IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 July 2023, accepted 1 August 2023, date of publication 7 August 2023, date of current version 15 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3302903

RESEARCH ARTICLE

Handwriting-Based ADHD Detection for Children Having ASD Using Machine Learning Approaches

JUNGPIL SHIN[1], (Senior Member, IEEE), MD. MANIRUZZAMAN[1], YUTA UCHIDA[1], MD. AL MEHEDI HASAN[2], (Member, IEEE), AKIKO MEGUMI[3], AND AKIRA YASUMURA[3]

[1]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan
[2]Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh
[3]Graduate School of Social and Cultural Sciences, Kumamoto University, Kumamoto 860-8555, Japan

Corresponding author: Jungpil Shin (jpshin@u-aizu.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI), Japan, under Grant JP20K11892, Grant JP21H00891, and Grant JP23KJ2124.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Interview Review Board (IRB) of Kumamoto University, Japan, under Approval No: 45, May 25, 2021, and performed in line with the Declaration of Helsinki.

ABSTRACT Attention deficit hyperactivity disorder (ADHD) for children is one of the behavioral disorders that affect the brain's ability to control attention, impulsivity, and hyperactivity and its prevalence has increased over time. The cure for ADHD is still unknown and only early detection can improve the quality of life for children with ADHD. At the same time, children with ADHD often suffer from various comorbidities like autism spectrum disorder (ASD), major depressive disorder (MDD), etc. Various researchers developed computational tools to detect children with ADHD depending on handwriting text. Handwriting text-based systems are depending on a specific language that causes problems for non-native speakers of that language. Moreover, very few researchers considered other comorbidities such as ASD, MDD, etc., in their studies to detect ADHD for children. In this study, handwriting patterns or drawing is assumed as an aspect to identify/detect ADHD children who have ASD using machine learning (ML)-based approaches. We collected handwriting samples from 29 Japanese children (14 ADHD with coexisting ASD children and 15 healthy children) using a pen tablet. We asked each child to draw two patterns, namely zigzag lines (ZL) and periodic lines (PL) on a pen tablet and repeated them three times. We extracted 30 statistical features from raw datasets and these features were analyzed using sequential forward floating search (SFFS) and selected the best combinations or subsets of features. Finally, these selected features were fed into seven ML-based algorithms for detecting ADHD with coexisting ASD children. These classifiers were trained with leave-one-out cross-validation and evaluated their performances using accuracy, recall, precision, f1-score, and area under the curve (AUC). The experimental results showed that the highest performance scores (accuracy: 93.10%; recall: 90.48%; precision: 95.00%; f1-score: 92.68%; and AUC: 0.930) were achieved by the RF-based classifier for the PL predict task. This study will be helpful and provide evidence of the possibility of classifying ADHD children having ASD and healthy children based on their handwriting patterns.

INDEX TERMS ADHD, ASD, detection, handwriting patterns, machine learning.

I. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is one of the behavioral disorders that affect the brain's ability to regulate attention, impulsivity, and hyperactivity [1]. It is mainly developed in childhood or preschoolers (aged 3-5 years) and becomes more acute/severe problems in school-aged children and can persist into adulthood [2], [3]. Children with ADHD suffer from various complications like poor academic performance and employment attainment [4], poor physical/mental health [4], and suicide attempts [4], [5], [6], [7]. Moreover, males are more affected by ADHD than females [8].

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

According to CDC, the number of children with ADHD has fluctuated over time: 4.4 million children with ADHD aged 3-17 years were diagnosed n 2003 [9], 5.4 million in 2007 and 6.1 million in 2016 [8]. This rate has gradually increased globally. Moreover, children or adults with ADHD also suffer from other psychiatric disorders such as autism spectrum disorder (ASD), major depressive disorder (MDD), etc [10], [11], [12]. Moreover, existing studies found that adults with ADHD had at least one coexisting psychiatric disorder [10], [11], [12].

Various diagnostic tools like MRI [13], [14], fNIRS [15], EEG [16], [17], [18], questionnaires-based [19], [20], and performance test [21], [22], etc. were widely used for detecting ADHD. Moreover, many existing works utilized handwriting text to detect children with ADHD [23], [24], [25]. Handwritten analysis can be performed in two ways: handwriting text and handwriting patterns which can be done both offline and online. Handwriting text-based systems are depending on a specific language that causes problems for non-native speakers of that language. Since non-native speakers do not know a language properly, they will face difficulties to write down a text of that language which will prevent them to generate actual signals to detect ADHD children. Patterns are common for all people and it creates an equal opportunity to draw the pattern for detecting ADHD.

Recently, handwriting patterns have also been used to classify age groups [26], [27], ADHD detection [28]. Moreover, very few researchers considered other coexisting comorbidities in their studies to detect ADHD for children [29]. In this work, we proposed a handwritten pattern for detecting ADHD children with coexisting ASD. Nowadays, various digital devices such as pen tablets allow us for recording the sequences of measurements from the tasks of handwriting. These recorded data are analyzed using statistical analysis and multiple algorithms based on machine learning (ML)-based approaches.

In this work, we also used handwriting patterns to detect children who have both ADHD and ASD problems. Two types of handwriting patterns such as zigzag lines (ZL) and periodic lines (PL) have two conditions: trace and prediction. At the same time, we tried to extract some statistical features and implemented sequential forward floating selection (SFFS) with seven ML-based approaches in order to select the best combination of effective and efficient features based on classification accuracy. These seven ML-based algorithms were support vector machine (SVM), random forest (RF), decision tree (DT), Gaussian naïve Bayes (GNB), k-nearest neighbors (k-NN), logistic regression (LR), and extra tree, respectively. This study found an excellent finding that handwritten PL predict pattern had discriminative power and was more capable to distinguish ADHD children with coexisting ASD from healthy children compared to other patterns. Moreover, Our proposed RF-based system produced the highest classification accuracy for detecting ADHD in children having ASD problems. In summary, the contributions of this study are as follows:

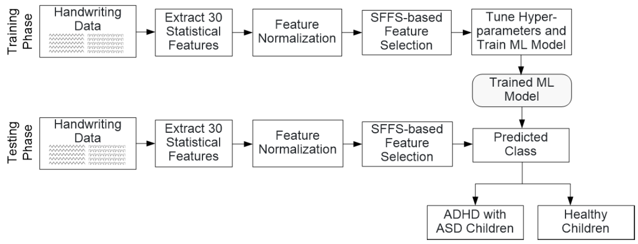

FIGURE 1. ML-based framework for predicting children with ADHD who had ASD.

- We proposed a novel handwriting pattern instead of handwriting text for detecting ADHD for children who have ASD.
- We have extracted statistical features from raw features.
- More effective and efficient features were selected using the SFFS method.
- Finally, seven machine learning algorithms were adopted for ADHD detection for children having ASD.

The remaining part of this study is organized as follows: Section II introduce materials and methods. This section includes the proposed methodology, subject and recruitment process, data collection device and its procedure, feature extraction, normalization, feature selection, and classification algorithms. Experimental design and performance metrics are presented in Section III. Moreover, the experimental results along with discussions are presented in Section IV. Finally, we summarize the conclusion and future direction of this study in Section VII.

