Machine Learning in Identifying Autism Spectrum Disorder

Hiren Kumar Praharaj, Sasmita Mishra

College Of Engineering Bhubaneswar,Biju Pattnaik University of Technology,Odisha, India

***Abstract*—** A severe neurological disorder affecting the entire brain system, autism spectrum disorder (ASD) has an effect on a person's cognitive, emotional, social, and physical health. They find it challenging to interact and communicate with other people. They are constantly in need of guidance and support from friends, parents, or other family members. Although there is currently no known cure for autism, early discovery can lead to more effective therapy. Autism disorder diagnosis can be made based on a person's behavioral patterns. This diagnostic approach is laborious and unproductive for identifying autism in its early stages. Therefore, in order to assist people in determining whether they should receive a clinical diagnosis and pursue therapy, quick and affordable ASD screening is required. As a result, we provide a quick and easy way to identify autism using machine learning.

Keywords: Simple Neural Network, WEKA Tools, Fuzzy Techniques, K fold cross validation.

I. INTRODUCTION

Autism is one of the major problems facing humanity. This has an impact on a person's general behavior. An individual's physical, mental, emotional, and social well-being will also be impacted. Individuals at all ages—toddlers, children, teenagers, adults, and senior citizens—can exhibit it. The autism screening exams cost money and take a lot of time. Based on the forecast or prognosis of the machine learning model, a machine learning-based method is suggested to assist the individual in determining whether or not to receive a formal clinical diagnosis. Although there is no known cure for ASD, early discovery might help identify more effective treatment strategies. This works really well and can save a lot of money on medical expenses.

II. RELATED WORK

1. Simple neural network models have been used by Madhura Ingalhalikar, Sumeet Shinde, Arnav Karmarkar, Archith Rajan, Dr. Rangaprakash, and Gopikrishna Deshpande (2021) to categorise their Classification Techniques have been utilised by M. S. Mythili and A. R. Mohamed Shanavas (2014) to research ASD. This paper's primary goals were to identify autism and its severity degrees. SVM and neural networks were two of the classification algorithms employed. WeKA tools and fuzzy techniques were also employed to examine the social interactionand conduct of the students.

2.In order to identify autism, J. A. Kosmicki1, V. Sochat, M. Duda, and D.P. Wall (2015) utilised a strategy of searching for the smallest collection of features. To assess the clinical assessment of ASD, the authors employed a machine learning methodology. The ADOS was applied to children's behavior that fell inside the autism spectrum. In this research, eight distinct machine learning algorithms from ADOS's four modules were applied. Stepwise backward feature recognition on score sheets from 4540 people was another aspect of

the study. With an overall accuracy of 98.27 percent and 97.66 percent, respectively, it used 9 out of the 28 behaviors from module 2 and it had employed 12out of the 28 behaviours from module 3 to detect an ASD risk.

3. Fadi Thabtah has suggested an ASD screening method that makes use of machine learning adaption and the DSM-5 (2017). In this article, the researcher discussed the benefits and drawbacks of the ASD Machine Learning categorization. He has also made an effort to call attention to the problems with current ASD screening techniques and the way they consistently rely on the DSM-IV rather than the DSM-5 manual.

III. SYSTEM DESIGN AND ARCHITECTURE

A. SYSTEM DESIGN

The architecture follows the following flow:

• Collection of Dataset The Autism Screening Datasets are used and include age groupings for adults, toddlers, and children. Datasets were gathered via the UCI Repository and Kaggle.

• Data Pre-processing the raw data will be cleaned by data pre-processing.

• Model development and Evaluation after pre-processing, the dataset gets split into testing and training sets. Multiple classifiers are developed using prominent machine learning algorithms (Decision Tree, Naive Bayes, KNN and SVM). In the training phase, the training data is given to train the classifier. In the testing phase, class predictions are made on the test dataset. The

classifiers are evaluated based on their performances in diagnosing autism. The test data is used for model evaluation to evaluate a model based on its performance and accuracy. The accuracies of the classification algorithms are compared. Using Voting classifier, a hybrid ensemble machine learning model is developed.
Accuracy and recall are the metrics that are computed during the model evaluation process. The performance metric accuracy is the ratio of correctly predicted observations to all observations. Recall is the ratio of correctly predicted positive observations to all positive observations (output label 1). The voting classifier is evaluated using Repeated Stratified K fold cross validation. From cross validation, accuracy scores and recall scores are obtained.

• Model deployment The model which has the highest accuracy is then deployed using Flask. In the frontend the user has to input their basic form details and according to the age category, the user attempts the ASD Screening test. At the backend, the user's input is pre-processed. The model is loaded from the pickle file and performs prediction with the processed input. The model's prediction is displayed at the user's front end.

