JOURNAL OF
EMERGING INVESTIGATORS

Article

The precision of machine learning models at classifying autism spectrum disorder in adults

Nandita Raj Kumar[1], Sasikala Raj Kumar[2]
[1]Silver Creek High School, San Jose, CA
[2]Special Education, Santa Clara Unified School District, Santa Clara, CA

SUMMARY

Autism spectrum disorder (ASD) is hard to correctly diagnose due to the very subjective nature of diagnosing it: behavior analysis. Due to this issue, we sought to find a machine learning-based method that diagnoses ASD without behavior analysis or helps reduce misdiagnosis. We tested the precision of many binary classification models such as the naive bayes, support vector machine, decision tree, random forest, and k-nearest neighbor models to compare their mean average precision in diagnosing individuals with ASD. Based on multivariable data, we hypothesized that the k-nearest neighbor model would be the best at diagnosing ASD accurately because this model is known to use data points that have related values to classify new data points. Upon training and testing of all the different models with an online dataset, the mean average precision of each model was analyzed along with its cross-validation scores, showing that the most accurate model at predicting whether an individual had ASD was the random forest model with a mean average precision of 0.92 and a mean cross-validation score of 0.86. The naive bayes model was the least accurate performer with a mean average precision of 0.80 and a mean cross-validation score of 0.64. Based on these results, the random forest model could aid in reducing the misdiagnosis of ASD. The usage of the random forest algorithm helps avoid bias in behavioral diagnosis by using objective data from screening tests rather than subjective data gathered by clinicians to classify ASD.

INTRODUCTION

The increasing prevalence of ASD presents a growing concern in modern society due to the various challenges that people with ASD face. According to the Centers for Disease Control and Prevention (CDC), nearly 1 in 36 children are diagnosed with ASD (1). ASD is a broad range of developmental conditions that can affect one's ability to communicate and behave as a cognitively normal individual (2). Apart from having potentially severe developmental problems, people with ASD face other mental health issues such as depression, anxiety, bipolar disorder, and increased likelihood of suicide attempts compared to individuals who do not have ASD (3). Moreover, people with ASD frequently have more physical health issues such as seizures, obesity, diabetes, and immune disorders than individuals without ASD do (3). The diagnosis of ASD is important but difficult as ASD shares numerous symptoms with other mental and physical illnesses, making it difficult to identify ASD (4).

Identifying ASD as early as possible is vital in ensuring support and resources for the individual's development into adulthood (5). According to a study by Lord et al., ASD is best diagnosed in children between the ages of two and nine (6). Having undiagnosed ASD as an adult has been shown to impact their quality of life in numerous ways including lowering their levels of independence and employment rates due to their abnormal communication skills (7). Detecting ASD in adults poses significant challenges due to the potential for symptoms to be less apparent by learned societal adaptations, therefore increasing the risk of misdiagnosis by healthcare professionals who may misinterpret ASD symptoms (8). A study has shown that individuals diagnosed with other psychiatric conditions, such as anxiety disorders or depression, may actually have undiagnosed ASD, leading to inappropriate treatment plans and delayed access to appropriate interventions (9). The present approach to diagnosing autism relies on DSM-5, which provides clinicians with a framework of criteria for diagnosing ASD. Given the high impact ASD has on people's lives, it begs the question of how diagnosing ASD for adults could be done with as much accuracy as possible while still having a higher accuracy in correct diagnoses by omitting the subjectivity of behavior analysis.

We thought that machine learning (ML) could be a solution to this issue of misdiagnosis of ASD in adults. Dating back to the 1970s when Stanford developed a program named MYCIN to detect bacterial infections, ML has been prevalent in solving numerous problems in the healthcare sector by aiding in the detection or classification of illnesses such as cancer through image-based analysis (10). ML models have been created to perform tasks such as the classification of illnesses through text and image inputs better than humans using training data (11). Through supervised learning, ML models can be trained with data and hyperparameters such as epochs and batch sizes to tweak the model and enhance the mean average precision of the model. Through this process, models become efficient in categorizing new data based on the model's patterns learned through the training data without the risk of misdiagnosis due to external factors such as the setting of diagnosis (12).

