



Exploration and analysis of On-Surface and In-Air handwriting attributes to improve dysgraphia disorder diagnosis in children based on machine learning methods

Jayakanth Kunhoth^{*}, Somaya Al Maadeed, Moutaz Saleh, Younes Akbari

Department of Computer Science and Engineering, Qatar University, Doha, Qatar

ARTICLE INFO

Keywords:

Learning disabilities
Dysgraphia
Handwriting
Screening tool
Machine learning

ABSTRACT

Dysgraphia is a type of learning disorder that affects children's writing skills. Poor writing skills can obstruct students' academic growth if it is undiagnosed and untreated properly in the early stages. The irregularity in the symptoms and varying levels of difficulty at each age level made the dysgraphia diagnosis task quite complex. This work focuses on developing machine learning-based automated methods to build the dysgraphia screening tool for children. The proposed work analyzes the various attributes of online handwritten data recorded by digitizing tablets during On-Surface (when the pen is on the tablet's surface) and In-Air activity (when the pen is away from the tablet's surface). The proposed work has considered feature extraction from the whole handwriting data in a combined manner instead of feature extraction from task-specific (word, letter, sentence, etc.) handwritten data separately to reduce the number of features. This approach has significantly reduced the number of features by about 85%. Extracted features are used to train and evaluate multiple machine learning classifiers such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random forest, and AdaBoost. Evaluation in a publicly available dataset indicates that the AdaBoost classifier achieved a classification accuracy of 80.8%, which is 1.3% more than the state-of-the-art method. Moreover, a deep analysis of different characteristics (kinematic, dynamic, temporal, spatial, etc.) of online handwriting is conducted to examine their significance in distinguishing normal and abnormal handwritten data. The analysis can help psychologists determine what attributes and methods should be considered for effective treatment.

1. Introduction

Learning disabilities are a broad generic categorization of learning disorders that hamper the skill acquisition activity of a human being. Learning disabilities and lack of intelligence are often got confused as similar. However, both are disparate, and studies have shown that human beings with learning disabilities possess no less than an average level of intelligence. Although self-intelligence is not that much affected due to learning disabilities, the latter can cause a lack of self-esteem and self-confidence in human beings, especially in children. Children are probably looking to learn novel information and acquire new skills day by day when they grow. Learning disabilities hinder or delay this skill acquisition in one way or another, resulting in children struggling to understand new things. This situation will make them feel inferior to their colleagues. And it can result in reduced self-esteem, confidence, and social-emotional and behavioral concerns. Students with learning disabilities account for a significant proportion of the "special educational needs" category [1]. Learning disabilities damage or harm the perception ability of the student and cause difficulties in one or many

tasks, including reading, writing, or doing math, etc. Dysgraphia is a learning disability mainly regarded as disarray in written expression. Dysgraphia can affect not only handwriting but also grammar, spelling, organization, etc. [2]. The available statistics about learning disabilities show that about 10%–30% of children in the world face difficulties in handwriting.

Usually, team-based assessments are conventionally practiced dysgraphia diagnosis methods in children. The team-based assessments include multiple specialists from different domains, such as education (teacher), psychology (occupational therapist), medicine (speech therapists, ophthalmologists), etc. These specialists jointly analyze the student's handwriting ability and other factors that can affect handwriting. Because the existence of some prevalent conditions (which can cause handwriting impairments), such as hearing loss, poor vision, lack of intelligence, or lack of training, should be ruled out before subjecting the child to the actual dysgraphia assessment. It is significant to contemplate various contributing factors of dysgraphia, such as speed and legibility of writing, inconsistency between spelling, ability,

^{*} Corresponding author.

E-mail address: j.kunhoth@qu.edu.qa (J. Kunhoth).

<https://doi.org/10.1016/j.bspc.2023.104715>

Received 17 November 2022; Received in revised form 31 January 2023; Accepted 14 February 2023

Available online 20 February 2023

1746-8094/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and verbal intelligence quotient, as well as how the pencil grip is held while writing and writing pose to assess the condition of the student effectively. And there is a lack of standard medical assessment methods for examining the existence of dysgraphia. Concise Evaluation Scale for Children's Handwriting (BHK) for French [3], Detailed Assessment of Speed of Handwriting (DASH) in Latin [4] are the commonly used popular standard assessments for dysgraphia diagnosis.

It is challenging to accurately diagnose any learning disorder since the assessment procedure has to consider or take into account multiple cues. The cues or symptoms of dysgraphia differ with the child's age. The symptoms may vary at each development stage or age range. More crucially, the predefined identifiers or indication should continue for at minimum six months with parallel intercession actions being administered [5]. The behavior of dysgraphia is often complex. Sometimes the child will have only dysgraphia, but in the worst case, comorbidity is observed. Children may have autism spectrum disorder (ASD), developmental coordination disorder (DCD) [6], or attention deficit hyperactivity disorder (ADHD) along with dysgraphia. It makes the assessment procedure bit difficult. It yields the need for early assessment and diagnosis of dysgraphia. Identification and intervention in the early stages can significantly decrease the effort and time required to treat the disorders.

The decision of the manual dysgraphia assessments performed by a team of specialists is merely based on the score or judgment of the handwritten product. Human bias and other factors (experience of the specialists) can positively or negatively affect the final results to an extent. Furthermore, manual assessments are time-consuming and require a lot of human resources. These limitations in manual assessments paved the way for the development of automated systems for dysgraphia diagnosis. The automated systems are focused on statistically analyzing the characteristics of handwriting acquired by digitizing tablets. Current technological advancements in the mobile industry have led to the development of tablets with built-in capabilities to extract various handwriting features/raw data. It includes the position of the pen tip, On-Surface/In-Air pen position, pen tip pressure, the azimuth angle of the pen to the tablet surface, the tilt of the pen, timestamp [7]. Multiple works in the literature have utilized digitizing tablets and information acquired by them to identify dysgraphia with the help of machine learning algorithms. Apart from methods based on digitizing tablet-based data acquisition and analysis of online handwritten features, handwritten image analysis-based methods [8] are proposed in the literature for dysgraphia screening.

Although dysgraphia diagnosis involves complex and multiple approaches, several machine learning-based automated diagnosis techniques have been proposed in the literature in recent years. Mekyska et al. proposed methods for screening the normally developing handwriting and dysgraphia handwriting using machine learning algorithms. Kinematics, dynamics, and non-linear dynamics of handwritten products are extracted to form the feature vectors for training the machine learning classifiers. Random forest and linear discriminant analysis algorithms are used on the extracted multiple features to discriminate the normal and abnormal handwritings with a sensitivity of 96%. The dysgraphia diagnosis system proposed in the [9] utilizes machine learning methods such as random forest, logistic regression, and naïve Bayes to classify the extracted online handwritten features. The significant features extracted from the handwritten data are slant, pressure, letter, word spacing, etc. In [10], Dimauro et al. proposed a software system for dysgraphia diagnosis by utilizing various document analysis algorithms and their modifications. The proposed approach has semiautomated the BHK test by automatically evaluating the nine relevant handwriting characteristics (writing size, skewed writing, letter spacing, etc.) along with the manual assessment of the other four relevant characteristics. The output of the software system generates a score similar to the BHK scale. The framework proposed in [11] uses image processing and pattern recognition on handwritten text samples for the automated detection of dyslexia symptoms. Although referred

to as dyslexia, only the writing aspect is considered in the work that yielded 73% classification accuracy.

The studies and experiments conducted in the literature [12–19] have shown that stroke dimensions, velocity, acceleration, jerk, pressure, tilt, temporal, azimuth angle, and the number of pen elevations have less or equal importance in discriminating the normally developing handwriting and dysgraphia. The most relevant features spotted in several works are kinematics of writing and pen tip pressure. Comparative analysis of relevant ML-based dysgraphia screening systems proposed in recent years is provided in Table 1.

Compared to the approaches merely based on handwriting images (offline), the tablet-based techniques (online) could explore more characteristics of handwriting, which turned out to be significant for the detection of dysgraphia [26]. The data acquisition in the latter approach involved writing with a standard pen or an electronic pen on paper overlaid on the tablet. The number of features used in the literature for dysgraphia diagnosis in most works is enormous (1000 or more). The features extracted from the handwritten data can generally be classified into two types based on the activities during handwriting: when the pen is touching the surface of the tablet or On-Surface activity (On-Surface features) and when the pen is away from the surface in between writing or In-Air activity (In-Air features). Most of the work in the literature has utilized features from two types of activities for classifying handwriting. However, a study on the significance of each activity's specific features or comparative analysis is not conducted in the literature, especially for the dataset we used to evaluate our method. Kinematic, dynamic, spatial, and temporal are the popular handwriting attributes/characteristics considered in the literature for distinguishing dysgraphia writing and normal writing [26]. However, no works in the literature have analyzed the significance of each attribute for distinguishing dysgraphia writing and normally developing writing. This work tries to address all these issues, and the contribution of the work is as follows:

- Design and develop a machine learning-based dysgraphia diagnosis system for preliminary assessment.
- Examine the effectiveness of different supervised machine learning algorithms (the ones which can handle nonlinear data, ranging from simple to kernel-based methods and ensemble learning methods) for classifying the online handwriting features.
- Analyze the effectiveness of On-Surface features alone, followed by the analysis of feature combination of On-Surface and In-Air features for dysgraphia diagnosis problem.
- Analyze the potential or significance of different attributes of handwriting activity or different categories of online handwritten features (Kinematic, dynamic, spatial, and temporal attributes for On-Surface activity, Kinematic and Temporal for In-Air activity) for discriminating the normally developing handwriting and dysgraphia.
- Develop methods that utilize fewer features than state-of-the-art method (to reduce the computational overhead) without compromising the classification performance. The proposed work has considered extracting features from the whole handwriting data combined instead of feature extraction from task-specific (word, letter, sentence, etc.) handwritten data separately to reduce the number of features.
- Propose a framework to assist psychologists in treating children with dysgraphia.

