International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

Diagnosis of Autism using Machine Learning as a Healthcare Technology

Camellia Ray

Abstract: ***Autism is one of the inborn disease, researchers are presently focusing on. The autistic child faces inflexibility in language, thinking and behavior together with the difficulties in understanding emotional states of others. There are lot of interventions going on to make them understand the feelings of others and vice-versa. Now a day, ASD became one of the quick spreading diseases all over the world. Therefore there is a huge need to provide a time-consuming and easy accessible diagnostic tool to detect autism at an early stage to help the clinicians in providing prior medications. Though there is no proper curability of autism, still easy detection helps to provide better therapy session and supports the autistic child to lead a comfort independent life. The thesis deals with the building up of a model where the parents and relatives of a suspected autistic child can easily detect if they are suffering from autism by providing their answers of some particular questions related to the characteristics of autism. In order to build that model, the data were collected manually from different autism therapy centers in India and those raw data are then classified by using three different classifiers namely Logistic Regression, Support Vector Machine and Random Forest with Python as a programming tool to find out the one with higher accuracy by various analyses after pre-processing. The Random Forest classifier with the highest accuracy is utilized in framing the question based model for the early discovery of autism which can be operated as a primary diagnostic model to assist medical professionals technologically.***

Keywords: Autism, Diagnosis, Random Forest, Logistic Regression, Support Vector Machine, 10-fold cross validation

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a brain developmental disorder which begins in the early childhood days and lasts throughout the entire life of a person. The "Spectrum" in Autistic Spectrum Disorder refers to the variations in its type and severity of symptoms which people experience. The disorder affects the acting and interacting skills of an ASD person and possesses restrictive interest and repetitive behaviors [22]. The affected person does not find themselves comfortable to look into the eyes of a person with whom they are talking to and often seem to be in their own world [8]. The symptoms for autistic spectrum disorder generally appear in the first two years of life. It may happen sometimes that some of the children develops normally but gradually starts losing their previously gained skills while attaining toddlerhood. According to the Centers for Disease Control (CDC), one of every 59 children is evaluated to have autism and they are three to multiple times more typical in boys than in girls [12].

Table- I: Characteristics of different types of Autistic Spectrum Disorders [7, 26, 27]

Disorders	Characteristics
Pervasive Developmental Disorder	Hesitation in Eye contact Prefer being alone Being aggressive unintentionally Poor imagination skills Interaction difficulties
Child Disintegrative Disorder	Lack of performing normal function like communication and social interaction Fails in interaction while want to do it Proficient knowledge and information Good language skills but lack in understanding irony of languages Delays in motor skills
Asperger Syndrome	Inability to begin or continue conversations Difficulty in establishing peer relations Repetitive behaviors Abnormal EEG signals Stops speaking, what they regularly used to
Rett Syndrome	Symptoms changes while getting older Problems in sitting, walking, crawling Speech delay Head of small size with slower brain growth Full of anxiety with less words to say
Autistic Disorder	Issues on understanding Lack in learning skills Avoid eye contact Difficulties in expressing what they need Behavioral changes due to stress and anxiety

A. Symptoms of Autism

As the name defines the degree of incapacity and the mix of manifestations fluctuates hugely from individual to individual. Those symptoms can change over time and may ranges from mild to severe [1]. As a result the severities can be sometimes difficult to determine. The categories can fall into certain points:

- Delayed speech and language skills
- Does not responds when call with their names [4]
- Has poor eye contact and lacks facial expressions
- Difficulties in to and fro conversation
- Faces difficulties in recognizing non-verbal cues
- Repeats words or phrases verbatim
- Unusual tone of robot-like voice
- Trouble in understanding the perspective of others and anticipate their activities
- Having enduring extraordinary enthusiasm for specific subjects

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author
Camellia Ray*, Computer Science, Birla Institute of Technology & Science Pilani, Hyderabad, India. Email: camellia.jhilik@gmail.com

- Trouble in adapting new routines
- Repetitive body developments like hand fluttering, shaking and turning [28]

B. Causes of Autism

Though the specific reason for Autism is obscure, still the researcher suggests that both genes and environment plays an important role. According to the brain scan report, the shape and structure of autistic patients are different from the neuro typical ones. Though the exact pattern of inheritance is still unclear, but there is a higher risk of autism in siblings or monozygotic twins [2]. There are several susceptible genes within specific chromosomal region like 7q22-q33 and 15q11-q13 which are still under research. Viral infections, air pollutants, medications or complications during pregnancies play a vital role in causing autism. Medical issues like gastrointestinal issue, seizures or dozing issue and emotional wellness challenges like tension, sadness and consideration may impact the advancement of autism [3]. Autism may also happen because of advanced aging of parents, illness during pregnancies, complications during birth and crisis of oxygen supply in the brain while birth or premature babies [5].

C. Diagnosis of Autism

Since there is no clinical test for autism, the analysis depends on the talking and acting aptitudes of people in contrast with other offspring of same age. Though autism can be detected at the age of 18 months and younger, but a reliable diagnosis by professional experts is suitable in the age of 2 years [6]. The American Academy of Pediatrics suggests that youngsters be screened for formative issue at well-kid preventive visits before age three where a pediatrician often takes a yes or no survey to find the signs for autism[25]. The early identification provides opportunities to take up appropriate interventions for maximizing the future potential of a child by reducing the risk of behavioral difficulties [19]. Mainly the diagnosis of autism takes place in two steps:

- Developmental Screening: Here the doctors interact with the parent and caregivers of child in their early month to know if they are facing delays in learning the basic skills when they should.
- Comprehensive diagnostic evaluation: It is considered to be the second step of developmental screening where the hereditary testing, neurological testing, hearing and vision screening and other clinical testing are processed along with the interviewing of parents for getting a brief study of developmental history.

