Identification of Autism Spectrum Disorder using Deep Neural Network

Ashima Sindhu Mohanty¹, ², Priyadarsan Parida²* and K C Patra¹

¹Department of Electronics, SUIIT, Sambalpur University, Sambalpur, Odisha, India
²Department of Electronics & Communication Engineering, GIET University, Gunupur, Rayagada, Odisha, India

*Corresponding Author: priyadarsanparida@giet.edu

Abstract: One of the acute neuro developmental disorders throughout the world today is the Autism Spectrum disorder (ASD). It is lifelong disorder which affects the behaviour and communication skill of an individual. According to world health organization 2019 report, the number of individuals diagnosed with ASD is increasing creating a threat as it is analogous to significant health care cost. Early recognition can considerably reduce the effect. In order to get rid of the time consuming and expensive diagnosis procedures for ASD, a mobile based ASD screening tool known as ASDTest app was developed. The app recorded over 1400 number of instances covering toddler, child, adolescent and adult. It is available publicly in Kaggle and UCI Machine Learning repository for research purpose. The paper gives a new approach for identification of ASD using a deep classifier. The ASD identification works in the following steps. Feature analysis explains ASD traits thereby improving the efficiency of screening process. Further, Machine Learning (ML) classifier models report ASD class type with evaluation parameters. In this analysis, an attempt is made for the incorporation of Principal Component Analysis (PCA) for feature dimension reduction followed by the usage of Deep Neural Network (DNN) for classification of ASD class type. The data upon which the techniques are applied are collected from Kaggle and UCI ML repository. The experiment result indicates that, PCA in combination with DNN provide clinically acceptable output for effective ASD identification.

Keywords: ASD, ML, DNN, Classification

1. Introduction

Autism spectrum disorder (ASD) is a lifelong neurodevelopment condition with impairments in socio-communicative skills and the presence of repetitive behaviors and interests[1][2]. Current study indicates that autism spectrum covers nearly 1.5 percent of world’s population of which many people with ASD remain undetected[3]. As per previous analysis, the early symptoms of ASD are observed during the initial 6 to 18 months in a toddler’s life span. Further the symptoms are followed by developmental regression with loss in verbal, social as well as communication ability followed by abnormal motor development between 18 to 36 months of the child’s life span[4]. Some of the inactive social and communication behaviors like:

- uncertain laugh and giggle
- unable to make eye contact and respond to sound
- insensitive towards physical pain
- no interest to cuddle with parents
like to remain solo
improper object attachment
Less sensitive towards sudden light, noise, etc.
Repeating words and sentences

In a child, it is easier to identify the behavioral changes easily by observation in comparison to adolescent and adult since in adolescent and adult cases; some signs of ASD may overlap with distinct mental health disorders. In order to improve the quality of life of people suffering from ASD, Early detection as well as treatment is the most crucial steps to be put forward.

This research emphasizes on classifying ASD traits from no ASD traits among toddlers (12 to 36 months), children (4 to 11 years), adolescents (12 to 16 years) and adults (17 years and above). The proposed approach uses Qualitative Checklist for Autism in Toddlers (Q-CHAT-10)[5][6]as well as Autism Spectrum Quotient (AQ-10)[7][8] based on distinct behavioral independent variables with 10 number of screening questions included in the data set. In AQ test, screening method assigns a point per question. If the individual scores more than 6, then the individual is found to be probable with ASD trait and referred for further diagnostic assessment. In adult category, for question items 1, 7, 8 and 10, if answer is either ‘slightly agree’ or ‘definitely agree’, a point is allotted and for remaining questions, a point is allotted for answer to be either ‘slightly disagree’ or ‘definitely disagree’. In adolescent category, for question items 1, 5, 8 and 10, if answer is either ‘slightly agree’ or ‘definitely agree’, a point is allotted and for standing questions, a point is allotted for the answer to be either ‘slightly disagree’ or ‘definitely disagree’. For child category, towards question items 1, 5, 7 and 10 if answer is ‘slightly agree’ or ‘definitely agree’, then a point is allotted and for rest of the questions, a point is allotted for the answer to be ‘slightly disagree’ or ‘definitely disagree’. The following table 1 indicates the AQ questionnaires for adult, adolescent and child category[9].

| Q.No. | AQ-10 Adult questionnaire | AQ-10 Adolescent questionnaire | AQ-10 Child questionnaire | Definitely Agree | Slightly Agree | Slightly Disagree | Definitely Disagree |
|-------|---------------------------|--------------------------------|----------------------------|-----------------|---------------|------------------|-------------------|
| 1     | “I mostly notice small sounds when others do not”. | He/ She notices patterns in things everytime. “He/she generally focuses more on the entire picture, rather than small details”. | “He/she mostly notice small sounds when others do not”. | | | | |
| 2     | “I generally focus on the entire picture, rather than small details”. | “In social group, he/she can easily keep track of different people’s conversations”. | “He/she generally focuses more on the entire picture, rather than small details”. | | | | |
| 3     | “I realize it to be easy for doing more than one thing at a time”. | “During an interruption, he/she can switch back very quickly to what he/she was doing”. | “In social group, he/she can easily keep track of different people’s conversations”. | | | | |
| 4     | “During an interruption, he/she can switch back very quickly to what he/she was doing”. | | | | | | |
"I realize it to be easy for reading between the lines while talking to someone".

