JCRES
Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 53

Comparative Analysis of Co-factors Used to Predict Autism Spectrum Disorder

Abstract—**Autism Spectrum Disorder(ASD) is a neurodevelopmental disease that instills in parents a trepidation and concern for their children. Hence, predicting if their child is autistic or not would help them cope with the situation. To identify the early symptoms of autism a number of co-factors of various types are used. This paper explores these co-factors by comparing them using several deep-learning algorithms. Each co-factor has some kind of challenge attached to it. Improvising and enhancing the data set using preprocessing techniques is a major requirement before applying the algorithms. A combination of all the co-factors would be an ideal data set, but doing some may create a cumbersome data set that might increase the complexity of the algorithm. Besides, each kind of co-factor has specific and nuanced algorithms designed for their data type. Hence by incorporating the required kind of algorithm for each co-factor, a comparison has been drawn to choose the most suitable co-factor to predict ASD. The two kinds of data sets used are behavioral and facial image data sets. The techniques used for the behavioral data set are feature selection, optimization techniques and ANN(artificial neural network), and the techniques used for the facial image data set are VGG16, transfer learning, and CNN(Convolutional neural network)**

I. Introduction

ASD stands for Autism Spectrum Disorder[1] which is neurodevelopmental disorder that affects social interaction, communication and behaviour. The term “spectrum” refers to the fact that the symptoms vary from person to person and mild to severe.

People with ASD may have difficulty with social skills and communication, such as understanding and using nonverbal cues like eye contact, tone of voice, and facial expressions[2]. They may also have repetitive behaviors or narrow interests and may struggle with sensory processing. ASD is typically diagnosed in early childhood, but sometimes symptoms may not become apparent until later. While there is no cure for ASD, early intervention and support can help individuals with ASD improve their communication, socialization, and behavior, and lead fulfilling lives.

Although the clinical approach would be to detect ASD using chromosomes[3], machine learning can also be used for the detection of ASD in kids and adults by employing ML algorithms on data sets. Machine learning algorithms can be trained on large data sets to identify patterns and correlations that may not be readily apparent to human observers. These algorithms can analyze complex data and generate predictive models that can help identify individuals at risk for autism or provide support for clinical decision-making.

Machine learning can also be used to integrate multiple sources of data, such as behavioral, physiological, and neuroimaging data, to improve the accuracy and reliability of autism detection. By combining data from different modalities, machine learning algorithms can capture complex interactions and relationships that may not be evident in individual data sources alone[4].

However, it's important to note that machine learning for autism detection[2] is still in the early stages of development and further research is needed to validate its accuracy, reliability, and ethical implications. It's also crucial to ensure that supervised machine learning algorithms[5] are transparent, interpretable, and adhere to ethical guidelines to ensure this technology's fair and responsible use in autism diagnosis and intervention. Nonetheless, machine learning holds promise as a potential tool for the early detection of autism, which can contribute to improved outcomes for individuals with autism and their families.

One approach to detecting autism using machine learning involves the analysis of behavioral data, such as speech, eye gaze, facial expressions, and motor movements. For example, machine learning algorithms can be trained to analyze speech patterns or eye gaze data during social interactions to identify characteristics that are indicative of autism[6].

Another approach involves using machine learning algorithms to analyze neuroimaging data, such as functional magnetic resonance imaging[7] (FMRI) or electroencephalography (EEG) data[8]. These algorithms can identify patterns of brain activity that are associated with autism, providing insights into the neural mechanisms underlying the disorder.

The kind of co-factors used can range from behavioral and facial image data sets to EEG signals and MRI scans. All the data sets have only a specific kind of information, hence an ideal data set would be a combination of all the co-factors. However, such a data set would increase the complexity and bring about other challenges pertaining to the compatibility of the data types of each co-factor.

In this paper, the comparative analysis of co-factors for ASD detection uses feature selection methods. Behavioral data set is obtained from Kaggle which contains the questionnaire, and other personal details such as age and ethnicity. The facial image data set is also obtained from Kaggle which contains 2450 train images, with 1225 images each in autistic and non-autistic categories.