Sadly, the waiting time for ASD conclusion is long and the methodology are not financially savvy. With the rapid growth of ASD cases all over the world, there is an urgent need of developing time-efficient ASD screening methods which will help the Health professionals to pursue clinical diagnosis [20]. Some of the diagnostic tools are as follows:

- Diagnostic Interview for Social and Communication Disorder (DISCO)
- Autism Diagnostic Interview Revised (ADI-R)
- The Developmental Dimensional and Diagnostic Interview (3di)
- Autism Diagnostic Observation Schedule (ADOS)

Several scales are also there for various psychometric tests to provide better diagnosis by assessing the cognitive and verbal ability [23]. The scales include:

- Wechler Scales
- Mullen Scales
- Merrill Palmer Scales
- Bayley Scales

Now a day, the handcrafted rules are replaced with the emergent of machine learning techniques for ASD detection with a predictive model. The technique helps to reduce the screening time by improving sensitivity and specificity [26].

II. METHODOLOGY

A. Dataset

The dataset consists of 3741 samples of children from 4 to 25 years of age with their 21 attributes who are suspected to be suffering from autism [9]. Out of those 21 attributes, 10 of them are the set of questionnaires relating to behavioral traits and 11 are the characteristics of individuals that have proved to be effective for diagnosis of autism [21].

Table- II: Name of the Table that justify the values [3]

Attribute	Question sets
A1_Score	Frequently hears voices that others do not hear
A2_Score	Spotlights on the comprehensive view by going out from little subtleties
A3_Score	Effectively follow the discussions of various individuals in a social gathering
A4_Score	Can without much of a stretch switch between various exercise
A5_Score	Doesn't have the foggiest idea how to chat with peers
A6_Score	Useful for ordinary short chats
A7_Score	Hard for the characters to comprehend the sentiments while perusing a story
A8_Score	Likes to mess around that should be imitated with other youngsters during pre-school instruction
A9_Score	Effectively comprehend what others think and feel by simply taking a gander at their countenances
A10_Score	Hard to make new friendships

Table- III: Attributes related to characteristics of individuals with their respective values

Attribute name	Values
Age	4-11
Gender	Male, Female
Ethnicity	White European, South Asian, Asian, Middle Eastern, Pasifika, Hispanic, Turkish, Latino, Black, Others, Unknown
Jaundice	Yes, No
Autism	Yes, No
Country of res	52 different countries
Used app before	Yes, No
Result	0-10
Age description	Range
Relation	Parent, Relative, Health care professionals, Unknown
Class	Yes, No

- ❖ Age refers to the number of years of an individual.
- ❖ Gender denotes the sex.
- ❖ Ethnicity means the national or cultural tradition.
- ❖ Jaundice indicates if the individual suffered from jaundice after birth.
- ❖ Autism signifies the genetic factor to know if any of the family members is autistic.
- ❖ Country of res marks the place from where the individual belongs.
- ❖ Used app before typify if any of the child has used autistic detection app before.
- ❖ Result coveys the total score of individual.
- ❖ Age description shows the range of age of the participant.
- ❖ Relation implies the correlation of the suspected child with the responder.
- ❖ And finally, the binary class represents if the person is suffering from autism or not labeled by Yes or No.

From the total of 3741 samples 3000 samples were obtained from the National Institute for Health Research (NHS). NHS is the biggest national clinical research funder in Europe that subsidizes top notch research to improve wellbeing. The reserve, prepares and underpins wellbeing specialists by giving world-class inquire about offices.

The following dataset was also used in a study conducted by [1].

Some 741 real time samples are regional and real-time data that were collected from the Centre for Autism Therapy, Counseling and Help (CATCH), Odisha and other hospitals in Kolkata and Odisha.

Most of the data were taken either from the parents or caregivers of the individual patients as very few of the individuals have responded by their own.

All the 10 questionnaires A1-A10 have potential answers: "Consistently, Usually, Sometimes, Rarely and Never" things' qualities that are mapped to "1" or "0" in the dataset. On the off chance that the reaction was Sometimes/Rarely/Never, 0 is doled out to the inquiry (A1-A10). Nonetheless, in the event that the reaction was Always/Usually/Sometimes, at that point "1" is alloted to that question.

B. Pre-processing

In the huge size of database there is a high chance of having noisy, missing and inconsistent data that leads to low quality mining. The data thus collected in raw format from multiple sources are not feasible for the analysis. So a transformation is necessary before feeding it to the algorithm to make a feasible model that is termed as preprocessing [10].

Since the dataset consists of attributes with mixed data, the categorical and nominal like attributes are fitted and transferred into numerical values by the help of Encoder for a better predictive model in python. Thus the attributes like Gender, Jaundice, Used app before, ethnicity and country of residence are replaced with their new encoded data since python is not comfortable in classification with categorical attribute for a predictive model.

