AUTISM SPECTRUM DISORDER DETECTION IN CHILDREN USING COMPUTER VISION STRATEGIES

[1]M. NIRMALA, [2]G. DHIVYA

[1]Assistant Professor, [2] II MCA Student
Department of Computer Applications
Hindusthan College of Engineering and Technology, Coimbatore, India,

Abstract- **Computer vision has been used to detect developmental abnormalities such as autism spectrum disorder (ASD).Computer vision used to pre diagnose various disorders. Facial analysis can be used to monitor vascular pulse, quantify discomfort, find facial paralysis, identify psychiatric problems, and use behavior imaging to distinguish people with ASD from those with normal development. We describe a technique for detecting an Autism's student level of engagement. Given the rise of distance learning in general and e-learning in particular, student participation is crucial and one of the most challenging issues for educators, academics, and policymakers. The main objective of the system is to monitor and determine the engagement of autism based student and their active participation during the E-learning session. We outline a method for assessing a student level of participation who has autism. It was created to work in real time and uses only information provided by the conventional built-in web-camera included in a laptop computer. We create a concentration index with three levels of engagement: "extremely engaged," "nominally involved," and "not engaged at all" by combining information on eye and head movements, as well as facial autism syndromes. The system was put to the test in a typical learning situation, and the findings demonstrate that it properly distinguishes "highly involved," "nominally engaged," and "not engaged at all" periods of time. Additionally, the results also show that the students with best scores also have higher concentration indexes.**

Keywords: Autism Spectrum Disorder, Behavior, Syndrome, student participation level, Computer Vision, Feature extraction.

1. INTRODUCTION

Humans have always had the intrinsic ability to discern between different faces and recognize them. Computers can now perform the same tasks. This offers a wide range of possibilities. Face detection and recognition can be used to improve access and security, similar to how the most recent Apple iPhone (see animation below) enables criminal identification, provides personalized healthcare, and offers other services. There are a tonne of materials online for face detection and identification, a subject that has been extensively studied. To identify the open source initiatives that are both straightforward to use and accurate, we tested a number of them. Additionally, we have developed a pipeline for autism syndrome identification, understanding, and recognition on any input image, syndrome of facial autism. The process of detecting human autism syndromes based on facial expressions is known as detection. This API can be used to create interactive video chat applications or to keep track of autistic symptoms linked to visual information published on social media or photo-sharing websites.

Facial autism disorders play a crucial role in human communication by assisting us in deciphering others' intentions. Typically, people use their facial expressions and voice to infer the emotional states of others, such as happiness, sadness, and rage, that are associated with autism syndrome. Diverse studies [1] and [2] have determined that verbal and nonverbal components make up one-third of human communication. In addition to other nonverbal elements, facial expressions are one of the primary information carriers in interpersonal communication, as they have special significance for individuals on the autism spectrum. Consequently, it is not surprising that research into facial autism has grown in popularity over the past few decades, with applications not only in perceptual and cognitive sciences but also in affective computing and computer animations. In our research and analysis of facial autism syndrome recognition, we take into account a classroom scenario in which we assess the students' moods before a lecture and assess these moods or autism syndromes to better understand the psychology of the students and what interests and bores them during a lecture. Teaching assessments in the classroom of today are straightforward, ungraded, and limited in their capacity to offer helpful input on kids' autism disorders in the classroom.

This research is being conducted because students' commitment and engagement during lectures are essential for comprehending the concepts being taught and can unquestionably improve their academic credentials. Although teachers can observe their students firsthand in the classroom, this cannot be used to measure their attentiveness. Furthermore, even in a classroom with constant supervision, a few students may become disinterested in the subject and lose their concentration. The creation of a technique to impartially assess and pinpoint pupils' attention and attentiveness lapses in the classroom is therefore required. Observing facial autism syndromes during class can therefore help predict students' general classroom behavior, such as their level of concentration (neutral), giggling during the lecture (happy), dozing off during the lecture (bored or melancholy), etc. The purpose of this study was to build on earlier research that identified mood trends in a classroom setting by observing students' facial expressions during lectures. This technique for analyzing the various forms of autism in a classroom setting can give students a better understanding of their autism syndrome states during lectures, which can aid in the development of teaching aids that improve students' concentration and the instructor's ability to convey information effectively.

We describe our unique approach that uses deep convolutional neural networks (CNN) to identify kids' classroom autism symptoms after analyzing prior research on the topic. The processing unit extracts the key frames from the video and employs facial feature extraction techniques to detect the faces in the video frame. The key frames from the live video of the students in the classroom as a whole are captured by a video camera in the classroom, and the captured video frames are presented to the processing unit. The most likely autism syndrome can be identified using the CNN model, which was trained on comparable databases of people and faces and can predict autism spectrum classes with high accuracy. In order to provide better teaching resources, improve lecture material, and improve the classroom environment, autism syndrome predictions may be used to conduct an evaluation. Autism syndrome classes such as happy, sad, surprise, disgust, furious, fear, and neutral autism syndromes are anticipated and assessed.

1) Facial Detection - Ability to detect the location of face in any input image or frame. The output is the bounding box coordinates of the detected faces
2) Facial Recognition - Compare multiple faces together to identify which faces belong to the same person. This is done by comparing face embedding vectors
3) Autism syndrome Detection - Classifying the autism syndrome on the face as happy, angry, sad, neutral, surprise, disgust or fear.