II. Related Works

Michael Lombardo and Simon Baron-book Cohen's[8] paper offers a thorough analysis of the many risk factors for ASD. Another study, by Julie Daniels et al., explores different risk factors for ASD based on a thorough analysis of the literature already in existence. [10]A study about early autism spectrum disorder detection by Costanzo.V et al looked at home movies of children with ASD and younger siblings to find early behavioral indicators that might indicate later ASD development.

In a study paper by P.Kavya Sree et al.[11] it is suggested that supervised learning algorithms be used to determine whether a kid will have ASD. The research makes use of a data set that includes behavioral and demographic data from kids with and without ASD. The effectiveness of various supervised learning techniques, such as logistic regression, decision trees, random forests, and support vector machines (SVM), in predicting ASD is evaluated by the authors. They also look into how feature selection affects how accurate the models are.

The paper by Vishal.V et al.[12] aims to compare the performance of different machine learning algorithms in predicting Autism Spectrum Disorder (ASD) using a data set of behavioral features. The study suggests that machine learning algorithms can be effective in predicting ASD using behavioral features, with SVM being the most accurate algorithm in this study. However, further research is needed to validate these results and explore the potential of other machine-learning techniques in ASD prediction.

The article [13] by Sadiq et al predicts the diagnosis of ASD in kids between the ages of 4 and 6. This study used machine learning algorithms to examine eye-tracking and EEG data.

To understand the risk of ASD in infants, a study [14] was conducted by Harel-Gadassi et al. The aim of the study was to examine the risk of ASD in the long term in individuals who are born preterm and full-term using both observational instruments and parental reports. Neonatal risk factors and developmental characteristics associated with ASD risk were also examined.

Oller et al article [15] described how automated speech analysis was utilized to find vocal characteristics that could help diagnose ASD in children between the ages of three and five.

Another research using a different co-factor was the paper[16] by M.Mishra et al. This paper explores the method of using morphometric features of T1 weighted MRI to detect Autism Spectrum Disorder.

In the study paper by Nafisa Sadaf Hriti et al[17], the use of visual and behavioral data for autism classification is examined. The study hopes to advance earlier research that classified autism using only one piece of data.

III. Algorithms and Methods

Two publicly available data sets containing information about symptoms that can be used to detect autism are taken for analysis. One data set contains information about ethnicity, gender, answers to the screening questions, etc. In contrast, the other data set contains images of autistic and non-autistic children in two folders, one of which has been categorized while the other is shuffled. A comparison is drawn between the two data sets based on their performance using machine learning models[4].The image data set contains 2450 train images, with 1225 images each in autistic and non-autistic categories.

A. For behavioral data set

1) *Feature Selection:* The Pearson Correlation method is used to select features for behavioral data. The Pearson correlation method is a statistical approach that evaluates the linear association between two variables. By utilizing Pearson correlation in feature selection[18], one can identify the most robust attributes that are linked to the target variable. To employ Pearson correlation for feature selection, an individual would need to calculate the correlation coefficient between each attribute and the target variable. The correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 denotes a perfect positive correlation, and 0 represents no correlation.

2) *Artificial Neural Network:* The Pearson Correlation method is used to select features for behavioral data. The Pearson correlation method is a statistical approach that evaluates the linear association between two variables. By utilizing Pearson correlation in feature selection[18], one can identify the most robust attributes that are linked to the target variable. To employ Pearson correlation for feature selection, an individual would need to calculate the correlation coefficient between each attribute and the target variable. The correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 denotes a perfect positive

correlation, and 0 represents no correlation.

3) *Genetic Algorithm:* In machine learning, genetic optimization can be applied to feature selection[18] or to determine the best hyperparameters for a particular model. Hyperparameters are variables that are set before the model is trained, such as the learning rate, regularization intensity, or number of hidden levels in a neural network. Finding the ideal hyperparameters can greatly enhance the model's efficiency.