II. MATERIALS AND METHODS

A. PROPOSED METHODOLOGY

In this study, we designed an ML-based framework for detecting ADHD children having ASD problems. Fig. 1 illustrates the overall framework of this work. To conduct this study, we performed the following steps. Firstly, the handwritten dataset was first taken as input data and then divided into two phases: the training and testing phases. In the testing phase, one subject was taken and the remaining (n-1) subjects were chosen for the training phase. In the training phase, training data was used to train ML-based models, and test data was utilized to predict children with ADHD having ASD problems. The second step was feature extraction from raw features, followed by feature normalization. This was designed to keep the extracted features within a similar scale. The next step was feature selection, which involved selecting the dominant features by removing irrelevant features. In this work, SFFS was employed as a feature selection method and selected the dominant features. These features were used to train ML-based framework algorithms and tuned their hyperparameters using the grid search method. In the training phase, these tuned parameters were utilized to train again ML-based algorithms. Moreover, the selected dominant features were also extracted from the test phase and fed into trained ML-based algorithms

TABLE 1. Summarization of utilized handwritten pattern datasets.

Class labels	Age (years)	No. of subjects	No. of task	No. of repeat	Total samples
ADHD with coexisting ASD	5-14	14	4	3	168 (14×4×3)
Healthy children	8-12	15	4	3	180 (15×4×3)

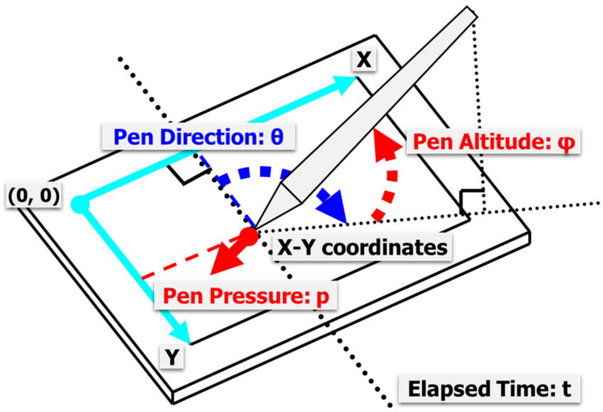

FIGURE 2. Pen tablet device.

to predict children into two classes: children with ADHD who coexist with ASD and healthy children.

B. SUBJECTS AND RECRUITMENT PROCESS

In this work, we included 29 Japanese school students or subjects aged 5-14 years. All subjects were diagnosed by a medical practitioner using the psychiatric disorders (ADHD or ASD) using the Autism-Spectrum Quotient (AQ) [30], [31] and the ADHD-Rating Scale-IV (ADHD-RS-IV) [32]. Depending on the rating scales, subjects were classified as ADHD children with coexisting ASD and healthy children. We found that 14 subjects had ADHD with coexisting ASD (age: 5-14 years) and the remaining 15 subjects had no disorder or were healthy (age: 8-12 years). The summarization of the dataset is presented in Table 1. All subjects were confirmed to be right-handed. Prior to participation in this study, we obtained written or oral consent from all subjects or their guardians (parents/grandparents/elder sisters or brothers).

C. DATA COLLECTION DEVICE

In this paper, we collected a handwriting pattern dataset based on a pen tablet (Cintiq Pro 16, Wacom Co. Ltd., Japan), which had a 15.6-inch screen with a resolution of 2560 × 1440 resolution. In this work, we used a stylus pen to draw handwritten on the pen tablet device and recorded their drawings at a resolution of 200 Hz. When subjects were drawn the handwritten patterns using a stylus pen on the pen tablet device, the pen tablet device provided us with six raw features such as x and y coordinates (in pixels), drawing speed, pen pressure, and pen tilt (horizontal/vertical angle), time which is shown in Fig. 2.

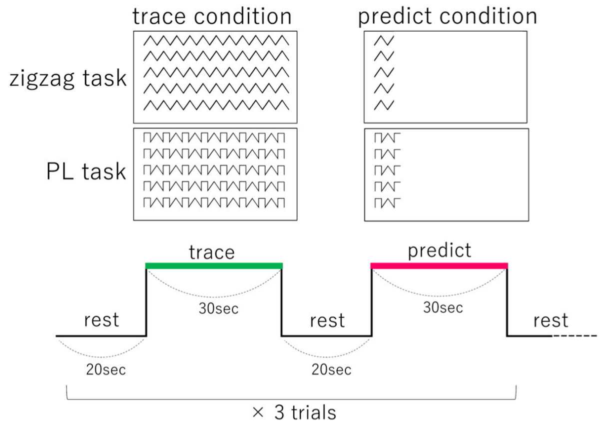

FIGURE 3. Handwritten data collection procedure for four tasks.

D. DATA COLLECTION PROCEDURE

We asked each subject to sit on a desk in a such way that their feet touched the floor and kept the pen tablet near the center of their body. At the same time, we set the distance between the pen tablet and their eyes at about 40 cm. In this work, two handwriting patterns were implemented to identify ADHD with coexisting ASD subjects and healthy subjects. One pattern was a continuous zigzag and another pattern was a periodic line (PL). Each pattern had two conditions including trace and predict. In trace condition, subjects need to use a pen to trace the entire or complete pattern that was visible on the screen of the pen tablet. Whereas, subjects continued their drawing pattern in predicted conditions, which was also visible on the screen of the pen tablet. The pattern was initiated at a distance of 2.1 cm from the left edge and 3.3 cm from the top of the pen tablet screen.

We set the following requirements in order to draw a zigzag line on the screen: 70 degrees apex angle, 80 degrees bottom angle, and 2.5 cm of each side, which was repeated 5 times. The space among these five lines was 3.5 cm, which was presented in gray color for the trace line. Whereas, only the 1st segmentation or parts were visible in the predicted condition. We asked subjects to draw predict or trace-based patterns without lifting the pen tip from the tablet and continue to the next line after finishing one line. To make a stimulus diagram of the PL line, we alternatively arranged a baseless quadrangle (height: 2 cm vs. width: 1.3 cm) and an isosceles triangle (vertex: 70 degrees). About 7.5 cycles were arranged in a row for triangles and squares. Five lines of PL line were presented in light gray on one screen.

To form a dataset, three blocks were executed in successive order. In order to perform tasks, we set the following condition for each block: 20-second rest, 30-second for the drawing task of trace and another 20-second rest, and a 30-second for predictive drawing. We instructed each subject to perform these tasks 3 times. The data collection procedure of handwritten is explained in Fig. 3.

E. FEATURE EXTRACTION

In this work, a set of 30 statistical features were extracted for each task to discriminate handwriting patterns by ADHD with coexisting ASD children from healthy children. These 30 features were computed from six raw features including x and y coordinates (in pixels), drawing speed, pen pressure, and pen tilt (horizontal/vertical angle). These statistical features are well-known and widely used in other domains [26], [27]. The list of extracted feature names, explanations, and their computational calculation formula is shown in Table 2.

F. FEATURE NORMALIZATION

Feature normalization is known as feature scaling or z-score normalization in the field of statistics and machine learning. In this work, we used z-score normalization in order to make a standardization transformation for feature normalization, which is computed using the following formulae:

$$z = \frac{X - \mu}{\sigma} \tag{1}$$

where X is the input feature; μ is the mean/average of the feature, and σ is SD. The value of z ranges from 0 to 1.

G. FEATURE SELECTION TECHNIQUE

Feature selection (FS) is a technique or process that reduces the dimension of the training set by selecting or identifying only the biomarkers that are associated with or relevant to the class or study variable (here, ADHD with coexisting ASD). This study excludes the biomarkers or features that: (i) have a lower or minimum discriminative power capability and (ii) are redundant or irrelevant to each other [34]. Because the selection of effective and efficient biomarkers or features may increase or improve learning algorithm efficiency, and predictive accuracy, and reduce computational time and cost. Moreover, the biomarkers or features that are fed into predictive or learning algorithms are hypothetically assumed to be associated with underlying class labels or diseases (here, ADHD+ASD). This study used the SFFS-based FS algorithm to determine efficient or potential biomarkers. The details of SFFS-based FS algorithms are explained in the following subsections:

1) SFFS-BASED ALGORITHMS

Sequential feature selection (SFS)-based methods are the set of greedy algorithms that are utilized to reduce feature dimension space [35]. In this study, SFFS was used to determine a proper subset or combination of biomarkers or features for ADHD with coexisting ASD. The pseudo-code of the SFFS-based algorithm is summarized in Fig. 4.

