Journal of Education for Pure Science- University of Thi-Qar
Vol.13, No.1 (March., 2023)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: http://doi.org/10.32792/utq.jceps.10.01.01

ADHD Detection Using Machine learning Algorithms and EEG brain signals

Aseel Tayeh Swadi [1] Firas Sabar Miften [2]

Msc21co2@utq.edu.iq firas@utq.edu.iq

[1, 2,] Department of Computer Science, College of Education for Pure Science, University of Thi-Qar, Iraq

Received 14/12/2022, Accepted 9/1/2023, Published / March /2023

Abstract:

Attention deficit hyperactivity disorder (ADHD) is a behavioral problem that can last into adulthood and affect children. Because it can show complicated brain activity, electroencephalography (EEG) plays a key role in determining the neurophysiology of ADHD. several statistical features are extracted from five frequency bands by using a discrete wavelet transform. The proposed system is evaluated by using K-means-based feature selection and 5 machine learning methods (Least-square support vector machine, k-nearest neighbor, Decision tree, and naive-Bayes classifier, support vector machine), so this system developed using a ten-fold cross-validation strategy and showed the testing accuracy for each classifier as (96.49%, 92.66%, 88.08%, 68.39%,53.71%), respectively.

Keywords: ADHD, EEG, machine learning, K-means.

1- Introduction:

One of the common behavioral diseases that affect children is Attention Deficit Hyperactivity Disorder (ADHD)[1]. Children with ADHD are unable to sit down, plan, complete tasks or understand what is happening around them. The prevalence of attention deficit hyperactivity disorder is estimated at 12.1% among males and 3.9% among females [2]. Adults with attention deficit hyperactivity disorder run the risk of engaging in antisocial conduct, abusing drugs and alcohol, and performing poorly in school and at work [3]. Early identification of children's attention deficit hyperactivity disorder results in early and effective therapy[4]. To confirm the observed behavior and symptoms, current diagnostic criteria for hyperactivity and attention deficit hyperactivity disorder are being tested [5]. Since these criteria are behavior-based, it can frequently be challenging to recognize preschool-aged children with ADHD, and the use of EEG is advised in these situations. EEG is a helpful technique that sheds light on brain activity in the background and measures the neural substrates of cognition and behavior [6]. The literature indicates that EEG plays a significant part in assessing the neurological function of children with ADHD [7]. In light of this, it might be a helpful tool for assessing and diagnosing the anomalous behavior of ADHD-affected kids.

2- RELATED WORKS

This section illustrates a few recent strategies in which many people with ADHD were identified. The estimation of ADHD based on EEG data has been done using a variety of methodologies. Laura Dubreuil-Vall et al. (2020) a method based on a convolutional neural network (CNN) with a four-layer architecture that combines filtering and clustering and achieved a classification accuracy of 88%[8]. M. Maya-Piedrahita et al (2020) proposed a supported methodology for the diagnosis of ADHD based on hidden Markov models (HMM) and dynamic EEG characterization of the probabilistic product nucleus (PPK). The proposed method has an accuracy rate of 90.0% [9]. Behrad TaghiBeyglou et al. (2020) proposed a method to detect ADHD by extracting the frequency and temporal characteristics and used KNN as a classifier to obtain an accuracy of 83.33% [10]. Arif Inizad et al. (2020) used graph signal processing (GSP) and graph learning (GL) techniques to extract structural and functional features. When they combined these combinations of features, they were able to achieve a high detection accuracy of 93.47% [11]. But Majid Moghaddari et al., (2020), created a deep learning model using a convolutional neural network, and then generated a three-channel RGB color image by extracting theta, alpha, beta, and gamma frequency bands from each segmented sample. The proposed model achieved an average accuracy of 98.48% [12]. Kathryn Joy et al. (2022) Based on several nonlinear entropy estimators and an artificial neural network classifier, investigate an efficient computer-assisted technical method for identifying ADHD from electroencephalography (EEG) signals [13]. Ali Al-Akhlasi et al. (2021), classified the two groups using the Effective Communication (EC) scale. Establishing an effective communication vector as a feature vector for the classification procedure relies on the ECM. Artificial neural networks, or multilayer artificial neural networks, were used during the classification process and the genetic algorithm was used to select the features. ECV g showed an accuracy of 89.1% based on ANN classification and GA results[14].

