REWARD-DRIVEN EMOTION DETECTION IN AUTISM SPECTRUM DISORDER WITH ATTENTION MECHANISM

REPORT

IT5712 - PROJECT I

Submitted by

Rahul Prasanth D (2020506070)

Bala Natesh R M (2020506018)

Sanjay G (2020506080)

Guided by

Dr. J.Dhalia Sweetlin

Associate Professor

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

MADRAS INSTITUTE OF TECHNOLOGY

ANNA UNIVERSITY: CHENNAI 600 044

NOV 2023

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects millions of children worldwide. One of the significant challenges faced by autistic children is the difficulty in expressing and understanding emotions, which can affect the effective communication and social interactions. Traditional methods of emotion recognition often fall short in addressing the unique needs and characteristics of these people.

Therefore, this project aims to create a specialized system that comprehensively identifies and understands emotions in autistic children, aiming to bridge the communication gap and provide essential insights into their emotional experiences which also helps the educators, therapists, and caregivers to take care of them.

II. PROBLEM STATEMENT:

The project addresses the challenge of understanding and effectively supporting emotional expression in autistic children. Autistic children often encounter difficulties in conveying their emotions. Our primary objectives include the development of a system that comprehensively identifies and understands the emotions exhibited by autistic children. This system will consider various modalities, including facial expressions, hand gestures, and other body movements, to provide an elaborate view of their emotional states. Additionally, the landmark points are detected, combined with the analysis of body movements, to enhance the accuracy of emotion prediction. It will help to understand how autistic children experience emotions by using information about where and when emotions happen on their faces, along with specific facial expressions, to develop ways to support and help them more effectively.

III. LITERATURE SURVEY

Facial Action Unit Detection via Adaptive Attention and Relation

Shao et al. (IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 32, 2023)

The authors propose an adaptive attention regression network by integrating the advantages of local attention predefinition and global attention learning, which can capture both predefined dependencies by landmarks in strongly correlated regions and facial globally distributed dependencies in weakly correlated regions. To simultaneously reason the specific pattern of each AU, the inter-dependencies among AUs, as well as the temporal correlations, they also propose an adaptive spatio-temporal graph convolutional network. Extensive experiments on benchmark datasets show that the approach achieves comparable performance in both constrained scenarios and unconstrained scenarios, and can accurately learn the regional correlation distribution of each AU. The Adaptive Attention Regression (AAR) method achieved an average F1-frame score of approximately 63.8 on the BP4D benchmark. The AAR network was tested on input images with misalignment errors and occlusions. If input images are severely misaligned AAR fails to precisely capture AU Region Of Interests (ROIs). The AAR does not explicitly process misalignment errors, such as explicitly learning rotation-invariant and scale-invariant features.

Action unit classification for facial expression recognition using active learning and SVM

Yao et al. (Multimed Tools Appl 80, 24287–24301 (2021))

This paper proposes a combination of active learning and SVM for the extraction of AUs and facial expression classification. Active learning uses the existing model to acquire new knowledge by simulating the process of human learning.

Based on continuously accumulated information, the existing model can be corrected to become more accurate. SVM was utilized to classify different AUs and ultimately map them to their corresponding facial expressions. Different facial expressions, regardless of being female or male, had different recognition rates ranging from 90% to 95% for females and from 83% to 95% for males. Of the seven facial expressions, five expressions (joy, sadness, anger, hate, and neutral) were recognized correctly. Regardless of gender in the samples, the hate and neutral facial expressions seem to be more difficult to recognize than the joy and surprise expressions.

Are 3D Face Shapes Expressive Enough for Recognising Continuous Emotions and Action Unit Intensities?

Kumar et al. (EEE TRANSACTIONS ON AFFECTIVE COMPUTING DOI 2023)

AU intensity estimation and dimensional emotion recognition models are trained based on the temporal dynamics of 3D facial expressions. The quality of 3D Morphable Models (3DMM) based expression features on the datasets of valence-arousal estimation and AU intensity estimation are extensively evaluated. A simple bi-directional Gated Recurrent Unit (GRU) network is applied to model the temporal dynamics of 3DMM expression features extracted from five dense 3D face alignment models: ExpNet, 3DDFA-V2, RingNet, DECA and EMOCA. The recognition performance of different 3D face shape models with the 2D face appearance baselines and models are compared. In the case of continuous emotion recognition 3D face expression features outperform the existing benchmarks as well as the 2D appearance baselines. On the task of AU intensity prediction 3D face shape models perform poorly compared to the existing state-of-the-art benchmarks based on 2D appearance features. The MSE values for 2D and 3D shapes were found to be 0.36 and 0.53 respectively. The

poor AU intensity estimation performance of the 3D face models might be due to the use of a global basis vector for expression modelling.