The rest of this paper is organized as follows; Section 2 presents the methodology of the proposed work. It explains how we built the proposed method, including all the details of the algorithm and the dataset used. Section 3 presents the results and findings. A discussion and new framework for assisting psychologists in treating students with dysgraphia are provided in Section 4. Finally, Section 5 presents the conclusion.

Table 1
Comparative analysis of state of the art dysgraphia classification methods.

Ref.	Age group	Subjects	Data collection task	Features	Classifiers	Performance
[12]	Grade 3	Total: 54 (27 positives)	Writing sequence of letters	Kinematic , 34 nonlinear dynamic, other 7 Features	Linear discriminant analysis, Random forest	Recall : 96%
[13]	Primary school	Total 298 (56 positives)	Copying the text	Static, Kinematic, Pressure, Tilt	Random Forest	Recall : 96.5%
[11]	Primary school	–	Images of handwritten text	OCR,MSER	ANN	Accuracy : 71%
[14]	Grade 3 to 4	Total:76 (15 positives)	Drawing few patterns/figures	Spatial, temporal, kinematic, dynamic, other — pen elevations and relative number of interruptions	XG-Boost	Specificity : 90%
[15]	Grade 3 to 4	Total:65 (33 positives)	Copy a short paragraph	Kinematic, temporal, spatial, and dynamic	Support vector machine and Random forest	Recall : 88%
[16]	Age:3	Total:104 (28 positives)	Draw shapes and symbols, questionnaire	Gesture smoothness, pressure(mean value), drawing kinematics	Logistic regression	Area under curve : 0.82
[17]	Age: 9 years	Total:280 (62 positives)	–	Static, kinematic, pressure, and tilt	linear regression	–
[18]	Age: 8 to 15	Total: 120 (57 positives)	write letters, words, sentence with varying speed	Dynamic, Spatiotemporal and kinematic features	AdaBoost	Accuracy : 79.5%
[20]	Age: 8 to 9	Total:90 (49 positives)	Writing write letters, words, and sentence	Spatiotemporal , dynamic, kinematic and other features	SVM	Accuracy : 90%
[21]	Grade: 3 to 6	Total:32	writing and drawing task	Spatial, temporal , dynamic and other features	SVM	Accuracy : 82.51%
[22]	Age: 10 to 13	Total:72 (36 positives)	Write letters, words, and sentence	Spatial, temporal , dynamic, kinematic and other features	SVM	Sensitivity : 75.5%
[23]	Age: 7 to 8	Total:40	drawing lines, tracing figures , encircling matching shapes	Not explicitly mentioned	Decision tree	Not mentioned
[24]	Age: 7 to 12	Total:60	Drawing and writing tasks	Spatiotemporal, dynamic and kinematic features	Random forest	Recall: 92.85%
[25]	Age: 8 to 15	Total: 120 (57 positives)	Writing tasks	Convolutional features	Convolutional neural network	Accuracy: 79.7%

2. Methodology

This section outlines the details of the dataset and methodology used to design and build the dysgraphia diagnosis classifiers. Fig. 1 provides an overview regarding how a handwriting analysis-based dysgraphia diagnosis system is developed and its workflow.

The major steps involved in designing a machine learning-based dysgraphia diagnosis system are conducting handwriting experiments to collect the raw data and extracting relevant features which outline the dynamic, kinematic, temporal, and spatial characteristics of raw handwritten data. Usually, children are advised to write certain things, such as letters, words, and sentences, or draw specific patterns on the tablet using a stylus pen. Recent technological advancements enable us to record the writing's trajectory, time, and dynamics using the digitized tablet. When a child writes on the tablet, the time, x , and y positions of the pen tip on the surface of the tablet, the pressure exerted by the pen tip on the tablet surface, altitude, and azimuth angle are recorded for further processing. Later the extracted features are stacked together to form feature vectors representing each individual's unique characteristics. Machine learning algorithms are trained using the extracted feature vectors to develop classifiers, and their performance is evaluated.

2.1. Dataset

The main challenge researchers face in developing the automated dysgraphia diagnostic system or proposing classification methods is the unavailability of raw data. Data is the backbone of machine learning-based decision systems, and sufficient data is required for algorithms to understand the unique underlying patterns. Although there are several handwritten datasets for the diagnosis of dysgraphia in the literature, only very few are accessible to the public. In this work, a dataset [18], which includes handwritten data from 120 children, was used to assess the performance of the proposed approaches. Children involved in data collection were between 8 and 15 years of age. Among 120 participants, 57 have symptoms of dysgraphia, and the remaining have normally developing handwriting. During data collection, the participants were asked to complete different handwriting tasks, including writing letters, words, and sentences on the tablet. The ground truth of the data is identified with the help of trained professionals and certified clinical therapists. The data is currently publicly available [27]. We retrieved this publicly available data for further analysis. For each individual, the raw data provides the x and y positions of the written expression and their respective time, pressure, azimuth, and altitude information. The dataset contains text data obtained using Slovak orthography. Fig. 2 shows the sample data from the dataset.

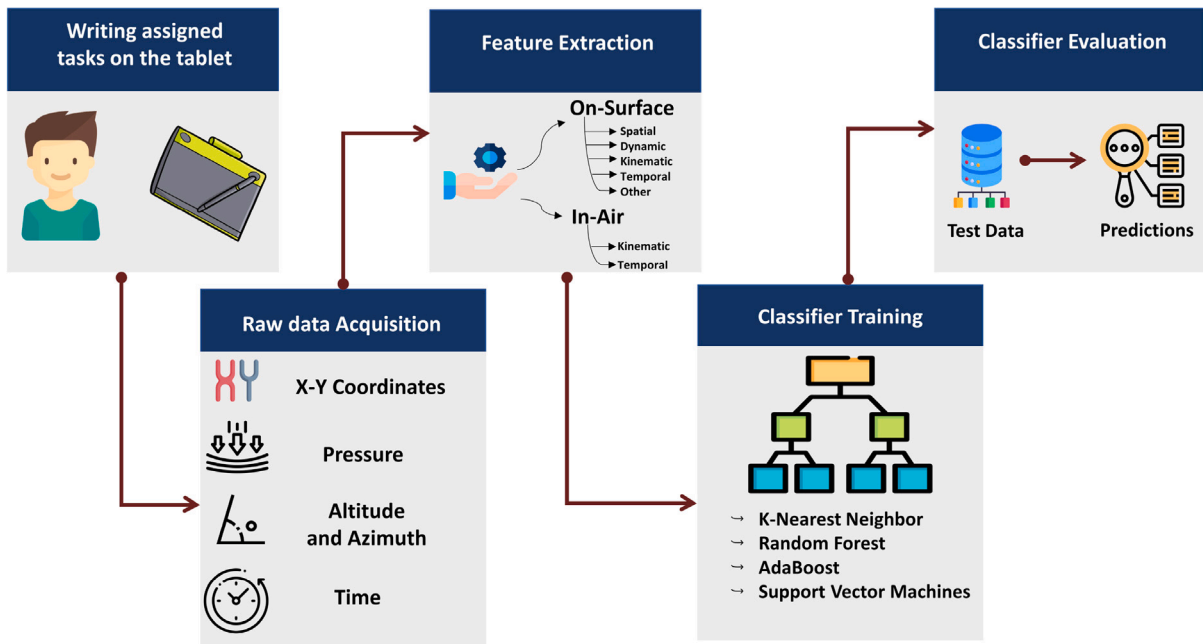


Fig. 1. Dysgraphia diagnosis system development : Workflow.

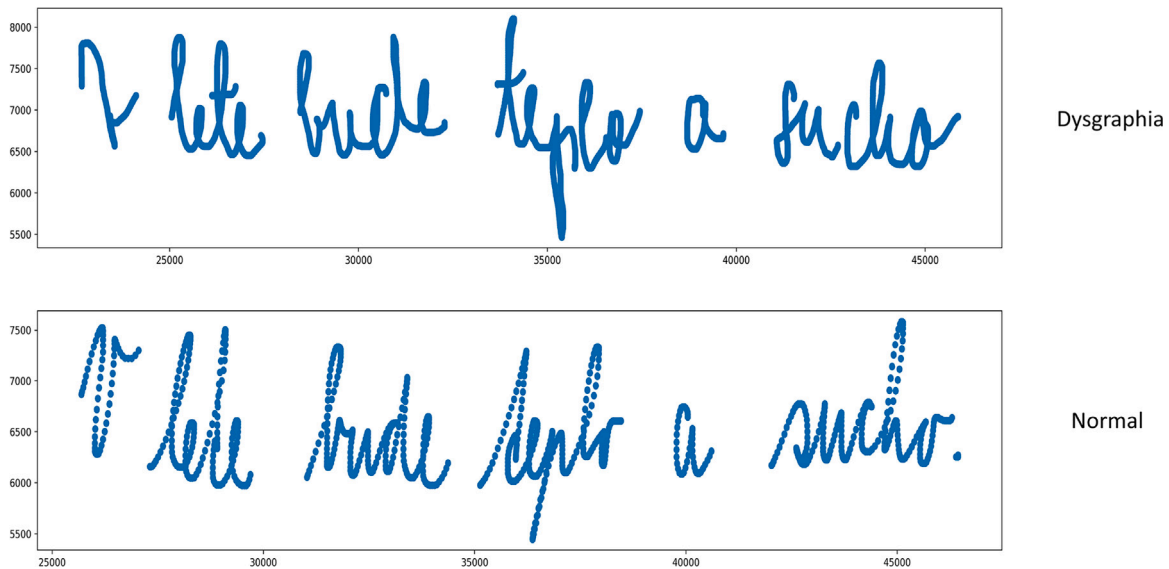


Fig. 2. Samples from the dataset.