In this study, K-means was used to decide which traits to compare between two groups of healthy children and children with ADHD. The least squares support vector machine was utilized as the classifier.

3- Dataset

In this paper, we have used the dataset named " EEG data for ADHD/Control children "[15]. Participants were 60 healthy controls and 61 children with ADHD (boys and girls, ages 7-12). The DSM-IV criteria have been used by a professional psychiatrist to define ADHD in children who have been taking Ritalin for up to six months. None of the children in the control group had any reports of high-risk behavior, epilepsy, or psychiatric illness in the past. EEG recordings were made using 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) at 128Hz sampling rate based on 10-20 standards. The reference electrodes located on the .earlobes were A1 and A2

The EEG recording procedure was based on visual attention tasks because one of the deficiencies in children with ADHD is visual attention. The children were asked to count a group of cartoon characters they were shown as part of the activity. Each image contained between 5 and 16 letters, chosen at random. The drawings were large enough for children to clearly see and count the characters. Every image was displayed instantly and seamlessly The infant's response to providing a continuous stimulus while recording the signal. Thus, the child's performance during this visual-cognitive task determined how long (i.e., speed of response) the EEG was recorded. An illustration of one of these images is shown in Figure .

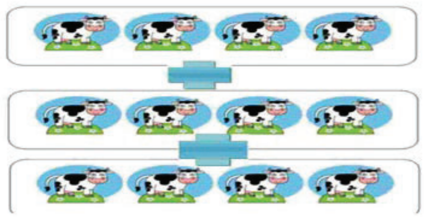

Figure 1: shows an illustration of an image that was given to the children while the cues were being recorded [16].

4- Methodology

The proper methodology is the key to successful research. The methodology used to conduct this study is shown in Figure 2

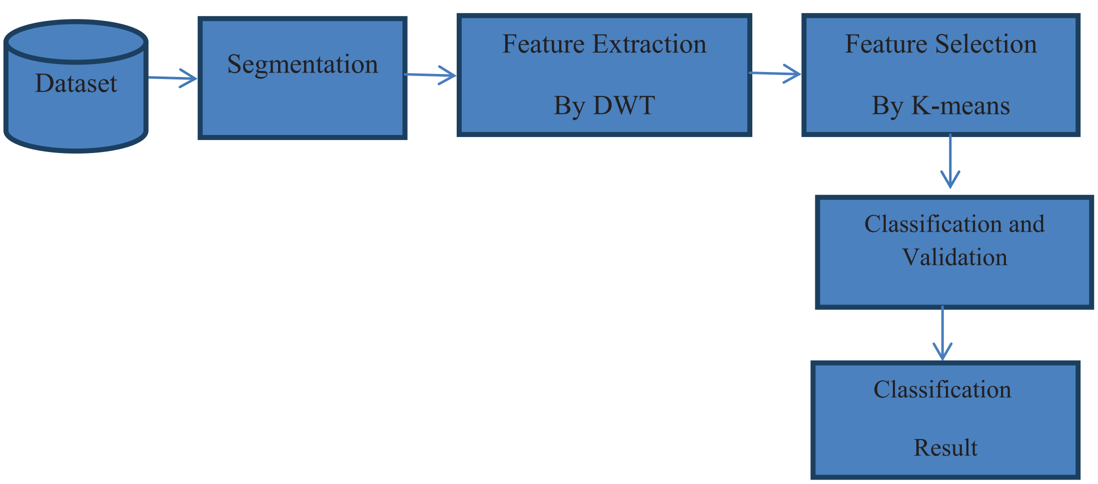

Figure 2 : shows the method's workflow for identifying ADHD+ Normal

4.1.Segmentation

The segmentation of EEG signals was done using a sliding window approach. This method used a sampling frequency of 128 HZ and a window size of 30 seconds. The EEG signal was divided into M windows, each of which included 3840 data points. These windows were stored in a matrix for the following step.

