AI based Classification for Autism Spectrum Disorder Detection using Video Analysis

1st Shivani Pandya
CSE Department
Nirma University
Ahmedabad, India
21ftphde57@nirmauni.ac.in

2nd Swati Jain
CSE Department
Nirma University
Ahmedabad, India
swati.jain@nirmauni.ac.in

3rd Jai Prakash Verma
CSE Department
Nirma University
Ahmedabad, India
jaiprakash.verma@nirmauni.ac.in

***Abstract*—Autism spectrum Disorder(ASD) is a complex neurobehavioral disorder that affects a person's ability to communicate and interact with others. It is also characterized by repetitive behaviors and restricted interests. There is no one-size-fits-all approach to autism, but early intervention and treatment can make a big difference in a person's life. Machine learning and deep learning are two promising areas of research that may help to improve our understanding of autism and lead for better treatments. Machine learning and Deep Learning approaches of artificial intelligence allows computers to learn from data without being explicitly programmed. These models could potentially be used to improve our ability to communicate with, and understand people with autism. Various machine-learning techniques are used to predict autism at an early stage. Support Vector Machine (SVM), Decision tree, Naïve Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbour are some of the machine learning techniques used in this research area. Various advancement in the field of machine learning and Artificial Intelligence (AI) has helped in the development of ASD Detection using Machine learning and Deep Learning. In this research work, the prediction of Autism Spectrum Disorder has been performed on a video dataset. The video dataset contains the video of Autistic and Non-Autistic kids performing four different actions. The video features have been extracted through Convolutional Neural Network(CNN) models such as InceptionV3 and Resnet50 and are trained through long Short Term Memory(LSTM) based models by using this we get 91% accuracy.**

***Index Terms*—ASD, Machine Learning, Deep Learning, Prediction**

I. Introduction

Autism Spectrum Disorder (ASD) is a complex neurological and developmental disorder [1] that affects a person's ability to communicate and interact with others. In the first three years of life, children with autism have developmental disorders. Autism appears as cognitive deficits in the three cognitive domains of communication (verbal and nonverbal), social interaction, and imagination, which can be observed in repetitive and constrained play or social activities. ASD can range from mild to severe and affects people of all races and ethnicity. The exact cause of ASD is unknown, but it is believed to be caused by a combination of genetic and environmental factors [2]. There is no cure for ASD, but early intervention and support can make a big difference in the lives of people with ASD and their families.

Currently, no treatment has been shown to manage to cope with the symptoms of ASD, but several interventions have been developed and studied for use with young children. Interventions in the form of medical help may reduce the harmfulness of the symptoms. It also helps to improve cognitive ability, the ability of the child to function of daily living style, and the participating in the community activities [3]. According to a survey in the united state 1 in 44 children identified with autism in 2016 [4], whereas in India, due to less digitization available data sets on the prevalence of autism are insufficient. As per [5], the prevalence of ASD was 1 in 125 in children 3-6 years and 1 in 85 in children 6-9 years of age, in rural areas it was 0.90%, 0.1% in tribal areas, 1.01% in urban areas, 0.61% in the coastal region and 0.6% in hilly regions. In India, a national NGO called 'Action For Autism', identified and announced that there are approximately 18 million people in India suffering from autism. The main impact of socio-economic and environmental risk factors on ASD development in India especially in the state of Tamil Nadu, South India is less reported. In addition, studies claimed in [6], the influence of environmental risk factors in autism is scarce from the Southern part of India. Autism Spectrum Disorder is a condition of development that requires attention. Today, machine learning is making life easier for people. Autism and other diseases require quick attention. This has led to the use of machine learning approaches in early autism prediction so that early treatment can be administered and damage can be reduced. This study focuses on the classification of ASD with help of machine learning and deep learning.

A. Contribution

1) Comprehensive literature survey in the area of ASD.
   - We have identified some studies on ASD prediction with the use of Machine learning and Deep learning.
   - Different approach for predicting autism 1. Video analysis [7], 2. Eye tracking [8], 3. MRI/fMRI [9], 4.EEG [10] and 5. Facial expression [11].
2) Implementation
   - As proposed research we work on video-based analysis which is used to predict autism.

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

- An extended work based on Zunino et al. [7].
- Implemented CNN Deep model InceptionV3 and Resnet-50 with 2-layer LSTM.

