Healthcare Analytics 4 (2023) 100211

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

A Multi-Classifier-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning

Anita Vikram Shinde [a,b], Dipti Durgesh Patil [c,*]

[a] Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India
[b] Department of Computer Engineering, Marathwada MitraMandal's College of Engineering, Pune, India
[c] Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Pune, Savitribai Phule Pune University, India

ARTICLE INFO

Keywords:
Machine learning
Autism spectrum disorder
Recommender system
Multi-classifier approach
Artificial Neural Network
Personalized recommendations

ABSTRACT

Efficient and effective medical diagnostic systems are needed for Autism Spectrum Disorder (ASD) detection and treatment. Healthcare specialists generates extensive remarks on patient behavioural assessment, which is time-consuming to process and record. Early detection of ASD means quality life with the help of appropriate treatment and care. Machine learning models can be utilized to investigate the feasibility of identifying the stated features and evaluating the presence or absence of autism. This study develops a recommender model with multi-classifiers to enhance precision in the prediction of ASD. Various machine learning algorithms are experimented to assess the model's performance. We show that Decision Trees and Random Forests exhibit improved performance if analyzed with other algorithms with respect to accuracy, precision, recall, and F1-score as evaluation metrics.

1. Introduction

Autism spectrum disorder (ASD), a kind of neurological disorder that is observed in children which mainly impacts their social behaviour and communication skills. It is detected at early childhood as critical mental disorder. It persists for their rest of life. Across the globe more that 1% of citizens facing this autism disorder [1,2]. The major research is necessary for early diagnosis of ASD so that best medical recommendations can be provided for better health of patient [3]. ASD mainly impacts on social interactions, communication and leads to recurring behaviour by patients. In many developed countries, autism in children is detected too late. Autism is the inadequate growth of patient's brain that obstructs their social interaction and behaviour. These children face problems in understanding and learning accurately and mainly find trouble in communication and interaction [4]. Autism disorder can appear in any national, cultural and tribal community but specifically boys are facing this problem than girls. It is feasible to enhance their communication and behavioural skills by applying suitable medication and sufficient training [5].

Patients with autism spectrum disorder (ASD) exhibit repetitive behaviour and social instability due to a neurodegenerative condition. Autism is caused by improper connections between brain cells, which results in abnormalities in brain structure or function. Autism Spectrum Disorder (ASD) affects people differently in terms of their symptoms and skills ASD symptoms mostly notice in the initial years. Three categories of autism have been established: severe autism, moderate autism, and mild autism.

ASD can be referred as dysfunction that can be made better but cannot be treated completely. In daily life, ASD patients need facilitator to assist them. Early diagnosis and treatment helps to improve their abilities thought it is lifetime disorder. Some probable reasons causing ASD includes premature delivery, ASD siblings etc.

Autism is a mental condition/ disorder that create hindrances such as deficient in communication skills, interaction in social life and excellent motor skills. ASD patients find it difficult to interact with parents, friends, caretakers and teachers. They experience problem in having eye contact with individuals that they interact. The main observed symptoms contain problem in uttering a word, exact daily routine execution (i.e. difficult to change habits), unaware about risk, no reply to name calling, recurring behaviour with hand or head, sensitive to sound, want to live alone etc. Paediatricians assess child in regular follow up visits and in case of existence of any autism symptoms, the expert opinion is suggested. Fig. 1 shows traditional process for ASD detection which is time consuming.

Due to significant rise in the rate of autism spectrum disorder in kids throughout world, there is demand of prior detection of ASD in children. Parents look for expertise help for better diagnosis of ASD at early age.

The psychologists investigate the child behaviour and utilize different screening tools such as Parents' Evaluation of Developmental

* Corresponding author.
E-mail addresses: anitavshinde8@gmail.com (A.V. Shinde), dipti.patil@cumminscollege.in (D.D. Patil).

https://doi.org/10.1016/j.health.2023.100211
Received 2 February 2023; Received in revised form 16 April 2023; Accepted 14 June 2023

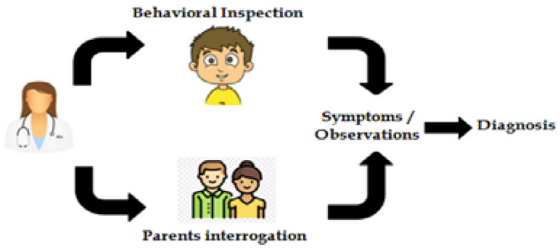

Fig. 1. Traditional method of ASD detection.

Status (PEDS), and Ages & Stages Questionnaires (ASQ), Communication and Symbolic Behaviour Scales (CSBS), Screening Tool for Autism in Toddlers and Young Children (STAT) to primarily verify autism symptoms. In case of positive test results, various diagnosis tools like Autism Diagnostic Observation Schedule (ADOS), Childhood Autism Rating Scale (CARS) are used for child evaluations. Early intervention program and particular education can be suggested by psychologist. It actually stands as a very critical part in life of patient to build essential expertise and defeat autism symptoms [6].

Because there is no blood or other medical test, diagnosing ASD is reliant on ASD historical record. ASD diagnosis has drawn a lot of attention. For fast diagnosis of autistic patients, early detection of this disorder required. These days, with remarkable development in Machine Learning, interdisciplinary researches are carried out in various domains such as medical (e.g. Disease diagnosis, medical imaging diagnosis, and medicine recommendation), military etc [7] to help society. Machine Learning techniques have emerged with various real time scenarios to assist artificial intelligence at human being level. With Machine Learning, building model for prediction and fastening the time consuming diagnosis with suitable treatments is achievable.

The goal of the proposed research work is to efficiently categorize patient (child, adolescent or adult) as yes or no for existence of autism spectrum disorder using multi-classifier based recommender system. The paper is organized into different sections as literature survey on existing algorithms used in ASD, Machine Learning methods used, research methodology, experimental results and conclusion with pointers to future enhancements.

2. Literature survey

2.1. Machine learning algorithms in autism

The machine learning techniques have been applied widely in diverse aspects of investigating and detecting autism spectrum disorder. Different Machine learning methods such as Support Vector Machine (SVM), Logistic Regression (LR), Ridge are used to investigate ASD sufferers by evaluating rs- functional magnetic resonance imaging (fMRI) data ABIDE data repository [8]. Classification between ASD and typical developing (TD) is done using Eigen values on fMRI and defining 264 features of human brain [9] with machine learning algorithms like K-Nearest Neighbour (KNN), SVM and LR.

