of the brain from fMRI scans after normalization and smoothing.	Features extracted from both left and right of the following brain areas: Broca's area, Wernicke's area, VWFA and MTG.
Usman and Muniyandi [58]	Phonological and cognitive features relating to GM, WM, and CSF were extracted from fMRI scans using CAT12 implemented in MATLAB 2017b. This was preceded by	Phonological and cognitive brain sub-system tissue features.

For each area, feature extracted included fold index, GM volume, cortical thickness, surface area, and mean curvature. The above features were selected using three distinct feature selection algorithms namely, t-test, Information Gain (IG), and Random Forest Variable Importance. All the algorithms were implemented in FreeSurfer image analysis suit.

In the study by Plonski *et al.* [53], optimum number of features were selected as the subset with the lowest LogLoss discriminatory power metric. These features comprise 25 most relevant features of cortical brain properties. Naïve Bayes, Site-Dependent Whitening (SDW), Site-Dependent

Extension (SDE) were modelled to extract these features at (p-value$<$0.05).

In another study, Tamboer *et al.* [3] utilized GM volume of voxels found in the left and right occipital fusiform gyri (LOFG, ROFG) and left inferior parietal lobule (LIPL) as most discriminative features for training SVM model. The volumes of these voxels were obtained by subtracting the average VBM transformed brain of controls from the dyslexic subjects' average VBM transformed brain, and z-transforming the resulting variations.

On the contrary to GM volumetric measures, WM volume, diffusion properties and FA features have been selected as discriminating features for dyslexia in a study by Cui *et al.* [56], [63]. These properties included mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD). The regions of selection of these features were group into three: putative reading system (which include the superior longitudinal fasciculus, inferior fronto-occipital fasciculus, thalamocortical projections, and corpus callosum); the limbic system (which include the cingulum and fornix); and the motor system (which include cerebellar peduncle, corona radiata, and corticospinal tract). Extraction and processing of these features were implemented using PANDA. Meanwhile, the most discriminative features were selected using nested-LOOCV, both inner and outer. These comprise 43 WM discriminative features selected from 12 bilateral WM regions in totality i.e., 31 WM volumes, 1 FA, 4 MDs, 4 ADs and 3 RDs. The procedure is preceded by non-rigid registration (MNI152), bias-field correction, spatial normalization and warping into Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) template.

Chimeno *et al.* [37] extracted both FA and activation pattern features from DTI and fMRI scans. FA features were extracted using Brain Extraction Tool (BET), FDT (FMRIB Diffusion Toolbox) and Track-Based Spatial Statistics (TBSS) functions while brain activation pattern features were extracted using FMRI Expert Analysis Tool version 5.98 (FEAT) functions and Brain Extraction Tool (BET). The extraction procedure was preceded by non-rigid registration and Gaussian smoothing to improve image tissue comparability.

E. MACHINE LEARNING TRAINING AND CLASSIFICATION

For classification of dyslexia and its biomarkers, construction and training of models is achieved using machine learning or deep learning algorithms. While early studies focused on a single traditional machine learning algorithm [3], [42], [43], [47], [50], [59], [60], other studies developed hybrid methods for dyslexia detection [32], [44], [61]. The implementation of deep learning algorithms are the most recent studies in dyslexia and its biomarkers detection researches [57], [58]. Meanwhile, majority of the studies compared different types of machine learning methods using the same dataset [32], [37], [44]–[46], [52], [55], [56], [62]. It is a usual practice to partition datasets into two during the training of machine learning models: training set (containing larger percentage of the entire data, $\geq$70% in most cases), testing or validation set (containing 30% or lesser of the dataset). Existing studies have used support vector machine (SVM), logistic regression (LR), artificial neural network (ANN), random forest (RF), linear discriminant analysis (LDA), k-nearest neighbour (KNN), k-mean classifier, Naïve Bayes, principal component analysis (PCA) and different architecture of convolutional neural network (CNN) as the machine learning algorithm to classify study-samples into binary class with significantly high classification accuracy.

As illustrations, Khan *et al.* [43] proposed KNN model for developing a diagnostic and classification system for children with dyslexia. Results from implementation of this machine learning model showed that 23% and 21% of the pupils sampled during the study were at the risk of dyslexia with 98% classification accuracy. ANN model proposed by Kohli and Prasad [42] achieved maximum accuracy of 75% after 10-fold cross-validation (CV) when implemented with student performance dataset obtained with the aid of structured questionnaire. Chimeno *et al.* [37] achieved highest classification performance for 3-layered ANN model amidst other machine learning classifiers that were compared against it. This result shows 94.87% accuracy, 94.73% sensitivity and 95% specificity respectively.

Tamboer *et al.* [3] present a study on finding the structural biomarkers in dyslexic by analyzing GM volume in the brain regions of the study participants using SVM classifier. SVM model yields an accuracy of 80%, sensitivity of 82%, specificity of 78% on the training and testing dataset. Cui *et al.* [56] proposed a Linear SVM (LSVM) model for investigating the biomarkers of dyslexia from WM volumetric and diffusivity features. Results from LSVM showed 83.61% accuracy, 75.00% sensitivity, 90.91% specificity, 87.50% TPR, 81.08% TNR and AUC of 0.86 (p-value$<$0.001) respectively, which is significantly higher than the baseline. This model shows improved performance compared against LR classifier which achieved 73.77% accuracy, 67.86% sensitivity, 78.79% specificity, 73.08% TPR, and 74.29% TNR when trained with the same features and datasets. Plonski *et al.* [55] compared three different machine learning methods for discriminating dyslexia biomarkers in schoolchildren. These methods include LR model, SVM with linear kernel, and RF classifier. Although, almost all combinations of classifiers give significantly above chance performance with classification accuracy up to 65%, but most stable feature selection in all CV iterations is LR which achieved 65% accuracy and AUC of 0.66 which are significant at (p-value$<$0.01).

Spoon *et al.* [40] proposed CNN model to automatically identify possible indications of dyslexia from the children's handwritings (behavioural biomarker). By employing a 5-fold CV method, the proposed CNN model achieved an average accuracy of 55.7$\pm$1.4% above the random baseline of 50%, which has been found to be higher than current detection rate. The model was further enhanced by Spoon *et al.* [41] to achieve an accuracy of 77.6% through the process of hyper-parameters fining tuning. Zahia *et al.* [57] proposed

3D CNN model and achieved highest accuracy of 72.7%, sensitivity of 75%, specificity of 71.4%, 60% precision and 67% F1-score while Usman and Muniyandi [58] proposed two-way cascaded CNN model and achieved highest accuracy of 84.6%, sensitivity of 76.5%, specificity of 78.2% and AUC of 0.76 respectively.