2. LITERATURE SURVEY / RELATED WORK

C.S.Paula et al.'s 3 pilot study provides preliminary data on the incidence of Pervasive Developmental Disorder (PDD) in South America. The prevalence of PDD was 27.2/10,000 (95% CI: 17.6–36.8) and some hypotheses were raised to explain this low frequency. Clinical findings were consistent with previous findings, with a preponderance of males, a greater proportion of children diagnosed with PDD-NOS than autistic disorder, and the majority being delivered to older mothers.

Carette et al, 4 ASD is a neurodevelopmental disorder with numerous subtypes, making diagnosis difficult. Eye-tracking systems enable the recording of precise eye focus on a screen. This paper focuses on automatic detection of ASD using eye-tracked data and an original Machine Learning approach.

E. Fombonne at al, 5 reviews 43 studies published since 1966 that estimated the prevalence of PDDs, including autistic disorder, Asperger disorder, PDD not otherwise specified, and childhood disintegrative disorder. The prevalence of autistic disorder has increased, while the prevalence of PDD not otherwise specified has decreased.

G. Yolks, I. Oztel, S. Kazan et al,6 Facial expressions are an important part of communication and can be used to diagnose and monitor neurological disorders. This paper presents a novel deep learning approach for automatic facial expression recognition, segmenting the facial components known to be important.

M. I. U. Haque and D. Valles 7 discusses the initial work of a research to teach young autistic children recognizing human facial expression using computer vision and image processing. The Kaggle's FER2013 dataset was used to train and experiment with a deep convolutional neural network model, which was modified with pictures of four different lighting conditions.

O. Rudovic, Y. Utsumi, J. Lee et al. 8 investigates the performance of deep learning models in the task of automated engagement estimation from face images of children with autism. They use video data of 30 children with different cultural backgrounds recorded during a single session of a robot-assisted autism therapy. They perform a thorough evaluation of the proposed deep architectures for the target.

M. S. Satu, F. Farida Sathi, M. S. Arifen, M. Hanif Ali, and M. A. Moni 9 conducted a study to explore the features of normal and autism in divisional regions in Bangladesh. They collected individual samples of children from their parents between 16-30 months using Autism Barta apps and preprocessed the data to categorize frequent features based on their individual regions.

Q. Guillon, N. Hadjikhani, S. Baduel, and B. Rogé 10 reviewed visual social attention in autism spectrum disorders (ASD) from infancy to adulthood. Results show that social orienting is not qualitatively impaired and decreased attention to faces does not generalize across contexts. This suggests that social orienting is not qualitatively impaired and decreased attention to faces does not generalize across contexts.

T. Akter, M. S. Satu, L. Barua, F. F. Sathi, and M. H. Ali 11 conducted an experiment to extract features of the fusiform facial area (FFA) from MRI images of normal and autistic brains. The images were preprocessed and extracted features to calculate area, perimeter, and eccentricity for further analysis. This experiment is one of the most widely studied aspects to detect autism.

S. Schelinski, K. Borowiak, and K. von Kriegstein 12 used two independent functional magnetic resonance imaging studies to investigate voice processing in a population with no voice-sensitive regions, autism spectrum disorder (ASD). The results showed that voice identity recognition impairments were not caused by dysfunction of these voice-sensitive regions. This suggests that dysfunction of these voice-sensitive regions may explain voice identity recognition impairments.

3. EXISTING SYSTEM

Deep learning approaches are currently having great success across several industries, including computer vision. In fact, a model for convolutional neural networks (CNN) can be trained to examine photos and recognize faces with autism syndrome. In this project, we develop a system that can identify students' facial features associated with autistic disorders. Three parts make up our system: face detection using Haar Cascades, normalization, and autism syndrome recognition using CNN on the FER 2013 database with seven different expression kinds. The obtained results demonstrate that face autism syndrome recognition is practical in education, and as a result, it can assist instructors in adapting their presentation to the autistic syndromes of their students. In this section, we describe our proposed system for analysing the facial expressions of students, which is based on the CNN architecture. The questionnaire method cannot be able to predict the real autism issues and its complex to design.

4. PROPOSED SYSTEM

In recent years, facial expression detection has become increasingly important in the field of human-computer interaction, and deep learning technology has made this possible. In this study, we develop a simple network model for the classification of the autism spectrum in real-time based on student facial expressions. Pre-activation in the residual block is employed to optimise the model and mitigate the effects of overfitting. It also uses Haar cascade for face identification and a lightweight CNN network model. Experimental results on the FER2013 database demonstrate that our model outperforms other cutting-edge expression classification techniques. In addition, our model requires fewer parameters, making network training simpler. When autism syndrome is observed in real time. By Open CV method, we can able to discover a real data from live input and Designing complexity is low comparing to the dataset method.

5. METHODOLOGY

Libraries Used

Matplotlib is a Python 2D plotting toolkit that generates publication-quality visuals in a variety of physical formats and cross-platform interactive settings. It can be used by four graphical user interface toolkits, the Python and IPython shells, the Jupyter notebook, web application servers, and Python scripts. Numpy is the Python programming language's Numpy module that supports large, multi-dimensional arrays and matrices, as well as a variety of high-level mathematical operations. Pandas is a free and open source scientific environment written in Python by and for scientists, engineers, and data analysts. Spyder is a free and open source scientific environment written in Python by and for scientists, engineers, and data analysts. The proposed EFF-TSC method framework is divided into three main parts: preprocessing, segmentation, and classification.

