PREDICTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING

Shivakiran Sastry G[1], Sanjay K[2], Sriram Manideep[3], Srividhya Ganesan[4]

Abstract- Autism spectrum disorder (ASD) is one of the neurocognitive disorders which is categorized by persistent difficulties since babyhood. The difficulties faced by an ASD individual commonly include less interaction with others, no response to the actions, reciprocal behaviors, hyperactive, etc. The research focused on the paper is concise to classification of an autistic children.

Keywords—Autism, Convolution Neural Network, Haar Cascade Classifier, Support Vector Machine, VGG16.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurocognitive disorder that deteriorates interaction, communication, and learning skills of a person. Normal people can understand and reply to their actions, but autistic children are poor at responding to the actions. Usually, Autistic children face different kinds of challenges like lack of concentration, learning impairment, psychological problems like anxiety, depression, sensory problems, locomotion problems etc.,

ASD diagnosis involves a significant amount of time and value. Prior detection of disorder can come to a great help by recommending patients with the appropriate medication at the premature phase. It improves the patient's illness from deteriorating further and would help to cut back on long-term expenses associated with overdue diagnosis.

The existing explosion rate of ASD around the world is numerous and is increasing at a high rate. According to World Health Organization, about one in out of 160 children is affected by autism. Some affected individuals can live on their own, while some require additional care and support. Therefore, time-efficient, accurate and straightforward predicting model is required which might predict autistic features in people and identify whether the individual require comprehensive assessment or not.

Autism seems to be strongly influenced by genetics. The research results show that in an identical twin the concordance rate is in between 60% - 90%, whereas concordance rates for fraternal twins & siblings lie in between 5%-10%. The autism is being implicated by various genes and gene mutations. The various genes involved are the ones which are responsible for the formation of synaptic circuits that enables the communication between different parts of the brain. Several environmental factors are thought to be related to an amplified risk of ASD, a minimum of partially, because they contribute to new mutations.

The research is aimed at the prediction of the autism spectrum disorder image dataset. The dataset of ASD is said to the pictures of kids. The dataset was collected from the "Kaggle website" URL: "https://www.kaggle.com/gpiosenka/autistic-children-data-set-traintestvalidate". Our Research focuses on classifying the facial image as autistic or non-autistic. After implementing classification model, based on the results obtained, it is possible to classify ASD through facial images.

II. METHODS / PROCEDURES

In our Research, an autism spectrum disorder dataset with the facial images were used. The image dataset is split into test, train and valid. The attributes of dataset are shown in the Table 1.

Attributes	Test	Train	Valid
Autistic	150	1268	50
Non-autistic	150	1268	50

Table 1: Features and Descriptions

Table 1 describes the image dataset, containing attributes such as Autistic and Non-autistic. Using this dataset, the model is trained with 2536 images (autistic: 1268, non-autistic: 1268), tested with 300 images (autistic: 150, non-autistic: 150), and validated with 100 images (autistic: 50, non-autistic: 50).

The images in the dataset are predicted using various ML algorithms.
They are:

1) Haar cascade classifier,
2) Support Vector Machine (SVM) Algorithm and
3) Convolution Neural network (CNN).

i. Haar Cascade Classifier.

Haar Cascade classifier is one of the prominent object detection approaches, proposed by Michael Jones and Paul Viola. It's a machine learning-based approach where a cascade function is trained from lots of positive and negative images.

Here we will work with facial images of children dataset. Firstly, the classifier training tool needs ample number of positive images (facial images of autistic children) and negative images (facial images of non-autistic children) to train the classifier. Figure (1) represents Haar features which are similar to the convolutional kernel. Every feature is a solitary value attained by subtracting the summation of pixels from white and black rectangles.

