UnLocking Emotion Recognition in ASD Children: Analyzing Facial Expressions

Anu Rose Joy
Department of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kerala
anurosejoy@amaljyothi.ac.in

Anuja Sara Mathew
Department of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kerala
anujasaramathew2002@gmail.com

Christeena Antony
Department of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kerala
christeenaantony436@gmail.com

Christeen Mareia Paul
Department of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kerala
christeenmarieapaul2024@cs.ajce.in

Dona Mariyam John
Department of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kerala
donamariyamjohn2024@cs.ajce.in

Abstract—**Children with Autism Spectrum Disorder (ASD) often face challenges in understanding and expressing emotions, which can hinder their social interactions and emotional de- velopment. This paper presents a study on using technology for emotion recognition in children with Autism Spectrum Disorder (ASD). We collect facial expression data from ASD children during different emotional states and apply machine learning to accurately identify emotions. Our results show high accuracy, indicating the potential for non-invasive, technology- based emotional support and intervention for ASD children. This research emphasizes the importance of multidimensional data in understanding emotions in ASD and offers promising prospects for personalized interventions.**

Keywords—**Autism Spectrum Disorder (ASD), Emotion Recognition, Technology-based Intervention**

I. INTRODUCTION

A. General Background

According to a 2021 study published in Indian Journal of Pediatrics, it estimates that ,1 in 68 children in India are thought to be affected by autism. The prevalence of autism in boys is higher than in girls, with a male-to-female ratio of roughly 3:1. Analyzing the emotions of autistic children poses challenges due to difficulties in interpreting their non-verbal cues, limited emotional vocabulary.Every child’s performance varies,thus generalization is very difficult.Thus it is important to analyze and interpret their emotions for their proper care and assistance.

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects communication and social interaction. One of the profound challenges faced by individuals with ASD is the ability to understand and express emotions effectively. These difficulties often hinder their social interactions and emotional development, making it essential to explore innovative solutions to support their emotional well-being and communication skills. ASD encompasses a spectrum of symptoms and severity levels, ranging from mild to severe, with each individual presenting unique challenges and strengths. Despite its prevalence and impact on individuals and their families, the exact causes of ASD remain largely unknown, though genetic, environmental, and neurological factors are believed to play significant roles.

B. Objective

The overarching goal of ASD research is to improve our understanding of the disorder and develop evidence-based interventions that improve the well-being and functioning of people with ASD. This includes improving early detection

DOI: 10.5281/zenodo.12548997

and diagnosis, offering tailored interventions to address core symptoms and associated challenges, fostering social inclusion and acceptance, and encouraging independence and autonomy. Furthermore, research seeks to empower people with ASD by identifying and leveraging their unique strengths and abilities. Collaboration among researchers, clinicians, educators, policy-makers, and community stakeholders is critical for furthering knowledge, putting research findings into practice, and advocating for more support and resources for people with ASD and their families. Finally, the goal is to create a more inclusiveand supportive society where people with ASD can drive and reach their full potential.

C. Scope

The scope of ASD research is broad, encompassing several domains such as psychology, neuroscience, education, and technology. Within this context, research frequently focuses on understanding the underlying mechanisms of ASD, identifying effective interventions and support strategies, and improving the quality of life for people with ASD and their families. Cognitive processes, social cognition, sensory sensitivities, communication skills, adaptive behaviors, and co-occurring conditions like anxiety and depression are all possible areas of investigation.

LITERATURE SURVEY

Fatma M Talaat et al. [1] proposed a real-time emotion identification system for autistic children that employs a Deep Convolutional Neural Network (DCNN) architecture for facial expression recognition. This paper makes use of a collection of cleaned images of autistic children expressing various emotions. Duplicated images and stock images have been removed. The study used 758 images for training (67 for anger, 30 for fear, 350 for joy, 48 for natural, 200 for sadness, and 63 for surprise) and 72 for testing (Anger, Fear, Joy, Natural, Sadness, Surprise).