3) Comparative Analysis
   - Zunino et al. [7] have designed the experiments and collected the video-based data which contains around 455 videos of ASD and TD(Typically Development) children.
   - As compare to Zunino et al. [7] work, one layer LSTM in which have achieved around 77% accuracy the proposed approach (ResNet50 + 2 Layer LSTM 0.7828) gives better accuracy at 91.43% in training and 78.28% in testing.

B. Paper Organization

The content of this paper is organized as The first section contains a detailed overview of various research work carried out in the field of Autism Spectrum Disorder detection using machine learning techniques. The next section provides a brief overview of the proposed work and methodology. The next section provides details of the experiments carried out and the results obtained. The final section provides the conclusion. In this research work, it is proposed to perform Prediction of Autism Spectrum Disorder on a video database.

II. Related Work

Autism Spectrum Disorder is a neurological condition that must be identified early to minimize its effects. More advancements and progress in this medical discipline are being made possible by numerous developments in the field of machine learning. To identify Autism at an early stage, many researchers have used various machine learning methods. They are using image categorization of MRI data, EEG data, Video-based data, facial Expression, and others. Various deep learning-based approaches, including CNN, ANN, and RNN have been used to identify autism in children.

In the study by Raj et al. [12], the authors suggested using techniques such as CNN, Support Vector Machine (SVM), and Artificial neural network (ANN). The missing values were dealt with first in this effort. Later, the authors replaced the SVM with a CNN-based classifier that incorporates all of the SVM's features and characteristics. Here, the accuracy of prediction for the ASD Child dataset was approximately 98.30% for both the SVM and CNN-based models. To manage and optimize loss functions, they have used the Adam optimizer.

In the work of Another researcher, M. S. Mythili et al. [13] have used classification techniques to detect Autism. In Neural Network methods, Support Vector Machine and Fuzzy Logic were used to analyze Autism in students. The dataset utilized in this research work consists of various attributes such as Language data, social skill data, Behaviour data, etc.

In another work by Omar et al. [14] the authors have proposed effective prediction of autism using various machine learning techniques. Here, the authors have also developed a mobile application for predicting autism in people of any age. In this work, the authors have performed predictions based on the Tree-CART classifier. In the beginning, the tree root consists of the whole dataset. Later on, the dataset is split based on feature selection. They have also performed prediction using the Random Forest-CART method. They have claimed that the Random Forest Method performs better than another classifier. J.A. Kosmicki et al. [15] have proposed to search the least set of features to detect autism. They have used a machine learning-based approach to detect autism. They have employed 8 different machine learning algorithms.

In another work by Misman et al. [13], the authors have used Deep Neural network architecture for the appropriate classification. In this work, the authors have developed a deep neural network using Keras Sequential Model API [16]. Here, the authors have used two main hyperparameters in the neural network which are the number of layers and the number of nodes. They later compared this deep neural network model with other machine learning algorithms such as SVM. The authors have claimed that the accuracy of this model is higher and provides accurate results.

In the work [11] Facial expressions promote social interaction: a smile can express interest, and a frown can express sympathy. People with autism have difficulty showing the right facial expression at the right time, according to an analysis of studies. Instead, they may remain expressionless or evoke an inexplicable appearance. Now computational analysis of facial expressions is an emerging research topic that could overcome the limitations of human perception, In the work, they focused on qualitative facial expression recognition; the suggested framework aims to computationally analyze how children with ASD and TD produce facial expressions.

In the work of [10] The identification of children with ASD may benefit from a machine learning technique that combines EEG and eye-tracking data. The classification of autistic children against typically developing children was done using the minimum redundancy maximum relevance (MRMR) feature selection method combined with SVM classifiers.

Computer vision has also developed a lot in the last few years. Computer vision is helping in various medical research areas. It is now being used in the analysis of Autism Spectrum disorder to perform early diagnosis and treatment of this disease. Along with these machine learning-based methods, visual attention-based methods were also implemented to detect autism in kids.

For this purpose, many researchers also developed eye tracking-based to detect Autism. In the research work by Wan et al. [8], the authors have utilized the eye-tracking method to detect autism. Eye tracking method helps in the early detection of Autism. Detection of Autism in young children using eye tracking small informative video was used. In this research work, the authors have used 37 ASD and 37 Typically developing children of age 4 to 6 and created a dataset by making them watch a 10-second video of female speaking. Reductions in fixation time were observed in ASD children.