Unsupervised learning algorithm neural network with auto encoders and decoders is employed for feature extraction and K-Means clustering is used detect correlation in gaze behaviour of the ASD victims [10]. Childhood Autism Rating Scale (CARS) is used as a standard method for comparison. In order to speed up autism diagnosis, gaze-based Arabic screening is integrated with intelligent method i.e. machine learning algorithm SVM in order to efficiently categorize ASD and non-ASD patients [11].

Region of interest (ROI) is separated with help of double- density dual-tree discrete wavelet transform (D3TDWT) as a first step and in second step, features are extracted using generalized autoregressive conditional hetero-scedasticity (GARCH). t-test is used for feature selection and classification is done with Support vector machines (SVM) [12]. The game has been constructed in order to decide the stages of autism varying from low to severe with help of fuzzy logic [13]. Unsupervised learning K-means cluster study [14] is used in order to investing challenging behaviour in ASD patients. Dominant challenging behaviour is recognized in maximum ASD patients. Machine learning Association Rule with minimum Redundancy Maximum Relevance (mRMR) method [15] is proposed to discover ASD patients. To support diagnosis prediction in decision making, Mutual Information Difference (MID) method is incorporated for identifying additional symptoms. Summary of Machine Learning algorithms used in ASD detection is presented in Table 1.

Questionnaires (based on ADI-R, ADOS) filled by parents and home videos [16] are used as machine learning autism screener. Random forest is used on the combination of questionnaires and home video screener. To address Heterogeneity in Autism Spectrum Disorder (ASD) [17], k-dimensional clustering along with univariate and multivariate analysis is described on ADI-R and ADOS information. Various Machine learning models along with computational linguistics [18] are applied on TalkBank datasets namely Eigsti & Nadig using Random Forest and Logistic regression to accurately discover ASD patients.

A comprehensive study of various investigation methods on exhaled breath in order to diagnose diseases of human being is presented [19]. It describes the biomarkers used for clinical assessment and diagnosis. A framework using multimodal approach that capturing facial characteristics, speech features and brain signals for depression analysis is described with enhanced prediction performance [20]. A novel method for prediction of depression using EEG signals of subject is proposed in [21] using discrete Fourier transform and fast Fourier transform.

In [22], the approaches are presented to discover concept drift in data stream and listed summary of available approaches for the concept drift issue. It mainly aimed on healthcare domain. Depression analysis using speech signal features is studied in [23] that mainly targeted on adolescence speech. Existing classification techniques for depression analysis are reviewed. Table 2 depicts Literature review of Machine Learning Algorithms in Autism Spectrum Disorder.

The study [24] proposed method to differentiate autistic and typically developed child using visual behaviour with eye tracking mechanism in virtual environment with 86% accuracy and 91% sensitivity. Autism detection using functional connectivity via selecting and interpreting input features was proposed in [25]. Support Vector Machine (SVM) trained on EEG was used as diagnosis tool for ASD. Federated learning was explained to extract high level features and to detect various neurological disorders [26] such as autism, schizophrenia, using functional magnetic resonance imaging. For automatic identification of different neurological disorders using EEG data, Computer aided diagnosis is introduced using convolutional neural network to help experts and doctors [27]. The Autism detection method [28] is proposed with Facial expressions as biomarker using deep learning CNN, Recurrent Neural network RNN and SVM. Federated learning (FL) model [29] was proposed for diagnosing ASD as neurological disorder using special behavioural traits, and the facial traits of patients.

Comparative analysis [30] of various machine learning algorithms namely KNN, LR, SVM and Naïve Bayes for autism prediction was experimented with highest accuracy for Naïve Bayes as 99%. Eye tracking method and machine learning algorithms [31]are combined to detect autism in high functioning adults using data acquired from web related task such as search, browse etc. Autism spectrum disorder analysis in children using hand gesture detection is explained in [32] using cosine similarity to examine gap between two fingers. Assistance system using Deep learning and Internet of Things for autistic patients proposed in [33] by examining stress factors such as facial expressions, pulse rate, body heat, skin conductance etc monitored remotely by caretakers. Eye tracking technology and machine learning are combined to diagnose autism in children [34]. The content centric prediction model [35] is built for ASD detection using various machine learning methods.

Table 1
Summary of Machine Learning Algorithms used in Autism Spectrum Disorder (ASD) detection survey.

Reference	Method	Dataset	Algorithms/ Statistical Techniques	Accuracy	Limitations/ Future scope
[8]	Resting-state functional magnetic resonance imaging (rs-fMRI)	ABIDE	SVM Logistic regression, ridge	71.40% 71.79% 71.98%.	Integration of ML classifiers with other ASD clinical features for accurate results
[9]	Functional magnetic resonance imaging (fMRI).	ABIDE	Linear Discriminant Analysis (LDA) KNN, SVM, LR	77.7% 73.7% 75.5% 76.6%	Use reinforcement and deep learning for better classification model
[10]	Clustering of eye-tracking scan path Figshare data repository	59 children data from French school	K-Means clustering	–	–
[11]	Remote eye tracking	Participant information	SVM	88.6%	Use of larger dataset to validate algorithm performance
[12]	Resting-state functional magnetic resonance imaging (rs-fMRI)	ABIDE	SVM	Male71.6% Female 93.7% as ASD	–
[13]	Human computer interaction	50 autistic children	Fuzzy logic	85%	Levels can be increased, more input parameters can be added
[14]	Cluster investigation	SKILL database of autism treatment services	K-Means algorithm	–	Incorporating functional component of challenging behaviour
[15]	Machine learning Association Rule (AR) with minimum Redundancy-Maximum-Relevance (mRMR)	Autism Therapy Counselling and Help (CATCH)	Association rule with min support and max confidence	83%	Time reduction in symptoms identification and accurate prediction
[16]	Screening tool (questionnaires and home videos)	ADI-R & ADOS	Random Forest	–	Use of screening tool for situation beyond autism
[17]	Ensemble model	208 ASD subjects Simons Simplex Collection (SSC)	k- dimensional Clustering	–	Use in other disorders with heterogeneity
[18]	Speech Transcripts analysis	TalkBank Eigsti & Nadig dataset	Logistic Regression & Random Forest	75%	conversation with Chat bots or robot assistant to be used

Machine learning algorithms such as SVM, NB and RF as classification methods are utilized to predict autism spectrum disorder that helps to develop mental health of child. Swarm intelligence based on binary firefly feature selection wrapper is utilized for ASD diagnosis on dataset available at UCI machine learning repository and obtained 97.95% of accuracy.