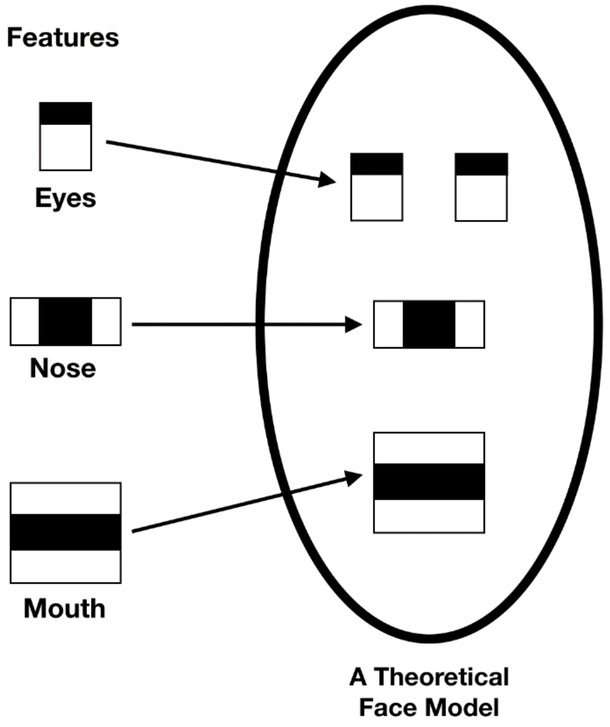

Figure (1)

The haar cascade classifier is generated using the **Cascade-Trainer-GUI** tool**.** Initially image dataset is classified into positive and negative image folders. A folder "P" with positive images (facial images of autistic children) and a folder with negative images (facial images of non-autistic children) at the ratio of 1:4. A custom haar cascade classifier is generated in XML format. Later the classifier is deployed in the code for predicting results.

ii. Support Vector Machine.

SVM is a supervised machine learning algorithm which is used for classification and regression problems. In SVM, classification of objects takes place swiftly with good precision and accuracy rate.

SVM has the strongest mathematical model for classification. In recent times, there have been numerous enhancements to SVM like, twin SVM, Lagrangian SVM, Quantum SVM, least square SVM, etc.

The below structure is produced by SVM which is linked with optimisation problem:

$$\min_{w,b,\xi}\left\{\frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i\right\}, \quad \text{constrains to}$$

$$y_i(w^T\phi(x_i)+b) \geq 1-\xi_i, \xi_i \geq 0,$$

where $(x_i, y_i), i = 1, \ldots, l$ is an instance-label pair, $x_i \in \mathbb{R}^n, y \in \{1,-1\}^l, \phi$ is a mapping function, $C > 0$ is a penalty parameter. The function $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$ is called the kernel function.

Types of Kernel functions and its structures are as follows:

Linear function: $K(x_i, x_j) = x_i^T x_j$,

Polynomial function: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$,

Radial basis function: $K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \gamma > 0$,

Sigmoid function: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$.

iii. Convolution Neural Network.

A Convolution neural network uses deep learning technique where an image is taken as input, allocates importance to varied details within the image and be ready to differentiate one from the opposite. The pre-processing required for ConvNet is lower when compared to other classification algorithms. CNN is most ordinarily applied to investigate visual imagery.

VGG16 is a convolution neural network (CNN) architecture named after the "Visual Geometry Group". To date it is still considered to be a wonderful vision model, although it's been somewhat outperformed by more modern advances like Inception and ResNet.

Among all fully-connected layers, only the last convolution layer is employed. The reason behind using the last convolution layer is to avoid fixed input size for the model i.e., 224×224. VGG is that the process of generating a sequence of values by analysing series of input values.

VGG16's last layer can detect up to 1000 categories of image. From **Table 1,** there are only 2 attributes to be categorized, i.e., autistic and non-autistic, and hence it is required to detect up to 2 categories only. Thus, last layer of VGG16 is eliminated.

<u>VGG Architecture</u>

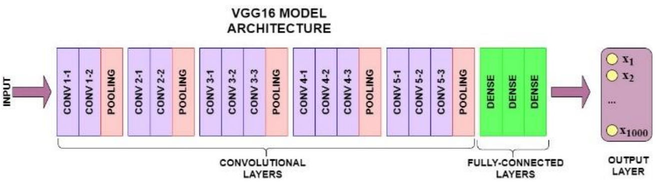

Figure (2)

VGG16 may be a classical convolutional neural spec. it absolutely was supported an analysis of a way to increase the depth of such networks.

