A Review of Machine Learning Models to Detect Autism Spectrum Disorders (ASD)

PRASENJIT MUKHERJEE[a], SOURAV SADHUKHAN[b], MANISH GODSE[c]
[a]Dept. of Technology
Vodafone Intelligent Solutions
Pune
INDIA

[a]Dept. of Computer Science
Manipur International University
Manipur
INDIA

[b]Dept. of Business Management
Pune Institute of Business Management
Pune
INDIA

[c]Dept. of IT
Bizamica Software
Pune
INDIA

Abstract: - Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that can manifest in a variety of ways. One common characteristic is difficulty with communication, which may manifest as difficulty understanding others or expressing oneself effectively. Social interaction can also be challenging, as individuals with ASD may struggle to comprehend social cues or adapt to new situations. Many machine-learning models have been developed or are in progress to detect ASD automatically. Three machine learning model-based frameworks have been studied and elaborated on, each with a clear concept of the detection of ASD among children and adults. This research paper has done a closer review of these frameworks and their datasets to diagnose ASD automatically. In the first framework, deep learning models such as Xception, VGG19, and NASNetMobile have been utilized for the detection of autism spectrum disorder (ASD). In addition, other models such as XGBoost, Neural Network, and Random Forest have been employed in the second framework to detect ASD from a clinical standard screening dataset for toddlers. Meanwhile, the third framework involves traditional machine learning models that have been trained using the UCI dataset for ASD. The accuracy of each model has been discussed and elaborated on.

Key-Words: - Deep Learning, Autism Spectrum Disorder, Machine Learning, ASD Detection, ML-based Framework, Traditional Machine Learning

Received: June 23, 2023. Revised: August 11, 2023. Accepted: September 14, 2023. Published: October 5, 2023.

1 Introduction

Autistic children often have difficulty understanding and responding to social cues, so they may not know how to start and maintain conversations. Additionally, they may have difficulty understanding abstract concepts and may be more comfortable with concrete concepts. They may also have trouble interpreting sensory information, such as touch or sound, which can lead to sensory overload. Finally, autistic children may be obsessed with certain topics or routines due to their difficulty processing changes in their environment. These difficulties have been attributed to the lack of reliable and valid screening instruments, the wide range of severity of ASD symptoms, and the overlap of symptoms with other disabilities. Additionally, early intervention can be expensive and may not always be available, depending on the particular situation of the family, as in [1]. ASD is a neurodevelopmental problem of the brain that has a wide range of symptoms and severity [2]. ASD has been included in the International Statistical Classification of Diseases and Related Health Problems (ISCDRHP) under the category of mental and behavioral disorders, as in [3]. The symptoms may appear in the first year of a toddler with less eye contact and poor responses, as in [4] [5] [6] [7]. People with autism may experience difficulties in social communication, such as difficulty understanding body language, facial expressions, and the meaning of words. They may also struggle with sensory processing, such as being over or under-sensitive to certain sounds, textures, lights, and tastes, as in [8]. It is characterized by difficulties with social interaction and communication, as well as restricted, repetitive behaviors. It is typically diagnosed in early childhood and can last throughout a person's lifespan. Symptoms of autism spectrum disorder are usually noticeable before the age of three and can range from difficulty communicating and interacting with others to repetitive behaviors and hypersensitivity to certain stimuli. These symptoms can vary greatly in severity and type between individuals. Machine learning algorithms can be used to analyze patterns in the behavior of children with autism and detect any abnormalities that might indicate the presence of autism. This can help clinicians diagnose the condition earlier and begin treatment sooner, which can improve the outcome for the child. Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects social communication, behavior, and sensory processing. Early identification and intervention are crucial to improving outcomes and quality of life for individuals with ASD. Some of the core features of ASD include difficulties with social interaction, communication, and repetitive behaviors or interests. These challenges can make it difficult for individuals with ASD to form and maintain relationships, understand social cues, and participate in everyday activities. Although the exact causes of ASD are not fully understood, research suggests that a combination of genetic and environmental factors may contribute to its development. While the condition is more prevalent in males, it is important to note that ASD affects individuals of all genders, races, and ethnicities, as in [9]. Diagnosing ASD can be a complex process that involves a thorough evaluation of a person's behavior, communication, and developmental history. However, access to a timely and accurate diagnosis can be limited, particularly for families in low-income communities. This can lead to delays in accessing appropriate services and support. Advances in technology, such as machine learning algorithms, have the potential to improve the accuracy and speed of ASD diagnosis. By analyzing large datasets, these algorithms can identify patterns and features that are characteristic of the condition, which may assist clinicians in making more accurate and efficient diagnoses. While these tools are not intended to replace clinical judgment, they may help supplement traditional assessment methods and increase access to diagnostic services for individuals and families affected by ASD, as in [10].

AI techniques can be used to analyze large amounts of data from various sources, such as genetics, medical records, and environmental factors. With AI, patterns can be identified and used to develop predictive models for ASD, which can help identify individuals at risk for the disorder and provide early interventions. The challenge arises because high-dimensional data has a large number of features and variables, which can make it difficult to identify meaningful patterns in the data. Furthermore, the sheer size of the data can make it difficult to process and analyze. As a result, the analysis of high-dimensional datasets requires specialized algorithms that can accurately identify patterns in the data. These algorithms must also be computationally efficient enough to process large amounts of data in a reasonable amount of time, as in [8].