4.2.Feature Extraction

After segmentation, the array containing all segments is delivered to a feature extraction module where features are extracted to reduce the amount of data while retaining enough discriminatory information needed for classification. This is a necessary and important part of any rating system. When the extracted features contain enough discriminatory information, the performance of the classifier improves. From the frequency domain, five bands delta (δ), theta (ϴ), alpha (α), beta (β), and gamma (ɣ) are extracted using (DWT), and ten statistical features are extracted from each band, so that the sum of features is fifty features and send them to K-means clustering to identify the features of interest. Below is an explanation of the features and how they are used.

4.2.1 Wavelet transform

Wavelet transform provides insightful information about signals by transforming them into new domains. WT separates the signal into a weighted set of scaled functions. [17]:

$$f(a,b) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\,\psi\left(\frac{t-b}{a}\right) \;\ldots\ldots\ldots\ldots(1)$$

where (a) specifies the scale, and (b) specifies the shift (b).

DWT is one of the primary methods for signal transformation employed in this study, with using of two filter banks, low pass and high pass filters, it effectively separates signals into wavelets. The noise must be removed to recover important information from signals, hence the approximation and detailed DWT coefficients are determined next[18]. DWT was applied with eight levels before statistical features were extracted. The db8 filter was used to perform the decomposition in DWT. The db8 wavelet was chosen as the best wavelet to employ since it most closely matched the signal.

The discrete Wavelet Transformation (DWT) is defined by the equation below (Equation 2)

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\,\psi_{m,n}(t)\mathrm{dt} \;\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

4.2.2 statistical features

The collection, organization, presentation, analysis and interpretation of data are part of statistics. When you apply statistics to a scientific, industrial, or social problem, you usually start by looking at a statistical ensemble or statistical model. [19]. The ten statistical characteristics used in this study are (minimum, maximum, mean, median, mode, root mean square (RMS), variance, standard deviation, skewness, and kurtosis). After being extracted from the five frequency bands, the values of each feature are normalized to the range of 0 to 1, allowing easy application of a robust feature selection method. **Table1** explains each feature.

Table1. A list of statistical features with explanation[20]

Feature name	Equation	Feature name	Equation
1- Minimum	$X_{Min} = Min[x_n]$	6- root mean square	$X_{rms}=\sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$
2- Maximum	$X_{max} = Max[x_n]$	7- Variance	$X_{Var} = \frac{\sum_{i=1}^{N}\sum_{i=1}^{N}(x_i - mean)^2}{N-1}$
3- Mean	$X_{Mean} = \frac{1}{n}\sum_{1}^{n} x_i$	8- standard deviation	$X_{SD} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - mean)^2}{N-1}}$
4- Median	$X_{Me} = (\frac{n+1}{2})^{th}$	9- Skewness	$X_{Skew} = \frac{\sum_{i=1}^{N}(x_i - mean)^3 / N}{\sigma^3}$
5- Mode	$X_{Mo}= L+(\frac{f_1-f_0}{2f_1-f_0})xh$	10- kurtosis	$X_{Kurt} = E[(\frac{X-\mu}{\sigma})^4]$

4.3.Feature Selection

Feature Selection is the procedure for picking crucial features from a broad selection of features. The accuracy of model categorization might be decreased by using incorrect features. The benefit of feature selection is that it lowers overfitting and increases classification precision[21]. The feature selection approach used in this paper is based on k-mean clustering.

4.3.1 K-Means clustering

We have proposed a feature selection algorithm based on k-means clustering. K-means clustering was able to select the most influential features and eliminate unrelated traits [22]. In this experiment, 20 features were selected and used as inputs for different machine learning algorithms.

4.4 Classification and validation

After extracting and selecting features using the K-means method, machine learning algorithms were used to classify ADHD and healthy controls. To further evaluate the effectiveness of our model in this study, we used a 10-fold cross-validation technique. After classification, three measures: precision, sensitivity, and specificity were used to evaluate the performance of the classifiers.