2.2. Feature extraction

Feature extraction is a pivot task while developing decision-making systems using traditional machine learning algorithms. The handwriting task will generate seven raw data values (x position, y position, time, pen position indicator, azimuth, altitude, pressure) for each instance. These raw data values must be transferred to meaning full value or vectors before training the machine learning classifier. In this work, we extracted 175 features from the raw data for each subject. Although the raw data available in the dataset represents multiple tasks (letter writing, word writing, sentence writing), we considered them a single handwriting task while extracting the feature. It means the features are not extracted for each task separately. Instead, the features are extracted from the combined handwritten content for each individual. This approach will decrease the number of features and reduce the computational overhead while training the classifiers and during prediction.

Detailed descriptions of the extracted features are provided in [Table 2](#). Among 175 extracted features, 119 are for On-Surface activity and 56 for In-Air activity.

2.2.1. On-Surface features

On-Surface features refer to the features extracted from the recorded data points when the stylus pen is touching the tablet’s surface. In other words, the feature extracted during the actual writing task. Five categories of On-Surface features are extracted from the data: kinematic, dynamic, temporal, spatial, and other. Dynamic features measure the characteristics of pressure exerted by the person on the surface while writing, characteristics of inclination of the pen, etc., while writing on the tablet. Spatial and temporal features extract the uniqueness associated with the space and time data. It means the size of the writing, letter or words is considered in the spatial data, and the time taken for writing activity is considered in the temporal data. Usually,

Table 2
Feature descriptions.

Category	Feature	Description	No. of features
Dynamic	Pressure	Pressure between the pen tip and tablet surface (Mean, Median, Max, Min, SD, 95th and 5th percentile)	7
	Azimuth	Angle of the pen to the vertical axis (Mean, Median, Max, Min, SD, 95th and 5th percentile)	7
	Altitude	Angle of the pen to the horizontal axis (Mean, Median, Max, Min, SD, 95th and 5th percentile)	7
Spatial	Stroke Length	Magnitude of length of stroke	5
	Horizontal Stroke Length	Stroke length in the horizontal direction (Mean, Median, Maximum, Minimum, SD)	5
	Vertical Stroke Length	Stroke length in the vertical direction (Mean, Median, Maximum, Minimum, SD)	5
	Y position of segments	Difference between y position of first and last stroke, second and penultimate stroke. Variance of y position of strokes (Mean, Minimum, Median, Maximum)	12
	Length	Length of the whole writing movement	1
	Vertical Length	Length of the whole writing movement in vertical direction	1
	Horizontal Length	Length of the whole writing movement in horizontal direction	1
	Length in vertical direction	(Mean, Median, Maximum, Minimum, SD)	5
	Length in the horizontal direction	(Mean, Median, Maximum, Minimum, SD)	5
	Temporal	Stroke time	Time taken to complete a stroke(Mean, Median, Maximum, Minimum, SD)
writing time		Time taken to complete whole writing task	1
On-Air time		Idle time between writing activity(Mean, Median, Maximum, Minimum, SD)	5
Kinematic	Velocity, vertical /horizontal	Change in position of stylus pen over time. Also, velocity is computed in both vertical and horizontal directions separately. (Mean, Median, Maximum, Minimum, SD, 95th percentile, 5th percentile)	21
	Acceleration	Change in writing speed over time (Mean, Median, Maximum, Minimum, SD, 95th percentile, 5th percentile)	7
	Jerk, vertical /horizontal	Change in acceleration while writing over time. Also, jerk is computed in both vertical and horizontal directions separately. (Mean, Median, Maximum, Minimum, SD, 95th percentile, 5th percentile)	21
	Extrema of velocity and acceleration	Number of extrema in velocity and acceleration values	2
Other	Pen Lift	Number of interruptions during writing	1

the subject with dysgraphia may take more time to complete the task. Furthermore, the spatial appearance of the written character will be worse in the case of dysgraphic subjects. Kinematic features explore the kinetic property of the writing, which include speed, acceleration, jerk etc.

2.2.2. In-Air features

In-Air features refer to the features extracted from the recorded data points when the stylus pen is not touching the tablet's surface. The In-Air features is extracted from the raw data representing the activity between writing each stroke, segment, or word. Two categories of In-Air features were extracted from the data. It includes 51 kinematic features (as mentioned in Table 2, but the difference is that it is computed when the pen is away from the tablet surface) and five temporal features (statistical values of On-Air time). On-Air time measures the time the student was not writing (or thinking about writing) from starting the writing task until completing it. Children with dysgraphia may take more time to think or may need more time to start a new stroke/segment after finishing the current stroke.

2.3. Machine learning classifiers

Machine learning classification algorithms are required to train and develop prediction models using the extracted features. The prediction model will be capable of classifying the extracted handwritten features into either normal or dysgraphia classes. From the data analysis perspective, this is a binary classification problem (supervised learning). Many supervised machine learning algorithms are available for binary classification problems ranging from simple KNN algorithms to complex deep neural network algorithms. Compared to traditional algorithms, deep neural network algorithms are data-hungry. Even though the selection of machine learning algorithms in this work is somewhat random and based on our previous experience, specific criteria were considered. The algorithms that can directly handle the non-linear data points and algorithms with different complexity levels are considered.

Furthermore, the performance of the machine learning algorithms needs to be analyzed (with different underlying working principles) for classifying the handwritten data. In that context, four supervised learning algorithms such as K-Nearest Neighbor (KNN) [28], Support Vector

Table 3

Definitions of true positives, true negatives, false positives and false negatives.

Actuals	Predictions	
	True	False
True	TP	FN
False	FP	TN

Machine (SVM) [29], RandomForest(RF) [30], and AdaBoost(AB) [31] are used in this work. These algorithms are prevalent and have been utilized in the literature for wide applications such as image classification, audio classification, text classification, etc.

KNN is a simple non-parametric algorithm based on computing the nearest neighbors by considering the distance metrics. KNN is included to examine how a simple algorithm works with handwriting features for dysgraphia diagnosis problems. KNN and decision trees are the popular simple machine-learning algorithms that can handle non-linear data points. KNN was considered above decision trees in this work since the latter will be used as base learners in Random forest and AdaBoost algorithms. SVM is a more complex algorithm, and its kernel feature enables it to handle non-linear data. Apart from simple and complex (kernel-based) algorithms, ensemble learning algorithms are considered. Ensemble learning algorithms are way more complex and require a longer training time. However, ensemble learning can reduce either variance or bias to an extent. Random Forest and Adaboost are the popular ensemble learning classifiers that constitute multiple decision trees. The random forest just integrates numerous decision trees to reduce the variance. Random forests are also known as bagged trees, meaning each tree is independent and can be implemented in parallel. On the other hand, AdaBoost, or adaptive boosting, implements a boosting approach to integrate multiple decision tree classifiers. In AdaBoost, the succeeding classifiers are modified based on the error in the preceding classifiers.

3. Evaluation and results

All of the experiments were carried out in Python. The classification models are built and trained on a laptop with configuration Intel(R) Core (TM) i7-7820HK CPU operating at 2.90GHZ (2901 MHz) with four cores and eight logical processors. The Sci-Kit library is used for training and evaluating the machine learning classifiers.

Evaluation of any model is critical to recognize how good the built model is. The commonly used evaluation metrics for multi-label classification methods are accuracy, precision, recall, and F1 score. First and foremost, four essential parameters for computing the evaluation metrics are true positives(TP), false positives(FP), true negatives(TN), and false negatives(FN). Table 3 defines TP, FP, TN, and FN.

Accuracy represents the percentage of correct predictions made by the classifier. In other words, it is the ratio between the number of accurate predictions and total predictions made by the classifier in percentage.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

In addition to accuracy, other evaluation metrics are used to measure the model's performance. Precision is the ratio of the true positives to total predictions.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall is the ratio of true positives to total true positives in the test data. Recall measures the sensitivity of the model.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Table 4

Hyperparameter configuration of classifiers.

Algorithm	Hyperparameters	Feature set	
		On-Surface	On-Surface + In-Air
KNN	No. of nearest neighbors , N	11	5
SVM	C	1	100
	gamma	1	0.01
	kernel	RBF	RBF
Random forest	Splitting criterion	Entropy	Gini
	No. of estimators	100	200
	Maximum depth	5	10
	Minimum samples leaf	5	5
	Minimum samples split	10	5
AdaBoost	Learning rate	0.1	0.5
	No. of estimators	120	265

F1 score is the weighted average of recall and precision. It considers both false negatives and false positives, similar to accuracy. In the unsymmetrical dataset, the F1 score is a better evaluation metric than the accuracy.

$$F1score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (4)$$

Moreover, the receiver operating characteristic (ROC) plot and the area under the ROC curve (AUC_ROC) are used to evaluate the performance of the proposed methods. ROC is the plot of the prediction where the y-axis represents sensitivity/recall/true positive rate, and x-axis represents false positive rate.

