Al-Nahrain Journal for Engineering Sciences **NJES** 25(4)142-151,2022.
http://doi.org/10.29194/NJES.25040142

Data Mining for Autism Spectrum Disorder detection among Adults

Sumia Hamad Jaafer[1], Israa F. Abdulazez[2], Noor Kamal Al-Qazzaz[2], Teba Yaseen Yousif[2]

Authors affiliations:
1*) Erbil Technical Medical Institute, Erbil Polytechnic University, Erbil - Iraq.
sumaya.hamad@epu.edu.iq

2) Department of Biomedical Engineering, Al-Khwarizmi College of Engineering, Baghdad University, Baghdad - Iraq.
asso193@yahoo.com
noorbme@kecbu.uobaghdad.edu.iq
teba.yaseen246@gmail.com

Paper History:

Received: 11nd Oct. 2022

Revised: 1st Nov. 2022

Accepted: 26th Nov. 2022

Abstract

Autism Spectrum Disorder (ASD) is one of the most common children's neurodevelopmental disorders (NDD) with an estimated global incidence of 1% to 2%. There are two aims for this research, first, to propose a data mining architecture that combines behavioural and clinical characteristics with demographic data. Second, to provide a quick, acceptable and easy way to support the ASD diagnosis. this can be performed by conducting a comparison study to determine the efficacy of four possible classifiers: logistic regression (LR), sequential minimum optimization (SMO), naïve Bayes, and instance-based technique based on k-neighbors (IBK). These classifiers have been performed with Waikato Environment for Knowledge Analysis (WEKA) tools to distinguish autistic adults from healthy, normal subjects. The results showed that, with 99.71%, SMO classification accuracy was 99.71, which exceeded the accuracy of other classifiers. The proposed architecture allows for early detection of ASD, distinguishing between ASD and healthy control subjects. This study could help doctors and clinicians by giving them a better idea of what the future holds for people with autism spectrum disorder (ASD) and by improving therapy programs, allowing people with ASD to live a long and happy life.

Keywords: Autism Spectrum Disorder, Classification, Data Mining, SMOTE, WEKA.

تحليل البيانات للكشف عن اضطراب طيف التوحد بين البالغين

سمية حمد جعفر ، اسراء فائق عبد العزيز ، نور كمال القزاز ، طيبة ياسين يوسف

الخلاصة:

يعد اضطراب طيف التوحد أحد أكثر اضطرابات النمو العصبية شيوعًا مع تقدير انتشار عالمي بنسبة ١ إلى ٢ ٪ عند الأطفال. هدفت هذه الدراسة إلى شقين ، أولاً ، التركيز على اقتراح بنية استخراج البيانات بناءً على تكامل السمات السلوكية والسريرية مع البيانات الديموغرافية. ثانيًا ، لإجراء دراسة مقارنة للكشف عن اضطراب طيف التوحد عن طريق اختبار فعالية استخدام أربعة مصنفات مرشحة وهي

Logistic regression (LR), Sequential Minimal Optimization (SMO), Naïve Baye, Instance Based method based on k-neighbours (IBK)

تم تنفيذ هذه المصنفات باستخدام أدوات لتصنيف البالغين المصابين بالتوحد من الأشخاص الطبيعيين الأصحاء.

تظهر النتائج أن تفوقت دقة التصنيف على دقة المصنفات الأخرى بنسبة ٩٩,٧١٪. تسمح البنية المقترحة بالتعرف المبكر لاضطراب طيف التوحد مع نتائج واعدة جدًا لتحديد اضطراب طيف التوحد مقارنةً بالمواضيع الصحية الاخرى. ستساعد هذه الدراسة الأطباء والأطباء السريريين في التخطيط وتقديم تنبؤ موثوق باضطراب طيف التوحد بالإضافة إلى البرنامج العلاجي الأمثل حتى تتمكن مرضى اضطراب طيف التوحد من الحصول على سنوات من جودة الحياة العالي

1. Introduction

A diagnosis of autism spectrum disorder (ASD) is made in 63% of children, according to WHO statistics. ASD develops in children and keeps spreading to adolescents and adults. Most often, symptoms start to manifest within the first five years of life [1]. ASD is a severe neurodevelopmental disease with highly expensive medical care. It is defined by a persistent lack of social contact and stereotyped behavior, which is frequently coupled with a general decline in communication skills [2]. Genetic and neurological factors are linked to ASD. The ability to think and envision, repetitive acts, and interpersonal communication issues are all ways that ASD-related behaviors are displayed [3].

ASD is considered as a one of numerous behavioral diseases caused by abnormal brain connections, which can occur locally, regionally, or both [4]. Numerous studies carried out on the anatomical brain connections in autism have found that autistic brains grow at a faster rate than control brains [5-7]. The rate of research that focuses on understanding the neurobiology of this complicated neurodevelopmental condition has increased significantly [8]. With the rising prevalence of ASD, neuroimaging and postmortem investigations have shown evidence of structural and functional connection abnormalities in the brains of ASD patients [8].