A brief description of the algorithm used and of the validation is given below.

4.4.1. LS-SVM Classifier

LS-SVM is one of the most powerful algorithms which can be used for classification [23]. Numerous types of data have been categorized using it. All of the training points are used by the LS-SVM learning model. The following is a definition of the LS SVM's classification formula:

$y(x) = sig\left(\sum_{i=1}^{n} y_{\propto_i} k(x, x_i) + b\right)$ ....(3)

Where:

ᴡ: is a weight vector.

ь: is the bias term.

ϕ(X): is a nonlinear mapping function.

in order to distinguish between ADHD in two classes, this study used LS-SVM as a classifier to classify features gathered by the k-means technique in ADHD data. The features picked by k-means are accepted as inputs by the LS-SVM algorithm. The radial basis function RBF kernel function was selected to improve the performance of the LS-SVM. To achieve the desired performance, it is necessary to carefully choose two important LS-SVM parameters, $\gamma$and ∂^2. To achieve the best results, the LS-SVM was trained using various combinations of the parameters γ and ∂^2. In our experiment, the best categorization results were obtained with γ= 1000 and ∂^2 = 1.

4.4.2. KNN Classifier

Nearest neighbor classifiers are based on learning by analogy, i.e. comparing a given test set with similar training sets. Training sets are described by n attributes. Each group represents a point in n-dimensional space. In this way, all training sets are stored in an n-dimensional pattern space. When an unknown set is provided, the KNN classifier searches the pattern space for the set of k training sets closest to the unknown set. These k training sets are the KNN of the unknown set[24].

A distance metric, like Euclidean distance, is used to define "closeness." Between two points or tuples, such as X_1 = $(x_{11}, x_{12}, \ldots .\, . , x_{1n})$ and X_2 = $(x_{21}, x_{22}, \ldots \ldots , x_{2n})$, the Euclidean distance is

$dis(x_1\,, x_2)$ = $\sqrt{\sum_{i=1}^{n}(x_{i1} - x_{i2})^2}$...........(4)

4.4.3. Decision Tree

A decision tree is a tree structure similar to a flowchart. Each node in the tree represents the test attribute, each branch represents the test output, and each leaf represents a class or distribution of classes. The top node is the root node and consists of the classification rule from the root node to the single leaf. Thus, the decision tree can be easily transferred to classification rules. It has many algorithms, but the main idea is to use the top-down induction method. The most important part is choosing what attribute will be the node and also evaluating if the tree is correct[21].

4.4.4. Naive-Bayes classifier

One powerful classification method developed from Bayes' theorem with independence between predictors as an assumption. The Naïve bays classifier assumes that there is no relationship between the presence of a particular feature compared to other features. The state of a particular feature does not affect the state of another feature[25]. The posterior probability is calculated using the following formula or equation:

P(c|x) = (P(x|c) * P(c)) / P(x) …………….(5)

4.4.5. SVM Classifier

A method of classification, where each data item in an n-dimensional space is plotted as a point. It falls under the supervised machine learning model. Support vectors are found near the margins of the workbook[24].

4.4.6. Validation

We have used a k-fold validation technique to stratify our data, and the dataset is divided into k distinct subsets in order of k-fold (or folds) across the cross-validation. The model is trained using K-1 folds (K minus 1) and tested using the residual fold [31]. The obtained average performance for each of the k-test folds is the final output for model evaluation. In this case, we used cross-validation 10 times. As a result, we trained the model to run 10 times using (333 samples), then tested using the remaining 10% of the data (36 samples).

Three measures (accuracy, precision, sensitivity, and specificity) were used post-classification to assess the performance of the classifiers.

Accuracy= $\frac{TP+TN}{TP+FN+TN+FP}$x 100% ………(6)

$\text{Precision} = \frac{TP}{(TP+FP)} \times 100\%$ …………..(7)

$\text{Sensitivity} = \frac{TP}{(TP+FN)} \times 100\%$.................(8)

$\text{Specificity} = \frac{TN}{(TN+FP)} \times 100\%$..................(9)

where the number of true positives, true negatives, false positives and false negatives, respectively, are denoted as TP, TN, FP and FN [26].