B. For image data set

1) *Convolutional Neural Network(CNN):* Machine learning includes convolutional neural networks, also known as convnets or CNNs. It is a subset of the several artificial neural network models that are employed for diverse purposes and data collection. [19]A CNN is a particular type of network design for deep learning algorithms that are utilized for tasks like image recognition and pixel data processing.

Three layers make up the traditional CNN architecture: a pooling layer, a convolutional layer, and a final fully connected layer. One of the fundamental layers of CNN, the convolutional layer performs the majority of the calculation effort[2]. This layer's primary function is to filter and convolution the initial data before passing it on to the final result of the following layer. The appropriate features from the entire input are then selected, and the filter is then applied to the following layer to form a feature map. The spatial reduction of the processing time and spatial representation is accomplished by the pooling layer, which is placed between the subsequent convolutions. To cut down on the computational overhead, this layer pools each of the separated inputs for the following convolutional layer. The Convolutional and pooling layers only choose the pertinent features from the input images. The final output, which is equal to the number of classes, is provided by the completely connected layer.[13]

2) *Transfer Learning:* Transfer learning is a machine learning technique in which a model is trained on a large data set and then reused on a smaller data set for a new task or domain. The idea behind transfer learning is that the model's knowledge accumulated on a large data set can be transferred to a new task or domain, resulting in faster training and improved performance.

3) *Pretrained Model:* The VGG16 model, which is a convolutional neural network with 16 layers, is also referred to as the VGG model.[19]. It has achieved exceptional accuracy in the top-5 test of the ImageNet data set, which contains almost thousand distinct categories and over fourteen million images. At ILSVRC-2014, it was one of the most frequently used models. Rather than utilizing filters with large kernel sizes, it employs a series of filters with kernel sizes of 3×3. As previously stated, the VGG16 comprises 16 layers and can categorize images into thousand classifications, including keyboards, pencils, and mice.

VGG16 is composed of 16 layers, in which 13 are convoluted and three are completely connected. The input is an image with dimensions of 224x224 and RGB color space. The convolutional layers employ filters of 3x3 size using a stride of one and padding of one. Meanwhile, the max pooling layers utilize a window of 2x2 size with a stride of 2.The quantity of filters increases as the model's depth increases, from 64 in the first layer to 512 in the final. The fully linked layers each have 4096 neurons, with a last output layer of 1000 neurons, matching the ImageNet data set's 1000 classifications.

4) *T-SNE Visualization:* The t-SNE technique is a statistical approach that enables the visualization of data with high dimensions by allocating a position to each data point in a map consisting of two or three dimensions.This is accomplished by using a Gaussian kernel assign a probability to each data point that is proportional to how similar it is to other points in the high-dimensional point.

For the application of t-SNE to image data, the first step is to extract significant features from the images. This can be achieved by using pre-trained CNNs like VGG, ResNet, or Inception, which have undergone training on extensive data sets like ImageNet and can extract a diverse range of image features.

After extracting the features from the images, t-SNE can be employed to represent these features in a lower-dimensional space.The resulting representation will demonstrate the similarities and distinctions between the images based on their extracted features.

IV. Implementation

The Image data set consists of images of children with autism and those without autism. The data set is split into the train, valid, and test folders.[20] As part of preprocessing, the images are resized uniformly and converted to a numpy array before feeding them to the model. Following this, the T-SNE visualization algorithm is applied to plot the distribution of the images in those 2 classes. Then the same process is applied to validation data.

For the task of image classification, a pre-existing CNN model named VGG16, which has been trained on a vast amount of image data, is employed by adopting a transfer learning technique.[21] Next, the pre-existing model is linked to a custom fully linked layer using the sigmoid activation

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 53

function. The model is subsequently compiled by designating the loss function, optimizer, and metrics for the model.