Sensors and AI system are integrated to fetch facial expressions and emotions as patient's state. It helps ASD patients in COVID-19 pandemic by providing alerts at regular interval. Machine learning approach referred Rule Machine Learning is proposed to diagnose ASD with generated rules used to recognize causes behind classification.

2.2. Severity stages of autism spectrum disorder (asd)

Autism is a neuro-developmental disorder that is not completely curable. It is about social interaction and social communication in various contexts. Early intervention will help to enhance the condition of the children that needs early identification of this disorder.

The American Psychiatric Association has made available the fifth edition of its Diagnostic and Statistical Manual of Mental Disorders (DSM-5). It is used as reference by healthcare suppliers in order to identify cerebral and behavioural conditions like autism. Table 3 describes severity stages of diagnostic criteria for ASD. The severity diagnostic criteria is segregated into three stages such as Stage 1– It is repetitive and limited behaviour that needs help, Stage 2 – Social interaction and conversation which involves significant help and Stage 3 – The most severe stage which needs high concentration and assistance.

2.3. ASD screening approaches

Since last more than 25 years, various assessment methods have been introduced to assist ASD detection at earlier stages of disorder In order to identify ASD at earlier stages, neuroscientists and psychologists have discovered different techniques and approaches. The maximum of these tools are developed to diagnose ASD patients using screening techniques.

Autism Behaviour Checklist (ABC): ABC is one of the strategies that aim on identifying ASD patients at earlier stage. This checklist evaluates the present of autism symptoms with help of caregivers in different situations and circumstances. It contains question list designed to evaluate patient's health.

Child Behaviour Checklist (CBCL): This is one of the most commonly employed tests in child psychology for evaluating emotional problems and abnormal behaviour. This questionnaire aims on habits like fear, over-control and hyperactive.

Social Communication Questionnaire (SCQ): This is fast, simple and inexpensive way to screen for autism spectrum disorders. This screening approach includes questionnaire that measures the severity of autism spectrum disorder over the age of 4 with a mental age over 2 years.

Table 2
Literature review of machine learning algorithms in Autism spectrum disorder.

Authors	Method	Dataset	Algorithms/ Statistical Techniques	Accuracy	Limitations/ Future scope
K. Vijayalakshmi et al. (2020)	Multi - classifier based regression (MCR)	ABIDE- I	Random Forest, Naïve Bayes, Logistic Regression	98% child dataset	Use of Version II dataset
Khozaei A et al.,(2020)	Cry-based screening	GARS-2	SubSet Instance (SSI) classifier	92.85%	Automating pre-processing work
Sherkatghanad et al. (2020)	functional magnetic resonance imaging (fMRI)	ABIDE- I	Convolutional Neural Networks (CNN)	70.22%	Work with more data to build robust model
Zhong Zhao et al. (2019)	Restricted kinematic features (RKF) identification	43 participants * 18 (features) matrix.	SVM, LDA, DT,RF, and KNN	Highest with KNN 88.37%	Implementation of ASD from ADHD
M. S. Satu et al. (2019)	Classification using Rule mining	Autism Barta app with	J48 Decision tree	98.44%	Different age limits data for further detection to be used
S. B. Shuvo et al. (2019)	Classification	ABIDE- I Adult dataset	Random Forest classifier	96%	Setting parameters of RF to get consistent results
K. S. Omar et al. (2019)	Decision tree	ABIDE- I Child, Adolescent, Adult Dataset	RF-CART RF-ID3	92.26%, 93.78%, 97.10%	Data collection from different sources to enhance accuracy of ML classifiers
O. Altay et al. (2019)	Classification	ABIDE- I Child dataset	Linear Discriminant Analysis (LDA) K-Nearest Neighbour (KNN)	90.8% 88.5%	–
A. S. Halibas et al. (2018)	Classification	ABIDE- I Child, Adolescent, Adult Dataset	Decision Tree, Naïve Bayes, k-nn, Random Tree, Deep Learning	85.87% 90.30% 88.89% 72.74% 96.38%	Use of other ML algorithms Work with More training data
W. Liu et al. (2018)	Eye movement analysis	Child dataset adolescents and young adults	Support vector machine (SVM)	92%	–
Wan G et al. (2019)	Eye tracking	Child dataset 37 ASD , 37 TP	Support vector machine (SVM)	85.1%,	Work on larger sample sizes, different age patients
LamyaaSadouk et al. (2018)	Deep learning	SMM dataset	Convolutional Neural Networks (CNN)	–	–
Duda M et al. (2017)	Classification	Survey data from parents	ElasticNet ,LDA	–	Incorporating novel data points Making a classifier more generalized
Gro“zekath ́ofer U et al. (2017)	Classification	Child (age 12 to 20) data from school	Decision Tree Support Vector Machine Random Forest	83% 86% >86%	Large dataset with different age range and gender
Nastaran et al.,(2017)	Deep learning	Simulated data Child (age 12 to 20) data from school	Convolutional Neural Networks (CNN)	–	Alteration of the system in an unsupervised manner to handle unlabelled data
M. F. Rabbi et al. (2021)	Image Classification	2940 Face images dataset from Kaggle	Multilayer Perceptron , Random Forest, Gradient Boosting Machine, AdaBoost Convolutional Neural Network	71.66% 72.78%, 75.23%, 74.56%, 92.31%	–

Table 3
Diagnostic criteria severity levels for ASD (autism spectrum disorder)
Source: Taken from (dsm-5) report.

Stage 1 Recurring and limited	Stage 2 Social Interaction	Stage 3 Major severe stage
➢Monotonous and limited behaviour	➢ Difficulty in admitting change simply	➢Main significant deficit in communication i.ie verbal & nonverbal.
➢Rigid in behaviour	➢Verbal and nonverbal social conversation ability	➢ Inadequate social relations
➢Trouble in commencing social relations	➢ Limited social interaction with no open up	➢ Less reply to social start by people

Autism Spectrum Quotient (AQ): It is most commonly used scales for assessing autistics features. AQ is the test used to evaluate how autism behaviour noticeable in patients with moderate intelligence

Childhood Autism Rating Scale (CARS-2): It is 15 elements rating scale used to discover autism children and differentiating them from patients with developmental disabilities.