DONG-HWAN LEE and JANG-HEE YOO [2] proposed a novel approach based on the divide-and-conquer CNN learning strategy. A new FER strategy involves detecting and standardizing facial areas. Optimizing ResNet-18.Use a confusion matrix to group similar expressions. Retraining and testing with grouped expressions.The goal of this approach is to improve FER accuracy by reducing intra-class distance while increasing inter-class distance. The experiments used four datasets: Tufts, RWTH, RAF, and FER2013, which each contained images with all facial expressions labeled in advance.

Ninad Mehendale [3] proposed sb approach to accurately detect and recognize facial emotions or expressions. The proposed method uses a two-level CNN framework for facial emotion recognition.In the first part of CNN the input image fed to the first-part CNN contains not only the face but also shapes, edges, textures, and objects in the background. To start, filters like edge detectors, circle detectors, and corner detectors are used to identify facial features. Once the face is detected, the second-part CNN focuses on capturing specific facial features like eyes, ears, lips, nose, and cheeks. This layer employs edge detection filters to recognize these features.Initially, used the Cohn-Kanade expression dataset, which had limited data and resulted in a maximum accuracy of 45 percentage. Divided your dataset into 70 percentage training images and 30 percentage testing images for evaluation. Performed 25 iterations, each time using a different set of 70 percentage training data.

Shuiyang Mao, P. C. Ching,Tan Lee [4]proposed Mel-frequency cepstral coefficients (MFCCs) and their derivatives and they are employed as features for emotion recognition in the methodology. The authors begin by labeling each segment with the corresponding utterance’s emotion. After labeling, a deep self-learning (DSL) method is used to correct the noisy labels that were produced. To increase the robustnessof the emotion segment model, the DSL framework uses a sequence of deep neural networks (DNNs) to gradually update the labels for network re-training. The authors also look into the application of soft labeling, which uses an emotion class distribution to represent each segment label and characterizes the underlying mixture of emotions.

Oreeti Khajuria [5]proposed the use of VGG-16 in conjunction with a transfer learning approach to achieve the best possible accuracy in facial emotion recognition (FER). Pre-processing the data, feature extraction, and facial emotion classification are some of the steps in the process. The dataset utilized in this study was gathered from the 35,887 samples available in the Kaggle repository; 28,821 of these samples were used for training, and 7066 were used for validation. The suggested model is assessed using two datasets that are accessible to the public and contrasted with other cutting-edge techniques. The accuracy of the suggested model, according to the results, is 91 percent, which is about 20 percent better than that of conventional machine learning models.

Morched Derbal,Mu’tasem Jarrah [6] propsed a novel approach to identifying whether a child between the ages of 3 and 10 has autism spectrum disorder is presented in this study. A child can become fully immersed in an intense and immersive environment through video games, causing them to display a wide range of emotions and facial expressions. The suggested approach makes use of deep learning algorithms to identify possible ASD cases by analysing gamers’ facial expressions in video games.A convolutional neural network (CNN) is used in the suggested method to extract facial features from the preprocessed data. Following feature extraction, the data is divided into two groups: typically developing and ASD. For classification, a support vector machine (SVM) classifier is employed.A number of metrics, including accuracy, sensitivity, and specificity, were used to assess the suggested approach. The suggested method's outcomes were contrasted with those of other machine learning-based techniques and conventional diagnosis techniques.

Hari Prasad Mal, Dr. Swarnalatha P [7] proposed method explains the system that is described in the paper, which obtains emotional sequences of different images by running a surveillance video through a model of a facial expression detector. Then, in order to account for the large rigid head motion, the sequence is represented as a single Emotional Avatar Image using the SIFT flow technique. The tensor perceptual colour framework achieves the highest recognition rate and performance when the method is evaluated based on recognition rate.

Taner Danisman, Marius Bilasco and Nacim Ihaddadene[8] suggested suggested method to extracts grayscale images from a given video and use a hybrid approach to identify facial feature points using artificial neural networks and image processing techniques (quantization, adaptive thresholding,and contrast en- hancement). Using a Viola-Jones face detector that has been tweaked, the system auto- matically finds faces in a given video. From grayscale photos, the system recogniz seven spots on the face: the corners of the mouth, the nose, the pupils, and the eyebrows.