The proposed research work is a review of AI applications to detect autism spectrum disorder among children and adults. Three frameworks that contain the machine learning models have been discussed with the dataset. The dataset plays an important role because each model uses datasets to train for predictions after getting new data. The facial images of ASD-detected children and general children have been taken as primary sources of data. Deep learning models like Xception, VGG19, and NASNetMobile have been applied to detect ASD, as in [11]. The second framework uses models like XGBoost, Neural Network, and Random Forest to detect ASD from the clinical standard ASD screening dataset of toddlers, as in [12]. The third framework uses traditional machine learning models that are trained with the UCI dataset of ASD, as in [13]. These frameworks and some other similar models have been elaborated on in Section III. The entire study is given in Section III, where the results of each model and observation have been discussed in Section IV and the application of the proposed study has been included in Section V. The proposed study ends with a conclusion in Section VI.

2 Related Works

Autism Spectrum Disorder (ASD) is a significant challenge for children's health today, and it has become a key area of focus in healthcare research. Many studies have explored the potential of artificial intelligence (AI) to address this disorder and other mental health-related issues. This section highlights some notable AI-based research on mental health that has been conducted.

By analyzing social media posts and biomedical images, doctors can identify patterns in behavior and physical symptoms that may be indicative of ASD. This data can then be used to accurately diagnose and treat the disorder. By recognizing facial features associated with ASD, it is possible to identify individuals with the disorder earlier in life. This can lead to earlier diagnosis and intervention, which can be beneficial for those affected by the disorder. In addition, this system could also be used to help identify individuals with ASD in social media posts, which can help connect those with the disorder with resources and support. Deep learning techniques rely on accurately identifying key facial features, such as eyes, nose, and mouth, and then mapping those features to a template. This allows the algorithm to recognize the face and identify the landmarks associated with it. The exception model achieved the highest accuracy result of 91%, followed by VGG19 (80%) and NASNETMobile (78%). The dataset used had a good variety of face images of different backgrounds, angles, and lighting conditions, which allowed the deep learning models to accurately perceive and recognize patterns and features of the faces. This enabled the three models to detect a wide variety of faces, which is why the exception model achieved the highest accuracy result. The application is designed to assess facial features from images of people's faces and compare them to a database of images of people with and without autism. The convolutional neural network is trained to recognize the differences between the two sets of images and categorize the images accordingly. The Flask framework then makes the application available online and allows users to easily interact with the system, as in [11]. It is also associated with difficulty processing sensory information and difficulty with motor skills such as handwriting or balancing. People with autism often have difficulty understanding and responding to social cues and may have difficulty forming relationships with others. Because autism is a spectrum disorder, it can manifest differently in each individual. This means that the symptoms can range from mild to severe, making it difficult to distinguish between typical development and autism. Furthermore, autism is often comorbid with other mental health issues, which can make it even more difficult to diagnose. Early screening and treatment can help identify and address any underlying health issues before they become severe. This can help reduce the risk of long-term symptoms as well as improve the overall quality of life. The goal of this research is to develop an automated pipeline that can quickly and accurately identify the signs of autism in toddlers and to use machine learning models to analyze the indicators of autism and determine which are the most significant for diagnosis. The dataset used for this research was curated from the UC Irvine Autism Spectrum Disorder dataset, which contains over 10,000 examples of autism-related features from children aged 4-5. The neural network model was designed to learn patterns from large datasets, while the random forest model was designed to identify relationships between variables. After they were trained on the data, they were tested on a new dataset to determine how accurately they could identify the presence of autism. LightGBM is an algorithm that measures the importance of each feature in a dataset by assigning a score to each one. We used this to identify which physical characteristics had the highest scores, indicating that they are most significant in giving rise to autism. To arrive at this conclusion, the study used a combination of genetic and physical features, including facial features, to create a machine-learning model to analyze the data. The model was then tested and validated against a set of data containing individuals with and without autism. The results indicated that the model was highly accurate at predicting the presence of autism, indicating the importance of physical characteristics in identifying autism. By catching signs of autism early, doctors can intervene and help the patient learn coping skills and manage the symptoms. This can help minimize the impact of autism on their lives and increase their quality of life, as in [12]. Background: Machine learning algorithms, when applied to data collected from patients with ASD, can help identify the features of the disorder, such as social and communication deficits, and thus enable more accurate and efficient diagnosis. With the help of machine learning, doctors will be able to better identify, diagnose, and treat patients with ASD. This is likely due to improved awareness and diagnosis of ASD, as well as an increase in environmental factors that can contribute to its development. Additionally, advances in technology and medical care have made it easier to identify the signs of ASD and diagnose it in a timely manner. Early diagnosis of ASD can have a major impact on the quality of life of individuals with ASD, as early interventions can be more effective and provide better outcomes. This study seeks to provide a simple and accurate way to classify ASD data, which can help with early diagnosis. By randomly splitting the data and running the experiments multiple times, we were able to identify the best method for each dataset. This allowed us to compare the performance of the different methods and determine which one was the most effective for each dataset. The accuracy of SVM and RF was compared to the other models and shown to be the highest. Additionally, the results indicated that SVM was better at generalizing and was more efficient in terms of training time, while RF was better at handling imbalanced data. This is likely because SVM is a discriminative classifier that tries to classify the data points by finding the optimal hyperplane that separates the two classes, while RF is an ensemble method that uses a collection of decision trees to achieve better performance than a single decision tree. Additionally, the RF method uses randomization to create diversity among its decision trees, which allows it to better handle imbalanced data. Random Forest (RF) is an ensemble machine-learning method that uses multiple decision trees to make predictions. Because it combines multiple models and considers the relationships between variables, it has been shown to outperform other machine learning methods when it comes to diagnosing ASD, as in [13]. Early detection and intervention are critical for helping children with autism get the most out of therapy and other interventions. If screening methods are easily implemented, it will allow for early detection, enabling families to get their children the help that they need as soon as possible. It is believed that ASD is the result of a combination of genetic, environmental, and biological factors. Research suggests that there may be distinct differences in the brain structure and function of individuals with ASD, which may explain their different behaviors and abilities. The logistic regression model is used because of its ability to accurately predict binary outcomes, such as whether a child has autism or not. The algorithm will be used to quickly process large amounts of data and make accurate predictions based on the data in the dataset. With machine learning, doctors can detect the disorder more quickly and accurately by using algorithms that look for patterns in the data. This can help them identify the disorder earlier and provide the necessary care to the toddler in a timely manner, improving their quality of life. These challenges include a lack of reliable data sets and data infrastructure, limited access to skilled personnel, and a lack of understanding of the legal and ethical implications of