From Table 3 and Figure 3, SVM is the most used machine learning model (13 studies, 34%), followed by ANN (5 studies, 13%), KNN (4 studies, 10.5%), CNN (4 studies, 10.5%), LR (3 studies, 8%), RF (3 studies, 8%), Naïve Bayes (2 studies, 5%) and k-means algorithm (2 studies, 5%). AdaBoost and PCA are less commonly used for dyslexia detection (1 study, 3% each). Worthy of note is fact that, most of the above models are instance-based interpretable machine learning algorithms. As an instance-based machine learning algorithms, KNN, K-means, Naïve Bayes and SVM construct assumptions directly from training instances (features) by adjusting their classifiers to the test dataset. While the asymptotic complexity of their assumptions increases in proportion to the volume of data, they possess the ability to store new features (instances) and discard old ones. Linear-R, LR, DTs, RF, SVM, ANN and deep CNNs used for dyslexia detection possess learned structures and learned parameters such as weights with specific interpretations [69], [70]. Also, sensitivity of their components to the new dataset as well as parameter adjustments can be analyzed. For example, components of CNN architectures implemented in both Zahia *et al.* [57]and Usman and Muniyandi [58] as well as generated feature maps can be visualized and interpreted at any layer of model activation during the training stage. Conclusively, the selection of suitable machine learning methods for dyslexia detection depend ultimately on the type of dataset for training the generated models. Therefore, potential studies should conduct a comparative performance of various advance machine learning methods in order to analyze the output of different classifiers rather than focusing on the performance of only one selected method.

F. PERFORMANCE EVALUATION

As shown in Table 4, existing studies have implemented machine learning or deep learning algorithms on MATLAB, WEKA, and Python environments using different Toolboxes and libraries functions such as LIBSVM, *train_test_split(*) function, scikit-learn, Keras and TensorFlow in order to construct predictive models. Evaluation of these models are based on different metrics of machine learning for dyslexia detection. This includes accuracy, sensitivity, specificity, precision, recall, mean square error (MSE), F-scores, positive predictive value (PPV), negative predictive value (NPV), receiver operating characteristic (ROC) curve and area under ROC curve (AUC). Accuracy is the percentage of dyslexic subjects correctly classified as positive. Sensitivity, also known as true positive rate (TPR) or recall, represents a proportion of correctly classified subjects with dyslexia. Specificity, or the true negative rate (TNR), is the proportion of the correctly classified subjects without dyslexia. The ROC curve

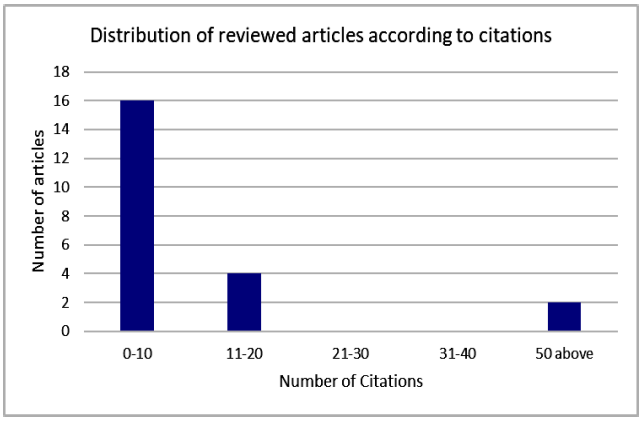


FIGURE 4. Distribution of literatures according to number of citations.

is a precision plot against recall, and thus provides a reflection of the trade-off between correctly classified dyslexic subjects and incorrectly classified dyslexic subjects from where AUC can be calculated directly.

In all the reviewed studies, evaluations of the machine learning model performance are based on k-fold cross validation (CV), where k=4, . . . , 10) [41], [55], [57], [58], [60]and leave-one-out-cross-validation (LOOCV) [55], [56], [62]. The most common type is 10-fold CV where the experimental dataset is randomly divided into 10 equal folds, and 9 of the folds used for model training while the remaining 1-fold is used for testing and validation. With exception of Plonski *et al.* [53], all other literatures reviewed in this study made use of accuracy metrics. The overall performance accuracy of machine learning models based on the data-type used is summarized thus: test-based methods ranges between 73.9%-99%, handwriting-based method ranges between 55.7%-77.6%, EEG-based methods ranges between 78%-95%, MRI-based methods ranges between 64%-94.8%, video-based methods ranges between 77.6%-97.8% and lastly, eye tracker-based method is 80.2%. It would therefore be interesting to understand how the performance of these machine learning methods changes when dataset from multiple sources and types are combined for model training. Figure 4 presents the distribution of the review articles in Table 4 according to the number of citations.

III. DISCUSSION AND POTENTIAL CHALLENGES

Machine learning algorithms are general-purpose methods of artificial intelligence that can learn patterns from data without the need to define them apriori and have been widely used for the detection of dyslexia over the past decades. The typical machine learning workflow for dyslexia prediction involves five distinct steps: data collection, data preparation and pre-processing, extraction and selection of features, model (classifier) training and simulation, as well as performance evaluation, thus forming the focus of this review. Although still in its infancy, the different architecture of deep learning algorithms is currently being applied in the field of dyslexia studies due to the inability of conventional machine learning

TABLE 4. Machine learning classifier performance for dyslexia biomarker detection.

Authors/ year	Objective	No. of subjects	Type of Data	Age range	Language	Method of data analysis	Machine/ Deep learning method	Classifier Performance
Kohli and Prasad (2010) [42]	To identify the possibility of dyslexia in schoolchildren early using artificial neural network.	-	Test score	-	English	MLP architecture of ANN was designed on NeuroSolution5 software. Performance evaluation was based on 10-folds cross-validation.	ANN	Accuracy=75%
Khan et al. (2018) [43]	To develop and evaluate an automated diagnostic system for dyslexia classification.	857	Test scores	6-7	Malay	*Train_test_split( )* function of scikit-learn libraries was used to implement KNN model for dyslexia classification.	KNN	Accuracy=99.0%
Abdul Hamid et al. (2018) [44]	To predict dyslexic students' engagement towards learning using machine learning based on BOF.	30	Face video and images	7-12	Malay	Three machine learning classifiers forming the basis of BOF were implemented in WEKA for predicting facial engagements and disengagements.	Linear SVM, RBF-SVM, Naïve Bayes, KNN	Accuracy: Linear SVM=97.8% RBF-SVM=97.1% Naïve Bayes=97.3% KNN=77.6%
Lakretz et al. (2015) [45]	To analyze patterns of reading errors in dyslexic individuals using graphical machine learning models.	313	Readin g test errors	7-62	Hebrew	Three probabilistic graphical models were used to construct classifiers using one Latent Dirichlet Allocation (LDA) and two Naïve Bayes.	LDA, Naïve Bayes	Accuracy: LDA=73.9% Naïve Bayes=80.1%
Al-Barhamtoshy and Motaweh (2017) [46]	To diagnose dyslexia early using a computational analysis classifier.	80	EEG scans	7-18	Arab	Machine learning algorithms were used to cluster data items based on features similarities.	K-means Algorithm, ANN, Fuzzy logic classifier.	Average accuracy: K-means=89.6% ANN=89.7% Fuzzy=85.7% Overall: Accuracy=81.1% Precision=62.0% Recall=100% F-score=76.6%
Perera et al. (2018) [47]	To explore the potency of machine learning methods in dyslexia detection using brain electrical signals obtained during writing and typing.	32	EEG	18+	English	Cubic SVM was used to construct a classifier for detecting activation pattern from the preprocessed EEG signals.	Cubic SVM	Accuracy=78.2% Sensitivity=88.2% Specificity=66.7%
Prabha and Bhargavi (2019) [32]	To propose a prediction model for differentiating dyslexics from non-dyslexics using their eye movement.	185	EEG scans	9-10	Swedish	PSO-based hybrid kernel SVM-PSO was used to fine-tune and predict dyslexic from high-level features extracted by PCA. Performance compared against linear SVM.	Linear SVM, Hybrid kernel SVM-PSO	Accuracy: Linear SVM=90% SVM-PSO=95%
Karim et al. (2013) [50]	To detect dyslexia marker from brain activities signals collected during resting state using Multilayer Perceptron (MLP).	6	EEG scans	4-7	Malay	MLP was used to construct a classifier for distinguishing dyslexic brain activities signals.	MLP	Accuracy: Eyes closed=85.0% Eye opened=86.0%
Frid and Breznitz (2018) [52]	To analyze differences between dyslexic and skilled readers ERP signals using machine learning.	32	EEG scans	Grade 6-7	Hebrew	Machine learning classifiers were constructed using SVM, ANN and PCA.	SVM, ANN, PCA	Accuracy=78.0%