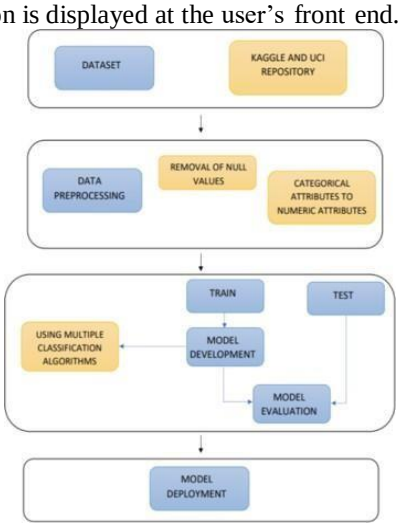


Fig. 1. System Architecture

IV. PROPOSED WORK

A. Pre-processing

Three datasets for different age categories of Adult, Children and Toddlers are used. The datasets are obtained from Kaggle and UCI repository. There are nearly 20 attributes in each of the dataset which consist of categorical, continuous, and binary values. The Class/ASD output label indicates if a person has ASD
(1) or not (0). Firstly, the raw data is pre-processed. Unnecessary columns are dropped, column names renamed so that they are uniform across all datasets. Null values removed. Repository.

B. Encoding

Encoding is performed on categorical values to convert string values into numerical values. For this purpose, Label Encoder is used. The label encoder is saved as a pickle file for further use. This pickle file is further used at the backend to encode the input obtained from the user.

C. Model development

After pre-processing, the dataset is split into testing set and training set. A train test split ratio of 80:20 is used. Multiple classification algorithms are used to develop the model. The classification algorithms used are K-nearest neighbors, Sup- port Vector Machine, Decision Tree, and Naive bayes. Using these classification algorithms, a hybrid ensemble machine learning model is developed using Voting classifier. Both hard voting and soft voting methods are used.

The Voting Classifier is an estimator that combines representations of many classification techniques along with individual confidence weights. The Voting estimator, which was created by integrating various classification models, is a powerful meta-classifier that effectively counteracts the limitations of the individual classifiers on a given dataset. Voting classifier assigns a class label to a record based on a majority vote and weights applied to the class or class probabilities.

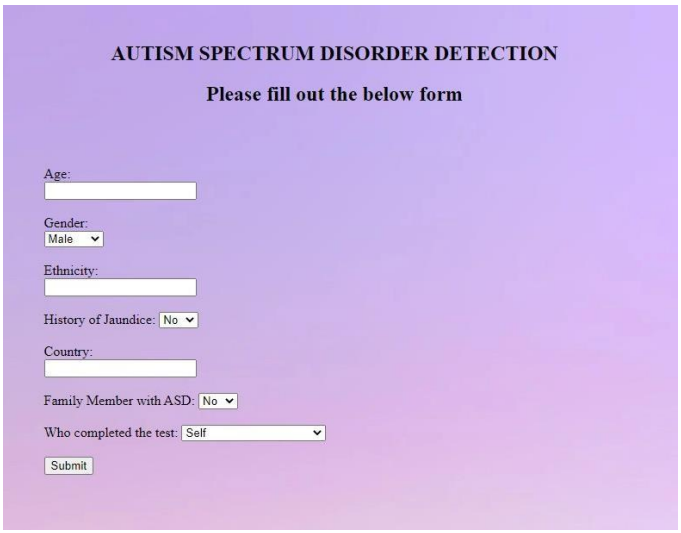
D. Model Evaluation

The Voting classifier is evaluated using Repeated Stratified K Fold cross validation that repeats Stratified K-Fold n times. Three repeats of stratified 10- fold cross-validation is performed. K Fold: Split dataset into k consecutive folds. Stratified: The folds are made by preserving the percentage of samples for each class. Repeats: Number of times cross validator needs to be repeated.

The metrics that are calculated are: Accuracy and recall The average of all the accuracies and recall scores are computed and the mean accuracy and recall score is produced.

E. Model Deployment

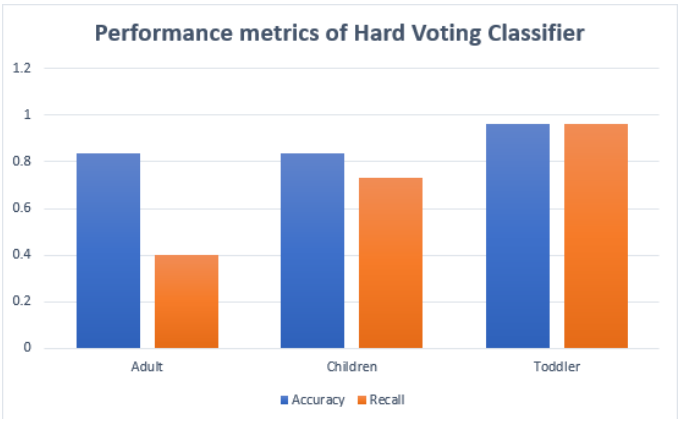
The flask webapp starts with the form page that asks the user to enter the required input. After the user submits the input through a form, the user attempts the ASD Screening test. After user submits the test, backend receives the input data. The user's input is first processed before performing the prediction. The encoder is loaded from the pickle file and is applied over the user's input. The model is loaded from the pickle file and performs prediction with the processed input. The model's prediction, the result of whether the user is autistic or not, and the class of the prediction are computed and displayed at the user's front end.