AI-powered applications, as in [14]. This is due to increased awareness of the condition and improved diagnostic tools, as well as a greater understanding of the condition and its effects on individuals' lives. More research is being done on the topic, leading to improved treatments and therapies. Some people with ASD may have difficulty with communication and forming relationships, while others may have only mild symptoms. Additionally, some people may have associated medical issues, such as seizures or sleep disturbances. Other common symptoms seen in those with autism include limited or inappropriate social interactions, difficulty with communication, restricted and repetitive behaviors, and sensory sensitivities. Diagnosis of autism can be done at any age through observation of these behaviors, physical examinations, cognitive testing, and genetic testing. This is to allow for a more accurate diagnosis of ASD as well as to enable early intervention to ensure that the symptoms do not worsen. This is done by using ML algorithms to analyze data such as patient records, behavior, and medical history to identify patterns that could indicate the presence of ASD. LR and SVM are two popular machine learning (ML) algorithms that can be used to classify data. The performance measure helps to compare the accuracy of the predictions made by the model with each algorithm. This can help users determine which algorithm provides more accurate results in a shorter amount of time, which can help them determine if they are suffering from ASD or not, as in [15]. ASD can cause difficulties in communication, social skills, and repetitive behaviors. It is believed to be caused by a combination of genetic and environmental factors and can affect people in different ways. Early intervention is key to reducing the effects of ASD, as it can help children learn the skills they need to better manage their symptoms and lead more independent lives. It also provides an opportunity for parents and caregivers to better understand their children and find ways to cope with and manage the disorder. This makes it difficult for physicians to accurately identify ASD symptoms and recognize them as being uniquely associated with ASD. As a result, the diagnosis is often delayed or missed entirely. Deep learning algorithms are able to uncover complex patterns in large amounts of data that may be too subtle or too complicated for a human expert to detect. By utilizing these algorithms, medical experts can be provided with more accurate and timely diagnoses of ASD, which can help improve treatment and outcomes for those affected by the disorder. With a larger dataset, machine learning algorithms can be used to develop better models for diagnosing ASD. The algorithms can also take into account subtle or nuanced symptoms that may not be easily detected by medical professionals, which can lead to more accurate diagnoses. The hybrid approach is beneficial because it combines the power of deep learning to extract complex patterns from data with the interpretability of XAI to explain why certain features are more important than others in predicting ASD. This helps to reduce the bias in the predictions and makes the results more trustworthy. The proposed framework combines data from both parents and clinicians to create a more comprehensive picture of a child's development. This data can then be used to make more accurate predictions about which children are likely to have ASD traits, allowing clinicians to provide earlier interventions and support, as in [16]. People with ASD typically have difficulty with social interaction, communication, and understanding language. They may also show restricted or repetitive behaviors, such as having difficulty transitioning from one activity to another. Children with autism can struggle with social interactions, body language, and understanding facial expressions. Early diagnosis can equip families with the necessary resources and interventions to help their child reach their full potential. With the prevalence of ASD increasing in recent years, it is becoming increasingly difficult for medical professionals to diagnose the condition in children without the help of automated methods. Automated methods can quickly and accurately detect signs of ASD in children, allowing medical professionals to make more informed decisions about diagnosis and treatment. We selected the AutoML method because it has the ability to automate the process of building, optimizing, and selecting the best-performing model with minimal manual effort. In addition, AutoML can also be used to identify important features in the dataset, which can then be further used to improve the accuracy of the machine-learning models. This is due to the fact that AutoML automates the process of selecting optimized feature combinations and hyperparameters, allowing us to quickly identify the optimal settings for our model. The combination of these techniques allowed us to achieve the highest accuracy with minimal effort, as in [17]. ASD is caused by a combination of genetic and environmental factors, including gene mutations and exposure to toxins. People with ASD may also have trouble forming social relationships, have difficulty with communication and language, and struggle with sensory sensitivity. MRI imaging modalities have the capability to detect subtle brain abnormalities that are associated with ASD, such as changes in the brain's structure, connectivity, and even chemistry. This makes it an invaluable tool for diagnosing and monitoring ASD. fMRI uses magnetic fields and radio waves to measure blood flow in the brain and identify any abnormalities or discrepancies in brain activity. sMRI uses high-resolution images to map the structure of the brain and detect any abnormalities in the brain's anatomy. These two modalities work together to help clinicians diagnose ASD with greater precision. These systems use AI to analyze brain images, such as MRI and fMRI scans, to assess an individual's brain structure and connectivity. The AI algorithms can detect subtle differences in brain structures, which can be used to diagnose ASD more accurately and quickly by specialists. ML algorithms are used to analyze the image data, identify the relevant features, and detect any abnormalities that could be indicative of ASD. DL applications are used to further analyze the data and identify patterns that may be indicative of ASD. This allows for more accurate and reliable diagnoses. Deep learning (DL) techniques employ large datasets of MRI images and AI algorithms to create models that can detect patterns in the images that are associated with ASD. These models can then be used to automate the diagnosis of ASD and provide more accurate and timely results. We compare the accuracy and training times of ML and DL models to show that DL models can learn faster and achieve higher accuracy. We also discuss the importance of feature selection and data pre-processing in improving the accuracy of the models. Finally, we suggest the potential of combining AI techniques with MRI neuroimaging to detect ASDs, as in [18]. It is usually diagnosed during early childhood, and symptoms can range from mild to severe. Common characteristics of ASD are difficulty with social interactions, difficulty with verbal and nonverbal communication, difficulty with sensory integration, and an overall difficulty in adapting to change. As a result, many healthcare providers are looking for more cost-effective ways to diagnose ASD, such as through the use of screening tools that can help identify the presence of ASD symptoms in a shorter amount of time. Additionally, research has found that early detection and intervention of ASD can have a significant impact on the child's development, so it is important to identify the disorder as quickly as possible. These methods are designed to provide a clearer picture of what is happening in the person's life, allowing for a more accurate diagnosis. The AQ and M-CHAT use standardized questions about social interaction, communication, and behavior to assess the individual's level of autism spectrum disorder. The user must be knowledgeable about the various items that need to be screened and be able to identify any discrepancies that could lead to inaccurate results. The screening items must be designed in such a way that they allow for accurate and efficient screening. ML algorithms can process large amounts of data quickly and efficiently. By taking advantage of such algorithms, we can greatly reduce the time needed to detect patterns, uncover trends, and identify anomalies in the data. These patterns and trends can then be used to make more accurate diagnoses, leading to improved accuracy and efficiency in the diagnostic process. RML is based on a combination of rule-based and ML techniques, which allows it to detect patterns in data that traditional ML techniques cannot. Furthermore, it provides users with interpretable rules that can be used to gain a better understanding of the data as well as identify potential areas for further research. This is likely due to RML's ability to learn from the data and identify patterns in the data that are not visible to traditional ML methods. Additionally, RML's ability to handle complex data and its ability to adjust to new data as it comes in make it a powerful tool for classification, as in [19].