From Figure (2), The kernel size is 3×3 and also the pool size is 2×2 for all the layers. The input to the VGG16 model is 224×224×3 pixels images, then we've two convolution layers with each 224×224×64 size, then we've a pooling layer which reduces the peak and width of the image to 112×112×64. Then we've two conv128 layers with each 112×112×128 size at that time we've got a pooling layer which again reduces the peak and width of the image to 56×56×128. Then we've three conv256 layers with 56×56×256 size, then again, a pooling layer reduces the image size to 14×14×512. Then again, we've three conv512 layers with every 14×14×521 layers, after that, we've a pooling layer with 7×7×521 then we've two dense or fully-connected layers with each of 4090 nodes. And ultimately, we've got a final dense or output layer with 1000 nodes of the scale which classify between 1000 classes of image net.

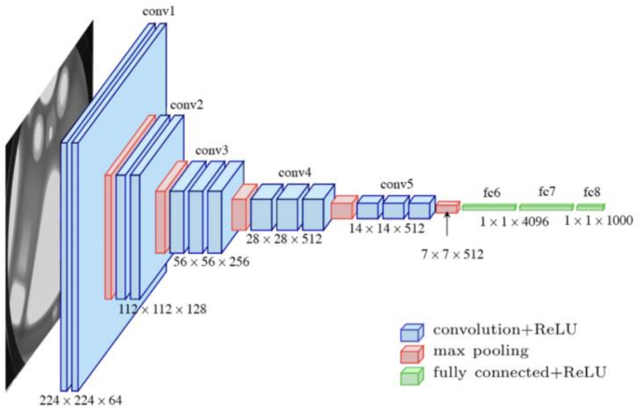

Figure (3)

VGG Working Example

In Image-Net dataset, the dimension of the image is fixed to 224×224 and it also contains RGB channels. Therefore, the tensor input is (224,224,3). The model is processed using above mentioned input and generates the 1000 vector values as output.

$$\hat{y} = \begin{bmatrix} \hat{y_0} \\ \hat{y_1} \\ \hat{y_2} \\ \hat{y_3} \\ . \\ . \\ . \\ . \\ .\hat{y_{999}} \end{bmatrix}$$

The below mentioned vector representation depicts the probability classification for the equivalent classes. Consider a model that predicts the image that belongs to 0^{th} class with probability 1, the probability of class 1= 0.05, class 2= 0.05, class 3= 0.03, class 780 = 0.72, class 999= 0.05 and every one the opposite class with 0.

$$\hat{y} = \begin{bmatrix} \hat{y_0} = 0.1 \\ 0.05 \\ 0.05 \\ 0.03 \\ . \\ . \\ . \\ \hat{y_{780}} = 0.72 \\ . \\ . \\ \hat{y_{999}} = 0.05 \end{bmatrix}$$

The SoftMax function is used in order to add all the probabilities to 1 and the function is shown below:

softmax function

$$P(y=j \mid \Theta^{(i)}) = \frac{e^{\Theta^{(i)}}}{\sum_{j=0}^{k} e^{\Theta_k^{(i)}}}$$

where $\Theta = w_0x_0 + w_1x_1 + ... + w_kx_k = \sum_{i=0}^{k} w_ix_i = w^Tx$

Now, 5 most probable classes are considered into the vector

$$C = \begin{bmatrix} 780 \\ 0 \\ 1 \\ 2 \\ 999 \end{bmatrix}$$

The following vector represents the ground truth vector:

$$G = \begin{bmatrix} G_0 \\ G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} 780 \\ 2 \\ 999 \end{bmatrix}$$

The Error function is shown below:

$$E = \frac{1}{n} \sum_k min_i d(c_i, G_k)$$
$$where\, d = 0\, if\, c_i = G_k\, else\, d = 1$$

For the example, the loss function is defined as :

$$E = \frac{1}{3}\left(min_i d(c_i, G_1) + min_i d(c_i, G_2) + min_i d(c_i, G_3)\right)$$

Therefore,

$$E = \frac{1}{3}(0 + 0 + 0)$$

$$E = 0$$

Finally, the loss becomes 0 when all the categories in ground truth are in the predicted top-5 matrix.

VGG Configuration

Table2 lists different VGG architecture. There are two different versions of VGG16 i.e., C & D. When compared between two variants of VGG16, there is no much difference, apart from only few layers of convolution. These two contains 1.34 crore and 1.38 crore parameters respectively.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

III. RESULTS

The implementation of predicting Autism Spectrum Disorder was carried out in Python IDE, Jupyter Notebook.