TABLE 4. *(Continued.)* **Machine learning classifier performance for dyslexia biomarker detection.**

Chimeno et al. (2014) [37]	To classify monocular vision subjects into dyslexics and controls groups.	56	fMRI scans, DTI scans	9-12	English	Classification system was design combinations of machine learning algorithms.	LDA, SVM, ANN, K-means, KNN, AdaBoost	ANN produced the best results: Accuracy=94.8% Sensitivity=94.7% Specificity=95%
Plonski et al. (2014) [53]	To examine the effect of MRI data distribution, site location and scanner field strength on dyslexia biomarker feature selection and machine learning classifier accuracy.	236	MRI scans	8.5-13.7	French, German and Polish	LR was modelled to discriminate between biomarkers of dyslexic and control subjects. The performance compared against RF and SVM.	LR, RF, SVM	AUC: Naïve LR=0.728 SDW LR=0.824 SDE LR=0.832 Naïve RF=0.640 SDW RF=0.690 SDE RF=0.660 Naïve SVM=0.616 SDW SVM=0.737 SDE SVM=0.651
Plonski et al. (2017) [55]	To investigate GM disruptions in dyslexic children using machine learning approach.	236	MRI scans	8.5-13.7	French, German and Polish	Three machine learning classifiers were trained and compared for discriminating dyslexia biomarkers.	Linear SVM, LR, RF	LR classifier achieved: Accuracy=65% AUC=0.66 for 10-fold CV Accuracy=64% AUC=0.65 for LOOCV
Cui et al. (2016) [56]	To discriminate dyslexia biomarker using WM neuroimaging features and machine learning methods.	61	MRI scans	10-14.7	Mandarin	LIBSVM Toolbox of MATLAB was used to implement LSVM biomarker classification. The classification was validated against LR classifier implemented in WEKA.	Linear SVM, LR	Linear SVM: Accuracy=83.6% Sensitivity=75.0% Specificity=90.9% PPV=87.5% NPV=81.1% AUC=0.86 LR classifier: Accuracy=73.8% Sensitivity=67.9% Specificity=78.8% PPV=73.1% NPV=74.3% AUC=0.80
Tamboer et al. (2016) [3]	To discriminate between dyslexic and non-dyslexic subjects based on anatomical differences using machine learning method.	925	MRI scans: DTI	18-21	Dutch	Libsvm of WEKA was employed to train SVM classifier to classifier voxels into binary class: with dyslexia and without dyslexia.	SVM	Accuracy=80% Sensitivity=82% Specificity=78%
Rello et al. (2018) [60]	To predict the likelihood of dyslexia in people interactions with a linguistic computer-based game using machine learning.	267	Eye tracker	7-60	English, Spanish	LIBSVM was utilized to model SVM with a Gaussian kernel for dyslexia prediction.	SVM	Accuracy=84.6%
Benfatto et al. (2016) [61]	To predict the risk of dyslexia in eye movement during reading using machine learning technique.	185	Eye tracker	9-10	Swedish	Automatic Recursive Feature Elimination was integrated with SVM to extract most relevant features from eye tracking signals and predict the risk of dyslexia. Performance of this model is compared against RF classifier.	SVM-RFE RF	SVM-RFE: Accuracy=95.6% Sensitivity=95.5% Specificity=95.7% RF classifier: Accuracy=95.3% Sensitivity=95.2% Specificity=95.5%

TABLE 4. *(Continued.)* **Machine learning classifier performance for dyslexia biomarker detection.**

Rezvani et al. (2019) [62]	To discriminate between neural-biomarkers of dyslexia from normal using local network features and machine learning classifiers.	44	EEG scans	Grade 3	Dutch	SVM with linear and polynomial kernels were trained using Machine Learning and Statistics Toolbox of MATLAB 10.8.0. The same toolbox was used to train KNN classifier which was compared with it.	SVM, KNN	Accuracy: SVM=95.4% KNN=86.0%
Rello and Ballestero (2015) [59]	To predict dyslexic readers using eye tracking measures and machine learning.	97	Eye tracker	11-54	Spanish	LIBSVM was used to train SVM with polynomial kernel for predicting dyslexia from readers eye tracking measures.	SVM	Accuracy=80.2%
Spoon et al. (2019) [40]	To identify dyslexia marker in schoolchildren's handwriting using neural network.	88	Handwriting images	K-6	Spanish, Latin	CNN was used to classify image of handwriting patches on Keras and TensorFlow	CNN	Accuracy=55.7%
Spoon et al. (2019) [41]	To identify possible indication of dyslexia in children's handwritings using advance machine learning method.	100	Handwriting images	K1-6	English	CNN was used to classify image of handwriting patches. Implemented of Keras and TensorFlow	CNN	Accuracy=77.6%
Zahia et al. (2020) [57]	To develop an automatic recognition system for children with dyslexia using fMRI data and CNN.	66	fMRI scans, BOLD functional images	9-12	Spanish	Three parallel 3D CNN was used to detect dyslexic brain activations. Keras 2.1 and TensorFlow libraries	3D CNN	Accuracy=72.7%, Sensitivity=75.0% Specificity=71.4% Precision=60.0% F1-score=67.0%
Usman and Muniyandi (2020) [58]	To design a privacy-preserving classification system for dyslexia biomarkers using CNN	45	MRI scans	15-23	-	Cascaded CNN model was used to classify encrypted image patches on MATLAB design interface	Two-way cascaded CNN	Accuracy=84.6% Sensitivity=76.3% Specificity=78.2% AUC=0.76

algorithms to process data in their natural raw form, coupled with other limitations.

Evidence emanating from the reviewed studies has shown that, collection of dyslexia datasets from the standardized psycho-educational tests and learners' handwritings (first category) is relatively cheap, as demonstrated in the studies by Spoon *et al.* [40], [41]. However, they may be difficult to collect until the child reaches school age. The child's brain structure remains unchanged from childhood to adulthood, although its pattern of activation changes due to the rapid formation of synapses as the child grows up, unless the child's brain is severely injured or critically ill. Studying dyslexia from image-based or eye pattern movement datasets (second and third categories) therefore, delivers more accurate results despite the fact that, they are currently relatively expensive to collect.