3 Machine Learning Models in ASD Detection

Today, artificial intelligence has established its presence in all sectors, including healthcare. Autism Spectrum Disorder (ASD) detection is a difficult challenge in the healthcare domain. Early detection of ASD is needed to start treatment to reduce all the symptoms of ASD. ASD is not curable, but it is possible to manage its symptoms. Parents have a crucial role in detecting ASD at the early age of a baby. In the detection of ASD, many types of research have been done or are in progress to use machine learning models and various kinds of datasets. In this section, a discussion has been done on the detection of ASD using machine-learning models. The discussion has progressed according to ASD detection cases. Each case has been described with the proper dataset, machine learning models, and model performance. The discussion about each framework has been given in the next section. Our primary aim is to understand the architecture of each system where simulation and numerical stability have been normalized, as in [11], [12], and [13].

3.1 ASD Detection Using Facial Images

Machine learning models are used to detect ASD among children. Machine learning models like Xception, VGG19, and NASNETMobile are very advanced image-based machine-learning models.

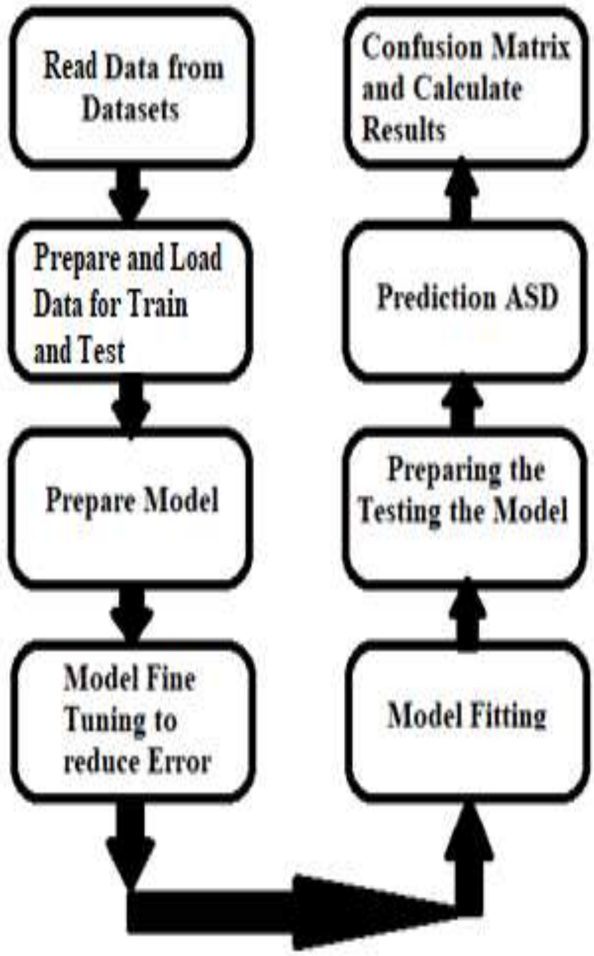

Fig. 1. Framework of ASD Detection using ML

3.1.1 Dataset

The dataset [11] has been prepared using images of autistic children and general children. The facial Images have been captured, which are the main input for the machine-learning models. The dataset has been prepared with 2940 facial images, of which half are of autistic children and the remaining half are of general children. The images have been collected from the social autism groups on Facebook, as in [11].