METHODOLOGY

Images and video frames are loaded and preprocessed using the 'preprocess image' function.The VGG16 model is loaded with pre- trained weights from the ImageNet dataset. The top layer (classification layer) is excluded ('include top=False'), and the input shape is specified.Features for the training data are extracted using the pre-trained VGG16 model. The features are flattened and stored in 'X train features'. A simple neural network classifier is built usingthe Sequential API. It consists of a dense layer with ReLU activation and an output layer with softmax activation.Theequation for softmax activation function is illustrated below:

$$SOFTMAX(Z)_i = e^{zi} + \sum_{J=1}^{N} e^{zj}$$

Here, z is the vector of raw outputs from the neural network. The value of e 2.718. The i-th entry in the softmax output vector softmax(z) can be thought of as the predicted probability of the test input belonging to class i. The last four layers of the pre-trained VGG-16 model areset to be trainable for fine-tuning. This allows the model to adapt to the specific dataset. The model is compiled withthe Adam optimizer, sparse categorical crossentropy loss, and accuracy as the metric.The classifier is trained with fine-tuning, learning rate scheduling.The model is evaluatedon the test set.

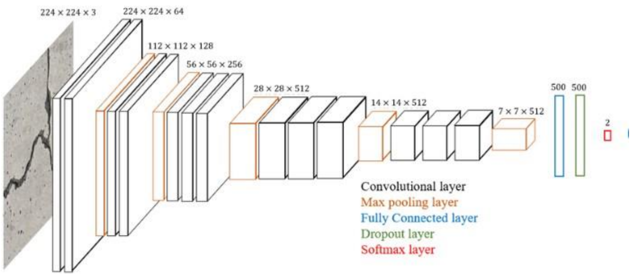

Fig. 1. Architecture of VGG16

A. DATA ACQUISITION

The two main features included in our project are Emotion Recognition using Facial Expressions from images and Videos.The six main facial expression includes natural,surprise,fear,anger,happy,sadness. Dataset was obtained from Kaggle. Used 758 images for training (Anger: 67, Fear: 30, Joy: 350, Natural: 48, Sadness: 200, Surprise: 63) and 72 images for testing(Anger: 3, Fear: 3, Joy: 42, Natural: 7, Sadness:14, Surprise: 6).20 percent from the training images is taken for validation.

(i) Input Images

Organizes the image file paths and corresponding labels for both the training and testing sets and writes this information to CSV files for later use in training a machine learning model.

D. DATA PREPROCESSING

The script starts by loading data from CSV files ('train data.csv' and 'test data.csv'). These files contain file paths and corresponding labels.Extracted file paths and labels from the loaded CSV files.A function is defined to preprocess images.It loads an image using the 'image.load img' and 'def predict emotions from video' function from TensorFlow ,converts the image to a NumPy array using 'image.img to array',adds a batch dimension using 'np.expand dims'and Normalizes pixel values to the range [0,1] by dividing by 255.0.

E. FEATURE EXTRACTION

The script loads the pre-trained VGG16 model from Keras applications with weights pre-trained on ImageNet. These 'include top=False' argument excludes the classification layers .Iterates through training image paths. For each image, preprocesses it using the 'preprocess image' function, Extracts features using the pre-trained VGG16 model ('base model'),Flattens the output features to create a 1D feature vector and Stores the flattened features in 'X train features'

F. BUILDING AND TRAINING THE CLASSIFIER

Uses ‘LabelEncoder()‘ to encode categorical labels (‘train labels‘) into numerical format (‘y train‘).Splits the data into training and validation sets using ‘train test split‘.A simple sequential model is defined with a dense hidden layer(256 units, ReLU activation) and an output layer with soft- max activation. The last layers of the base model (VGG16)are frozen (‘layer.trainable = False‘) to fine-tune only theadded classifier. The model is compiled using the Adam optimizer, sparse categorical cross-entropy loss, and accuracy as the metric.Defines a learning rate schedule and early stopping callbacks .Trains the classifier with fine-tuning, learning rate scheduling, and early stopping.

G. MODEL EVALUATION

The trained classifier is saved to a specified path. Iterates through test image paths. Preprocesses images and extracts features using the pre-trained VGG16 model.Flattens the features and stores them in ‘X test features‘.Evaluates the classi- fier on the test set using ‘classifier.evaluate‘. Prints the test ac-curacy and Plots the training accuracy and validation accuracyover epochs.