Another researcher utilized video analysis to detect autism. In the research work by Zunino et al. [7], the authors have created a video analysis dataset to detect autism. This work

by the researcher focuses on providing objective support to the doctor for the early assessment of diagnosis. Here, the researchers have created a specific experiment to record videos of the children. They recorded the patient performing simple gesture like grasping a bottle. They have processed the video data using recurrent deep neural networks. This resulted in proper discrimination between the two classes.

Another research work by Sherkatghanad et al. cites sherkatghanad2020automated proposed Autism detection using a Convolutional Neural Network. Here, the researchers have utilized the Brain imaging dataset to perform the detection process. They have used CNN on the dataset containing brain images. The resting-state functional magnetic resonance imagining (fMRI) data from the dataset named Autism Brain imaging Exchange (ABIDE) were used to detect Autism patients. Various patterns were utilized by authors to classify patients. They achieved an accuracy of 70.22 % using the ABIDE dataset.

In another similar work by Anibal Heinsfeld et al. [17], the authors identified autism spectrum disorder among patients using a large brain imaging dataset by using deep learning algorithms on them. They have also used the ABIDE dataset. The authors have performed a study on patterns of functional connectivity. These patterns are useful in identifying ASD participants using functional brain imaging data. The authors also tried to identify and study the neural patterns that emerged from the classification. The model developed by the authors achieved a 70% accuracy in the identification of ASD versus normal patients in the dataset. Anti-correlation of brain function between anterior and posterior areas of the brain was obtained from the patterns obtained through classification.

In the work by Liu et al. [18], the authors to perform early identification of autism in children have developed a computer vision-based system. They have considered reduced levels of response to languages as the base to detect autism in children. Here, the subjects are recorded while they play with the toys. The response of subjects to instructions is measured by vision-based techniques like hand tracking and object recognition. Based on this the subject is classified into the appropriate class.

From the above research studies, it is clear that both deep learning and computer vision-based methodologies as well as various traditional machine learning techniques are leading the way for an early autism diagnosis. In the field of medicine, this early detection is quite beneficial.

III. Proposed Work

In this section, we include an introduction to the proposed research work. We focused on a CNN-based approach for video-based classification with Resnet-50 and InceptionV3 deep learning network architecture. Previous works have utilized CNN and SVM-based models. In this particular research work, we tried working with CNN + LSTM-based model [7].

Our proposed architecture is shown in fig1. The first process that was performed to process on the video database was frame extraction. The frame extraction was performed using the FFmpeg library. after that, As shown in the below figure1, the frames will be passed through the convolutional model and the features will be extracted for each frame. Here, in this research work, InceptionV3 and Resnet50 were used to extract the frame features. These two CNN models are trained on ImageNet Database. InceptionV3 is much faster in the process of feature extraction but Resnet on the other hand is a very deep model. It has also introduced a Residual unit to solve the degradation problem. Feature of the frames are extracted from the average pooling layer of Resnet50 and InceptionV3 and were saved in .npy file.

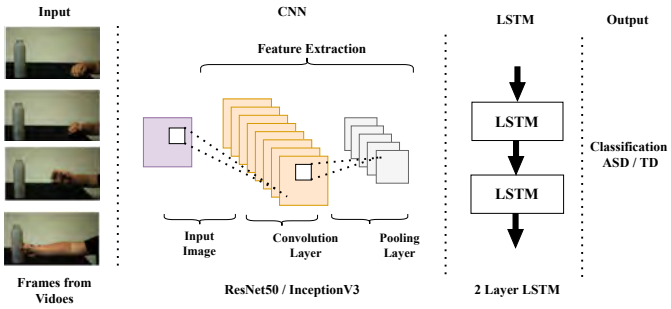


Fig. 1: Proposed Method

1) Data collection: Training and testing - The whole dataset has been split into two parts i.e., one part is training the dataset and the other one is testing the dataset with a ratio of 80:20 respectively. Here, the dataset contains around 1837 videos of 4 different actions which are 1. Placing, 2. Pouring, 3. Pass-to-place, and 4. Pass-to-pour. Each of these video datasets is divided into training and testing.
2) Feature Extraction: After dividing the data into training and testing, the features of each of the frames of the video are extracted using different CNN algorithms. There are various CNN algorithms such as InceptionV3, Resnet50, etc. which can be utilized for this task. These CNN models are trained on the ImageNet dataset. The frame extraction was performed using the ffmpeg library of python.

