

Autism detection based on autism spectrum quotient using weighted average ensemble method

Lawysen, Nelsen Anggara, Abba Suganda Girsang

Department of Computer Science, BINUS Graduate Program - Master of Computer Science, BINUS University, Jakarta, Indonesia

Article Info

Article history:

Received Feb 21, 2024

Revised May 2, 2024

Accepted May 12, 2024

Keywords:

Autism spectrum disorder

Autism-spectrum quotient

Classification

Machine learning

Weighted average ensemble

ABSTRACT

Autism spectrum disorder (ASD) is a condition that occurs in an individual, wherein it is accompanied by various symptoms such as difficulties in socializing with others. Early detection of ASD patients can assist in preventing various symptoms caused by ASD. The focus of this research is to automate the diagnosis of ASD in an individual based on the results of the autism spectrum quotient (AQ) using weighted average ensemble method. Initially, preprocessing is carried out on the dataset to ensure optimal performance of the resulting model. In the preprocessing step, the filling of missing values and feature selection occurs, where the feature selection method being utilized is p-value. The model in this research uses the weighted average ensemble method, which is the model that combines three machine learning classification algorithms. Eight classification algorithms are tested to identify the three algorithms with the best performance, namely gaussian Naïve Bayes (NB), logistic regression (LR), and random forest (RF). Following the testing, the model constructed using the weighted average ensemble method exhibits the highest performance compared to the model built using a single classification algorithm. The performance matrix used to measure the model's performance is area under the curve (AUC)/receiver operating characteristic (ROC), with the developed model achieving an AUC/ROC value of 0.912.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Lawysen

Department of Computer Science, BINUS Graduate Program - Master of Computer Science

BINUS University

Jakarta, Indonesia

Email: lawysen@binus.ac.id

1. INTRODUCTION

Autism spectrum disorder (ASD) is a developmental disorder that occurs in an individual, characterized by difficulties in social interaction and communication [1], [2]. ASD experienced by individuals varies in severity and manifestations. It is a condition present from birth [3]. Diagnosing autism in children is generally easier than in adults because it requires a multitude of factors such as medical records, developmental history, and behavioral information from close individuals like parents, friends, colleagues, or teachers [4].

Hence, to aid in detecting and measuring the level of ASD in adults, the autism spectrum quotient (AQ) instrument can be employed [5]. This tool can differentiate individuals with ASD from those without within a population [3]. AQ is a questionnaire that assesses various symptoms considered characteristic of ASD [5]. Respondents answer these questions with agree or disagree responses [6], [7]. The AQ was developed by Simon BaronCohen and colleagues at the Autism Research Center, Cambridge, UK. In this study, the shorter version of AQ, the AQ10, consisting of only 10 questions, is being utilized [7].

The aim of this study is to determine whether an individual is indicative of having ASD or not based on their AQ questionnaire responses using machine learning algorithms. In addition to evaluating the AQ questionnaire responses, an assessment of the correlation between profile, background and expected results will be carried out. This study holds significance due to the increasing prevalence of ASD over time [8]. Early diagnosis of ASD in individuals is crucial as it allows for prompt management of associated developmental, communication, and social interaction challenges [9].

2. LITERATURE REVIEW

ASD is a mental disorder that causes individuals to experience various impediments in communication, social interaction, and physical abilities, leading those with ASD to struggle in interacting with others, such as parents, friends, family, and teachers [10]. Therefore, ASD must be detected as early as possible so that various measures can be promptly taken to enhance the abilities of ASD patients, enabling them to overcome these emerging obstacles. There are various methods to diagnose ASD in individuals, such as conducting screenings to investigate patient behavior using various screening tools administered by a psychologist [10], observing health and developmental history, interviewing individuals close to the patient like parents, friends, and teachers [4], and employing instruments such as the AQ, autism diagnostic interview-revised (ADI-R), and autism diagnostic observation schedule (ADOS) [3], [5]-[7], [11].