Khan *et al.* [43] proposed a KNN model for the classification of children with dyslexia and achieved a 98% accuracy. The ANN model proposed by Kohli and Prasad [42] achieved a maximum accuracy of 75% after 10-fold cross-validation (CV). Chimeno *et al.* [37] achieved the highest classification performance for the 3-layer ANN model with 94.9% accuracy, 94.7% sensitivity and 95% specificity. The ANN model proposed by Karim *et al.* [50] achieved a classification accuracy of 85% and 86% when trained with a backpropagation algorithm.

The performance of the SVM model proposed by Tamboer *et al.* [3] was found to be 80% accuracy, 82% sensitivity, 78% specificity on a training and a testing dataset samples containing 49 neuroimages. A significant decrease in model performance was observed when the model was validated on new datasets comprising 816 neuroimaging samples with 59% accuracy, 67% sensitivity and 59% specificity. This was due to high rate of false alarms, class rarity (imbalanced dataset) and undiagnosed cases of dyslexia in the sample, among others. Abdul Hamid *et al.* [44] proposed linear SVM, RBF-SVM, Naïve Bayes and KNN models and achieved the highest classification accuracy of 97.8%, 97.1%, 97.3% and 77.6% using the same image-based dyslexia dataset. Perera *et al.* [47] proposed a cubic SVM model with 78.2% accuracy, 88.2% sensitivity and 66.7% specificity using the brain pattern signal features extracted from the

EEG dataset. Prabha and Bhargavi [32] proposed SVM and SVM-PSO models for linear and hybrid kernels, respectively. These models achieved the highest classification accuracy of 90% and 95% respectively. In addition, Rello *et al.* [60] achieved the highest classification accuracy of 84.6% with the SVM classifier. Benfatto *et al.* [61] achieved 95.6% accuracy, 95.5% sensitivity and 95.7 specificity using the SVM-RFE model for the classification of dyslexia. Rello and Ballestero [59] achieved 80.2% with the SVM model, while Rezvani *et al.* [62] achieved 95.4% and 86% accuracy for SVM and KNN respectively.

Cui *et al.* [56] proposed a linear SVM model and achieved 83.6% accuracy, 75% sensitivity, 90.9% specificity, 87.5% TPR, 81.1% TNR and 0.86 AUC, respectively, which is significantly higher than the baseline. This model shows improved performance compared to the LR classifier, which achieved 73.8% accuracy, 67.9% sensitivity, 78.8% specificity, 73.1% TPR and 74.3% TNR when trained using the same extracted features. Plonski *et al.* [55] compared three different methods of machine learning, including LR model, SVM with linear kernel and RF classifiers. Although, almost all combinations of classifiers give significantly high classification performance with accuracy of up to 65%. The most stable feature selection in all CV iterations was found in the LR model, which achieved 65% accuracy and AUC of 0.66. With the SVM model, Frid and Breznitz [52] achieved 78% accuracy. This performance was found to be better than the equivalent ANN and PCA models respectively. Chimeno *et al.* [37] proposed models LDA, SVM, ANN, K-means, KNN, and AdBoost. Analysis of the six proposed models showed that the ANN model achieved the best performance with 94.8% accuracy, 94.7% sensitivity and 95% specificity.

Lakretz *et al.* [45] proposed LDA and Naïve Bayes and achieved the highest classification accuracy of 73.9% and 80.1% respectively. Al-Barhamtoshy and Motaweh [46] achieved classification accuracy of 89.6%, 89.6% and 85.7% respectively with K-means, ANN and fuzzy logic classifiers. The overall accuracy of the three models was found to be 81.1% accuracy, 62% precision, 100% recall and 76.6% F-score. Most of the traditional machine learning algorithms discussed above for the detection of dyslexia suffer from a number of problems ranging from data sparsity [37], [40], [44], [46], [47], [51], [56], [59], [62] curse of dimensionality [42], [43], [45], underfitting [32], [51], overfitting [3], [53], [62], high false alarm rates [3] and imbalanced classification [40], [60], [61] to mention but a few.

The CNN model proposed by Spoon *et al.* [40] achieved a test accuracy of 55.7% when validated using a 5-fold CV method. However, this accuracy was found to be higher than the current baseline detection rate. The accuracy of the model was further improved by Spoon *et al.* [41] in order to achieve an accuracy of 77.6% through the process of hyper-parameter fine-tuning. Zahia *et al.* [57] proposed a 3D CNN model and achieved 72.7% accuracy, 75% sensitivity, 71.4% specificity, 60% precision and 67% F1-score, while Usman and Muniyandi [58] proposed a two-way cascaded CNN model and achieved 84.6% accuracy, 76.5% sensitivity, 78.2% specificity and AUC of 0.76, respectively.

While the use of deep learning models, specifically CNN models is beginning to gain attention in dyslexia detection researches, it is therefore pertinent to discuss some of the potential challenges users of contemporary architectures of deep CNN in dyslexia analysis will need to address in order to maximize the state-of-the-art benefits of these models.

1. Achieving biologically-interpretable biomarker features remain one of the prominent challenges in MRI-based studies due to inconsistent and varied image acquisition protocols. Since CNN models learn high-level abstract features directly from the images, potential CNN model users need to device a mechanism for improving the homogeneity and comparability of this kind of dataset before using them to train CNN models. Such mechanism could be smoothing and normalization and are capable of improving the state-of-the-art deep learning performance.
2. Accuracy of the state-of-the-art deep CNN classifier is largely dependent on the type of preprocessing tasks performed on the input dataset before feeding them into the CNN-based models. This may be negatively affected if not properly handled for any kind of dataset. It is therefore, recommended that potential deep learning users for dyslexia biomarker analysis need to devise an algorithm that will improve existing data preprocessing methods in order to enhance common interpretability of relevant features to be studied. Consequently, the suggested improvement plan will, in no doubt, enhance the performance of the state-of-the-art deep CNN models. Furthermore, adapting existing supervised learning algorithms for training deep learning can also improve the interpretability of relevant dyslexia features if attempted.
3. Privacy of both experimental dataset and CNN model hyper-parameters is also a major potential challenge in achieving state-of-the-art classification performance. Increasing generation of dyslexia dataset across multiple sources has prompted the need for cloud-based classifications, a concept otherwise, referred to as privacy-preserving classification. A typical example of this setup is Microsoft Azure Machine Learning Studio and Google Cloud Machine Learning Engineering. The protection of patient medical records has become highly necessary in this situation [71], particularly, when such data is related to learning disability such as dyslexia. While remote classification can play a significant role in enhancing the quality and timely availability of medical decision by reducing the cost, such data may be exposed to various forms of security, privacy, confidentiality, and integrity threats. It is therefore recommended that potential deep CNN users should devise an encryption system for both the dataset and the model parameters if this kind of situation is

inevitable. The use of homomorphic encryption [72] system can be helpful in this regard.

