Machine Learning Based Model for Prediction of Autism Spectrum Disorder

Sandhya A Kulkarni [1]  
Asst Professor  
Department of Computer Science  
KSSEM  
Bangalore, India  
sandhya.a.kulkarni@kssem.edu.in

Dhanush Vasudevaraju [2]  
Student  
Department of Computer Science  
KSSEM  
Bangalore, India  
soccerhanush@gmail.com

C Sai Ranganath [5]  
Student  
Department of Computer Science  
KSSEM  
Bangalore, India  
challasairanganath@gmail.com

C Ramesh Chandra [3]  
Student  
Department of Computer Science  
KSSEM  
Bangalore, India  
rameshchalla2001@gmail.com

Dinesh M [4]  
Student  
Department of Computer Science  
KSSEM  
Bangalore, India  
dinesh96061@gmail.com

Abstract — Autism spectrum disorder is a neurodevelopmental disorder that affects a person's interaction, communication and learning skills. Although diagnosis of autism can be done at any age, its symptoms generally appear in the first two years of life and develop through time. Autism patients face different types of challenges such as difficulties with concentration, learning disabilities, mental health problems such as anxiety, depression, motor difficulties, sensory problems, and many others. Diagnosis of autism requires significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis. Thus, an efficient, accurate and easy screening test tool is very much required which would predict autism traits in an individual. The main idea behind this project is to detect autism spectrum disorder in an individual (male/female). This project is implemented by making use of a Machine Learning model using parameters such as an individual's age, gender, ethnicity, Autism Quotient Tool. The detection derived from this project will help an individual to get required diagnosis in time to prevent further complications of developing Alzheimer's disease.

Keywords—machine learning, autism spectrum disorder, parameters, autism quotient tool

I. INTRODUCTION

This paper focuses on proposing a model which would assist in prediction of ASD in an individual, so that diagnosis can be done and further treatments may be followed. The dataset used is the Autistic Spectrum Disorder Screening Data [3]. The datasets provide insights into various factors affecting the prediction of the disorder. Machine learning algorithms like decision tree, random forest, logistic regression, support vector classifier and artificial neural networks are used for finding out the optimal model for each dataset. Several performance metrics are used in order to analyze and compare each model from every angle possible.

II. RELATED WORK

Machine Learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Classical machine learning is often categorized by how an algorithm learns to become accurate in its predictions. There are four basic approaches:

A. Supervised learning

In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is used.

B. Unsupervised learning

This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.

C. Semi-supervised learning

This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.
D. Reinforcement learning

Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

To further describe machine learning, its life-cycle is described below:-

![Machine Learning Life-Cycle Diagram](image)

1. Gathering Data

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.

2. Data preparation

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

3. Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

4. Data Analysis

Now the cleaned and prepared data is passed on to the analysis step. This step involves:
- Selection of analytical techniques
- Building models
- Review the result

5. Train Model

In this step we train our model to improve its performance for better outcome of the problem. We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and features.

6. Test Model

In this step, we check for the accuracy of our model by providing a test data set to it. Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7. Deployment

The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system. If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not.

IV. THE PROPOSED METHOD

A. Random Forest Algorithm

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. The existing base papers make use of this algorithm.


The objective of this work is to propose an autism prediction model using ML techniques and to develop a mobile application that could effectively predict autism traits of an individual of any age. In other words, this work focuses on developing an autism screening application for predicting the ASD traits among people of age groups 4-11 years, 12-17 years and for people of age 18 and more.


This paper uses ASD screening dataset for analysis and prediction of probable cases in adults, children and adolescents. The dataset for each of the age groups are analyzed and inferences are drawn from them. Machine learning algorithms like Random Forest.

B. Support Vector Machines

Support Vector Machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. It is formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. The existing base papers make use of this algorithm.


Various classification techniques were implemented with these transformed ASD datasets and assessed for their performance. SVM showed the best performance. The results of these analytical approaches indicate that, when appropriately optimised, machine learning methods can provide good predictions of ASD status. This suggests that it may possible to apply these models for the detection of ASD in its early stages.
Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework [17]

In this paper, objective imaging-based biomarker discovery for psychiatric conditions is critical for accurate diagnosis and treatment. Using a machine learning framework, this work investigated the utility of brain’s functional network topology extracted from functional magnetic resonance imaging (fMRI) functional connectivity (FC) as a viable biomarker of autism spectrum disorder (ASD). A battery of complex network features were computed from the FC network using graph theoretic techniques. Recursive-Cluster- Elimination Support Vector Machine algorithm was employed to compare the predictive performance of three independent feature sets, (i) FC, (ii) complex network measures, and (iii) both combined. The study found that FC could diagnose ASD with 67.3% accuracy and graph measures with 64.5% accuracy, while the combined feature set could diagnose with 70.1% accuracy (all accuracies were significantly different, p < 10^-8).

C. K-Nearest Neighbour

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another. The existing base papers make use of this algorithm.


In this paper, the classification method for ASD diagnosis was used in children aged 4-11 years. The K-Nearest Neighbor (KNN) algorithm is used for classification. To test the algorithms, 30 percent of the data set was selected as test data and 70 percent as training data. As a result of the work done, the accuracy of the KNN algorithm is 85.5%. For the For KNN algorithm, these values are calculated as 0.9762 and 0.80. F-measure values are calculated as 0.8913 for the KNN algorithm.


The objective of the research is to foresee Autism Spectrum Disorder (ASD) in toddlers with the help of machine learning algorithms. Of late, machine learning algorithms play vital role to improve diagnostic timing and accuracy. This research work precisely compares and highlights the effectiveness of the feature selection algorithms K-Nearest Neighbor (KNN), and to improve the efficiency of Random Tree classification algorithm while modelling ASD prediction in toddlers. And to improve the efficiency of Random Tree classification algorithm while modelling ASD prediction in toddlers. Analysis results uncover that the Random Tree dependent on highlights chosen by Extra Tree calculation beat the individual methodologies.

V. CONCLUSION

This research comes under machine learning which is an advanced technique at present. This research provides a comparative view among different ML approach in terms of their performance. The results showed that Random Forest algorithm provides highest accuracy compared to other machine learning algorithms such as Support Vector Machines and K-Nearest Neighbour. Different existing data mining procedures and its application were considered or explored. Utilization of machine learning algorithms was connected in various medical data sets. Machine learning strategies have diverse power in different medical data sets. Previously mentioned conventional machine learning techniques gave less exact outcome and results additionally shifts in light of the procedures has been utilized for the prediction. The existing system cannot properly represent the diversity of autism classes with complex intra-class variability and inter-class similarity and it is less effective for autism classification.

REFERENCES


Informatics, Electronics and Vision (ICIEV), doi:10.1109/iciev.2016.7760073