3.1.2 Framework

A clear framework to detect ASD has been given, where each section is described in Fig. 1. According to Fig. 1, the given framework [11] shows that data will be read from the dataset and split into the train and test parts in the first and second steps. The model will be prepared to train with training data. After the completion of training, the model will be fine-tuned to reduce the error in the next step. After reducing the error, the model will be fitted and tested with test data for validation. After completion of this step, the model will be ready to predict using new data from the user side. The predicted result of the test data will be utilized to calculate the accuracy, precision, recall, and confusion matrix in the final step. Three advanced machine-learning models have been used to detect autism from facial images, as in [11]. The input dataset contains the image data for training and testing these models. The result of these models has been discussed in section IV.

3.2 ASD Detection of Toddlers Using Machine Learning Models

This work [12] has been proposed to detect ASD in toddler children. The age range of toddler children is between 12 months and 3 years. This is a good time to detect ASD among children because early detection helps to start ASD therapies according to the need. AI already accepts this challenge to find out the solution to early detection of ASD among children. Many models have been developed that are useful in the detection of ASD. XGBoost, neural networks, and random forest models have been used to detect ASD among toddlers, as in [12].

3.2.1 Dataset

The dataset [12] has been collected from Kaggle.com, which is an open-source repository of machine learning. The autism dataset was prepared by the University of California, Irvine. This dataset contains the screening data for toddlers. The dataset contains 1054 records with 18 variables that point to different attributes. 10 variables are questions that determine ASD among toddlers. These 10 variables are questions related to autism. The questions are set from A1 to A10. If the answer to Questions A1 to A9 is “sometimes”, "rarely," or "never," then the value will be assigned as 1, and 0 will be the opposite of these answers. If the answer to question A10 is "always,", "usually,", or "sometimes,", then the value will be assigned as 1, and 0 will be the opposite of the answer. The scores of these questions and other attributes have been used to train the models for the prediction of ASD, as in [12].

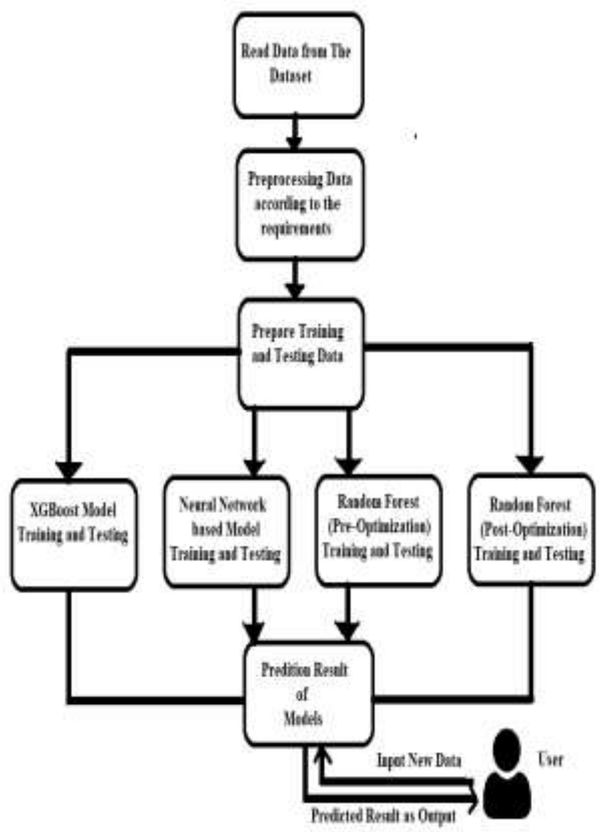

Fig. 2. Framework of ASD detection among Toddlers using ML

3.2.2 Framework

Fig. 2 shows the framework of this system [12], which is equipped with machine learning models like XGBoost, neural networks, and random forests. The data will be read from the dataset. The preprocessing task will be applied when it requires some cleaning in the second step. According to the third step, the data will be split into training and testing parts. Now, each model, like XGBoost, neural networks, and random forests, will use training data to train and understand the pattern, as in [12]. After the completion of training, each model will be evaluated using testing data. In the end, models are ready to predict results according to the user’s input. The Random Forest model has been used with pre-optimization and post-optimization. The XGBoost model is an ensemble model that is equipped with many weak models. XGBoost stands for ‘Extreme Gradient Boosting’ and is the most popular machine learning model that accepts large datasets, and its overall performance is good and stable. The neural network has been developed by inspiring the human brain. Neural networks are used to solve complex machine-learning problems because of their ability to compute quickly and generate responses quickly. The other model is the

Random Forest, which is the most popular model to solve classification problems in machine learning. The random forest model has been developed by the decision trees. A decision tree is the key point of the Random Forest algorithm. The results of these models have been discussed in Section IV.

3.3 ASD Detection Using Traditional Machine Learning Models

The Support Vector Machine (SVM), K Nearest Neighbour (KNN), and Random Forest models are used to detect Autism using the UCI dataset as in [13].

3.3.1 Dataset

Three datasets [13] have been used to solve the ASD detection problem. These three datasets have been taken from the UCI database. The three datasets are AQ-10-Adult for adults, AQ-10-Adolescence for adolescents, and AQ-10-Child for children. The data has been classified into train data and test data with different values. The values will be selected randomly. The score of each subset of data is measured with average accuracy, average sensitivity, average F-measure, and average AUC, as in [13].