We aimed to find the best ML model that provides an accurate prediction of whether an individual has ASD based on screening data from questionnaires. We believed that the best classification model would be one that uses all of the features and models the feature with an equation to diagnose ASD. In the field of ML, there are numerous types of models such as naive bayes, k-nearest neighbor, decision tree, support vector machine, and a random forest model. Each of these models and algorithms is based on different concepts and theorems. The naive bayes model is based on the Bayes

https://doi.org/10.59720/23-176

theorem which assumes that all features are independent of each other (13). The k-nearest neighbor model predicts the label of a data point based on the closest data point in its training or sample data (14). The decision tree model creates a tree of conditional statements that leads to the final classification of a data point (15). The support vector machine model uses a hyperplane to split the data into different categories in a way that the most similar support vectors have a maximum margin with respect to the hyperplane (16). The random forest model uses multiple decision trees that represent each category to classify data (17).

We hypothesized that a model that classifies data based on closely related information such as the k-nearest neighbor model would have the best mean average precision when diagnosing ASD. After training and testing all models with the same data, we found that the random forest model was the most accurate in diagnosing ASD with a mean average precision of 0.92 and a mean cross-validation score of 0.86. The fairly high mean average precision result of this model shows the potential that ML has in the healthcare sector in diagnosing ASD and possibly even reducing the number of misdiagnoses.

RESULTS

To determine the classification algorithm with the highest precision for diagnosing autism, we trained and tested five different classification models using screening data from the AQ-10 test, an official 10 question diagnostic assessment, as well as their demographic information. We proposed that a classification model such as the k-nearest neighbor algorithm, which sorts data based on closely related information, would yield the highest mean average precision in diagnosing ASD. All five of the models were trained with screening data that was compiled from the results of a screening test for ASD. The utilized data originated from a publicly available online dataset hosted on Kaggle by REVA Academy for Corporate Excellence. This dataset comprised screening outcomes of individuals undergoing ASD testing alongside their final diagnoses. The data went through the same procedures for each model to maintain consistency (**Figure 1**). Some models such as the support vector machine and the k-nearest neighbor models, which rely on the distance of the data points or vectors, had to have their features scaled using normalization methods. The other models did not use normalization because they did not depend on the distance of data points to classify data.

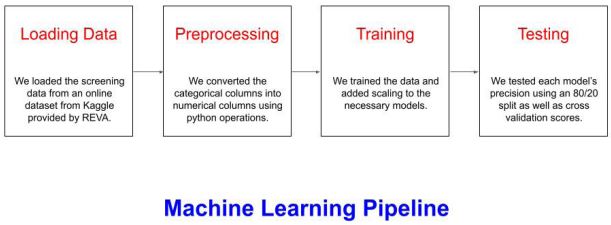

Figure 1. ML pipeline for ASD Classification. The process of this experiment starts with the loading of data through screening questions and answers in a comma-separated values (CSV) file into variables. The data is then processed to remove data with missing information and convert all data to numerical data. The various models were then trained. The models were then tested with an 80/20 test split and cross-validation to output the precision values.