5. Classification results

After applying all of the classifiers, we can present the classification results for 19 channels after applying 10-fold cross-validation in a set of tables. Table 2 shows the classification results for LS-SVM Classifier, and Table 3 shows the results for the KNN classifier. Table 4, Table 5, and Table 6 show classification results for the Decision tree, Naive Bayes, and SVM classifiers ,respectively.

Table 2: shows the classification results for LS-SVM Classifier

Channel Number	Channel Labels	Accuracy	Precision	Sensitivity	Specificity
1	Fp1	95.28	94.12%	88.89%	94.44%
2	Fp2	95.83	73.33%	100%	84%
3	F3	96.67	100%	93.75%	100%
4	F4	96.11	94.12%	94.12%	94.74%
5	C3	93.89	100%	95.24%	100%
6	C4	96.39	100%	94.12%	100%
7	P3	98.06	100%	100%	100%
8	P4	97.78	100%	100%	100%
9	O1	96.67	93.33%	100%	95.45%
10	O2	96.67	100%	93.33%	100%
11	F7	95.83	94.44%	100%	94.74%
12	F8	96.11	93.33%	93.33%	95.24%
13	T7	97.22	100%	100%	100%
14	T8	93.61	94.44%	89.47%	94.12%
15	P7	98.06	100%	81.82%	100%
16	P8	97.78	100%	100%	100%
17	Fz	96.67	90.91%	100%	0.9615%
18	Cz	96.67	86.67%	100%	91.30%
19	Pz	98.06	94.74%	100%	94.44%

Table 3: shows the classification results for KNN Classifier

Channel Number	Channel Labels	Accuracy	Precision	Sensitivity	Specificity
1	Fp1	90.83%	100%	92.31%	100%
2	Fp2	92.50%	94.44%	89.47%	94.12%
3	F3	92.50%	88.24%	100%	90.48%
4	F4	%93.06	100%	93.75%	100%
5	C3	90.00%	80.00%	92.31%	86.96%
6	C4	91.94%	82.35%	93.33%	85.71%
7	P3	95.56%	94.74%	100%	94.44%
8	P4	93.61%	85.00%	94.44%	93.33%
9	O1	93.33%	93.75%	100%	95.24%
10	O2	91.39%	100%	84.21%	100%
11	F7	92.22%	100%	100%	100%
12	F8	90.28%	76.47%	100%	82.61%
13	T7	93.33%	100%	83.33%	100%
14	T8	91.67%	71.43%	71.43%	81.82%
15	P7	96.67%	100%	93.75	100%
16	P8	92.22%	88.89%	94.12%	89.47%
17	Fz	93.06%	93.33%	93.33%	95.24%
18	Cz	92.22%	72.22%	100%	78.26%
19	Pz	94.17%	86.67%	92.86%	90.91%

Table 4: shows the classification results for Decision tree Classifier

Channel Number	Channel Labels	Accuracy	Precision	Sensitivity	Specificity
1	Fp1	84.44%	83.33%	71.43%	90.91%
2	Fp2	87.78%	94.44%	100%	94.47%
3	F3	88.33%	57.89%	84.62%	65.22%
4	F4	88.61%	100%	80.00%	100%
5	C3	87.22%	88.24%	93.75%	90.00%
6	C4	86.39%	86.67%	76.47%	89.47%
7	P3	89.72%	80.00%	80.00%	85.71%
8	P4	83.06%	93.33%	82.35%	94.74%
9	O1	88.89%	89.47%	100%	89.47%
10	O2	87.50%	100%	93.75%	100%
11	F7	87.50%	82.35%	93.33%	85.71%
12	F8	90.83%	75.00%	100%	76.19%
13	T7	90.83%	81.25%	100%	86.96%
14	T8	83.89%	93.75%	75.00%	93.75%
15	P7	89.72%	78.75%	91.67%	87.50%
16	P8	92.50%	94.12%	94.12%	94.74%
17	Fz	86.94%	86.67%	81.25%	90.00%
18	Cz	90.56%	100%	87.50%	100%
19	Pz	88.89%	60.00%	60.00%	84.62%