The trained classifier is saved to a specified path. Iterates through test image paths. Preprocesses images and extracts features using the pre-trained VGG16 model.Flattens the features and stores them in ‘X test features‘.Evaluates the classifier on the test set using ‘classifier.evaluate‘. Prints the test accuracy and Plots the training accuracy and validation accuracy over epochs.

The trained classifier is saved to a specified path. Iterates through test image paths. Preprocesses images and extracts features using the pre-trained VGG16 model.Flattens the features and stores them in ‘X test features‘.Evaluates the classifier on the test set using ‘classifier.evaluate‘. Prints the test accuracy and Plots the training accuracy and validation accuracy over epochs.

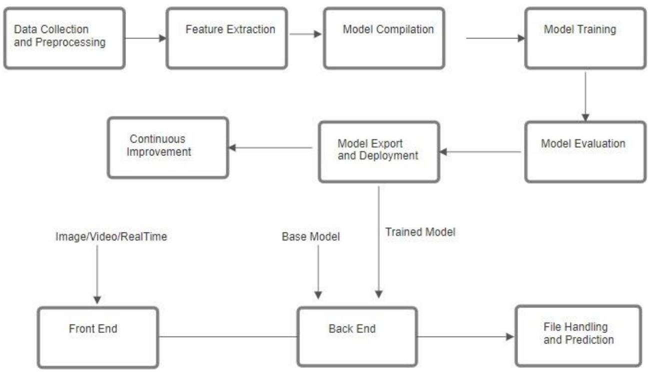

Fig. 2. Architecture Diagram of Proposed Model

TABLE I
VGG16 MODEL SUMMARY

Layer (type)	Output Shape	Param #
input 2 (InputLayer)	[(None, 224, 224, 3)]	0
block1 conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1 conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1 pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2 conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2 conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2 pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3 conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3 conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3 conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3 pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4 conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4 conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4 conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4 pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5 conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5 conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5 conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5 pool (MaxPooling2D)	(None, 7, 7, 512)	0
Total params: 14714688 (56.13 MB)		
Trainable params: 14714688 (56.13 MB)		
Non-trainable params:0(0.00 Byte)		

I. CONTINUOUS IMPROVEMENT

Continuous improvement for the Emotion Recognition sys- tem involves ongoing efforts to enhance its performance, efficiency, accuracy, and usability over time. This can be achievedthrough strategies such as data augmentation to increase dataset diversity, fine-tuning model hyperparameters, exploring different pre-trained models, adopting ensemble methods, establishing a feedback loop for user input, implementing automated testing and monitoring, and maintaining version control and experiment tracking. By continuously iterating and refining the system, it can adapt to changing requirements and improve its effectiveness in recognizing emotions from images and videos.

II. RESULTS AND DISCUSSIONS

A. DATASET

The Kaggle platform is used to collect data for a dataset on Face Emotion Recognition. This paper uses clean images of autistic children expressing various emotions. Duplicated images and stock images have been removed. The dataset wascategorized into six facial emotions: anger, fear, joy, natural, sadness, and surprise. Figure 3 depicts the six most commonly used emotions. This study utilized 758 images for training (Anger: 67, Fear: 30, Joy: 350, Natural: 48, Sadness: 200, Surprise: 63) and 72 images for testing (Anger: 3, Fear: 3, Joy: 42, Natural: 7, Sadness: 14, Surprise: 6). Validation involves using 20% of the training images.

Fig. 3. The six primary used emotions

. *EVALUATION OF PROPOSED MODEL*

The performance of the VGG with the used data without optimization is shown in Fig. 4.

