

# NEWLY PROPOSED TECHNIQUE FOR AUTISM SPECTRUM DISORDER BASED MACHINE LEARNING

Dr. Sherif Kamel<sup>1</sup> and Rehab Al-harbi<sup>2</sup>

<sup>1</sup>Associate Professor Department of Communication and Computer Engineering, October University for Modern Sciences and Arts, Giza, Egypt

Head of Computer Science, Arab East Colleges for Graduate Studies, Riyadh, KSA

<sup>2</sup>Master of Computer Science, Arab East Colleges for Graduate Studies, Riyadh, KSA

## ABSTRACT

*The rapid growth in the number of autism disorder among toddlers needs for the development of easily implemented and effective screening methods. In this current era, the causes of Autism Spectrum Disorder (ASD) do not know yet, however, the diagnosis and detection of ASD is based on behaviours and symptoms. This paper aims to improve ASD disease prediction accuracy among toddlers by using the Logistic Regression model of Machine Learning, through the collected health care dataset and by using an algorithm for rapid classification of the behaviours to check whether the children are having autism diseases or not according to information in the dataset. Therefore, Machine Learning decreasing the time needed to detect the disorder, then providing the necessary health services early for infected toddlers to enhance their lifestyle. In healthcare, most machine learning applications are in the research stage, and to take the advantage of emerging software tools that incorporate artificial intelligence, healthcare organizations first need to overcome a variety of challenges.*

## KEYWORDS

*Artificial Intelligence, Machine Learning, Autism Spectrum Disorder.*

## 1. INTRODUCTION

Machine Learning has become one of the important fields and fastest-growing areas of computer science and technology; it is a subfield of Artificial Intelligence (AI). Nowadays, machines are capable of learning like humans, and they are designed to perform tasks. Generally, Machine Learning used to understand the structure of data and fit that data into models that can be understood and utilized by people [1-2].

The field of ML can be implemented in healthcare to make better decision making about patients diagnoses and providing necessary needs for treatments. It aims to decrease the time needed to detect diseases, minimize physicians and health care professionals efforts in the detection of diseases and reducing progression through early discover. Predictive analysis by using ML algorithms help to predict the disease more correctly through processing a huge medical datasets. Many diseases became easy to detect and recognize by using ML such as Heart Diseases Diagnosis and Autism Spectrum Disorder (ASD). ASD is known as a developmental disorder that affects behaviour and communication, this autism can be diagnosed at any age. Generally appear in the first two years of life, in others, symptoms may not show up until 24 months or later. It can be a lifelong disorder, but symptoms may improve over time. Some of the children with ASD seem to develop normally until 18-20 months then stop gaining new skills or they lose the skills they once had. Some people with ASD need a lot of help in their daily lives; others need less. They have difficulties with social and communication interaction with others, these

difficulties can appear as unusual behaviours in people infected. The common behaviours shown on toddlers are they do not respond to name by 12 months of age, avoids eye contact, and do not share interests with others. In addition, the common communication behaviours showed as, delayed speech and language skills, they do not point or respond to pointing and repeat words over and over, some of them can speak well, and others can't speak at all or only very little. There are different behaviours with ASD and variant from child to another, but most will show several. Briefly, health care professionals are still trying to understand the cause of ASD. Research indicates that there are some factors that could lead to ASD, include having a sibling with ASD, having certain genetic conditions, and very low birth weight [3-4].

By using ML we can recognize these behaviours that are related to ASD. Therefore, it contributes to the provision of health services and the necessary needs early for infected toddlers, and early intervention can improve a child's overall development, improve and enhance communication skills and interact with individuals. Early detection can give an autistic child the essential treatment that may include behaviour therapy, speech therapy, occupational therapy, and treatments. Hence, an autistic child can have a better life. The goal is to help children to better communicate, play with others and learn social skills, improve learning and take care of their bodies and health [5].

The following sections will discuss the literature review, the newly proposed Model, System design and implementation with testing and finally the conclusion and the future work.

## **2. REVIEW OF LITERATURES**

In [6], this research presented, Autism Diagnostic Observation Schedule-Generic (ADOS) which is widely used for behaviours evaluation of autism spectrum disorder. It consists of four modules for a specific group of individuals. This module may take between 30 and 60 min to deliver. In ADOS a series of machine learning algorithms used to study the module of ADOS in genetic autism. ADOS-R was sufficient to classify autism with 100% accuracy.

Although autism could have a significant genetic component, it could be diagnosed through behaviours. Diagnosing autism can be indicative through language and communication, social interactions or interests and activities. The ADOS is exam consists of a variety of activities related to autism behaviours. The ADOS is run by clinical professionals, it's duration from 30 to 60 min. The long length of the ADOS exam may contribute to delay in diagnosis. Families may wait as long as 13 months between initial screening and diagnosis. These delays directly affect the delivery of speech and behaviours therapies. There was a desire for the design abbreviated screening examinations that are meant for more rapid diagnosis. This research study focused on the statistical and data-driven of the ADOS questions that could result in an abbreviated and accurate instrument for the classification of autism.

By this mean, the behavioural diagnosis of autism can be effective but time-consuming, the study found that the ADTree performance is more accurate in determining individuals with autism.

In [7], this research focused on recent studies that used machine learning in ASD classification. There are a number of diagnostic tools for ASD, i.e. clinical diagnosis method is Autism Diagnostic Observation Schedule-Revised (ADOS-R) and Autism Diagnostic Interview (ADI). In some studies the goals were reducing the screening time and improving sensitivity, to enhance ASD diagnosis accuracy, by using machine learning methods it offers effective and efficient classification for ASD problems. This study discussed various issues related to ASD screening and focused on recent studies of ML to develop new methods for detecting and classifying of ASD.

