International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

ASD SCREENING USING MACHINE LEARNING

Sagar Sharma
Scholar Student, CSE, HMR Institute of Technology and Management
Delhi, India

Mandeep Singh
Assistant professor, CSE, HMR Institute of Technology and Management
Delhi, India

Rishabh Singh
Scholar Student, CSE, HMR Institute of Technology and Management
Delhi, India

Rohit Prajapati
Scholar Student, CSE, HMR Institute of Technology and Management
Delhi, India

Ashish Rana
Scholar Student, CSE, HMR Institute of Technology and Management
Delhi, India

Abstract*—Autism spectrum disorder (ASD) is a condition that creates challenges for the person affected in performing tasks like social interaction, speech and nonverbal communication, and they might show restrictive/repetitive behaviors. ASD starts showing symptoms in early childhood with many of the most-obvious signs presenting around the age of 2-3 years old, but in some cases, children with autism develop just like any other child until toddlerhood when they stop acquiring or lose previously gained skills. ASD screening is the first step in diagnosing and treatment of the disorder. While there is no medicinal cure for ASD, early treatment can help reduce autism symptoms and improve quality of life. With the advancements and use of machine learning techniques in the medical and healthcare sector, this project is an attempt to predict ASD in children based on the symptoms shown by them. Dataset being used comprises 21 attributes which includes test score, age, gender, autism, jaundice, etc. Datasets are publicly available on UCI Machine Learning Repository. We have tried to select the best features and combine them with some of the best Machine Learning techniques to produce results with accuracy.

I. INTRODUCION

Machine learning, described in simple words, is computer algorithms using the data to reason over and improve the results automatically by learning the patterns found in the data which is not visible to human eyes. It is a subset of Artificial Intelligence which allows systems to learn from experience (historical data) without being programmed explicitly. The goal was to provide computers the ability to learn and improve over time, an attempt to imitate and reproduce human ability. But over time, with substantial growth in the sector of computing power, machine learning opened the path for doing things and uncovering patterns which were unattainable to humans. And since then, machine learning is being used in different fields every day to achieve results once we thought were never possible.

Machine learning has made a huge impact in the medical and healthcare sector by analyzing and predicting many diseases like tumors, cancer, heart attack and many more which helps in early detection and treatment saving lives. Many diseases have medical tests and diagnosis which can confirm if a patient has that specific disease but some disorders still cannot be diagnosed by medical tests. They need careful evaluation and screening of the patient to determine if they are suffering from that disease or not. One of these problems is Autism Spectrum

Disorder (ASD). According to the data collected by CDC, 1 in 68 children have been identified with ASD. The disorder ranges from all racial, ethnic and socioeconomic groups, and has quadrupled the frequency among males (1 in 42) compared to females (1 in189). But we can use the data collected by previous patients, who may or may not have ASD, and Machine learning to try to uncover hidden relations in the data of patients and then try predicting if a patient is suffering from ASD or not. It is called a "spectrum" disorder because people with ASD can have a range of symptoms. It is very difficult to predict ASD because there are several other mental disorders such as attention deficit hyperactivity disorder (ADHD). with few symptoms very similar to ASD.

Patients suffering from autism face difficulty in most basic day-to-day tasks like social communication and interaction showing limited to no interest and displaying repetitive behaviors. Listed below are some examples of the recurrent behaviors among people diagnosed with ASD:

* Having very little/inconsistent eye contact with no proneness to look at or listen to people
* Almost never displaying the fondness of objects like toys or surroundings by pointing at them to others.
* Being slow or failing to respond when their name is called or any other attempts are made to catch their attention.
* Either having difficulties participating in a back and forth conversation, or lecturing about a random topic without giving any regards to others interest in the conversation.
* Making inappropriate facial expressions or gestures that do not fit to what is being said or talking in an abnormal tone like a robot or sing-song for example.
* Not understanding others point of view or feelings, or not being able to anticipate or understand actions of other people.
* Precise repetition of words or phrases spoken by another person (known as echolalia) or repetition of certain behaviour.
* Having unusually intense interest in subjects like facts and numbers, or in moving objects.
* Being unsettled by small changes in their routine or surroundings.
* Showing more reactiveness or less reactiveness compared to others towards sensorial input like light, sound, or temperature.
* Experiencing problems related to sleep and irritability.

As of today, here is no standard treatment for ASD. There are a lot of ways to increase the ability of ASD affected children to learn new skills. Starting early after the diagnosis can lead to better results in most of the cases with the help of some treatments like behavior and communication therapies, skills training, and medicines to control symptoms.