Early diagnosis and treatment of ASD are possible [9]. Clinical diagnosis of ASD is greatly aided by early discovery of the disorder [10], therefore, the quality of life for children with ASD and their families can then be improved by using this diagnosis to guide individualized treatment plans [11]. Unfortunately, determining an ASD diagnosis can be time-consuming and expensive. The recent rise in ASD cases worldwide has prompted medical professionals and scientists to look for improved screening techniques. The development of modern technology has allowed us to store a vast amount of data. Making conclusions based on the gathered data is a crucial activity called data mining. Indeed, machine learning has made significant advancements recently and is becoming more important in applied sciences like biology and biomedicine [12-15]. Machine learning techniques are employed or advised to support the interpretation of data in clinical decision-making and diagnostics [16, 17].

Numerous number of studies on ASD have been performed [2, 18], although they have some drawbacks for instance, ASD data collection is the exclusive focus of some studies [19, 20], while others analyze brain imaging data using the rs-fMRI technique (resting-state functional magnetic resonance imaging) [11, 21] and others concentrate on a specific geographic area [22, 23]. Therefore, it is crucial to develop a rapid, simple, and effective way to aid in the early identification of ASD in which families of ASD patients may find it helpful to seek out professionals for care [9].

The classification of the UCI database's ASD datasets is the main goal of this effort. The clinical diagnosis of ASD in people of all ages is covered by the widespread and sizable ASD databases. Surveys on a mobile application called "ASD Tests" were used to gather data from several different nations. The iOS and Android operating systems both support this application that invented by Dr Fadi Fayez from the Nelson Marlborough Institute of Technology. But the data that has been gathered is not complete. Some data records may not have values since they are optional. As a result, the information gathered is not trustworthy enough to be used to directly make therapeutic judgments. The ASD dataset from the UCI database was utilized in this experiment (higher than 18 years old, AQ-10-Adult). The datasets include 20 features that are utilized for additional research, particularly for identifying ASD and enhancing the precision of ASD classification.

Our contributions are focused on two key areas: first, we present a data mining architecture that combines demographic data with characteristics from the clinical and behavioural sciences. Second, to offer a quick, palatable, and simple means of proving the ASD diagnosis. To do this, a comparative study comparing the performance of four potential classifiers can be carried out: logistic regression (LR), sequential minimum optimization (SMO), nave Bayes, and instance-based approach based on k-neighbors (IBK). The UCI database's datasets for ASD was pre-processing and classifying using Waikato Environment for Knowledge Analysis (WEKA) technologies. The datasets came from surveys that were conducted. The surveys ask about personal information and include some ASD screening questions. To speed up processing, the acquired data from the UCI datasets will be transformed into numerical data. To rectify the data, a synthetic oversampling approach (SMOTE) was used because some of the recorded data was missing. The classification techniques are then put into action. Doctors can diagnose ASD more precisely, promptly, and simply based on the gathered results following classification.

2. Related Work

Various data mining algorithms have been employed in the past in a number of biomedical investigations for classification, grouping, and association. Researchers have investigated how data mining algorithms are utilized in healthcare and biomedicine, offering recommendations on how to employ these algorithms and the potential applications of data mining in the healthcare sector [24]. Additionally, a work published in [25] shows the use of several data mining techniques for the identification of the most significant genes and gene sequences in a collection of gene expression microarrays. Qasem et al. discussed how to cope with learning models for forecasting patient health and the difficulty of utilizing predictive data mining in clinical medicine [25]. These models can be quite helpful in assisting doctors with activities related to diagnosis, treatment, or monitoring.

Machine learning was utilized by Raj et al. [26] to speed up the observation-based screening and diagnosis of autism. They used a variety of machine learning techniques, including feature-selection-based machine learning, to examine the entire set of marks for the assessment of ASD behaviour [27].

In an fMRI study of the theory of mind (ToM), 15 high-functioning adolescents and adults with autism and 15 typically developing control participants were compared to determine the causal effects of one brain region on another (effective connectivity, thought to be an explanatory model for autism) [28].

Given the significance of early ASD detection, numerous studies have examined the primary characteristics of the condition. Individual and family-based characteristics, as well as a number of geographic determinants, were examined by van Buitenen et al. [29]. A two-tiered screening procedure with improved quality evaluation, interagency policy collaboration, and coordination was adopted, according to Snijder et al. [30]. An outline of the main components of ASD diagnosis, including age, has been provided by researchers [23]. Finally, Bent et al. [31] also examined the potential delays in ASD identification. They compared children with ASD to those with an intellectual disability or developmental delay in terms of child age at first parental worry and age at first parental discussion of concerns with a health care practitioner. They investigated if delays in ASD diagnosis were related to how a clinician handled parental concerns.

3. Methods and Materials

The objectives of this work are to firstly propose a data mining architecture employing behavioral, clinical, and demographic data and, secondly, carry out a comparison analysis for the identification of ASD by evaluating the performance of four distinct classifiers (**Fig. 1**).

3.1 Dataset and Experimental Procedure

The UCI database's adult adults (17 and older) ASD datasets were used in this investigation [32]. Table 1 displays the dataset's characteristics. Ten attributes are employed in the datasets for our training method, and the type and description of each attribute are used to store the results of the ASD diagnosis (i.e., ground truth). All the characteristics of the datasets [33] are listed in Table 2, including the ground truth that is used to assess the specificity, sensitivity, and accuracy of the classifiers. The binary values assigned by "YES" or "NO" are contained in the ground truth. Ten attributes—from 1 to 10—are for screening questions and personal information.