4. Inadequacy of dataset, parameter optimization and possibility of overfitting have remained a serious problem in deep learning implementation from the inception. At present, there exist problem of paucity of dataset, particularly those of the brain anatomy and activation patterns for dyslexia study. Nearly all MRI-based studies reviewed suffer from inadequate MRI dataset leading to model overfitting in some cases. Overfitting also occurs in situation when a deep learning architecture, particularly CNN models, contains large number of hyper-parameters [73]. It is therefore, suggested that potential users of CNN for dyslexia detection from image-based dataset employ approaches recommended in the existing deep learning studies to minimize the effect of overfitting. Such approaches include dropout technique [74], data dimensionality reduction e.g., creation of smaller image patches [40], 58], parameters fine-tuning [41], data augmentation [75], and regularization (L_1 and L_2) to mention but a few.
5. Selection of hyper-parameters is also a potential challenge considering their impact on deep CNN performance. A minor change in the values for hyperparameter will have significant effect on CNN's overall efficiency. Careful selection of parameters is therefore a possible design problem that needs to be addressed when using deep CNN to evaluate dyslexia-related dataset.
6. Finally, efficient deep CNN training requires powerful hardware tools like Graphics Processing Units (GPUs). The GPU-base processors are powerful in terms of processing speed and memory usage. They are up to 100 times faster than their Central Processing Units (CPUs) equivalent, with an enlarged arithmetic computational capability [76] and have been used to achieve reduced processing time in studies such as [77] and [78]. Implementing deep learning algorithms on a system based on this type of hardware resources will help in boosting the classification speed for dyslexia and its biomarkers detection.

IV. CONCLUSION AND FUTURE DIRECTION

This study focuses on reviewing existing machine learning and deep learning algorithms which have been implemented for dyslexia and its biomarker detection. Dyslexia, an exceedingly complicated brain developmental disorder, has attracted considerable interest in modern neuroscience and machine learning field in recent times. Although, considerable machine learning methods have been applied in this domain of study during the last two decades, the use of deep learning algorithms is still at its infant stage. Also, from our review, it can be deduced that SVM is the most used machine learning method for dyslexia detection and prediction. Data for dyslexia detection and analysis have been collected from multiple heterogeneous sources.

For the purpose of this review, selected articles for critical analysis were discovered through major academic databases using some combinations of search keywords and criteria. They were screen and selected following the steps of PRISMA flow diagram and certain pre-defined inclusion criteria. Consequent upon the above selection process, only 22 machine learning based studies for dyslexia detection were thoroughly reviewed and summarized in Table 4.

Results from all machine learning methods reviewed are promising and attest to the fact that, dyslexia is a heterogeneous disorder caused by alterations in brain tissues characteristics. By extension, they also complement the existing computer-based interventions in that, most of these interventions are AI-based computer games. With the increasing usage of deep learning in modern medical field coupled with recent state-of-the-art performance shown by CNN models, this review holds the view that achieving high classification performance with CNN models for dyslexia and its biomarkers detection is feasible through tackling the identified potential challenges. As an example, many recent studies and reviews [79]–[82] have shown how different CNN model architectures have been used to analyze MRI datasets of various critical brain disorders such as brain tumor, Alzheimer's disease, Parkinson's disease and schizophrenia. For CNN to compete favourably with other existing machine learning methods and attain clinically acceptable level, it is therefore suggested that serious consideration should be accorded to the identified challenges by future researches. Meanwhile, there is a need for future studies to develop a language-independent data collection protocol for machine learning dyslexia detection since this learning disability is not specific to a region, language, or culture. It would be an interesting research, if future studies can evaluate the impact of ensemble methods, a situation where prediction capabilities from multiple machine learning models are hybridized to improve performance of the machine learning methods. In addition to the above, potential dyslexia detection studies should also consider the implementation of multimodal DL frameworks [80]–[85] which allow the capitalization of different dyslexia datasets from multiple heterogeneous sources, as well as the use of multi-task techniques to manage auxiliary tasks for improved performance.

APPENDIX

See supplementary file with name **Supplementary1.pdf** for more details about artificial neural network and deep learning.

ACKNOWLEDGMENT

The authors would like to appreciate the support of Universiti Kebangsaan Malaysia (UKM).

REFERENCES

[1] A. Elnakib, A. Soliman, M. Nitzken, M. F. Casanova, G. Gimel'farb, and A. El-Baz, "Magnetic resonance imaging findings for dyslexia: A review," *J. Biomed. Nanotechnol.*, vol. 10, no. 10, pp. 2778–2805, Oct. 2014, doi: 10.1166/jbn.2014.1895.

[2] J. M. Fletcher *et al.*, "Classification of learning disabilities: An evidence-based evaluation," in *Identification of Learning Disabilities: Research to Practice*, R. Bradley, L. Danielson, and D. P. Hallahan, Eds. Washington, DC, USA: Erlbaum Associates Publishers, Jan. 2002, pp. 185–250.

[3] P. Tamboer, H. C. M. Vorst, S. Ghebreab, and H. S. Scholte, "Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia," *NeuroImage: Clin.*, vol. 11, pp. 508–514, 2016, doi: 10.1016/j.nicl.2016.03.014.

[4] S. O. Wajuihian, "Neurobiology of developmental dyslexia Part 1: A review of evidence from autopsy and structural neuro-imaging studies," *Optometry Vis. Develop.*, vol. 43, no. 3, pp. 121–131, 2012.

[5] S. O. Wajuihian and K. S. Naidoo, "Dyslexia: An overview," *Afr. Vis. Eye Health*, vol. 70, no. 2, pp. 89–98, Dec. 2011, doi: 10.4102/aveh.v70i2.102.

[6] B. Sarah, C. Nicole, H. Ardag, M. Madelyn, S. K. Holland, and H. Tzipi, "An fMRI Study of a Dyslexia biomarker," *J. Young Investigators*, vol. 26, no. 1, pp. 1–4, 2014.

[7] Association International Dyslexia, Baltimore, MD, USA. (2017). *Dyslexia in the Classroom—What Every Teacher Needs to Know*. [Online]. Available: www.DyslexiaIDA.org

[8] N. A. M. Yuzaidey, N. C. Din, M. Ahmad, N. Ibrahim, R. A. Razak, and D. Harun, "Interventions for children with dyslexia: A review on current intervention methods," *Med. J. Malaysia*, vol. 73, no. 5, pp. 311–320, Oct. 2018.

[9] S. Kaisar, "Developmental dyslexia detection using machine learning techniques: A survey," *ICT Exp.*, vol. 6, no. 3, pp. 181–184, 2020, doi: 10.1016/j.icte.2020.05.006.

[10] F. Ramus, I. Altarelli, K. Jednoróg, J. Zhao, and L. Scotto di Covella, "Neuroanatomy of developmental dyslexia: Pitfalls and promise," *Neurosci. Biobehav. Rev.*, vol. 84, pp. 434–452, Jan. 2018, doi: 10.1016/j.neubiorev.2017.08.001.

[11] T. R. Vidyasagar, "Visuo-spatial attention," *Clin. Exp. Optom.*, vol. 87, no. 1, pp. 4–10, 2004.