3.3.2 Framework

The framework [13] is equipped with three models: SVM, KNN, and Random Forest. All these models are best for the classification problem. The first model is SVM, which creates the best line or decision boundary to classify the n-dimensional space for plotting new data points in the correct category. SVM uses the vectors to create the hyperplan. The optimum hyperplan segregates the vectors that define the classes. The KNN is another supervised machine learning algorithm that can be used for classification or regression. The K is the nearest neighbor that has been used by the KNN algorithm. A majority vote for a particular class determines that a new observation should be inside it. Larger values of K refer to stable decision boundaries for classification, whereas small values of K refer to decision boundaries that are not better than a larger K value. Random Forest is a popular model in classification. This model contains a number of decision trees according to the various subsets of the dataset, and it will calculate the average for prediction. The greater number of decision trees refers to the higher accuracy that prevents the overfitting problem, as in [13].

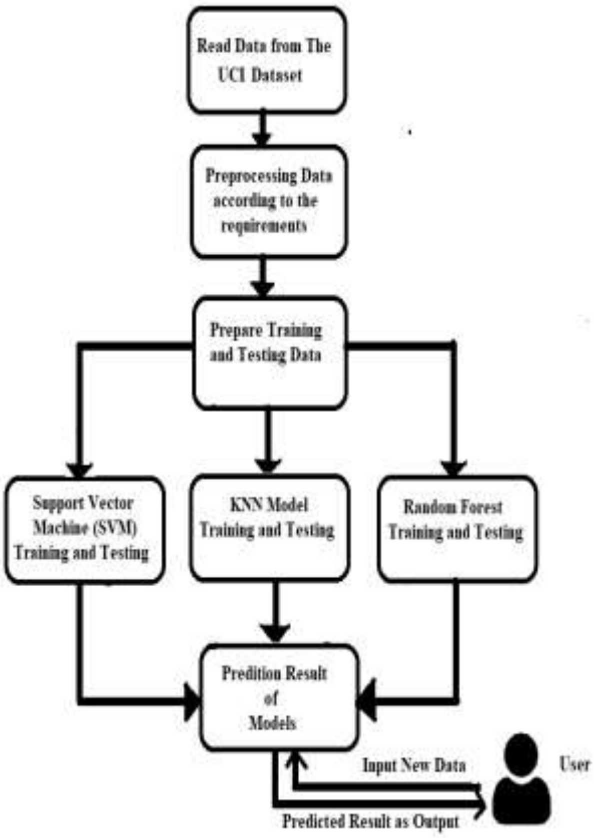

Fig. 3. Framework of ASD detection using Traditional ML

According to Fig. 3, first, the data is read from the UCI dataset [13] and split into training and testing sets. Then, a model is prepared for training using the training data. After the model is trained, it is fine-tuned to reduce errors and increase accuracy. The model is then fitted with the data and ready for testing using the test data. Once testing is complete, the model can be used to predict results based on the new data. The machine learning models SVM, KNN, and Random Forest will be trained using the UCI data as in [13]. The accuracy of each model has been discussed in Section IV.

4 Results and Discussion

The models have been discussed with a framework in Section III. This section refers to the discussion about the results of each model. The first framework is ASD detection using facial images of autistic children and general children, as in [11]. The second framework is the ASD detection of toddlers using screening data [12], and the third frame is the ASD detection from the UCI dataset as in [13]. Each framework contains machine learning models, and these models have been used for prediction after successful training and testing.

4.1 Result of the Models of ASD Detection Using Facial Images

The first framework [11] is about ASD detection using facial images. Three deep learning models—Xception, VGG19, and NASNetMobile—have been implemented to recognize ASD among children using their facial images as data. The accuracy of the Xception model has scored 91%, whereas the specificity and sensitivity of this model are 94% and 88%, respectively. The VGG19 model has scored 80% accuracy, and its specificity and sensitivity values are 83% and 78%, respectively. The NASNetMobile model has 78% accuracy, 75% specificity, and 82% sensitivity. Specificity, on the other hand, measures the ability of a model to correctly identify negative instances of a given category. It is calculated as the number of true negatives divided by the sum of true negatives and false positives. Sensitivity is a metric that measures the ability of a model to correctly identify positive instances of a given category. It is calculated as the number of true positives divided by the sum of true positives and false negatives. The accuracy, specificity, and sensitivity of each model have been given in Table 1 as in [11].

Table 1. Accuracy, Specificity, and Sensitivity of Each Model

Sl. No.	Models	Specificity	Sensitivity	Accuracy
1	Xception	0.94	0.88	0.91
2	VGG19	0.83	0.78	0.80
3	NASNETMobile	0.75	0.82	0.78

4.2 Result of ASD Detection of Toddlers Using Screening Data

The baseline XGBoost model [12] has performed well to detect ASD among toddlers. The toddler's dataset contains 18 variables, where A1 to A10 are questions that need answers to train the model. The other machine learning models are also used to detect ASD among toddlers. Neural networks and Random Forest pre- and post-optimization are the models that are used, and their performance has been given in Table 2 as in [12].

Table 2. Performance Scores of ASD Detection Models among Toddlers

Sl. No.	Model	Precision	Recall	F1	Accuracy
1	Neural Network	100%	100%	100%	100%
2	Random Forest(Pre-Optimization)	98.15%	98.10%	98.09%	98.10%
3	Random Forest(Post-Optimization)	100%	100%	100%	100%
4	XGBoost				97.04% - Mean Accuracy and 1.78% Standard Deviation

The performance scores of each model can be seen in Table 2, where the neural network model has 100% accuracy with 100% precision and a 100% recall value. The Random Forest (post-optimization) model has the same scores as the neural network model. It has 100% scores in precision, recall, and accuracy. The Random Forest (pre-optimization) scored 98.15% in precision, 98.10% in the recall, and 98.10% in accuracy, whereas XGBoost has a 97.04% accuracy score with a standard deviation value of 1.78%, as in [12].