Multitask, Multilabel, and Multidomain Learning With Convolutional Networks for Emotion Recognition

Pons et al. (IEEE TRANSACTIONS ON CYBERNETICS, VOL. 52, NO. 6, JUNE 2022)

A datasetwise selective sigmoid cross-entropy loss function is formalized to simultaneously train a multitask, multilabel and multidomain model. The authors utilize two popular convolutional neural network (CNN) architectures, VGG-16 and Resnet-50, for their experiments. These networks serve as the basis for training models for emotion recognition and AU detection. Dedicated individual networks are trained for each task and dataset separately. The soft-max cross-entropy loss function is used for emotion recognition tasks, while the sigmoid cross-entropy function is employed for AU detection due to the multilabel nature of the latter task.

The accuracy was found to be 5 3% for Single RestNet-50 and 8 2% for SJMT RestNet-50. This method addresses one of the challenges with discrete emotion recognition in the wild, which is the lack of large public labelled datasets.

Exploring Complexity of Facial Dynamics in Autism Spectrum Disorder

Krishnappa Babu et al. (IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 14, NO. 2, APRIL-JUNE 2023)

This work focuses on analyzing the complexity of spontaneous facial dynamics of toddlers with and without ASD. Toddlers watched developmentally-appropriate and engaging movies presented on a smart tablet. Simultaneously, the frontal camera of the tablet was used to record the toddlers' faces, providing the opportunity for the automatic analysis via CV. The facial landmarks' dynamics

of the toddlers with ASD versus TD were studied specifically. An iPad-based application (app) was designed that displayed strategically designed, developmentally appropriate short movies involving social and non-social components. The device's front-facing camera was used to record the children's behavior and capture ASD-related features. The children's facial dynamics were exploited from the eyebrows and mouth regions using multiscale entropy (MSE) analysis to study the complexity of such facial landmarks'
dynamics. Distinctive landmarks' dynamics were captured in children with ASD, characterized by a significantly higher level of complexity in both the eyebrows and mouth regions when compared to typically-developing children. In Cross-Validation using Decision Tree model the accuracy for each video shown is found to be Video1=77.5%, Video2=7 3%, Video3=73.8%, Video4=67.2%. The study sample
has a limited number of ASD participants and did not have sufficient power to determine the impact of demographic characteristics on the results. Other measures of complexity might be more robust than the MSE in their ability to discriminate children with and without ASD.

Real-time facial emotion recognition system among children with autism based on deep learning and IoT

Talaat (F.M. Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Computing & Applications 35, 12717–12728 (2023))

This system proposes an enhanced deep learning (EDL) technique to classify the emotions using convolutional neural network. The proposed emotion detection framework takes the benefit from using fog and IoT to reduce the latency for real-time detection with fast response and to be a location awareness. The architecture outperforms earlier convolutional neural network-based algorithms and does not

require any hand-crafted feature extraction. A total of six emotions are detected by the propound system: anger, fear, joy, natural, sadness, and surprise. The training accuracy was found to be 0.963 and validation accuracy 0.88. The limitation of the proposed technique is that it uses small dataset (limited scale) as the large number of real dataset is not available.

Facial Action Unit Detection With Transformers

Jacob et al. (2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR))

The paper outlines an innovative approach for the detection and analysis of facial expressions, with a specific focus on Facial Action Units (FAUs) associated with muscle activations. It highlights the significance of facial expressions as a primary means of conveying nonverbal information, noting that while some expressions are universally understood, others are individualized, necessitating the use of the Facial Action Coding System (FACS). FAU detection is framed as a multi-label binary classification problem, with some approaches considering the degree of FAU activation. The model architecture incorporates attention-based techniques, including separate attention maps for each action unit, and leverages multi-task learning to exploit task relationships. The model's framework involves feature extraction, attention learning, and multi-task modules, with novel loss functions for feature discrimination and multi-label classification. The approach is capable of end-to-end training, achieving state-of-the-art performance on public datasets and undergoing comprehensive evaluation, including ablative studies to assess design choices.