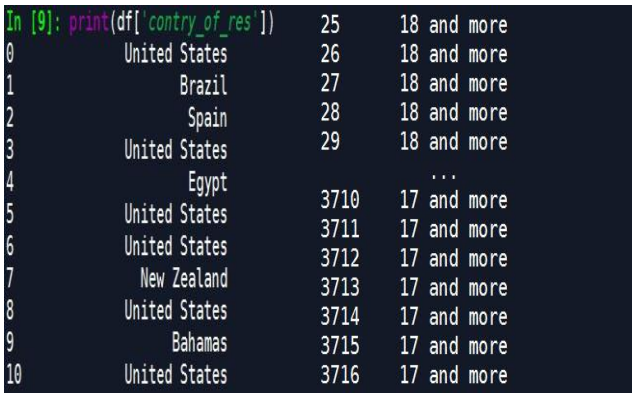
Figure 1 shows the encoded values of Jaundice, Gender, Country of Residence

	jundice		gender		contry_of_res
0	0	0	2	0	113
1	0	1	1	1	24
2	1	2	1	2	98
3	0	3	2	3	113
4	0	4	2	4	41
5	1	5	1	5	113
6	0	6	2	6	113
7	0	7	1	7	77
8	0	8	1	8	113
9	1	9	1	9	15
		10	1	10	113
				11	27

Fig. 1. New encoded values of Jaundice, Gender, Country of Residence

Handling missing data is equally important as machine learning algorithm does not support missing values to exhibit inaccuracy and error in building the model. Thus the missing values in the dataset were replaced with the most frequently occurring values of the particular column.

Figure 2 gives the view of samples after replacing the missing data



	contry_of_res		
0	United States	25	18 and more
1	Brazil	26	18 and more
2	Spain	27	18 and more
3	United States	28	18 and more
4	Egypt	29	18 and more
5	United States		...
6	United States	3710	17 and more
7	New Zealand	3711	17 and more
8	United States	3712	17 and more
9	Bahamas	3713	17 and more
10	United States	3714	17 and more
		3715	17 and more
		3716	17 and more

Fig. 2. Missing data replacement of Country of Residence and Age-description

The selection of a subset of relevant features helps in simplification of the model to train faster by reducing the complexity and improving accuracy.

To provide with the better performance of the model by reducing over fitting and curse of dimensionality, a chi square test is required to determine the existence of relationship between the target variable and each input feature [29]. In order to discard the independency of the input variables from the target one, a chi square statistics is also calculated by finding the square between each feature and its target. The higher the value of chi square, there is a further chance of correlation between feature and its class.

Let the feature variable be X.

A: Various positive occurrences that contains instance X

B: number of negative examples
that contains include X

C: number of positive example that don't contain include X

D: number of negative examples that don't contain include X

A, B, C, D are considered as observed value whereas E_A, E_B, E_C, E_D are the expected value.

$$\text{Chi}^2 = \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}$$

$$\frac{1}{d} = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

$$= \frac{(A-E_A)^2}{E_A} + \frac{(B-E_B)^2}{E_B} + \frac{(C-E_C)^2}{E_C} + \frac{(D-E_D)^2}{E_D} \qquad (1)$$

Where, Observed frequency = number of observations of class= Ek

Expected frequency = number of expected observation of class if there was no relationship between the feature and target = Ok

Feature selection method came up with the 8 best attributes found to be relevant with the output variable from the total of 21 with their consecutive statistical score.

Figure 3 shows the eight best features with their respective statistical scores after feature selection technique.

	Specs	Score
17	result	2832.370195
15	contry_of_res	1085.247913
8	A9_Score	743.906779
5	A6_Score	669.651347
4	A5_Score	428.898999
3	A4_Score	370.204436
6	A7_Score	306.259069
2	A3_Score	250.675062
12	ethnicity	222.554249
1	A2_Score	197.371027

Fig. 3.8 best attributes and their statistical scores using feature selection techniques

After pre-processing of the data set, the category of a new observation can be identified based on a training set with set of observations and their categories. This can only be possible by finding a model that will distinguish the classes to predict the class labels.

The data set of 3741 samples were splitted into 70% training data to build and fit the model with predefined data points and their respective categories. The remaining 30% were used as testing data to validate the model through new data points. Random state is thereby used for repetitive result.

C. Three classifiers

To build up a model, there is a need of various classifiers for proper classification which is mainly based on Machine learning [15]. Some of them are Logistic Regression, Naive Bayes Classifiers, Decision Trees, Boosted Trees, Random Forest [16], Neural Network, Nearest Neighbor etc [11]. After the brief analysis of literature survey, it has been conclude that though various machine learning algorithm were used in different researches, but their combination has not yet been applied [13]. So from the different classifiers, three of them were selected to build a model and identify whether the suspected children are truly autistic or not [14].

D. Classification of dataset without cross-validation

After preprocessing of the dataset, the category of a new observation can be identified based on a training set with set of observations and their categories. This can only be possible by finding a model that will distinguish the classes to predict the class labels [18].

The dataset of 3741 samples were splitted into 70% training data to build and fit the model with predefined data points and their respective categories. The remaining 30% were used as testing data to validate the model through new data points. Random state is thereby used for repetitive result.