In this diagnosis process, the input will be a training of dataset of previous cases, after the data is processed then ML algorithms can be applied. The output of this phase is used to evaluate the effectiveness of the ML algorithms to determine the type of diagnosis. Thus, this study pointed to the need for updating the screening of ASD methods to meet the new criteria of DSM-5 ASD.

ADOS-R is the most screening tool used in clinical to diagnose of ASD due to reliability, validity, and sensitivity, it evaluates an individual's behaviours based on social skills, languages and plays. The examination of ADOS-R takes 30-45 minutes and responses are recorded by the examiner. ADI normally is an interview conducted by clinical professionals with the parent of individuals over 18 months old. In this interview, the parents are given 93 questions about the observed behaviour of the child. It takes one to two hours to be conducted, it aims to shorten the time to be appropriate for children. Decision trees can be employed as a ML method and contain fewer features than using ADOS-R. To shorten the time related to medical diagnosis, this process takes around 30-120 minutes based on, the clinical diagnosis procedure, and the expertise of the clinical professionals.

In [8], this research presented, detecting autism using ML by predicting if a child has ASD through observe developmental delays, speech, and learning disability as ASD behaviours. Previous research showed that ASD can be detected reliably in children as young as 18 months by using 2 stages of screening strategies, and showed that the average age of the first diagnosis around 3-6 years. The ASD can be detected at 18 months, so there will be a gap between possible age at first diagnosis. It is better to develop a decision support system, to help physicians estimate the developmental status of children more accurately. Another study showed that children with very low birth weight (<1500 g) had 3.2 times higher odds of autism/ASD than normal birth weight children. They also determined that children born prematurely had 2.3 times higher odds of autism/ASD than term children. The conclusion that ASD relates to developmental delay, premature birth, and very low birth weight.

In this study, many classification algorithms that are used are Naive Bayes, Random Forest, and Support Vector Machine to ensure the results are reliable. The goal was to identify whether the screening children has ASD or not and how severe the ASD might be, which can be used by physicians through mentioned attributes. With decision making the detecting is more easier and earlier than currently.

In [9], it is worthwhile to consider ML played an important role in terms of data analytics. It can be implemented in the health care field, this will lead to the enhancement of ML because of huge data being analysed and processed. By using techniques of ML it can create a model for prediction and identification of ASD. There is a demand for automated processing to minimize the time and cost of any field. Healthcare is one of these important fields. Nowadays, diagnosing time is a big challenge for many health cases, especially autism. The diagnosis of autism takes 6 months, ML aims to accelerate the process, speeding the diagnosing time for early intervention helps to improve Autistic children. A huge amount of health and medical data can be used in predictive modelling. This study used the supervised learning methods. Machine Learning allows computer programs to access data and learn from themselves without human intervention.

In [10], this research showed that ML has been successfully applied to shorten the duration of the diagnosis and to increase the performance of the diagnosis of the ASD disease and identify the behavioural disorders. This study used support vector machines (SVM), k-nearest neighbors (kNN), and random forest (RF) ML methods to perform a fast and accurate diagnosis of ASD. Autism varies greatly depending on age and skill. The prevalence of autism seen as a rare event in the past. The frequency of occurrence in males is 4 times higher than that of females. Often ASD results dependent on language and intelligence and it varies from person to another. Besides

clinical methods, there are parent-based nonclinical methods such as Autism Spectrum Quotient (AQ). ML has tremendous abilities in screening and intervention studies in behavioural sciences and in order to provide faster access to health services, increase diagnostic performance. This study aimed to provide a quick preliminary diagnosis through a questionnaire contains effective questions in the diagnosis of ASD and answered by healthcare professionals.

In this study, SVM, kNN and RF machine learning methods are used to analyse the answers given to a questionnaire which includes questions proved to be effective in differentiating the ASD cases. Accuracy, sensitivity, and specificity performance metrics are used for the evaluation of the trained models. Accuracy rates of 95%, 89%, and 100% were achieved as a result of binary classification using SVM, kNN and RF methods, respectively. The lowest performance in the classification of ASD is kNN method. The results obtained with the RF method showed that this method can classify ASD cases with complete success.

Most of the above research techniques provides an accurate performance in determining individuals with Autism but are of time consuming. The newly proposed system will be applied to predict accurately and save time in predicting the Autistic cases. Table 1 shows the pros and cons of the logistic regression model used in the newly proposed system.

Table 1. Pros and Cons of the Logistics Regression Model

Advantages	Disadvantages
It is easier to implement and very efficient to train.	It is tough to obtain complex relationships using logistic regression.
Good accuracy for many simple data sets.	In Linear Regression independent and dependent variables are related linearly
Logistic regression is less inclined to over-fitting.	It constructs linear boundaries.
Very fast at classifying records.	If the number of observations is lesser than the number of features, Logistic Regression should not be used, it may lead to overfitting.