AQ is a self-assessment tool in the form of a questionnaire recommended as an efficient and economical method to estimate ASD in individuals, where the results from AQ can guide the referral process [7]. AQ has two versions: AQ50, consisting of 50 questions divided into 5 sections social skills, attention shifting, attention to detail, communication, and imagination each containing 10 questions [7]. There is a shorter version, AQ10, comprising a total of 10 questions derived from the 50 questions in AQ50. The 10 questions selected from AQ50 are not solely from one section like communication or imagination but are a combination from all five sections in AQ50. The sequential questions in AQ10 are numbers 5, 28, 32, 37, 27, 31, 20, 41, 36, and 45 [7]. Detailed questions on AQ10 can be seen in Table 1.

Table 1. Questions of AQ10

| No | Questions of AQ10 for adult |
|----|--|
| 1 | I often notice small sounds when others do not |
| 2 | I usually concentrate more on the whole picture, rather than the small details |
| 3 | I find it easy to do more than one thing at once |
| 4 | If there is an interruption, I can switch back to what I was doing very quickly |
| 5 | I find it easy to 'read between the lines' when someone is talking to me |
| 6 | I know how to tell if someone listening to me is getting bored |
| 7 | When I'm reading a story, I find it difficult to work out the characters' intentions |
| 8 | I like to collect information about categories of things (e.g., types of cars, types of bird, types of train, and types of plant) |
| 9 | I find it easy to work out what someone is thinking or feeling just by looking at their face |
| 10 | I find it difficult to work out people's intentions |

The aim of AQ10 is to identify individuals who might be suffering from undiagnosed ASD and refer them for further assessment. AQ10 can also be used to track changes in autistic symptoms over time [11]. AQ10 was developed by Allison, Auyeung, and Baron-Cohen in 2012, based on a large sample of 1,000 cases and 3,000 controls. They selected 10 items with the highest discriminatory power between ASD and non-ASD groups from AQ50. They also validated AQ10 against gold standard diagnostic instruments for ASD, such as the ADI-R and ADOS [11].

Calculating AQ10 results is similar to AQ50, summing the points for each answer, where "strongly agree" and "somewhat agree" score 1 point, and "strongly disagree" and "somewhat disagree" score 0 points. In the case of AQ10, if the total score ranges from 0 to 10, higher scores indicate a higher likelihood of the participant being part of the ASD group. A score of 6 or more signifies whether someone has ASD or not [11]. Research has been conducted to compare the performance of AQ50 and AQ10, involving 476 adult participants. The results showed that the performance of both AQ50 and AQ10 is highly similar, indicating that both can be equally utilized to aid in diagnosing ASD in individuals [7].

2.1. Related works

Research utilizing machine learning algorithms to process the results of AQ for detecting ASD has been widely conducted. Most studies have leveraged AQ10 outcomes. Some research not only examines AQ results in adults but also discusses children and adolescent groups. Here is a discussion of various previous studies related to detecting ASD.

Vakadkar *et al.* [12] proposes the utilization of various machine learning algorithms to enhance the accuracy and speed of diagnosing ASD in individuals. The dataset used in this paper is the autism screening data for toddlers obtained from the Kaggle website, totaling 1,054 data points and 18 attributes. This paper utilizes multiple machine learning classification algorithms such as support vector machines (SVM), random forest (RF) classifier, Naïve Bayes (NB), logistic regression (LR), and K-nearest neighbor (KNN), comparing the performance of all these classification algorithms. Before modeling, the authors of this paper preprocess the dataset, including noise removal, encoding, normalization, handling missing values, and outliers, ensuring that the resulting model performs well. The performance matrix used to measure the models in this paper includes accuracy and F1-score, where the LR algorithm exhibits the highest accuracy and F1-score at 97.15% and 0.98, respectively.

Hossain *et al.* [13] proposes a solution to automate the process of diagnosing ASD in patients using feature selection algorithms and machine learning classification. The dataset utilized in this paper is the ASD dataset of toddlers, children, adolescents, and adult's version 2, comprising a total of 23 attributes. This paper employs the relief of algorithm to classify significant attributes and 27 classification algorithms to find the best-performing one. Evaluation matrices used in this paper are accuracy and F1-score, where the multi layer perceptron (MLP) algorithm outperforms the other 26 classification algorithms with an average accuracy of 99.7% and an F1-score of 0.995.