"He/she frequently finds about not knowing to keep a conversation going".

"He/she has no idea of how to continue with a conversation with his/her peers".

"He/she is fine with social chit-chat".

"When he/she is through a story, it is hard for him/her to work out the characters’ intention".

"I like collecting information about group of things (e.g. types of: bird car, plant, train, etc)".

"During his/her stay in preschool, he/she enjoyed playing games with children which involved pretending".

"He/she realizes it to be tough visualizing how would it be to be someone else".

"He/she feels easy to work upon what someone thinks or feels just by looking at his/her face".

"He/she realizes it to be hard in making new friends".

In toddler category, for question items 1 to 9, if option in column C, D or E is encircled then 1 point is allotted per question. For question item 10, encircling the option in column A, B or C gains a point. All the points are summed up. For the score more than 3, ASD diagnostic assessment is cited for the toddler individual. The following Table 2 indicates the questionnaires in Q-CHAT 10 for toddler category[9].

**Table 2. Q-CHAT questionnaires for toddlers.**

| Q. No. | Q-CHAT toddler questionnaires | A | B | C | D | E |
|--------|--------------------------------|----|----|----|----|----|
| 1      | “Does your child gaze at you when his/her name is called by you”? | “Always” | “Usually” | “Sometime s” | “Rarely” | “Never” |
| 2      | “How far is it easy for you to come in eye contact with your child”? | “Very use” | “Quite easy” | “Quite difficult” | “Very difficult” | “Impossible” |
| 3      | “Does your child point for indicating that he/she wants something”? (e.g. pointing a toy which is out of reach from him/her) | “Many times a day” | “A few times a day” | “A few times a week” | “Less than once a week” | “Never” |
| 4      | “Does your child point for sharing interest with you”? (e.g. pointing towards an captivating sight) | “Many times a day” | “A few times a day” | “A few times a week” | “Less than once a week” | “Never” |
“Does your child pretend”? (e.g. caring for dolls, talking on a toy phone)

“Many times a day”
“A few times a day”
“A few times a week”
“Less than once a week”
“Never”

“Does your child follow your sight”?;

“Many times a day”
“A few times a day”
“A few times a week”
“Less than once a week”
“Never”

“Does your child show signs of comforting you or someone else in the family when you or someone else is upset”? (e.g. stroking your or their hair, hugging you or them)

“Always”
“Usually”
“Sometimes”
“Rarely”
“Never”

“You would describe your child’s first words as”

“Very typical”
“Quite typical”
“Slightly unusual”
“Very unusual”
“My child doesn’t speak”

“Does your child use basic gestures”? (e.g. waving good bye)

“Many times a day”
“A few times a day”
“A few times a week”
“Less than once a week”
“Never”

“Does your child gaze at nothing without any purpose”?;

“Many times a day”
“A few times a day”
“A few times a week”
“Less than once a week”
“Never”

The dataset utilized in this research was given by Fadi Fayez Thabtah using a mobile application, ASDTests for screening ASD in all category of individual. The toddler data set consists of 18 distinct items with one output class and the child, adolescent and adult data set comprises of 21 distinct items with one output class in the data sets.

For diagnosing ASD, a classifier model is built using deep learning to classify ASD and no ASD class. The classifier model is designed from training data and gets evaluated on testing data. Present research on ASD classification emphasizes:

(i) Evolution of latest deep learning approaches to classify ASD and no ASD class.
(ii) Transforming and reducing the number of features followed by turning down the diagnosis time for ASD.
(iii) Improvement of classification performance parameters.

The performance parameters are accuracy, error, sensitivity, specificity, Area under Curve (AUC), positive predictive value (PPV), negative predictive value (NPV), etc. The confusion matrix for ASD Screening is provided below:

| Actual class type | ASD          | No ASD       |
|-------------------|--------------|--------------|
| ASD               | True Positive (TP) | False Negative (FN) |
| No ASD            | False Positive (FP) | True Negative (TN) |

a) True Positive (TP): Actual data: positive, predicted data: positive.
b) False Positive (FP): Actual data: negative, predicted data: positive.
c) False Negative (FN): Actual data: positive, predicted data: negative.
d) True Negative (TN): Actual data: negative, predicted data: negative.

To validate the effectiveness of deep learning predictive model, cross validation is implemented[10]Firstly, the input training data set is split into N number of partitions, where N is generally adjusted to 10. Then the classifier model is trained upon (N – 1) number of partitions followed by testing on the remaining partitions and the procedure is repeated N times by randomly partitioning the training dataset. During the cross validation, stratification occurs in which during the data splitting, random shuffling is implemented for the availability of each class in each partition.

There are numerous neurological disorders whose few symptoms match with those of ASD thereby making the detection of ASD a difficult and challenging task. But at the same time early
detection of ASD is essential for maintaining the individual’s mental as well as physical health. With the upliftment of Machine Learning as well as Deep Learning classifier models, prediction and early detection of neurological diseases based upon physiological parameters now seems to be much possible.