### 3. PROPOSED SYSTEM MODEL

#### 3.1. Overview

The goal of the proposed model is to build and apply ML algorithms and tools to accelerate the process of predicting the ASD and to alert healthcare professionals on the probabilities of the autism disease existence. The proposed model is based on using the data sets collected based on patients' historical records and common behaviours of autism shown on toddlers between 6 and 24 months of age. By using supervised ML methods the datasets of the model will be trained and tested to generate accurate and correct detection of autism and this will contribute to shortening the time of screening and providing early behaviour-oriented therapeutic programs that will help the child to get rid of unwanted behaviours and modify their lifestyle without any help. The model is built to address a variety of challenges through the latest ML algorithms and methods, in order to improve the healthcare of autism. When deploying this model that will lead to help and serve healthcare professionals and save their effort during screening and also reduce the necessary time for diagnosis, in addition, to help parents to use this screening model if they notice any unusual behaviours appear in the toddler.

Figure.1 shows the block diagram of the proposed model. It shows the schematic and the visual representation of the model parts and components. It consists of blocks that represent components and connected by lines to show the relationship between them.

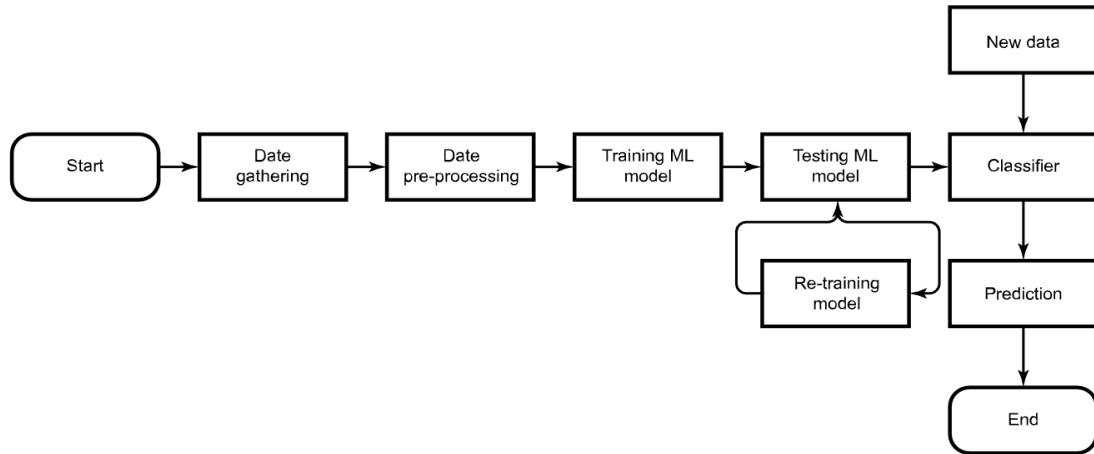


Figure 1. Block Diagram of System Model

**Data gathering:** Data gathering has a significant factor in solving the problem and this will allow capturing the record of previous cases of autism. The classifier can be more efficient as it will be based on the dataset from which it was built.

**Data pre-processing:** Data pre-processing is an important step in ML, the data requires pre-processing in order to yield useful insight into the data before feeding it to the algorithms. This process includes data cleaning, data reduction and data transformation in order to achieve better results from the model.

**Training model:** The model is trained using a training dataset by a supervised learning method, typically contains examples or samples used to fit the model which learns from these data. This involves ML algorithms that used to build a predictive model.

**Testing model:** The test model is used to evaluate the performance and estimates how well the model is trained and better fits. Typically, provides the final estimate of validating the model.

**Re-training model:** re-training comes after the testing model if the model does not fit well with data and has a desire to re-train the dataset to meet the goals and tasks. Therefore, choosing the right models of prediction and classification data contribute in building an accurate and reliable model.

**New data:** This is the actual input data that must be ready to classify. These data are transformed into the classifier in order to get the correct prediction based on historical and trained data.

**Classifiers:** They are algorithms tending to predict the class of data. The classes are called labels/targets or categories, it uses supervised learning to decide which data learned from the input data.

**Prediction:** It is the final step for the processes, expected to give the right predictions of the input data and try to predict labels and create the desired classifications for each entered data.

### 3.2. System Flowchart

Figure.2 shows the flowchart of the proposed model and provides clear information about the sequence of operations of the ML model that must be followed in order to build an efficient model.