Table 5: shows the classification results for Naïve-Bayes Classifier

Channel Number	Channel Labels	Accuracy	Precision	Sensitivity	Specificity
1	Fp1	65.83%	33.33%	57.14%	72.41%
2	Fp2	65.56%	33.33%	75.00%	57.14%
3	F3	66.39%	47.37%	81.82%	60.00%
4	F4	70.83%	58.33%	78.77%	81.48%
5	C3	73.61%	58.82%	100%	73.08%
6	C4	65.56%	40.00%	54.55%	64.00%
7	P3	71.39%	35.29%	75.00%	60.71%
8	P4	67.78%	40.00%	54.55%	64.00%
9	O1	66.94%	33.33%	83.33%	66.67%
10	O2	68.06%	42.11%	66.67%	54.17%
11	F7	68.335	40.00%	75.00%	67.86%
12	F8	66.39%	41.18%	70.00%	61.54%
13	T7	69.44%	41.18%	70.00%	61.54%
14	T8	66.67%	37.50%	85.71%	65.52%
15	P7	73.06%	42.86%	75.00%	71.43%
16	P8	66.94%	29.41%	71.43%	58.62%

17	Fz	69.17%	40.00%	66.67%	66.67%
18	Cz	65.28%	28.57%	44.44%	62.96%
19	Pz	72.22%	50.00%	71.43%	82.76%

Table 6: shows the classification results for SVM Classifier

Channel Number	Channel Labels	Accuracy	Precision	Sensitivity	Specificity
1	Fp1	58.61%	06.67%	50.00%	58.82%
2	Fp2	55.28%	83.33%	40.00%	81.82%
3	F3	59.17%	73.68%	58.33%	58.33%
4	F4	60.83%	91.67%	61.11%	94.44%
5	C3	62.50%	16.67%	75.00%	53.13%
6	C4	54.72%	83.33%	47.62%	86.67%
7	P3	63.33%	25.00%	100.00%	51.61%
8	P4	47.22%	75.00%	48.00%	63.64%
9	O1	45.56%	26.67%	57.14%	62.07%
10	O2	45.00%	28.57%	66.67%	66.67%
11	F7	44.44%	40.00%	53.33%	42.86%
12	F8	55.56%	66.67%	33.33%	16.67%
13	T7	67.50%	33.33%	33.33%	33.3%
14	T8	58.33%	40.00%	53.33%	42.86%
15	P7	69.72%	26.67%	57.14%	62.07%
16	P8	68.33%	89.47%	70.83%	83.33%
17	Fz	58.89%	40.00%	53.33%	42.86%
18	Cz	55.28%	83.33%	40.00%	81.82%
19	Pz	43.61%	20.00%	30.00%	53.85%

6- Conclusion and Discussion

ADHD is a behavioral problem that can last into adulthood and affect children. An early diagnosis of the disorder can aid in its treatment and management. We applied popular machine learning algorithms, such as LS-SVM, KNN, Decision Tree, Nave Bayes, and SVM to data of ADHD patients based on the important features selected by K-means. The performance validation metrics are accuracy, precision, Sensitivity, and Specificity. showed that the LS-SVM algorithm performed best in diagnosing ADHD with a mean accuracy of 96.49 percent, followed by the KNN algorithm with an accuracy of 92.66 percent. Table 7 and Figure 4 illustrate the mean of accuracy, precision, Sensitivity, and Specificity for each classifier. Thus, based on the data, it is concluded that the LS-SVM algorithm is the best suitable algorithm for ADHD diagnosis.

Table 7. Typical performance test results for machine learning algorithms using Selected features by K-mean cluster.