Oma *et al.* [14] suggests creating a model to detect ASD across all age groups by combining RF-classification and regression trees and RF-iterative dichotomiser 3 models. The dataset used comprises AQ-10 datasets divided into three parts: child dataset with 292 data points, adolescent dataset with 104 data points, and adult dataset with 704 data points. Performance metrics used to measure the model's performance include accuracy, specificity, sensitivity, precision, and false positive rate (FPR), with a focus on accuracy emphasized by this paper. The accuracy yielded by the model developed in this paper reaches 97.10% for the adult dataset.

Sujatha *et al.* [15] presents a new approach to classifying autism using various machine learning algorithms. The study aims to facilitate early prediction and treatment of individuals predicted to have autism, which is beneficial in preventing worsening conditions. It utilizes four ASD datasets from the UCI machine learning repository, encompassing data with complete values. The datasets cover different age groups, ranging from adults, teenagers, children, to toddlers. The study applies multiple classification algorithms such as SVM, KNN, RF, NB, stochastic gradient descent (SGD), adaptive boosting (AdaBoost), and CN2 rule induction. The classification accuracy produced is remarkably high, with the SGD model achieving an accuracy of 99.7% for the adult dataset.

Thabtah and Peebles [16] introduces a new machine learning model called rules-machine learning (RML) for autism detection. This model offers classification with higher prediction accuracy. Data were collected using the ASDTests mobile application, employing four different methods to detect autism. The dataset used in this paper was gathered from September 2017 to January 2018, covering children, teenagers, and adults. Various features in the data include age, gender, and ethnicity. The RML method is based on covering classification, conducting evaluations to eliminate redundancy. Results on three datasets indicate that RML achieves classification with a prediction accuracy of approximately 95%, which is higher compared to other machine learning approaches such as boosting, bagging, decision tree (DT), and rule induction.

Thabtah [17] introduces ASDTests, a mobile application designed to assist in efficiently and easily detecting autism. This research offers an easily usable platform for healthcare professionals and individuals to conduct screening, providing multiple language options. The application has collected over 1,400 cases, allowing data analysis for autism levels. It employs machine learning algorithms like NB and LR to analyze features and enhance the accuracy of autism prediction outcomes. The study uses shortened versions of renowned screening methods like AQ and Q-CHAT tailored based on age groups. The machine learning classification results from this research show the highest accuracy level in the LR model, reaching 99.85% in the adult population.

3. RESEARCH METHOD

Figure 1 illustrates the four methodological stages employed in this research: preprocessing to optimize the dataset before utilization, selection of the classification algorithms to construct the model, model creation using the weighted average ensemble method, and finally, assessing the performance of the resultant model. These stages encompass the comprehensive approach taken to develop and evaluate the model in this study. The detailed explanation of each stage provides insight into the methodology and its execution.

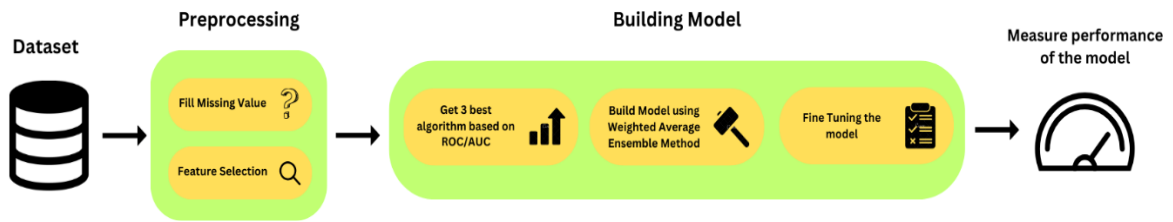


Figure 1. Flow of the proposed method

3.1. Dataset

The dataset utilized is sourced from Kaggle [18]. This dataset comprises 800 entries across 22 columns. Within these 800 entries, 161 are diagnosed autism cases while 639 are diagnosed as non-autism cases. Hence, considering these figures, it can be inferred that this dataset is imbalanced due to a significant disparity in the quantity between the classes. All patient data in this dataset belong to the adult age group. Specific details regarding the data set used are shown in Table 2.