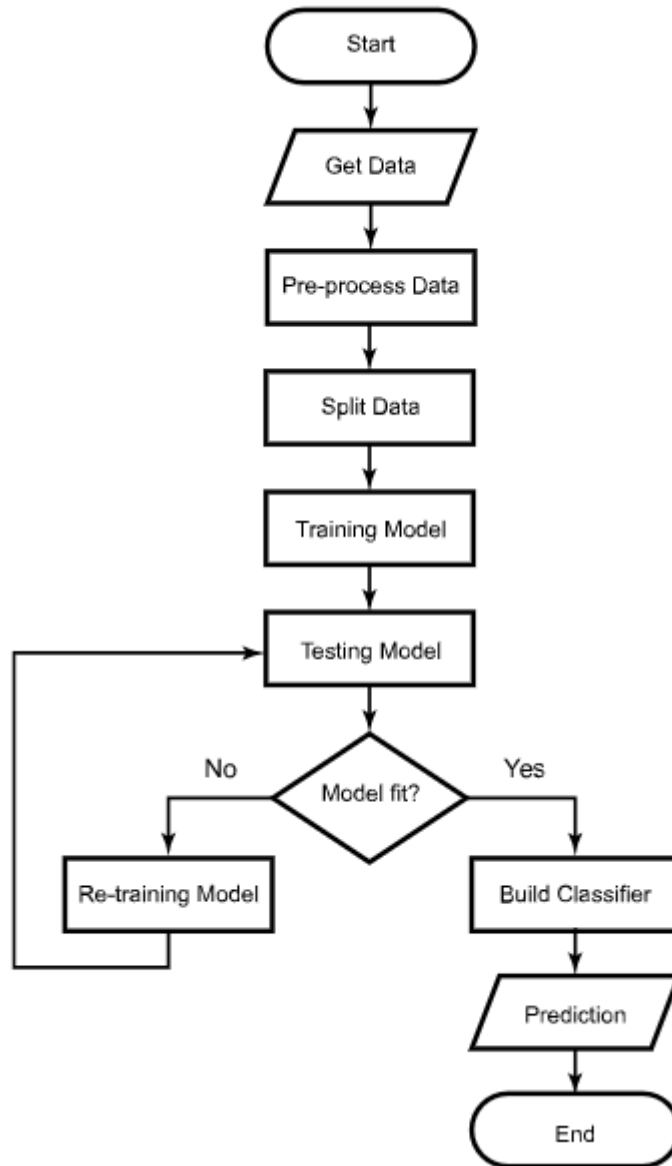


Figure 2. Flowchart of System Model

The proposed model is based on building a new system model that will help in detecting the ASD because detection via screening is time-consuming. With the capabilities of ML, autism can be predicted at an early stage and by using classification algorithms we can get an accurate and efficient result. The proposed model assumed to show the results in terms of accuracy and high performance.

At the beginning, the dataset should be gotten from the repositories or sources about ASD, then these data will be prepared for building the model by using algorithms to classify the outcomes of the new data. These data must be processed before fitting into the model. Then, under supervised learning in ML, the dataset will be split into training and testing sets. The training set is used to validate the actual model algorithms and it is typically between 75%-80% of data. The testing dataset is used to test the accuracy of the model algorithms and it typically contains 20%-25% of data.

Re-training the model will be done once again with other algorithms to get the best results if the algorithms don't fit well. After that, the classifier will be built to take the new input data from the user and then classify to check whether the child has ASD or not. Finally, the prediction process of the model must give us correct and accurate classification.

## **4. SYSTEM DESIGN AND IMPLEMENTATION**

### **4.1. System Tools**

This section is dedicated to the system hardware and software requirements that are required to implement and configure the system in order to run the proposed model efficiently.

#### **4.1.1. System Hardware**

Computer hardware is a technical description of the computer capabilities and components that used to execute and implement the system or the model. The proposed system needs a computer with the following specifications:

- Operating system: Windows 10 64-bit
- Processing speed: Intl (R) Core (TM) i5-7200U CPU @ 2.50GHz 2.71 GHz
- Random Access Memory: 8.00 GB
- Network (Wi-fi): 802.11b WLAN.

#### **4.1.2. System Software**

Software tools include the programming languages that will be used to develop and construct the system or the model to meet the objectives, these tools are:

- Anaconda navigator: is a free and open-source Python distribution made for large-scale data processing, predictive analytics, and scientific computing. It is available for Mac OS, Windows, and Linux [11]. It contains a variety of built-in libraries and provides wide and sophisticated tools for data science, machine learning and deep learning and more.
- NumPy: it is a fundamental library in Python; it contains a collection of tools that can be used in a mathematical problems functionality for multidimensional arrays, to faster access to items NumPy being more efficient and more convenient.
- Pandas: is a high-level data manipulation built on NumPy using data structure called the DataFrame. Simply, a pandas DataFrame is a table, similar to an Excel spreadsheet that allows converting tabular data into rows and columns.
- Seaborn: it is a matplotlib library for data visualization, it provides a drawing and information for a statistical graphics.
- Python it has a power of general-purpose programming languages with the ease of use of domain-specific scripting languages like MATLAB or R. Python that contains important

libraries for data such as loading, visualization, statistics, natural language processing, image processing, and more.

- Scikit-learn: it contains a number of machine learning algorithms, as well as documentation about each algorithm. It is very popular tool and the most common Python library for machine learning [11].
- Spyder: in addition to the built-in libraries, Anaconda has a Spyder which is a powerful scientific python development environment, it was written in Python and designed for scientists, engineers and data analysts, with advanced editing, interactive testing and analysis.
- Flask: it is a web micro-framework with python for building web applications and provides tools, libraries, and technologies that make the processes for designing web applications simpler. Flask will be helpful for deploying a machine learning model into the web for use by individuals.
- Hypertext Markup Language - revision 5 (HTML5): is a software solution that defines the structures, properties and behaviors of web pages contents.
- Cascading Style Sheets (CSS5): is used for presenting a contents written in a markup language and set the style of user interfaces.
- Bootstrap: is an open-source CSS framework designed for front-end web development. It contains designing templates for forms, buttons, navigation and other interface components.

#### 4.2. System Implementation and Testing

Because ASD diagnosis takes a long time, in addition to, the rapid growth of the number of cases in ASD, the detection of this disease became easier than before according to machine learning by recognizing the common behaviors that appear on toddlers. The early diagnosis and detection of ASD can aid to change the lifestyle of toddlers by using the ML model. In the proposed system the Logistic Regression is used to predict the outcome of model as well as to determine the most related risk factors of ASD disease. The Logistic Regression is a type of statistical analysis ,it predicts categorical output based on dependent variable ( Target Variable) from an independent variable ( Feature Variable). In this type of machine learning algorithm that uses supervised learning, the dependent variable must be binary as the result should be in the form of 0 or 1, Yes or No. It also used to calculate the probability. The model is chosen based on the dataset, and the data used in training had two values either 0 or 1 which is categorical data. The prediction or output of model should be binary either 0 or 1, based on observed characteristics [12]. The implementation will be as Shown in figure 3:



Figure 3. Model Sequence



The dataset as shown in table 2 is screening for a group of toddlers related to autism that contained common features or behaviours to be used in analysis and recognized autistic. In this dataset, ten recorded behaviours show the effectiveness in determining the ASD. The aims of this classification are to predict the child has a probability of ASD diseases, and to contribute to early intervention. The dataset contains 1054 instances, 18 attributes and it completely does not contain any missing values.