Screening Tool for Autism in Toddlers and Young Children (STAT): This is an interactive measure designed to screen for ASD in patients between 24 to 36 months of age. It is developed for community service provider.

Quantitative Checklist for Autism in Toddlers (Q-CHAT-10): It contains 10 questions with answers mapped to binary values. The values

Table 4
Technology used in Autism detection.

Technology	Function
Augmented Reality	Facial expressions and emotions are used with non-verbal facial signs storybooks
Virtual Reality	Real world is presented to autistic children with help of Virtual Reality. Simulation is used for crafts work and understanding behaviour in social situations
Mobile learning	Various Activities and exercises are used to enhance ABA (Applied Behavioural Analysis) using behavioural model
Social Robots	Social robot and human robot are used by therapists for communication, Helping children with less involvement
Serious Gaming	Learning with games for vocabulary in autistic children
Edutainment	Graphics material and objects are used to enhance communication ability, assisting autistic children
Virtual learning environment	Digital social stories are created, virtual environment is used for social interaction.

are received at data gathering process by giving answers to questionnaire.

2.4. Technology used in autism

There is person to person variation in intensity and amalgamation of the symptoms in autism. Therefore, parents/ guardians can do timely detection if the kid is facing conversation problems, complexity in relating to people, events and recurring behaviours difficulty. Now days, various nations have paid considerable awareness to autism in community area and various financial supports have been made available to promote research and the public facilities for ASD patients.

Technology can be used as bridge to maintain expertise and capability improvement. It is utilized for ASD patients to target on conversation with various mediums. Similarly, they have very little focus duration hence usually fight for concentration. Different tools with audio and visual simulations are utilized to maintain them involved in learning and preserving their focus. They are organized to evaluate their behaviour, social interaction. Table 4 lists down various technologies incorporated in Autism detection. It mainly includes augmented reality, Virtual reality, Mobile learning, virtual learning etc.

2.5. Treatment therapies for autism

There are various therapies that can be used to help autistic patients to enhance their skills and resolve their symptoms. Initiating therapy early increases the possibility of patient's success in terms of better health. Some of therapies are as:

Socially Assistive Robotics (SAR):SAR is robot that assists the child in their social interaction. It is one of the tools that teaches or shows socially useful behaviour to help children in expressing themselves to others due to weak communication and social skills as an effect of ASD. It helps children in social, emotional and communication deficits

Nutrition Therapy: Balanced diet supplies various different nutrients essential for the proper body functioning. Assimilation of vitamins, minerals and necessary fatty acids from the diet is very important. It facilitates to take care of symptoms linked to autism with help of fair and stabilized diet.

Occupational Therapy: It helps autistic patients to increase their physical, sensory, cognitive and motor skills. It boosts their confidence and attainment level. It helps to improve activities of daily living and the work with everyday objects. In occupational therapy, patient's requirements and objectives are targeted.

Speech Therapy: Speech therapy is a critical part of treatment for autism. It is helpful in resolving communication and speech issues. It contains non-verbal skills, taking part in conversation and recognizing gestures. Children can make use of sign language or images to communicate. It enhances the clarity in sound production

3. Machine learning models used

For performing our experimental analysis, we have employed several machine learning models. For selecting model, data pre-processing, over-fitting and tuning are taken into consideration. Data pre-processing includes cleaning, handling missing values and encoding of categorical attributes to numerical data. Machine learning models used are as:

a. **Random Forest (RF):** While handling binary classification, decision tree is commonly used classification method. The main drawback of decision tree is over-fitting which can be handled by employing Random Forest. Ensemble of trees has led to considerable enhancement in accuracy of classification and permitting to vote for most accepted class. Random forest creates forest as a collection of many decision trees. Random forest helps to improve prediction accuracy by sampling of dataset [36]. With voting, top scored tree can be chosen from forest. It also handles over-fitting. Random forest works as follows:

1. Choose random N data samples from the training dataset
2. Construct the decision tree using chosen data samples
3. Select the number of decision trees to be built.
4. On arrival of new data sample, Classifier predicts final decision using majority of votes given by decision constructed trees.

b. **Naïve Bayes (NB):** Bayes theorem principles are the foundation of Naïve Bayes classifier. It assumes class conditional independence rule amongst features while predicting class label hence eases learning process.

$$P(A|B) = P(B|A) * P(A)/P(B) \quad (1)$$

It is considered to be efficient predictive method with better performance in prediction for dataset containing lost values [37].

c. **Logistic Regression (LR):** Logistic regression is a predictive algorithm mainly incorporated while handling relationship between binary class label and one or more predictor variables. It is one of the statistical methods used to work with predictive problems considering probability of an event [38]. It performs classification of categorical based class label using independent variables.

d. **Decision tree (DT):** Decision tree is a supervised machine learning algorithm used for classification examples [39]. It is actually a flow-chart where every internal node represents check on an attribute,

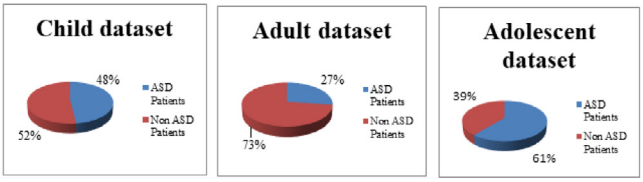

Fig. 2. Distribution of data within classes in three datasets.

each branch shows the output of that check and every leaf node represents a class label which is final decision taken. Decision tree algorithm is as follows:

1. Start tree with the root node which includes the whole dataset
2. Identify the best attribute in the dataset with attribute selection measures
3. Split the dataset into subsets.
4. Create the decision tree node using the best attribute
5. Repeatedly construct new decision trees with the subsets created. Repeat above steps until no more classification of node can be done.
   Final nodes will be leaf nodes holding decision.

e. **Support Vector Machine (SVM):** SVM is generally used in classification problems. It is a high-speed classification machine learning algorithm with excellent accuracy. In SVM, we find hyper-plane in N-dimensional space which categorizes the samples into binary classes in the finest feasible way. The distance of the closest training samples from the hyper-plane is called margin. SVM finds the optimal separating hyper-plane in such a way that the hyper-plane has the maximum margin [40].

f. **Artificial Neural Network (ANN):** ANN is neural network having connection with multiple neurons [41]. Each neuron cell has input values and related weights. This network mainly has three main layers namely, input layer, hidden layer and last is output layer. There is no loop or cycle in this network.