This paper is organized in a way where, section 2 presents the research work where some related developments with classifier models are discussed. Section 3 focuses on the data sets collected for the research followed by description of the proposed methodology in section 4. The results obtained from the experimentation on each category of individual is discussed in section 5 followed by the conclusion and future scope in section 6 and 7 respectively.

2. Literature Review

The author put forward a time efficient mobile-based ASD screening tool known as ASDTest[11]. Followed by Q-CHAT 10 and all AQ-10 based screening, the author collected 1452 instances within a period of 4 months using the ASDTest tool enclosing all 4 categories of individuals. The toddler data set was unbalanced and hence dropped from the investigation leaving 1100 instances and 21 features corresponding to child, adolescent and adult. The adult data set comprises of 704 instances followed by adolescent one containing 104 cases and finally child data set comprising of 292 number of instances. Some missing values in two features: “ethnicity” and “who_is_taking_the_test” were found during the analysis. The wrapping filtering method extracted some of the features from the data sets. From the adult one, 12 influential features got filtered from all features: question 1 to 10 in AQ-10 Adult questionnaire, “gender” and “used app before”. In adolescent category, 8 features got selected: question 2, 5, 9 and 10 from AQ-10 Adolescent questionnaire, “gender”, “born with jaundice or not” and “used app before”. Lastly in child dataset, 4 features got selected from all features: question 1, 4, 8 and 10 in AQ-10 Child questionnaire. Two ML algorithms, Naïve Bayes (NB)[12] and Logistic Regression (LR)[13] were applied to classify the ASD data. The performance parameters: accuracy, sensitivity and specificity were determined for all data sets. Adult dataset achieved higher rates than adolescent as well as child datasets as it contains more number of instances in comparison to the other two datasets. From the two, LR proved to be better in comparison to NB classifier. In adult dataset, LR outperformed NB in terms of accuracy, sensitivity and specificity by 4.12, 4.2 and 3.01 percent respectively.

The researchers with an aim towards effective screening identified further fewer as well as influential features in the ASD screening process[14]. The analysis covered the same data as used by the author in[11].The toddler dataset was dropped off from the entire dataset due to its unbalanced nature. In the overall dataset of 1100 instances excluding toddler one, there are 707 number of instances with ASD class and 393 with no ASD class present. The authors put forwarded Variable Analysis (VA) which takes into account feature to class correlation as well as decreases feature to feature correlation. VA result was compared with other filtration methods: CHI-SQ[15], IG[16], Correlation Feature Set (CFS)[17] in addition with Correlation Attribute Evaluation analysis[18]. VA analysis chose 6, 8 and 8 items out of 21 from AQ-10 adult, AQ-10 adolescent as well as AQ-10 child dataset respectively. VA results got verified by two ML classifiers, Repeated Incremental Pruning to produce Error Reduction (RIPPER) as well as C4.5 (Decision Tree) in terms of specificity, sensitivity, PPV, NPV, and accuracy. To train the data set, ten-fold cross validation was used where arbitrarily the data set was split into 10 parts. In all the data sets, VA reduced the features in the most effective way as compared to other filtering methods. The authors compared the result of VA features with features of IG, CHI, correlation, CFS together with original 21 features by RIPPER and C4.5 algorithms. For adolescent data set, the accuracy rate of VA was good with IG, CHI, correlation as well as CFS with a slight fall in the rate for child and adult data set. In adolescent case, when compared with the original 21 features, the RIPPER and C4.5 classifiers derived from VA features revealed higher accuracy by 10 as well as 6 percent respectively. The specificity rate was found to be the highest in adult data set for VA features but was 2.8, 1.9, 1.4, 3.6 and 3.0 percent less than the features of no feature selection, IG, CHI, Correlation and CFS respectively. The sensitivity rate of 87.30 percent from the RIPPER classifier was the highest for VA features in adolescent case with a slight fall in adult and child cases. In adolescent case, the RIPPER algorithm derived from VA features resulted in better PPV as well as
NPV than other feature sets with a slight fall but acceptable rates in adult and child cases. VA analysis selected much limited number of features from all data sets in comparison to the other filtering methods followed by classifier models processing which yielded acceptable rate of performance parameters.

The author in [19] proposed a method for identifying ASD with an intention to achieve better classification accuracy with minimum feature subsets. The child data set with 21 features and 292 instances was obtained from the UCI ML repository assessed with swarm intelligence based binary firefly feature selection wrapper. Using the feature selection wrapper, the author selected 10 features from 21 features of ASD dataset to distinguish between ASD and no ASD class type: “A1_Score”, “A2_Score”, “A3_Score”, “A4_Score”, “A5_Score”, “A7_Score”, “A8_Score”, “A9_Score”, “A10_Score”. Followed by dimensionality reduction, for classifying ASD and no ASD class type, the authors used ML classifier models: NB, J48 Decision tree[20], Surface Vector Machine (SVM)[21], K Nearest Neighbors (KNN)[22] and Multilayer Perceptron (MLP)[23]. The results obtained by the approach showed average accuracy in the range of 92.12% to 97.95% with optimum feature subset selection and training the classification models with minimum behaviour sets thereby validating the performance of classification models.

