Classification of adult autistic spectrum disorder using machine learning approach

Nurul Amirah Mashudi, Norulhusna Ahmad, Norliza Mohd Noor
Department of Engineering, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Malaysia

Article Info

ABSTRACT

Autism spectrum disorder (ASD) is a neurological-related disorder. Patients with ASD have poor social interaction and lack of communication that lead to restricted activities. Thus, early diagnosis with a reliable system is crucial as the symptoms may affect the patient's entire lifetime. Machine learning approaches are an effective and efficient method for the prediction of ASD disease. The study mainly aims to achieve the accuracy of ASD classification using a variety of machine learning approaches. The dataset comprises 16 selected attributes that are inclusive of 703 patients and non-patients. The experiments are performed within the simulation environment and analyzed using the Waikato environment for knowledge analysis (WEKA) platform. Linear support vector machine (SVM), k-nearest neighbours (k-NN), J48, Bagging, Stacking, AdaBoost, and naïve bayes are the methods used to compute the prediction of ASD status on the subject using 3, 5, and 10-folds cross validation. The analysis is then computed to evaluate the accuracy, sensitivity, and specificity of the proposed methods. The comparative result between the machine learning approaches has shown that linear SVM, J48, Bagging, Stacking, and naïve bayes produce the highest accuracy at 100% with the lowest error rate.

Corresponding Author:

Norliza Mohd Noor
Dept. of Engineering, Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia Kuala Lumpur Campus
Jalan Sultan Yahya Petra, 54100 Kuala Lumpur
Email: norliza@utm.my

1. INTRODUCTION

Autism spectrum disorder, or mainly known as ASD, is a neurological disease either in children or adults. Poor social interaction and communication have become more prevalent due to neurological disease [1]. Three psychological aspects of ASD are clinically diagnosed such as speech and language, mutual communication, and limited activities. ASD is identified in one lifespan and claimed as a psychological disorder in which the symptoms have occurred during the first two years [2]. Commonly, the beginning of ASD symptoms is during childhood and remains until the entire lifetime.

Furthermore, the potential factors for ASD are biological and environmental. Numerous diagnosis approaches have been applied for ASD such as autism diagnostic observation schedule-revised (ADOS-R) and autism diagnostic interview (ADI) [3], [4] autism quotient trait (AQ) [5], and social communication questionnaire (SCQ) [6]. Most of these approaches have employed mathematical formulas to diagnose accuracy. Thus, reliable clinical methods are highly demanded enhancing the accuracy and significant period to diagnose the disease [7].

Nevertheless, recent studies of ASD using machine learning did not foresee the conceptual, implementation, analysis, validation, and challenges. These challenges are not restricted to forms in which

diagnostic tool features are used. The implementation of machine learning must depend on various data cases that conflict with features, noise processing, feature extraction and selection, evaluation, and imbalance of training dataset imbalances of diagnostic types. Machine learning approaches are utilized as a clinical method to enhance the classification accuracy of the ASD disease as well as able to reduce the diagnostic time. Multiple objectives for ASD screening employ machine learning adaptation and diagnostic and statistical manual of mental disorders, 5th edition (DSM-5) as a screening tool [8]. The benefits and lacks of the screening have highlighted the issue related to the accuracy and ASD screening tools of DSM-5.

The machine learning classification method such as convolutional neural network (CNN), artificial neural network (ANN), k-nearest neighbours (k-NN), support vector machine (SVM), logistics regression, and naïve bayes used to evaluate and predict the ASD among children, adult, and adolescents [2]. The data was collected from the University of California, Irvine (UCI) repository. According to the analysis of results, CNN has produced a significant accuracy at 99.53% for the ASD adult dataset and 96.88% for ASD adolescent dataset. In comparison, 98.30% of the accuracy for ASD children's data was achieved by logistics regression, SVM, ANN, and CNN.