To build up a model, there is a need of various classifiers for proper classification which is mainly based on Machine learning. Some of them are

After splitting, the total dataset was classified with three different classifiers namely Logistic Regression, Support Vector Machine and Random Forest in order to build a model and predict discrete categories. While talking about the different classifiers, Logistic Regression is used when the target variable is found to be categorical like whether the email is spam or not or to find a tumor is malignant or not. The s-shaped curve takes any genuine esteemed number and maps it into a value somewhere in the range of 0 and 1.

Equation for Logistic Regression,

$$y = \frac{e^{(b_0+b_1+x)}}{1+e^{(b_0+b_1+x)}} \qquad (2)$$

Where, y= predicted output, b0= bias or intercept term and b1= co-efficient for the single value output which is x.

Figure 4 describes the linear model to be unbounded, whereas the logistic model representing as an s-shaped curve that ranges from 0 to 1 through mapping.

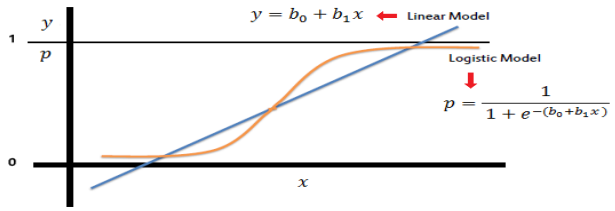


Fig. 4.Two different curves of Linear Model and Logistic Regression Model

The second classifier is the Support Vector Machine [30] that separates the data points with different categories through a hyper plane which gives the largest minimum distance to the training example to isolate consummately. The new examples are later mapped into same space to predict their belonging category.

Figure 5 shows the three hyper planes to classify the data points distinctly. C with the maximum distance between nearest data point is chosen as the right hyper plane.

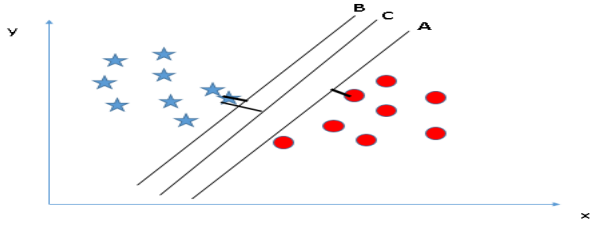


Fig. 5.Hyperplanes in Support Vector Machine

Last is the Random Forest classifier that builds multiple decision trees of a series of question sets and their conditions organized in a tree structure. The multiple decision trees are later merged to get the majority of final class label for further accuracy and stable prediction by selecting a random subset of features for splitting the nodes [17].

Figure 6 shows the three different decision trees where the test conditions are applied by following the appropriate branch to get the outcome of the test in the leaf node. The major-voting of the class predicts the final class of the test data.

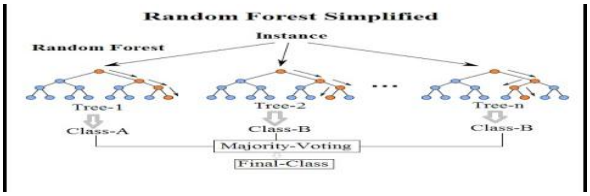


Fig. 6.Random Forest Classifier

The training and testing data of the dataset are thereafter classified with three different classifiers to build a model for predicting whether a child is suffering from autism or not and fit them accordingly.

Figure 7 displays the classification accuracy of the three classifiers before cross-validation.

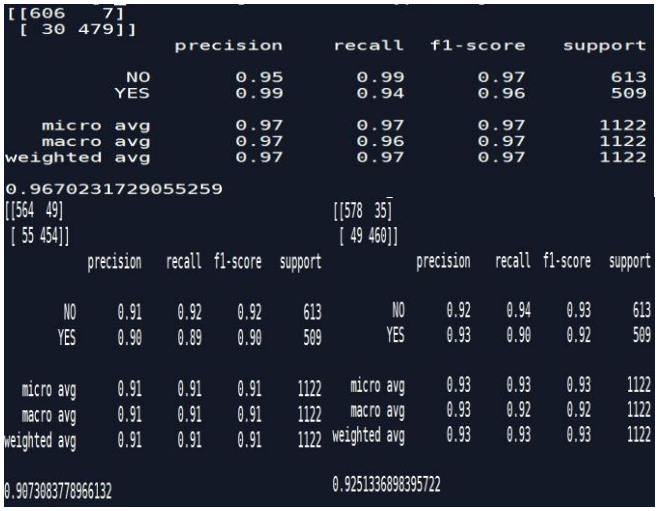


Fig. 7.Classification accuracy of Logistic Regression, SVM and Random Forest without cross-validation

Thus the classification of Logistic Regression provides with an accuracy of 93% whereas that of Support Vector Machine is 96%. Finally while classifying with the Random Forest by merging 200 different decision trees; it came up with an accuracy of 100%. The various evaluation measures like precision, recall, f-measures are used to find the classification report for testing the classification algorithm.

E. Classification of dataset with cross-validation

A resample system is important to evaluate the expertise of AI models to be less one-sided for the forecast of inconspicuous information that were not utilized during the preparation of a model. During the k-cross approval, the complete examples are haphazardly divided into k equivalent size of subsamples. Out of those subsamples, one of them will be considered as approval information for testing and the remainder of k-1 will be utilized for preparing which will be rehashed k times with the utilization of approval information precisely once in the k subsamples. They are later found the middle value of to create a solitary estimation by giving a favorable position of utilizing all the perception as preparing and approval. Moving forward with the huge advantage, the dataset of autism suspected patients were 10-times cross validated before classification. Furthermore, the accuracy of cross validation were estimated for the 10 subsamples and averaged to find the final cross-validated accuracy for three different classifiers that are likely to be Logistic Regression, Support Vector Machine and Random Forest. Figure 8 displays the classification accuracy of the three classifiers after cross-validation.