Using behavioral, clinical, and demographic data, this study developed a data mining architecture. It also intends to conduct a comparison study for ASD detection by comparing the efficacy of four distinct classifiers.

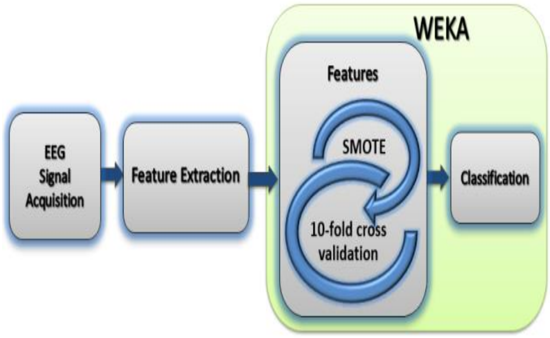

Figure 1: The block diagram of the proposed method.

Table 1: The demographic data and clinical features.

clinical features	Count
Number	704
Age	17-55
Gender	F(337), M(367)
Control	515
Autistic	189

The information came from adult ASD screening data that comprised 20 ASD-related features. Further study was carried out in an attempt to discover the fundamental autistic features and improve the categorization of autism spectrum disorder (ASD) patients depending on these traits. Furthermore, as shown in Table 2, those datasets comprise 10 behavioral traits (AQ-10) of persons that have been found to be beneficial in identifying ASD patients, as well as 10 attributes defining healthy people.

. **Table 2:** The characteristics and their explanations for the adult ASD screening dataset.

Attribute	Type	Description
Age	Number	Years of age
Gender	String	Female or Male
Ethnicity	String	White-European, Asian, Black, Latino, Middle Eastern, South Asian, Hispanic, Turkish, and Pasifika
Born with jaundice	Boolean (yes or no)	Whether or not the patient had jaundice when he or she was born.
PDD in a family member	Boolean (yes or no)	Whether or not anybody in your immediate family suffers from a PDD
Who is responsible for finishing the test?	String	Self, parent, clinician, medical staff, caregiver, etc.

Country of residence	String	United States, Spain, Brazil, Egypt, Bahamas, New Zealand, Burundi, Argentina, Austria, Jorden, United Arab Emirate, Ireland, Afghanistan, United Kingdom, Lebanon, Italy, South Africa, Pakistan, Chile, Bangladesh, France, Australia, China, Canada, Netherland, South Arabia, Romania, Tonga, Sweden, Oman, Philippines, India, Sierra Leone, Sri Lanka, Ethiopia, Viet Nam, Costa Rica, Iran, Mexico, Germany, Russian, Iceland, Nicaragua, Armenia, Hong Kong, Ukraine, Kazakhstan, Japan, Uruguay, Serbia, Portugal, Malaysia, Ecuador, American Samoa, Niger, Bolivia, Aruba, Belgium, Finland, Nepal, Indonesia, Turley, Angola, Czech Republic, Azerbaijan, and Cyprus.
The screening application has been used before	Boolean (yes or no)	If a screening application was used by the user
Screening Method Type	Integer (0,1,2,3)	Age-based screening techniques (0=toddler, 1=child, 2=adolescent, 3=adult)
Question 1 Answer	Binary (0, 1)	When others don't, I often detect little sounds
Question 2 Answer	Binary (0, 1)	Generally, I like to concentrate on the big picture instead of the details.
Question 3 Answer	Binary (0, 1)	It's simple for me to accomplish many things at once.
Question 4 Answer	Binary (0, 1)	I can quickly resume what I was doing if there is an interruption.
Question 5 Answer	Binary (0, 1)	When somebody is speaking to me, I can readily read between the lines.
Question 6 Answer	Binary (0, 1)	I can sense when someone is bored while listening to me.
Question 7 Answer	Binary (0, 1)	It's difficult for me to figure out what the characters' motivations are while I'm reading a novel.
Question 8 Answer	Binary (0, 1)	I value gaining information on a range of topics (e.g. types of car, types of bird, types of train, types of plant, etc)
Question 9 Answer	Binary (0, 1)	Just simply glancing at someone's face, I can tell what they're thinking or feeling.
Question 10 Answer	Binary (0, 1)	The screening procedure determines the response code for the question.
Screening score	Integer	It's tough for me to decipher people's motives.

3.2 Preprocessing of Features in WEKA

To carry out the experiments, data for ASD screening in adults was processed by using the WEKA package (version 3.8.4). WEKA is a machine learning program that allows for the combined use of a variety of tools in order to conduct extensive comparisons of different methodologies.

The data sets were analyzed using the popular open-source data mining application. The dataset was used to investigate the performance of a variety of classification procedures (classifiers). The test was performed on an HP Windows 10 Enterprise system with a 2.40 GHz Intel® CoreTM i7-4500U CPU. The datasets used are comparable in size, especially in terms of the number of attributes. The sufferer's gender, age, race, if they were born with jaundice, and whether they have a family member with PDD are all included in the adult autism data. Other details include who is taking the exam, their country of residence, if they have used the screening app before, and the type of screening approach they are employing.