Table 2. Snapshot of Distribution Dataset.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Qchat_10_Score	Sex	Ethnicity	Jaundice	Family_mem_with_ASD	Who_completed_the_test	Class_ASD
0	0	0	0	0	0	1	1	0	1	28	3	f	middle eastern	yes	no	family member	No
1	1	0	0	0	1	1	0	0	0	36	4	m	White European	yes	no	family member	Yes
1	0	0	0	0	0	1	1	0	1	36	4	m	middle eastern	yes	no	family member	Yes
1	1	1	1	1	1	1	1	1	1	24	10	m	Hispanic	no	no	family member	Yes
1	1	0	1	1	1	1	1	1	1	20	9	f	White European	no	yes	family member	Yes
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0	0	0	0	0	0	0	0	0	1	24	1	f	White European	no	yes	family member	No
0	0	1	1	1	0	1	0	1	0	12	5	m	black	yes	no	family member	Yes
1	0	1	1	1	1	1	1	1	1	18	9	m	middle eastern	yes	no	family member	Yes
1	0	0	0	0	0	0	1	0	1	19	3	m	White European	no	yes	family member	No
1	1	0	0	1	1	0	1	1	0	24	6	m	asian	yes	yes	family member	Yes

Figure 4 shows, the visualization of the classification or the target class of toddlers with ASD in the dataset. This classification is represented by yes or no. The figure shows the toddlers that have ASD are 728 toddlers, as well as 326 toddlers have no autism spectrum.

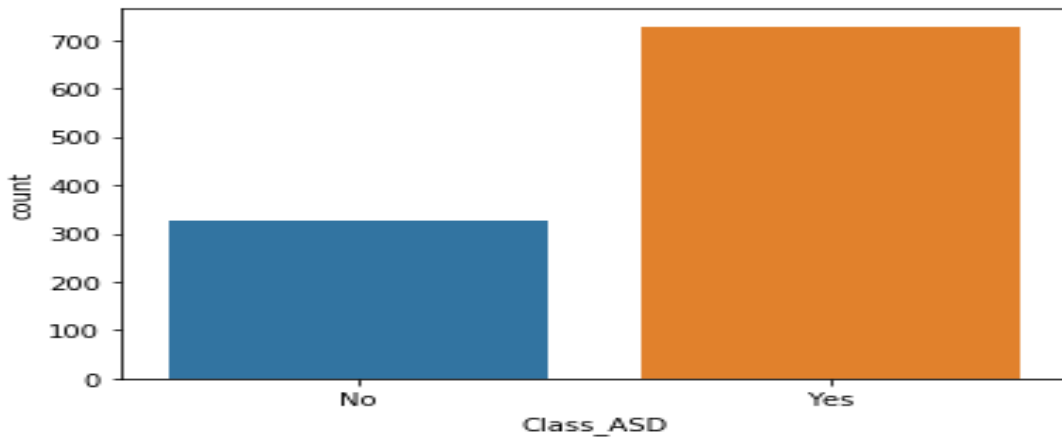


Figure 4. Total of toddlers with ASD

Figure 5 shows, whether the toddlers families have more than one member with ASD or not. The goal of this classification is to know and analyse the factors for this disease and whether it is related to the inheritance factor or not. As shown in the figure, many families do not have a member with autism, where 0 represents no and 1 represents yes. The count of this feature is 884 families have no members, and 170 families have more than one member infected with ASD. As shown it is observed that autism may not an inheritance factor.

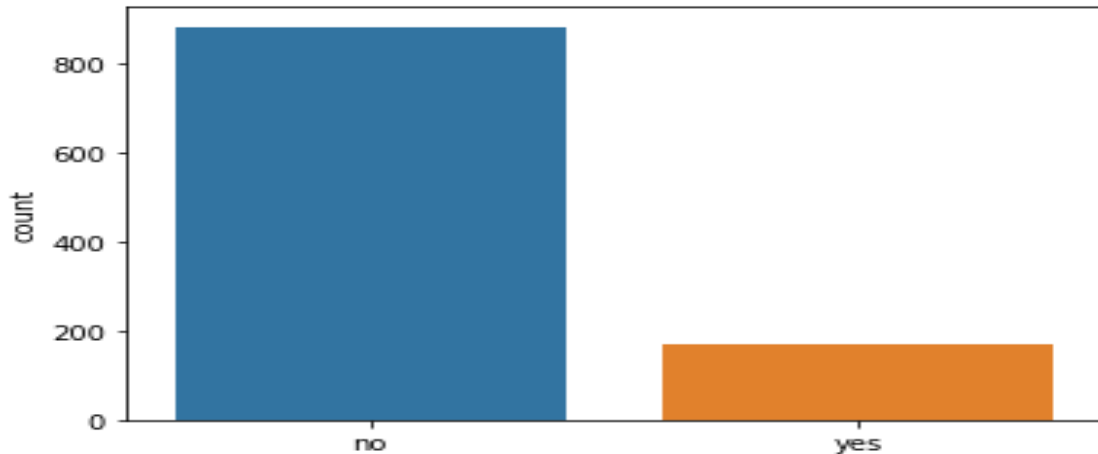


Figure 5. Total of families member with ASD

Figure 6 shows the ratio of the females infected with ASD is less than males, the ratio is about 4 males for every 1 female diagnosed. Recently, the researches in the United States mentioned that the ratio gets higher to 1 case in every 42 males and 1 case in every 189 females. Therefore, the data in the figure confirm that the males have a higher risk of infection with ASD.