The exact causes of ASD is still not known to the researchers who are trying to find more about the root of this disorder. Conducted studies imply that genes can act in an uncommon way under the influences from the surroundings which might affect development of a person in ways, ultimately leading to ASD.

With the causes of autism unknown as of yet, there are some factors that might increase the danger of developing ASD like having older parents or having a brother or sister with autism. Certain conditions like Down syndrome and Rett syndrome can also increase the chances of ASD. But any of these factors are not a guarantee that the patient will have ASD.

II. RELATED WORK

Following is the brief description on works done prior related to Predicting ASD using Machine Learning. Hyde et al [2] presented a review paper contributing to research literature providing a comprehensive review of 45 papers which utilized supervised machine

learning in predicting ASD. Their work gives guidance for others interested in researching in the field of ASD.

In [1] Omar et al used different machine learning models for predicting ASD and decided that the prediction model created by merging Random Forest-CART (Classification and Regression Trees) and Random Forest-ID3 (Iterative Dichotomiser 3) produced the better results in terms of accuracy, specificity, sensitivity, precision and false positive rate (FPR) for datasets used.Using their proposed model, a mobile application was also developed. Halim et al [3] took a different approach by taking two different methods, one predicting based on questions related to child's behavior answered by parents/guardians and other predicting by monitoring the behavior of children from short videos, and combining both of their outcomes to produce a much better end result with higher accuracy. A multi-center clinic study was also conducted which showed significantly better performance suggesting that these mobile methods using machine learning are reliable for detection of ASD.

Thabtah [4] also used machine learning and tackled the task of diagnosing ASD as a classification problem where predictive models are built on collected data. They highlighted a problem in ASD screening tools which were using the DSM-4 rather than DSM-5 manual. They also suggested amends to the tools to meet the new DSM-5 ASD criteria. Thabtah [5] also collected the datasets of child, adolescent and adult containing 292, 104 and 704 instances respectively.

Raj and Masood [6] attempted to implement various machine learning techniques like Naïve Bayes, SVM, Logistic regression, CNN, etc. and compare their results to find the best results in predicting ASD. After the research, it was concluded that CNN based predictions were better than others with accuracy of 99.53%, 98.30%, 99.8% in ASD screening for adult, children and adolescents respectively using the datasets [6] for training their model. For the same purpose, Bone et al [7] implemented SVM (Support Vector Machine) and achieved 89.2% sensitivity and 59% specificity using data of 1726 individuals (1264 and 462 with ASD and NON-ASD traits respectively) but their research was not accepted because of a wide range of age of people whereas Hauck and Kliewer [8] took a relatively different approach which consisted of identifying more important screening questions to be used in methods like Autism Diagnostic Observation Schedule and Autism Diagnostic Interview Revised, and showed that ADI-R and ADOS screening test provide improved results when combined together. Heinsfeld et al [9] choose a different approach for identifying ASD in patients with the help of a large brain imaging dataset from the Autism Imaging Data Exchange (ABIDE I) and applying a neural network on it. They were able to achieve 70%, 65% and 63% mean classification accuracy using deep learning, SVM and random forest classifier respectively.

Li B, Sharma, Meng et al. [10] used machine learning classifiers to detect autism in adults by imitation method. Their dataset was based on 16 ASC participants and their hand movements. Wall, Dally, Luyster et al. [11] tried to lower the screening time and more rapid detection of ASD traits using the Alternating Decision Tree. Linstead, E., German et al. [12] worked on a sample size of 726 children applied artificial neural networks for the task of detecting autism and proved that it outperforms linear regression models used in previous studies. Hyde et al. [13] researched on the effect of ASD on the employment of individuals suffering from "high-functioning" ASD. Their study also analyzed the factors that contribute to whether or not they will be recruited.

Research done by Wiggins et al. [14] was based on the SEED algorithm which utilized s the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) to classify children with ASD with results supporting their work. Wall et al. [15] reached the conclusion that ADTree and the functional tree produced better results at higher sensitivity and accuracy while working on classifying ASD with short screening tests and validation. Bennet and Goodall [16] did a meta-analysis study which showed that people diagnosed

with DSM-IV or DSM-IV-TR ASD had a 35 and 37 % reduction respectively for a DSM-5 ASD diagnosis.

From all the work mentioned above in the Literature Survey, it can be concluded that despite a number of researches done in this field, there is still not a very reliable tool for predicting ASD using machine learning. Many works have shown good results but only a handful can be integrated with a mobile application and used at home by the parents.