1. **Input:** $Y = \{y_1, y_2, \ldots, y_d\}$
2. **Output:** $X_k = \{X_k | j = 1, 2, \ldots, k;\ x_j \in Y\}, where\ k = (0, 1, 2, \ldots, d)$
3. **Initialization:** $X_0 = \emptyset, k = 0$
4. **Select the subset of features:**
   $\mathrm{x}^+ = \mathrm{argmax}\ \mathrm{J}(\mathrm{X_k} + \mathrm{x}), \text{where } \mathrm{x} \in \mathrm{Y} - \mathrm{X_k}$
   $\mathrm{X_{k+1}} = \mathrm{X_k} + \mathrm{x}^+;\ \mathrm{k} = \mathrm{k} + 1$
5. **Select the worst features:**
   $\mathrm{x}^- = \mathrm{argmax}\ \mathrm{J}(\mathrm{X_k} - \mathrm{x}), \text{where } \mathrm{x} \in \mathrm{X_k}$
6. **If** $\mathrm{J}(\mathrm{X_k} - \mathrm{x}) > J(\mathrm{X_k})$, then
   $\boldsymbol{X_{k-1} = X_k - x^-;\ k = k + 1}$
   Go to step 5
   Else
   Go to step 4

FIGURE 4. Pseudo code of SFFS-based algorithm.

H. CLASSIFICATION MODEL

Seven ML-based algorithms such as SVM, RF, DT, GNB, k-NN, LR, and ET were employed to distinguish ADHD with coexisting ASD children from healthy children, which are shortly discussed in the next subsections:

1) SUPPORT VECTOR MACHINE

SVM is a powerful supervised learning method that can be used to solve problems in classification and regression [36], [37]. The main purpose of SVM is to find boundaries (hyperplanes) that can easily be separated the class label (yes/no) by solving the following constraints:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(z_i,\ z_j) \tag{2}$$

Subject to

$$\sum_{i=1}^{n} {y_i}^T \alpha_i = 1, 0 \leq \alpha_i \leq C, i = 1, \ldots, n \ \&\ \forall\ i = 1, 2, 3, \ldots, n \tag{3}$$

The final discriminate function is written as:

$$f(z) = \sum_{i=1}^{n} \alpha_i K(z_i,\ z_j) + b \tag{4}$$

where, b is the bias term and $K(z_i,\ z_j)$ is a gram or kernel matrix, which needs to be chosen or defined during performing SVM. In this work, we used a radial basis kernel, which is mathematically defined ass:

$$K(z_i,\ z_j) = \exp(\text{-}\gamma \left\| z_i\text{-}z_j \right\|^2) \tag{5}$$

In this work, the value of cost (C) and gamma (γ) are optimized using a grid search method. At the same time, we chose the optimum value of C and γ at which points SVM provides the highest classification accuracy. In the case of this study, we have used the following steps in order to perform SVM for predicting ADHD with coexisting ASD children:

Step 1: Spilt the dataset into training and test set. Whereas, 1 subject was taken as a test set, and the remaining (n-1) subjects are used as the training set.

Step 2: Select a suitable kernel on the basis of the characteristics of the training set. Here, we chose the radial basis function.

Step 3: Select the hyperparameters of regularization parameter C and gamma (γ) of the kernel function using the grid search method.

Step 4: After optimizing hyperparameters (C and γ), trained SVM with RBF model on the training set.

TABLE 2. List of 30 extracted statistical features from the raw dataset and their computational formula .

SN	Feature	Descriptions	Formula
1	Width	Range of X coordinate	Max (X)-Min (X)
2	Height	Range of Y coordinate	Max (Y)-Min (Y)
3	Length (L)	Total length of the drawing	$\sum_{k=1}^{n} a_k$, Where, $a = \{x \epsilon \text{ Distance}\}$
4	Velocity (V)	Total velocity of the drawing	L/Total drawing time
5	Max_V_P	Maximum velocity of the Peak	Max$\{x\epsilon\text{V}\}$
6	Min_V_P	Minimum velocity of the Peak	Min$\{x\epsilon\text{V}\}$
7	Max_A_P	Maximum acceleration of the Peak	Max$\{x\epsilon\text{A}\}$
8	Min_A_P	Minimum acceleration of the peak	Min$\{x\epsilon\text{A}\}$
9	Mean_GA_H_Mean	Mean of grip angle for horizontal	$a_{\bar{a}h} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Angles}\}$
10	Mean_GA_V	Mean of grip angle for vertical	$a_{\bar{a}v} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Angles}\}$
11	SD_GA_H	SD of Grip angle for horizontal	SD $(ah) = \sqrt{\frac{\sum_{i=1}^{n}(a_i - a_{\bar{a}h})^2}{n}}$, Where, $a = \{x \in \text{ Angles}\}$
12	SD_GA_V	SD of Grip angle for vertical	SD (av) $= \sqrt{\frac{\sum_{i=1}^{n}(a_i - a_{\bar{a}v})^2}{n}}$, Where, $a = \{x \in \text{ Angles}\}$
13	Mean_Press	Mean of recorded pressure	$a_{\bar{p}} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Pressure}\}$
14	SD_Press	SD of recorded pressure	SD $(p) = \sqrt{\frac{\sum_{i=1}^{n}\left(a_i - a_{\bar{p}}\right)^2}{n}}$, Where, $a = \{x \in \text{ Pressure}\}$
15	Mean_Pos_C_Pres	Mean of positive change in pressure	$a_{\bar{p}pc} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Pressure}\} > 0$
16	SD_Pos_C_Pres	SD of positive change in pressure	SD (ppc) $= \sqrt{\frac{\sum_{i=1}^{n}(a_i - a_{\bar{p}pc})^2}{n}}$, Where, $a = \{x \in \text{ Pressure}\} > 0$
17	Max_Pos_C_Pres	Maximum of positive change in pressure	Max(k), where, k=x $\in$ Pres>0
18	Mean_Neg_C_Pres	Mean of negative change in pressure	$a_{\bar{n}pc} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Pres}\} < 0$
19	SD_Neg_C_Pres	SD of negative change in pressure	SD$(npc) = \sqrt{\frac{\sum_{i=1}^{n}(a_i - a_{\bar{n}pc})^2}{n}}$, Where, $a = \{x \in \text{ Pres}\} < 0$
20	Max_Neg_C_Pres	Maximum of negative change in pressure	Max(k), where, k={x$\in$ Pressure}<0
21	Error	No. of outliers and triangle square errors based on angles	(Square Angle-Triangle Angle) <0
22	Mean_Peak_Pres	Mean of Peak pressure at minima	$a_{\bar{p}p} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Pressure}\}$
23	ErrorStopTime	Mean of starting time at minima point before error	$\frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Time}\}$
24	Mean_Angle_Mean	Mean of angles at maxima and minima	$a_{\bar{A}} = \frac{\sum_{k=1}^{n} a_k}{n}$, Where, $a = \{x \in \text{ Angles}\}$
25	Angle_Var	Angle variance	Var $(A) = \frac{\sum_{i=1}^{n}(a_i - a_{\bar{n}pc})^2}{n}$, Where, $a = \{x \in \text{ Pressure}\}$
26	RegL_Slope	Slope of regression model	[33]
27	RegL_Inter	Intercept of the regression model	[33]
28	LoopCount	Spent of writing time partitioned by the number of peaks	$\sum_{k=1}^{n} a_k$, Where $a = \{x \epsilon \text{ Peaks}\}$
29	Angle_Velocity	Mean of velocities at the edge of the peaks and valleys	Distance/Time
30	Error_Rate	Rate of Error	Error/Peaks

Step 5: Use the trained SVM with RBF kernel to predict the class label (ADHD with coexisting ASD vs. HC) of the test set.