Keras:

Keras is an open-source neural network library written in Python. It can be used in conjunction with Theano, Microsoft Cognitive Toolkit, or Tensor Flow. Its key design goals are user-friendliness, modularity, and extensibility, with the purpose of facilitating speedy experimentation with deep neural networks. It was developed as part of the research project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), and its major inventor and maintainer is Francois Chollet, a Google engineer. It provides a higher-level, more intelligible collection of abstractions that enable constructing deep learning models easier, regardless of the computational backend used. Microsoft added a CNTK backend to Keras with CNTK v2.0. This project component will introduce the proposed system, which uses a deeply separable convolutional architecture to assess pupils' facial expressions. The Haar cascade classifier is used to identify faces in the input image before cropping and normalising the obtained face information to a size of 4848. The facial images are then fed into CNN. The end result is a list of all potential facial expressions, such as pleased, sad, angry, afraid, astonished, disgusted, and normal.

For this project, create a new model framework. CNN is primarily concerned with three factors: memory usage, recognition accuracy, and speed training. The majority of the network is made up of seven convolutional layers, a maximum pooling layer, and a global average pool. Modelling and idea generation for algorithms.

ALGORITHM USED

Convolutional Neural Networks:

Deep learning algorithms make use of convolutional neural networks (CNN). This network may contain convolutional layers, pooling layers, ReLU layers, layers with full connectivity, and loss layers. Each convolutional layer in a typical CNN architecture is followed by a Rectified Linear Unit (ReLU) layer, a Pooling layer, one or more convolutional layers, and one or more fully connected layers. Instead of pooling and completely connected layers, similar convolutional layers are proposed. This may lead to inter-feature dependencies and an increase in the number of parameters, but it can be mitigated by employing smaller convolutional layers within the network, which also functions as a form of regularization. Using a large collection of labelled images, CNNs are taught to recognize the patterns and characteristics associated with particular objects or classes. CNNs can be trained to classify new images and used to extract features for tasks such as object detection and image segmentation.

CNNs have attained state-of-the-art performance on a variety of image recognition tasks, including object categorization, object identification, and image segmentation. They are extensively used in computer vision, image processing, and other related fields, and have been implemented in a variety of applications including security systems, autonomous vehicles, and medical imaging.

CNNs, or deep learning neural networks, are neural networks designed for processing structured data arrays, such as images.

- The design components in the input image, such as lines, gradients, circles, or even eyes and faces, are picked up incredibly well by CNN.
- Convolutional neural networks are so reliable for computer vision because of this property.
- CNN requires no preprocessing and can be applied directly to an underdone image.
- A convolutional neural network is a feed forward neural network with up to 20 neurons.
- The strength of convolutional neural networks is due to the convolutional layer, a type of layer.
- CNN consists of many convolutional layers piled on top of one another, each capable of recognizing increasingly complex structures.
- Handwritten digits can be detected with three or four convolutional layers, however human faces require 25 layers.

LAYERS OF CNN

Convolutional layers:

A convolutional layer consists of clusters of neurons called kernels. Despite their small size, the profundity of the kernels is always the same as the input. In the case of high-dimensional inputs such as images, where connecting all neurons to all antecedent outputs is extremely wasteful, neurons from a kernel are connected to a limited portion of the input known as the receptive field. The convolution procedure gives the convolutional layer its name. In mathematics, convolution is an operation performed on two functions that results in a third function that is a modified (convoluted) version of one of the original functions.

Pooling strata: On the one hand, aggregating layers are utilised to reduce the processing and spatial dimensions of the network's representation. Pooling layers are also used to manage overfitting. The most common pooling layer filters are 2 x 2 with a 2 pixel stride. Consequently, the input is virtually halved.Pooling layers reduces the extent of the feature maps. As a consequence, both the number of parameters to be learned and the amount of computation performed by the network are decreased. The pooling layer summarizes the features in a region of the convolution layer-generated feature map. Consequently, succeeding operations are performed on summarised features as opposed to precisely positioned features generated by the convolution layer. Consequently, the model is more resistant to variations in the position of the input image's features.

Completely linked layers:

Layers in a typical neural network are interconnected. Each output from a layer with complete connectivity is linked to each neuron in that layer. Convolutional layers are created using the same techniques as fully connected layers. Consequently, conversion between them is feasible. Data augmentation is a technique for generating novel and unusual "data." This provides two benefits: first, it permits the synthesis of "more data" from sparse data, and second, it precludes overfitting.

A lack of training samples could lead to overfitting. One method is to create fictional data and add it to the training set. We randomly divided the original dataset into two parts: training and testing. The training and test sets each employ half of the original dataset. We have a large dataset, but less data is normally utilised for training and more data is used for testing.