III. EXPLORING THE STRATEGIC BASES OF ASD

A. Prediction model

AQ-10 dataset [5] was used for developing the machine learning based prediction model which contains 10 questions where only 1 or 0 can be scored for each answer recorded. These screening questions are mainly based on a child's day-to-day behavior like social interaction, answering when called, pointing at things or paying attention. Dataset consisted of 21 attributes, which are both in numerical or categorical form, including age, gender, ethnicity, country of residence, family member with PDD, answers for Q1-10, etc.

Data was cleaned by removing the irrelevant columns and filling in the missing values to make the prediction model more accurate. A number of machine learning algorithms including but not limited to Logistic Regression, Naïve Bayes, SVM, KNN, Decision Trees and CNN were used to train the model with cleaned data. After training the models and comparing their results, CNN (Convolutional Neural Networks) was chosen as the final prediction model and it was implemented using TensorFlow library, which is an open-source machine learning library best known for training deep neural networks.

Convolutional neural network, which is classified as a class of deep neural networks, is a popular pattern recognition algorithm and is very efficient at doing so. Its architecture, as shown in figure 1., comprises input layer, neurons (hidden layer) and output layer. The neurons are connected to each other forming a network which calculates the weighted sum of inputs for each neutron and produces activation values as output. When input is passed to the network of neurons, the first layer extracts the features from the input and its output is passed to the second layer of the network to extract more complex features. The networks can be built as deep as the high-level feature extraction required from them.

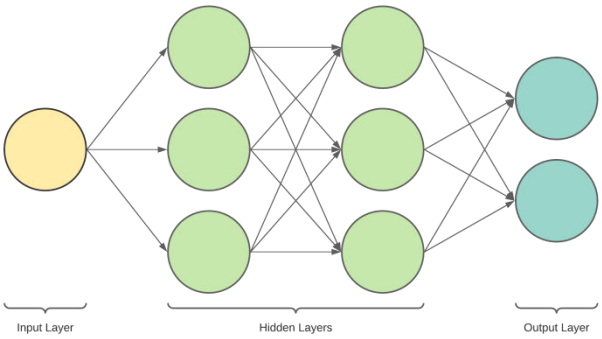

Figure 1. Architecture of a basic convolutional neural network

B. Mobile application

An easy-to-use mobile application, shown in figure 2, was developed with a user-friendly and very informative interface using Flutter, which is an open-source UI SDK created by Google. Applications developed using Flutter can run on multiple platforms including but not limited to Android, iOS, Windows, Linux and web from a single codebase. Flutter's engine is written in C++ which makes the execution speed fast whereas the applications in Flutter are written in Dart language.

TensorFlow Lite was used for the ease of running TensorFlow models on mobile and features like on-device machine learning inference with low latency and a small binary size. Models can be trained easily using transfer learning and then converted using TensorFlow Lite converter. Inference is run with TensorFlow Lite interpreter. It also uses device's hardware like GPU for acceleration in machine learning operations and depending on the device, speed-up can be substantial. At last, the model is optimized using the tools provided in TensorFlow Lite known as Model Optimization Toolkit.

The goal of the application was to record the response of parents/guardians for the AQ-10 questions and then use that data as input for the prediction model embedded in the application to make predictions on whether there is a possibility of ASD or not using machine learning.

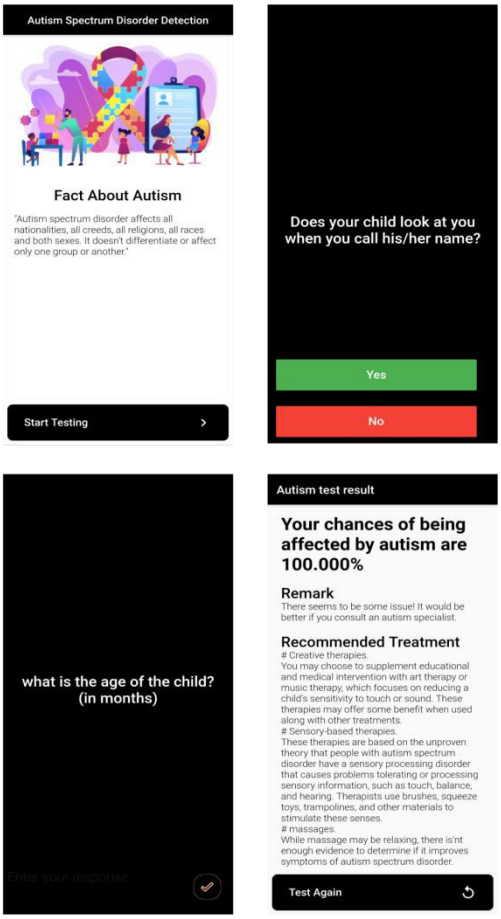

Figure 2. User interface of application

IV. PROPOSED METHODOLOGY

Whole process takes place in 3 stages – Data Preprocessing, Model Training and Testing, and Integration with Front-End. First two stages are done completely using Python 3 and integration is done using TensorFlow Lite and Flutter. Overall flow of work can be easily understood from figure 3.