To establish a better performance, the authors proposed a new Rule based Machine Learning (RML)[24] which in addition with detecting autistic cases also offered rules to be used for analyzing the reasons behind classification. Except toddlers, the research covered datasets of child, adolescent and adult category collected from UCI ML repository with 1100 instances and 21 features. To evaluate the performance, the RML result got further compared against eight algorithms: RIPPER, RIDOR, Nnge, Bagging, Boosting, CART, C4.5, and PRISM[25][26][20]. In case of all the data sets, performance of RML with rate of accuracy, error, sensitivity specificity and harmonic mean was satisfactory in comparison to the considered ML algorithms. The only limitation of the research is excluding the toddler instances because of their unbalanced nature as well as rare availability.

The researchers collected adult, adolescent, child and toddler ASD data sets from UCI ML repository with 2009 instances[27]. The data sets contain 1054 number of instances in toddler, 248 number of instances in child, 98 instances in adolescent and 609 number of instances in adult data sets. Nine ML classifier models: Adaboost[28] Flexible Discriminant Analysis (FDA)[29], C5.0, Boosted Generalized Linear Model (Glmboost)[30], LDA[31], Mixture Discriminant Analysis (MDA)[32], Penalized Discriminant Analysis (PDA) [33], SVM and CART got implemented for classifying ASD class. Performance parameters: Accuracy, AUROC, Kappa Statistics, Specificity, Sensitivity and Logloss were evaluated to account for the experimental results. Prior to classification, three feature transformation (FT) methods: log, Z-score and sine were used with statistical features like mean, median and maximum. According to the analysis, it was found that for toddler data set, the SVM classifier showed maximum performance with more than 95 percent performance rate, for child data, the Adaboost classifier showed the best performance rate with more than 95 percent, while Glmboost model performed the best in case of adolescent data set and finally for adult data set, Adaboost once again proved to be the best classifier. In addition, the FT which showed the best classification for toddler data set was the sine function and Z-score for child, adolescent and adult data set.

The authors by employing four ML algorithms: NB, LR, SVM and KNN along with ANN and CNN predicted the possibility of ASD in child, adolescent and adult[34]. The data sets got collected from UCI ML repository covering 1100 number of instances and 21 features excluding toddler category. The performance was evaluated in terms of accuracy, sensitivity and specificity. In adult dataset, the CNN produced good rate of performance parameters of nearly 99 percent which outperformed other classifier models. For adolescent dataset, CNN outperformed other classifier models in terms of accuracy and sensitivity rate with nearly 95 percent. But the specificity rate yielded was slightly low but acceptable. Finally, in the child dataset, CNN resulted in the highest accuracy and specificity rate of nearly 99 percent but sensitivity rate was slightly low though acceptable. The analysis proved CNN as one of the better options for predicting and classifying ASD in all category of individuals.
3. Data Collection

The ASD data sets are collected from the Kaggle and UCI ML Repository which is publicly available for research purpose[35][36][37][38]. The data are collected by the ASDTest app developed by the author in[11]. In this research all the four types of the data sets belonging to toddler, child, adolescent and adult are used. The details of the data set are summarized below in Table 3:

| Sl. No. | Name of data set | Source | Attribute type | Number of attributes | Number of cases | Individuals with ASD class | Individuals with no ASD class |
|---------|------------------|--------|----------------|---------------------|----------------|---------------------------|-----------------------------|
| 1       | Toddler          | Kaggle [35] | Categorical, continuous and binary | 18 | 1054 | 728 | 326 |
|         |                  | UCI ML repository [36] | Categorical, continuous and binary | 21 | 292 | 141 | 151 |
| 2       | Child            | UCI ML repository [36] | Categorical, continuous and binary | 21 | 104 | 63 | 41 |
| 3       | Adolescent       | UCI ML repository [37] | Categorical, continuous and binary | 21 | 704 | 189 | 515 |
| 4       | Adult            | UCI ML repository [38] | Categorical, continuous and binary | 21 | 704 | 189 | 515 |

Except the toddler data set, in the rest three data sets, there are some missing values mainly in the attributes: “age”, “ethnicity” and “Who_completed_the_test” The detail is shown below in Table 4:

| Class name | Adult Data set | Instances after dropping missing value | Child Data set | Instances after dropping missing value | Adolescent Data set | Instances after dropping missing value |
|------------|----------------|----------------------------------------|----------------|----------------------------------------|----------------------|----------------------------------------|
| ASD        | 189            | 180                                    | 141            | 126                                    | 63                   | 62                                     |
| no ASD     | 515            | 429                                    | 151            | 123                                    | 41                   | 36                                     |
| Total      | 704            | 609                                    | 292            | 249                                    | 104                  | 98                                     |

Table 5 gives the attribute details in toddler and rest data sets respectively.

| Attribute Name          | Type of Attribute | Descriptions of Attributes                                      |
|-------------------------|-------------------|-----------------------------------------------------------------|
| Case Number             | Numeric           | Number of cases in dataset                                      |
| Question item 1-10 Answer | Either 0 or 1     | The questions in Q-CHAT-10 and AQ-10 based on possible answer. |
| Age_Mons                | Numeric           | Ages of toddler (in month), child, adolescent and adult (in years) |
| Score                   | Numeric           | Screening score from Q-CHAT-10 and AQ-10 questionnaire.         |
| Sex                     | String            | Male/ Female                                                    |
| Ethnicity               | String            | Ethnicity List in text format                                   |
| Jaundice                | Boolean (Yes/No)  | born with jaundice or without jaundice                          |
| Family_mem_with_ASD     | Boolean (Yes/No)  | Whether any immediate family member had                         |
### Who completed the test
- **String**
  - The individual performing ASD test