ASD patients formed the minority class in this research. A synthetic oversampling approach (SMOTE) was utilized to correct the data [34]. Using a grid search approach and 10-fold cross-validation, the classifier parameters and quantity of oversampling were determined [12]. As a result, overfitting and bias in classification analyses were avoided. The provided dataset was divided into ten subgroups of equal size. One subset was used as a test set, while the other nine were pooled to create a training set for learning the classifier. This procedure was done 10 times, providing ten degrees of accuracy. The average of these accuracies revealed that learning from such a dataset was 10-fold cross-validation efficient [35].

3.3 Classification in WEKA

The data mining methodology is a method of identifying data trends. The patterns revealed must be useful in the sense that they result in a gain. The goal of data mining is to extract information from a data set and transform it into understandable data to help users making better decisions [36]. In data mining, pre-processing and categorization are two essential methodologies. Classification is an example of a supervised learning approach. It's a method for estimating group membership for a given data set [34]. Banking, hospitals, insurance, and health informatics are just a few of the businesses that use data mining. In the field of health informatics, data mining assists physicians in discovering effective pharmaceuticals and patients in obtaining better and more cost-efficient health care [37, 38].

The work is currently utilizing WEKA to classify the data for adults. The open-source machine learning application was created by the University of Waikato in New Zealand [39]. In WEKA, classification is the process of identifying a model or function that describes and distinguishes data classes with the purpose of using the model to predict the class of unknown objects in WEKA. The four candidates explored in this work are the LR, SMO, naïve Bayes, and IBK classifiers.

Based on cross-validation, machine learning divides datasets into two subsets. The first subset is referred to as the training data; it is a portion of the actual dataset that is fed to the machine learning model in order for it to discover and learn patterns. In this manner, our model is trained. The second subset is referred to as the testing data.

LR is an approach of data categorization that divides data into distinct categories. LR employs the logistic sigmoid function to convert its output to a probability value which may subsequently be transferred to two or more discrete classifications [40].

The SMO method is a decomposition technique in which a multi-variable optimization problem is divided into a series of sub-problems. Every sub-problem optimizes an optimal solution with a limited number of variables, usually just one, while the rest of the variables are treated as constants [40]. The SMO method is built by taking the decomposition strategy to its logical conclusion and optimizing a tiny subset of only two points at each iteration [40].

Naïve Bayes implements the probabilistic naïve Bayes classifier. In Naive Bayes Simple, the normal distribution is utilized to model numeric characteristics. Kernel density estimators, which can be used with naive Bayes, are responsible for determining whether the normality assumption is valid. Guided discretization may also be used to manage numeric characteristics. The incremental version of naïve Bayes Updateable handles one request at a time.

If the normality assumption is generally valid, kernel density estimators can be used with naïve Bayes to develop the classifier's performance. Numeric characteristics may also be managed through guided discretization. One request is processed at a time by the incremental version of Nave Bayes Updateable [40].

The IBk technique may choose the suitable value of K. It may also compute distance weighting by locating the training instance nearest to the given test instance and using a basic distance measure to forecast the same class as this training instance. When multiple instances are located within a reasonable distance of the test instance, the first one found is used [40].

3.4 Performance Measures

The mean classification performance and confusion matrix were used to evaluate the performance of the proposed system.

3.4.1 The degree of categorization accuracy on average

The average classification accuracy of the recommended approach was used to define the outcomes of the autism classification. Equations 1 and 2 calculate the average categorization accuracy as a percentage:

$$Average\ Classification\ Accuracy = (Number\ of\ correctly\ classified\ instances / Total\ number\ of\ instances) * 100 \quad \ldots\ldots.. (1)$$

3.4.2 Confusion matrix

The use of the confusion matrix is another method to demonstrate whether a classification is effective in terms of accuracy, sensitivity (recall), and specificity (precision). These confusion matrices demonstrate the performance of the classifier used to categorize the traits represented by the projected set of features (Table 3). These matrices show how often some traits are confused with others.

Table 3 Definition of the confusion matrix

	True Condition		
Predicted Condition	Total population	Condition Positive	Condition Negative
	Predicted Condition Positive	True Positive (TP)	False Positive (FP)
	Predicted Condition Negative	False Negative (FN)	True Negative (TN)

The ambiguity matrix's diagonal classification accuracy values are sufficient for two classes, but those outside the diagonal show between-class classification errors. To get the classifiers' precision, sensitivity (recall), and accuracy, utilize Equations 2-4:

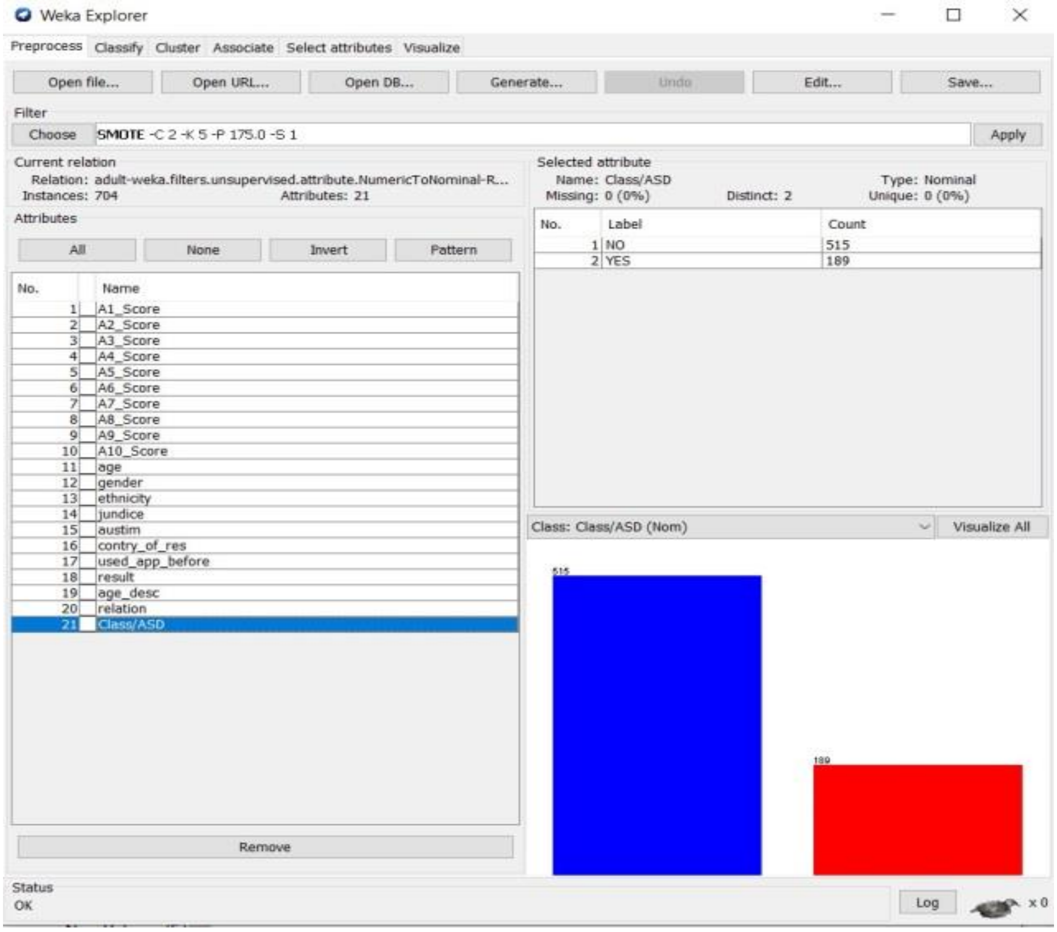

Figure 2: The control and the ASD data representation in WEKA

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \ldots(2)$$

$$Sensitivity\ (recall) = \frac{TP}{TP + FN} \quad \ldots(3)$$

$$Specificity(precision) = \frac{TN}{TN + FP} \quad \ldots(4)$$

4. Results and Discussion

This work presents the results of the proposed technique. The acts of healthy and autistic patients were shown and examined using the classification accuracy analysis and confusion matrix.

4.1 Dataset Acquisition

The dataset from 515 healthy control subjects and 189 ASD patients with the 20 different attributes under study are shown in Figure 2.

The characteristics under investigation include 10 behavioral traits (AQ-10) of persons and 10 attributes that characterize them; these attributes have been shown to be useful in identifying ASD patients, as shown in Table 2. Furthermore, these characteristics were used for additional study in defining influential autistic symptoms and enhancing ASD patient classification.

Analyzing ASD in Adults with WEKA

In this study, patients with ASD were the minority group. SMOTE, which is a synthetic oversampling approach in the minority class, was used to correct the data imbalance in the ASD patients with autism with a ratio of (175%), as shown in Figure 3. The classifier variables and oversampling quantities were built utilizing a grid search approach and 10-fold cross-validation to avoid overfitting and bias in classification studies.

The provided dataset was separated into ten disjoint subgroups of equal size. The remaining nine subsets were merged to form a training dataset for learning the classifier, with one serving as a test set. This process was repeated 10 times, resulting in a ten-degree accuracy. The 10-fold cross-validation accuracy of learning from this data is the average of these accuracies. The fraction of oversampling was supplied with the parameters since SMOTE impacts the dataset. As a result, the parameters estimated using various SMOTE percentages may differ. The SMOTE was used to equalize the frequency of the work using only the training set.

Classification in WEKA

This study includes a comparative analysis of three datasets in order to classify the ASD and healthy control patients.

1. **Logistic regression:** Builds linear logistic regression models with a classification accuracy of 98.16 %, sensitivity of 98 %, and specificity of 99 %, cases were correctly identified. Table 4 shows the confusion matrix using the LR classification algorithm for ASD patients and healthy controls.