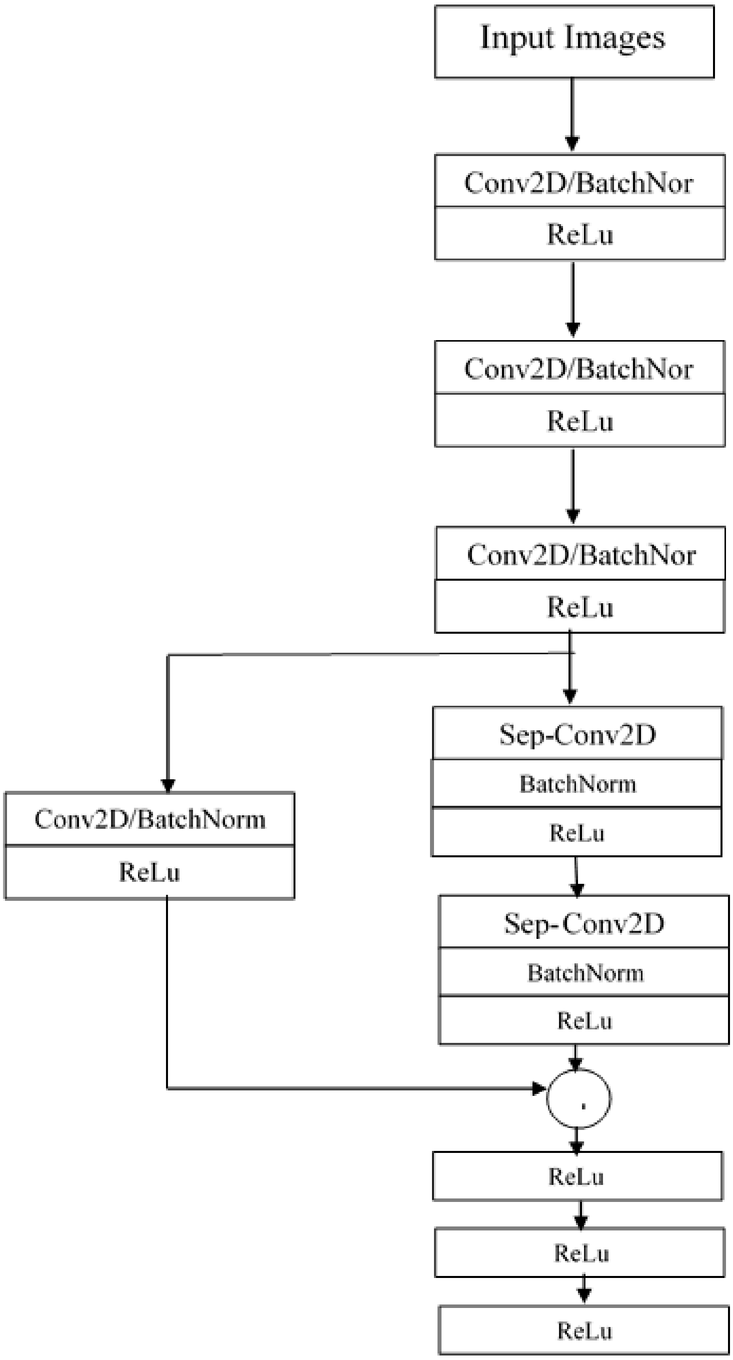

Figure 1 Keras

The first step in the input layer is a preprocessed grayscale image of a face with a size of 4848. Following that, 8 convolution kernels of size 33 are selected, and convolution processing is done to the input image to generate a local expression feature map. Use the ReLU function for nonlinear activation. Second, when convolving the upper layer to output the feature map, two convolution kernels with varied widths of 33 and 55 are used to obtain a more complete extraction of face expression images at various scales. Third, this network employs four residual depth separable convolutions to reduce the cost and amount of convolutional parameters.

Figure 1 depicts the distinction between depth separable and standard convolution. Depth separable convolution is made up of two phases: depth wise convolution and point wise convolution. Instead of convolving all M channels in the first stage of a standard CNN, the depth operation only applies convolution to one channel at a time. During the point wise convolution procedure, an 11 convolution operation is performed on M channels. As a result, the convolution kernel size for this operation is increased. The output size is determined by the number of convolution kernels used.

The number of multiplications in the depth separable convolution for the same input picture is equal to the total number of multiplications in the two stages. The multiplication and total number of training parameters are reduced by comparing the calculation amount of the depth separable convolution to the calculation amount of the standard convolutionis depicted in Figure 2.

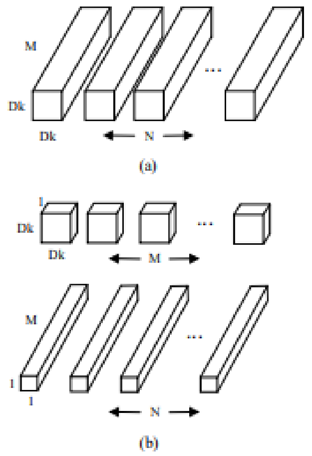

Figure 2 Standard Convolution & Depth Separable Convolution

Fourth, when the gradient signal from the error function propagates back to an earlier layer in deep neural networks, it diminishes exponentially, which is known as the "vanishing gradients" problem. The network in this project employs the gradient signal in the residual module to accomplish the output of one layer as input to the other layer of the design via "skip connections" or "shortcut paths," which are triggered by adjustment. The activation function is placed before the final output, which means that after obtaining the input, the value is first normalized, then activated, activated multiplied by the weight layer, and finally output. Make the network's parameters easier to understand and is depicted in Figure 3.