Five machine learning approaches which are naïve bayes, RBF SVM, linear SVM, C4.5, and random forest have been implemented to classify the ASD for the adult in particular kinematic parameters and experimental conditions [9]. The data used was a series of hand movements for 16 respondents of autism spectrum condition (ASC) adults approved by Manchester University. The result has shown the highest classification accuracy for linear SVM at 86.7%, respectively. Swarm intelligence based on the feature selection namely binary butterfly wrapper has been employed in the ASD dataset for the child, which consists of 21 data [10]. The purpose of their study was to boost the performance of classification accuracy by utilizing the naïve bayes, J48, SVM, k-NN, and multilayer perceptron (MLP). Based on the result, SVM has illustrated the highest classification accuracy of 97.95% after implementing the wrapper.

Data mining techniques such as decision tree, random forest, SVM, logistics regression, categorical lasso, and linear discriminant analysis (LDA) have been used to evaluate the classification accuracy based on the AUC using a dataset obtained from Simons Simplex Collection, Boston Autism Consortium, and Autism Genetic Resource Exchange, which contains 2775 subjects with autism [11]. The result obtained according to the feature selection methods was an accuracy of 96.5% using SVM. ASD screening in children dataset has been utilized to diagnose ASD disease in [12]. The dataset consists of 292 subjects with 141 patients diagnosed with ASD. LDA and k-NN have been used for the classification accuracy of the data. According to the result, LDA has shown better accuracy than k-NN at 90.80%, respectively.

ANN has also been employed to predict the classification accuracy of ASD [13]. The authors only used 14 attributes from the ASD screening in the adult dataset, which are age, gender, jaundice, 1-10 answers according to the screening questions, and the class/ASD. The result has predicted 100% accuracy for the ANN approach with a 0% error rate. The author in [14] has adopted a deep learning algorithm to identify the pattern between autistic children and normal children using electroencephalograms (EEG). The author has gained access to the dataset from University King Abdul Aziz, Jeddah. There are twenty files, twelve normal and eight autistics with the age between nine to sixteen years old, in the dataset. The subjects must be in a calm and relax condition to capture artifact-free EEG data. According to the result, the deep learning algorithm using CNN has produced a consistent accuracy result of 80%. Nonetheless, the CNN model implies a significant ability to improve the algorithms for a complex deep learning model.

Machine learning methods used in this study can substantially contribute new methods to diagnosis cases related to ASD. Furthermore, the methods can significantly minimize the features of current ASD methods without affecting the performance of specificity and accuracy, specifically for ASD screening in adults. Other than that, the three varieties of cross validations used in this study may produce the validity and usability of the machine learning approaches based on its performance.

The objective of the study is to achieve the accuracy of ASD classification based on the data collected from mobile apps ASD test [15], [16]. Individuals were required to answer the questions posed in the mobile apps ASD test. The data obtained were analyzed using machine learning approaches such as k-NN, linear SVM, naïve bayes, J48 decision tree, AdaBoost, Bagging, and Stacking, using ASD screening in adult data. The paper is structured by the following sections: section 2 provides the materials and methods used for each classifier, section 3 is the result and discussion for each classifier according to the classification performance, and section 4 is the conclusion.

2.   MATERIALS AND METHODS

The ASD for the adult dataset is justified including the ten questions of a personality questionnaire used to classify the symptoms of autism in this section. A brief description of the software platform, k-fold cross validation, data pre-processing, and performance evaluation are explained in the following sub-

sequence section in more detail. Last of all, the classification methods used in this study are described concisely.

2.1. ASD for adult dataset

In this study, we used data ASD screening data for adults obtained from the UCI machine learning repository [16]. The data was extracted using mobile application called ASD test conducted by Thabtah [8]. The dataset consists of 703 subjects with 21 features of adults' screening data for autism. The response class is categorized into two classes in which adults with ASD have 189 subjects, and adults without ASD have 515 subjects. Ten behavioral features are proven to be effective and reliable in differentiating ASD cases from the controls and 10 individual features. The features used in this study along with their types and description are given in Table 1.