Step 6: Repeat **Step 1** to **Step 5** into n times.

Step 7: Compute performance metrics such as accuracy, recall, precision, f1-score, and AUC.

2) RANDOM FOREST

Random forest (RF) is an ensemble learning that integrates multiple weak learners based on a decision tree to improve generalization ability. It is also used for both regression and classification problems. RF generates multiple decision or classification trees during the training phase. Each tree is created using bootstrapping sampling from the original data and the classification tree method. After forming a forest, a new object is placed on each tree for classification. The forest is selected according to the class that provides the maximum votes for the object. RF is performed as follows:

Step 1: Draw the number of trees (n_{tree}) bootstrap samples from the n training samples.

Step 2: Construct classification trees for each bootstrap sample by taking m_{try} of predictors and choosing the best split from among these variables.

Step 3: Predict the new class by combining the prediction of the (n_{tree}) trees.

Error rate or classification accuracy can be assessed in two ways: training and test set. In this paper, we trained the RF-based model for the training set and evaluated its performance for the test set. The hyperparameters were optimized using a grid search method and the value of (n_{tree}) and m_{try} were selected at the points that provided the highest classification accuracy or lowest error rate, respectively.

3) DECISION TREE

Decision tree (DT) is a tree-structure-based technique that is widely used in data mining for solving regression and classification tasks [38]. The objective of DT is to build a model that can predict the study variable by learning or training simple decision rules from input features. In DT, there are three nodes: the internal node, the decision node, and the leaf node. Here, internal nodes are the set of input features, decision nodes are utilized to make any decision by learning or training and leaf nodes are the output of these decisions. Nowadays, it is widely used in different fields like healthcare, medical imaging, and so on.

4) GAUSSIAN NAÏVE BAYES

Gaussian Naïve Bayes (GNB) is a classification method that is also widely used in ML. GNB assumes that the distribution of each input feature must follow a normal or Gaussian distribution. Assume that we have a set of input features x_i $(i = 1, 2, \ldots, k)$ and y_k be the outcome or class label (k = 0, 1). Here, "0" stands for healthy children and "1" stands for ADHD with coexisting ASD children and. First, the input training set is segmented by class label and computes or estimates the mean and standard deviation (SD) of each input feature for each class using the following formula:

$$\text{Mean}: \mu_k = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6}$$

$$\text{SD}: \sigma_k = \sqrt{\left(\frac{1}{n}\sum_{i}^{n}(x_i - \bar{x})^2\right)} \tag{7}$$

Suppose that we have taken some observations value "v". The probability density function (pdf) of "v" given y_k is computed as follows:

$$p(x = v|y_k) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{v-\mu_k}{\sigma_k}\right)^2}; \quad \sigma_k > 0 \; and \; -\infty \le v, \; \mu_k \le \infty \tag{8}$$

For a given testing data point, we compute the likelihood or posterior probability based on the estimated value of μ_k and σ_k for each class (1/0). The predicted class label is expected to belong to the class with the highest posterior probability.

5) K-NEAREST NEIGHBORS

k-nearest neighbors (k-NN) is a popular ML-based technique developed by Fix and Hodges in 1951, later expanded by Cover and Hart [39] that can be utilized to solve regression and classification problems. It is a distance-based learning algorithm to measure feature vector similarity. This paper used Euclidean distance to compute the distance to all training data points and select the value of k. At the same time, we determine the majority class among k-neighbors that are treated as predicted classes. In this paper, we tune or optimize the value of k using a grid search technique to achieve better performance.

6) LOGISTIC REGRESSION

Logistic regression (LR) is a statistical method that is used for binary classification tasks. It is usually used to estimate or predict the probability of binary class labels based on input feature vectors. Whereas, input features can be continuous or categorical, and the class label is binary either "1" or "0". Here, "1" stands for ADHD with coexisting ASD children, and "0" stands for healthy children. During the training phase, LR used logit or sigmoid function, which is defined as:

$$\sigma(\mathrm{w}) = \frac{1}{1 + \exp(-\mathrm{w})} \tag{9}$$

where, $\mathrm{w} = \sum_{i=0}^{k} \alpha_i \mathrm{X_i}$ with $\alpha_0 = 1$. Here, $\mathrm{X_i}$ is the set of ith input features, and α_i is the set of ith unknown coefficients, which needs to be estimated. In this paper, we estimated these coefficients by MLE during the training phase and these coefficients were used to predict the outcome or class label as ADHD with coexisting ASD children and healthy children.

7) EXTRA TREE

An extra tree (ET) is an ensemble of DTs that perform classification or regression depending on a tree-based algorithm. Unlike the RF-based algorithm, the ET-based algorithm also constructs multiple DTs based on random subsets of training sets and features. At the same time, it randomly chooses the thresholds for each feature. The ET-based algorithm maintains its optimization ability while adding an additional layer of randomization [40].

III. EXPERIMENTAL SETUP AND PERFORMANCE METRICS

A. EXPERIMENTAL SETUP

In this paper, we used Python version 3.10 to perform all experiments. As an operating system, Windows 10 version 21H1 (build 19043.1151) 64-bit is configured and Intel(R) Core (TM) i5-10400 with 16 GB RAM is used in terms of hardware. In this work, we have used the leave-one-out cross-validation (LOOCV) protocol. Whereas, the dataset was divided into two sets: a training set and a test set. Here, one subject was used for the test set and the remaining (n-1) subject was used for the training set. During training predictive models, we set the initial interval or range for each hyperparameter of each predictive model, which is shown in Table 3. For example, SVM used "RBF" kernel with cost (C) of {0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000}; and gamma (γ): {0.00001, 0.0001, 0.001, 0.01, 0.1, 1}. RF: the max_depth: {5, None}; the n_estimator: {50, 100, 200, 300}; the min samples_split: {2, 3}, the min samples leaf: {1, 3}, bootstrap: {True, False}, and the criterion: {"gini", "entropy"}. DT: the max depth of {1, 2, 3, 4, 5}; the min samples leaf of {1 to 10}, the min samples split of {2,3,4,5}. k-NN: the value of n neighbors from 1 to 20; the weight of {'uniform', 'distance'}, 'p': {1, 2}. LR: cost (C) of {10**i for i in range (-4,4)}, the penalty of {"l1", "l2"}, and the solver of 'liblinear'. ET: the max depth of {3,4,5}, the min samples leaf of {1,4,7}, the min samples split of 2. In the training phase, we used the initial parameters of each classifier and tuned these parameters using the grid search method. After optimizing the parameters of each classifier, we trained again all classifiers, which were used to predict the class label on the test set. This entire process was repeated n times. Moreover, we also computed the predicted class of each trial and its probability for all subjects over four tasks.

B. PERFORMANCE METRICS

Performance metrics are used to assess the effectiveness or efficiency of predictive models in making predictions or

TABLE 3. Set hyperparameters of classifiers.