Table- IV: Different classifiers with their cross validation accuracy

Machine Learning Classifier	Cross-validation score
Logistic Regression	87.13%
Support Vector Machine	91.71%
Random Forest	95.50%

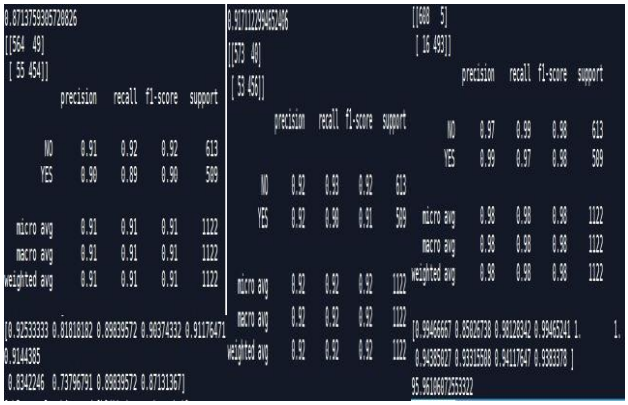


Fig. 8.Classification accuracy of Logistic Regression, SVM and Random Forest after cross-validation

F. Prediction through two different predictive models

From the 21 attributes for autistic detection, the samples of set of ten questionnaires were separately splitted for training and testing with re-sampling through 10 cross validation. The sampled preprocessed data were moved to Random Forest classifier for classification in order to build a model that helps to find out the disease from the suspicious ones. On the other way, the samples for rest of the attributes consisting of characteristics of individuals are again splitted for training and testing in a separate way. The 10 cross validation were further used for training and validating the ten equal size of subsamples which are later classified individually to establish a second model for autism detection. The accuracy validated from the two different models was then averaged to ascertain the final accuracy. Figure 9 defines the accuracy of three classifiers by merging the models. Figure 10 displays the flow chart of different predictive models

Table- V: Different models with accuracies for different classifiers

Machine Learning Algorithm	Accuracy of First Model	Accuracy of Second Model	Final accuracy
Logistic Regression	87.2%	86.2%	86.5%
Support Vector Machine	89.21%	87.96%	88.0%
Random Forest	85.69%	94.49%	89.5%

```
For first 6 attributes---------------------------->
0.8724483051616941
For rest 4 attributes---------------------------->
0.8620054555968135
Accuracy of Logistic Regression is---------------->
86.5
For first 6 attributes---------------------------->
0.8921568627450981
For rest 4 attributes---------------------------->
0.8796791443850267
Accuracy of SVM is---------------->
88.0
For first 6 attributes---------------------------->
85.66664602180137
For rest 4 attributes---------------------------->
94.49262032085561
Accuracy of Random Forest is---------------->
89.5
```

Fig. 9. Classification accuracy of Logistic Regression, SVM and Random Forest after merging two models

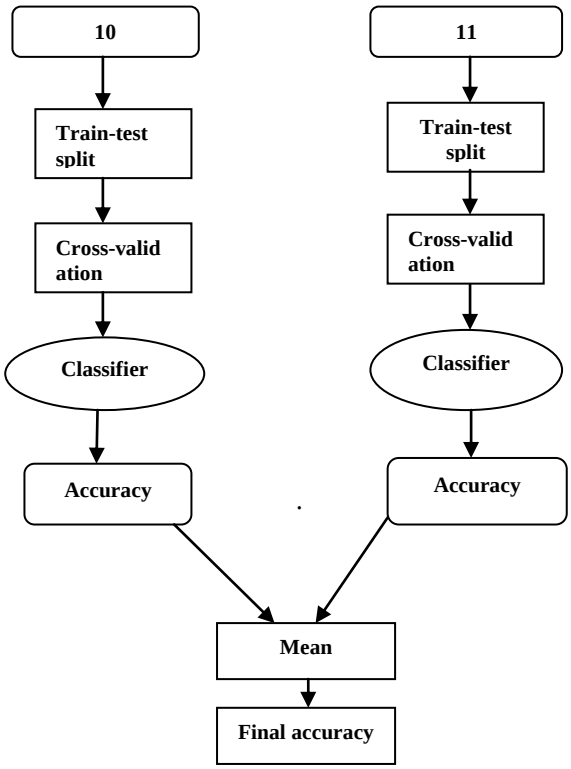


Fig. 10. Flowchart of different predictive models

G. Classification through 8 best feature selection technique

The total dataset of 3741 autistic conjectured children were classified in two ways- with the total of 21 features in one hand and that of 8 best features on the other. After proper classification of the preprocessed data, the Logistic Regression classifier provides an accuracy of 90.73% with 21 attributes and 95.72% by selecting the 8 best features through chi square statistics. Whereas Support Vector Machine to be 95.72 % and that of Random Forest classifier with 97.77% accuracy.