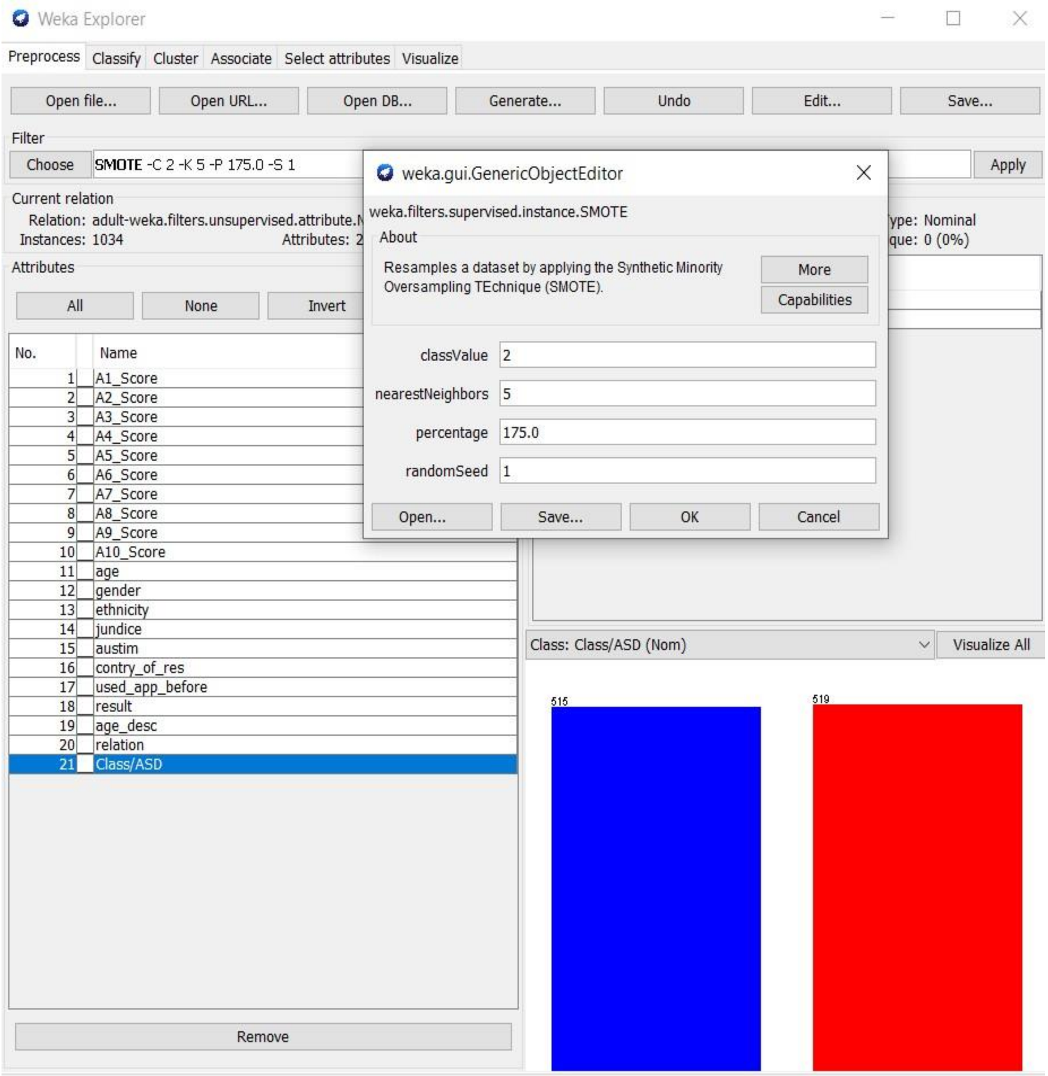

Figure 3. The control and the ASD data representation in WEKA after SMOTE

Table 4: Confusion matrix for ASD patients and healthy control subjects with LR classification technique

Predicted \ True	Control	ASD
Control	503	12
ASD	7	512

From the confusion matrix for ASD patients and controls using the LR classifier, it can be observed that the accurate identification on the main diagonal illustrated that 503 out of 515 were correctly classified as control subjects, whereas 512 out of 519 were correctly classified as ASD patients.

2. **SMO:** Support vector classification using a sequential minimum optimization approach gives a classification accuracy of 99.71 %, a sensitivity of 100 %, and a specificity of 99 %.

Table 5: Confusion matrix for ASD patients and healthy control subjects with SMO classification technique

Predicted \ True	Control	ASD
Control	515	0
ASD	3	516

From the confusion matrix for ASD patients and controls using the SMO classifier, it can be seen that the accurate identification on the major diagonal revealed that 515 were correctly categorized as control subjects while 516 out of 519 were correctly diagnosed as ASD patients.

3. **Naïve Bayes:** With a classifier of accuracy 98.74 %, sensitivity of 98 %, and specificity of 99 %, cases were correctly identified.

Table 6: Confusion matrix for ASD patients and healthy control subjects with Naïve Bayes classification technique

Predicted \ True	Control	ASD
Control	505	10
ASD	3	516

From the confusion matrix for ASD patients and controls using the Naïve Bayes classifier, it can be observed that the accurate identification on the main diagonal illustrated that 3 of the control subjects were misclassified as ASD patients, whereas 10 of the ASD patients were incorrectly classified as control subjects.

4. **IBK:** k-nearest-neighbors classifier with a classifier of accuracy 97.39 %, sensitivity of 96 %, and specificity of 99 %, cases were correctly identified.

Table 7: Confusion matrix for ASD patients and healthy control subjects with IBK classification technique

True / Predicted	Control	ASD
Control	493	22
ASD	5	514

The accuracy identification on the major diagonal of the confusion matrix for ASD patients and controls using the IBK classifier showed that 5 of the control participants were wrongly categorized as ASD patients, whereas 22 of the ASD patients were incorrectly classified as control subjects.

In this work, Figure 4 illustrates a comparative plot of the classification of ASD patients and healthy control subjects. Given that SMO's classification accuracy outperformed that of the other classifiers (LR, SMO, Nave Bayes, and IBK), it can be shown that SMO has the potential to considerably improve present categorisation.