CT	Search Range of Parameters
SVM	C: {0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000}; γ: {0.00001, 0.0001, 0.001, 0.01, 0.1, 1}
RF	max_depth: {5, None}, n_estimators: {50, 100, 200, 300}, min_samples_split: {2, 3}, min_samples_leaf: {1, 3}, bootstrap: {True, False}, criterion: {"gini", "entropy"}
DT	max_depth: {1, 2, 3, 4, 5}, min_samples_leaf: {1 to 10}, min_samples_split: {2, 3, 4, 5}
GNB	None
KNN	n_neighbors: np.arange (1, 20), weights: {'uniform', 'distance'}, 'p': {1, 2}
LR	C: {10**i for i in range (-4,4)}, penalty: {"l1", "l2"}, random_state: [1], solver: ['liblinear']
ET	max_depth: [3, 4, 5], min_samples_leaf: [1, 4, 7], min_samples_split: [2]

classifications. In this work, different performance metrics like accuracy (ACC), precision (Preci), recall (Rec), F1 score (FS), and ROC-AUC score were used to evaluate the performance of seven predictive models. These metrics provide insight into how well the predictive models are able to accurately identify ADHD in children with ASD based on the input data. These performance metrics were computed based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which are presented in Table 4. ACC, Preci, Rec, and FS are computed using the following formula:

Accuracy:

Accuracy measures the proportion of correct predictions. The ratio of correctly predicted classes to the total predicted classes is mathematically expressed as:

$$\text{ACC }(\%) = \frac{\text{TP} + \text{TN}}{N} \times 100 \tag{10}$$

Recall:

Recall measures the proportion of true positive predicted class out of all actual positive classes, which is mathematically expressed as:

$$\text{Rec }(\%) = \frac{\text{TP}}{R_1} \times 100 \tag{11}$$

Precision:

Precision measures the proportion of true positive predictions (i.e., correctly identifying ADHD with coexisting ASD) out of all positive predictions classes and is mathematically expressed as:

$$\text{Prec }(\%) = \frac{\text{TP}}{C_1} \times 100 \tag{12}$$

F1-score:

FS is computed based on the value of recall and precision using the following formula:

$$\text{FS }(\%) = \frac{2TP}{2TP + FP + FN} \times 100 \tag{13}$$

IV. EXPERIMENTS RESULTS

We organized the experimental results section into two subsections. Firstly, we described the statistical baseline characteristics of ADHD children with coexisting ASD and healthy children. Subsequently, we discussed the experimental results for the classification models that included a comprehensive performance analysis of the classifiers by considering all features and selecting optimal features.

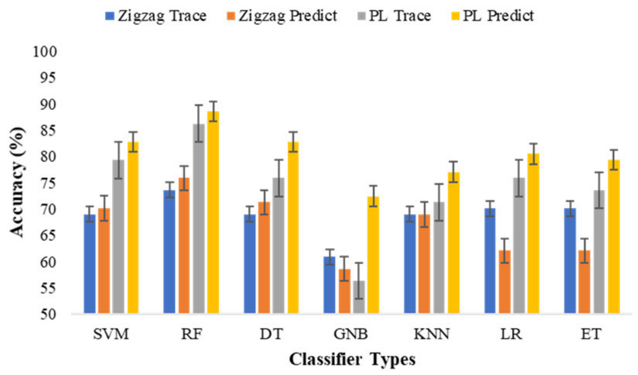

FIGURE 5. Classification accuracy of seven classifiers for all features over four tasks.

A. STATISTICAL BASELINE CHARACTERISTICS OF ADHD WITH COEXISTING ASD AND HC CHILDREN

The statistical baseline characteristics of ADHD with coexisting ASD and healthy children are presented in Table 5. As shown in Table 5, the overall prevalence of ADHD children with coexisting ASD and healthy children were 48.28% and 51.72% and their average ages were 8.57$\pm$ 2.24 and 9.92 $\pm$ 1.23 years. Moreover, we found that male children were more likely to have ADHD with coexisting ASD problems compared to females. Furthermore, about 57.89% of male children had ADHD with coexisting ASD problems and the remaining male children were healthy. It was observed that age and gender had no significant difference between ADHD children with coexisting ASD and healthy children.

B. EXPERIMENTAL RESULTS OF CLASSIFICATION MODELS

In this study, we performed two experiments for predicting ADHD children with coexisting ASD: (1) all feature-based performance analysis for predicting ADHD children with coexisting ASD and (ii) significant feature-based performance analysis for detecting ADHD children with coexisting ASD. The results of these two experiments are more clearly explained in the following subsections:

1) EXPERIMENT-I: ALL FEATURE-BASED PERFORMANCE ANALYSIS

The purpose of this experiment was to examine the performance of seven classifiers for identifying ADHD children with coexisting ASD by considering all 30 features, which were extracted from six raw features (See in Table 2). Next, seven ML-based classifiers (SVM, RF, DT, GNB, KNN, LR, and ET) with LOOCV were trained and the hyperparameters of each classifier were optimized to detect ADHD children with coexisting ASD. At the same time, we computed the classification accuracy of each classifier over four tasks. The classification accuracy of seven classifiers over four tasks is

TABLE 4. Confusion matrix.

		Predicted Class		Total
		ADHD with coexisting ASD	Healthy children	
Actual Class	ADHD with coexisting ASD	TP	FN	R_1=TP+FN
	Healthy children	FP	TN	R_2=FP+TN
	Total	C_1=TP+FP	C_2=FN+TN	N= R_1+ R_2= C_1+ C_2

TABLE 5. Baseline characteristics of ADHD with coexisting ASD and healthy children.

Variables	Overall	ADHD with coexisting ASD	Healthy children	Statistics	p-value[1]
Total, n (%)	29	14 (48.28)	15 (51.72)		
Age, Mean ± SD	9.27 ± 1.90	8.57 ± 2.24	9.92 ± 1.23	t=2.02, df=27	0.053
Gender, Male (%)	19 (70.4)	11 (57.89)	8 (42.11)	X^2=2.04, df=1	0.121

df: degrees of freedom; [1]p-value is gained from independent t-test for age variable and chi-square test for gender variable

TABLE 6. Classification accuracy (in %) of seven classifiers for all features over four tasks.

CT	Zigzag Trace	Zigzag Predict	PL Trace	PL Predict
SVM	68.97	70.11	79.31	82.76
RF	73.56	75.86	86.21	88.51
DT	68.97	71.26	75.86	82.76
GNB	60.92	58.62	56.32	72.41
KNN	68.97	68.97	71.26	77.01
LR	70.11	62.07	75.86	80.46
ET	70.11	62.07	73.56	79.31

CT: Classifier Types and PL: Plain Line

shown in Table 6, and their corresponding results are also shown in Fig. 5

As shown in Table 6 and Fig. 5, we observed that the RF-based model achieved the highest classification accuracy for all tasks compared to other classifiers. More specifically, RF obtained 73.56% accuracy for zigzag trace, 75.86% accuracy for zigzag predict, 86.21% accuracy for PL trace, and 88.51% accuracy for PL predict, respectively. We also observed that the RF-based model produced a higher accuracy (88.51%) performance for PL predict tasks than the rest of the other tasks and classifiers.

2) EXPERIMENT-II: SIGNIFICANT FEATURE-BASED PERFORMANCE ANALYSIS

This section showed the performance analysis of seven classifiers for the identification of ADHD children with coexisting ASD by selecting optimal or significant features. The optimal features were selected by SFFS. In this work, we employed SFFS with seven classifiers (SVM, RF, DT, GNB, KNN, LR, and ET). Moreover, we trained these classifiers separately with LOOCV and also optimized their hyperparameters. At the same time, we computed the classification accuracy of each classifier and tried to determine the most effective combination of features that yielded the highest classification accuracy. The classification accuracy results of SFFS with seven classifiers for four tasks are presented in Table 7. Their correspondence results are also illustrated in Fig. 6.