Furthermore, we use t-sne visualization to visualize the output of the last layer in a low-dimensional space and map the predicted labels to the t-sne visualization to see how the CNN is making its predictions and to ensure that the results are robust and reliable. The test data is resized and converted into a numpy array. Model validation is done by checking the prediction of the sample test images with the ones already classified in the valid folder.[4]

The behavioral data set contains the screening questionnaire consisting of 10 behavioral features along with 10 individual characteristics.[22] The preprocessing of this data set is done by label encoding and checking for missing values. Following this feature selection is done using Pearson correlation to select features that are highly correlated with the output. Next, the data set is split into training and testing sets, with the training set used to train the ANN.

Then the model is built by adding the input layer and first hidden layer and the activation function and fits into a training set, then it is compiled using an Adam optimizer, loss function, and evaluation metrics.[23] Next, the evaluation of the model's performance is carried out using the test set. Furthermore, we employed a genetic approach to reduce overfitting and boost the model's adaptability. The fitness function is created and combined with ANN and genetic algorithms using the pygad library. The parent type is then defined, after which the mutation, crossover for genetics, and best fitness value are evaluated. Finally, the overall performance of the model is evaluated.

V. Discussion

To choose the best set of co-factors used to detect autism using machine learning, just one algorithm on each data set would not be sufficient.[24] A composition of all co-factors there is- Image Data, Video Data, EGG Signals, Parental Questionnaire, MRI, and Ultrasound scans would be ideal but the complexity of the process and the time taken would be extremely high. A well-detailed CSV file would suffice, provided that it is balanced and contains all required parameters to detect autism. Here, we have chosen two data sets- image and behavioral to compare and pick which would be better to detect autism[20]. The results of applying an artificial neural network combined with a genetic optimization algorithm are compared with existing models SVM and random forest. ANN seems to have the highest accuracy compared to the other two, with an accuracy of 98%.

Then the results of applying the VGG16 algorithm are compared with existing models GoogleNet and MobileNet-224 algorithms[20]. The accuracy of the VGG16 model seems to be high compared to the other two at 80%, still resulting less than the ANN model using behavioral data.

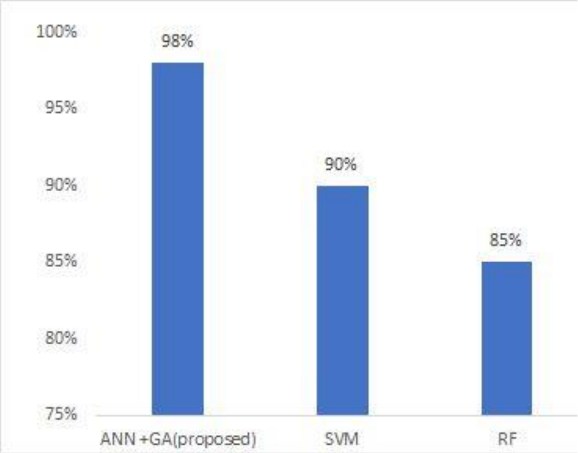

Fig 1: Comparison of proposed model with existing models for behavioral data.

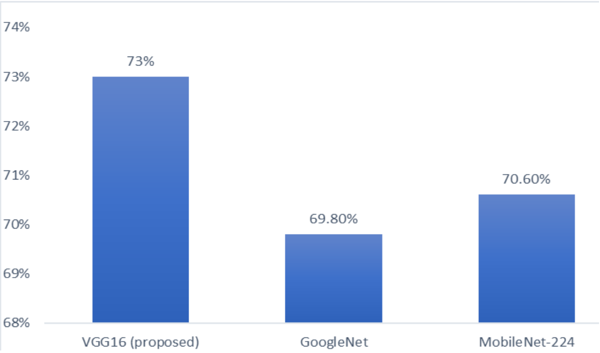

Fig 2: Comparison of the results of the proposed system against existing models for Image data.

VI. Conclusion

The results of the study suggest that using behavioral data as input for machine learning models would provide the most significant results. Although EEG signals data provides a closer accuracy, it is a difficult data set to obtain on top where overfitting and robustness exist as problems.