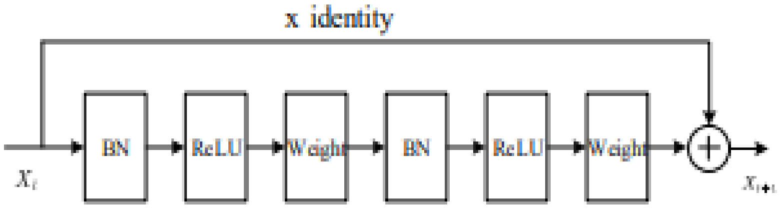

Figure 3 Pre-active residual block

Project Pipeline

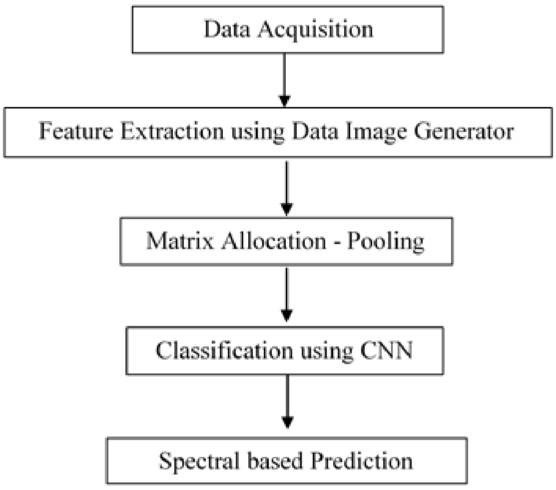

Figure 4 Project Pipeline

The process of sampling signals that measures real-world physical phenomena and converting them into a digital form that can be manipulated by a computer and software. This feature can be used to detect features in imagery. Each layer can extract one or more unique feature in an image and that can be even displayed through mapping. It is a key-step in convolutional based systems that reduces the dimensionality of the feature maps. The objective is to reducing a value or an image dimensionality and allowing for assumptions to be done. A kind of network architecture for deep learning and is specially used for image recognition and tasks that involve the processing of pixel data. An exploratory data analysis technique that reduces complex multidimensional datasets into clusters of similar data in fewer dimensions.

ARCHITECTURE DIAGRAM

One of the most significant modelling tools is the flow diagram. The system's component models are created using it. These elements include the system's operation, the data it uses, a third party that engages with it, and the way information moves through it.

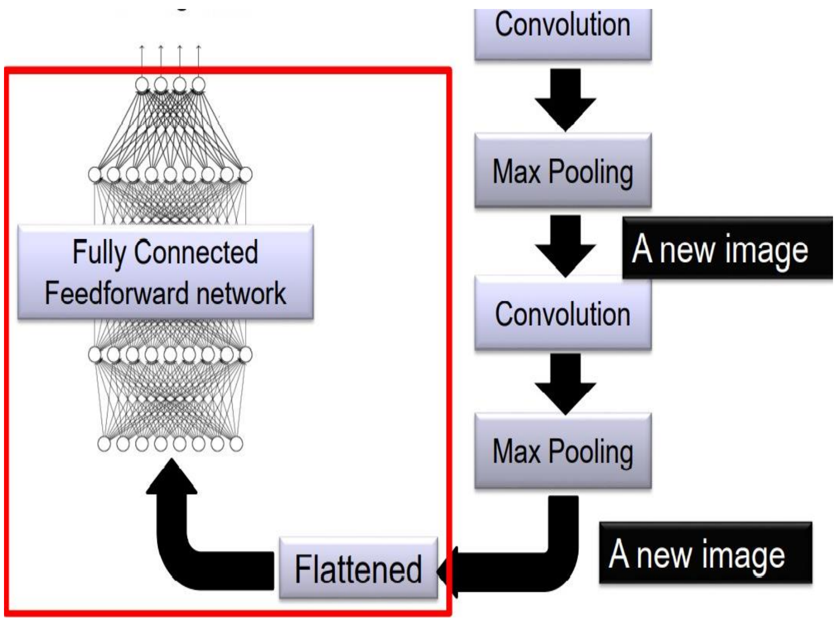

Figure 5 Flow Diagram

SYSTEM ARCHITECTURE

The architecture of a system reflects how the system is used and how it interacts with other systems and the outside world. It describes the interconnection of all the system's components and the data link between them. The architecture of a system reflects the way it is thought about in terms of its structure, functions, and relationships. In architecture, the term "system" usually refers to the architecture of the software itself, rather than the physical structure of the buildings or machinery. The architecture of a system reflects the way it is used, and therefore changes as the system is used.

Convolutional layers are stacked on top of Max Pooling layers to build a CNN. To avoid being too oversized, we also offer Dropout. We next add a Soft Max layer that is immediately followed by a fully linked (Dense) layer. The basic CNN model is built using the most popular hyper-parameters. Six classes are initially all that are used for the model's training and evaluation because we need to iteratively change the hyperparameters and, if necessary, the architecture of the model. The model is trained iteratively while the following hyper-parameters are adjusted, selecting the best 6 classes based on the quantity of training data available. Then, there are 12, 15, and soon, twenty lessons per day.

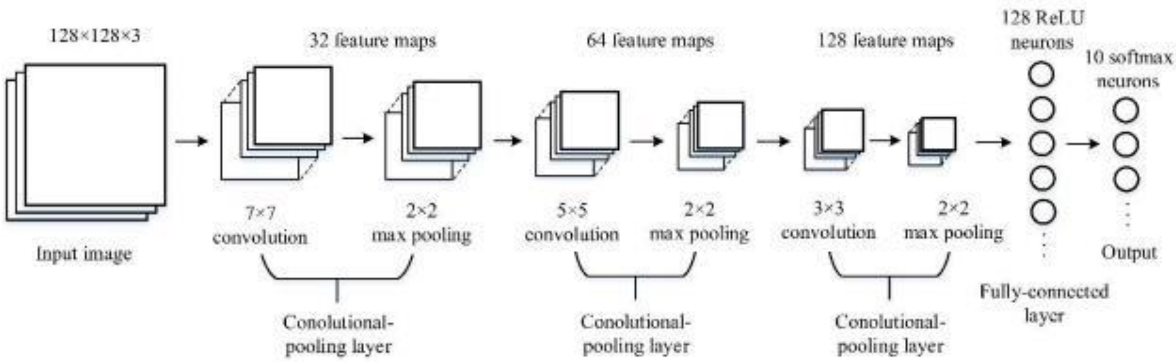

Figure 6 SYSTEM ARCHITECTURE DIAGRAM

The model was initially built using just 3 convolutional layers, 2 pool layers, and 2 thick layers. When there were only 6 classes, adding more convolutional layers did not significantly increase the accuracy. As anticipated, accuracy decreased as more classes were added. Therefore, a few extra layers are added to the model to achieve the highest level of accuracy. The optimal performance was ultimately achieved by using 4 convolutional layers, 4 pooling layers, 1 normalization layer, 2 dropout layers, and 2 dense layers. An epoch shows how many times the machine learning algorithm has cycled through the full training dataset. Here, 5 epochs have been employed.