S.N.	Classifier	Accuracy	Precision	Sensitivity	Specificity
1	LS-SVM	96.49%	95.23%	96.00%	91.54%
2	KNN	92.66%	89.89%	93.08%	92.55%
3	Decision tree	88.08%	85.52%	86.59%	89.44%
4	Naïve Bayes	68.39%	40.68%	71.39%	65.92%
5	SVM	56.52%	50.00%	54.13%	59.83%

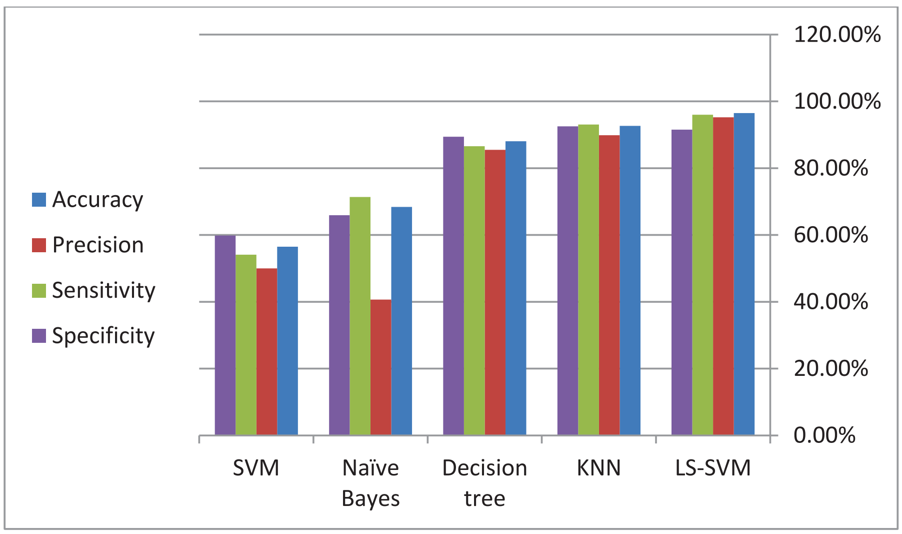

Figure 4: Typical performance test results for machine learning algorithms using Selected features by K-mean cluster.

REFERENCES

[1] **K. Miyoshi and Y. Morimura**, "Clinical manifestations of neuropsychiatric disorders," in *Neuropsychiatric disorders*, Springer, 2010, pp. 1–14.

[2] **K. H. Taber, R. A. Hurley, and S. C. Yudofsky,** "Diagnosis and treatment of neuropsychiatric disorders," *Annu. Rev. Med.*, vol. 61, pp. 121–133, 2010.

[3] **L. C. Fonseca, G. M. A. S. Tedrus, C. de Moraes, A. de V. Machado, M. P. de Almeida, and D. O. F. de Oliveira**, "Epileptiform abnormalities and quantitative EEG in children with attention-deficit/hyperactivity disorder," *Arq. Neuropsiquiatr.*, vol. 66, pp. 462–467, 2008.

[4] **R. C. Kessler *et al*.**, “The effects of temporally secondary co-morbid mental disorders on the associations of DSM-IV ADHD with adverse outcomes in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A),” *Psychol. Med.*, vol. 44, no. 8, pp. 1779–1792, 2014.

[5] **R. B. Tenenbaum et al**., “Response inhibition, response execution, and emotion regulation among children with attention-deficit/hyperactivity disorder,” *J. Abnorm. Child Psychol.*, vol. 47, no. 4, pp. 589–603, 2019.

[6] **F. Lenzi, S. Cortese, J. Harris, and G. Masi**, “Pharmacotherapy of emotional dysregulation in adults with ADHD: a systematic review and meta-analysis,” *Neurosci. Biobehav. Rev.*, vol. 84, pp. 359–367, 2018.

[7] **A. Thapar, M. Cooper, R. Jefferies, and E. Stergiakouli**, “What causes attention deficit hyperactivity disorder?,” *Arch. Dis. Child.*, vol. 97, no. 3, pp. 260–265, 2012, doi: 10.1136/archdischild-2011-300482.