4. Dataset used in experiment

The dataset used in research purpose can be used from UCI Machine Learning Repository made available by Fadi Fayez Thabtah. It is used for various classification and regression models. ASD screening child dataset has 292 (141 ASD:151 non-ASD) records [42], Adolescent Dataset has 104 (63 ASD:41 non-ASD) records and Adult Dataset has 704(189 ASD:515 non-ASD) records along with 21 properties in which 10 questions (AQ1-AQ10) are about behavioural information and 10 demographic attributes with one class label as either yes or no. Table 5 shows questions and their values in dataset. Fig. 2 shows distribution of data within classes in three datsets.

Research Methodology: The steps in Autism detection solution are shown in Fig. 3.

The autism dataset is selected for experimentation purpose. Data pre-processing is carried out on raw dataset using feature engineering. Recommender model is constructed by applying various classification algorithms on refined dataset. The flow of the proposed system architecture is depicted in Fig. 4.

Performance of each classifier is evaluated based on ASD prediction done using various metrics such as accuracy, precision, recall and f1-score. Finally, recommendations are generated to help patients in decision support systems.

Recommender systems [44] are broadly classified into 3 categories namely content based, collaborative filtering and hybrid method. Content based method uses the history of active user to generate recommendations. Collaborative filtering incorporates only similar users for generating recommendations. Hybrid method takes advantages of both methods. Recommendations are generated using collaborative filtering [45] method which is based on similarity amongst the patients. After identifying similar patients, the suggestions given to those patients are ranked and top recommendations are given to active user. In autism context, recommendations can be therapy, exercise, diet, social organizations etc. It will help them in enjoying healthy life style like typically developed human beings.

Table 5
Questions in dataset [43].

Questions	Values
Able to hear sound that others can't hear.	(0–1)
Targets on large picture than small details.	(0–1)
Pursue the discussions of people in a community group	(0–1)
Can effectively switch between available actions	(0–1)
No idea about how to chitchat with peers.	(0–1)
Fine for day-today small chats.	(0–1)
Finding hard to understand emotions while reading a book	(0–1)
Fond of playing role plays as part of pre-school education.	(0–1)
Recognize their experience by observing facial expressions	(0–1)
Hard to create new friends	(0–1)
Age	Nominal
Gender	(1, 2)
Background('White-European', 'South Asian', 'Asian', 'Middle Eastern ', 'Pasifika', 'Hispanic', 'Turkish','Latino', 'Black', 'Others', 'Unknown')	(1, 11)
Born with jaundice (yes, no)	(1, 2)
Family member with PDD (yes, no)	(1, 2)
Country	(1, 52)
Familiar with the screening app (yes, no)	(1, 2)
Score	(0, 10)
Age_desc	Nominal
Who has attempted the test	(1, 5)
Class/ASD	(0–1)

Results and Discussions:

This section relates to the outcome and discusses the hardware and software requirements needed to run the model effectively. Operating system: 64-bit Windows 7. Processing speed: In the Google Colaboratory web application environment, the pandas tool was used using an Intel(R) Core(TM) i5-2450M CPU running at 2.50 GHz, 2501 Mhz, on two cores, and 4 GB of RAM. On the aforementioned dataset, the proposed work applies machine learning algorithms to improve prediction quality. To evaluate the performance of the model and make comments on the best approach, the analysis of experiments is done using evaluation metrics including confusion matrix, accuracy, precision, recall, and F1-score.

Evaluation Metrics: Confusion matrix is a table with real and predicted values as dimensions as shown in Table 6.

Relevance of the model prediction is assessed in terms of Precision as an evaluation metric whereas recall treated as true positive rate and sensitivity that tells proportion of correctly classified samples.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-score is test's accuracy. It is given by harmonic mean of precision and recall.

$$F1 - Score = \frac{2(Recall * Precision)}{Recall + Precision} \tag{4}$$

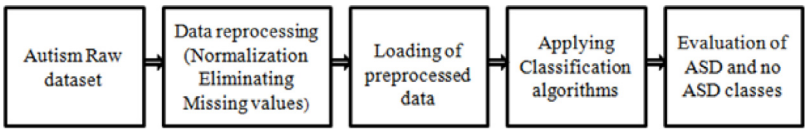

Fig. 3. Steps in autism detection solution.

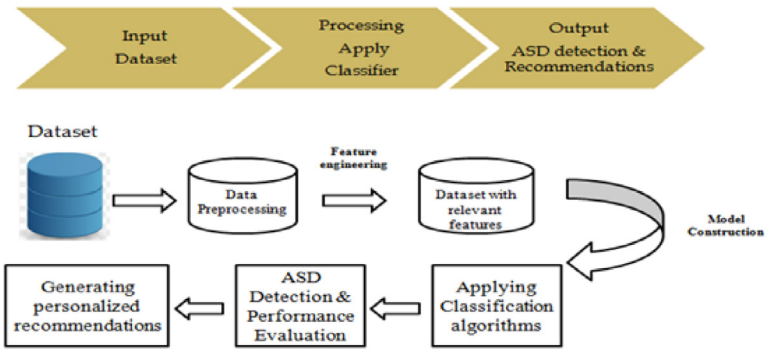

Fig. 4. Proposed system architecture.

Table 6
Confusion matrix for ASD prediction.

Predicted values	Actual values	
	Patient suffers from ASD	Patient does not suffer from ASD
Prediction as ASD is present	True Positive(TP)	False Positive (FP)
Prediction as ASD is not present	False Negative (FN)	True Negative (TN)

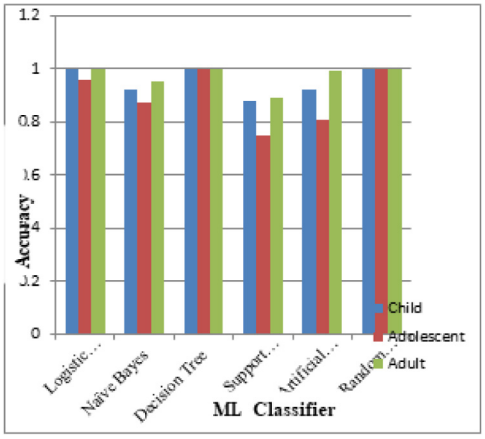

Fig. 5. Accuracy graph of classifiers.