[12] L. Danelli, M. Berlingeri, G. Bottini, F. Ferri, L. Vacchi, M. Sberna, and E. Paulesu, "Neural intersections of the phonological, visual magnocellular and motor/cerebellar systems in normal readers: Implications for imaging studies on dyslexia," *Hum. Brain Mapping*, vol. 34, no. 10, pp. 2669–2687, Oct. 2013, doi: 10.1002/hbm.22098.

[13] U. Kuhl, N. E. Neef, I. Kraft, G. Schaadt, L. Dörr, J. Brauer, I. Czepezauer, B. Müller, A. Wilcke, H. Kirsten, F. Emmrich, J. Boltze, A. D. Friederici, and M. A. Skeide, "The emergence of dyslexia in the developing brain," *NeuroImage*, vol. 211, May 2020, Art. no. 116633, doi: 10.1016/j.neuroimage.2020.116633.

[14] S. M. Irby and R. G. Floyd, "Test review: Wechsler abbreviated scale of intelligence, second edition," *Can. J. School Psychol.*, vol. 28, no. 3, pp. 295–299, Sep. 2013, doi: 10.1177/0829573513493982.

[15] C. A. Dell, B. Harrold, and T. Dell, "Test Review: Wilkinson, G. S., & Robertson, G. J. (2006). wide range achievement Test-Fourth Edition. Lutz, FL: Psychological Assessment Resources. WRAT4 Introductory Kit (includes manual, 25 test/response forms [blue and green], and accompanying test mat," *Rehabil. Couns. Bull.*, vol. 52, no. 1, pp. 57–60, 2008, doi: 10.1177/0034355208320076.

[16] B. J. Wending, F. A. Schrank, and A. J. Scmitt, "Educational interventions related to the Woodcock-Johnson III test of achievement," in *Woodcock-Johnson III Assessment Service Bulletin Number*, vol. 8, no. 8. Rolling Meadows IL, USA: Riverside Publishing, A Houghton Mifflin Company, 2007, pp. 1–30.

[17] J.-J. Mitchell, "Comprehensive test of phonological processing," *Assessment Effective Intervent*, vol. 26, no. 3, pp. 57–63, Apr. 2001, doi: 10.1177/073724770102600305.

[18] R. M. Bruno and S. C. Walker, "Comprehensive test of phonological processing (CTOPP)," *Diagnostique*, vol. 24, nos. 1–4, pp. 69–82, Mar. 1999, doi: 10.1177/1534508499024014-408.

[19] N. L. Bell, K. S. Lassiter, T. D. Matthews, and M. B. Hutchinson, "Comparison of the peabody picture vocabulary test—Third edition and Wechsler adult intelligence scale—Third edition with university students," *J. Clin. Psychol.*, vol. 57, no. 3, pp. 417–422, 2001, doi: 10.1002/jclp.1024.

[20] B. Srivastava and M. T. U. Haider, "Personalized assessment model for alphabets learning with learning objects in e-learning environment for dyslexia," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 7, pp. 809–817, 2020, doi: 10.1016/j.jksuci.2017.11.005.

[21] A. Alsobhi, N. Khan, and H. Rahanu, "An empirical approach to validate the dyslexia adaptive E-learning (DAEL) framework," in *Proc. 10th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Jul. 2015, pp. 253–258, doi: 10.1109/ICCSE.2015.7250252.

[22] M. S. Abtahi, "Interactive multimedia learning object (IMLO) for dyslexic children," *Procedia-Social Behav. Sci.*, vol. 47, pp. 1206–1210, 2012, doi: 10.1016/j.sbspro.2012.06.801.

[23] R. Skiada, E. Soroniati, A. Gardeli, and D. Zissis, "EasyLexia: A mobile application for children with learning difficulties," *Procedia Comput. Sci.*, vol. 27, pp. 218–228, 2014, doi: 10.1016/j.procs.2014.02.025.

[24] P. Tzouveli, A. Schmidt, M. Schneider, A. Symvonis, and S. Kollias, "Adaptive reading assistance for the inclusion of students with dyslexia: The AGENT-DYSL approach," in *Proc. 8th IEEE Int. Conf. Adv. Learn. Technol.*, 2008, pp. 167–171, doi: 10.1109/ICALT.2008.236.

[25] M. Haziq, L. Abdullah, and S. Hisham, "MyLexics: An assistive courseware for dyslexic children to learn basic Malay language," *ACM SIGACCESS Access. Comput.*, no. 95, pp. 3–9, Sep. 2009, doi: 10.1145/1651259.1651260.

[26] S. M. Daud and H. Abas, "'Dyslexia Baca' mobile app–the learning ecosystem for Dyslexic children," in *Proc. Int. Conf. Adv. Comput. Sci. Appl. Technol.*, Dec. 2013, pp. 412–416, doi: 10.1109/ACSAT.2013.87.

[27] M. Rauschenberger, L. Rello, R. Baeza-Yates, and J. P. Bigham, "Towards language independent detection of Dyslexia with a Web-based game," in *Proc. Internet Accessible Things*, 2018, pp. 1–10, doi: 10.1145/3192714.3192816.

[28] E. Ojanen, M. Ronimus, T. Ahonen, "GraphoGame–a catalyst for multi-level promotion of literacy in diverse contexts," *Frontiers Psychol.*, vol. 6, p. 671, May 2015, doi: 10.3389/fpsyg.2015.00671.

[29] U. Richardson and H. Lyytinen, "The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read," *Human Technol., Interdiscipl. J. Humans ICT Environ.*, vol. 10, no. 1, pp. 39–60, May 2014, doi: 10.17011/ht/urn.201405281859.

[30] E. Borleffs, T. K. Glatz, D. A. Daulay, U. Richardson, F. Zwarts, and B. A. M. Maassen, "GraphoGame Si: The development of a technology-enhanced literacy learning tool for standard Indonesian," *Eur. J. Psychol. Educ.*, vol. 33, no. 4, pp. 595–613, Oct. 2018, doi: 10.1007/s10212-017-0354-9.

[31] F. Benmarrakchi, J. E. Kafi, and A. Elhore, "Communication technology for users with specific learning disabilities," *Procedia Comput. Sci.*, vol. 110, pp. 258–265, 2017, doi: 10.1016/j.procs.2017.06.093.

[32] A. Jothi Prabha and R. Bhargavi, "Prediction of Dyslexia from eye movements using machine learning," *IETE J. Res.*, pp. 1–10, Jun. 2019, doi: 10.1080/03772063.2019.1622461.

[33] A. Jothi Prabha and R. Bhargavi, "Predictive model for Dyslexia from fixations and saccadic eye movement events," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105538, doi: 10.1016/j.cmpb.2020.105538.

[34] J. Ashburner and K. J. Friston, "Voxel-based morphometry—The methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000, doi: 10.1006/nimg.2000.0582.

[35] A. Elnakib, A. El-Baz, M. F. Casanova, G. Gimel'farb, and A. E. Switala, "Image-based detection of corpus callosum variability for more accurate discrimination between autistic and normal brains," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4337–4340, doi: 10.1109/ICIP.2010.5652409.

[36] F. Falahati, E. Westman, and A. Simmons, "Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging," *J. Alzheimer's Disease*, vol. 41, no. 3, pp. 685–708, Jul. 2014, doi: 10.3233/JAD-131928.