4.2 Result of ASD Detection Using Traditional Machine Learning Models

The UCI dataset [13] has been taken as the main data source to detect ASD among children. The three most popular traditional machine learning algorithms have been used for ASD prediction, according to the new input. These models are KNN, SVM, and Random Forest (RF). The performance of each model has been given in Table 3 as in [13].

Table 3. Performance of Traditional Machine Learning Models to Detect ASD among Children

Sl. No.	Models	Case	AUC
1	KNN	AQ-10-Adults for the Case of Complete Data	0.94
2	SVM	AQ-10-Adults for the Case of Complete Data	1.00
3	RF	AQ-10-Adults for the Case of Complete Data	1.00
4	KNN	AQ-10-Adults for the Case of Missing Data	0.93
5	SVM	AQ-10-Adults for the Case of Missing Data	1.00
6	RF	AQ-10-Adults for the Case of Missing Data	1.00
7	KNN	AQ-10- Adolescence for the Case of Complete Data	0.87
8	SVM	AQ-10- Adolescence for the Case of Complete Data	0.97
9	RF	AQ-10- Adolescence for the Case of Complete Data	1.00
10	KNN	AQ-10- Adolescence for the Case of Missing Data	0.85
11	SVM	AQ-10- Adolescence for the Case of Missing Data	0.98
12	RF	AQ-10- Adolescence for the Case of Missing Data	1.00
13	KNN	AQ-10-Child for the Case of Complete Data	0.85
14	SVM	AQ-10-Child for the Case of Complete Data	0.89
15	RF	AQ-10-Child for the Case of Complete Data	0.99
16	KNN	AQ-10-Child for the Case of Missing Data	0.85
17	SVM	AQ-10-Child for the Case of Missing Data	0.91
18	RF	AQ-10-Child for the Case of Missing Data	1.00

Table 3, Table 3 shows the performance graph of the KNN, SVM, and RF, which is based on the AUC scores. Six cases have been classified as: 1. AQ-10-Adults for the Case of Complete Data, 2. AQ-10-Adults for the Case of Missing Data, 3. AQ-10: Adolescence for the Case of Complete Data; 4. AQ-10: Adolescence for the Case of Missing Data; 5. AQ-10: Child for the Case of Complete Data; and 6. AQ-10: Child for the Case of Missing Data. The AUC score has been calculated by the true positive rate and the false positive rate, as in [13]. The AUC scores according to the first case of KNN, SVM, and RF are 0.94, 1.00, and 1.00. In the second case, the AUC scores of KNN, SVM, and RF are 0.93, 1.00, and 1.00. The AUC scores of KNN, SVM, and RF in the 3rd case are 0.87, 0.97, and 1.00, whereas the AUC scores are 0.85, 0.98, and 1.00 in the 4th case. The AUC scores of KNN, SVM, and RF are

0.85, 0.89, and 0.99 in the 5th case, whereas 0.85, 0.91, and 1.00 in the 6th case, as in [13].

Each framework with machine learning models has been discussed with figures and tables that indicate the procedure for ASD detection among children or adults. These models are good for detecting ASD, but it is more important that models detect the kinds of symptoms in ASD individuals. Early detection of ASD among children is much more important than ASD detection among adults. You will get a good result if you start therapies on early ASD-detected children. If ASD is detected after a certain age, then it will be difficult to get any good improvement results for ASD individuals. The discussed frameworks will work on ASD detection, but ASD symptoms are also needed for further understanding of the need for therapies. Sometimes fixed attributes are used in the dataset, or some questions regarding ASD have been used in the dataset with answers, but according to the ASD problem, the best data source for ASD is parents because an ASD individual spends most of the time with the parent. Facial images are also used to detect autism, but it is difficult to segregate ASD and ADHD children through images. A child may have an ASD problem or ADHD problem, as well as a Global Development Delay (GDD) problem, which can be there. Facial images cannot segregate this problem correctly. If there are hundreds of ASD children, and if we find patterns, then it is possible to get a hundred different patterns from ASD children. There are no fixed patterns to detect ASD among children, but a few may be common. The study of these three frameworks has given a clear understanding of the strong role of AI in ASD detection, where a hybrid approach can be executed.

5 Application of the Proposed Study

The application of the proposed study is to understand the various techniques of ASD detection among children. Today, autism is a major issue among children, according to the World Health Organization (WHO, https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders). Many applications have been developed using machine learning and natural language processing to detect autism in the early stages, but the research method may not be cost-effective and the data regarding this problem is not up to par. The proposed study can be useful in identifying space for further research, like parent-child dialogue with an autistic child. The deep learning models can be applied to understand the symptoms of autism at an early age from the parent's dialogue.

6 Conclusion

Three frameworks have been discussed with machine learning models to detect ASD among children, toddlers, and adults. The first approach is ASD detection using the facial images of ASD children and general children. Advanced machine learning models have been used to detect ASD. These models are trained with facial images, and their accuracy is very high for the detection of ASD. The second framework is about ASD detection among toddlers. This framework used some screening data from toddlers to train the machine learning models. ASD detection at an early age is a good option to start therapies to reduce the symptoms of ASD. Each model has been trained with the ASD dataset, and performance scores are high according to the predictions. The third framework contains some traditional machine-learning models that are popular machine-learning models for classification problems. These models are able to predict ASD among children, adolescents, and adults with high accuracy after training with complete and missing data. These three kinds of frameworks have elaborated on AI applications in the healthcare domain with strong results. The deep learning models can be applied to the parent-child dialogues of an autistic child. The parent's dialogues are nothing but textual information about their children, and this data can be utilized for identifying the symptoms of autism. After the detection of symptoms from the parents' dialogues, the symptoms can be analyzed according to the severity of autism, and this task will be a future enhancement.