Figure 11 (a) and (b) shows the classification through 10 best features.

```
[[564  49]
 [ 61 448]]
              precision    recall  f1-score   support

          NO       0.90      0.92      0.91       613
         YES       0.90      0.88      0.89       509

   micro avg       0.90      0.90      0.90      1122
   macro avg       0.90      0.90      0.90      1122
weighted avg       0.90      0.90      0.90      1122

0.9019607843137255
```

```
[[592  21]
 [ 27 482]]
              precision    recall  f1-score   support

          NO       0.96      0.97      0.96       613
         YES       0.96      0.95      0.95       509

   micro avg       0.96      0.96      0.96      1122
   macro avg       0.96      0.96      0.96      1122
weighted avg       0.96      0.96      0.96      1122

0.9572192513368984
```

Fig. 11(a). Classification accuracy of Logistic Regression and SVM through 8 best-features

```
[[606   7]
 [ 18 491]]
              precision    recall  f1-score   support

          NO       0.97      0.99      0.98       613
         YES       0.99      0.96      0.98       509

   micro avg       0.98      0.98      0.98      1122
   macro avg       0.98      0.98      0.98      1122
weighted avg       0.98      0.98      0.98      1122

0.9777183600713012
```

Fig. 11(b). Classification accuracy of Random Forest through 10 best-features

While classifying the total dataset through Random forest classifier, the 21 attributes takes 0.00200 seconds while the best 10 attributes of 21 takes 0.00150 seconds for completing the training and testing process in order to assess the autistic child among the sceptical ones.

Figure 12 presents the comparision chart of classifying 21 vs. 10 attributes

```
Time taken for training and testing with 21 attributes
0.004008769989013672
[[606   7]
 [ 20 489]]
              precision    recall  f1-score   support

          NO       0.97      0.99      0.98       613
         YES       0.99      0.96      0.97       509

   micro avg       0.98      0.98      0.98      1122
   macro avg       0.98      0.97      0.98      1122
weighted avg       0.98      0.98      0.98      1122

Time taken for training and testing with 10 best attributes
0.0026397705078125
[[605   8]
 [ 15 494]]
              precision    recall  f1-score   support

          NO       0.98      0.99      0.98       613
         YES       0.98      0.97      0.98       509

   micro avg       0.98      0.98      0.98      1122
   macro avg       0.98      0.98      0.98      1122
weighted avg       0.98      0.98      0.98      1122
```

Fig. 12. Comparision of timing for classification through 21 and best 10 attributes

H. Evaluation metrices for machine learning algorithm

Sometimes it is necessary to understand the quality of a model or selecting the most acceptable model by examining their efficiency and performance to assign a class for new unlabelled examples. The various evaluation measures like precision, recall and f-scores are there that will be good enough to assess the performance of a model to solve the problems [24].

Precision is the portion of applicable examples among the recovered cases.

$$Precision = \frac{No.of\ True\ Positive}{No.of\ True\ Positive + No.of\ False\ Positive} \quad (3)$$

Whereas; Recall is the part of applicable occurrence over the all out number of significant occasions.

$$Recall = \frac{No.of\ True\ Positive}{No.of\ True\ Positive + No.of\ False\ Negative} \quad (4)$$

And finally, F-score is characterized as the weighted harmonic mean of the precision and recall of the test.

$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

Table- VI: Different classifiers with their respective Precision, Recall and F-scores

Classifiers	Precisions (No)	Precisions (Yes)
Logistic Regression	0.91	0.90
Support Vector Machine	0.92	0.93
Random Forest	0.95	0.99

Classifiers	Precisions (No)	Precisions (Yes)
Logistic Regression	0.92	0.89
Support Vector Machine	0.94	0.90
Random Forest	0.99	0.94

Classifiers	Precisions (No)	Precisions (Yes)
Logistic Regression	0.91	0.90
Support Vector Machine	0.92	0.92
Random Forest	0.95	0.96

I. The proposed model

The model was thus build by training with the 10 best features and re-sampled with 10 fold cross validation through Random Forest classifier. The user with autistic suspicious patients can provide inputs for the 10 best features in the model, which after prediction yields with the answer to be Yes or No means if they are suffering from autism or not.

Thus a predictive model for assessing the autistic disorder patient were build through Random Forest Classifier in such a way, so that the caretaker, parents and health care providers, care takers, parents and relatives of suspected child can give answers of those 10 best features through user input on behalf of autistic child and get a suitable answer of 'YES' to be autistic or 'No' for not to be autistic easily.

Figure 13 shows the list of questionnaires asked to parents or caregivers of autistic suspected child and the answers they are providing to detect autism.

```
1. Does your child easily follow the conversations of different
people in a social group??
2.Can yor child easily switch between different activities??
3.Does your child not knows how to chat with peers??
4.   Is your child good for everyday short chats??
5.Does he faces difficulties to understand the feelings while
reading a story??
6.Does he easily understands what others feel by just looking at
their faces??
7.Ethnicity??
8.Country_of_residence??
9.Result??
OTHER QUESTIONS:
1. Does your child often hears voices that others do not hear??
2.Does your child focuses on the big picture by going out from
small details??
3.Does your child likes to play games with other children in
preschool??
4. Does your children faces difficulties to make new friendships
Enter the entries in a single line (separated by space):

1 0 0 1 1 1 Asian India 9
['YES']
```

Fig. 13. List of questionnaires asked and answered during Autism detection

III. RESULTS AND DISCUSSIONS

Now a day there is a huge need of providing a time-consuming and easy accessible diagnostic tool to detect autism at an early stage to help clinicians in providing early medication. Therefore autistic and non-autistic real time data were collected from different hospitals and therapy centers in Odisha and Kolkata which were further pre-processed and later classified with the three different classifiers by using 10 best features out of 21. The k cross-validation was also used for training and building the model more accurately with the folding of 10 times by dividing the samples equally.