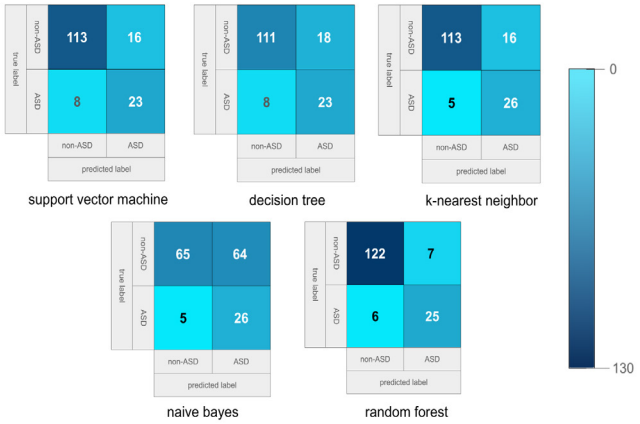

Figure 2. The confusion matrices for all five classification models. The number of true positive, false positive, true negative, and false negative labels are shown. The false positive (quadrant 1) is a type 1 error, and the false negative (quadrant 3) is a type 2 error. The random forest model has the highest true positive and lowest type 1 error while naive bayes model has the lowest true positive and highest type 1 error. The color scale shows higher values with darker shades of blue and lower values with lighter shades of blue.

We determined the effectiveness of the models using the mean average precision, average precision for ASD, and the average precision for non-ASD. The confusion matrix was also used to see the categories (ASD or non-ASD) with which the model had trouble. The confusion matrix shows the true positive, true negative, false positive, and false negative labels that the model has classified (**Figure 2**). For a good model, the predicted labels and the actual labels should have the highest value. Mean average precision, average precision for ASD, and average precision for non-ASD were then calculated from the confusion matrix values by using different ratios of the four matrix values. To assess overfitting, the cross-validation averages for 10 folds were included for each of the models. The cross-validation averages help us eliminate the chances of getting a lucky split in which the model overtrains. All three of the different types of precision scores came from just testing the model on the 80/20 split (80% training data and 20% testing data) whereas the cross-validation average came from a 10-fold cross-validation.

In our case, the random forest model had the greatest number of cases that match both predicted and actual labels while naive bayes had the greatest number of cases that did not match the predicted and actual labels (**Figure 2**).

Another important attribute of the models was their features and relative importance. The feature importance of the dataset was collected from the random forest model to examine its performance in matching both predicted and actual labels. The residence feature had the highest feature importance for the random forest model whereas the feature which determined whether the individual had used the app that held the screening questionnaire before had the least importance in the random forest model (**Figure 3**). The main AQ-10 scores had similar or only slightly better importance to most other features such as jaundice, gender, and relation (**Figure 3**). Of the 10 different AQ scores, the A3 score or the answer to the third question in the AQ-10 screening test, which was the individual's ability to socialize with others, had the most importance. The importance of these features plays

https://doi.org/10.59720/23-176

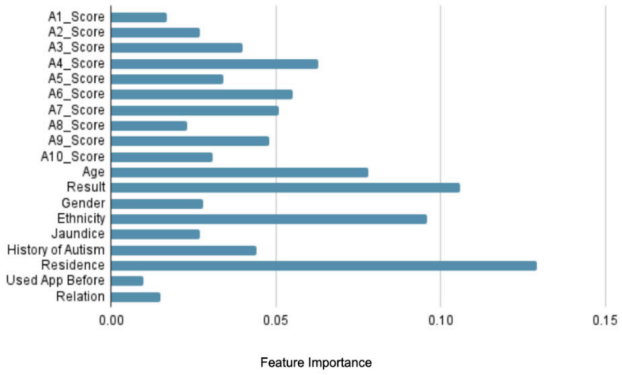

Figure 3. Feature importance relative to each other for random forest-based ASD classification. The relative importance of each feature is displayed. A higher number means a higher level of influence the feature has in the model to classify the data. This feature importance data was collected from the random forest model.

a big part in the performance of tree-based models such as the decision tree and the random forest model (18).