### Used the screening app before
- **Binary**
  - Yes/No
  - If the user used a screening app beforehand

### Country of residence
- **String**
  - List of countries in text

### Screening method
- **Numeric**
  - Kind of screening method chosen based on age description

### Class/ASD Traits
- **Boolean**
  - Yes/No
  - Whether the case has ASD/no ASD

## 4. Methodology

The proposed method is represented by Figure 1 according to which the input ASD data (Toddler, Child, Adolescent, and adult), collected from Kaggle and UCI ML data repository are applied for pre-processing prior to classification of ASD class. Prior to standardization, the missing data are dropped from the data sets to be analyzed. In the first stage of pre-processing, the input samples are standardized into desired values. After standardization in second step, the inputs are encoded to binary codes for satisfying activation function boundaries. Following, the third step is continued with reduction in the dimension of data by using Principle component analysis (PCA)[39]. In the next stage, 10-fold cross validation is used for separating the training and testing data. In further stage the training data is fed to the DNN for classification of ASD class. After training, at last stage, the predicted outputs are found from the model and are compared with the targets to calculate the performance parameters like Accuracy, Sensitivity, Specificity and f-measure.

### 4.1 Preprocessing

The collected data from the repository are preprocessed before feeding to the algorithm. The steps involved in preprocessing stage in this research are: (i) Standardization, (ii) Dimension reduction using PCA.

#### 4.1.1 Standardization

From the dataset used in this research, it is found that all the attributes are not in the same scale. Hence it is essential to bring the input attributes to a proper scale using the process of standardization. To carry out the process of standardization, mean and standard deviation are computed for a particular attribute. The standardized data is mathematically represented as:

\[
Sta_XX = \frac{(x - x_{\text{mean}})}{(x_{\text{std}})} \tag{1}
\]

where,
- \(x\) indicates the current value of the input \(X\).
- \(x_{\text{mean}}\) indicates the mean value of \(X\).
- \(x_{\text{std}}\) indicates the standard deviation of the input \(X\).

#### 4.1.2 Dimension reduction using PCA

When the number of input attributes is greater than 10, it is desirable to implement dimension reduction where the number of input attributes are reduced. The attributes which have less impact upon the output class type are dropped from the data set such that there is no effect on the output class. In this research, PCA is utilized for the dimension reduction in all data sets. In all the data sets, the dimension is reduced to 6 number of attributes based upon the PCA score. The PCA applied on all the data sets is shown in Figure 2 following which the best principal components (PC) are shown in Table 6.

### 4.2 K-fold cross validation

Cross-validation is a common procedure to split the training and testing data for evaluating machine learning/deep learning models on data sample. K refers to the number of groups that a given data sample is to be split into. Depending on the value of K the number of folds of the data is decided. For an example if K = 5, the complete data is randomly divided into 5 folds, out of which 4 folds are used for training and rest 1 fold is used for testing data. The procedure for K-fold cross validation is given below:
a) Splitting complete ASD dataset into k groups randomly
b) From k groups of data, taking one group as test group and rest (k-1) group as training group.
c) Fitting the deep learning model on training set and evaluating it on the testing data.

In this research, the value of K is chosen as 10. So it leads to 10-fold cross validation. In addition, the training and testing data are assigned with 80 percent as well as 20 percent weightage respectively.

![Architecture of the proposed method](image1.png)

**Figure 1.** Architecture of the proposed method

![PCA upon all data sets](image2.png)

**Figure 2.** PCA upon all data sets

| Data set   | PC1    | PC2    | PC3    | PC4    | PC5    | PC6    |
|------------|--------|--------|--------|--------|--------|--------|
| Toddler    | 20.80986 | 12.57357 | 12.05018 | 10.81484 | 9.904125 | 9.1673883 |
| Child      | 16.89853 | 12.13555 | 11.73338 | 10.63186 | 9.227409 | 8.506796 |
| Adolescent | 17.40421 | 13.84336 | 11.55226 | 10.66255 | 8.597747 | 6.878896 |
| Adult      | 16.74009 | 11.19603 | 10.09836 | 9.640852 | 8.870905 | 7.996994 |
4.3 Deep Neural Network Model
The DNN used in this analysis consists of five layers viz. sequence input layer, Long-Short Term Memory (LSTM) layer[40], fully connected layer, Softmax layer and classification layer as shown in the Figure 3

![Figure 3. Architecture of proposed method](image)

The sequence input layer takes the sequential ASD data in sequence to label classification. Secondly, the LSTM layer utilized in this deep network has a size of 6 as per the number of attributes found from PCA and produced 200 outputs through sigmoid activation function. The layer uses learning factor of 2, bias L2 factor of 0 together with d weights L2 learning factor of 1. Then after the fully connected layer (fc-Layer) used in this work takes the 200 sequential outputs produced from the LSTM layer and produces 2 number of output. This layer uses weight learning rate of 1, bias learning rate of 1, weight L2 factor of 1 and bias L2 factor of 0. Following, the softmax layer applies a softmax unit activation function on the output found from fully connected layer. Finally, the classification output layer produces two output classes (0 and 1) by computing the cross entropy loss with two mutually exclusive classes.