The whole experiment is categorized into two sections. The first section focused on comparing and contrasting the performance of On-Surface features with the On-Surface features + In-Air features combination. And they are followed by the comparison of the proposed methods with the state-of-the-art techniques. So, initially, the machine learning classifiers are trained with On-Surface features, and performance is recorded. Later the feature vectors are extended by adding the In-Air features with the On-Surface features. Training and evaluation of machine learning classifiers are conducted with new feature vectors, and performance is recorded. Some of the features among the new feature vectors derived from the raw data have higher magnitude values than other derived features. Training and evaluating the machine learning classifiers using the features in this form may negatively affect the performance. Because the classification algorithm may be biased towards features with higher magnitude.

Since machine learning algorithms like SVM, KNN, and neural networks utilize the spatial relationship between the data samples to generate the decision boundary, they are sensitive to feature sets with a wide magnitude range. So the features have to be pre-processed (feature scaling) before training the classifiers to avoid this issue. On the other hand, machine learning algorithms such as decision trees, AdaBoost, Random forests, etc., are not affected since their working principle or decision boundary generation principle is different. Either without scaling or with scaling, the tree-based algorithms will perform normally. This work uses tree-based algorithms and other algorithms such as SVM and KNN, to develop classifiers. The Min-Max scaling method is utilized to scale the feature values to a specific range. To optimize the performance as well as reduce the selection bias, the hyperparameters of all the classifiers are tuned. Each classifier was trained and evaluated with the best hyperparameter sets. The training and evaluation followed 10-fold cross-validation. The details of hyperparameters used in each classifier are shown in Table 4. The performance of each classifier with On-Surface features and On-Surface features + In-Air is provided in Fig. 3.

In Fig. 3, accuracy, precision, recall, and f1-score of the developed machine learning classifiers are provided. A total of 8 classifiers are

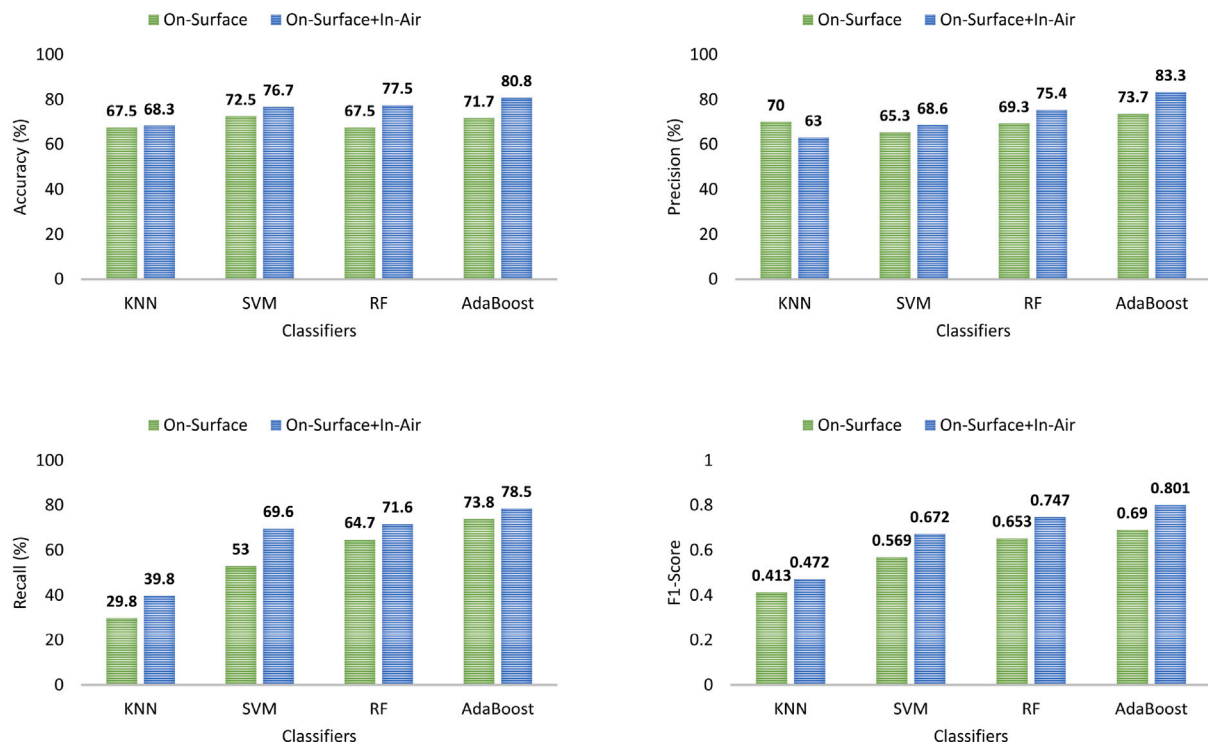


Fig. 3. Performance of the proposed methods, left top: Accuracy, right top : Precision, left bottom : recall, right bottom: F1-Score.

built and evaluated, where four of them are trained with On-Surface features while the rest are trained with On-Surface + In-Air features. With On-Surface features, the SVM classifier has generated the best classification performance in terms of accuracy (72.5%). But accuracy alone cannot be considered for a problem like medical diagnosis. More significant metrics like precision and recall must be considered in this case. Precision indicates how many are correctly predicted as positive from all the positive predictions. When the precision is higher, the false positive rate is less (falsely classified as positive).

Similarly, we also have to consider the false negatives (falsely classified as negatives). In preliminary diagnosis, false negatives are more or equally significant than false positives. The recall value metric will take account of false negatives. Considering precision and recall, it is clear that SVM classifiers are not good as AdaBoost or RF classifiers when trained with On-Surface features. The AdaBoost classifier has recorded a maximum precision of 73.7% and a recall of 73.8%. In some instances, classifiers can have higher precision but less recall or vice versa. So better metrics like the F1-score are used for analyzing the classifier's performance. F1-score computes the harmonic mean of precision and recall. In terms of F1-score also AdaBoost classifier surpassed all other classifiers trained with On-Surface features.

By analyzing the accuracy of each classifier, it is comprehensible that adding In-Air features along with On-Surface features has significantly improved the classification performance. It shows the significance of In-Air features for classifying normally developing and dysgraphia writing. In On-Surface + In-Air features set, the performance of SVM, AdaBoost, and RF classifiers has drastically increased. Among all the proposed methods, AdaBoost with On-Surface + In-Air feature set has yielded the best accuracy of 80.8%, which is acceptable. The superiority of the AdaBoost classifier with the On-Surface + In-Air feature set is visible not only in terms of accuracy but also while considering the precision, recall, and F1-metric. AdaBoost with On-Surface + In-Air feature set yielded a precision score of 83.3%, a recall score of 78.5%, and an F1-score of 0.801. The precision score of 83.3% indicates that if the AdaBoost classifier has predicted ten samples as positive, more than eight samples are true positives, which is

reasonably good. Moreover, the obtained recall score can be interpreted as if there are ten positive samples in the test set; the classifiers can predict almost 8 of them as positive, which is reasonably acceptable.

Better metrics like ROC and AUC are computed to analyze the performance of classifiers and features more deeply. ROC and AUC compare the classifier's true positive rate (recall) against the false positive rate. Higher values of AUC indicate that there is a more significant distinction between True positives and True negatives. The ROC and AUC of the classifiers with On-Surface features set and On-Surface + In-Air features set are provided in Figs. 4 and 5, respectively.

The AUC of the SVM, RF, and AdaBoost classifiers trained with the On-Surface + In-Air feature set is more than 0.8, which is reasonably good. When the feature set is shrunk to On-Surface alone, the AUC of the classifiers is between 0.76–0.77, which is not good enough. Although there is no standard acceptable AUC value, anything more than 0.85 is usually considered good or excellent. In this context, the RF and AdaBoost trained with On-Surface + In-Air feature set displayed its effectiveness for diagnosing dysgraphia from handwritten data.

Since the number of samples is comparatively less in the evaluated dataset, there is the possibility of overfitting. The performance of classifiers (in terms of accuracy) in both train and test sets is compared to examine the possibility of overfitting. The accuracy of classifiers in the train and test set is provided in Table 5. The classification accuracy in the train set is always higher than in the test set. So, it is impossible to deny the possibility of slight overfitting. However, a few other things, such as fewer data samples or fewer samples in the test set, can also be the reason for observed results. Fewer data samples in the training set can lead to overfitting. Furthermore, the number of data samples in the test data in each fold was about 12. So even misprediction of at least two data samples during testing can reduce the accuracy by about 18%

The performance of the proposed methods is compared with state-of-the-art dysgraphia diagnosis methods evaluated in the same dataset to show the effectiveness. The performance comparison of the proposed method with state of the art method is provided in Table 6. The methods highlighted in bold are the proposed methods. The number