Table 2. Dataset description

| No | Questions of AQ10 for adult | Description |
|----|-----------------------------|---|
| 1 | ID | Patient ID |
| 2 | A1_Score to A10_Score | The answer key for the 10-item screening tool AQ is as follows: Definitely agree and slightly agree (1) Slightly disagree and definitely disagree (0) |
| 3 | Age | The patient’s age in years |
| 4 | Gender | The patient’s gender |
| 5 | Ethnicity | The patient’s ethnicity |
| 6 | Jaundice | Did the patient suffer from jaundice at birth? |
| 7 | Autism | Has any family member of the patient been diagnosed with autism? |
| 8 | Country_of_residence | The country where the patient resides |
| 9 | Used_app_before | Has the patient ever had a screening test before? |
| 10 | Result | The scores obtained for AQ1-10 by the patient |
| 11 | Age_description | The description of the patient’s age |
| 12 | Relation | Relation of patient who completed the test |
| 13 | Class | The diagnosis result of the patient (autism (1)/not (0)) |

3.2. Preprocessing

Before constructing the model using various classification algorithms, a check was performed on the dataset that will be used. This was done with the aim of ensuring that the resultant model would exhibit good performance. The following is an explanation of several steps carried out at the preprocessing stage. Fill missing value is the first action carried out at the preprocessing stage. During the check for null data in the dataset, it was found that there were no empty entries across all attributes in this dataset. However, upon closer examination, some entries containing the value ‘?’ were discovered. These ‘?’ values were found in the ‘relation’ and ‘ethnicity’ features, leading to the assumption that these values were used as replacements for null values. Therefore, these ‘?’ values were replaced with “others”, aiming to align more consistently with the other data present in those features. Feature selection is an action that is carried out after the fill missing value work has been completed. This stage aims to choose which features have a direct impact on the output, allowing the utilization of various features in crafting a model that yields optimal performance. Features that are not selected will not be used in model creation. P-values were utilized for feature selection across various features in the dataset.

P-value serves as a statistical indicator used to measure the strength of evidence against the null hypothesis [18]. Its function involves comparing whether the p-value of a feature is greater or smaller than a predetermined value. If a feature’s p-value is smaller than the designated threshold, it is chosen for model construction. Conversely, if the p-value of a feature exceeds the set threshold, it is not used as it lacks significant influence on the resulting output. The p-value threshold is 0.05.

Following the results presented in Table 3, the decision was made to exclude the features “used_app_before” and “gender” from the model. This choice was informed by their respective p-values exceeding the established threshold of 0.05, with values of 0.37 and 0.97, respectively. The exclusion of these features ensures that they will not be used in the construction of the model.

Table 3. P-values each feature

| No | Fitur | p-values | status |
|----|-----------------|------------------------|--------|
| 1 | Gender | 0.9758243168741388 | Reject |
| 2 | Ethnicity | 1.4561482608094145e-33 | Accept |
| 3 | Jaundice | 0.00013300658957470307 | Accept |
| 4 | Autism | 1.0060560058593027e-23 | Accept |
| 5 | Contry_of_res | 2.8611111937550227e-19 | Accept |
| 6 | Used_app_before | 0.3745543476430917 | Reject |
| 7 | Relation | 0.03206590795264572 | Accept |
| 8 | A1_Score | 4.104487536920418e-17 | Accept |
| 9 | A2_Score | 1.3998012922364413e-25 | Accept |
| 10 | A3_Score | 2.4007562062687566e-38 | Accept |
| 11 | A4_Score | 4.8840206399536454e-45 | Accept |
| 12 | A5_Score | 1.7931000962761736e-38 | Accept |
| 13 | A6_Score | 1.3536803601031668e-52 | Accept |
| 14 | A7_Score | 5.621312869749746e-37 | Accept |
| 15 | A8_Score | 2.181610421234002e-18 | Accept |
| 16 | A9_Score | 9.762852300543611e-39 | Accept |
| 17 | A10_Score | 5.880205552224773e-22 | Accept |