Our study highlighted variations in performance among the classification algorithms. The random forest model attained the highest mean average precision of 0.92 and a cross-validation score of 0.86 by employing an ensemble learning technique that constructs multiple decision trees during training (**Table 1**). This suggests that the ensemble learning technique employed by the Random Forest Model effectively captured complex patterns in the data, leading to more accurate predictions. In contrast, the naive bayes model exhibited the lowest precision, with an average precision of 0.80 and a cross-validation score of 0.64 by relying on the assumption of feature independence given the class label (**Table 1**). This indicates that the Naive Bayes Model struggled to capture the dependencies between features, which are crucial for accurately classifying ASD cases. Meanwhile, the decision tree, support vector machine, and k-nearest neighbor models demonstrated moderate performance, with differing levels of precision and cross-validation scores (**Table 1**). Apart from overall precision for classifying both ASD and non-ASD, the models had different precision when comparing their performance on how well they solely classified ASD or solely classified non-ASD in a collection of both ASD and non-ASD. Furthermore, when examining precision for identifying ASD and non-ASD cases separately, distinctive trends emerged. The k-nearest neighbor model performed well in identifying non-ASD cases, achieving an average precision of 0.96 (**Table 1**). Conversely, for ASD cases, the random forest model demonstrated better with an average precision of 0.78, while the naive bayes model had an average precision of 0.29 (Table 1). The higher the number for the average precision for ASD the better the model is at classifying people as ASD in a collection of both ASD and non-ASD. The higher mean average precision values show the combined precision for both classes (ASD and non-ASD). Whereas the average precision for a specific class shows the ratio in which the model diagnosed that specific class correctly compared to all the data in that class.

	K-nearest neighbor	Naive bayes	Support vector machine	Decision tree	Random forest
Mean Average Precision	0.89	0.80	0.87	0.86	**0.92**
Cross-Validation Score (Average of 10 Folds)	0.83	0.64	0.85	0.84	**0.86**
Precision in classifying individuals as ASD (average precision of ASD)	0.62	0.29	0.59	0.56	**0.78**
Precision in classifying individuals as non-ASD (average precision of non-ASD)	**0.96**	0.93	0.93	0.93	0.95

Table 1. Comparison of each model's testing scores. The mean average precision and the cross-validation scores are calculated and used as a way to test the progress and validity of a model. The precision of calculating each data point into a category is shown for all the models. The highlighted values represent the scores of the best-performing model in each category.

DISCUSSION

This study aimed to assess the effectiveness of different classification algorithms in classifying ASD. Five algorithms, including the k-nearest neighbor, naive bayes, decision tree, support vector machine, and random forest models, were evaluated using average precision and mean cross-validation score. We hypothesized that models emphasizing the relationships between features would outperform others such as the k-nearest neighbor model have a higher precision than other models. Contrary to the initial hypothesis favoring k-nearest-neighbor, random forest model emerged as the top performer, with an average precision of 0.92 and a mean cross-validation score of 0.86. Whereas the naive bayes model demonstrated lower precision with an average precision of 0.80 and a mean cross-validation score of 0.64.

All the models used were classification models that used features and their values to try to predict the label or class of the data point. All five of our models have a weighted mean average precision of 0.80 or higher showing that our models can diagnose ASD. Nevertheless, some of the models could be improved. The mean average precision of some of the models had specific reasons as to why they were so low or high. For instance, the naive bayes model is a purely theoretical model that does not have much effect in practice. The naive bayes model is known for assuming that all the features are independent of each other; in other words, the prediction made by the model does not make any relationship connections between the features to predict the class. This makes it a poor model and theory for our study because the only way to predict a class would be to find the relationship between the features to make the best classification. Thus, this explains why the mean average precision for naive bayes is very low for classifying individuals as having autism. Our decision tree model was the second worst model but still had a high mean average precision of 0.86. The decision tree

https://doi.org/10.59720/23-176

	K-nearest neighbor	Naive bayes	Support vector machine	Decision tree	Random forest
Hyperpara-meters	n_neighbors=10	used GuassianNB() w/o hyperparameters	kernel='linear' gamma = 'auto' C= 2	criterion='entropy' max_depth=2 random_state=1	n_estimators=100, criterion='entropy', class_weight=None