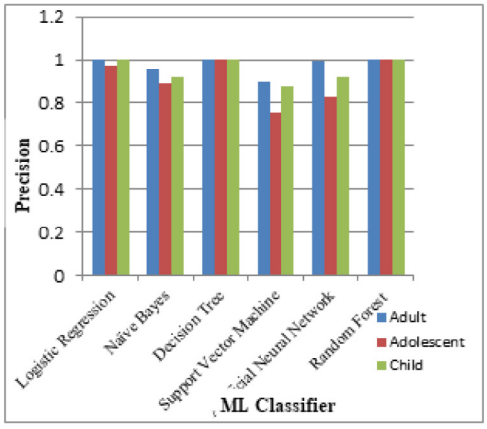

Fig. 6. Precision graph of classifiers.

The performance of various machine learning algorithms is compared as shown in Fig. 5, 6, 7 and 8 for metrics accuracy, precision, recall and F1-score respectively. It has been observed that Decision tree and Random Forest has shown better performance (see Table 7).

The three datasets are split as 70% training data and 30% testing data Different Machine learning classifiers such as LR (Logistic Regression), SVM(Support Vector Machine), RF(Random Forest), KNN(K-Nearest Neighbour), XGBoost are implemented in and various evaluation parameters such as accuracy, precision, recall and F1-score are measured.

In child dataset, the accuracy ranges from 97.59% to 100% as shown in Fig. 9 and best prediction performance is shown by LR, Adaboost and XBG. In case of Adolescent dataset, the accuracy ranges 89.25% to 100% and best prediction performance is shown by LR with 100% precision, recall and F1-score as depicted in Fig. 8. In Adult dataset, accuracy ranges from 96.45% to 100%. As shown in Fig. 11 best prediction performance is given by Logistic Regression. Fig. 9, 10 and 11 shows comparison graphs of various evaluation parameters on child dataset, Adolescent and Adult dataset respectively. Logistic Regression gives highest accuracy only for adult dataset. Logistic Regression results are compared with DT and RF. It has been observed that highest accuracy is achieved in Decision Tree and Random Forest. Multi-classifier approach is constructed by integrating results obtained by using various machine learning algorithms to enhance accuracy in prediction.

5. Conclusion

Autism is one of the highest raising developmental and neurological disorders in children therefore the investigation on its early diagnosis along with the classification models will definitely assist to a larger scope while predicting accurate assessment. The machine learning algorithms are widely incorporated in various domains due to accurate

Table 7
Comparison of result with different algorithms using evaluation metrics.

Dataset	Logistic regression				Naïve Bayes				Decision tree			
	Accuracy	Precision	Recall	f1 score	Accuracy	Precision	Recall	f1 score	Accuracy	Precision	Recall	f1 score
Adult	1	1	1	1	0.95	0.96	0.96	0.96	1	1	1	1
Adolescent	0.96	0.97	0.97	0.97	0.875	0.89	0.88	0.88	1	1	1	1
Child	1	1	1	1	0.92	0.92	0.92	0.92	1	1	1	1
Dataset	Support vector machine				Random forest				Artificial neural network			
	Accuracy	Precision	Recall	f1 score	Accuracy	Precision	Recall	f1 score	Accuracy	Precision	Recall	f1 score
Adult	0.89	0.9	0.89	0.88	1	1	1	1	0.98	0.99	0.99	0.99
Adolescent	0.75	0.76	0.75	0.75	1	1	1	1	0.81	0.83	0.81	0.82
Child	0.875	0.88	0.88	0.87	1	1	1	1	0.92	0.92	0.92	0.92

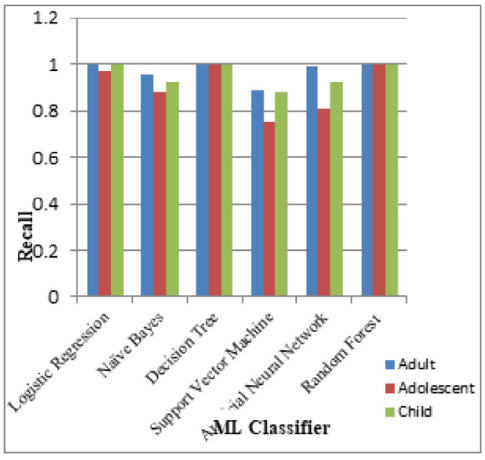

Fig. 7. Recall graph of classifiers.

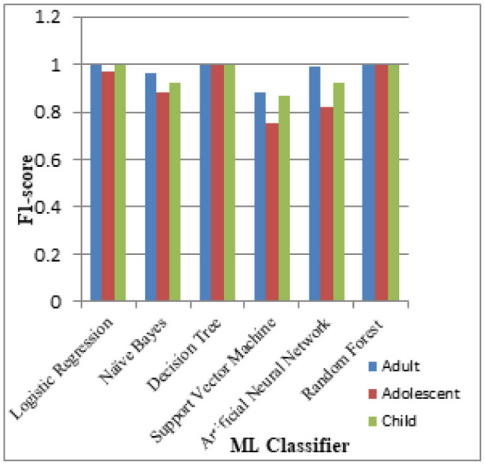

Fig. 8. F1-score graph of classifiers.

and précised results. The proposed work has focused on building the classification models with machine learning algorithms on available dataset. The performance is evaluated with help metrics such as accuracy, precision, recall & f1-score. The better performance is given by Decision tree and Random Forest if compared with other methods for ASD data categorization. It is observed that the early detection of ASD is absolutely achievable. The size of dataset is playing important role here because if complete and large amount of data is available, higher accuracy of ASD detection by machine learning algorithms will be attained. Hence, ASD can be efficiently treated with improved life quality of ASD patients and their family.

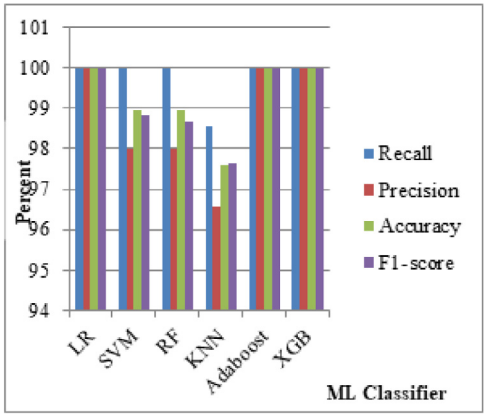

Fig. 9. Evaluation metrics graph for child dataset.