3.3. Building model

Tests are underway for the upcoming model creation, evaluating 8 classification algorithms. The aim is to determine the 3 best classification algorithms for subsequent model construction based on their area under the curve (AUC)/receiver operating characteristic (ROC) values. Table 4 displays the results of testing the 8 classification algorithms, with the selected algorithms being those that exhibit the highest AUC/ROC values.

Table 4. Classification algorithm performance

| No | Classification algorithm | AUC/ROC |
|----|--------------------------|--------------------|
| 1 | LR | 0.9069129020892985 |
| 2 | SVM | 0.8646327672952457 |
| 3 | Gaussian NB | 0.9090057424018732 |
| 4 | MLP | 0.8993940931833094 |
| 5 | KNN | 0.7861399776865605 |
| 6 | DT | 0.7247017418277262 |
| 7 | RF | 0.9055184053253995 |
| 8 | GBoost | 0.9005207168262348 |

From the conducted tests, it was found that LR, gaussian, and RF are the algorithms with the highest AUC/ROC values. Therefore, the upcoming model will leverage these three algorithms. LR is one of the popularly used classification algorithms, serving as a predictive analysis technique reliant on probabilities. It models the relationship between covariates and categorical variables. LR is a linear model more suitable for classification issues than regression problems [19]. Gaussian NB is a simple learning algorithm that applies bayes' theorem, assuming that each input feature is independent. Despite its tendency to overly assume independence among all features and outcomes, this algorithm can still compete with other classifications due to its computational efficiency and low variance. Its strength lies in its effectiveness with large datasets and its ability to incrementally update with new training data. Typically used in classification, this algorithm computes the posterior probability of each class based on the provided features [20]. RF is an algorithm that combines multiple DT. It creates n trees from random subsets of features from the available data. Once the forest, comprising several trees, has been constructed, new data for classification is presented to each tree in the forest. The classification process is executed by each tree, and eventually, the class with the most votes from all the votes generated by each tree is chosen as the class for the new data [21].

Once the three classification algorithms to be used have been obtained, the next step is to build a model using weighted average ensemble method using these three algorithms. The model will be constructed utilizing the weighted average ensemble method. The working principle of this algorithm involves combining several existing classification algorithms, with each classification algorithm assigned its respective weight. The weight of each algorithm is determined based on the results obtained from the previous testing phase. The algorithm with the highest AUC/ROC value will be given a higher weight compared to the other algorithms.

In the model being developed, the greatest weight is held by the gaussian algorithm at 0.6, while the same weight of 0.2 is shared by the other two algorithms, LR and RF. For cross-validation, the StratifiedKFold method is employed. The StratifiedKFold method is a cross-validation technique used for evaluating a model. A set of folds is created, ensuring that each fold maintains an equal proportion of data from each existing class, which is particularly beneficial for datasets with imbalanced proportions [22].

The functioning of StratifiedKFold involves dividing the dataset into K folds. In each iteration, one-fold is used for testing, while the remaining 'K-1' folds are used for training the existing model. This process is repeated 'K' times, with each fold utilized precisely once for testing. The 'stratified' aspect ensures that the class distribution within each fold reflects the overall class distribution, thereby providing more reliable performance metrics for the created model [23]. The value of k for StratifiedKFold was set to 7, determined through the hyperparameter tuning process.

3.4. Measure performance of the model

The performance matrix utilized to measure the model's performance is AUC/ROC. AUC/ROC is a graphical representation for evaluating the performance of a classification model, depicting the relationship between sensitivity and specificity [24]. The formulas for calculating sensitivity in (1) and specificity in (2) are as:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2)$$

Markoulidakis *et.al.* [25] proposes the confusion matrix as a key tool in data analysis, offering a concise way to evaluate classification model performance. Categorizing predictions into true positives, true negatives, false positives, and false negatives, this matrix enhances precision in assessing model accuracy and errors. The proposal contributes to more informed decision-making across various domains, marking a significant advancement in data analysis.