Table 2. Hyperparameters of the 5 classification models. This table displays the different hyperparameters for the classifiers. All of the models had hyperparameters except the Naive Bayes model, though it is important to note that the specific classifier was a GaussianNB() classifier which is a type of Naive Bayes classifier.

model uses nodes that split the dataset into subsets based on the outcome of the decision. The decision is usually a key identifier of a feature that distinguishes the data point from other data points. Although decision trees are good at building a tree and path for classifying data points for known data, decision trees have a difficult time dealing with new data as shown by the results of our model. This is a result of overfitting or using the training data too much to an extent where it can only work with the training data and not new datasets such as the testing data. The next best model was the support vector machine. The support vector machine was only slightly better than the decision tree when using the mean average precision to compare the models. Unlike the previous two models, this model uses the distance between vectors (data points) and the hyperplane (the plane that separates the two classes of data) to classify the new data points. Thus, we needed to normalize the data using a scaler before training the model. This model had no obvious disadvantage due to the type of model itself but may have had slight result differences due to hyperparameters such as the kernel type and gamma type. Although we tuned the hyperparameters using the grid search method that is commonly used to tune hyperparameters (19), it is possible that other hyperparameters would have been a better fit for the type of data we had (**Table 2**).

The second-best model in terms of precision was the k-nearest neighbor model which used the distance between the closest "neighbors" to make a prediction thus the data was normalized before the training of the model. Like our reasoning behind the hypothesis, the model's success might be because people who have ASD share traits that are given by the AQ score which will allow the model to predict the right diagnosis because they share a similar or close data point with each other. Lastly, the best model in terms of precision was the random forest model with a 0.92 mean average precision. The random forest model uses an ensemble of decision trees which helps eliminate one of the disadvantages of decision trees which is overfitting by taking the mean of the decision tree results. This is supported by its cross-validation score of 0.86 even though this value is significantly lower than its mean average precision. This inadvertently might have been the biggest factor in making it the best model for diagnosing ASD.

Other than the disadvantages and advantages of the types of models themselves, another factor that could have affected all our models' mean average precision might have been the data. The data we used underrepresented numerous ethnicities and places of residence, which may have skewed the results of the feature importance causing the model to interpret the data incorrectly. In our dataset, the US and the United Kingdom were overrepresented places of residence, and the White-European and Asian ethnicities were overrepresented ethnicities. The underrepresented countries might make the model assume that individuals from these countries are less likely to be diagnosed with ASD, but this is not true. In ML, this skewness of certain features is considered a bias (19). This skewness of data surely exists in the other features too. Unless we receive more balanced data, our model may have a bias in its predictions.

It is also notable that the results of the model also depend on the average precisions of ASD and non-ASD predictions rather than just the mean average precision. This is because the mean average precision might disproportionately show the performance of a model when the model is good at classifying individuals as ASD but not as non-ASD or vice versa. In our study, all of our models had average precision for non-ASD that was greater than 0.90 most likely due to the fact that data had more individuals with non-ASD than ASD (**Table 1**). It is possible that the average precision for ASD could be improved if more data for individuals who are classified as ASD were provided.

Another issue we need to address is the subjectivity of our data. This model was aimed to reduce the subjectivity of humans when diagnosing ASD so it makes sense that our model should be completely objective. Unfortunately, this is not the case as the data collected was self-reported information. Although some information such as gender, place of residence, and whether the individual had jaundice is objective, the main results from the AQ-10 test are subjective as the individuals themselves have to answer these questions. This might lead to the data itself being subjective and thus will result in the prediction of the model being as subjective as the data it is received. The current method of diagnosing autism is based on DSM-5, a set of criteria that helps clinicians diagnose ASD (20). Those criteria list that in order for individuals to be diagnosed with ASD, they must have persistent communication issues and social impairments (20). In a similar approach, the AQ-10 questions aim to figure out whether or not individuals meet the DSM-5 criteria. Therefore, this screening test could be as subjective as normal diagnosing methods depending on how accurately the screening test is filled out.