The app enables users to analyze ASD behaviors using four modules [17]. One of the early autism screening tools intended to help adults to identify autistic symptoms in a personality questionnaire is the autism spectrum quotient (AQ) [18]. Initially, fifty questions of the AQ test concerned the cognitive skill of autism. Each question has four possibilities which are definitely agree, slightly agree, slightly disagree, and definitely disagree, however, there is a score on each question [17]. A shortened AQ test namely AQ-10-Adult that consisted of 10 questions was proposed by Allison *et al.* [19]. Nevertheless, during the screening process, users should choose the similar four options for each question in the AQ-10-Adult test to compute the score using diagnostic rules. An autistic person can be classified if any individual is scored more than six. There would be either 0 or 1 point for each question; a point is added when the answer to questions 1, 7, 8, and 10 are either "Slightly Agree" or "Definitely Agree". Additionally, there is either slightly or definitely disagree with the answer to question 2-6, and 9 [17].

Table 1. Features and its descriptions

Features	Type	Description
Age	Number	The age of the subjects
Gender	String	The individuality can be female or male
Ethnicity	String	The ethnicity of the subject
Jaundice	Boolean (yes or no)	If the case was diagnosed with jaundice
Autism	Boolean (yes or no)	If the close relatives have PDD
Relation	String	The person who completed the test such as the individual, parents, caretakers, and physicians
Country of residence	String	The country residence of the subject
Used app before	Boolean (yes or no)	If the person has used the screening application
AQ-1	Binary (0, 1)	The response is clarified based on the screening process
AQ-2	Binary (0, 1)	The response is clarified based on the screening process
AQ-3	Binary (0, 1)	The response is clarified based on the screening process
AQ-4	Binary (0, 1)	The response is clarified based on the screening process
AQ-5	Binary (0, 1)	The response is clarified based on the screening process
AQ-6	Binary (0, 1)	The response is clarified based on the screening process
AQ-7	Binary (0, 1)	The response is clarified based on the screening process
AQ-8	Binary (0, 1)	The response is clarified based on the screening process
AQ-9	Binary (0, 1)	The response is clarified based on the screening process
AQ-10	Binary (0, 1)	The response is clarified based on the screening process
Age description	Text	Age category
Screening score	Integer	The total score was determined using the implementation of the screening algorithm
Class/ASD	Boolean (yes or no)	The result is shown after the test

2.2. Methods

The process of the classification system in this study is illustrated in Figure 1, to present a better understanding of the implementation of the process. The explanation of each task is given in the sub-sequence section.
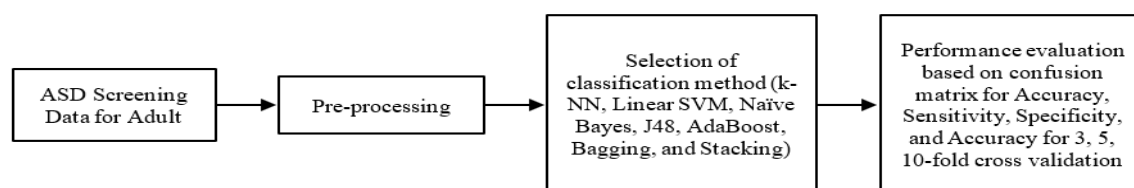

Figure 1. A classification system for ASD for the adult dataset

2.2.1. Software platform

The Waikato environment for knowledge analysis (WEKA) was used in this study to perform data pre-processing and classification of ASD for adult dataset [20]. WEKA is an open-source machine learning software using the JAVA programming language. Most researchers used the software as it supports numerous data mining functions for instance classification, clustering, association, data pre-processing, feature selection, and regression.

2.2.2. k-Fold cross validation

The ASD is divided into k subsets for adult data. In general, the data $(k-1)/k$ is used for training, and the data $1/k$ is used for the testing. Then, the process is reiterated k-times. As a final point, the validation result of mean k-time is selected as the last rate estimation. In this study, the performance is measured by 3, 5, and 10-folds cross validation, which is the ratio of training and testing at 67:33, 80:20, and 90:10, respectively.

2.2.3. Data pre-processing

In this study, all missing values were substituted for nominal and numerical data to tackle the issues of inadequate and incompatible data with missing values. Furthermore, data are filtered into nominal features using discretization to generate strong results for a variety of numerical features in data. The discretize equation is written is being as,

$$x[n] = x(n-1) + N\Delta t \tag{1}$$

where Δt is known as the step size or time step.