The following parameter is applied to evaluate the performance of classification models.

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN}$$

i. Haar Cascade Classifier.

A custom haar cascade classifier was generated using "Cascade-Trainer-GUI" and the same was used in implementing the prediction results.

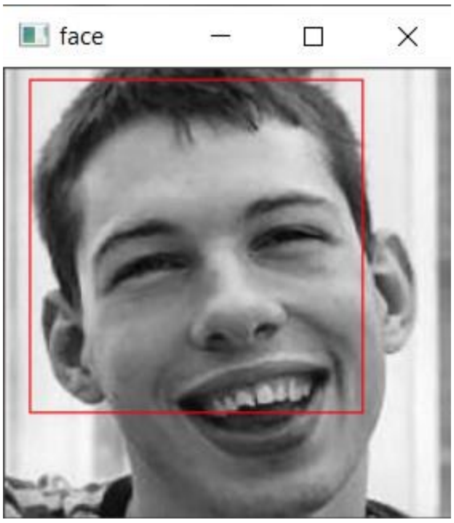

Figure (4)

Figure (4) represents the prediction of an autistic image. As the rectangular box in the image depicts that the image is predicted as autistic.

ii. Support Vector Machine.

An SVM model is generated by training against the dataset using "Kernel Regularization Function" with L2 norm parameter by setting value to 0.01.

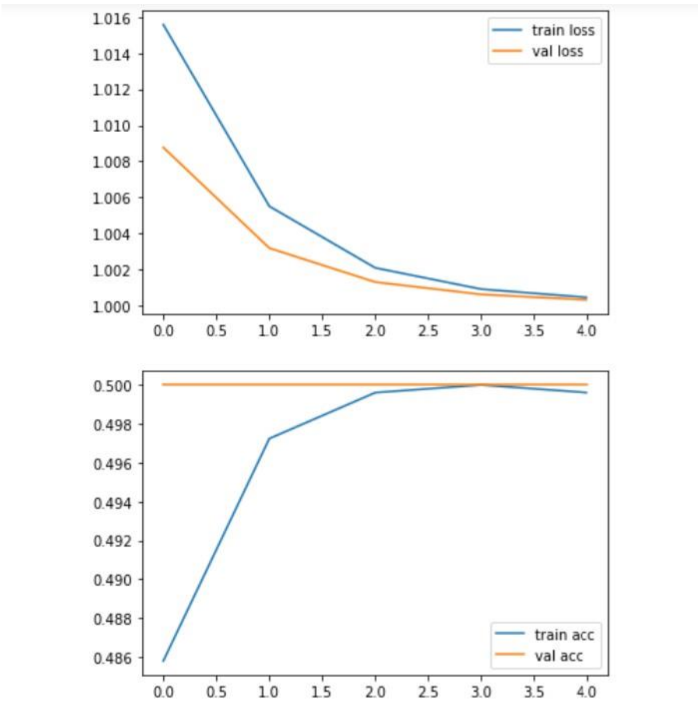

Figure (5)

The **Figure (5)** represents training accuracy, training loss, validation accuracy, value loss for SVM model.

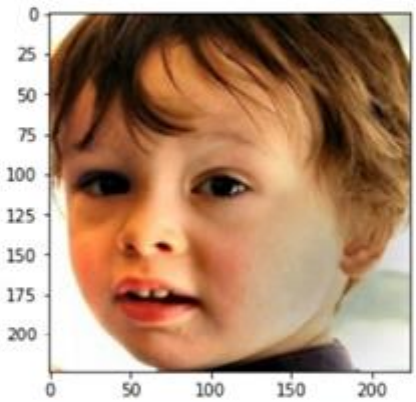

Figure (6)

The **Figure (6)** is predicted as autistic image using the SVM model.

iii. Convolution Neural Network.

A CNN model is generated using VGG16 by collecting the weights from ImageNet database.