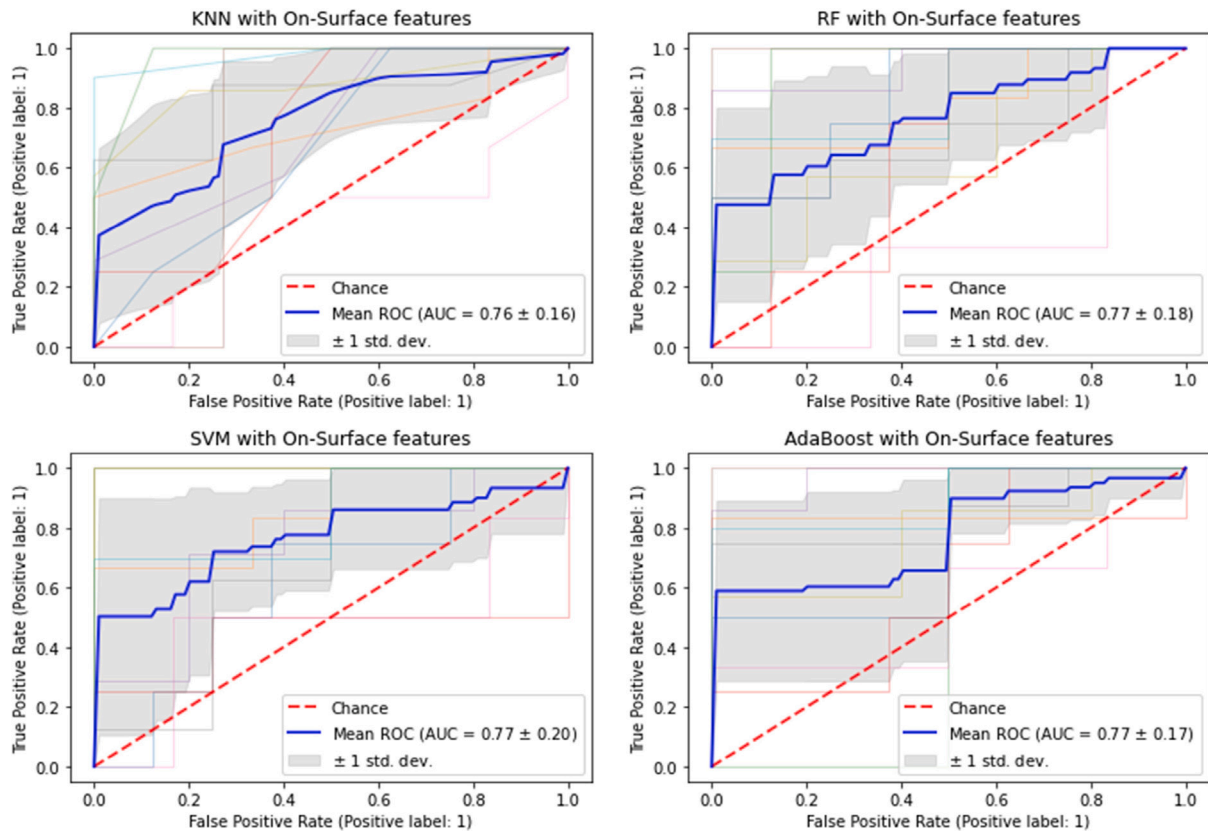


Fig. 4. ROC and AUC of classifiers trained with On-Surface feature set.

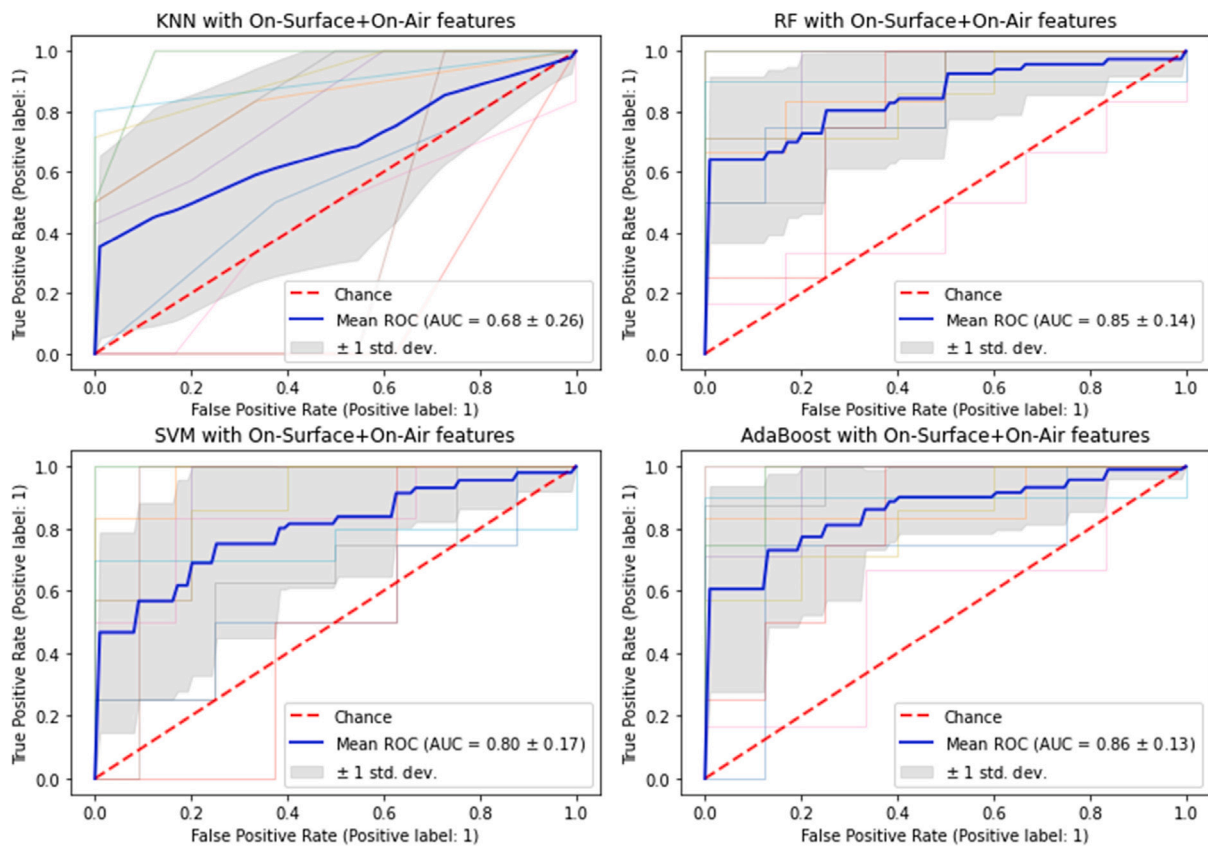


Fig. 5. ROC and AUC of classifiers trained with On-Surface+In-Air feature set.

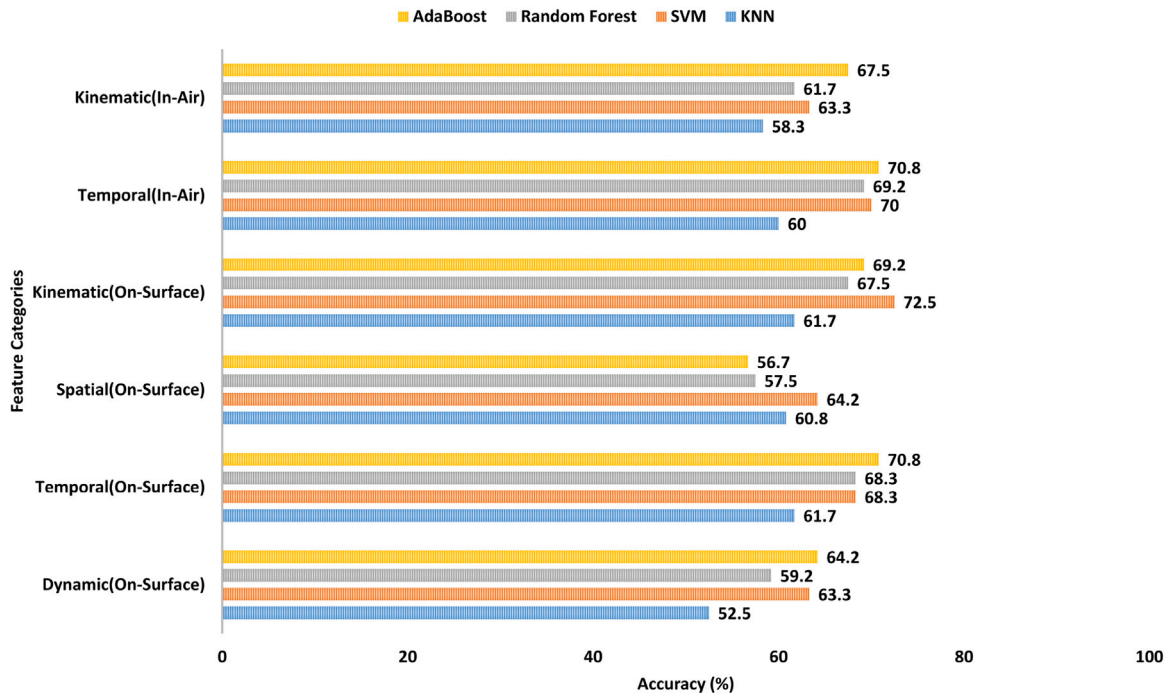


Fig. 6. Classification accuracy of each feature category with different classifiers.

Table 5 Performance (accuracy) comparison of classification algorithms in train and test set.

Algorithm	Feature set			
	On-Surface		On-Surface+In-Air	
	Train set	Test set	Train set	Test set
KNN	71.6	67.5	74.81	68.3
SVM	91.7	72.5	92.3	76.7
Random forest	91.4	67.5	94.2	77.5
AdaBoost	90.9	71.7	95.8	80.8

Table 6 Comparison with state-of-the-art methods.