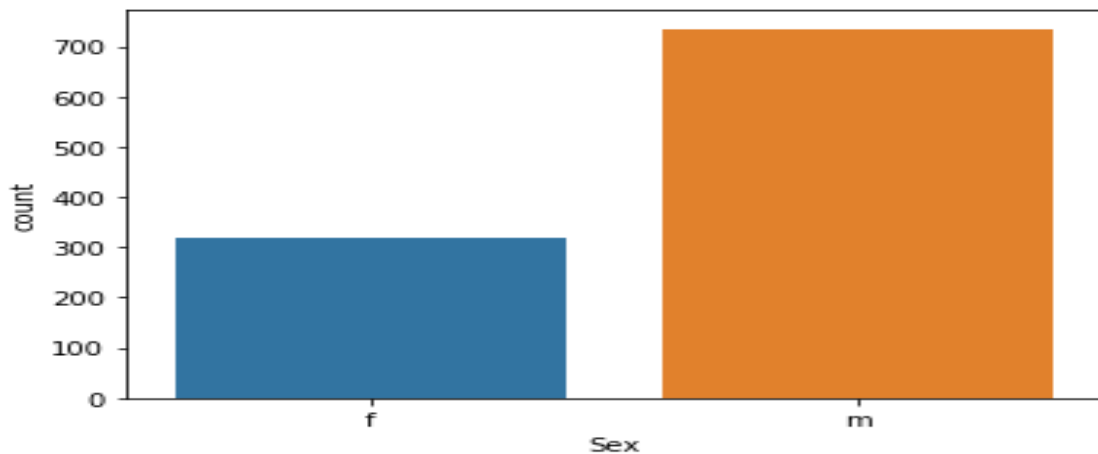


Figure 6. Total of toddlers with ASD based on sex

Figure 7 shows, the infected toddlers based on age in months. As shown the toddlers above 30 months have a higher probability to be infected with ASD than others which is between two and three years. The researchers found that the probability of autism spectrum symptoms and behaviours appears in the first two years of toddlers. From the graph, it is observed that, the screening for toddlers in this age is very important due to the high probability of autistic.

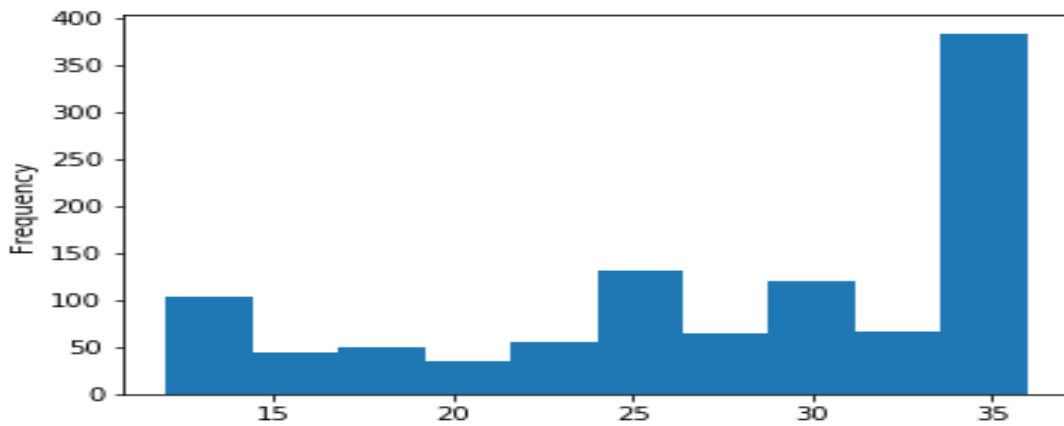


Figure 7. Total of autistic toddlers based on age

Before start building the data should be processed to make sure it is free of null values, complete and clean. This step is very important in order to get the best accuracy of the machine learning model and predict the resulting from the classifier. The model used a dataset contain 1054 rows and 19 columns. As shown in figure 8, the dataset is complete and has no null values.