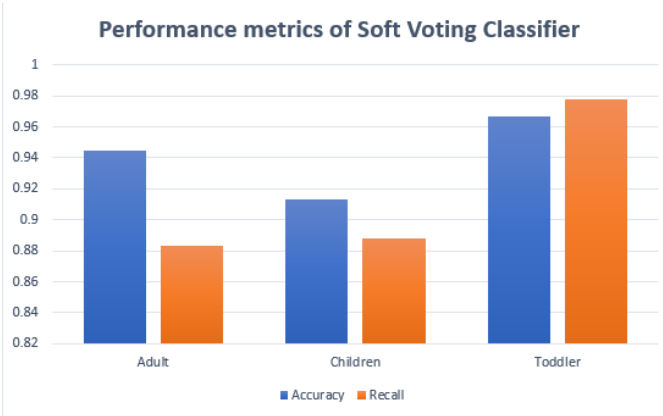
`Fig. 2: Web page

V. RESULTS

The performance metrics of the Hard Voting Classifier obtained are: accuracy of 83.76% for adults, accuracy of 84.48% for children, and accuracy of 96.39% for toddlers. The Recall values of the Hard Voting classifiers are 39.88% for adults, 73.11% for children, 96.11% for toddlers.



The performance metrics of the Soft Voting Classifierobtained are: accuracy of 94.45% for adults, 91.30% for children and 96.71% for toddlers. The Recall values of the Soft Voting classifiers are 88.35% for adults,88.77% for children, 97.85%for toddlers.



In comparison, the Soft Voting Classifier has betterperformance metrics than Hard Voting Classifier.

VI. CONCLUSION

The findings imply that the Hard Voting classifier is not as effective as the Soft Voting classifier. The Naive Bayes classifier outperforms the ensemble model on the adult dataset by a little margin when the classification methods and the ensemble model are compared. For many datasets, the Soft Voting classifier performs better in terms of accuracy than other approaches. If an autism diagnosis is made, the patient is referred to support groups that address the needs of autistic people and urged to obtain a formal clinical diagnosis. It is advised to provide a list of suitable establishments for every age group, which should include special education schools and other pertinent facilities.

REFERENCES

[1] M. Ingalhalikar, S. Shinde, A. Karmarkar, A. Rajan, D. Rangaprakash and G. Deshpande, "Functional Connectivity-Based Prediction of Autism on Site Harmonized ABIDE Dataset," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 12, pp. 3628-3637, Dec. 2021,doi: 10.1109/TBME.2021.3080259.

[2] M. F. Rabbi, S. M. M. Hasan, A. I. Champa and M. A. Zaman, "A Convolution Neural Network Model for Early-Stage Detection of Autism Spectrum Disorder," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 110-114, doi: 10.1109/ICICT4SD50815.2021.9397020.

[3] Mythili, M. S., and AR Mohamed Shanavas. "An Analysis of students' performance using classification algorithms." IOSR Journal of Computer Engineering 16.1 (2014): 63-69.

[4] Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Translationalpsychiatry, 5(2), e514- e514.

[5] Thabtah, Fadi. "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." Proceedings of the 1st International Conference on Medical and health Informatics 2017. 2017.

[6] Vaishali, R., and R. Sasikala. "A machine learning based approach to classify autism with optimum behavior sets." International Journal of Engineering & Technology 7.4 (2018): 18.

[7] Vakadkar, Kaushik, Diya Purkayastha, and Deepa Krishnan."Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques." SN Computer Science 2.5 (2021): 1-9.Cavus, Nadire, Abdulmalik A. Lawan, Zurki Ibrahim, Abdullahi Dahiru, Sadiya Tahir, Usama Ishaq Abdulrazak, and Adamu Hussaini. "A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder." Journal of Personalized Medicine 11, no. 4 (2021): 299.

[8] S.B. Shuvo, J. Ghosh and A. S. Oyshi, "A Data Mining Based Approach to Predict Autism Spectrum Disorder Considering Behavioral Attributes," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-5, doi:10.1109/ICCCNT45670.2019.8944905.

[9] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, United States, 2008. morbidity and mortality weekly report.61." Centers for Disease Control and Prevention, 2012.

[10] S. E. Bryson, L. Zwaigenbaum, and W. Roberts, "The early detection of autism in clinical practice," Pediatrics & child health, vol. 9, no. 4, pp. 219– 221, 2004.

[11] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," and "A complete guide to the random forest algorithm," Built In, vol. 16, 2019.

[12] Haishuai Wang, Li LiLianhua Chi, Ziping Zhao, "Autism Screening Using Deep Embedding Representation," International Conference on Computational Science, Lecture Notes in Computer Science, vol 11537, pp. 160-173, jun 2019.

[13] Muhammad Nazrul Islam, Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on Electrical, Computer and Communication Engineering, feb 2019.