The model shows the best average F1-score of 61.5. The major challenge with the dataset used here is the severity of the class imbalance and the variation in the head pose and expression.

Discriminative Few Shot Learning of Facial Dynamics in Interview Videos for Autism Trait Classification

Zhang et al. (IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 14, NO. 2, APRIL-JUNE 2023)

This paper attempts to fill this gap by developing a novel discriminative few shot learning method to analyze hour-long video data and exploring the fusion of facial dynamics for the trait classification of ASD. The model first extracts the spatio-temporal features of the video and uses the combination of K-SVD with MFA to get more discriminative representations. A few-shot learning module is designed to further improve classification performance It achieves the best performance with an accuracy of 91.72% by fusing the seven selected scenes that are comparable to the standardized diagnostic scales. This Experiment adopts a scene-level feature fusion strategy, which requires manually splitting entire hour-long videos into 15 separate scenes by time markers and extracting facial-dynamics features of each scene.

An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children with Autism Spectrum Disorder

Nie et al. (IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 29, 2021)

The text discusses the development of a computer-mediated system, referred to as the C3I system, designed to enhance the Interactional Joint Attention (IJA) skills of young children with Autism Spectrum Disorder (ASD). The C3I system aims to fill this gap by involving caregivers in the training process. A long video clip is shown to distract the child's attention away from the caregiver. When the child is sufficiently distracted, the caregiver presses a button on a tablet to pause the video. At this point, the expectation is that the child will look back at the caregiver and initiate joint attention. If the child does not respond as expected, the caregiver can press another button to alert the system. In response, the system

displays a non-social audio-visual cue, such as a bouncing ball with sound effects, to guide the child's attention either toward the caregiver or one of the target monitors. The system's real-time tracking and response to the child's behavior are central to its operation and its goal of promoting Interactional Joint Attention (IJA) skills in children with ASD.

The feasibility study had only one session per dyad with repetitive trials. In addition, the sample size was small and there was no control group. As such, the results of this feasibility study need to be considered with caution until a larger study verifies its generalizability.

IV. NOVELTY

The project adopts a multimodal approach to emotion recognition in autistic children, combining facial expressions and body gestures for a more comprehensive understanding of their emotions. Additionally, the implementation of LSTM (Long Short Term Memory) with the attention mapping and reward mechanism offers a more refined analysis of emotions, significantly enhancing accuracy. It is designed specifically to look into the unique challenges of emotion recognition in autistic children. This customization contributes to a profound understanding of how autistic children experience emotions, potentially paving the way for early diagnosis of autism based on emotion recognition. This application of machine learning in the healthcare field holds promise for improving the well-being and support of autistic individuals.