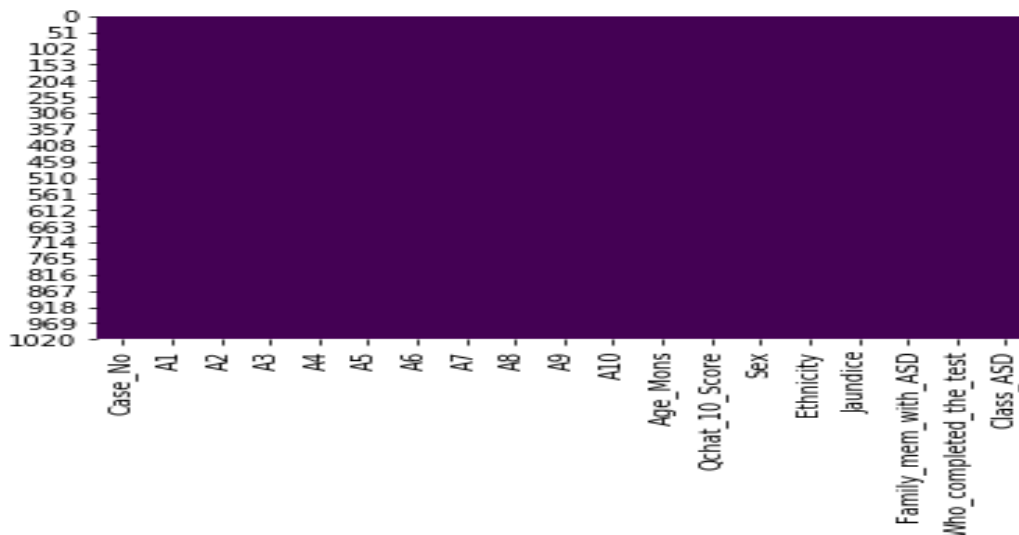


Figure 8. Plotting of dataset

The model achieves an accuracy equal 1, which is effective and has high efficiency in determining the autistic cases. Therefore, this model is very helpful for both individuals and healthcare professionals.

The ML model turned into Application Programming Interface (API) using Flask as a web service in order to predict the new data. The predictor or classifier should return whether the child has ASD or not. Figure 9 represents the prediction API in which a new data will be passed to predict function in the model, it contains 10 questions about the common behaviours for autism, the answers must be 1 or 0, then the prediction shows the result of these answers.

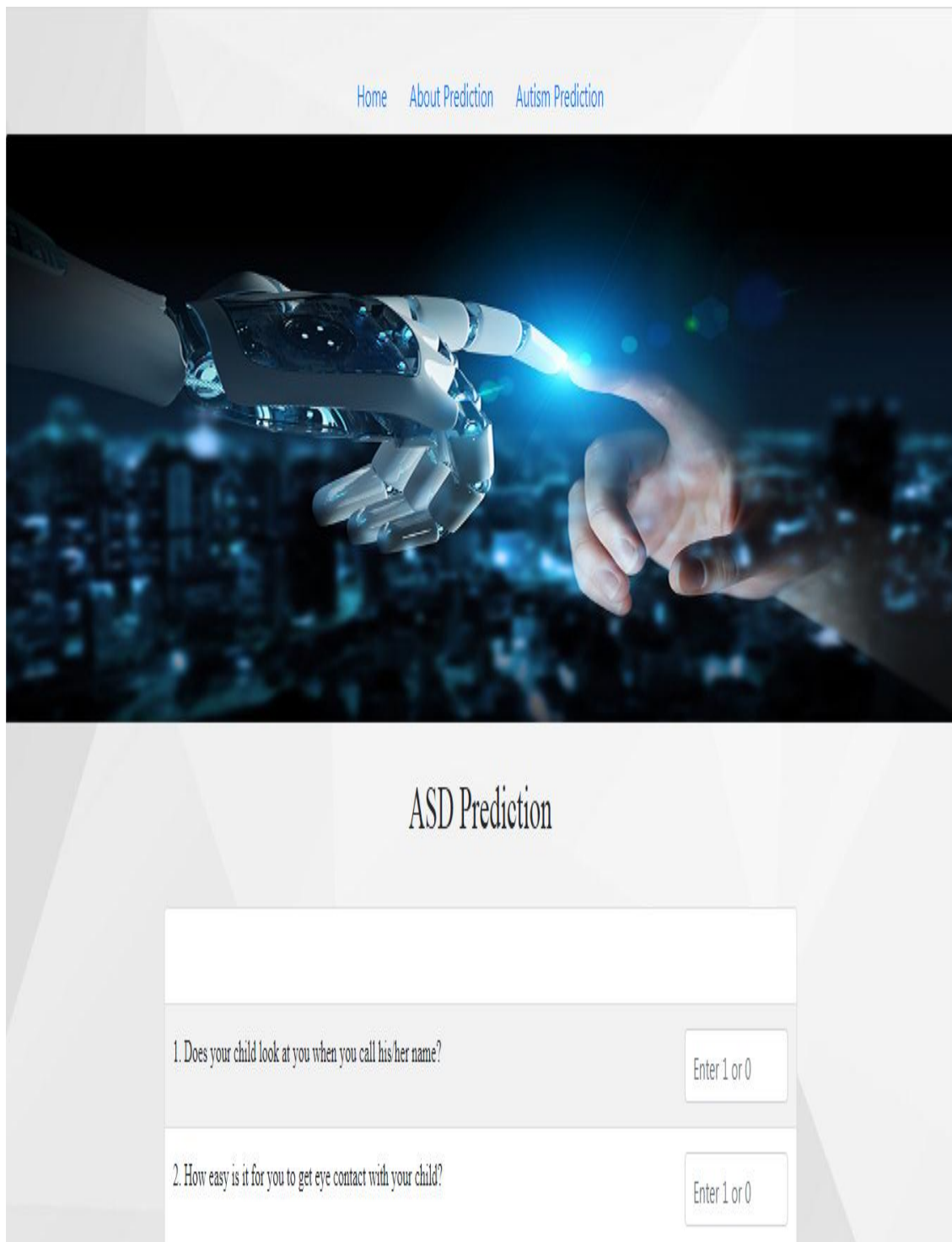


Figure 9. Snapshot- Predictor of ASD

Figures (10, 11) illustrate the results of two different cases. It is based on the score of questions, if it is more than 3 then the child has ASD, else not has ASD. This results according to a previous training medical history dataset to identify the risk factors. The predicted model has an accuracy 1.0.