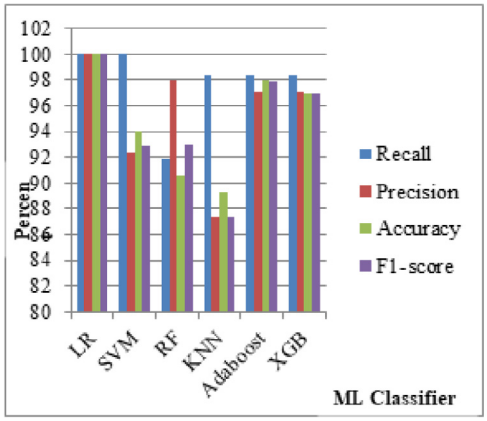

Fig. 10. Evaluation metrics graph for adolescent dataset.

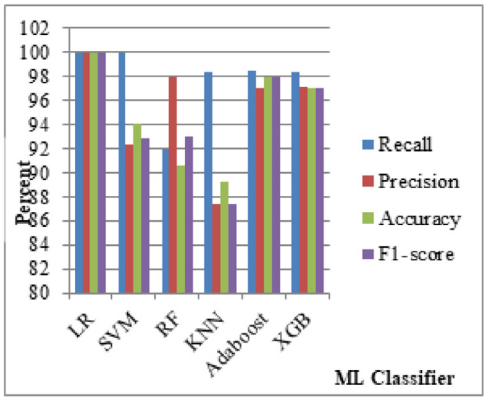

Fig. 11. Evaluation metrics graph for adult dataset.

In future, experimentation can be carried out on different datasets for fast, effective ASD detection and for generating recommendations. Parallel computing with GPUs and Deep learning algorithms can be incorporated while working with massive datasets to speed-up the processing time.

CRediT authorship contribution statement

Anita Vikram Shinde: Conceptualization, Writing – original draft. **Dipti Durgesh Patil:** Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] American Psychiatric Asso., ASD 299.00 (F84.0) DSM of Mental Disorders, Fifth Edition, American Psychiatric Publishing., VA, 2013, pp. 50–59.

[2] Gbd 2015 disease and injury incidence and prevalence, collab, 2016.

[3] F. Hauck, N. Kliewer, Machine learning for autism diagnostics: Applying support vector classification, Int. Conf. Heal. Informatics Med. Syst. (2017) 120–123.

[4] American Psychiatric Asso., ASD, in: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, American Psychiatric Association, VA, 2013, p. 50.

[5] J.S. Handleman, S. Harris, Preschool Education Programs for Children with Autism, 2nd Ed TX: Pro-Ed., 2000.

[6] K. Karre, Y. Rama Devi, IOP Conf. Ser.: Mater. Sci. Eng. 1055 (2021) 012101.

[7] A.R. Olivera, et al., Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - ELSA- Brasil: accuracy study, Sao Paulo Med. J. 135 (3) (2017) 234–246.

[8] X. Yang, M.S. Islam, A.M.A. Khaled, Functional connectivity magnetic resonance imaging classification of ASD using the multisite ABIDE dataset, IEEE EMBS Int. Conf. BHI (2019) 1–4, http://dx.doi.org/10.1109/BHI.2019.8834653.

[9] S. Mostafa, L. Tang, F. Wu, Diagnosis of ASD based on eigen values of brain networks, IEEE Access 7 (2019) 128474–128486, http://dx.doi.org/10.1109/ACCESS.2019.2940198.

[10] M. Elbattah, R. Carette, G. Dequen, J.-L. Guérin, F. Cilia, Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder, in: 4 1st Annual Int. Conf. of EMBC, 2019, pp. 1417–1420, http://dx.doi.org/10.1109/EMBC.2019.8856904.

[11] A.B. Dris, A. Alsalman, A. Al-Wabil, M. Aldosari, Intelligent gaze-based screening system for autism, in: 2nd Int. Conf. ICCAIS, 2019, pp. 1–5, http://dx.doi.org/10.1109/CAIS.2019.8769452.

[12] S. Sartipi, M.G. Shayesteh, H. Kalbkhani, Diagnosing of ASD based on GARCH variance series for rs-fMRI data, in: 9th, IST, 2018, pp. 86–90, http://dx.doi.org/10.1109/ISTEL.2018.8661147.

[13] Iyer, et al., Assess autism level while playing games, in: 2nd CSCITA, 2017, pp. 42–47, http://dx.doi.org/10.1109/CSCITA.2017.8066573.

[14] E. Stevens, et al., A cluster analysis of challenging behaviours in autism spectrum disorder, in: 16th IEEE ICMLA, 2017, pp. 661–666, http://dx.doi.org/10.1109/ICMLA.2017.00-85.

[15] S.R. Dutta, S. Giri, S. Datta, M. Roy, A machine learning-based method for autism diagnosis assistance in children, in: 2017 ICIT, 2017, pp. 36–41, http://dx.doi.org/10.1109/ICIT.2017.26.

[16] H. Abbas, F. Garberson, E. Glover, D.P. Wall, Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening, in: IEEE Int.Conf. on Big Data, 2017, pp. 3558–3561, http://dx.doi.org/10.1109/BigData.2017.8258346.

[17] K. Al-jabery, T. Obafemi-Ajayi, G.R. Olbricht, T.N. Takahashi, S. Kanne, D. Wunsch, Ensemble statistical and subspace clustering model for analysis of ASD phenotypes, in: 38th Annual Int. Conf. of the IEEE EMBC, 2016, pp. 3329–3333, http://dx.doi.org/10.1109/EMBC.2016.7591440.

[18] R. Vikram, A. Rida, Detecting ASD with machine learning models using speech transcripts, 2021, arXiv:2110.03281.

[19] N.M. Mule, D.D. Patil, M. Kaur, A comprehensive survey on investigation techniques of exhaled breath (EB) for diagnosis of diseases in human body, Inform. Med. Unlocked (2021) http://dx.doi.org/10.1016/j.imu.2021.100715.

[20] T. Mantri Shamla, Dipti D. Patil, V.M. Pankaj Agrawal, Wadhai real time multimodal depression analysis, IJITEE (ISSN: 2278-3075) 8 (9) (2019) 2298–2304.

[21] P. Sharwin, T. Mantri Shamla, Dipti D. Patil, Vijay wadhai cognitive depression detection methodology using EEG signal analysis, Intell. Comput. Inform. Commun. (2018) 557–566, http://dx.doi.org/10.1007/978-981-10-7245-1_55.