Figure 14 cites the graphical representation of different classifiers without and with cross validation together with 10 best features.

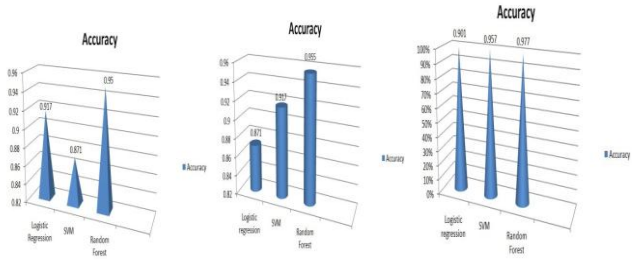


Fig. 14(a). Graphical representation of different classifiers without cross validation (b). Graphical representation of different classifiers with cross validation (c). Graphical representation of different classifiers without 10 best features

The various evaluation measures were estimated to test the classification algorithm and analyze them. Random Forest provided with the highest accuracy than Logistic Regression and SVM.

Figure 15 gives a graphical view of different classifiers with their evaluation metrices.

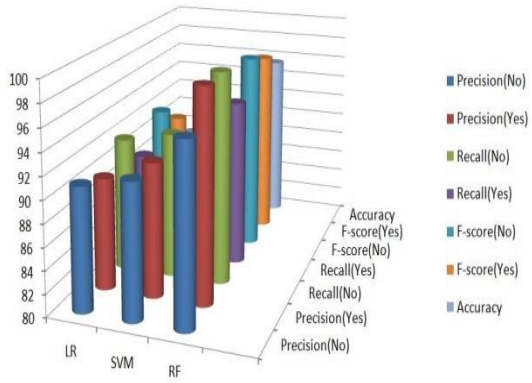


Fig. 15. Graphical representation of different classifiers without their evaluation measures

The predictive model was thus later build up with the help of Random Forest based classifier to diagnose if the suspected patient is truly suffering from autism or not by providing with different questions to the attainder on behalf of the autistic child. Their answers to the particular questions were put into the model to get the respective result.

IV. CONCLUSION AND FUTURE SCOPE

Since the chances of children being diagnosed with autistic spectrum disorder have increased dramatically, there is a huge need to provide a time-consuming and easy accessible diagnostic tool to detect autism at an early stage that encourages the clinicians to provide earlier medications. The dataset with 3741 samples, assembled from several autism therapy centers in India and their 21 attributes were preprocessed by handling missing values and mapping the categorical values. The dataset were further classified through three different classifiers namely Logistic Regression, Support Vector Machine and Random Forest algorithm simultaneously by selecting 8 best features through model selection technique. Among all the classifiers, Random forest serves up with the highest accuracy by assessing the evaluation metrices. The model thus prepared after training and testing were used by the autistic suspicious patients to know whether they are suffering from autism by providing user input for those very attributes.

After proper diagnosis of autism, there is a need to focus on future research initiatives in the area of behavioral analysis of autism as the autistic child possesses multiple limitations including social interaction, communication, and restrictive behaviors. Since they are more sensitive with little issues, miscommunication and misunderstanding results in frustration and discouragement, the emotion detection of autistic child will help them in providing better and comfort life.

ACKNOWLEDGMENT

The paper is one of the pieces of my execution to diagnosing the medically introverted kid. I need to express my most significant gratefulness to every single one of the people who has given me the probability to finish the paper. My special thanks are extended to the authorities and all the parents of autistic child in the Centre for Autism Therapy, Counselling and Help (CATCH) in Bhubaneswar Odisha for their continuous support during autistic data collection.