[37] Y. García Chimeno, B. García Zapirain, I. Saralegui Prieto, and B. Fernandez-Ruanova, "Automatic classification of dyslexic children by applying machine learning to fMRI images," *Bio-Med. Mater. Eng.*, vol. 24, no. 6, pp. 2995–3002, 2014, doi: 10.3233/BME-141120.

[38] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and G. Prisma, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Phys. Therapy*, vol. 89, no. 9, pp. 873–880, 2009, doi: 10.1371/journal.pmed.1000097.

[39] H. Perera, M. F. Shiratuddin, and K. W. Wong, "Review of EEG-based pattern classification frameworks for Dyslexia," *Brain Inform.*, vol. 5, no. 4, pp. 1–14, 2018, doi: 10.1186/s40708-018-0079-9.

[40] K. Spoon, D. Crandall, and K. Siek, "Towards detecting Dyslexia in children's handwriting using neural networks," in *Proc. Int. Conf. Mach. Learn. AI Social Good Workshop*, 2019, pp. 1–5.

[41] K. Spoon, K. Siek, D. Crandall, and M. Fillmore, "Can we (and should we) use AI to detect Dyslexia in children's handwriting?" in *Proc. Artif. Intell. Social Good (NeurIPS)*, 2019, pp. 1–6.

[42] M. Kohli and T. V. Prasad, "Identifying dyslexic students by using artificial neural networks," *world Congr. Eng.*, vol. 1, no. 1, pp. 1–4, 2010.

[43] R. U. Khan, J. L. A. Cheng, and O. Y. Bee, "Machine learning and Dyslexia: Diagnostic and classification system (DCS) for kids with learning disabilities," *Int. J. Eng. Technol.*, vol. 7, no. 3, pp. 97–100, 2018.

[44] S. S. A. Hamid, N. Admodisastro, N. Manshor, A. Kamaruddin, and A. A. A. Ghani, "Dyslexia adaptive learning model: Student engagement prediction using machine learning approach," in *Recent Advances on Soft Computing and Data Mining: Advances in Intelligent Systems and Computing*, vol. 700, R. Ghazali, M. Deris, N. Nawi, and J. Abawajy, Eds. Cham, Switzerland: Springer, 2018, pp. 372–384.

[45] Y. Lakretz, G. Chechik, N. Friedmann, and M. Rosen-Zvi, "Probabilistic graphical models of Dyslexia," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 1919–1928, doi: 10.1145/2783258.2788604.

[46] H. M. Al-Barhamtoshy and D. M. Motaweh, "Diagnosis of Dyslexia using computation analysis," in *Proc. Int. Conf. Inform., Health Technol. (ICIHT)*, Feb. 2017, pp. 1–7, doi: 10.1109/ICIHT.2017.7899141.

[47] H. Perera, M. F. Shiratuddin, K. W. Wong, and K. Fullarton, "EEG signal analysis of writing and typing between adults with Dyslexia and normal controls," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 5, no. 1, p. 62, 2018, doi: 10.9781/ijimai.2018.04.005.

[48] H. Perera, M. F. Shiratuddin, K. W. Wong, and K. Fullarton, "EEG signal analysis of passage reading and rapid automatized naming between adults with Dyslexia and normal controls," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017, pp. 104–108, doi: 10.1109/ICSESS.2017.8342874.

[49] C. W. N. F. Che Wan Fadzal, W. Mansor, K. Y. Lee, S. Mohamad, N. Mohamad, and S. Amirin, "Comparison between characteristics of EEG signal generated from Dyslexic and normal children," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci.*, Dec. 2012, pp. 843–846, doi: 10.1109/IECBES.2012.6498210.

[50] I. Karim, W. Abdul, and N. Kamaruddin, "Classification of Dyslexic and normal children during resting condition using KDE and MLP," in *Proc. 5th Int. Conf. Inf. Commun. Technol. Muslim World (ICT4M)*, Mar. 2013, pp. 4–8, doi: 10.1109/ICT4M.2013.6518886.

[51] A. Frid and Z. Breznitz, "An SVM based algorithm for analysis and discrimination of dyslexic readers from regular readers using ERPs," in *Proc. IEEE 27th Conv. Elect. Electron. Eng. Israel*, Nov. 2012, pp. 1–4, doi: 10.1109/EEEI.2012.6377068.

[52] A. Frid and L. M. Manevitz, "Features and machine learning for correlating and classifying between brain areas and dyslexia," 2018, *arXiv:1812.10622*. [Online]. Available: https://arxiv.org/abs/1812.10622

[53] P. Płoński, W. Gradkowski, A. Marchewka, K. Jednoróg, and P. Bogorodzki, "Dealing with the heterogeneous multi-site neuroimaging data sets: A discrimination study of children dyslexia," in *Brain Informatics and Health, BIH* (Lecture Notes in Computer Science), vol. 8609, D. Slezak, A. Tan, J. F. Peter, and L. Schwabe, Eds. Cham, Switzerland: Springer, 2014, pp. 471–480.

[54] K. Jednoróg, A. Marchewka, I. Altarelli, A. K. Monzalvo Lopez, M. van Ermingen-Marbach, M. Grande, A. Grabowska, S. Heim, and F. Ramus, "How reliable are gray matter disruptions in specific reading disability across multiple countries and languages? Insights from a large-scale voxel-based morphometry study," *Hum. Brain Mapping*, vol. 36, no. 5, pp. 1741–1754, May 2015, doi: 10.1002/hbm.22734.

[55] P. Płoński, W. Gradkowski, I. Altarelli, K. Monzalvo, M. van Ermingen-Marbach, M. Grande, S. Heim, A. Marchewka, P. Bogorodzki, F. Ramus, and K. Jednoróg, "Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia," *Hum. Brain Mapping*, vol. 38, no. 2, pp. 900–908, Feb. 2017, doi: 10.1002/hbm.23426.

[56] Z. Cui, Z. Xia, M. Su, H. Shu, and G. Gong, "Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach," *Hum. Brain Mapping*, vol. 37, no. 4, pp. 1443–1458, Apr. 2016, doi: 10.1002/hbm.23112.

[57] S. Zahia, B. Garcia-Zapirain, I. Saralegui, and B. Fernandez-Ruanova, "Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging," *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, Art. no. 105726, doi: 10.1016/j.cmpb.2020.105726.

[58] O. L. Usman and R. C. Muniyandi, "CryptoDL: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network," *Symmetry*, vol. 12, no. 5, p. 836, 2020, doi: 10.3390/sym12050836.

[59] L. Rello and M. Ballesteros, "Detecting readers with dyslexia using machine learning with eye tracking measures," in *Proc. 12th Web Conf.*, May 2015, pp. 1–8, doi: 10.1145/2745555.2746644.

[60] L. Rello, E. Romero, M. Rauschenberger, A. Ali, K. Williams, J. P. Bigham, and N. C. White, "Screening dyslexia for english using HCI measures and machine learning," in *Proc. Int. Conf. Digit. Health*, Apr. 2018, pp. 80–84, doi: 10.1145/3194658.3194675.