On Table 5, it can be seen that in a classification, there are four possibilities of an event, namely:

- True positives (TP) = when positive data is classified as positive
- True negatives (TN) = when negative data is classified as negative
- False positives (FP) = when negative data is classified as positive
- False negatives (FN) = when positive data is classified as negative

Table 5. Confusion matrix

| | Real positif | Real negative |
|---------------------|----------------|----------------|
| Predictive positif | True positive | False negative |
| Predictive negative | False positive | True negative |

The AUC/ROC metric was chosen to evaluate the model due to the imbalance in the dataset used for its creation, where class comparisons were uneven. This metric is recognized as a reliable performance measure for models developed with imbalanced datasets. Weng and Poon [26] validates the appropriateness of using AUC/ROC in such situations.

4. RESULTS AND DISCUSSION

In model development, testing was conducted by attempting to create models, setting the value of K for StratifiedKFold from 5 to 10. The K value resulting in the highest performing model will be utilized. The test results for the K value in model creation are provided, where the weight values for each algorithm are 0.6 for the gaussian algorithm and 0.2 for the LR and RF algorithms.

On the test results listed in Table 6, it was found that a better-performing model was achieved when a K value of 7 was employed compared to other values. Therefore, 7 was established as the K value for StratifiedKFold. Following the evaluation to determine the optimal K value, tests were also conducted to set the weight values for each algorithm used in the model. During these tests, a higher weight was assigned to the gaussian algorithm compared to the other two algorithms, because during the evaluation to determine the chosen algorithm for model creation, this algorithm was outperformed by the others. The test results for the weight values assigned to each algorithm are presented here.

Table 6. The results of the K value testing in creating the AUC/ROC model

| K | AUC/ROC |
|----|--------------------|
| 5 | 0.9027589346963731 |
| 6 | 0.9078098146163743 |
| 7 | 0.9122867410742107 |
| 8 | 0.9098038633966246 |
| 9 | 0.9070494134012499 |
| 10 | 0.9084339074171336 |

Based on the test results listed in Table 7, it's evident that the combination of weight values, 0.6 for gaussian and 0.2 for LR and RF, yields the highest-performing model compared to other weight value combinations. Hence, this weight combination will be employed in model creation. However, the testing results indicate that the performance difference among these weight combinations is not significant. After completing multiple tuning processes to determine the optimal values for K and weights, tests will be conducted using various ensembling methods to compare the performance of the constructed models.

Table 7. The results of the weight value testing for each algorithm

| Gaussian NB | LR | RF | AUC/ROC |
|-------------|------|------|--------------------|
| 0.4 | 0.3 | 0.3 | 0.9095120380564615 |
| 0.6 | 0.2 | 0.2 | 0.9122867410742107 |
| 0.5 | 0.25 | 0.25 | 0.9083272250298985 |
| 0.65 | 0.15 | 0.2 | 0.9113890430390565 |
| 0.65 | 0.2 | 0.15 | 0.9094563956162661 |

Table 8 it is evident that the performance of the model generated using the weighted average ensemble method is superior when compared to other ensembling methods. Thus, it is proven that models created through the weighted averaging ensembling method outperform models created using other ensembling methods and models created solely by leveraging a single classification algorithm. The performance comparison results for the model constructed against all previously tested models are presented in Table 9.