4.4 Classifier output based on performance parameters
The effectiveness of the proposed method in terms of classifier Output, ASD/no ASD is measured through the performance parameters like Accuracy (ACC), Specificity (Spe), Sensitivity (SN) and F-measure(F1) by DNN model. All the performance parameters are calculated from TP, TN, FP and FN values. The mathematical equations of the performance parameters are given in equation (2) to equation (5)

\[
Acc = \frac{(TP + TN)}{(TP + TN) + (FP + FN)} \tag{2}
\]

\[
SN = \frac{(TP)}{(TP + FN)} \tag{3}
\]

\[
Spe = \frac{(TN)}{(TN + FP)} \tag{4}
\]

\[
F_1 = \frac{(TP)}{TP + \frac{1}{2}(FP + FN)} \tag{5}
\]

5. Results and Discussion
The proposed method is applied on all category of toddler, child, adolescent and adult dataset in successive stages to find the performance measures Following the stage of pre-processing with 6 number of reduced attributes, the training of the DNN model is carried out in MATLAB 2016 environment using a PC with Intel Corei3 1.99GHz processor with 12GB RAM.
The training process for toddler, child, adolescent and adult dataset has a common learning rate of 0.005, mini batch size of 16 and learns rate drop period of 125 is shown in the figure 4, Figure 5, Figure 6 and Figure 7:

Figure 4. Training for Toddler data set

Figure 5. Training for child data set

Figure 6. Training for adolescent data set
The training details are summarized in the following Table 7:

| Data set     | Training data | Testing data | Number of epoch | Training duration(Second) |
|--------------|---------------|--------------|-----------------|----------------------------|
| Toddler      | 80 percent    | 20 percent   | 200             | 235                        |
| Child        | 500           | 185          |                 |                            |
| Adolescent   | 500           | 69           |                 |                            |
| Adult        | 500           | 296          |                 |                            |

Among all categories, the toddler data set contains the highest number of instances which results in the longest training duration among toddler, child and adolescent data set. Though the adult data set contains less number of instances in comparison to toddler one, but still due to higher number of epochs in adult data set, its training duration is the highest among all. From the training process it is evident that the training accuracy is found to be 100% in all the categories.

Following the training, the trained models achieved are tested on the testing data of all categories found from 10-fold cross validation model to yield performance measures given in Table 8. The confusion matrix obtained for all test data sets are shown in Fig. 8.

**Table 8. Performance analysis for ASD classification by DNN**

| Data set    | Performance measures | Calculated values |
|-------------|-----------------------|-------------------|
| Toddler ASD data | Accuracy            | 0.8524            |
|             | Sensitivity           | 0.7048            |
|             | Specificity           | 1                 |
|             | F1 score              | 0.82              |
|             | Accuracy              | 0.8421            |
|             | Sensitivity           | 1                 |
| Adolescent ASD data | Specificity           | 0.7273            |
|             | F1 score              | 0.8421            |
|             | Accuracy              | 0.8571            |
|             | Sensitivity           | 1                 |
| Child ASD data | Specificity           | 0.6818            |
|             | F1 score              | 0.8852            |
|             | Accuracy              | 0.8926            |
|             | Sensitivity           | 1                 |
| Adult ASD data | Specificity           | 0.8434            |
|             | F1 score              | 0.8539            |
Due to more instances and age group, adult test data showed the highest accuracy of 0.8926. Toddler category also showed good accuracy rate but due to its unbalanced nature, somehow its accuracy is bit reduced. Child and adolescent test data also yielded acceptable rate of accuracy i.e., more than 80 percent. In order to validate the performance analysis on ASD classification, the error is also calculated as per the actual and predicted value. As per the test data, the toddler data set has 210 number of instances out of which 27 instances could not achieve their target class. The child test data set has 49 number of instances out of which 16 instances could not achieve their target class. Further, in adolescent case, the test data has 19 instances out of which 4 could not achieve their target class. Finally, in adult data set, out of 121 number of instances, 22 instances could not achieve their target class. From table 7, it can be observed that 85.24 percent of toddler data is classified accurately by using the proposed DNN model leaving rest 14.76 percent of the respective data which could not achieve their target class. It is also found that in case of adolescent data, accuracy rate achieved is 84.21 percent leaving rest 15.79 percent of the respective data which didn’t achieve the target class. In child data set, 85.71 percent of accuracy rate is achieved with 14.29 percent of data which could not achieve their target class. Finally in case of adult data set, the highest accuracy rate of 89.26 percent is achieved with 10.74 percent of data which didn’t achieve the target class. Hence for the misclassified data, mean error and root mean square error (RMSE) are evaluated as shown in Table 9, 10, 11, 12.