[8] **L. Dubreuil-Vall, G. Ruffini, and J. A. Camprodon**, “Deep Learning Convolutional Neural Networks Discriminate Adult ADHD From Healthy Individuals on the Basis of Event-Related Spectral EEG,” *Front. Neurosci.*, vol. 14, no. April, pp. 1–12, 2020, doi: 10.3389/fnins.2020.00251.

[9] **M. C. Maya-Piedrahita, D. Cárdenas-Peña, and A. A. Orozco-Gutierrez**, “Diagnosis of attention deficit and hyperactivity disorder (ADHD) using hidden Markov models,” in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1205–1209.

[10] **B. Taghibeyglou, N. Hasanzadeh, F. Bagheri, and M. Jahed,** “ADHD diagnosis in children using common spatial pattern and nonlinear analysis of filter banked EEG,” *2020 28th Iran. Conf. Electr. Eng. ICEE 2020*, pp. 0–4, 2020, doi: 10.1109/ICEE50131.2020.9260711.

[11] **A. Einizade, M. Mozafari, M. Rezaei-Dastjerdehei, E. Aghdaei, A. M. Mijani, and S. Hajipour Sardouie**, “Detecting ADHD children based on EEG signals using Graph Signal Processing techniques,” *27th Natl. 5th Int. Iran. Conf. Biomed. Eng. ICBME 2020*, no. November, pp. 264–270, 2020, doi: 10.1109/ICBME51989.2020.9319456.

[12] **M. Moghaddari, M. Z. Lighvan, and S. Danishvar**, “Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG,” *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, doi: 10.1016/j.cmpb.2020.105738.

[13] **R. Catherine Joy, S. Thomas George, A. Albert Rajan, and M. S. P. Subathra**, “Detection of adhd from eeg signals using different entropy measures and ann,” *Clin. EEG Neurosci.*, vol. 53, no. 1, pp. 12–23, 2022.

[14] **A. Ekhlasi, A. M. Nasrabadi, and M. Mohammadi**, “Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals,” *Front. Biomed. Technol.*, vol. 8, no. 2, pp. 115–122, 2021, doi: 10.18502/fbt.v8i2.6515.

[15] **A. M. Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi**, "EEG data for ADHD/Control children," *IEEE Dataport*, 2020.

[16] **A. Allahverdy, A. K. Moghaddam, M. R. Mohammadi, and A. M. Nasrabadi**, "Detecting ADHD Children using the Attention Continuity as Nonlinear Feature of EEG," 2016.

[17] **W. M. Lwin, T. Aung**, and K. K. Wai, "Applications of Wavelet Transform," 2019.

[18] **A. Teolis and J. J. Benedetto**, *Computational signal processing with wavelets*, vol. 182. Springer, 1998.

[19] **S. Agarwal,** *Data mining: Data mining concepts and techniques*. 2014. doi: 10.1109/ICMIRA.2013.45.

[20] **A. K. Jain, R. P. W. Duin, and J. Mao**, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, 2000.

[21] **S. Natek and M. Zwilling**, "Student data mining solution–knowledge management system related to higher education institutions," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6400–6407, 2014.

[22] **R. Xu and D. Wunsch**, "Survey of clustering algorithms," *IEEE Trans. neural networks*, vol. 16, no. 3, pp. 645–678, 2005.

[23] **J. A. K. Suykens, T. Van Gestel, J. De Brabanter,** B. De Moor, and J. Vandewalle, "Least squares support vector machines, World Scientific Publishing, Singapore," 2002.

[24] **J. Han, M. Kamber, and J. Pei,** "Data Mining: Concepts and Techniques [Internet]. Waltham." Elsevier, 2011.

[25] **A. Burkov,** *The hundred-page machine learning book*, vol. 1. Andriy Burkov Quebec City, QC, Canada, 2019.

[26] **R. Kohavi, "Glossary of terms,"** *Spec. issue Appl. Mach. Learn. Knowl. Discov. Process*, vol. 30, no. 271, pp. 127–132, 1998.