Additionally, the data we used only had screening information for 704 individuals. This may seem like a large sampling of data, but ML models need substantial amounts of data to accurately classify new data (21). Improvements to our models could increase our mean average precision further in diagnosing ASD.

Furthermore, the results from the data were initially collected using a test split of 80/20. Although research does not indicate what type of split is best, some research has suggested that a good split is close to a 70/30 split (22). This method of testing our model may have resulted in overfitting our data. To counteract this, we also used cross-validation folds which allows us to account for the overfitting of models. The cross-validation scores provide a better picture of the model's accuracy. Compared to the model's mean average precision, some models such as the naive bayes and the random forest model had significantly lower mean cross-validation scores (**Table 1**). This is because these models had a split which ended up overfitting the model. Despite this, many of our models had mean cross-validation scores very

https://doi.org/10.59720/23-176

close to the 80/20 split mean average precision.

Our study has shown a promising start in integrating ML into the process of diagnosing ASD, especially considering the challenges adults face in obtaining a diagnosis. By addressing these difficulties and enhancing diagnostic accuracy through ML, we anticipate significant improvements in the quality of life for individuals navigating ASD as adults. Our models could be further improved if more research is done to better fit the data or larger sample sizes are used to train the models. Further research should also delve into the incorporation of other features such as brain scans or more medical information about the patient such as the age when they first started speaking. With those changes, our models could have a better mean cross-validation score than 0.86. ML models might not eliminate the need for a behavior analysis altogether, but they could certainly help professionals have a second input and therefore help reduce the number of misdiagnoses each year.

MATERIALS AND METHODS

Data

The various models were trained on data that would help the machine learning models diagnose ASD such as screening data. The ASD screening data was collected from an official test recommended by the United Kingdom National Institution for Health Care and Excellence called the Autism Quotient-10 which poses ten questions that help determine whether an individual has ASD (23). The dataset collected includes the test results of 704 individuals who are all adults (**Figure 4**).

Screening Data Information

Male v.s female distribution	
Male	Female
367	337

Range of ages	
Youngest	Oldest
18	64

Jaundice distribution	
Yes	No
69	635

AQ-10 answer distribution

A1		A2		A3		A4		A5		A6		A7		A8		A9		A10	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
508	196	319	385	322	382	349	355	351	353	200	504	294	410	247	457	228	476	404	300

AQ-10 score distribution										
0	1	2	3	4	5	6	7	8	9	10
14	33	74	110	131	83	70	57	55	47	30

History of autism distribution	
Yes	No
91	613

Top 5 ethnicity distribution	
White-European	233
Asian	123
Middle-Eastern	92
Black	43
South Asian	36

Top 5 residence distribution	
USA	113
United Arab Emirates	82
India	81
New Zealand	81
United Kingdom	77

"Used app before?" response distribution	
Yes	No
12	692

ASD classification distribution	
ASD	non-ASD
189	515

Top 5 relation to individual distribution	
Self	522
Parent	50
Relative	28
Others	5
Health care professional	54

Figure 4. Frequency distribution for all features in screening data. All of the features' distribution is displayed. The total number of individuals is 704 and distributions that do not add up to 704 signifies missing data. For the places of residence, ethnicity, and relation to individual features, the top 5 category distributions were shown.

Of the 704 individuals, 515 individuals were classified as not having ASD and 189 individuals were classified as having ASD. The screening data included various features that were prominent such as the AQ scores and place of residence as well as less prominent features like whether the person has used this app for screening before. The data used was from a publicly accessible online dataset from Kaggle provided by REVA Academy for Corporate Excellence that had compiled the screening results of people testing ASD and the actual diagnosis they received (24). The dataset was downloaded and then loaded onto a data frame with the help of the Pandas library.