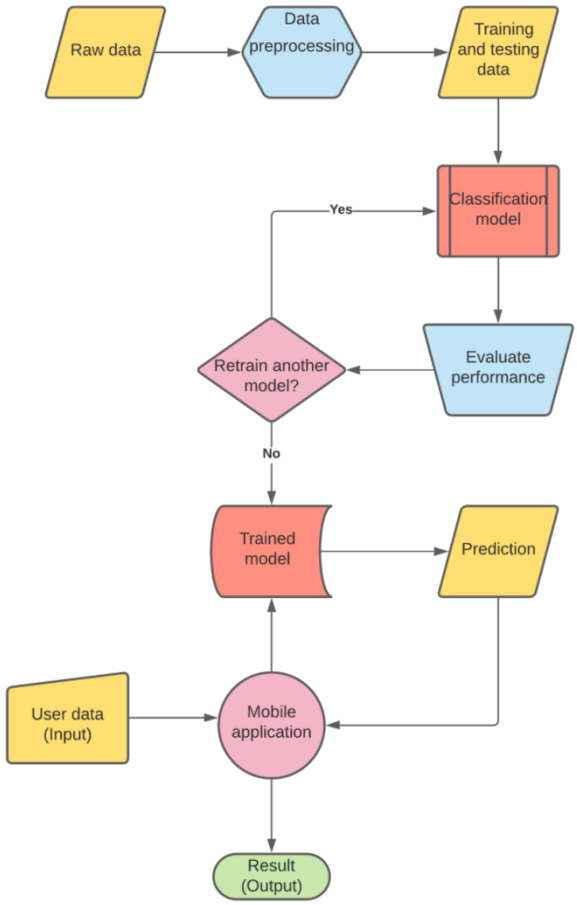

Figure 3. Flow of work

A. Data Cleaning and Analysis

Data cleaning is a part of data pre-processing and it includes preparing the data beforehand for the purpose of ease in analysis by removing and/or modifying data that is defective, incomplete, useless, or repeated. It is done to increase the accuracy of the dataset without removing any data points if possible. The main goal of data cleaning is to create standardized datasets which can be analyzed easily.

Data analysis is also a part of data pre-processing and it consists of inspecting, cleansing, transforming, and modeling data which was cleaned with the goal of gaining useful information and helping in choosing the right machine learning model which is best suited for the data. Sometimes problems can be solved just by cleaning and analyzing the provided data by finding

some hidden insights in the data using other tools and visualization techniques.

B. Model Training and Testing

The dataset being used is divided in two parts – training set and testing set. Testing set is kept untouched during the training of the model to check the accuracy and result afterwards. This helps in checking if problems like over-fitting exist in our model or not. Different machine learning algorithms are used to determine the best one suited for predicting ASD with best accuracy and result.

C. Integration with Front-End

Flutter is a portable SDK (software development kit) developed by Google to create natively compiled applications for different platforms like mobile, web and desktop from a single codebase. It is an open-source program which gives developers more flexibility with their creations. The apps are written in Dart language which allows them to use advanced features. Here, it is used to develop a mobile application where the users can answer some questions, based on which the algorithm will predict if there are chances of ASD in child or not.

TensorFlow Lite, as the name suggests, is the lighter version of TensorFlow. It is a set of tools to help developers run machine learning models on mobile, embedded, and IoT devices. Apps developed on TensorFlow Lite have good performance and small binary size. It is used to integrate the model with application so that model can be saved and used on mobile devices.

V. DATASET

Dataset used for training the machine learning models is available publicly on UCI Machine Learning repository. Table 1 and 2 describes the dataset and its features.