[22] Dipti D. Patil, G. Mudkanna Jyoti, Dnyaneshwar Rokade, Vijay M. Wadhai, Article: Concept adapting real-time data stream mining for health care applications IJCA proceedings on ICCIA, 2012.

[23] P. Mantri S., S.S. Agrawal, Dorle D. Patil, V.M. Wadhai, Clinical depression analysis using speech features, in: 6th Int. Conf. on Emerging Trends in Engineering and Technology, 2013, pp. 111–112, http://dx.doi.org/10.1109/ICETET.2013.32.

[24] Alcañiz, et al., Eye gaze as a biomarker in the recognition of ASD using virtual reality and machine learning: A proof of concept for diagnosis, Autism Res. 15 (1) (2022) 131–145, http://dx.doi.org/10.1002/aur.2636, Epub 2021 Nov 22. PMID: 34811930.

[25] M.S. Kabir, S. Kurkin, Combination of machine learning and functional networks concept for diagnosis of autism spectrum disorder, in: Fourth Int. Conf. (CNN) Kaliningrad, Russian Federation, 2022, pp. 63–67, http://dx.doi.org/10.1109/CNN56452.2022.9912484.

[26] Z.-A. Huang, et al., Federated multi-task learning for joint diagnosis of multiple mental disorders on MRI scans, IEEE Trans. Biomed. Eng. 70 (4) (2023) 1137–1149, http://dx.doi.org/10.1109/TBME.2022.3210940.

[27] M.N.A. Tawhid, S. Siuly, K. Wang, H. Wang, Automatic and efficient framework for identifying multiple neurological disorders from EEG signals, IEEE Trans. Technol. Soc. 4 (1) (2023) 76–86, http://dx.doi.org/10.1109/TTS.2023.3239526.

[28] A. Sharma, P. Tanwar, Identification of autism spectrum disorder (asd) from facial expressions using deep learning, in: Int. Conf. on COM-IT-CON, Faridabad, India, 2022, pp. 478–484, http://dx.doi.org/10.1109/COM-IT-CON54601.2022.9850892.

[29] H. Shamseddine, S. Otoum, A. Mourad, On the feasibility of federated learning for neurodevelopmental disorders: ASD detection use-case, in: GLOBECOM, IEEE, 2022, pp. 1121–1127, http://dx.doi.org/10.1109/GLOBECOM48099.2022.10001248.

[30] V. Vishal, A. Singh, Y.B. Jinila, K. C, S.P. Shyry, J. Jabez, A comparative analysis of prediction of asd using machine learning, in: 6th Int.Conf. on ICOEI, Tirunelveli, India, 2022, pp. 1355–1358, http://dx.doi.org/10.1109/ICOEI53556.2022.9777240.

[31] K.-F. Kollias, C.K. Syriopoulou-Delli, P. Sarigiannidis, G.F. Fragulis, Autism detection in high-functioning adults with the application of eye-tracking technology and machine learning, in: 11th Int. Conf. on MOCAST, Bremen, Germany, 2022, pp. 1–4, http://dx.doi.org/10.1109/MOCAST54814.2022.9837653.

[32] G. Shushma, I.J. Jacob, Autism spectrum disorder detection using ai algorithm, in: Second Int. Conf. on Artificial Intelligence and Smart Energy (ICAIS) Coimbatore, India, 2022, pp. 1–5, http://dx.doi.org/10.1109/ICAIS53314.2022.9743011.

[33] S. K, et al., Deep learning and iot based assistance system for asd people, in: 4th IEEE MENACOMM, Amman, Jordan, 2022, pp. 83–88, http://dx.doi.org/10.1109/MENACOMM57252.2022.9998222.

[34] Z.M. Fadhel, Q.M. Hussein, Detecting autism spectrum disorder in children using eye tracking and machine learning, in: Int. Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) Ankara, Turkey, 2022, pp. 1–3, http://dx.doi.org/10.1109/HORA55278.2022.9800014.

[35] Shinde Anita, Patil Dipti, Content-centric prediction model for early autism spectrum disorder (ASD) screening in children, 2022, http://dx.doi.org/10.1007/978-981-19-5331-6_38.

[36] E. Uğur, N.H. Dang, Autism spectrum disorder detection with machine learning methods, Curr. Psych. Res. Rev. 15 (4) (2019) 297–308, http://dx.doi.org/10.2174/2666082215666191111121115.

[37] Kaushik Vakadkar, Diya Purkayastha, Deepa Krishnan, Detection of autism spectrum disorder in children using machine learning techniques, SN Comput. Sci. 2 (2021) 386, http://dx.doi.org/10.1007/s42979-021-00776-5.

[38] K. Vijayalakshmi, M. Vinayakamurthy, Anuradha, A hybrid recommender system using multiclassifier regression model for autism detection, in: Int. Conf. (ICSTCEE), 2020, pp. 139–144, http://dx.doi.org/10.1109/ICSTCEE49637.2020.9277034.

[39] T. Akter, M.I. Khan, M.H. Ali, M.S. Satu, M.J. Uddin, M.A. Moni, Improved machine learning based classification model for early autism detection, in: 2nd ICREST, 2021, pp. 742–747, http://dx.doi.org/10.1109/ICREST51555.2021.9331013.

[40] T.Y. Rashme, L. Islam, A.A. Prova, S. Jahan, Autism screening disorder: Early prediction, in: IEEE 4th Int. Conf. on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1–6, http://dx.doi.org/10.1109/GUCON50781.2021.9573547.

[41] R. Suman, M. Sarfaraz, Analysis and detection of asd using machine learning techniques, Procedia Comput. Sci. (ISSN: 1877-0509) 167 (2020) 994–1004, http://dx.doi.org/10.1016/j.procs.2020.03.399.

[42] F.F. Thabtah, Autistic spectrum disorder screening data for children, 2017, https://archive.ics.uci.edu/ml/machine-learning-databases/00419/.

[43] O. Altay, M. Ulas, Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and k-nearest neighbor in children, in: 6th Int. (ISDFS), 2018, pp. 1–4, http://dx.doi.org/10.1109/ISDFS.2018.8355354.

[44] Anita Shinde, Reena Pagare, Article: A study of recommender system techniques, IJCA 47 (16) (2012) 1–4.

[45] Anita Shinde, Dipti D. Patil, A comprehensive survey on recommender systems techniques and challenges in big data analytics with IoT application, J. Pharmaceut. Negative Results 13 (4) (2022) 1492–1505, http://dx.doi.org/10.47750/pnr.2022.13.04.211.