Max Pooling:

The input feature map is compressed in order to primarily introduce the pooling layer. On the one hand, the network's computational complexity is made simpler by the shrinking feature map. On the other hand, the important features can be extracted with the help of feature compression. In this stage, we set the maximum pool size to two and the stride to three. The input region of the image was finally normalized using the Norm layer. It can improve the learning pace of the model while also lowering the risk of overfitting. It improves the standardization of the extracted features.

EXPERIMENTED RESULTS

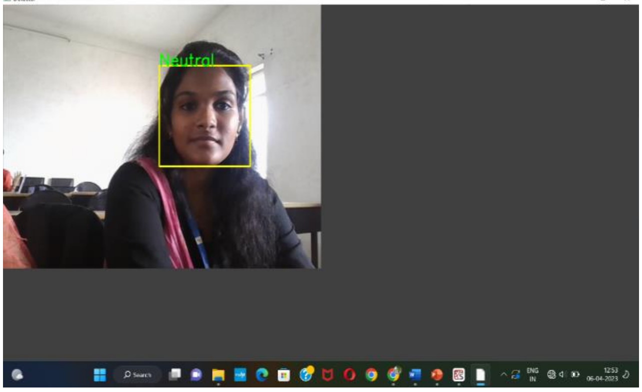

Figure 7 Display Neutral Child

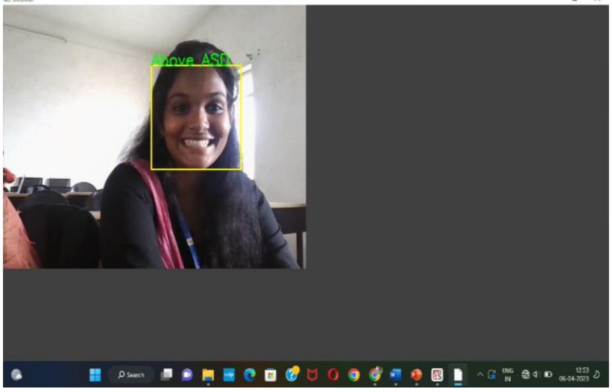

Figure 8 Display Above ASD Child

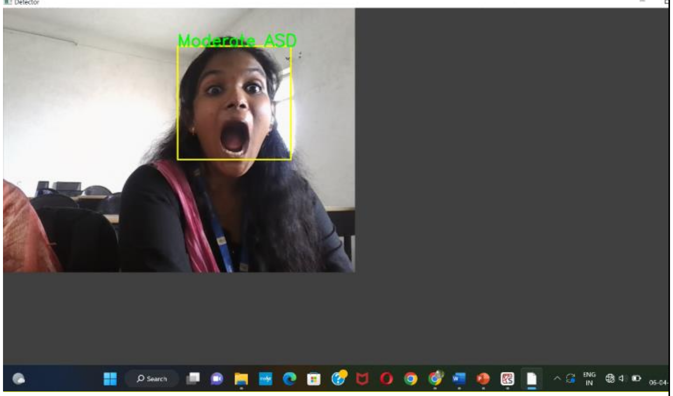

Figure 9 Display Moderate ASD Child

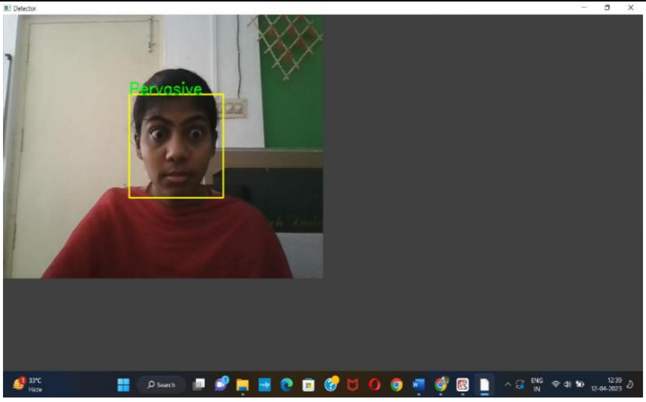

Figure 10 Display Pervasive Child

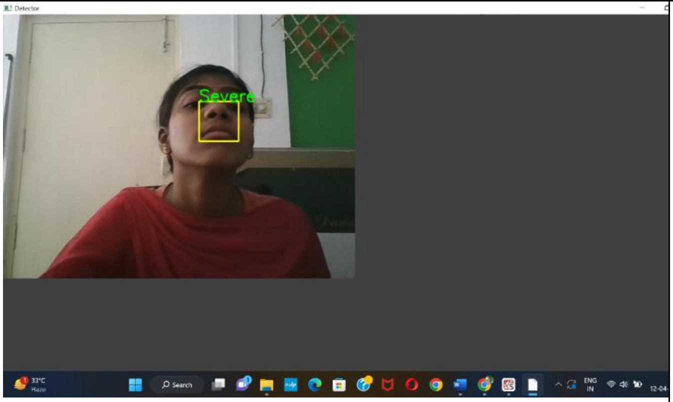

Figure 11 Display Severe Child

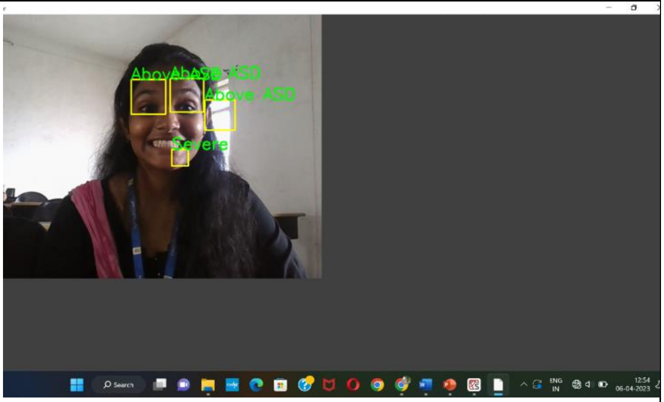

Figure 12 Display Separation of Child

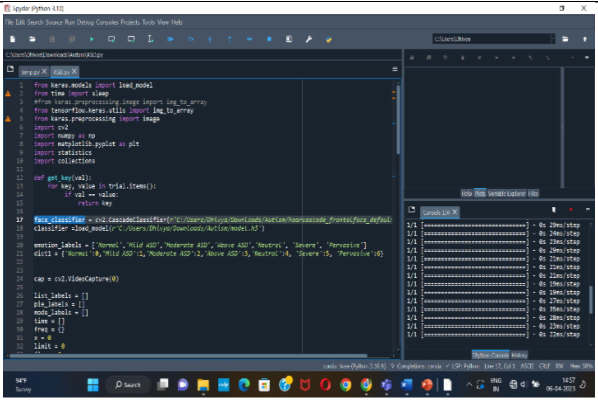

Figure 13 Displaying Plots Output

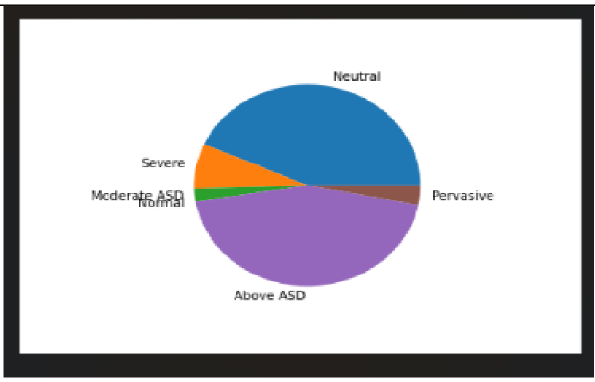

Figure 14 View Pie Chart

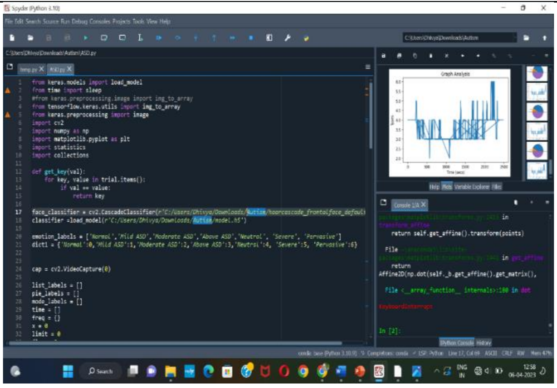

Figure 15 Display Graph Analysis

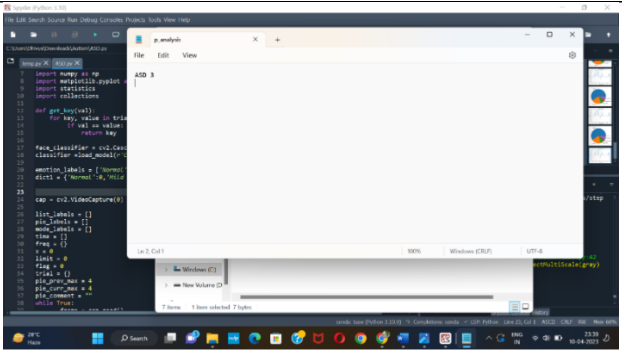

Figure 16 Display Result Analysis

Figure 7 represents the Neutral Level. Repetitive play or verbal expressiveness (echolalia). A fixation on certain activities, ideas, or concepts. A reluctance to engage in new experiences or to disrupt routines. Aversion to certain forms of interaction, especially hugging or cuddling.
Figure 8 represents the Above ASD Level of Autism child. On the most severe end of the spectrum is Level 3 which requires very substantial support. Signs associated with both Level 1 and Level 2 are still present but are far more severe and accompanied by other complications as well.
Figure 9 represents the Moderate Level of Autism child. Children with moderate autism may or may not interact with peers. They generally struggle to make eye contact, interpret body language and emotions, and understand figures of speech, and they may simply walk away from conversations that don't involve their favorite topics or interests.
Figure 10 represents the Pervasive Level of Autism child. They are hitting, biting, scratching, pinching, pulling hair, spitting in people's faces, slapping, punching, kicking and generally using physical force. This term is also used if their child is biting their own hand, head banging or slapping their own head, and other self-injurious activities.
Figure 11 represents the Severe Level of Autism child. ASD level 3 is characterized by severe challenges in social communication as well as extremely inflexible behavior. Children with level 3 autism will be nonverbal or have the use of only a few words of intelligible speech. Initiation of social interaction is very limited, as well as response to others.
In this report, we primarily concentrate on two scientific advances in detecting ASD issues:
1) First-person perspective analysis of ASD 2) video-based ASD behavior recognition.