V. ARCHITECTURE

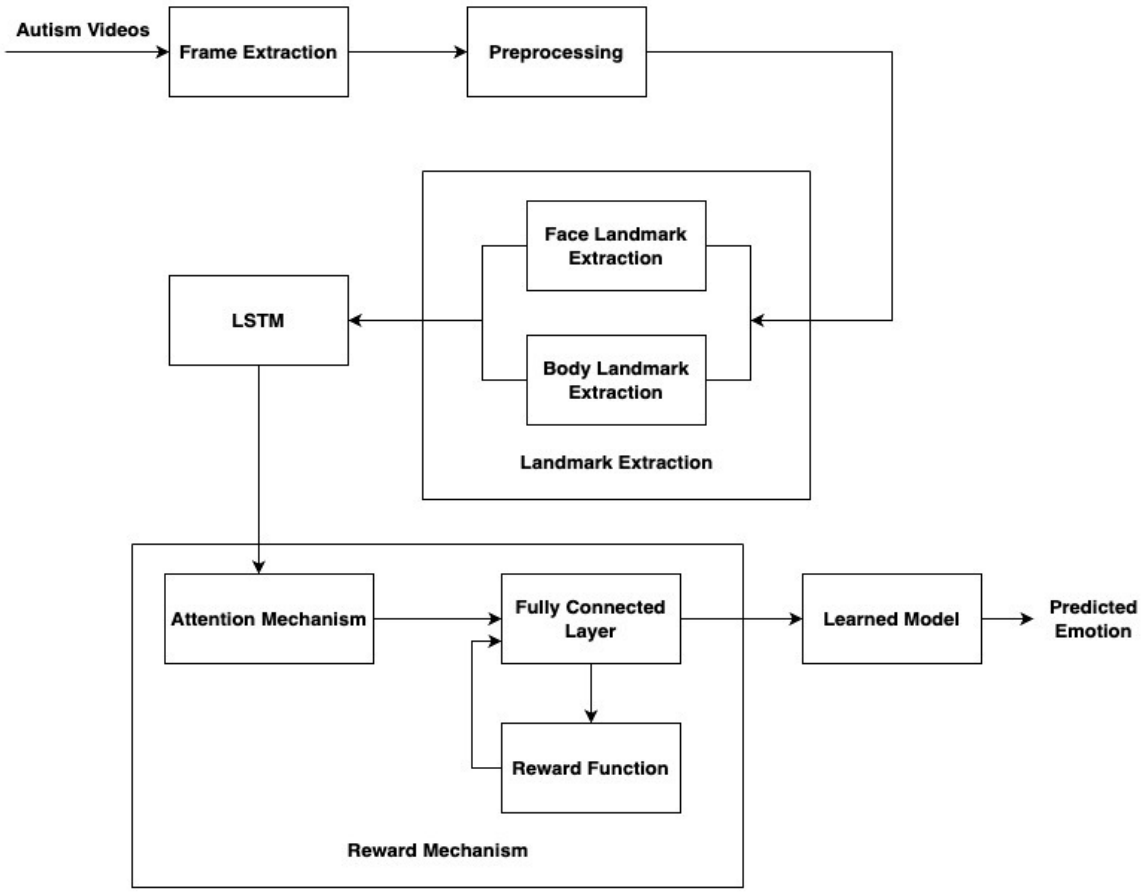

Figure 1: Architecture of the project

VI. IMPLEMENTATION

LSTM

The proposed work's network architecture comprises multiple layers designed for sequence processing and attention-based feature extraction. The model begins with an LSTM layer, configured for 30 time steps and 1662 features in each sequence, followed by a attention layer to emphasize relevant features. Two

additional LSTM layers capture temporal dependencies, leading to a final Dense layer stack for classification. The architecture is structured to leverage the strengths of LSTM units in sequence learning, with attention mechanisms enhancing the network's focus on crucial information. The model aims to discern patterns within input sequences, crucial for tasks like emotion detection, where temporal features are vital. Optimization is achieved through the use of rectified linear unit (ReLU) activations, and the final softmax layer facilitates multi-class classification. This comprehensive design seeks to extract meaningful representations from sequential data, leveraging attention mechanisms for enhanced discrimination, making it particularly suitable for tasks requiring understanding of temporal dynamics, such as emotion recognition in diverse datasets

Attention mechanism

The attention mechanism algorithm involves several key steps, and a commonly used version is the scaled dot-product attention. The algorithm is given below

Mathematical representation is given by

$$\text{Attention}(q,k,v)=\text{Softmax}\frac{q.(k^T)}{\sqrt{d}} \text{ . v} \quad (1)$$

Reward function

In this proposed work, four distinct reward functions are integrated into the training process. These reward mechanisms are strategically employed during loss propagation, contributing to the model's learning process. By incorporating diverse reward signals, it aims to reinforce precise emotion recognition and enhance the overall performance of the system.

The reward functions are given below,

FUNCTION 1:

The rewards are calculated using the equation 2 and the rewards are given in form of losses that is the positive reward makes the loss value lesser and the negative reward makes the loss value greater than the existing loss value.

$$Reward = (pred\_correct) - (0.1 * diversity\_penalty) - (0.2 * weighted\_penalty) + 0.3 \quad (2)$$

$$pred\_correct = \sum_{i=1}^{n} y_{true,i} - y_{pred,i} \quad (3)$$

$$diversity\_penalty = ||mean\_pred\_axis\_0 - mean\_pred\_axis\_1|| \quad (4)$$

$$weighted\_penalty = \sum_{i=1}^{n} (y_{true,i} - y_{pred,i})^2 * weight[i] \quad (5)$$

Where pred_correct in the equation (2) is the loss between the true value and the predicted value which is computed using the mathematical equation (3), diversity_penalty is to increase the diversity in the prediction of emotions to prevent the overfitting problem and it is computed by the equation (4) which considers the mean predictions along the different axes and the weighted_penalty accounts for the mistakes made by the model, the difference between the predicted emotion and true emotion represents the errors for each class. The penalties are weighted by the corresponding class weights which is given by the equation (4). All these parameters are combined along with the bonus value the reward will be calculated