Preprocessing

The data was then preprocessed using the get_dummies function which turns all the categorical columns into numerical columns. To maximize results, models need numerical data to measure, calculate, and compare data. So, the get_dummies function turns the categorical data into columns that represent each category and uses 0 and 1 to label the data points which indicate the category a data point is to be in (25). For models that classify data based on the distance between the vectors or data points, the MinMaxScaler from the Scikit package was applied for the normalization of data. Normalization of data occurs by scaling all the numerical data to fit between -1 and 1. This is to ensure that numbers like age and test score which have different units are not given more importance due to their larger number.

Training and Testing

The Scikit package was used to split the data into 80% training data and 20% testing data. Each model was then trained using Scikit's model packages. Each model had some different parameters to tune the model. For example, the k-nearest-neighbor model has a parameter for the number of neighbors and the random forest model has a parameter specifying the number of trees. All these parameters were tuned for maximum mean average precision. Finally, the model was tested and assessed based on the confusion matrix result and the classification report used from the metrics Scikit package. To access the results of the model with a better method, we used cross-validation scores instead of just splitting the testing and training data. The results were

	Mean average precision (mAP)	Average Precision of ASD (AP_{ASD})	Average Precision of non-ASD ($AP_{non\text{-}ASD}$)
Formula	$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k$ where… n is the number of class k is the class (ASD or non-ASD)	$AP_{ASD} = \frac{TP}{TP+FP}$	$AP_{non\text{-}ASD} = \frac{TP}{TP+FP}$
What it means in this experiment	The mean of the average precision for the two classes so (AP_{ASD} + $AP_{non\text{-}ASD}$)/2	Data points the model classified as ASD correctly divided by all the data points in the dataset that are ASD.	Data points the model classified as non-ASD correctly divided by all the data points in the dataset that are non-ASD.

Table 3. Explanation of the terms used to explain model results. The mean average precision, average precision for ASD, and average precision for non-ASD are all calculated using the confusion matrix results. Each score represents and measures a different aspect of the model.

https://doi.org/10.59720/23-176

then analyzed using average precision values. Precision is the number of true positives over the total positive classifications (**Table 3**). In other words, the mean average precision is the number of correct predictions over the total number of cases. Whereas the average precision for a specific class (ASD or non-ASD) is the number of correct predictions for that specific class over the total number of cases that were in a specific class (**Table 3**).

Code Repository: https://github.com/NanditaRK/JEI-Autism-Classification-Models
Dataset: https://www.kaggle.com/c/autismdiagnosis/data