It was noticed that SFFS with an RF-based classifier produced better classification accuracy for all tasks than other classifiers. But, SFFS with RF-based classifier selected 9 features and produced the highest classification accuracy of 93.10% for the PL prediction task compared to other tasks

TABLE 7. Classification accuracy (in %) of seven classifiers of individual task for optimal features.

CT	Zigzag Trace		Zigzag Predict		PL Trace		PL Predict	
	NSF	ACC	NSF	ACC	NSF	ACC	NSF	ACC
SVM	10	77.01	6	80.46	16	87.36	5	86.21
RF	5	87.36	5	85.06	7	88.51	9	93.10
DT	2	82.76	16	82.76	14	86.21	3	91.95
GNB	8	78.16	2	66.67	10	79.31	16	82.76
KNN	6	82.76	8	82.76	3	81.61	5	90.80
LR	13	75.86	5	62.07	17	81.61	5	88.51
ET	2	77.01	5	77.01	5	80.46	10	88.51

NSF: Number of Selected Features; CT: Classifier Types; PL: Periodic Line

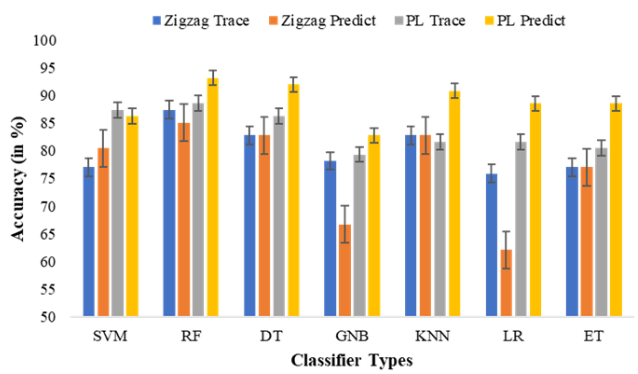

FIGURE 6. Classification accuracy of seven classifiers for optimal features over four tasks.

and classifiers. For PL predict task and selected 9 features, the classification accuracy of 86.21%, 91.95%, 82.76%, 90.80%, 88.51%, and 88.51% were obtained by SVM, DT, NB, k-NN, LR, and ET, respectively. On the other hand, SFFS with an LR-based classifier selected the combination of 5 features and obtained the lowest classification accuracy of 62.07% for the zigzag prediction task.

Moreover, the confusion matrix of the RF-based classifier for four tasks is presented in Fig. 7. Four performance evaluation parameters including recall, precision, FS, and AUC of SFFS with RF-based classifier for four tasks are presented in Table 8. We observed that the RF-based model obtained the highest performance scores (precision: 95.00%; FS: 92.68%; and AUC: 0.930) except recall was obtained by RF for predicting PL tasks compared to other tasks. We observed that the PL predict task with an RF-based model has more

TABLE 8. Four Performance parameters of proposed RF-based model.

Task Types	NSF	Precision (%)	Recall (%)	F1-Score (%)	AUC
Zigzag Trace	5	89.74	83.33	86.42	0.872
Zigzag Predict	5	85.37	83.33	84.34	0.850
PL Trace	7	84.78	92.86	88.64	0.887
PL Predict	9	95.00	90.48	92.68	0.930

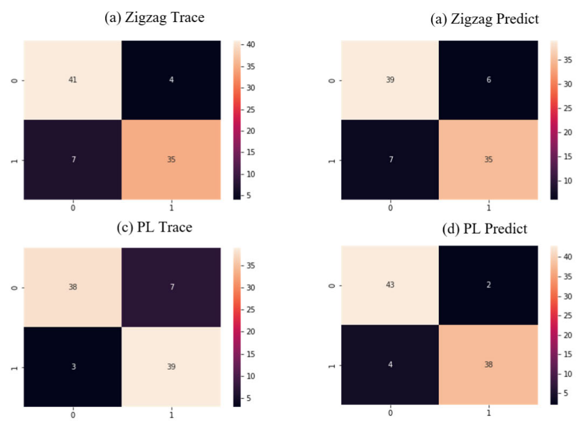

FIGURE 7. Confusion Matrix of RF-based model for: (a) Zigzag Trace; (b) Zigzag Trace; (c) PL Trace; and (d) PL Predict. Here, "1" represents ADHD children with coexisting ASD, and "0" represents healthy children.

discriminative power to discriminate ADHD with coexisting ASD children from healthy children.

V. DISCUSSION

ADHD for children is one of the most common psychiatric and behavioral disorders. Children with ADHD also suffered from various comorbidities and its prevalence has increased globally. Moreover, the cure for ADHD with other comorbidities were still unknown and only early detection can improve the quality of life. Existing studies were designed only for ADHD detection systems based on handwritten text [24], [25] and other diagnostic tools [41], [42], [43]. These handwritten-text-based systems were developed based on offline systems that required a specific language. As a result, non-native speakers faced a little bit problem to write down the text of these languages. In order to solve these problems, we designed a handwritten pattern with an ML-based algorithm to discriminate children with ADHD having ASD from healthy children.

In order to design this system, we performed some steps. Firstly, we asked children to draw four handwritten patterns (zigzag trace, zigzag predict, PL trace, and PL predict) on the pen tablet device using a stylus pen and repeat them three times. As a result, the pen tablet device generated six raw features, which are already discussed in the data collection procedure section. From these six raw features, we extracted thirty statistical features. Subsequently, we performed two experiments to conduct this study. The first experiment was to examine the performance of classifiers by considering all features. In order to perform this experiment, we adopted seven ML-based algorithms and trained these models with LOOCV protocol for all features. Our experimental results showed that PL predict-based patterns with an RF-based algorithm achieved outstanding performance than other classifiers and other tasks. The second experiment was to examine the discriminative power of classifiers by selecting significant features. We employed SFFS-based algorithms in order to select the subset of the relevant features, which were used in seven classifiers to discriminate ADHD children with coexisting ASD from healthy children. We also trained these classifiers with LOOCV protocol for four tasks and their results are shown in Table 7. Our experimental results also confirmed that the PL predict task with an RF-based algorithm also obtained outstanding performance compared to other algorithms and other tasks. Finally, we concluded that our proposed system has high discriminative power to detect ADHD children with coexisting ASD.

VI. FUTURE WORK DIRECTION

Despite this study obtaining promising results and its had still limitations. For example, this study used a relatively small number of subjects whose all subjects were confirmed to be right-handed. In our future work, we will extend this study by including more subjects and also considering left-handed subjects. Moreover, we will also implement deep learning-based algorithms to detect ADHD with coexisting ASD children.

VII. CONCLUSION

This paper proposed an ML-based ADHD with coexisting ASD detection system from handwriting patterns. In order to build this system, we performed the following experiments. First, we extracted 30 statistical features from the raw data and normalized them by Z-score. Second, we employed SFFS to determine relevant features and implemented a grid search technique to select the optimal parameters of the classification algorithms. This study used seven classification algorithms to discriminate ADHD children with ASD from healthy children for each task. The experimental results illustrated that the RF-based algorithm achieved 93.10% accuracy for the PL prediction task, which was comparatively higher than other classifiers and other tasks. This study suggests that the PL predict pattern with an RF-based classifier has a high discriminative to detect ADHD children with coexisting ASD. This study will be helpful for medical practitioners/physicians to detect children with ADHD having ASD at an early stage.

INSTITUTIONAL REVIEW BOARD STATEMENT

All procedures followed were according to the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. Ethical approval for this dataset was granted by the Interview Review Board (IRB) of Kumamoto University, Japan. (Approval Number: 45, Approval Date: 25 May 2021).