2.2.4. Performance evaluation

The performance of the classification model is measured by the amount of test data that are formulated using a confusion matrix based on correctly and incorrectly predicted models. The measurement of accuracy, sensitivity, and accuracy is then calculated from the confusion matrix. The confusion matrix of ASD and No-ASD is shown in Table 2.
− True positive (TP) is the data that the patient has identified with ASD.
− False positive (FP) is the data that non-patient has identified with ASD.
− False negative (FN) is the data that the patient has not identified with ASD.
− True negative (FN) is the data that the non-patient has not identified with ASD.

The equation of accuracy, sensitivity, and specificity are being as,

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \tag{2}$$

$$\text{Sensitivity} = \frac{TN}{FP+TN} \tag{3}$$

$$\text{Specificity} = \frac{TP}{TP+FN} \tag{4}$$

Table 2. Confusion matrix table

		Predicted class	
		ASD	No-ASD
Actual Class	ASD	TP	FN
	No-ASD	FP	TN

2.3. Classification methods
2.3.1. k-Nearest neighbours

k-Nearest Neighbours is also known as k-NN. It is a supervised machine learning technique to overcome challenges in classification and regression [21]. The number of classes in the dataset with a small value and positive integer is the initial value of the input parameter. The majority of neighbours are classified as the input data. The k-NN algorithm needs to run several times with different K values and choose the K that reduces the number of errors and maintains the prediction accuracy. Thus, in this case, the input parameter K of the ASD dataset is 3. A brute force search algorithm is implemented by using the Euclidean distance function for the nearest neighbour search as in (5). The function of Euclidean distance is used to compute the distance between instances that is good for numeric data on the same scale.

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5}$$

2.3.2. Linear support vector machine

Support vector machine or SVM is a supervised learning technique used for classification [22] and regression. Principally, the function of SVM to classify the outcomes by mapping data between input vectors to a huge perspective space. Thus, linear SVM aims to fully utilize the distance between the decision hyperplane and the marginal distance, which is the nearest data point [23]. In this study, SVM has implemented John Platt's sequential minimal optimization algorithm for training a support classifier as in (6). SVM is used to obtain the performance accuracy for ASD screening models. The linear regression is selected as the calibrator in the SVM classifier with the M5 selection method. Furthermore, PolyKernel is chosen as the kernel function.

$$min_\alpha \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum \alpha_i \tag{6}$$

2.3.3. Naïve bayes

Naïve bayes is one of the supervised machine learning approaches that are mainly known as Bayesian algorithms with a simple probability distribution [24]. The main principle of naïve bayes is focused on the expectations of freedom, which indicates less training time to be compared to the SVM approach. Furthermore, naïve bayes is also known as numeric estimator precision values that are chosen based on the analysis of the training data. The equation of naïve bayes is stated in (7).

$$P(c|x) = \frac{P(X|C)P(c)}{P(x)} \tag{7}$$

2.3.4. J48 decision tree

J48 decision tree is a comprehensive machine learning approach [25], which has been used by most researchers nowadays. Generally, J48 is used to develop a classification tree based on a hierarchical tree system, in which the decision results have illustrated the attributes and terminal nodes. The visual classification of the J48 approach is effective and efficient. Nevertheless, J48 is vulnerable to the noise in the data [26]. Variety of decision trees algorithms used for classification such as classification and regression tree (CART), chi-square automatic interaction detector (CHAID), ID3, and C4.5. Therefore, J48 is implemented in this study as one of the classification accuracies approaches.