[61] M. N. Benfatto, G. Seimyr, J. Ygge, T. Pansell, A. Rydberg, and C. Jacobson, "Screening for dyslexia using eye tracking during reading," *PLoS ONE*, vol. 11, no. 12, 2016, Art. no. e0165508, doi: 10.1371/journal.pone.0165508.

[62] Z. Rezvani, M. Zare, M. Bonte, V. Der Molen, and G. González, "Machine learning classification of dyslexic children based on EEG local network features," *BioRxiv Prepr. Serv. Biol.*, pp. 1–23, Mar. 2019, doi: 10.1101/569996.

[63] Z. Cui, S. Zhong, P. Xu, Y. He, and G. Gong, "PANDA: A pipeline toolbox for analyzing brain diffusion images," *Frontiers Human Neurosci.*, vol. 7, pp. 1–16, 2013, doi: 10.3389/fnhum.2013.00042.

[64] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. J. Behrens, "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data," *NeuroImage*, vol. 31, no. 4, pp. 1487–1505, Jul. 2006, doi: 10.1016/j.neuroimage.2006.02.024.

[65] L. Xing and Y. Qiao, "Deepwriter: A multi-stream deep CNN for text-independent writer identification," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 584–589.

[66] T. Asvestopoulou, V. Manousaki, A. Psistakis, I. Smyrnakis, V. Andreadakis, I. M. Aslanides, and M. Papadopouli, "DysLexML: Screening tool for dyslexia using machine learning," 2019, *arXiv:1903.06274*. [Online]. Available: http://arxiv.org/abs/1903.06274

[67] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020, doi: 10.1109/access.2020.3029154.

[68] P. Patronik and S. J. Piestrak, "Design of RNS reverse converters with constant shifting to residue datapath channels," *J. Signal Process. Syst.*, vol. 90, no. 3, pp. 323–339, Mar. 2018, doi: 10.1007/s11265-017-1238-6.

[69] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable machine learning–a brief history, State-of-the-Art and challenges," 2020, *arXiv:2010.09337*. [Online]. Available: http://arxiv.org/abs/2010.09337 and [Online]. Available: https://www.infoq.com/presentations/ml-interpretability-lime

[70] F. Doshi-Velez and B. Kim, "Considerations for evaluation and generalization in interpretable machine learning," in *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Cham, Switzerland: Springer, 2018, pp. 3–17.

[71] A. Mahmood, T. Hamed, C. Obimbo, and R. Dony, "Improving the security of the medical images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 9, pp. 137–146, 2013, doi: 10.14569/ijacsa.2013.040922.

[72] N. Dowlin, R. Gilad-Bachrach, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, 2016, pp. 342–351. [Online]. Available: http://proceedings.mlr.press/v48/gilad-bachrach16.pdf%0A and http://research.microsoft.com/apps/pubs/?id=260989

[73] A. Khan, A. Sohail, U. Zahoora, and A. Saeed Qureshi, "A survey of the recent architectures of deep convolutional neural networks," 2019, *arXiv:1901.06032*. [Online]. Available: http://arxiv.org/abs/1901.06032

[74] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[75] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, and H. Kiya, "Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1–5. [Online]. Available: http://arxiv.org/abs/1905.01827.

[76] M. A. Rahman and R. C. Muniyandi, "Review of GPU implementation to process of RNA sequence on cancer," *Informat. Med. Unlocked*, vol. 10, pp. 17–26, 2018, doi: 10.1016/j.imu.2017.10.008.

[77] A. Maroosi and R. C. Muniyandi, "Accelerated execution of P systems with active membranes to solve the N-queens problem," *Theor. Comput. Sci.*, vol. 551, pp. 39–54, Sep. 2014, doi: 10.1016/j.tcs.2014.05.004.

[78] A. Maroosi, R. C. Muniyandi, E. Sundararajan, and A. M. Zin, "Parallel and distributed computing models on a graphics processing unit to accelerate simulation of membrane systems," *Simul. Model. Pract. Theory*, vol. 47, pp. 60–78, Sep. 2014, doi: 10.1016/j.simpat.2014.05.005.
[79] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Inform.*, vol. 7, p. 11, Oct. 2020, doi: 10.1186/s40708-020-00112-2.
[80] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018.
[81] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019, doi: 10.1016/j.zemedi.2018.11.002.
[82] H. M. Ali, M. S. Kaiser, and M. Mahmud, "Application of convolutional neural network in segmenting brain regions from MRI data," in *Brain Informatics, BI* (Lecture Notes in Computer Science), vol. 11976, P. Liang, V. Goel, and C. Shan, Eds. Cham, Switzerland: Springer, 2019, pp. 136–146.
[83] G. Aceto, D. Ciuonzo, and A. Montieri, "MIMETIC: Mobile encrypted traffic classification using multimodal deep learning," *Comput. Netw.*, vol. 165, Dec. 2019, Art. no. 106944, doi: 10.1016/j.comnet.2019.106944.
[84] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Toward effective mobile encrypted traffic classification through deep learning," *Neurocomputing*, vol. 409, pp. 306–315, Oct. 2020, doi: 10.1016/j.neucom.2020.05.036.
[85] M. Mahmud, M. S. Kaiser, T. M. McGinnity, and A. Hussain, *Deep Learning in Mining Biological Data*. New York, NY, USA: Springer, 2020.

OPEYEMI LATEEF USMAN received the B.Sc. degree and master's degree in computer science education from the Tai Solarin University of Education, in 2009 and 2014, respectively. He is currently pursuing the Ph.D. degree in computer science with the Research Centre for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia. Since 2011, he has been an Academic Staff of the Computer Science Department, Tai Solarin University of Education, Ogun State, Nigeria. He has good number of publications in reputable journals both locally and internationally. His research interests include neural networks, deep learning, predictions, residue number systems, and security.

RAVIE CHANDREN MUNIYANDI received the M.I.T. degree and the Ph.D. degree in computer science from the Universiti Kebangsaan Malaysia (UKM), in 1997 and 2011, respectively. He is currently an Associate Professor with the Research Center for Cyber Security (SIBER), Faculty of Information Science and Technology, UKM. He has authored over 90 articles and above 400 citations. His research interests include programming technologies, software security, bio-inspired computing, machine learning, and membrane computing.

KHAIRUDDIN OMAR received the B.Sc. and M.Sc. degrees in computer science from the Universiti Kebangsaan Malaysia (UKM), in 1986 and 1989, respectively, and the Ph.D. degree from the Universiti Putra Malaysia, in 2000. He is currently a Professor with the Research Center for Artificial Intelligence Technology (CAIT), Pattern Recognition Research Group, Faculty of Information Science and Technology, UKM. He has authored over 250 articles and over 2200 citations. His research interests include artificial intelligence, pattern recognition, image processing, machine learning, and feature extraction.

MAZLYFARINA MOHAMAD received the Ph.D. degree in biomedical science from the Universiti Kebangsaan Malaysia. She is currently a Senior Lecturer with the Center for Diagnostic, Therapeutic and Investigative Studies, Faculty of Health Sciences, Universiti Kebangsaan Malaysia. She has many publications to her credit both locally and internationally. Her research interests include cognitive neuroscience, functional MRI analysis, and medical physics.