Autism detection using images is a complex task, and extracting features from them is difficult, but behavioral data not only provides information about autism but also aids in the early diagnosis of the disorder in children. And in identifying the most important features that contribute to the prediction. This could help to build trust in the model and increase its clinical utility. We conclude that behavioral data is more significant in detecting autism and is valuable information for both parents and doctors in early autism identification and diagnosis.

Future work would focus on collecting more data from various sources and on improving the proposed machine-learning classifiers to enhance their accuracy. Adding to that, building a web application for the proposed model would make it feasible for future researchers to understand and access information. Furthermore, pertaining to the co-factors, the ASD prediction process can be improved by integrating behavioral function and EEG data-based analysis.

REFERENCES

[1] Hodges H, Fealko C, Soares N. Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. Transl Pediatr. 2020 Feb;9(Suppl 1):S55-S65. doi: 10.21037/tp.2019.09.09. PMID: 32206584; PMCID: PMC7082249.

[2] Zeyad A. T. Ahmed, Theyazn H. H. Aldhyani, Mukti E. Jadhav, Mohammed Y. Alzahrani, Mohammad Eid Alzahrani, Maha M. Althobaiti, Fawaz Alassery, Ahmed Alshaflut, Nouf Matar Alzahrani, Ali Mansour Al-madani, "Facial Features Detection System To Identify Children With Autism Spectrum Disorder: Deep Learning Models", Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 3941049, 9 pages, 2022. https://doi.org/10.1155/2022/3941049

[3] Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapduram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet. 2008 Feb;82(2):477-88. doi: 10.1016/j.ajhg.2007.12.009. Epub 2008 Jan 17. PMID: 18252227; PMCID: PMC2426913.

[4] Raj, S., & Masood, S. (2020, April 16). “Analysis and detection of autism spectrum disorder using machine learning techniques”. Procedia Computer Science. , Vol. 167, 2020, pages 994-1004. https://doi.org/10.1016/j.procs.2020.03.399

[5] P. Kavya Sree1, A. Phani Sai Kumar2, R.Nandini,” Predicting Autism Spectrum Disorder using Supervised Learning Algorithms”, International Research Journal of Engineering and Technology (IRJET) ISSN: 2395-0056 Volume: 09 Issue: 04 | Apr 2022 https://www.irjet.net/archives/V9/i4/IRJET-V9I4357.pdf

[6] Rashid, Ali & Hameed, Shaimaa. (2022). Review of Autistic Detection Using Eye Tracking and Vocalization Based on Deep Learning. Journal of Algebraic Statistics. 13. 286-297.

[7] Bahathiq, R. A., Banjar, H., Bamaga, A. K., & Jarraya, S. K. (2022). Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. *Frontiers in Neuroinformatics*, *16*. https://doi.org/10.3389/fninf.2022.949926

[8] M C Lai, M V Lombardo, Prof s Baron-Cohen, “Autism”, Lancet, Seminar, Vol. 383, ISSUE 9920, September 2013. https://doi.org/10.1016/S0140-6736(13)61539-1/

[9] Chaste P, Leboyer M. Autism risk factors: genes, environment, and gene-environment interactions. Dialogues Clin Neurosci. 2012 Sep;14(3):281-92. doi: 10.31887/DCNS.2012.14.3/pchaste. PMID: 23226953; PMCID: PMC3513682.

[10] Costanzo V, Chericoni N, Amendola FA, Casula L, Muratori F, Scattoni ML, Apicella F. Early detection of autism spectrum disorders: From retrospective home video studies to prospective 'high risk' sibling studies. Neurosci Biobehav Rev. 2015 Aug;55:627-35. doi: 10.1016/j.neubiorev.2015.06.006. Epub 2015 Jun 17. PMID: 26092266.