2.3.5. AdaBoost

Adaptive boosting known as AdaBoost was developed by Freud and Schapire [27]. AdaBoost is a supervised learning algorithm of machine learning application. The core idea of AdaBoost is to match a sequence of weak learner models that are more effective than random guessing. Each instance in the training dataset is weighted to determine the accuracy either it is classified correctly or incorrectly. The decision stump is used as a classifier for AdaBoost models. The primary purpose of the decision stump is to boost the AdaBoost M1 nominal classifier. Only minor class problems can be tackled. The final prediction is then obtained from the combination of the predicted model based on a weighted majority vote (classification) or weighted sum (regression). In (8) shows the formula of AdaBoost.

$$E_t = \sum_i E[F_{i-1}(x_i) + a_t h(x_i)] \tag{8}$$

2.3.6. Bagging

Bagging is one of the most popular techniques in ensemble methods and is known as bootstrap aggregation. Bagging is the earliest and simplest algorithm developed by Breiman [28]. This method can be used to reduce the variance for the algorithms that have high variance such as decision trees. In this study, bagging is used to predict ASD disease. The equation of the bagging method is stated in (9). The fast decision tree learner algorithm is used as the default classifier to enhance the classification accuracy. The algorithm generates a decision tree and prunes it with a reduced-error with back fitting. The lack of values was coped with by dividing the corresponding instances into bits. The final decision tree was obtained as a composition of all base classifiers with the maximum votes.

$$f(x) = \frac{1}{M}\sum_{m=1}^{M} f_m(x) \tag{9}$$

2.3.7. Stacking

Stacking is an ensemble machine learning approach used to integrate either diversified classification or regression through meta-classifiers. The features on the results of the base level are prepared using a

proper training set that contains various machine learning approaches. Thus, stacking is a stratified approach. In this study, stacking is employed for ASD disease. Various classifiers are implemented in stacking such as 0-R, naïve bayes, logistic regression, sequential minimal optimization (SMO), k-NN ($k=3$), PART, fast decision tree learner, and J48 decision tree. PART decision list is selected as a meta-classifier in this study.

3. RESULTS AND DISCUSSION

Data pre-processing is the initial stage to be performed prior to the simulation for all models. According to the ASD for the adult dataset, there are few missing values in the features and one value for gender feature that contains an irrational number has caused inconsistent value. Thus, data with missing values have been omitted. Several features or attributes that do not contribute to autism were omitted to enhance the classification accuracy of ASD data such as ethnicity, country of residence, used app before, age description, and relation. Hence, the number of features used in this study has been reduced to 16 features which were age, gender, jaundice, autism, screening score, 1-10 questions related to autism behavioral features, and class/ASD. Once feature selection was executed, the numerical features were filtered into the nominal features using discretization that required all attribute indices. The classification process was then implemented using k-NN, linear SVM, naïve bayes, J48, AdaBoost, Bagging, and Stacking.

The confusion matrix was computed for each model to obtain the significant prediction of class. Confusion matrices associated with the seven different machine learning approaches are tabulated in Table 3. A 10-fold cross validation was carried out to predict the results. The confusion matrices Table shows some machine learning approaches used in this study that have produced the highest predicted class for ASD and non-ASD patients who have or have not identified with ASD disease. The machine learning approaches that influenced the best accuracy of the ASD class are linear SVM, naïve bayes, J48, Bagging, and Stacking.

The accuracy, sensitivity, and specificity of the classification methods were compared by the k-fold cross validation as tabulated in Table 4. The discretization techniques were performed to all k-fold cross validation throughout the pre-processing process. The classification accuracy results for each approach have increased when the k-fold cross validation was escalated. AdaBoost reported a similar result with 98.3% for 3-fold and 10-fold cross validation. Nevertheless, the 10-fold cross validation result presented better performance with the lowest error rate compared to the smaller k-fold cross validation. The finding indicates several proposed machine learning approaches have produced the best classification accuracy at 100%, respectively. Furthermore, the classification accuracy for Stacking and k-NN ($k=3$) methods were boosted from 99.7% to 100% and 98.6% to 99.2%, respectively, as the k-fold cross validation increases.

The classification accuracy for all machine learning approaches with 10-fold cross validation is demonstrated in Figure 2. The machine learning approaches; Stacking, Bagging, J48, and linear SVM have produced 100% without error rate. However, naïve bayes has produced accuracy results at 100% with a minimum error rate of 0.0028. As the result of these approaches have shown better performance for $k = 3$ and $k = 5$, thus the performance testing is sufficient to achieve at 3-fold cross validation.