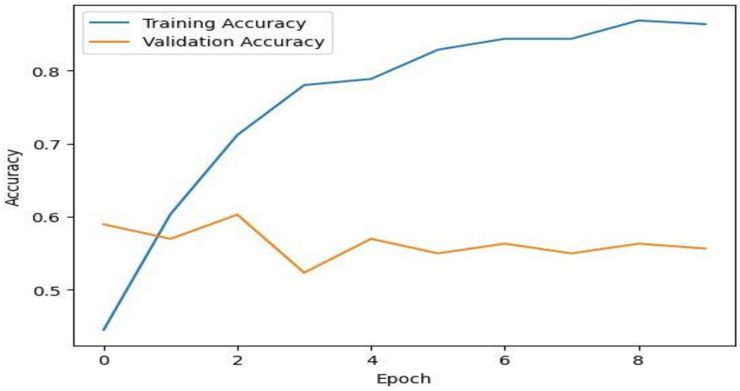

Fig. 4. The Result of Training Dataset

As shown in figure 4, we can see a blue line,representing training accuracy,it indicates how well the model predicts using the training data that is presently being trained on.The orange line,which representing the validation accuracy,shows how accurate the model is on different validation dataset.

The training and validation accuracy of the proposed model is shown which achieves approx. 86.33 percent and 55.63 percent respectively.

In this, we discussed about the accuracy comparison result between it and related studies. The ratio of the number of accurate predictions to the total number of predictions is called accuracy.

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

The test accuracy is a measure of the model's performance on this test set and is usually calculated as the proportion of correctly classified instances in the test set.The training accuracy of our modal is 70 percent.

III. CONCLUSION

In this research, EMOSYNC presents a comprehensive approach to detecting emotions from images and videos using machine learning techniques. Through this endeavor, we have successfully developed a system capable of recognizing six primary emotions: natural, surprise, fear, anger, happy, and sadness. Leveraging a dataset obtained from Kaggle, consisting of 758 training images and 72 testing images, we employeda variety of methodologies to preprocess the data, extract features using a pre-trained VGG16 model, build and train a classifier, and evaluate its performance. Our project embodies the principles of continuous improvement, as we have implemented various strategies to enhance the system's accuracy, efficiency, and usability over time.This includes exploring techniques such as data augmentation, hyperparameter tuning, transfer learning with different pre-trained models, ensemble methods, and establishing feedback loops for user input. Moving forward,we envision further advancements in the Emotion Recognition system by expanding the dataset to encompass a broader range of emotions and demographics,and integrating feedback mechanisms to adapt to evolving user needs. By embracing continuous improvement and pushingthe boundaries of what is possible in emotion recognitiontechnology, our project paves the way for transformative advancements in artificial intelligence and human-computer interaction.

REFERENCES

[1] Talaat, F.M. Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Comput &Applic 35, 12717–12728 (2023). https://doi.org/10.1007/s00521-023-08372-9

[2] D. -H. Lee and J. -H. Yoo, "CNN Learning Strategy for Recognizing Facial Expressions," in IEEE Access, vol. 11, pp. 70865-70872, 2023, doi: 10.1109/ACCESS.2023.3294099.

[3] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020). https://doi.org/10.1007/s42452-020-2234-1

[4] S. Mao, P. C. Ching and T. Lee, "Enhancing Segment-Based Speech Emotion Recognition by Iterative Self-Learning," in IEEE/ACM Trans- actions on Audio, Speech, and Language Processing, vol. 30, pp. 123- 134, 2022, doi: 10.1109/TASLP.2021.3133195.

[5] O. Khajuria, R. Kumar and M. Gupta, "Facial Emotion Recognition using CNN and VGG-16," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 472-477, doi: 10.1109/ICICT57646.2023.10133972.

[6] Morched Derbali, Mutasem Jarrah and Princy Randhawa, "Autism Spectrum Disorder Detection: Video Games based Facial Expres- sion Diagnosis using Deep Learning" International Journal of Ad- vanced Computer Science and Applications(IJACSA), 14(1), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140112

[7] H. P. Mal and P. Swarnalatha, "Facial expression detection using facial expression model," 2017 International Conference on Energy, Commu- nication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 1259-1262, doi: 10.1109/ICECDS.2017.8389644.

[8] Taner Danisman, Ioan Marius Bilasco, Nacim Ihaddadene, Chaabane Djeraba. Automatic Facial Feature Detection for Facial Expression Recognition. Fifth International Conference on Computer Vision Theory and Applications (VISAPP) 2010, May 2010, Angers, France. pp.407- 412. ffhal-00812308f