Table 8. Result ensembling method

| Ensembling method | AUC/ROC |
|-------------------|--------------------|
| Weighted average | 0.9122867410742107 |
| Averaging | 0.8571662269677393 |
| Voting | 0.7830267558528428 |

Table 9. The results of the weight value testing for each algorithm

| No | Algorithm | AUC/ROC |
|----|----------------------------------|---------|
| 1 | Weighted average ensemble method | 0.912 |
| 2 | Gaussian NB | 0.909 |
| 3 | LR | 0.906 |
| 4 | RF | 0.905 |
| 5 | GBoost | 0.900 |
| 6 | MLP | 0.899 |
| 7 | SVM | 0.864 |
| 8 | Ensemble averaging | 0.857 |
| 9 | Voting ensemble method | 0.783 |
| 10 | DT | 0.724 |




5. CONCLUSION

In this research, a model was developed to identify autism in adults using a dataset of individuals with ASD. Feature selection was done using the p-value method before building the model. Initial tests were carried out with the top 3 performing algorithms from 8 classification methods to select the best algorithm. The final model was created using the weighted average ensemble method combining these top three algorithms, outperforming models based on a single algorithm. However, the study's limitations include a small dataset size and class imbalance. Future research should focus on using larger, more balanced datasets and exploring deep learning techniques to improve the model's performance.




REFERENCES

- [1] R. Loomes, L. Hull, and W. P. L. Mandy, "What Is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 56, no. 6, pp. 466–474, Jun. 2017, doi: 10.1016/j.jaac.2017.03.013.
- [2] M. C. Lai *et al.*, "Improving autism identification and support for individuals assigned female at birth: clinical suggestions and research priorities," *The Lancet Child and Adolescent Health*, vol. 7, no. 12, pp. 897–908, Dec. 2023, doi: 10.1016/S2352-4642(23)00221-3.
- [3] D. Bemmouna, S. Weibel, M. Kosel, R. Hasler, L. Weiner, and N. Perroud, "The utility of the autism-spectrum quotient to screen for autism spectrum disorder in adults with attention deficit/hyperactivity disorder," *Psychiatry Research*, vol. 312, p. 114580, Jun. 2022, doi: 10.1016/j.psychres.2022.114580.
- [4] P. Ghanouni and L. Seaker, "What does receiving autism diagnosis in adulthood look like? Stakeholders' experiences and inputs," *International Journal of Mental Health Systems*, vol. 17, no. 1, p. 16, Jun. 2023, doi: 10.1186/s13033-023-00587-6.
- [5] S. G. M. Wouters and A. A. Spek, "The use of the Autism-spectrum quotient in differentiating high-functioning adults with autism, adults with schizophrenia and a neurotypical adult control group," *Research in Autism Spectrum Disorders*, vol. 5, no. 3, pp. 1169–1175, Jul. 2011, doi: 10.1016/j.rasd.2011.01.002.
- [6] J. Dolah, W. A. J. W. Yahaya, T. S. Chong, and A. R. Mohamed, "Identifying autism symptoms using autism spectrum quotient (ASQ)," *Procedia - Social and Behavioral Sciences*, vol. 64, pp. 618–625, Nov. 2012, doi: 10.1016/j.sbspro.2012.11.072.
- [7] K. L. Ashwood *et al.*, "Predicting the diagnosis of autism in adults using the autism-spectrum quotient (AQ) questionnaire," *Psychological Medicine*, vol. 46, no. 12, pp. 2595–2604, Sep. 2016, doi: 10.1017/S0033291716001082.
- [8] J. Zeidan *et al.*, "Global prevalence of autism: a systematic review update," *Autism Research*, vol. 15, no. 5, pp. 778–790, May 2022, doi: 10.1002/aur.2696.
- [9] N. Salari *et al.*, "The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis," *Italian Journal of Pediatrics*, vol. 48, no. 1, p. 112, Dec. 2022, doi: 10.1186/s13052-022-01310-w.
- [10] A. V. Shinde and D. D. Patil, "A multi-classifier-based recommender system for early autism spectrum disorder detection using machine learning," *Healthcare Analytics*, vol. 4, p. 100211, Dec. 2023, doi: 10.1016/j.health.2023.100211.
- [11] 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 and Adolescent Psychiatry*, vol. 51, no. 2, pp. 202–212.e7, Feb. 2012, doi: 10.1016/j.jaac.2011.11.003.
- [12] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of autism spectrum disorder in children using machine learning techniques," *SN Computer Science*, vol. 2, no. 5, p. 386, Sep. 2021, doi: 10.1007/s42979-021-00776-5.