6. CONCLUSION

The review of the framework for facial expression recognition has been highlighted in this research. It has been demonstrated that the framework has good applicability in actual activities and helps to address issues like students' lack of accountability and teachers' inability to receive timely feedback. In the end, it will help raise the standard of online education. Additionally, due to the enormous number of participants in the online courses, it is impossible to guarantee that everyone maintains a high level of focus, and as a result, the expressions of the students may not accurately reflect their autistic symptoms. Setting thresholds can be used to filter out some irrelevant data and highlight the key autistic symptoms in an image.

7. FUTURE ENHANCEMENT

Following are potential suggestions for our suggested model:
1. In each class, we only utilized a few photographs. To create a more reliable model, more data must be used.
2. When the entire dataset of all patients is fed into the model, the time complexity of the model should be reduced.

3. In this study, the effects of two characteristics (sex and average age) must be taken into account.
4. With balanced data, the performance might increase.

REFERENCES:

1. Mehrabian A. Communication without words. Psychol. Today. 1968;2:53–56. [Google Scholar]
2. Kaulard K., Cunningham D.W., Bülthoff H.H., Wallraven C. The MPI facial expression database—A validated database of emotional and conversational facial expressions. PLoS ONE. 2012;7:e32321. doi: 10.1371/journal.pone.0032321.
3. C. S. Paula, S. H. Ribeiro, E. Fombonne, and M. T. Mercadante, “Brief report: prevalence of pervasive developmental disorder in Brazil: a pilot study,” Journal of Autism and Developmental Disorders, vol. 41, no. 12, pp. 1738–1742, 2011.
4. R. Carette, F. Cilia, G. Dequen, J. Bosche, J.-L. Guerin, and L. Vandromme, “Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach,” in Proceedings of the International Conference on IoT Technologies for Healthcare, pp. 75–81. 4, Springer, Angers, France, 24–25 October 2017.
5. E. Fombonne, “Epidemiology of pervasive developmental disorders,” Pediatric Research, vol. 65, no. 6, pp. 591–598, 2009.
6. G. Yolcu, I. Oztel, S. Kazan et al., “Deep learning-based facial expression recognition for monitoring neurological disorders,” in Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1652–1657, Kansas City, MO, USA, 13–16 November 2017.
7. M. I. U. Haque and D. Valles, “A facial expression recognition approach using DCNN for autistic children to identify emotions,” in Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 546–551, Vancouver, Canada, 1–3 November 2018.
8. O. Rudovic, Y. Utsumi, J. Lee et al., “CultureNet: a deep learning approach for engagement intensity estimation from face images of children with autism,” in Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 339–346, Madrid, Spain, 1–5 October 2018, ISSN 2153-0866.
9. M. S. Satu, F. Farida Sathi, M. S. Arifen, M. Hanif Ali, and M. A. Moni, “Early detection of autism by extracting features: a case study in Bangladesh,” in Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 400–405, Dhaka, Bangladesh, 10–12 January 2019.
10. Q. Guillon, N. Hadjikhani, S. Baduel, and B. Rogé, “Visual social attention in autism spectrum disorder: insights from eye tracking studies,” Neuroscience & Biobehavioral Reviews, vol. 42, pp. 279–297, 2014.
11. T. Akter, M. S. Satu, L. Barua, F. F. Sathi, and M. H. Ali, “Statistical analysis of the activation area of fusiform gyrus of human brain to explore autism,” International Journal of Computer Science and Information Security, vol. 15, pp. 331–337, 2017.
12. S. Schelinski, K. Borowiak, and K. von Kriegstein, “Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition,” Social Cognitive and Affective Neuroscience, vol. 11, no. 11, pp. 1812–1822, 2016.
13. X. Jiang and Y.-F. Chen, “Facial image processing,” in Applied Pattern Recognition, H. Bunke, A. Kandel, and M. Last, Eds., pp. 29–48, Springer, Berlin/Heidelberg, Germany, 2008.
14. M. Duda, R. Ma, N. Haber, and D. P. Wall, “Use of machine learning for behavioral distinction of autism and ADHD,” Translational Psychiatry, vol. 6, no. 2, Article ID e732, 2016.
15. G. Deshpande, L. E. Libero, K. R. Sreenivasan, H. D. Deshpande, and R. K. Kana, “Identification of neural connectivity signatures of autism using machine learning,” Frontiers in Human Neuroscience, vol. 7, p. 670, 2013.
16. M. N. Parikh, H. Li, and L. He, “Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data,” Frontiers in Computational Neuroscience, 2019.
17. F. Thabtah and D. Peebles, “A new machine learning model based on induction of rules for autism detection,” Health Informatics Journal, vol. 26, no. 1, pp. 264–286, 2020.
18. M. H. Al Banna, T. Ghosh, K. A. Taher, M. S. Kaiser, and M. Mahmud, “A monitoring system for patients of autism spectrum disorder using artificial intelligence,” in Proceedings of the International Conference on Brain Informatics, pp. 251–262, Springer, Padua, Italy, July 2020.
19. F. B. Pokorny, B. W. Schuller, P. B. Marschik et al., “Earlier identification of children with autism spectrum disorder: an automatic vocalisation-based approach,” in Proceedings of the Annual Conference of the International Sp, August 2017.
20. F. Eyben, K. R. Scherer, B. W. Schuller et al., “The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing,” IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190–202, 2016
21. A. Jack, “Neuroimaging in neurodevelopmental disorders: focus on resting-state fMRI analysis of intrinsic functional brain connectivity,” Current Opinion in Neurology, vol. 31, no. 2, pp. 140–148, 2018.
22. C. H. Y. Fu and S. G. Costafreda, “Neuroimaging-based biomarkers in psychiatry: clinical opportunities of a paradigm shift,” Canadian Journal of Psychiatry, vol. 58, no. 9, pp. 499–508, 2013.