Table 3. Confusion matrices of five different machine learning techniques

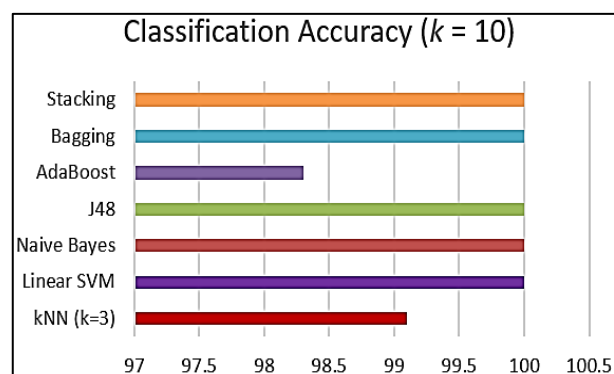| Methods	Predicted class (ASD)		Actual class
	No	Yes	
kNN (k=3)	512	2	N
	4	185	Y
Linear SVM	514	0	N
	0	189	Y
Naive Bayes	514	0	N
	0	189	Y
J48	514	0	N
	0	189	Y
AdaBoost	502	12	N
	0	189	Y
Bagging	514	0	N
	0	189	Y
Stacking	514	0	N
	0	189	Y

Figure 2. Classification accuracy of machine learning approaches for 10-fold cross validation

Altay and Ulas [12] have applied the k-NN method in the ASD for child dataset for comparative analysis. The study has conducted 70% of the training dataset and 30% of the testing dataset. Based on Table 5, the accuracy of our proposed k-NN method has shown significant accuracy for the adult dataset at

99.1% as compared to the accuracy result for the child dataset, due to different approaches for cross validation and various numbers and subjects in the dataset.

The implementation of SVM according to the literature has presented a variety of outcomes, including our proposed approach. The comparative result of the SVM method is classified in Table 6 shows our proposed approach in linear SVM has produced the highest accuracy compared to the methods in the literature. Li *et al.* have collected data for 16 autism spectrum condition (ASC) adults and 16 healthy adults [9]. The study was conducted for 40 means and standard deviations with eight situations and three questions. Nevertheless, the study has implemented two types of SVM, which are RGF and linear. The result has shown that linear SVM has high accuracy compared to RBF SVM. Thus, the different procedures and approaches in this study have influenced the outcome of accuracy. Vaishali and Sasikala [10] used ASD for the child dataset with 10-fold cross validation. However, the authors have performed binary firefly as a feature selection to optimize the performance process. The contradiction of subjects in the dataset and feature selection leads to the classification accuracy that produces 97.8% compared with our proposed approach. Moreover, to compare the linear SVM [10] and k-NN [9] result as both authors have implemented similar dataset, linear SVM has shown a better performance by using the selected feature.

Table 4. Comparison of the classification accuracy based on each k-fold cross validation

Methods	Accuracy (%)		
	$k = 3$	$k = 5$	$k = 10$
kNN (k=3)	98.6	99	99.2
Linear SVM	100	100	100
Naive Bayes	100	100	100
J48	100	100	100
AdaBoost	98.3	98.6	98.3
Bagging	100	100	100
Stacking	99.7	99.7	100

Table 5. Comparison of the classification accuracy using k-NN approach

Author(s)	Accuracy (%)
O. Altay and M. Ulas [12]	90.8
This study	99.1

Table 6. Comparison of the classification accuracy using SVM approach

Author(s)	Accuracy (%)
B. Li *et al.* [9]	86.7
R. Vaishali and R. Sasikala [10]	97.95
M. Duda *et al.* [11]	96.5
This study	100