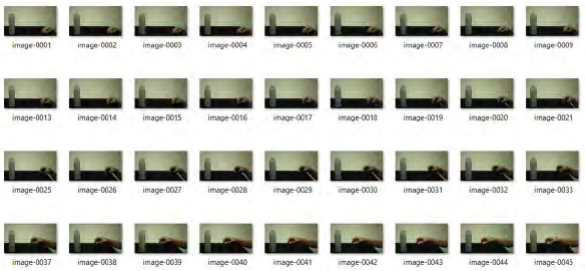


Fig. 2: Extracted Images

3) Training the features: The extracted features will later be trained using appropriate models. Initially, the features were passed through a normal MLP model. After Training the Resnet50 and InceptionV3 features for 5

epochs we get the result analysis. In this research work, a comparison has been drawn between the based model and the CNN-LSTM model.

IV. Methodology

We work on the Pre-trained CNN models Resnet-50 and InceptionV3.

A. Convolution Neural Network: Resnet-50

ResNet50 [19] is a deep residual network developed by Microsoft Research Asia and Beijing Jiaotong University. It was introduced in December 2015 in a paper titled "Deep Residual Learning for Image Recognition" and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2016. ResNet50 is a 50-layer convolutional neural network trained on more than a million images from the ImageNet database. The network is 50 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. The ResNet50 network is based on the ResNet-34 network, which was also introduced in the "Deep Residual Learning for Image Recognition" paper. ResNet-34 is a 34-layer convolutional neural network trained on the ImageNet database. The ResNet50 network was designed by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun [19]. Resnet-50 architecture is shown in Figure3.

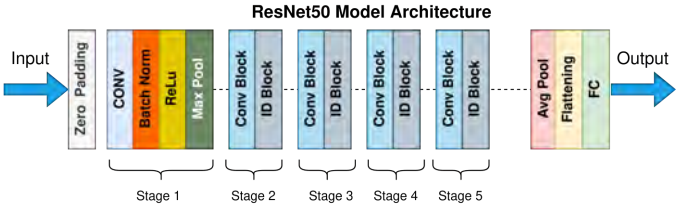


Fig. 3: Resnet-50 Architecture

B. Convolution Neural Network: InceptionV3

InceptionV3 was first proposed in the year 2015 by Szegedy, et. al [20]. It is used for image analysis and object detection. It has a total of 42 layers and a lower error rate. It is a widely used image recognition model that has been observed to achieve an accuracy rate of better than 78.1% on the ImageNet dataset. Convolutions layers with maximum pooling, layers with average pooling, dropouts, and fully connected layers are some of the symmetric and asymmetric building pieces that make up the Inception-v3 model. The inceptionV3 model is the result of numerous concepts that have been created by various researchers throughout the years [20]. InceptionV3 architecture is shown in Figure4.

C. Long Short-Term Memory

A recurrent neural network called Long Short Term Memory (LSTM) is built in such a way that it naturally retains information for a long period. We used 2 layer LSTM for Our experimental system contains CNN based system to extract the features of the data and an LSTM classifier. The extracted

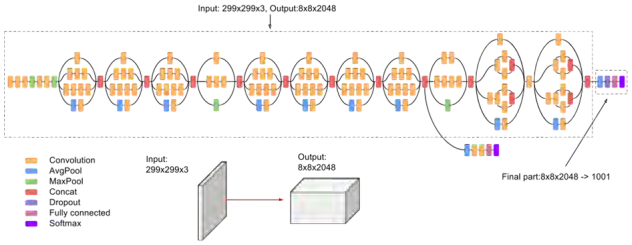


Fig. 4: InceptionV3 Architecture [21]

features are passed through the LSTM system to perform classification [22]. We extracted the image features using different CNN models such as ResNet50 and InceptionV3.

V. Experimental Analysis

The first process that was performed to process on the video database was frame extraction. The operating system we used was Windows 10. Tensorflow, Keras Numpy, Sklearn, and Matplotlib were implemented in the experiments. Frames were extracted from the video dataset before the experiment. All the frames were annotated in the appropriate folder names.