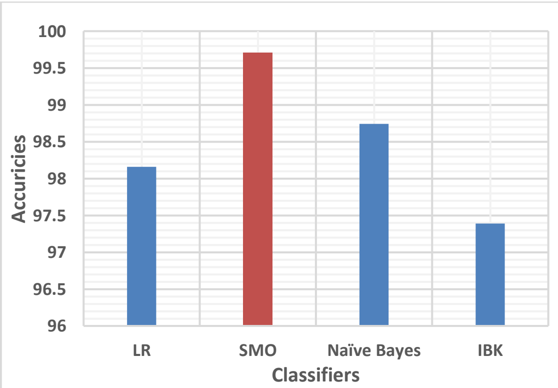

Figure 4. A comparative plot of the effective classifiers for ASD patients and control subjects

There are certain limitations to this study that should be mentioned. Because of the Covid-19 Virus, the workgroup was unable to collect data from the hospital. As a result, we revert to using an online dataset of adult autism screening. This study will help to improve the lives of ASD children in general by making an indirect contribution to improving the patients' lifestyles among the ASD population, which is important and beneficial to research and society. As a result, future work will focus on completing the research and attempting to improve its accuracy.

4. Conclusion

Data mining using WEKA was used in this experiment to differentiate ASD patients from healthy control volunteers. We achieved our goal of evaluating and investigating four selected WEKA-based classification algorithms while data mining the adult dataset. Based on the autistic data, the best system was the SMO classifier, which had a 99.71% accuracy. These findings suggest that the naive Bayes, LR, and IBK classifiers, among the machine learning algorithms tested, have the potential to considerably enhance existing classification methods for use in the medical business. As among machine learning algorithms studied, SMO has the potential to considerably improve traditional categorization methods for medical applications. Our future study will focus on the understanding of EEG signals, which will encompass larger datasets.

5. References

[1] J. Zeidan, E. Fombonne, J. Scorah, A. Ibrahim, M. S. Durkin, S. Saxena, et al., "Global prevalence of autism: a systematic review update," Autism Research, vol. 15, pp. 778-790, 2022.

[2] P. Hlavatá, T. Kašpárek, P. Linhartová, H. Ošlejšková, and M. Bareš, "Autism, impulsivity and inhibition a review of the literature," Basal Ganglia, vol. 14, pp. 44-53, 2018.

[3] F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening," International journal of medical informatics, vol. 117, pp. 112-124, 2018.

[4] W. Cao, H. Zhu, Y. Li, Y. Wang, W. Bai, U. Lao, et al., "The development of brain network in males with autism spectrum disorders from childhood to adolescence: Evidence from fNIRS study," Brain sciences, vol. 11, p. 120, 2021.

[5] J. Gao, M. Chen, Y. Li, Y. Gao, Y. Li, S. Cai, et al., "Multisite autism spectrum disorder classification using convolutional neural network classifier and individual morphological brain networks," Frontiers in Neuroscience, vol. 14, p. 629630, 2021.

[6] S. H. Jaafer, "Hurst Exponent and Tsallis Entropy Markers for Epileptic Detection from Children," Al-Khwarizmi Engineering Journal, vol. 17, pp. 34-42, 2021.

[7] N. K. Al-Qazzaz, I. F. Abdulazez, and S. A. Ridha, "Simulation recording of an ECG, PCG, and PPG for feature extractions," Al-Khwarizmi Engineering Journal, vol. 10, pp. 81-91, 2014.

[8] E. Shephard, F. S. McEwen, T. Earnest, N. Friedrich, I. Mörtl, H. Liang, et al., "Oscillatory neural network alterations in young people with tuberous sclerosis complex and associations with co-occurring symptoms of autism spectrum disorder and attention-deficit/hyperactivity disorder," Cortex, vol. 146, pp. 50-65, 2022.

[9] U. Erkan and D. N. Thanh, "Autism spectrum disorder detection with machine learning methods," Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews, vol. 15, pp. 297-308, 2019.

[10] M. Liao, H. Duan, and G. Wang, "Application of Machine Learning Techniques to Detect the Children with Autism Spectrum Disorder," Journal of Healthcare Engineering, vol. 2022, 2022.

[11] D.-Y. Song, C.-C. Topriceanu, D. C. Ilie-Ablachim, M. Kinali, and S. Bisdas, "Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-