A comparative analysis of the classification accuracy between the proposed method and in the literature is presented in Figure 3. The proposed method used in this study almost surpassed the other methods proposed by the author in literature. Linear SVM, naïve bayes, J48, bagging, and stacking have the same accuracy rate as a result in [13], which is 100% accuracy. Thus, the approaches have enhanced the accuracy rate from 3-fold cross validation to 10-fold cross validation.
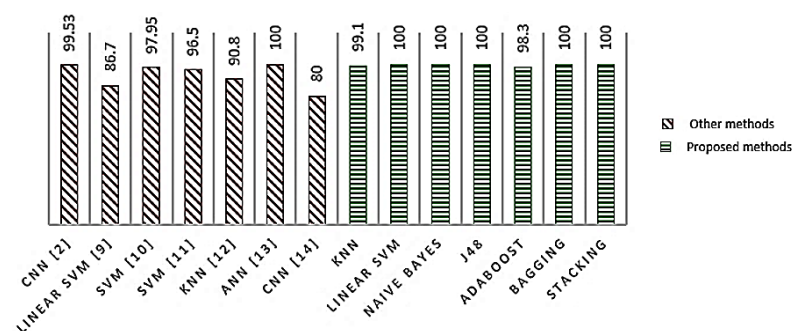

Figure 3. Classification accuracy based on the literature

Classification of adult autistic spectrum disorder using machine learning approach (Nurul Amirah Mashudi)

4. CONCLUSION

A cognitive disorder that prevents verbal and speech development, analytical, and social skills, is known as ASD. The potential factors for ASD are biological and environmental. A significant challenge for autism is upgrading the performance of the diagnostic forms in the current screening tools to minimize the diagnostics time effectively without affecting the validity or sensitivity of the test. The proposed study adopted the ASD screening data for adults to predict the classification model of ASD disease specifically for the adult who is ASD patient and non-patient that has or has not classified with ASD disease. Cross-validation was implemented with 3, 5, and 10-folds into the dataset. Thus, to evaluate the classification accuracy with other methods in the literature, only 10-fold cross validation was used. The data-preprocessing stage was performed through the dataset by replacing missing values and discretization afterward. Few features were omitted which have no significant value for the classification process. In addition to ASD studies, machine learning approaches showed strong findings in various applications. Machine learning approaches such as Bagging, Stacking, AdaBoost, linear SVM, naïve bayes, J48, and k-NN used to classify data correctly have therefore been proposed. According to the results, Bagging, Stacking, linear SVM, naïve bayes, and J48 have achieved a significant accuracy at 100%, respectively. The accuracy results in this study were compared to the previous works that used a variety of ASD repositories. Besides, accuracy, specificity, and sensitivity were also counted in this study to find the number of patients with ASD disease and without ASD disease. Therefore, machine learning methods used in this study can significantly contribute new methods to diagnosis cases related to ASD and minimize the features of current ASD methods without affecting the performance of specificity and accuracy of the test.

REFERENCES

[1] R. Sabzi *et al.*, "Recognizing Persian handwritten words using deep convolutional networks," *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Shiraz, Iran, 2017, pp. 85-90, doi: 10.1109/AISP.2017.8324114.

[2] S. Raj and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Computer Science,* vol. 167, pp. 994-1004, 2020, doi: 10.1016/j.procs.2020.03.399.

[3] C. Lord *et al.*, "The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders,* vol. 30, no. 3, pp. 205-223, 2000, doi: 10.1023/A:1005592401947.

[4] C. Lord, M. Rutter, and A. Le Couteur, "Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *Journal of autism and developmental disorders,* vol. 24, no. 5, pp. 659-685, 1994, doi: 10.1007/BF02172145.

[5] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians," *Journal of autism and developmental disorders,* vol. 31, no. 1, pp. 5-17, 2001, doi: 10.1023/A:1005653411471.

[6] S. K. Berument, M. Rutter, C. Lord, A. Pickles, and A. Bailey, "Autism screening questionnaire: diagnostic validity," *The British Journal of Psychiatry,* vol. 175, no. 5, pp. 444-451, 1999, doi: 10.1192/bjp.175.5.444.

[7] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics for Health and Social Care,* vol. 44, no. 3, pp. 278-297, 2019, doi: 10.1080/17538157.2017.1399132.

[8] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*, 2017, pp. 1-6, doi: 10.1145/3107514.3107515.

[9] B. Li, A. Sharma, J. Meng, S. Purushwalkam, and E. Gowen, "Applying machine learning to identify autistic adults using imitation: An exploratory study," *PloS one,* vol. 12, no. 8, Art. No. e0182652, 2017, doi: 10.1371/journal.pone.0182652.