Table 9. Error analysis in toddler test data

| Sl. No. | Actual class | Predicted class | Error | Mean Error | RMSE  |
|--------|--------------|-----------------|-------|------------|-------|
| 1      | 0            | 0.0004          |       | 0.0004     |       |
| 2      | 0            | 0.0111          |       | 0.0111     |       |
| 3      | 0            | 0.2049          |       | 0.2049     |       |
| 4      | 0            | 0.0124          |       | 0.0124     |       |
| 5      | 0            | 0.0003          |       | 0.0003     | 0.0001819 0.0761 |
| 6      | 0            | 0.0692          |       | 0.0692     |       |
| 7      | 0            | 0.301           |       | 0.301      |       |
| 8      | 0            | 0.005           |       | 0.005      |       |
### Table 10. Error analysis in Child test data

| Sl. No. | Actual class | Predicted class | Error Mean | Error RMSE |
|---------|--------------|-----------------|------------|------------|
| 1       | 0            | 0.0211          | 0.0211     |
| 2       | 0            | 0.0578          | 0.0578     |
| 3       | 0            | 0.0619          | 0.0619     |
| 4       | 0            | 0.0015          | 0.0015     |
| 5       | 0            | 0.0011          | 0.0011     |
| 6       | 0            | 0.0106          | 0.0106     |
| 7       | 0            | 0.0001          | 0.0001     |
| 8       | 1            | 0.9989          | 0.0011     | 0.0632    | 0.17601769 |
| 9       | 1            | 0.8507          | 0.1493     |
| 10      | 1            | 0.3177          | 0.6823     |
| 11      | 1            | 0.9967          | 0.0033     |
| 12      | 1            | 0.997           | 0.003      |
| 13      | 1            | 0.9999          | 0.0001     |
| 14      | 1            | 0.9955          | 0.0045     |
| 15      | 1            | 0.9897          | 0.0103     |
| 16      | 1            | 0.9967          | 0.0033     |

### Table 11. Error analysis in Adolescent test data

| Sl. No. | Actual class | Predicted class | Error Mean | Error RMSE |
|---------|--------------|-----------------|------------|------------|
| 1       | 0            | 0.019           | 0.019      |
| 2       | 0            | 0.6599          | 0.6599     | 0.17255   | 0.9908 |
| 3       | 0            | 0.0021          | 0.0021     |
| 4       | 1            | 0.0092          | 0.9988     |
Table 12. Error analysis in Adult test data

| Sl. No. | Actual class | Predicted class | Error Mean | RMSE  |
|---------|--------------|-----------------|------------|-------|
| 1       | 0            | 0.0035          | 0.0035     |       |
| 2       | 0            | 0.0049          | 0.0049     |       |
| 3       | 0            | 0.8666          | 0.8666     |       |
| 4       | 0            | 0.0415          | 0.0415     |       |
| 5       | 0            | 0.0001          | 0.0001     |       |
| 6       | 0            | 0.0025          | 0.0025     |       |
| 7       | 0            | 0.7726          | 0.7726     |       |
| 8       | 0            | 0.0007          | 0.0007     |       |
| 9       | 0            | 0.0008          | 0.0008     |       |
| 10      | 0            | 0.8067          | 0.8067     |       |
| 11      | 0            | 0.0004          | 0.0004     |       |
| 12      | 0            | 0.0021          | 0.0021     |       |
| 13      | 0            | 0.0001          | 0.0001     |       |
| 14      | 1            | 0.9987          | 0.0013     |       |
| 15      | 1            | 0.9917          | 0.0083     |       |
| 16      | 1            | 0.198           | 0.802      |       |
| 17      | 1            | 0.5405          | 0.4595     |       |
| 18      | 1            | 0.9965          | 0.0035     |       |
| 19      | 1            | 0.7974          | 0.2026     |       |
| 20      | 1            | 0.5105          | 0.4895     |       |
| 21      | 1            | 0.293           | 0.707      |       |
| 22      | 1            | 0.9996          | 0.0004     |       |

6. Conclusion

The proposed work emphasizes on early ASD detection. In this work, using PCA, there is a reduction in the number of attributes followed by use of 10-fold cross validation to train the data and finally deep learning technique to detect ASD on all categories of individuals with age group corresponding to toddler, child, adolescent and adult. Based upon contribution of minimal benefit, the attributes are reduced in the data set. The different evaluation parameters such as accuracy, sensitivity, specificity and F-measure yielded clinically acceptable results using DNN like other works whose results are also well acceptable. In other words, it can be assured that a deep learning model can be implemented for detecting ASD in addition with other conventional ML classifier model as earlier suggested by researchers. Generally in most of the research analysis, toddler data set is dropped because of its unbalanced nature making the detection of disorder difficult in that category. In this study, along with other categories, toddler data set is also analyzed thereby making it successful to predict ASD in toddlers.