DATA AVAILABILITY

All codes of classification models and the data presented in the present study are publicly available at the following link: https://www.u-aizu.ac.jp/labs/is-pp/pplab/ADHD-ASD-HANDWRITING/ADHD-ASD-Dataset.zip. To ensure the anonymity of participants, the authors removed demographic characteristics (age and gender) from their data set.

REFERENCES

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, vol. 5, no. 5. Washington, DC, USA: American Psychiatric Association, 2013.

[2] I. Lazzaro, E. Gordon, W. Li, C. L. Lim, M. Plahn, S. Whitmont, S. Clarke, R. J. Barry, A. Dosen, and R. Meares, "Simultaneous EEG and EDA measures in adolescent attention deficit hyperactivity disorder," *Int. J. Psychophysiology*, vol. 34, no. 2, pp. 123–134, Nov. 1999.

[3] M. Altýnkaynak, N. Dolu, A. Güven, F. Pektaş, S. Özmen, E. Demirci, and M. Izzetoğlu, "Diagnosis of attention deficit hyperactivity disorder with combined time and frequency features," *Biocybernetics Biomed. Eng.*, vol. 40, no. 3, pp. 927–937, Jul. 2020.

[4] V. A. Harpin, "The effect of ADHD on the life of an individual, their family, and community from preschool to adult life," *Arch. Disease Childhood*, vol. 90, no. suppl_1, pp. i2–i7, Feb. 2005.

[5] R. Barkley, K. Murphy, and M. Fischer, *ADHD in Adults: What the Science Says*. Guilford Press, 2010.

[6] A. Stickley, A. Koyanagi, V. Ruchkin, and Y. Kamio, "Attention-deficit/hyperactivity disorder symptoms and suicide ideation and attempts: Findings from the adult psychiatric morbidity survey 2007," *J. Affect. Disorders*, vol. 189, pp. 321–328, Jan. 2016.

[7] M. Impey and R. Heun, "Completed suicide, ideation and attempt in attention deficit hyperactivity disorder," *Acta Psychiatrica Scandinavica*, vol. 125, no. 2, pp. 93–102, 2012.

[8] M. L. Danielson, R. H. Bitsko, R. M. Ghandour, J. R. Holbrook, M. D. Kogan, and S. J. Blumberg, "Prevalence of parent-reported ADHD diagnosis and associated treatment among US children and adolescents, 2016," *J. Clin. Child Adolescent Psychol.*, vol. 47, no. 2, pp. 199–212, Mar. 2018.

[9] Centers for Disease Control and Prevention (CDC), "Mental health in the United States. prevalence of diagnosis and medication treatment for attention-deficit/hyperactivity disorder–United States, 2003," *Morbidity Mortality Weekly Rep.*, vol. 54, no. 34, pp. 842–847, 2005.

[10] T. Torgersen, B. Gjervan, and K. Rasmussen, "ADHD in adults: A study of clinical characteristics, impairment and comorbidity," *Nordic J. Psychiatry*, vol. 60, no. 1, pp. 38–43, Jan. 2006.

[11] E. Sobanski, D. Brüggemann, B. Alm, S. Kern, M. Deschner, T. Schubert, A. Philipsen, and M. Rietschel, "Psychiatric comorbidity and functional impairment in a clinically referred sample of adults with attention-deficit/hyperactivity disorder (ADHD)," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 257, pp. 371–377, Oct. 2007.

[12] L. Reale, B. Bartoli, M. Cartabia, M. Zanetti, M. A. Costantino, M. P. Canevini, C. Termine, and M. Bonati, "Comorbidity prevalence and treatment outcome in children and adolescents with ADHD," *Eur. Child Adolescent Psychiatry*, vol. 26, no. 12, pp. 1443–1457, Dec. 2017.

[13] C. J. Vaidya, X. You, S. Mostofsky, F. Pereira, M. M. Berl, and L. Kenworthy, "Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders," *J. Child Psychol. Psychiatry.*, vol. 61, no. 1, pp. 51–61, 2020.

[14] X. Zhou, Q. Lin, Y. Gui, Z. Wang, M. Liu, and H. Lu, "Multimodal Mr images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning," *Frontiers Neurosci.*, vol. 15, pp. 710133–710147, Sep. 2021.

[15] A. Yasumura, M. Omori, A. Fukuda, J. Takahashi, Y. Yasumura, E. Nakagawa, T. Koike, Y. Yamashita, T. Miyajima, T. Koeda, M. Aihara, H. Tachimori, and M. Inagaki, "Applied machine learning method to predict children with ADHD using prefrontal cortex activity: A multicenter study in Japan," *J. Attention Disorders*, vol. 24, no. 14, pp. 2012–2020, Dec. 2020.

[16] M. Maniruzzaman, M. A. M. Hasan, N. Asai, and J. Shin, "Optimal channels and features selection based ADHD detection from EEG signal using statistical and machine learning techniques," *IEEE Access*, vol. 11, pp. 33570–33583, 2023.

[17] M. Maniruzzaman, J. Shin, M. A. M. Hasan, and A. Yasumura, "Efficient feature selection and machine learning based ADHD detection using EEG signal," *Comput., Mater. Continua*, vol. 72, no. 3, pp. 5179–5195, 2022.

[18] A. Müller, S. Vetsch, I. Pershin, G. Candrian, G.-M. Baschera, J. D. Kropotov, J. Kasper, H. A. Rehim, and D. Eich, "EEG/ERP-based biomarker/neuroalgorithms in adults with ADHD: Development, reliability, and application in clinical practice," *World J. Biol. Psychiatry*, vol. 21, no. 3, pp. 172–182, Mar. 2020.

[19] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with ADHD using behavioral activity: A machine learning analysis," *Appl. Sci.*, vol. 12, no. 5, pp. 2737–2749, 2022.

[20] S. Itani, M. Rossignol, F. Lecron, and P. Fortemps, "Towards interpretable machine learning models for diagnosis aid: A case study on attention deficit/hyperactivity disorder," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0215720.

[21] M. Adamou, S. L. Jones, L. Marks, and D. Lowe, "Efficacy of continuous performance testing in adult ADHD in a clinical sample using QbTest+," *J. Attention Disorders*, vol. 26, no. 11, pp. 1483–1491, Sep. 2022.

[22] M. O. Ogundele, H. F. Ayyash, and S. Banerjee, "Role of computerised continuous performance task tests in ADHD," *Prog. Neurol. Psychiatry*, vol. 15, no. 3, pp. 8–13, May 2011.

[23] M. B. Racine, A. Majnemer, M. Shevell, and L. Snider, "Handwriting performance in children with attention deficit hyperactivity disorder (ADHD)," *J. Child Neurol.*, vol. 23, no. 4, pp. 399–406, Apr. 2008.

[24] R. A. Langmaid, N. Papadopoulos, B. P. Johnson, J. G. Phillips, and N. J. Rinehart, "Handwriting in children with ADHD," *J. Attention Disorders*, vol. 18, no. 6, pp. 504–510, 2014.

[25] R. Cohen, B. Cohen-Kroitoru, A. Halevy, S. Aharoni, I. Aizenberg, and A. Shuper, "Handwriting in children with attention deficient hyperactive disorder: Role of graphology," *BMC Pediatrics*, vol. 19, no. 1, pp. 1–6, Dec. 2019.

[26] J. Shin, M. Maniruzzaman, Y. Uchida, M. A. M. Hasan, A. Megumi, A. Suzuki, and A. Yasumura, "Important features selection and classification of adult and child from handwriting using machine learning methods," *Appl. Sci.*, vol. 12, no. 10, pp. 5256–5270, 2022.