[10] R. Vaishali and R. Sasikala, "A machine learning based approach to classify Autism with optimum behaviour sets," *International Journal of Engineering and Technology,* vol. 7, no. 4, pp. 1-6, 2018.

[11] M. Duda, R. Ma, N. Haber, and D. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," *Translational psychiatry,* vol. 6, no. 2, pp. e732-e732, 2016, doi: 10.1038/tp.2015.221.

[12] O. Altay and M. Ulas, "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children," *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, Antalya, Turkey, 2018, pp. 1-4, doi: 10.1109/ISDFS.2018.8355354.

[13] I. M. Nasser, M. Al-Shawwa, and S. S. Abu-Naser, "Artificial Neural Network for Diagnose Autism Spectrum Disorder," 2019.

[14] A. A. Nur, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," *IAES International Journal of Artificial Intelligence,* vol. 9, no. 1, pp. 91-99, 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[15] F. Thabtah, "ASDTests. A mobile app for ASD screening," 2017.

[16] F. F. Thabtah, "Autism Screening Adult Data Set," UCI Machine Learning Repository, 2017.

[17] J. Abuel-Reesh, "A knowledge based system for diagnosing shortness of breath in infants and children," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 1, no. 4, pp. 102-115, 2017.

[18] S. S. Abu-Naser, A. Almasri, Y. S. Abu Sultan, and I. S. Zaqout, "A prototype decision support system for optimizing the effectiveness of elearning in educational institutions," 2011.

[19] C. Allison, B. Auyeung, and S. Baron-Cohen, "Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls," *Journal of the American Academy of Child & Adolescent Psychiatry,* vol. 51, no. 2, pp. 202-212, 2012, doi: 10.1016/j.jaac.2011.11.003.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, no. 1, pp. 10-18, 2009, doi: 10.1145/1656274.1656278.

[21] P. Hall, B. U. Park, and R. J. Samworth, "Choice of neighbor order in nearest-neighbor classification," *The Annals of Statistics,* vol. 36, no. 5, pp. 2135-2152, 2008, doi: 10.1214/07-AOS537.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[23] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal,* vol. 13, pp. 8-17, 2015, doi: 10.1016/j.csbj.2014.11.005.

[24] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," *arXiv preprint arXiv:1302.4964,* 2013.

[25] V. P. Bresfelean, "Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment," *2007 29th International Conference on Information Technology Interfaces*, Cavtat, Croatia, 2007, pp. 51-56, doi: 10.1109/ITI.2007.4283743.

[26] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems,* vol. 24, no. 7, pp. 1024-1032, 2011, doi: 10.1016/j.knosys.2011.04.014.

[27] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119-139, 1997, doi: 10.1006/jcss.1997.1504.

[28] L. Breiman, "Heuristics of instability and stabilization in model selection," *The annals of statistics,* vol. 24, no. 6, pp. 2350-2383, 1996, doi: 10.1214/aos/1032181158.

BIOGRAPHIES OF AUTHORS

Nurul Amirah Mashudi received a Bachelor of Computer Science (Graphics and Multimedia Software) in 2016 from Universiti Teknologi Malaysia, Johor Bahru. Currently, she is pursuing her Master's Degree of Science (Systems Engineering) at Universiti Teknologi Malaysia, Kuala Lumpur. Her current research areas cover image processing, image analysis, machine learning application.

Dr. Norulhusna Ahmad received her Master's degree in Electrical Engineering (Telecommunication) and Ph.D. in Electrical Engineering in 2003 and 2014, respectively from Universiti Teknologi Malaysia (UTM). Currently, she is a Senior Lecturer at Razak Faculty of Technology and Informatics, UTM Kuala Lumpur. Her current research areas cover Future Communication Theory, Massive IoT Technologies, UAV Communication, Machine Learning Application, Multiple Access Techniques and Green Energy Communication.

Dr. Norliza Mohd Noor is a Professor at Razak Faculty of Technology and Informatics, UTM Kuala Lumpur Campus. Her research areas are image analysis and machine learning.