Received: June 22, 2023
Accepted: December 13, 2023
Published: June 28, 2024

REFERENCES

1. CDC. "ADDM Network Community Report on Autism." *Centers for Disease Control and Prevention*, 23 Mar. 2023, www.cdc.gov/ncbddd/autism/addm-community-report/index.html, Accessed 21 November 2023.
2. Maenner, Matthew J. "Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020." *MMWR. Surveillance Summaries*, vol. 72, 2023. www.cdc.gov, doi.org/10.15585/mmwr.ss7202a1.
3. Sala, Regina, et al. "Bridging the Gap Between Physical Health and Autism Spectrum Disorder." *Neuropsychiatric Disease and Treatment,* vol. 16, June 2020, pp. 1605–18. Dovepress, https://doi.org/10.2147/NDT.S251394.
4. 'Autism Spectrum Disorder' National Institute of Mental Health (NIMH), www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd, Accessed 14 July 2023.
5. Elder, Jennifer Harrison, Kreider, Consuelo Maun, et al. 'Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships'. *Psychology Research and Behavior Management*, vol. 10, 24th Aug. 2017, pp. 283–292, https://doi.org/10.2147/PRBM.S117499.
6. Lord, Catherine, Risi, Susan, et al. 'Autism From 2 to 9 Years of Age'. *Archives of General Psychiatry*, vol. 63, no. 6, 1st Jun. 2006, pp. 694–701, https://doi.org/10.1001/archpsyc.63.6.694.
7. Smith DaWalt, Leann, Hickey, Emily, et al. 'Impact of Working Together for adults with autism spectrum disorder: a multifamily group intervention'. *Journal of Neurodevelopmental Disorders*, vol. 13, no. 1, 8th Oct. 2021, pp. 44, https://doi.org/10.1186/s11689-021-09395-w.
8. 'Autism signs and characteristics: checklist for adults' The Spectrum, thespectrum.org.au/autism-diagnosis/checklist-adults/, Accessed 14 July 2023.
9. Fusar-Poli, Laura, et al. "Missed Diagnoses and Misdiagnoses of Adults with Autism Spectrum Disorder." *European Archives of Psychiatry and Clinical Neuroscience*, vol. 272, no. 2, Mar. 2022, pp. 187–98. Springer Link, https://doi.org/10.1007/s00406-020-01189-w.
10. Bush, Jonathan. "How AI Is Taking the Scut Work Out of Health Care." *Harvard Business Review*, 5 Mar. 2018. hbr.org, hbr.org/2018/03/how-ai-is-taking-the-scut-work-out-of-health-care.
11. Davenport, Thomas and Kalakota, Ravi 'The potential for artificial intelligence in healthcare'. *Future Healthc J*, vol. 6, no. 2, 1st Jun. 2019, pp. 94–98, https://doi.org/10.7861/futurehosp.6-2-94.
12. Hurwitz, Judith. *Machine Learning For Dummies®, IBM Limited Edition*. 2018.
13. "1.9. Naive Bayes." *Scikit-Learn*, scikit-learn/stable/modules/naive_bayes.html. Accessed 18 July 2023.
14. "1.6. Nearest Neighbors." *Scikit-Learn*, scikit-learn/stable/modules/neighbors.html. Accessed 18 July 2023.
15. "1.10. Decision Trees." *Scikit-Learn*, scikit-learn/stable/modules/tree.html. Accessed 18 July 2023.
16. "1.4. Support Vector Machines." *Scikit-Learn*, scikit-learn/stable/modules/svm.html. Accessed 18 July 2023.
17. "1.11. Ensemble Methods." *Scikit-Learn*, scikit-learn/stable/modules/ensemble.html. Accessed 18 July 2023.
18. Ronaghan, Stacey. "The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-Learn and Spark." *Medium*, 1 Nov. 2019, towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3.
19. Mehrabi, Ninareh, et al. *A Survey on Bias and Fairness in Machine Learning*. arXiv:1908.09635, arXiv, 25 Jan. 2022. arXiv.org, arxiv.org/abs/1908.09635.
20. CDC. "Diagnostic Criteria | Autism Spectrum Disorder (ASD) | NCBDDD | CDC." *Centers for Disease Control and Prevention*, 2 Nov. 2022, www.cdc.gov/ncbddd/autism/hcp-dsm.html.
21. *Sample Size in Machine Learning and Artificial Intelligence – Perioperative Data Science*. sites.uab.edu/periop-datascience/2021/06/28/sample-size-in-machine-learning-and-artificial-intelligence/. Accessed 21 July 2023.
22. Nguyen, Quang Hung, et al. "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil." *Mathematical Problems in Engineering*, edited by Yu-Sheng Shen, vol. 2021, Feb. 2021, pp. 1–15. DOI.org (Crossref), https://doi.org/10.1155/2021/4832864.
23. *Overview | Autism Spectrum Disorder in Adults: Diagnosis and Management | Guidance | NICE*. NICE, 27 June 2012, www.nice.org.uk/guidance/CG142. Accessed 21 November 2023
24. REVA University. *Autism Screening Challenge Data*. www.kaggle.com/c/autismdiagnosis/data. Accessed 12 Dec. 2023.
25. Pandas.Get_dummies — Pandas 2.1.3 Documentation. pandas.pydata.org/docs/reference/api/pandas.get_dummies.html#pandas.get_dummies. Accessed 22 Nov. 2023.