[27] J. Shin, M. A. M. Hasan, M. Maniruzzaman, A. Megumi, A. Suzuki, and A. Yasumura, "Online handwriting based adult and child classification using machine learning techniques," in *Proc. IEEE 5th Eurasian Conf. Educ. Innov. (ECEI)*, Feb. 2022, pp. 201–204.

[28] A. Megumi, A. Suzuki, J. Shin, and A. Yasumura. (2022). *Developmental Changes in Writing Dynamics and Its Relationship With ADHD and ASD Tendencies: A Preliminary Study*. [Online]. Available: https://doi.org/10.21203/rs.3.rs-1616383/v1

[29] V. Johansson, S. Sandin, Z. Chang, M. J. Taylor, P. Lichtenstein, B. M. D'Onofrio, H. Larsson, C. Hellner, and L. Halldner, "Medications for attention-deficit/hyperactivity disorder in individuals with or without coexisting autism spectrum disorder: Analysis of data from the Swedish prescribed drug register," *J. Neurodevelopmental Disorders*, vol. 12, no. 1, pp. 1–12, Dec. 2020.

[30] A. Wakabayashi, Y. Tojo, S. Baron-Cohen, and S. Wheelwright, "The autism-spectrum quotient (AQ) Japanese versions: Evidence from high-functioning clinical group and normal adults," *Jpn. J. Psychol.*, vol. 75, no. 1, pp. 78–84, 2004.

[31] A. Wakabayashi, S. Baron-Cohen, T. Uchiyama, Y. Yoshida, Y. Tojo, M. Kuroda, and S. Wheelwright, "The autism-spectrum quotient (AQ) children's version in Japan: A cross-cultural comparison," *J. Autism Develop. Disorders*, vol. 37, no. 3, pp. 491–500, Mar. 2007.

[32] G. J. DuPaul, T. J. Power, A. D. Anastopoulos, and R. Reid, *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. New York, NY, USA: Guilford Press, 1998.

[33] (2023). *Linear Regression*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

[34] S. Okser, T. Pahikkala, A. Airola, T. Salakoski, S. Ripatti, and T. Aittokallio, "Regularized machine learning in the genetic prediction of complex traits," *PLoS Genet.*, vol. 10, no. 11, Nov. 2014, Art. no. e1004754.

[35] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[36] M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ''Support vector machine and random forest modeling for intrusion detection system (IDS),'' *J. Intell. Learn. Syst. Appl.*, vol. 6, no. 1, pp. 45–52, 2014.

[37] S. U. Jan, Y.-D. Lee, J. Shin, and I. Koo, ''Sensor fault classification based on support vector machine and statistical time-domain features,'' *IEEE Access*, vol. 5, pp. 8682–8690, 2017.

[38] O. Z. Maimon and L. Rokach, *Data Mining With Decision Trees: Theory and Applications*, vol. 81. Singapore: World Scientific, 2014.

[39] T. Cover and P. Hart, ''Nearest neighbor pattern classification,'' *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[40] P. Geurts, D. Ernst, and L. Wehenkel, ''Extremely randomized trees,'' *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.

[41] H. Chen, W. Chen, Y. Song, L. Sun, and X. Li, ''EEG characteristics of children with attention-deficit/hyperactivity disorder,'' *Neuroscience*, vol. 406, pp. 444–456, May 2019.

[42] G. Guney, E. Kisacik, C. Kalaycioglu, and G. Saygili, ''Exploring the attention process differentiation of attention deficit hyperactivity disorder (ADHD) symptomatic adults using artificial intelligence on electroencephalography (EEG) signals,'' *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 29, no. 5, pp. 2312–2325, Sep. 2021.

[43] Y. Chen, Y. Tang, C. Wang, X. Liu, L. Zhao, and Z. Wang, ''ADHD classification by dual subspace learning using resting-state functional connectivity,'' *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101786.

JUNGPIL SHIN (Senior Member, IEEE) received the B.Sc. degree in computer science and statistics and the M.Sc. degree in computer science from Pusan National University, South Korea, in 1990 and 1994, respectively, and the Ph.D. degree in computer science and communication engineering from Kyushu University, Japan, in 1999, under a scholarship from the Japanese Government (MEXT). He was an Associate Professor, a Senior Associate Professor, and a Full Professor with the School of Computer Science and Engineering, The University of Aizu, Japan, in 1999, 2004, and 2019, respectively. He has coauthored more than 350 papers published in widely cited journals and conferences. His research interests include pattern recognition, image processing, computer vision, machine learning, human–computer interaction, non-touch interfaces, human gesture recognition, automatic control, Parkinson's disease diagnosis, ADHD diagnosis, user authentication, machine intelligence, and handwriting analysis, recognition, and synthesis. He is a member of ACM, IEICE, IPSJ, KISS, and KIPS. He served as the program chair and a program committee member for numerous international conferences. He serves as a reviewer for several major IEEE and SCI journals. He serves as an Editor for IEEE journals, Springer, Sage, Taylor & Francis, MDPI *Sensors* and *Electronics*, and Tech Science.

MD. MANIRUZZAMAN received the B.Sc., M.Sc., and M.Phil. degrees in statistics from the Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh, in 2013, 2014, and 2021, respectively. He became a Lecturer and an Assistant Professor with the Statistics Discipline, Khulna University, Khulna, Bangladesh, in September 2018 and November 2020, respectively. Currently, he is a Ph.D. Fellow with the Pattern Processing Laboratory, School of Computer Science and Engineering, The University of Aizu, Japan, under the direct supervision of Dr. Jungpil Shin. He has coauthored more than 40 papers published in widely cited journals and conferences. His research interests include bioinformatics, artificial intelligence, pattern recognition, medical image, signal processing, machine learning, data mining, and big data analysis.

YUTA UCHIDA received the bachelor's degree in computer science and engineering from The University of Aizu (UoA), Japan, in March 2022, where he is currently pursuing the master's degree. He joined the Pattern Processing Laboratory, UoA, in April 2020, under the direct supervision of Dr. Jungpil Shin. He is working on the analysis and recognition of ADHD. His research interests include computer vision, pattern recognition, and deep learning.

MD. AL MEHEDI HASAN (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer science and engineering from the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh, in 2005, 2007, and 2017, respectively. He became a Lecturer, an Assistant Professor, an Associate Professor, and a Professor with the Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology (RUET), Rajshahi, in 2007, 2010, 2018, and 2019, respectively. Recently, he has completed the postdoctoral research with the School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Japan. He has coauthored more than 130 publications published in widely cited journals and conferences. His research interests include bioinformatics, artificial intelligence, pattern recognition, medical image, signal processing, machine learning, computer vision, data mining, big data analysis, probabilistic and statistical inference, operating systems, computer networks, and security.

AKIKO MEGUMI received the B.Sc. degree in Japanese literature from the Doshisha Women's College of Liberal Arts, Japan, in 2004, and the M.Sc. and Ph.D. degrees in social and cultural sciences from Kumamoto University, Japan, in 2019 and 2023, respectively. She has clinical experience as a Licensed Clinical Psychologist and a Speech-Language-Hearing Therapist. She has coauthored nine papers published in widely cited journals and conferences. Her research interests include brain function and developmental disorders. She has received several awards.

AKIRA YASUMURA received the B.Sc. and M.Sc. degrees in technical education from the University of Teacher Education Fukuoka, Japan, in 2001 and 2003, respectively, and the Ph.D. degree in arts and sciences from The University of Tokyo, Japan, in 2013, under a scholarship from the Japanese Government (MEXT). He has been an Associate Professor of psychology with Kumamoto University, Japan, since 2018. He is a Licensed Clinical Psychologist. He has coauthored more than 40 papers published in widely cited journals and conferences. He has received numerous awards and holds patents. His research interests include brain function, developmental disorders, and machine learning.