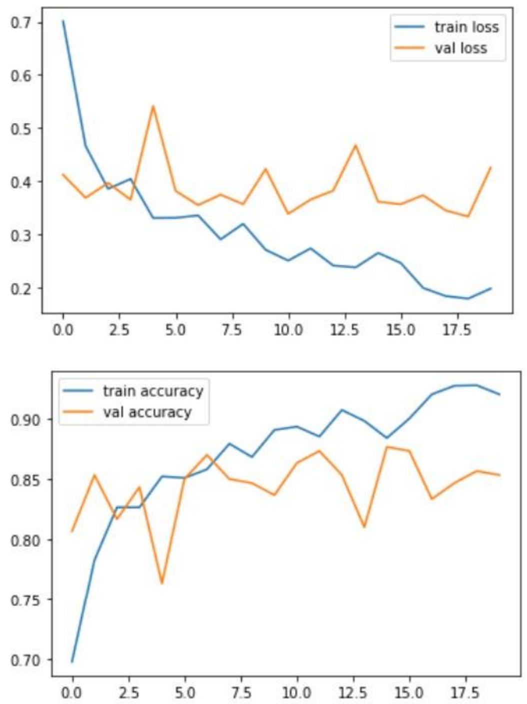

Figure (7)

The **Figure (7)** represents training accuracy, training loss, validation accuracy, value loss for VGG16 of CNN model.

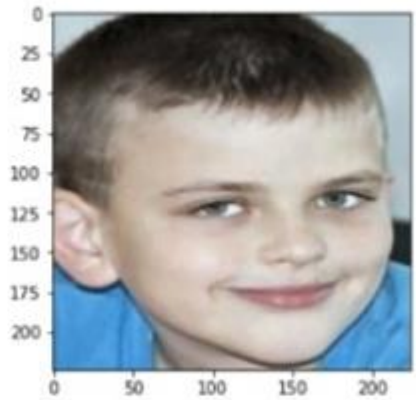

Figure (8)

The **Figure (8)** is predicted as autistic image using the VGG16 of CNN model.

IV. DATA ACCESSIBILITY

Data used for the discoveries of paper can be accessed from "Kaggle Machine Learning Repository" website at "https://www.kaggle.com/gpiosenka/autistic-children-data-set-traintestvalidate".

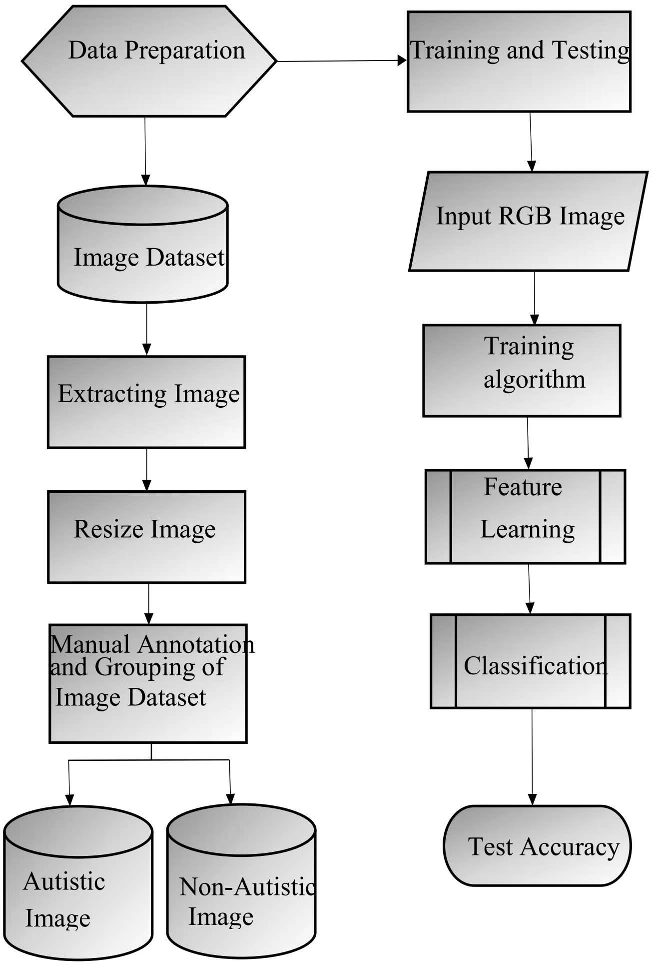

Figure (9): Implementation Flowchart

V. CONCLUSION

Algorithms	Accuracy
Haar Cascade Classifier	72%
Support Vector Machine	65%
Convolution Neural Network (VGG16)	90%