REFERENCES

1. Li Y, Adel Said Elmaghraby A "A Framework for using Games for Behavioral Analysis of Autistic Children", Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES), IEEE 20-23, 2014.
2. Bretagne Abirached, Yan Zyang, J.K. Agarwal, Birgi Tamersoy et al. "Improving Communication Skills of Children with ASDs through interaction with Visual Characters", 1st International Conference on Serious Games and Applications for Health (SeGAH), IEEE 16-18, 2011.
3. Nazih Heni, Habib Hamam "Design of Emotional Educational System Mobile Games for Autistic Children", 2nd International Conference on Advanced Technologies for Signal and Image Processing, IEEE 21-23, 2016.
4. Agnes Lacroix, Michele Guidette, Bernadette Roge, Judy S. Reilly "Facial emotion recognition in 4- to 8-year-olds with autism spectrum disorder: A developmental trajectory approach", Research in Autism Spectrum Disorders, Volume 8 Issue 9, September 2014.
5. Chien-Hsu Chen, I-Jui Lee, Ling-Yi Lin "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders", Research in Developmental Disabilities, Volume 36 Issue 8, January 2014.
6. Maha Jazouli, Aicha Majda, Arsalane Zarghili "A $P Recognizer for Automatic Facial Emotion Recognition using Kinect Sensor", Intelligent Systems and Computer Vision (ISCV), IEEE 17-19, 2017.
7. Fadi Thabtah, Firuz Kamalov, Khairan Rajab "A new computational intelligence approach to detect autistic features for autism screening", International Journal of Medical Informatics, Volume 117, September 2018.
8. Patricia Howlin "Autism Spectrum Disorders", Psychiatry, Volume 5, Issue 9, September 2006.
9. Osman Altay, Mustafa Ulas "Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K-Nearest Neighbor in Children", 6th International Symposium on Digital Forensic and Security (ISDFS), 2018.
10. Mark Hall, Eibe Frank, Geoffrey Holmes et al. "The WEKA data mining software: an update", ACM SIGKDD Explorations Newsletter. ACM Digital Library, Volume 11 Issue 1, November 2009.
11. Saurabh Mukherjee, Dr. Neelam Sharma "Intrusion Detection using Naive Bayes Classifier with Feature Reduction", Procedia Technology, Volume 4, September 2012
12. Rita Barone, Salvatore Alaimo, Marrianna Messina et al. "A Subset of Patients with Autism Spectrum Disorders Show a Distinctive Metabolic Profile by Dried Blood Spot Analyses", Frontiers in Psychiatry, Volume 9, December 2018.
13. Radha V "Neural Network Based Face Recognition Using RBFN Classifier", Proceedings of the World Congress on Engineering and Computer Science, Volume1, October 2011.
14. Dayana C, Tejera Hernandez "An Experimental Study of K* Algorithm", Information Engineering and Electronic Business (IJIEEB), Volume 2, March 2015.
15. John G. Cleary, Leonard E. Trigg "K*: An Instance-based Learner Using an Entropic Distance Measure", 12th International Conference on Machine Learning, July 1995.
16. Meng Li, Qing Song, Qinjun Zhao "A Fuzzy Adaptive Rapid-Exploring Random Tree Algorithm", 3rd International Conference on Materials Science and Mechanical Engineering (ICMSME), DEStech 167-171, 2016.
17. Mariana Belgiu, Lucian Dragut "Random forest in remote sensing: A review of applications and future directions", Journal of Photogrammetry and Remote Sensing, Volume 114, April 2016.
18. E Feczko, NM Balba, O Miranda-Dominguez et al. "Subtyping cognitive profiles in Autism Spectrum Disorder using a Functional Random Forest algorithm", NeuroImage, Volume 172, May 2018.
19. Halim Abbas, Ford Garberson, Eric Glover, Dennis P Wall "Machine Learning approach for early detection of autism by combining questionnaire and home video screening", Scholarly Journal of Informatics in Health and Biomedicine, Volume 25 Issue 8, August 2018
20. Alokananda Rudra, Saoni Banerjee, Nidhi Singhal et al. "Translation and Usability of Autism Screening and Diagnostic Tools for Autism spectrum Conditions in India", Autism Research, Volume 7, October 2014.
21. Alokananda Rudra, Mathhew K Belmonte, Parmeet Kaur Soni et al. "Prevalence of Autism Spectrum Disorder and Autistic Symptoms in a School-Based Cohort of Children in Kolkata, India", Autism Research, Volume 10 Issue 10, December 2017.
22. Qandeel Tariq, Jena Daniels, Jessey Nicole Schwartz et al. "Mobile detection of autism through machine learning on home video: A development and prospective validation study", PLOS, Volume 15 Issue 11, November 2018.
23. Kayleigh K. Hyde, Marlena N. Novack, Nicholas LaHaye et al. "Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review", Review Journal of Autism and Developmental Disorders, Volume 6 Issue 2, February 2018.
24. Fadi Thabtah "Autism spectrum disorder screening: Machine learning adaptation and dsm-5 fulfillment", Proceedings of the 1st International Conference on Medical and Health Informatics, Volume 5, May 2017.
25. Lorna Goddard, Lucy A Henry Hill et al. "Experiences of diagnosing autism spectrum disorders: A survey of professionals in the United Kingdom", National Autistic Society, Volume 20 Issue 7, December 2016.

26. Christensen DL, Braun KVN, Baio J et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities. MMWR Surveillance Summaries. 2018; 65 (13): 1-23.
27. Fadi Thabtah "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. Informatics for Health & Social Care". Volume 44 Issue 3, September 2017.
28. JA Kosmicki, V Sochat, M Duda, DP Wall "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning", Translational Psychiatry, Volume 5 Issue 2, February 2015.
29. Peter C. Austin, Juan Merlo "Intermediate and advanced topics in multilevel logistic regression analysis", Statistics in Medicine, Volume 36, May 2017.
30. Alessandra Retico, Ilaria Gori, Alessia Giuliano, Fillipo Muratori, Sara Calderoni "One-Class Support Vector Machines Identify the Language and Default Mode Regions As Common Patterns of Structural Alterations in Young Children with Autism Spectrum Disorders", Child and Adolescent Psychiatry, Volume 10, June 2016.

AUTHOR PROFILE

Camellia Ray received her B. Tech degree in Computer Science from University of Engineering and Management, Jaipur in 2017. Later she completed her M. Tech in Computer Science from KIIT University Bhubaneswar in the year 2019 with her research domain Machine Learning and Data Science. Currently she is a Junior Research Fellow in BITS Pilani, Hyderabad working with Artificial Intelligence. She has two publications in Springer regarding Autism.