- [13] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets," *Health Information Science and Systems*, vol. 9, no. 1, p. 17, Dec. 2021, doi: 10.1007/s13755-021-00145-9.
- [14] K. S. Oma, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, "A machine learning approach to predict autism spectrum disorder," in *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, Feb. 2019, pp. 1–6, doi: 10.1109/ECACE.2019.8679454.
- [15] R. Sujatha, S. L. Aarthy, J. M. Chatterjee, A. Alaboudi, and N. Z. Jhanjhi, "A machine learning way to classify autism spectrum disorder," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 6, pp. 182–200, Mar. 2021, doi: 10.3991/ijet.v16i06.19559.
- [16] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health Informatics Journal*, vol. 26, no. 1, pp. 264–286, Mar. 2020, doi: 10.1177/1460458218824711.
- [17] F. Thabtah, "An accessible and efficient autism screening method for behavioural data and predictive analyses," *Health Informatics Journal*, vol. 25, no. 4, pp. 1739–1755, Dec. 2019, doi: 10.1177/1460458218796636.
- [18] J. Gao, "P-values - A chronic conundrum," *BMC Medical Research Methodology*, vol. 20, no. 1, p. 167, Dec. 2020, doi: 10.1186/s12874-020-01051-6.
- [19] A. H. Ataya, "Early detection of diabetes using machine learning techniques," in *Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023*, Feb. 2023, pp. 886–891, doi: 10.1109/ICAIS56108.2023.10073861.
- [20] M. V. Anand, B. Kiranbala, S. R. Srividhya, K. C., M. Younus, and M. H. Rahman, "Gaussian naïve bayes algorithm: a reliable technique involved in the assortment of the segregation in cancer," *Mobile Information Systems*, vol. 2022, pp. 1–7, Jun. 2022, doi: 10.1155/2022/2436946.
- [21] S. Pavithra, R. Vanithamani, and J. Justin, "Computer aided breast cancer detection using ultrasound images," *Materials Today: Proceedings*, vol. 33, pp. 4802–4807, 2020, doi: 10.1016/j.matpr.2020.08.381.
- [22] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified K-fold cross validation optimization on machine learning for prediction," *Sinkron*, vol. 7, no. 4, pp. 2407–2414, Oct. 2022, doi: 10.33395/sinkron.v7i4.11792.
- [23] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Frontiers in Nanotechnology*, vol. 4, Aug. 2022, doi: 10.3389/fnano.2022.972421.
- [24] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.
- [25] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass confusion matrix reduction method and its application on net promoter score classification problem," *Technologies*, vol. 9, no. 4, p. 81, Nov. 2021, doi: 10.3390/technologies9040081.
- [26] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," *Conferences in Research and Practice in Information Technology Series*, vol. 87, pp. 27–32, 2008.




BIOGRAPHIES OF AUTHORS

Lawysen    was born in March 15, 2002 in the city of Jakarta, Indonesia. He obtained a Bachelor of Computers degree from Bina Nusantara University in 2024 and is currently studying to obtain a Master of Information Technology degree at Bina Nusantara University. His research interests are intelligent systems/machine learning. He can be contacted at email: lawysen@binus.ac.id.



Nelsen Anggara    was born in Mei 12, 2002 in the city of Jakarta, Indonesia. He obtained a Bachelor of Computers degree from Bina Nusantara University in 2024 and is currently studying to obtain a Master of Information Technology degree at Bina Nusantara University. His research interests are intelligent systems/machine learning. He can be contacted at email: nelsen.anggara@binus.ac.id.



Abba Suganda Girsang    his first degree, namely a Bachelor's degree in Electrical and Electronics Engineering in 2000 from Gajah Mada University, Indonesia. He also earned a Master of Computer Science from Gajah Mada University in 2008 and a Ph.D. from National Cheng Kung University in 2014. His main research interests are business intelligence and analytics, intelligent systems/machine learning, and deep learning. He can be contacted at email: agirsang@binus.edu.