Autism is taken into account together of the rapidly-growing neuro cognitive disorders among the children, Therefore the research for the primary diagnosis with the help of classification models would contribute in solving the matter of creating an accurate valuation. The research is focused on the event of classification models using the VGG16 algorithm of CNN, SVM classifier, Haar Cascade classifier using OpenCV. Thus, we are able to predict autism spectrum disorder using facial details using above mentioned models. Therefore, achieved better accuracy from VGG16 (90%) compared to other classification models

REFERENCES

[1] Srividhya Ganesan, Sustainable Development of Green Healthcare Communities for Prediction of Autism Spectrum Disorder using Machine learning approach, Journal of Green Engineering, Scopus Indexed Journal, Volume-10, Issue -9, September 2020.

[2] Srividhya Ganesan, Sustainable Development of Green Healthcare Communities for Prediction of Autism Spectrum Disorder using Machine learning approach, Journal of Psychology and Education, January 2021

[3] Kos Micki JA, Sochat V, Duda M, Wall DP. Attempting to find a minimal set of behaviours for autism detection through feature selection-based machine learning. Trans Psychiatry.2015;5(2): e514.

[4] Riley M, Karl J, Chris T. A study of early stopping, assembling, and patch working for cascade correlation neural networks. IAENG Int J Applied Mathematics. 2010;40(4):307-16.

[5] Allison C, Baron-Cohen S, Charman T, Wheelwright S, Richler J, Pasco G, et al. The QCHAT (quantitative checklist for autism in toddlers): A normally distributed quantitative measure of autistic traits at 18-24 months of age: preliminary report. J Autism Dev Disord.2008;38(8):1414-25

[6] Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioural distinction of autism and ADHD. Trans Psychiatry. 2016;6(2): e732.

[7] Bone D, Goodwin MS, Black MP, Lee CC, Audhkhasi K, Narayanan S. Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. J Autism Dev Disord. 2015;45(5):1121-36.

[8] Yonekura A, Kawanaka H, Surya Prasath VB, *et al.*

[9] Escudero J, Ifeachor E, Zajicek JP, Green C, Shearer J, Pearson S. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. IEEE Trans Biomed Eng 2013; 60(1): 164-8. [http://dx.doi.org/10.1109/TBME.2012.2212278]

[10] Liu S, Liu S, Cai W, *et al.* Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans Biomed Eng 2015; 62(4): 1132-40. [http://dx.doi.org/10.1109/TBME.2014.2372011] [PMID: 25423647]

[11] Abdar M, Zomorodi-Moghadam M, Zhou X, *et al.* A new nested ensemble technique for automated diagnosis of breast cancer. Pattern Recognit Lett [Internet] 2018.

[12] Adem K, Kiliçarslan S, Cömert O. Classification and diagnosis of cervical cancer with softmax classification with stacked autoencoder. Expert Syst Appl 2019; 115: 557-64. [Internet]. [http://dx.doi.org/10.1016/j.eswa.2018.08.050]

[13] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992; 43(3): 175-85.

[14] Dua D, Graff C. UCI Machine Learning Repository. Available at: http://archive.ics.uci.edu/mlIrvine.

[15] Thabtah F. Autism Spectrum Disorder screening: Machine learning adaptation and DSM-5 fulfilment. ICMHI '17 Proceedings of the first International Conference on Medical and Health Informatics 2017; Taichung City, Taiwan May 20-22.

[16] Thabtah F. ASDTests. A mobile app for ASD screening. [Internet]. 2017 [cited 2018 Dec 20]. Available from: ww.asdtests.com

[17] Thabtah F. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. Informatics Heal. Soc. Care 2019; 44(3): 278-97.

[18] Liu, B., Hsu, W., Ma, Y, ―Integrating classification and association rule mining‖, Proceedings of the fourth international conference on knowledge discovery and data mining, pp.80–86, 1998.

[19] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder", International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, pp. 1-6, 2019.

[20] Alwidian J., Hammo B., and Obeid N, ―Enhanced CBA algorithm Based on Apriori Optimization‖, The 28th IBIMA conference on Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth, 2016.