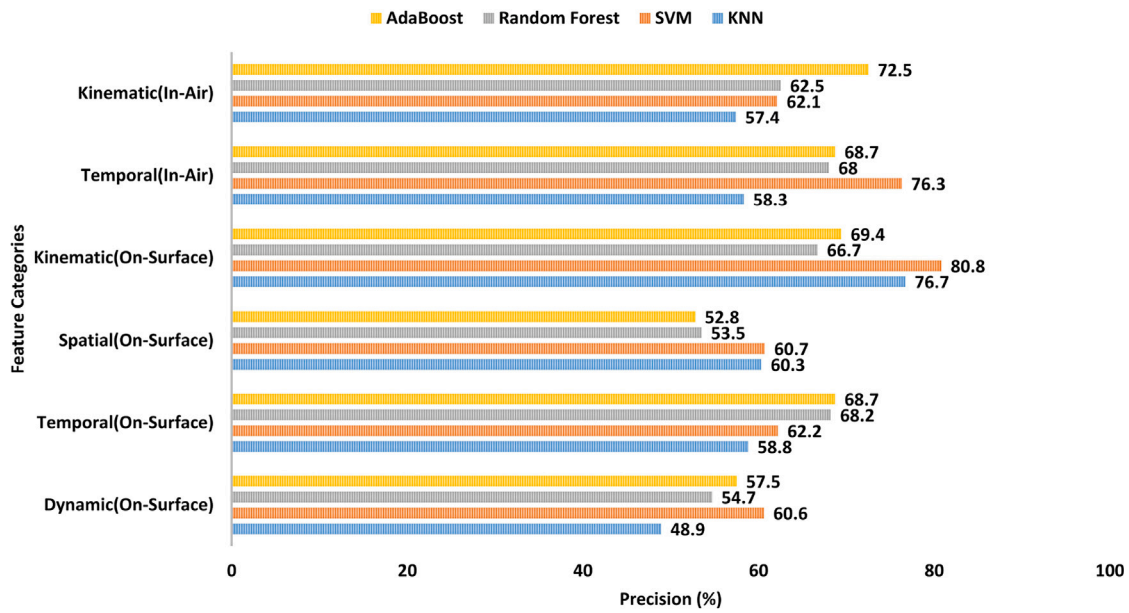
Methods	Features	Accuracy
Adaboost [18]	On-Surface + In-Air features (1176)	79.5%
SVM [18]	On-Surface + In-Air features (1176)	78.8%
RF [18]	On-Surface + In-Air features (1176)	77.6%
CNN [25]	-	79.7%
SVM	On-Surface + In-Air features (175)	76.7%
RF	On-Surface + In-Air features (175)	77.5%
AdaBoost	On-Surface + In-Air features (175)	80.8%

of features used in each method is provided in the 2nd column of the table.

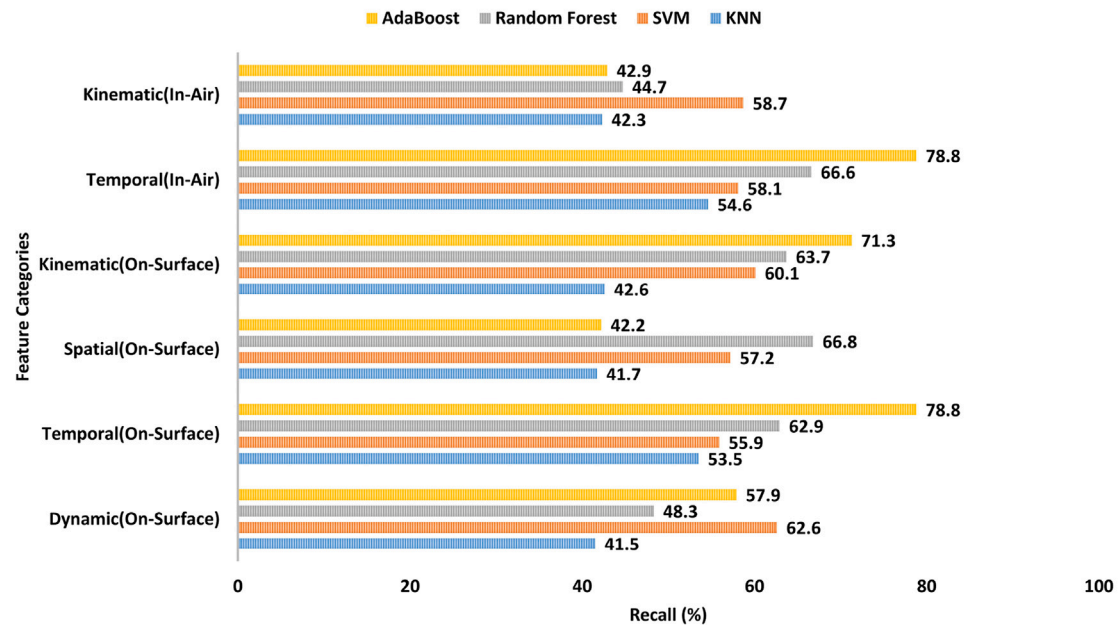
Even though the proposed work has used similar feature types (kinematic, dynamic, temporal, spatial) compared to the state-of-the-art methods, but generated similar performance or better performance by

using a very less number of features. The best result achieved in one of the state-of-the-art methods is 79.5% accuracy for AdaBoost [18] with On-Surface + In-Air features. However, it has been trained with 1176 features. On the other hand, the proposed AdaBoost method trained with On-Surface + In-Air features achieved a classification accuracy of 80.8%, which is almost 1.3% more than the state-of-the-art methods. And this performance is achieved by using fewer features, only 175, which is just about 15% of the total features used in the state-of-the-art methods. The 1176 features set include 1064 On-Surface features and 112 In-Air features. But our feature set contains only 119 On-Surface features and 56 In-Air features. Not only with the AdaBoost Classifiers but also with the Rf classifier, our approach has generated competitive results. It displays the superiority of our method over the state-of-the-art method. The dataset consists of raw data for eight separate writing tasks. The state-of-the-art method has extracted features for each task separately. On the other hand, the proposed method extracted the feature for the whole raw data without task separation, reducing the number of features to a large extent. Fewer features will reduce the computational overhead associated with the training and prediction. And another advantage is the proposed method has displayed superior performance even with fewer features. [25] utilized the convolutional neural network algorithm for classifying normal and dysgraphia writing. They achieved an accuracy of 79.7%, which is almost similar to the performance of the method proposed in [18]. Compared to [25] also, our method performed better. Moreover, the [25] utilized the CNN algorithm, which is computationally expensive compared to the traditional machine learning algorithm.

The second section of the experiment is focused on the analysis of each feature category (kinematic, dynamic, temporal, spatial, temporal (In-Air), kinematic(In-Air)) for the classification of handwritten data. The main aim of this experiment was to study the relevance or significance of each feature category for discriminating the normally developing handwriting and dysgraphia handwriting. A total of 24 classifiers (four classifiers and six feature categories) are trained and evaluated in this experiment. Similar to the first experiment, the second experiment also followed hyperparameter tuning and 10-fold cross-validation during the classifier training and evaluation task. The accuracy of classifiers trained with each feature category is provided in Fig. 6.



(a) Precision



(b) Recall

Fig. 7. Classification performance of each feature category with different classifiers.

Considering the accuracy obtained for each category features with different classifiers, the kinematic, temporal (In-Air), and temporal features are superior for classifying handwriting data into normal or dysgraphia classes. The SVM classifier trained with kinematic features yielded the highest classification accuracy of 72.5%. The second-highest accuracy of 70.8% was recorded in AdaBoost Classifier trained with temporal features. In all classifiers except KNN, the temporal (In-Air and On-Surface) and kinematic features generated a minimum

accuracy of not less than 60% and not less than 65%, respectively. Among all feature categories, the kinematic features are more in number. In usual cases, more number features generate better results. However, temporal features are significantly fewer in number (only six) but displayed better performance. It shows that time-related feature is very significant in the dysgraphia diagnosis problem. In general, it can be interpreted that features' quality is more important than quantity in traditional machine learning applications. By analyzing these values,

it can be concluded that kinematic and temporal features significantly distinguish normally developing and dysgraphia handwriting.

Similar to the first experiment, the precision and recall of the classifiers are computed to analyze their classification performance more deeply. The precision is calculated to examine how well the classifiers trained with each feature category predict the positive class. And the recall is computed to examine how many times the classifiers trained with each feature category predicted the positive class accurately. The precision and recall of classifiers trained with each feature category are provided in Fig. 7.

In terms of precision, Kinematic features are superior to the rest. SVM classifier trained with the kinematic feature recorded a precision of 80.8%. It means the false positive predictions are fewer. However, the recall value of SVM trained with kinematic features is 60.1%. The maximum recall obtained with kinematic features is 71.3%. Although the accuracy and precision were good, the low recall value indicates that the kinematic features are not very significant as the temporal features in this problem. Because the proposed solution is for preliminary diagnosis and false negatives should be reduced. In this context, temporal features are better and achieved a recall value of 78.8% with AdaBoost classifiers. And further, the performance of each feature category varies with each classifier. It can be due to the inductive bias and assumption of the machine learning classifiers considered for the experiment. Based on the analyses, we can conclude that the more significant features for distinguishing the handwritten data are temporal, kinematic, and the rest of the feature categories, respectively.

4. Discussion

The proposed work focused on developing machine-learning methods to build preliminary dysgraphia diagnosis systems for children. And the main aim of the work was to develop handwritten classification methods or machine learning algorithms to distinguish the normally developing handwriting and dysgraphia handwriting with fewer features than the state-of-the-art methods. To an extent, the proposed method has achieved the aim of handwriting classification with fewer features (85% lesser compared to the state-of-the-art method). Even though the number of features is less, one of the proposed methods (AdaBoost classifier trained with 175 features) outperformed the state-of-the-art with a classification accuracy of 80.8%. Moreover, a feature selection approach based on the XGBoost algorithm is implemented to reduce the number of features further. This approach generated 36 relevant features. The accuracy of machine learning classifiers trained with 36 relevant features is not good enough compared to those trained with 175 features. Although the proposed method has outperformed the state-of-the-art method, the achieved classification accuracy is not excellent considering the task is a binary classification problem. We can point out multiple underlying reasons for not achieving excellent classification performance in the evaluated dataset.