A. Dataset Description

In this research work, the video database available on this (https://pavis.iit.it/index.php/datasets/autism-spectrum-disorder-detection-dataset) Andrea Zunino et al. cites zunino2018video created a video gesture dataset. The authors created this dataset with the help of 20 children with ASD and 20 typically developed children. These children were brought to the Child Neuropsychiatry Unit of the IRCCS Giannina Gaslini Hospital and primary schools in Genova. The dataset was created by conducting the test of children individually in a quiet room. These children were seated on an adjustable chair. They were asked to rest their right elbow and wrist on the table. The children were asked to assume a position to perform the testing. Using that particular position, they were asked to reach toward the object and grasp it to place it in a box. Other actions involved pouring some water into a glass or passing of bottle to another person, who would either place the bottle in a box or pour some water. Based on this they were asked to perform movements and their movements were recorded. The final dataset contains 1837 video sequences. The dataset contains a total of 1837 video files. Children of two groups (ASD and Non-ASD) were asked to grasp the bottle in order To perform four different actions: 1). Placing 2). Poring 3). Passing to Place and 4). Passing to Pour.

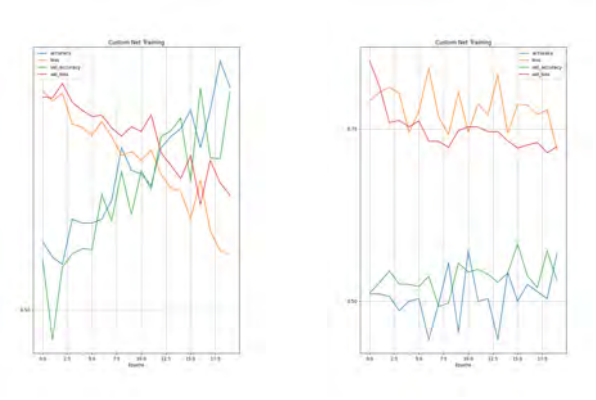
We have used Python Programming language.

- Various Python libraries will be used such as The open-CV library for the computer vision-based task.
- NumPy for array-based operations.
- TensorFlow and Keras libraries.
- KNN models can be used.
- CNN models such as InceptionV3 and Resnet50 are used to extract image features.

- Various other RNN models such as LSTM can be used to train those features

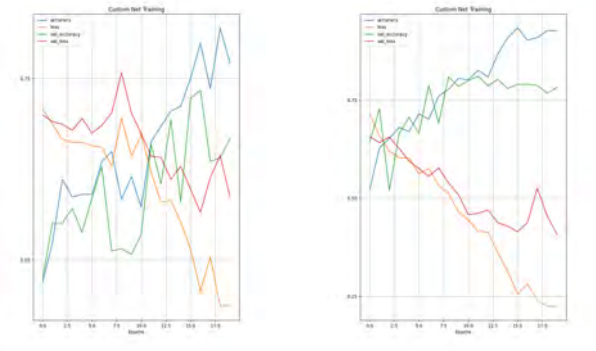
VI. Result and Discussion

In this research, we have compared the basic CNN+LSTM model with a CNN + 2-layer LSTM model. The other parameters of model training were kept the same. From the results, it is clear that the training and test accuracy has increased when a 2-layer LSTM model is used as a classifier for the extracted features. As shown in below tableIII.



(a) (b)

Fig. 5: Visualization of (a).InceptionV3 and (b).Resnet-50 in one Layer LSTM



(a) (b)

Fig. 6: Visualization of (a).InceptionV3 and (b).Resnet-50 in Two Layer LSTM

TABLE I: Training Phase Accuracy

Epochs	20
ResNet50+LSTM	0.6217
InceptionV3+LSTM	0.7128
ResNet50 + 2 Layer LSTM	0.9143

A deeper level Network can help extract higher-level semantic features and thus here are helping in increasing the training and testing accuracy. The CNN-based model can help extract complex features and the LSTM model helps in processing those deep features to perform classification.

TABLE II: Testing Phase Accuracy

Epochs	20
ResNet50+LSTM	0.5297
InceptionV3+LSTM	0.7008
ResNet50 + 2 Layer LSTM	0.7828

It can be observed that with single Layer LSTM, Training on InceptionV3 features gives better result. But, upon increasing the number of LSTM layers, the Training on Resnet50 features provides better result.