TABLE I. Description of dataset

Name of dataset	Dataset characteristics	Number of attributes	Number of instances
Autistic Spectrum Disorder Screening Data for Children Data Set	Multivariate	21	292

TABLE II. Features of dataset and their description

Attribute	Description
Age	Age of child (in months)
Sex	Whether child is male or female
Ethnicity	Ethnicity of child
Born with jaundice	Whether the child had jaundice or not when born
Family member with PDD	Whether any family member is affected by similar disorder
Who is completing the test?	Parent, caregiver, self, medical staff, etc.
Country	Country of residence
Used the screening app	Whether user has used a screening app or not
Screening method type	The type of screening method chosen based on age category
Question 1 – 10 answers	The answer code of the question based on the screening method used
Screening score	The final score obtained on the screening algorithm

VI. LIMITATIONS

The ultimate goal is to develop a platform which can help the parents detect chances of ASD in their child based on their observation of the child's behavior and early symptoms being displayed by the child. Currently, the data present is still not sufficient to train reliable models which can make predictions that can be used for medical purposes. Future work solely depends on gathering more relevant data so that prediction model can be improved further. Feedback

provided by the users will be valuable for improving user experience such as usability and user interface of the mobile application. ASD can develop at a very young age and some parents might take it as if their child is just being a child, acting silly and they ignore the early signs. It can't be stressed more that the early detection of ASD can help massively in helping the child grow as close to normal as possible. So there is also a need to increase the awareness of people towards what ASD is.

VII. CONCLUSION AND FUTURE SCOPE

This work was an attempt to build a classifier using machine learning techniques to detect (op predict) ASD in children in the earlier stages so they can be provided with better care and treatment. Model was implemented on ASD Screening Data for Children dataset and accuracy was used as the evaluation metrics to determine the best model for prediction. After comparing, CNN classifier was found best suited for building the application which can help parents/guardians of a child to detect whether their child is suffering from ASD or not. As mentioned earlier, there is no medical way of determining if a child has ASD or not. So, this application only predicts the chances of ASD and recommends the user to visit an autism specialist for further treatment and help as soon as possible.

VIII. REFERENCES

[1] Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K., & Islam, M. N. (2019). A Machine Learning Approach to Predict Autism Spectrum Disorder. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). doi:10.1109/ecace.2019.8679454

[2] Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., & Linstead, E. (2019). Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review. Review Journal of Autism and Developmental Disorders. doi:10.1007/s40489-019-00158-x

[3] Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2018). Machine learning approach for early detection of autism by combining questionnaire and home video screening. Journal of the American Medical Informatics Association, 25(8), 1000–1007. https://doi.org/10.1093/jamia/ocy039

[4] F. Thabtah. Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017. ACM, 2017.

[5] F. Thabtah. UCI machine learning repository, 2017. [Online]. Available:

[6] Raj, S., & Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. Procedia Computer Science, 167, 994–1004. doi:10.1016/j.procs.2020.03.399

[7] Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. Journal of Child Psychology and Psychiatry, 57(8), 927–937. doi:10.1111/jcpp.12559

[8] F. Hauck and N. Kliewer. Machine learning for autism diagnostics: Applying support vector classification.

[9] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 16–23. doi:10.1016/j.nicl.2017.08.017

[10] Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. (2017). Applying machine learning to identify autistic adults using imitation: An exploratory study. PLOS ONE, 12(8), e0182652. https://doi.org/10.1371/journal.pone.0182652

[11] Wall, D. P., Dally, R., Luyster, R., Jung, J. Y., & DeLuca, T. F. (2012). Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. PLoS ONE, 7(8), e43855. https://doi.org/10.1371/journal.pone.0043855

[12] Linstead, E., German, R., Dixon, D., Granpeesheh, D., Novack, M., & Powell, A. (2015). An Application of Neural Networks to Predicting Mastery of Learning Outcomes in the Treatment of Autism Spectrum Disorder. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). doi:10.1109/icmla.2015.214

[13] Hyde, K., Griffiths, A.-J., Giannantonio, C., Hurley-Hanson, A., & Linstead, E. (2018). Predicting Employer Recruitment of Individuals with Autism Spectrum Disorders with Decision Trees. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). doi:10.1109/icmla.2018.00222

[14] Wiggins, L. D., Reynolds, A., Rice, C. E., Moody, E. J., Bernal, P., Blaskey, L., Rosenberg, S. A., Lee, L. C., & Levy, S. E. (2014). Using Standardized Diagnostic Instruments to Classify Children with Autism in the Study to Explore Early Development. Journal of Autism and Developmental Disorders, 45(5), 1271–1280. https://doi.org/10.1007/s10803-014-2287-3

[15] Wall, D. P., Kosmicki, J., DeLuca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. Translational Psychiatry, 2(4), e100. https://doi.org/10.1038/tp.2012.10

[16] Bennett, M., & Goodall, E. (2016). A Meta-analysis of DSM-5 Autism Diagnoses in Relation to DSM-IV and DSM-IV-TR. Review Journal of Autism and Developmental Disorders, 3(2), 119–124. https://doi.org/10.1007/s40489-016-0070-4