First and foremost, the sample size in the dataset is small. The dataset consists of handwriting samples collected from 120 children. It is strenuous to achieve ideal classification performance in machine learning algorithms when trained with a limited number of samples. Furthermore, training the machine learning models with a limited number of samples can result in overfitting. The generalization ability of the machine learning model will be affected by the lack of sample size in the dataset. The number of testing samples in each cross-validation fold is much less (about 12). Even if 2 or 3 predictions became wrong among 12, the accuracy would be between 83% and 75%. The age range of students who participated in the dataset collection is 8 to 15, relatively wide. It means the dataset consists of handwriting data from students of ages ranging from 8–15. Since handwriting is a continuously developing and altering skill during these years, the manifestation of abnormality may differ with age. Also, the relevant features for discrimination may vary with age. The subsets of the dataset (based on different age ranges) need to be analyzed to study the relationship

between age and relevant features. But the number of total samples is comparatively less (only 120) in the available dataset. So the analysis of the subset of data is not practically possible, or it is insufficient to represent the actual population.

Apart from proposing machine learning algorithms to distinguish the normally developing handwriting and dysgraphia handwriting with fewer features, multiple contributions are made in this work. The performance of four machine learning algorithms (KNN, SVM, Random forest, and AdaBoost) is analyzed for classifying normal and dysgraphia writing. KNN is a simple but effective method that has been prevalently used for various classification and regression problems in the literature. But in the dysgraphia diagnosis problem, the performance of the KNN was abysmal. SVM is a complex algorithm compared to KNN, and the possible choice of different kernels for data analysis makes it popular for many machine-learning tasks. SVM has generated decent classification performance. Even though our selection of machine learning algorithms is somewhat random and based on our previous experience, we managed to pick the algorithms with different outlying classification principles. So, after choosing a simple and complex machine learning algorithm (KNN and SVM), we went behind ensemble learning methods, Random forest, and AdaBoost. Although the basic learner in both ensemble learning methods is the decision tree algorithm, each ensembling approach is entirely different. AdaBoost has displayed the best performance among the four machine-learning classification algorithms. It shows the effectiveness of the boosting-based ensemble learning method for the dysgraphia diagnosis problem.

Furthermore, this work analyzed the On-Surface features' effectiveness for classifying handwritten data. The obtained results showed that the On-Surface features alone are insufficient for effectively distinguishing normal and dysgraphia writing. Combining In-Air features with On-Surface features has significantly improved the classification performance. One of the crucial attributes/subfeatures of the In-Air feature set is On-Air time. The On-Air time feature considers the statistical values of idle time that the subject has taken after finishing a segment to starting a new segment. This time can be more in the case of students with dysgraphia. Since there is a higher probability that they require more time to think or they may be confused about writing the next segment. Moreover, different characteristics of handwriting (kinematics, dynamics, temporal and spatial) during On-Surface activity and In-Air activity are analyzed in this work. Observed results indicated that the temporal and kinematics attributes of handwriting have more significance in distinguishing normal and dysgraphia handwriting.

The main aim of the automated dysgraphia diagnosis system is to replace the time-consuming manual preliminary analysis conducted by the occupational therapist and other trained professionals. Psychologists or occupational therapists can utilize this system to diagnose the existence of dysgraphia in children within 5 to 10 min. Moreover, they can also benefit from the diagnosis system throughout treating subjects with dysgraphia. Combining the preliminary diagnosis system with psychological analysis can introduce more intelligent and effective treatment procedures. As part of a tracking system that uses both the psychological approach and machine learning systems, treatment can be considered by analyzing the handwriting attributes and their improvement, as shown in Fig. 8. By exploring what treatment can be helpful to improve the characteristics/attributes (On-Surface and In-Air) that are explored in the research, such artificial intelligence systems can assist psychologists in designing and obtaining a better treatment system. The analysis provided in the experiment and results section shows that temporal and kinematic attributes can distinguish dysgraphia vs. normal cases better. It means the temporal and kinematic aspects of students with dysgraphia are very distinguishable from normally developing students. Quickly identifying these aspects affecting handwriting quality enables the psychologist to design a new treatment strategy by focusing on the specific attributes and their gradual improvements.

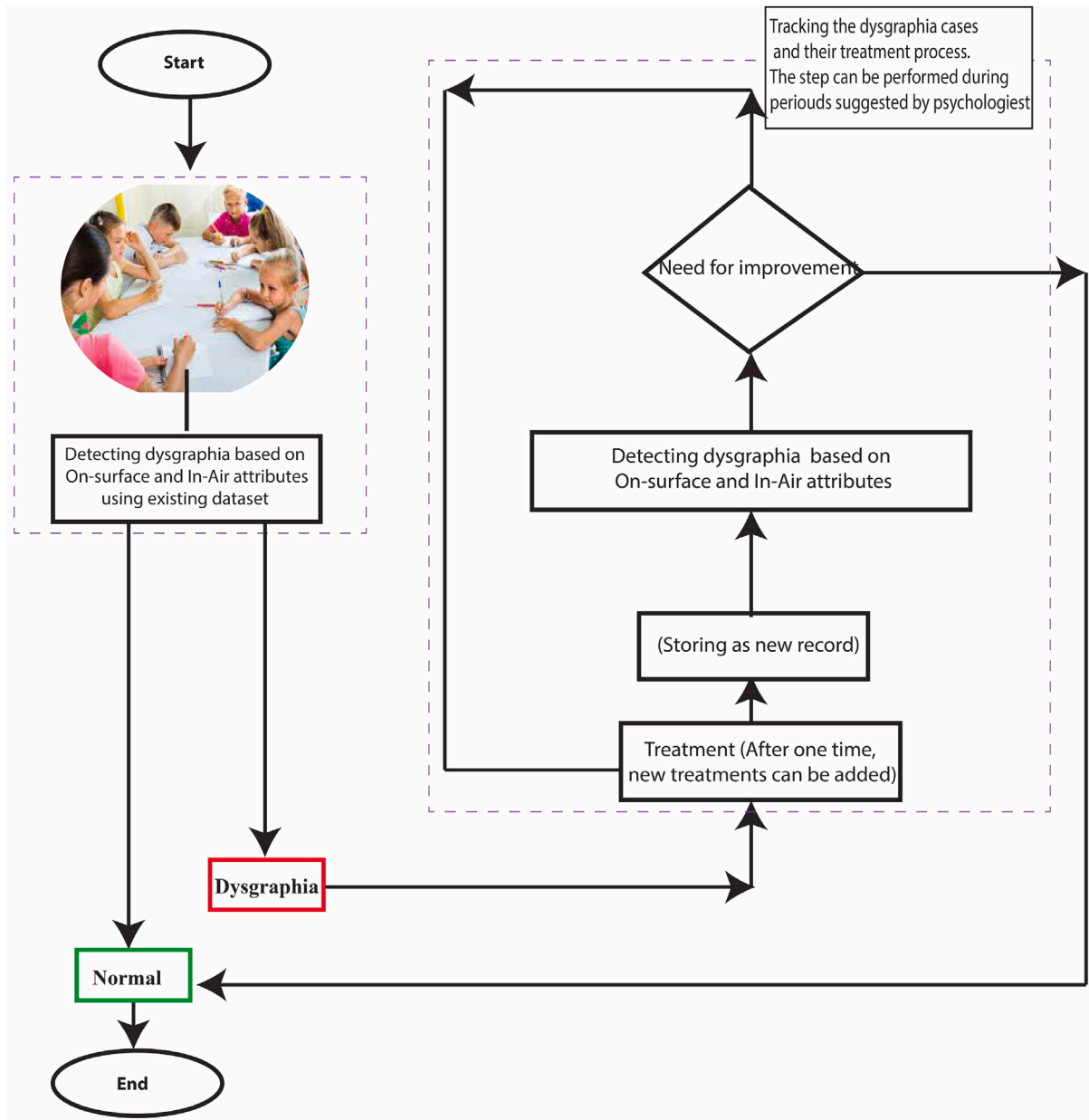


Fig. 8. Overview of the proposed system to assist psychologists in the treatment process.

The proposed work in this article and most of the literature considered dysgraphia diagnosis as a binary classification problem. These systems can identify whether a student has dysgraphia or not. The severity of dysgraphia may differ from person to person. Grading the severity of dysgraphia is more challenging than classifying the handwritten data into either dysgraphia or normal class (binary classification). In that context, the scope of more intelligent systems can be considered in the future that grade the severity of dysgraphia. Grading the levels or severity of dysgraphia using automated systems can assist the psychologist in preliminary diagnosis and further treatment after diagnosing the dysgraphia. Although the dysgraphia diagnosis systems utilize supervised machine learning algorithms for classification, grading levels or severity of dysgraphia can be achieved by using unsupervised algorithms such as traditional k means clustering algorithm or more advanced clustering algorithms [32–34]. It eliminates the need for further labeling the ground truth of the data with different severity levels.

Most of the works in the literature have used online handwritten data for dysgraphia diagnosis compared to the very few that have used handwriting images. Different features can be extracted from the images (offline data) compared to online handwritten data. Future works can be focused on creating a public image handwritten dataset and analysis of the same for diagnosis of dysgraphia. Also, the online handwritten data captured using digitizing tablets can be converted to offline image data. In that context, combining online and offline data can be considered to enhance classification performance. The dataset used in this work is the only publicly available handwritten dataset available for the dysgraphia diagnosis problem in children. But the scalability of the prediction algorithms trained with this dataset is limited since this handwritten dataset is available in Slovak orthography. And Slovak alphabet uses several letters in addition to the 26 letters used in the English alphabet. So the prediction algorithms trained with this dataset can only be used for diagnosis in students who know Slovak or who are native Slovak speakers. This limitation yields the scope for a

global public dataset (an ensemble of multiple orthographies) for the dysgraphia diagnosis problem.