[11] P. Kavya Sree, A. Phani Sai Kumar, R.Nandini, Tanay Lodh,N Sai Susmitha Naidu,”Predicting Autism Spectrum Disorder using Supervised Learning Algorithms”,International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 09 Issue: 04 , Apr 2022 https://www.irjet.net/archives/V9/i4/IRJET-V9I4357.pdf

[12] V. Vishal, A. Singh, Y. B. Jinila, K. C, S. P. Shyry and J. Jabez, "A Comparative Analysis of Prediction of Autism Spectrum Disorder (ASD) using Machine Learning," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1355-1358, doi: 10.1109/ICOEI53556.2022.9777240.

[13] Marjane Khodatars, Afshin Shoeibi, Delaram Sadeghi, Navid Ghaasemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem, Roohallah Alizadehsani, Assef Zare, Yinan Kong, Abbas Khosravi, Saeid Nahavandi, Sadiq Hussain, U. Rajendra Acharya, Michael Berk, “Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review”, Computers in Biology and Medicine, Volume 139, 2021, 104949, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2021.104949

[14] Harel-Gadassi A, Friedlander E, Yaari M, Bar-Oz B, Eventov-Friedman S, Mankuta D, Yirmiya N. Risk for ASD in Preterm Infants: A Three-Year Follow-Up Study. Autism Res Treat. 2018 Nov 11;2018:8316212. doi: 10.1155/2018/8316212. PMID: 30534432; PMCID: PMC6252203.

[15] D. K. Oller, P. Niyogi, S. Gray, S. F. Warren,”Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development”,PNAS ISSN:107 (30) 13354-13359 Vol. 107 Issue No: 30 |July 19, 2010 https://doi.org/10.1073/pnas.1003882107

[16] M. Mishra and U. C. Pati, "Autism Spectrum Disorder Detection using Surface Morphometric Feature of sMRI in Machine Learning," 2021 8th International Conference on Smart Computing and Communications (ICSCC), Kochi, Kerala, India, 2021, pp. 17-20, doi: 10.1109/ICSCC51209.2021.9528240.

[17] Nafisa Sadaf Hriti, Karishma Shaer, Farhan M Nafis Momin, Hasan Mahmud, Md. Kamrul Hasan," Autism Classification using Visual and Behavioral Data",medRxiv 2022.10.28.22281655; doi: https://doi.org/10.1101/2022.10.28.22281655

[18] M. B. Mohammed, L. Salsabil, M. Shahriar, S. S. Tanaaz and A. Fahmin, "Identification of Autism Spectrum Disorder through Feature Selection-based Machine Learning," 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2021, pp. 1-6, doi: 10.1109/ICCIT54785.2021.9689805.

[19] Fawaz Waselallah Alsaade, Mohammed Saeed Alzahrani, "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8709145, 10 pages, 2022. https://doi.org/10.1155/2022/8709145

[20] Ahmed ZAT, Aldhyani THH, Jadhav ME, Alzahrani MY, Alzahrani ME, Althobaiti MM, Alassery F, Alshaflut A, Alzahrani NM, Al-Madani AM. Facial Features Detection System To Identify Children With Autism Spectrum Disorder: Deep Learning Models. Comput Math Methods Med. 2022 Apr 4;2022:3941049. doi: 10.1155/2022/3941049. PMID: 35419082; PMCID: PMC9001065.

[21] V. Kavitha and R. Siva, "Review of Machine Learning Algorithms for Autism Spectrum Disorder Prediction," *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, 2022, pp. 608-613, doi: 10.1109/ICACRS55517.2022.10029201.

[22] S. B. Shuvo, J. Ghosh and A. S. Oyshi, "A Data Mining Based Approach to Predict Autism Spectrum Disorder Considering Behavioral Attributes," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944905.

[23] Hosseini MP, Beary M, Hadsell A, Messersmith R, Soltanian-Zadeh H. Deep Learning for Autism Diagnosis and Facial Analysis in Children. Front Comput Neurosci. 2022 Jan 20;15:789998. doi: 10.3389/fncom.2021.789998. PMID: 35126078; PMCID: PMC8811190.

[24] Vakadkar, K., Purkayastha, D. & Krishnan," Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques", SN COMPUT. SCI. 2, 386 (2021). https://doi.org/10.1007/s42979-021-00776-5