Acknowledgement:
The authors extend their appreciation to the Manipur International University, Imphal, India for supporting this research work on Autism.

References:

[1] Maria Lai, Jack Lee, Sally Chiu, Jessie Charm, Wing Yee So, Fung Ping Yuen, Chloe Kwok, Jasmine Tsoi, Yuqi Lin, Benny Zee, A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder, *EClinical Medicine*, 2020, pp. 1-20.

[2] C. S. Paula, S. H. Ribeiro, E. Fombonne, and M. T. Mercadante, Brief report: prevalence of

pervasive developmental disorder in Brazil: a pilot study, *Journal of Autism and Developmental Disorders*, vol. 41, no. 12, 2011, pp. 1738–1742.

[3] L. C. Nunes, P. R. Pinheiro, M. C. D. Pinheiro et al., A Hybrid Model to Guide the Consultation of Children with Autism Spectrum Disorder, A. Visvizi and M. D. Lytras, Eds., *Springer International Publishing*, View at: Google Scholar, 2019, pp. 419–431.

[4] Apa–American Psychiatric Association, Diagnostic and statistical manual of mental disorders (DSM –5), 2020, https://www.psychiatry.org/psychiatrists/practice/dsm.

[5] R. Carette, F. Cilia, G. Dequen, J. Bosche, J.-L. Guerin, and L. Vandromme, Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach, *the International Conference on IoT Technologies for Healthcare*, Springer, Angers, France, 2017, pp. 75–81.

[6] L. Kanner, Autistic disturbances of affective contact, *Nerv. Child*, Vol. 2, 1943, pp. 217–250.

[7] E. Fombonne, Epidemiology of pervasive developmental disorders, *Pediatric Research*, Vol. 65, no. 6, 2009, pp. 591–598.

[8] D. Aarthi, M. Udhayamoorthi, G. Lavanya, Autism Spectrum Disorder Analysis using Artificial Intelligence: A Survey, *International Journal of Advanced Research in Engineering and Technology,* Vol. 11(10), 2020, pp. 235-240.

[9] N. Ajaypradeep, R. Sasikala, Child Behavioral Analysis: Machine Learning based Investigation for Autism Screening and Early Diagnosis, *International Journal of Early Childhood Special Education,* Vol. 13(2), 2021, pp. 1199-1208.

[10] N. V. Ganapathi Raju, Karanam Madhavi, G. Sravan Kumar, G. Vijendar Reddy, Kunaparaju Latha, K. Lakshmi Sushma, Prognostication of Autism Spectrum Disorder (ASD) using Supervised Machine Learning Models, *International Journal of Engineering and Advanced Technology (IJEAT)*, Vol. 8(4), 2019, pp.1028-1032.

[11] Fawaz Waselallah Alsaade and Mohammed Saeed Alzahrani, Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms, *Computational Intelligence and Neuroscience*, 2022, pp. 1-10.

[12] Arjun Singh, Zoya Farooqui, Branden Sattler, Unyime Usua, Michael Helde, Using Machine Learning Optimization to Predict Autism in Toddlers, *11th Annual International Conference on Industrial Engineering and Operations Management*, Singapore, 2021, pp. 6920-6931.

[13] Uğur Erkan1, Dang N.H. Thanh, Autism Spectrum Disorder Detection with Machine Learning Methods, *Current Psychiatry Research and Reviews*, Vol. 15(4), 2019.

[14] Dr. Sherif Kamel, Rehab Al-harbi, Newly proposed technique for autism spectrum disorder based machine learning, *International Journal of Computer Science & Information Technology (IJCSIT),* Vol. 13(2), 2021.

[15] Sriram Dhanyatha , A. Greeshma, Gouthami, M. Yeshwanth, Y Shobha, Prediction of Autism Spectrum Disorder based on Machine Learning Approach, *International Research Journal of Engineering and Technology (IRJET),* Vol. 8(7), 2021, pp. 2907-2917.

[16] Anupam Garg, Anshu Parashar, Dipto Barman, Sahil Jain, Divya Singhal, MehediMasud, Mohamed Abouhawwash, Autism Spectrum Disorder Prediction by an Explainable Deep Learning Approach, *Computers, Materials & Continua*, Vol. 71(1), 2022, pp. 1459-1471.

[17] Basma Ramdan Gamal Elshoky, Eman M. G. Younis, Abdelmgeid Amin Ali, Osman Ali Sadek Ibrahim, Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images, *ETRI Journal*, 2021, pp. 613-623.

[18] P. Moridian1, N. Ghassemi, M. Jafari, S. Salloum-Asfar, D. Sadeghi, M. Khodatars, A. Shoeibi, A. Khosravi, S. H. Ling, A. Subasi, R. Alizadehsani, J. M. Gorriz6, Sara A Abdulla, U. Rajendra Acharya, Automatic Autism Spectrum Disorder Detection Using Artificial Intelligence Methods with MRI Neuroimaging: A Review, *Frontiers in Molecular Neuroscience*, Vol. 15, 2022, pp. 1-51.

[19] Fadi Thabtah, David Peebles, A New Machine Learning Model based on Induction of Rules for Autism Detection, *Health Informatics Journal*, 2020, pp. 1-23.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.