5. Conclusion

This work focuses on the development of machine learning-based methods for the diagnosis of dysgraphia. The proposed approach utilizes handwritten data from the digitized tablet and its kinematic, temporal, dynamic, and spatial characteristics to distinguish normal and abnormal handwriting. The proposed method achieved a state-of-the-art classification performance and a 1.3% increment in accuracy compared to the literature. Our approach is superior to the methods in the literature in terms of accuracy. Further, our method reduced the computational overhead by decreasing the number of features used for training the classifiers. Moreover, this work deeply investigated the significance of different feature categories for handwriting analysis-based dysgraphia diagnosis problem. The observed result indicates that the handwritten data's temporal and kinematic characteristics are more significant than others (dynamic, spatial), etc.

Funding

This publication was supported by Qatar University Graduate Assistant Grant. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of Qatar University.

Ethics approval

This paper complies with the ethical standards of research and methodology

CRediT authorship contribution statement

Jayakanth Kuntho: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization. **Somaya Al Maadeed:** Supervision, Conceptualization, Writing – review & editing, Project administration, Funding acquisition. **Moutaz Saleh:** Supervision, Writing – review & editing. **Younes Akbari:** Conceptualization, Writing – review & editing, Visualization.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.bspc.2023.104715>.

Data availability

Data belongs to a public database and the code of method will be shared up on request.

References

- [1] M. Knickenberg, C.L.A. Zurbruggen, M. Venetz, S. Schwab, M. Gebhardt, Assessing dimensions of inclusion from students' perspective—measurement invariance across students with learning disabilities in different educational settings, *Eur. J. Special Needs Educ.* (2020) <http://dx.doi.org/10.1080/08856257.2019.1646958>.
- [2] R.K. Deuel, Developmental Dysgraphia and motor skills disorders, *J. Child. Neurol.* (1995) <http://dx.doi.org/10.1177/088307389501005103>.
- [3] L. Hamstra-Bletz, J. DeBie, B.P.L.M. Den Brinker, *Concise Evaluation Scale for Children's Handwriting*, Swets 1 zeitlinger ed., Lisse, 1987.
- [4] A.L. Barnett, S.E. Henderson, B. Scheib, J. Schulz, Development and standardization of a new handwriting speed test: The detailed assessment of speed of handwriting, *Br. J. Educ. Psychol.* (2009) <http://dx.doi.org/10.1348/000709909x421937>.
- [5] American Psychiatric Association, American psychiatric association. DSM-5 task force, in: *Diagnostic and Statistical Manual of Mental Disorders : DSM-5*, American Psychiatric Association, 2013.
- [6] C. Lopez, C. Hemimou, B. Golse, L. Vaivre-Douret, Developmental dysgraphia is often associated with minor neurological dysfunction in children with developmental coordination disorder (DCD), *Neurophysiol. Clin.* (2018) <http://dx.doi.org/10.1016/j.neucli.2018.01.002>.
- [7] M. Faundez-Zanuy, J. Fierrez, M.A. Ferrer, M. Diaz, R. Tolosana, R. Plamondon, Handwriting biometrics: Applications and future trends in e-security and e-health, *Cogn. Comput.* (2020) <http://dx.doi.org/10.1007/s12559-020-09755-z>.
- [8] A. Devi, G. Kavva, Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches, *Prog. Neuro-Psychopharmacol. Biol. Psych.* (2022) 110647.
- [9] G. Richard, M. Serrurier, *Dyslexia and Dysgraphia prediction: A new machine learning approach*, 2020, arXiv.
- [10] G. Dimauro, V. Bevilacqua, L. Colizzi, D. Di Pierro, TestGraphia, a software system for the early diagnosis of dysgraphia, *IEEE Access* 8 (2020) 19564–19575, <http://dx.doi.org/10.1109/ACCESS.2020.2968367>.
- [11] I.S. Isa, W.N. Syazwani Rahimi, S.A. Ramlan, S.N. Sulaiman, Automated detection of Dyslexia symptom based on handwriting image for primary school children, *Procedia Comput. Sci.* 163 (2019) 440–449, <http://dx.doi.org/10.1016/j.procs.2019.12.127>.
- [12] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, S. Rosenblum, Identification and rating of developmental dysgraphia by handwriting analysis, *IEEE Trans. Hum.-Mach. Syst.* 47 (2) (2017) 235–248, <http://dx.doi.org/10.1109/THMS.2016.2586605>.
- [13] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, P. Dillenbourg, Automated human-level diagnosis of dysgraphia using a consumer tablet, *npj Digit. Med.* 1 (1) (2018) <http://dx.doi.org/10.1038/s41746-018-0049-x>.
- [14] J. Mekyska, J. Bednarova, M. Faundez-Zanuy, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, J.M. Havigerova, Computerised assessment of graphomotor difficulties in a cohort of school-aged children, in: *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops 2019*, 2019, <http://dx.doi.org/10.1109/ICUMT48472.2019.8970767>.
- [15] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, P. Brezina, New approach of dysgraphic handwriting analysis based on the tunable Q-factor wavelet transform, in: *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, 2019, pp. 289–294, <http://dx.doi.org/10.23919/MIPRO.2019.8756872>.
- [16] L.G. Dui, F. Lunardini, C. Termine, M. Matteucci, N.A. Stucchi, N.A. Borghese, S. Ferrante, A tablet app for handwriting skill screening at the preliteracy stage: Instrument validation study, *JMIR Serious Games* 8 (4) (2020) e20126.
- [17] T. Gargot, T. Asselborn, H. Pellerin, I. Zammouri, S.M. Anzalone, L. Casteran, W. Johal, P. Dillenbourg, D. Cohen, C. Jolly, Acquisition of handwriting in children with and without dysgraphia: A computational approach, *PLoS One* 15 (9 September) (2020) 1–22, <http://dx.doi.org/10.1371/journal.pone.0237575>.
- [18] P. Drotár, M. Dobeš, Dysgraphia detection through machine learning, *Sci. Rep.* 10 (1) (2020) 1–11, <http://dx.doi.org/10.1038/s41598-020-78611-9>.
- [19] T. Asselborn, M. Chapatte, P. Dillenbourg, Extending the spectrum of Dysgraphia: A data driven strategy to estimate handwriting quality, *Sci. Rep.* 10 (1) (2020) 1–11, <http://dx.doi.org/10.1038/s41598-020-60011-8>.
- [20] S. Rosenblum, G. Dror, Identifying developmental dysgraphia characteristics utilizing handwriting classification methods, *IEEE Trans. Hum.-Mach. Syst.* 47 (2) (2016) 293–298.
- [21] S.W. Sihwi, K. Fikri, A. Aziz, Dysgraphia Identification from HandWriting with Support Vector Machine Method, in: *Journal of Physics: Conference Series*, vol. 1201, 2019, <http://dx.doi.org/10.1088/1742-6596/1201/1/012050>.
- [22] Z. Dankovicova, J. Hurtuk, P. Fecilak, Evaluation of digitalized handwriting for dysgraphia detection using random forest classification method, in: *SISY 2019 - IEEE 17th International Symposium on Intelligent Systems and Informatics*, Proceedings, 2019, <http://dx.doi.org/10.1109/SISY47553.2019.9111567>.
- [23] A. Devi, G. Kavva, M.J. Therese, R. Gayathri, Early diagnosing and identifying tool for specific learning disability using decision tree algorithm, in: *2021 Third International Conference on Inventive Research in Computing Applications, ICIRCA, IEEE*, 2021, pp. 1445–1450.
- [24] S. Kedar, et al., Identifying learning disability through digital handwriting analysis, *Turk. J. Comput. Math. Educ. (TURCOMAT)* 12 (1S) (2021) 46–56.
- [25] J. Skunda, B. Nerušil, J. Polec, Method for dysgraphia disorder detection using convolutional neural network, in: *Proceedings - WSCG 2022: 30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Václav Skala-UNION Agency*, 2022.
- [26] J. Kuntho, S. Al-Maadeed, S. Kuntho, Y. Akbari, Automated systems for diagnosis of dysgraphia in children: A survey and novel framework, 2022, arXiv preprint [arXiv:2206.13043](https://arxiv.org/abs/2206.13043).
- [27] P. Drotár, M. Dobeš, Dysgraphia detection dataset, 2020, <https://github.com/peet292929/Dysgraphia-detection-through-machine-learning>.

- [28] L.E. Peterson, K-nearest neighbor, *Scholarpedia* 4 (2) (2009) 1883.
- [29] D.A. Piner, D.M. Schnyer, Support vector machine, in: *Machine Learning*, Elsevier, 2020, pp. 101–121.
- [30] G. Biau, Analysis of a random forests model, 2012.
- [31] R.E. Schapire, Explaining adaboost, in: *Empirical Inference*, Springer, 2013, pp. 37–52.
- [32] B. Diallo, J. Hu, T. Li, G.A. Khan, X. Liang, Y. Zhao, Deep embedding clustering based on contractive autoencoder, *Neurocomputing* 433 (2021) 96–107.
- [33] M.A. Khan, G.A. Khan, J. Khan, M.R. Khan, I. Atoum, N. Ahmad, M. Shahid, M. Ishrat, A.A. Alghamdi, Multi-view clustering based on multiple manifold regularized non-negative sparse matrix factorization, *IEEE Access* 10 (2022) 113249–113259.
- [34] G.A. Khan, J. Hu, T. Li, B. Diallo, Y. Zhao, Multi-view low rank sparse representation method for three-way clustering, *Int. J. Mach. Learn. Cybern.* 13 (2022) 233–253.