analysis," Neuroradiology, vol. 63, pp. 2057-2072, 2021.
[12] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, and J. Escudero, "Classification enhancement for post-stroke dementia using fuzzy neighborhood preserving analysis with QR-decomposition," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 3174-3177.
[13] N. K. Al-Qazzaz, S. H. B. Ali, S. A. Ahmad, K. Chellappan, M. Islam, and J. Escudero, "Role of EEG as biomarker in the early detection and classification of dementia," The Scientific World Journal, vol. 2014, 2014.
[14] N. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. Ahmad, M. Islam, and J. Escudero, "Automatic artifact removal in EEG of normal and demented individuals using ICA–WT during working memory tasks," Sensors, vol. 17, p. 1326, 2017.
[15] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Discrimination of stroke-related mild cognitive impairment and vascular dementia using EEG signal analysis," Medical & Biological Engineering & Computing, pp. 1-21, 2017.
[16] N. K. Al-Qazzaz, S. Ali, M. S. Islam, S. A. Ahmad, and J. Escudero, "EEG markers for early detection and characterization of vascular dementia during working memory tasks," in 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2016, pp. 347-351.
[17] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Entropy-based markers of EEG background activity of stroke-related mild cognitive impairment and vascular dementia patients," in 2nd International Conference on Sensors Engineering and Electronics Instrumental Advances (SEIA 2016), Barcelona, Spain, 2016.
[18] J. McDermott, D. Study, J. Clayton-Smith, and T. Briggs, "The TBR1-related autistic-spectrum-disorder phenotype and its clinical spectrum," European journal of medical genetics, vol. 61, pp. 253-256, 2018.
[19] X.-y. Liu and S.-m. To, "Personal growth experience among parents of children with autism participating in intervention," Journal of autism and developmental disorders, vol. 51, pp. 1883-1893, 2021.
[20] S. R. Shahamiri, F. Thabtah, and N. Abdelhamid, "A new classification system for autism based on machine learning of artificial intelligence," Technology and Health Care, pp. 1-18, 2021.
[21] C. P. Santana, E. A. de Carvalho, I. D. Rodrigues, G. S. Bastos, A. D. de Souza, and L. L. de Brito, "rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis," Scientific reports, vol. 12, pp. 1-20, 2022.
[22] X. Hu, J. Wang, L. Wang, and K. Yu, "K-Nearest Neighbor Estimation of Functional Nonparametric Regression Model under NA Samples," Axioms, vol. 11, p. 102, 2022.
[23] X. Li, J. Hu, X. Liu, J. Yu, and C. C. Feng, "Adaptive digital elevation models construction method based on nonparametric regression," Transactions in GIS.
[24] U. M. Obeta, O. R. Ejinaka, and N. S. Etukudoh, "Data Mining in Medical Laboratory Service Improves Disease Surveillance and Quality Healthcare," in Prognostic Models in Healthcare: AI and Statistical Approaches, ed: Springer, 2022, pp. 459-481.
[25] S. N. Qasem and F. Saeed, "Hybrid Feature Selection and Ensemble Learning Methods for Gene Selection and Cancer Classification," International Journal of Advanced Computer Science and Applications, vol. 12, 2021.
[26] 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.
[27] A. H. B. Noruzman, N. A. Ghani, and N. S. A. Zulkifli, "A Comparative Study on Autism Among Children Using Machine Learning Classification," in International Conference on Emerging Technologies and Intelligent Systems, 2021, pp. 131-140.
[28] E. Grossi, G. Valbusa, and M. Buscema, "Detection of an autism EEG signature from only two EEG channels through features extraction and advanced machine learning analysis," Clinical EEG and Neuroscience, vol. 52, pp. 330-337, 2021.
[29] N. van Buitenen, J. Meijers, C. van den Berg, and J. Harte, "Risk factors of violent offending in mentally ill prisoners with autism spectrum disorders," Journal of psychiatric research, vol. 143, pp. 183-188, 2021.
[30] M. I. Snijder, S. P. Kaijadoe, M. van 't Hof, W. A. Ester, J. K. Buitelaar, and I. J. Oosterling, "Early detection of young children at risk of autism spectrum disorder at well-baby clinics in the Netherlands: Perspectives of preventive care physicians," Autism, vol. 25, pp. 2012-2024, 2021.
[31] C. A. Bent, J. Barbaro, and C. Dissanayake, "Parents' experiences of the service pathway to an autism diagnosis for their child: What predicts an early diagnosis in Australia?," Research in Developmental Disabilities, vol. 103, p. 103689, 2020.
[32] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," Journal of autism and developmental disorders, vol. 45, pp. 1121-1136, 2015.
[33] R. E. Rosenberg, R. Landa, J. K. Law, E. A. Stuart, and P. A. Law, "Factors affecting age at initial autism spectrum disorder diagnosis in a national survey," Autism research and treatment, vol. 2011, 2011.
[34] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Discrimination of stroke-related mild cognitive impairment and vascular dementia using EEG signal analysis,"

Medical & biological engineering & computing, vol. 56, pp. 137-157, 2018.

[35] N. K. Al-Qazzaz, S. H. M. Ali, and S. A. Ahmad, "Differential Evolution Based Channel Selection Algorithm on EEG Signal for Early Detection of Vascular Dementia among Stroke Survivors," in 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018, pp. 239-244.

[36] F. Gullo, "From patterns in data to knowledge discovery: what data mining can do," Physics Procedia, vol. 62, pp. 18-22, 2015.

[37] N. K. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Automatic artifact removal in EEG of normal and demented individuals using ICA–WT during working memory tasks," Sensors, vol. 17, p. 1326, 2017.

[38] N. K. Al-Qazzaz, M. K. Sabir, S. H. B. M. Ali, S. A. Ahmad, and K. Grammer, "Electroencephalogram profiles for emotion identification over the brain regions using spectral, entropy and temporal biomarkers," Sensors, vol. 20, p. 59, 2020.

[39] I. H. Witten, E. Frank, L. E. Trigg, M. A. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical machine learning tools and techniques with Java implementations," 1999.

[40] T. Ghosh, M. H. Al Banna, M. S. Rahman, M. S. Kaiser, M. Mahmud, A. S. Hosen, et al., "Artificial intelligence and internet of things in screening and management of autism spectrum disorder," Sustainable Cities and Society, vol. 74, p. 103189, 2021.