References

[1] American Psychiatric Association 2013 *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association)

[2] Ruzich E, Allison C, Smith P, Watson P, Auyeung B, Ring H and Baron-Cohen S 2015 Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol. Autism* 6 1–12
[3] Brugha T S, McManus S, Bankart J, Scott F, Purdon S, Smith J, Bebbington P, Jenkins R and Meltzer H 2011 Epidemiology of Autism Spectrum Disorders in Adults in the Community in England Arch. Gen. Psychiatry 68 459–66

[4] Al Backer N B 2015 Developmental regression in autism spectrum disorder. Sudan. J. Paediatr. 15 21–6

[5] Robins D L, Fein D, Barton M L and Green J A 2001 The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. J. Autism Dev. Disord. 31 131–44

[6] Allison C, Auyeung B and Baron-Cohen S 2012 Toward Brief “Red Flags” for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls J. Am. Acad. Child Adolesc. Psychiatry 51 202–12

[7] Baron-Cohen S, Hoenstra R A, Knickmeyer R and Wheelwright S 2006 The Autism-Spectrum Quotient (AQ)—Adolescent Version J. Autism Dev. Disord. 36 343–50

[8] Auyeung B, Baron-Cohen S, Wheelwright S and Allison C 2008 The Autism Spectrum Quotient: Children’s Version (AQ-Child) J. Autism Dev. Disord. 38 1230–40

[9] Anon https://www.autismresearchcentre.com/

[10] Abdelhamid N and Thabtah F 2014 Associative Classification Approaches: Review and Comparison J. Inf. Knowl. Manag. 13 1–30

[11] Thabtah F 2019 An accessible and efficient autism screening method for behavioural data and predictive analyses Health Informatics J. 25 1739–55

[12] John G H and Langley P 1995 Estimating continuous distributions in Bayesian classifiers Proceedings of the Eleventh conference on Uncertainty in artificial intelligence pp 338–345

[13] Thabtah F, Abdelhamid N and Peebles D 2019 A machine learning autism classification based on logistic regression analysis Heal. Inf. Sci. Syst. 7 1–11

[14] Thabtah F, Kamalov F and Rajab K 2018 A new computational intelligence approach to detect autistic features for autism screening Int. J. Med. Inform. 117 112–24

[15] Huan Liu and Setiono R Chi2: feature selection and discretization of numeric attributes Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence (IEEE Comput. Soc. Press) pp 388–91

[16] Witten I and Frank E 2011 Data Mining: Practical Machine Learning Tools and Techniques (Elsevier)

[17] Hall M 1999 Correlation-Based Feature Selection for Machine Learning (The University of Waikato)

[18] Quinlan J R 1986 Induction of decision trees Mach. Learn. 1 81–106

[19] R V and R S 2018 A machine learning based approach to classify Autism with optimum behaviour sets Int J Eng Technol 8 4216–9

[20] Salzberg S L 1994 C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993 Mach. Learn. 16 235–40

[21] Keerthi S S, Shevade S K, Bhattacharyya C and Murthy K R K 2001 Improvements to Platt’s SMO Algorithm for SVM Classifier Design Neural Comput. 13 637–49

[22] Aha D W, Kibler D and Albert M K 1991 Instance-based learning algorithms Mach. Learn. 6 37–66

[23] Pal S K and Mitra S 1992 Multilayer perceptron, fuzzy sets, and classification IEEE Trans. Neural Networks 3 683–97

[24] Thabtah F and Peebles D 2020 A new machine learning model based on induction of rules for autism detection Health Informatics J. 26 264–86

[25] W. C W 1995 Fast effective rule induction Proceedings of the 12th International Conference on Machine Learning pp 115–23

[26] Gaines B R and Compton P 1995 Induction of ripple-down rules applied to modeling large databases J. Intell. Inf. Syst. 5 211–28

[27] Akter T, Shahriare Satu M, Khan M I, Ali M H, Uddin S, Lio P, Quinn J M W and Moni M A 2019 Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders IEEE Access 7 166509–27
[28] D. Mease, A. J. Wyner and A B 2007 Boosted classification trees and class probability/quantile estimation *J. Mach. Learn. Res.* **8** 409–39

[29] Zhao H, Fu L, Gao Z, Ye Q, Yang Z and Yang X 2019 Flexible non-greedy discriminant subspace feature extraction *Neural Networks* **116** 166–77

[30] Hofner B, Mayr A, Robinzonov N and Schmid M 2014 Model-based boosting in R: a hands-on tutorial using the R package mboost *Comput. Stat.* **29** 3–35

[31] Pourghasemi . A, Arabameri and H. R. 2019 Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R *Spat. Model. GIS R Earth Environ. Sci. Netherlands Elsevier* 299–321

[32] Tibshirani T H and R 1966 Discriminant Analysis by Gaussian Mixtures *J. R. Stat. Soc. Ser. B* **58** 155–76

[33] Hastie T, Buja A and Tibshirani R 1995 Penalized Discriminant Analysis *Ann. Stat.* **23** 73–102

[34] Raj S and Masood S 2020 Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques *Procedia Comput. Sci.* **167** 994–1004

[35] Thabtah F 2018 https://www.kaggle.com/fabdelja/autism-screening-for-toddlers/version/1.

[36] Thabtah F F 2017 https://archive.ics.uci.edu/ml/machine-learning-databases/00419

[37] Thabtah F F 2017 https://archive.ics.uci.edu/ml/machine-learning-databases/00420/

[38] Thabtah F F 2017 https://archive.ics.uci.edu/ml/machine-learning-databases/00426/

[39] Hotelling H 1933 Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24** 417–41

[40] Hochreiter S and Schmidhuber J 1997 Long Short-Term Memory *Neural Comput.* **9** 1735–80