VII. Conclusion

Machine learning algorithms have made it easier to identify Autism in children, adults, and adolescents. Various approaches have been proposed to detect autism, including video gesture analysis, image classification, and eye movement analysis. The experiments for this study were run on a database of videos. When applying LSTM to train the features in the analysis of video datasets, it was found that the results were superior to those obtained with a standard MLP-based model. With Resnet50 extracted features, the Two-layer LSTM model delivered higher results. Early work in this area has shown that machine learning can be used to identify ASD-related patterns in behavior, as well as to predict future behaviors. These studies suggest that machine learning may be a valuable tool for understanding ASD and providing personalized interventions. In the Near Future model can be built to predict autism based on other video communication where symptoms like: Repeating certain behaviors like repeating words or phrases much time, Using Echo words, losing interest in talks, and even eye movement or blinking actions capture can help detect autism

References

[1] D. S. Hasin, C. P. O'brien, M. Auriacombe, G. Borges, K. Bucholz, A. Budney, W. M. Compton, T. Crowley, W. Ling, N. M. Petry, *et al.*, "Dsm-5 criteria for substance use disorders: recommendations and rationale," *American Journal of Psychiatry*, vol. 170, no. 8, pp. 834–851, 2013.

[2] C. Lord, M. Elsabbagh, G. Baird, and J. Veenstra-Vanderweele, "Autism spectrum disorder," *The lancet*, vol. 392, no. 10146, pp. 508–520, 2018.

[3] R. M. Joseph and H. Tager-Flusberg, "An investigation of attention and affect in children with autism and down syndrome," *Journal of autism and developmental disorders*, vol. 27, no. 4, pp. 385–396, 1997.

[4] A. Knopf, "Autism prevalence increases from 1 in 60 to 1 in 54: Cdc," *The Brown University Child and Adolescent Behavior Letter*, vol. 36, no. 6, pp. 4–4, 2020.

[5] M. Juneja and S. Sairam, "Autism spectrum disorder—an indian perspective," *Recent Advances in Autism*, 2018.

[6] B. Geetha, C. Sukumar, E. Dhivyadeepa, J. K. Reddy, and V. Balachandar, "Autism in india: a case–control study to understand the association between socio-economic and environmental risk factors," *Acta Neurologica Belgica*, vol. 119, no. 3, pp. 393–401, 2019.

[7] A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, and V. Murino, "Video gesture analysis for autism spectrum disorder detection," in *2018 24th international conference on pattern recognition (ICPR)*, pp. 3421–3426, IEEE, 2018.

[8] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, *et al.*, "Applying eye tracking to identify autism spectrum disorder in children," *Journal of autism and developmental disorders*, vol. 49, no. 1, pp. 209–215, 2019.

[9] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, and V. Salari, "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in neuroscience*, vol. 13, p. 1325, 2020.

[10] J. Kang, X. Han, J. Song, Z. Niu, and X. Li, "The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data," *Computers in biology and medicine*, vol. 120, p. 103722, 2020.

[11] M. Leo, P. Carcagnì, C. Distante, P. Spagnolo, P. L. Mazzeo, A. C. Rosato, S. Petrocchi, C. Pellegrino, A. Levante, F. De Lumè, *et al.*, "Computational assessment of facial expression production in asd children," *Sensors*, vol. 18, no. 11, p. 3993, 2018.

[12] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[13] M. F. Misman, A. A. Samah, F. A. Ezudin, H. A. Majid, Z. A. Shah, H. Hashim, and M. F. Harun, "Classification of adults with autism spectrum disorder using deep neural network," in *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pp. 29–34, IEEE, 2019.

[14] 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," in *2019 International conference on electrical, computer and communication engineering (ECCE)*, pp. 1–6, IEEE, 2019.

[15] J. Kosmicki, V. Sochat, M. Duda, and D. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," *Translational psychiatry*, vol. 5, no. 2, pp. e514–e514, 2015.

[16] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[17] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

[18] J. Liu, K. He, Z. Wang, and H. Liu, "A computer vision system to assist the early screening of autism spectrum disorder," in *International Conference on Cognitive Systems and Signal Processing*, pp. 27–38, Springer, 2018.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[22] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.