### ASD Prediction

1. Does your child look at you when you call his/her name?	<input type="text" value="Enter 1 or 0"/>
2. How easy is it for you to get eye contact with your child?	<input type="text" value="Enter 1 or 0"/>
3. Does your child point to indicate that s/he wants something? (e.g. toy that is out of reach)	<input type="text" value="Enter 1 or 0"/>
4. Does your child point to share interest with you? (e.g. pointing at an interesting sight)	<input type="text" value="Enter 1 or 0"/>
5. Does your child pretend? (e.g. care for dools, talk on a toy phone)	<input type="text" value="Enter 1 or 0"/>
6. Does your child follow where you're looking?	<input type="text" value="Enter 1 or 0"/>
7. If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them)	<input type="text" value="Enter 1 or 0"/>
8. Would you describe your child's first words as:	<input type="text" value="Enter 1 or 0"/>
9. Does your child use simple gestures? (e.g. wave goodbye)	<input type="text" value="Enter 1 or 0"/>
10. Does your child stare at nothing with no apparent purpose?	<input type="text" value="Enter 1 or 0"/>

The score indicates, the child has ASD

Figure 10. Snapshot of the Case with ASD

## ASD Prediction

1. Does your child look at you when you call his/her name?	<input type="text" value="Enter 1 or 0"/>
2. How easy is it for you to get eye contact with your child?	<input type="text" value="Enter 1 or 0"/>
3. Does your child point to indicate that s/he wants something? (e.g. toy that is out of reach)	<input type="text" value="Enter 1 or 0"/>
4. Does your child point to share interest with you? (e.g. pointing at an interesting sight)	<input type="text" value="Enter 1 or 0"/>
5. Does your child pretend? (e.g. care for dolls, talk on a toy phone)	<input type="text" value="Enter 1 or 0"/>
6. Does your child follow where you're looking?	<input type="text" value="Enter 1 or 0"/>
7. If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them)	<input type="text" value="Enter 1 or 0"/>
8. Would you describe your child's first words as:	<input type="text" value="Enter 1 or 0"/>
9. Does your child use simple gestures? (e.g. wave goodbye)	<input type="text" value="Enter 1 or 0"/>
10. Does your child stare at nothing with no apparent purpose?	<input type="text" value="Enter 1 or 0"/>

The score indicates, the child has no ASD

Figure 11. Snapshot of the Case with no ASD

## 5. CONCLUSION AND FUTURE WORK

This paper aims at developing a screening model using ML techniques for detecting and assessing the behaviors associated with ASD among toddlers age 12 - 30 months. Early detection ASD can help to limit the challenges that are facing children such as communications, social skills, learning disabilities, daily problems and many others. The goal is to apply supervised learning algorithms for a dataset from previous medical cases that can predict the autistic among toddlers. The uses of the new screening methods can encourage the use of technology in clinical environments for autism and empower clinicians with tools that provide useful knowledge for better decision making. The future work will be dedicated to the prediction using improved technique and algorithm for classification and comparing the results with other different models. In addition to that using more data and features including more common behaviours for detecting ASD in accurate manner.

## REFERENCES

- [1] Vaishali, R., and R. Sasikala. "A Machine Learning Based Approach to classify Autism with Optimum Behaviour Sets", (2018) International Journal of Engineering & Technology 7(4): 18.
- [2] Thabtah Fadi, Firuz Kamalov, Khairan Rajab "A New Computational Intelligence Approach to Detect Autistic Features for Autism Screening", International journal of medical informatics, 117 (2018), pp. 112-124.
- [3] Thabtah, Fadi. "Machine Learning in Autistic Spectrum Disorder Behavioural Research: A review and ways forward", (2018), Informatics for Health and Social Care: 1-20.
- [4] M.S. Mythili, AR Mohamed Shanavas "A study on Autism Spectrum Disorders using Classification Techniques", International Journal of Soft Computing and Engineering (IJSCE), 4 (2014), pp. 88-91.
- [5] Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, Todd F. DeLuca" Use of Artificial Intelligence to Shorten the Behavioural Diagnosis of Autism. PloS one, 7 (8) (2012), p. e43855.
- [6] Dennis Paul Wall, J. Kosmicki, T.F. Deluca, E. Harstad, Vincent Alfred Fusaro "Use of Machine Learning to Shorten Observation-based Screening and Diagnosis of Autism", Translational psychiatry, 2 (4) (2012), p. e100.
- [7] F. Thabtah, "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment", Journal of ACM Digital Library, No. 6, May 2017.
- [8] Bekerom, "Using Machine Learning for Detection of Autism Spectrum Disorder", 26th Twente Student Conf. on IT Feb 3th, No. 7, February 2017.
- [9] H. Alarif and G. Young, "Using Multiple Machine Learning Algorithms to Predict Autism in Children", proc. of International Conf. on Artificial Intelligence (ICAI'18), No. 4, 2018.
- [10] A. DEMIRHAN, "Performance of Machine Learning Methods in Determining the Autism Spectrum Disorder Cases", Mugla Journal of Science and Technology, No. 6, June 2018.
- [11] A. Muller, S. Guido, Introduction to Machine Learning with Python, O'Reilly Media, Sebastopol, 2016.
- [12] David W. Hosmer jr, Stanley Lemeshow, Rodney X. Sturdivant , "Applied Logistic Regression", 3rd Edition – 2013.