FUNCTION 2:

The Reward function 2 is the modified version of the equation (2), it is also used to update the loss values for increasing the efficiency of the model. Equation (6) gives the mathematical representation.

$$Reward = (pred\_correct) - (0.1 * diversity\_penalty) - (0.3 * weighted\_penalty) + 0.4 + (0.2 * average\_attention\_score) \quad (6)$$

$$avg\_attention = mean(attention\_score) \quad (7)$$

Where the pred_correct represents the correctly predicted value which is already given in the equation (3), the diversity_penalty is given in the equation (4) and the weighted_penalty also same as the reward function 1, its mathematical representation is given by the equation (5). The difference between the reward function 1 and reward function 2 is the consideration of the average attention score which is computed by the equation (7).

3.3 FUNCTION 3:

The reward function 3 is the modified version of equation (6) where additional parameters are included to increase the performance of the model. The modified reward function in given in the equation (8).

$$Reward = (pred\_correct) - (0.1 * diversity\_penalty) + (0.2 * average\_attention\_score) - (0.3 * weight\_penalty) + n\_term \quad (8)$$

$$pred\_correct = \sum_{i=1}^{n} y_{true,i} * y_{pred,i} * (1 + \log(external)) \quad (9)$$

$$weight\_penalty = \sum_{i=1}^{n} \left(y_{true,i} - y_{pred,i}\right)^2 * weight[i] * attention\_score[i] \quad (10)$$

$$n\_term = (0.2 * external\_metric) + (0.3 * mean(\sqrt{attention\_score}) \quad (11)$$

Where the pred_correct is the cross entropy between the true value and the predicted value along with a logarithmic transformation which is given in the equation (9). Then diversity_penalty encourages the model to make the diverse predictions by penalizing the similarity between the mean prediction along different axes and the mathematical representation is given by the equation (4), weight_penalty is to penalize predictions errors in a weighted manner which takes both class weight and the attention score and it is represented in equation (10). Then the n_term acts as a bonus to balance the reward with the help of the external_metric and the attention score and it is given by the mathematical equation (11).

FUNCTION 4

The reward function 4 includes the temporal regularization components to focus on the temporal dynamic. This is mathematically represented in equation (12).

$$Reward = (pred\_correct) - (0.1 * diversity\_penalty) - (0.3 * weighted\_penalty) + bonus + (0.2 * avg\_attention) + n\_term + temporal\_reg \quad (12)$$

$$diversity\_penalty = \left|\left|mean\_pred\_axis\_0 - mean\_pred\_axis\_1\right|\right| * (1 - mean(attention\_score)) \quad (13)$$

$$avg\_attention = \sqrt{mean(attention\_score)} \quad (14)$$

$$temporal\_reg = e^{-mean(y_{pred})} \quad (15)$$

Where pred_correct is the cross entropy between the true values and the predicted values and the mathematical representation is given by the equation (9). diversity_penalty is different from other equations by including the mean of the attention score. It will penalize if the attention score is not distributed properly and it is computed by the equation (13). The avg_attention represents the average attention score and it is mathematically represented by equation (14), the use of square root and logarithmic transformation introduces non-linearities and scaling effects that can influence the contribution of each component. The term temporal_reg represents the temporal regularization which influences the model to have lower temporal influence when the mean predicted values are high. This can help prevent the model from being overly influenced by strong temporal patterns, potentially reducing overfitting.

VII. RESULTS AND EVALUATION:

Preprocessing

The frames are extracted from the videos and undergo transformation in shape, sharpness of the image to maintain the consistency of the dataset. Colour normalization will be applied to mitigate variations in lighting conditions, promoting robustness in feature extraction

Figure :2 Frame without preprocessing

Figure :3 Preprocessed frame

Landmark points extraction

Landmark point extraction from children involves capturing detailed information about facial, hand, and body features. This comprehensive set of 468 face landmarks, along with 21 hand and 33 body landmarks, enables a thorough representation of expressive behaviours. Facial landmarks, intricately mapping facial features, contribute to nuanced emotion analysis. Simultaneously, hand and body landmarks provide insights into gestural and postural aspects, enhancing the model's understanding of expressive cues. If the hands of the children is not visible in the frame, then they won't be considered for the emotion detection

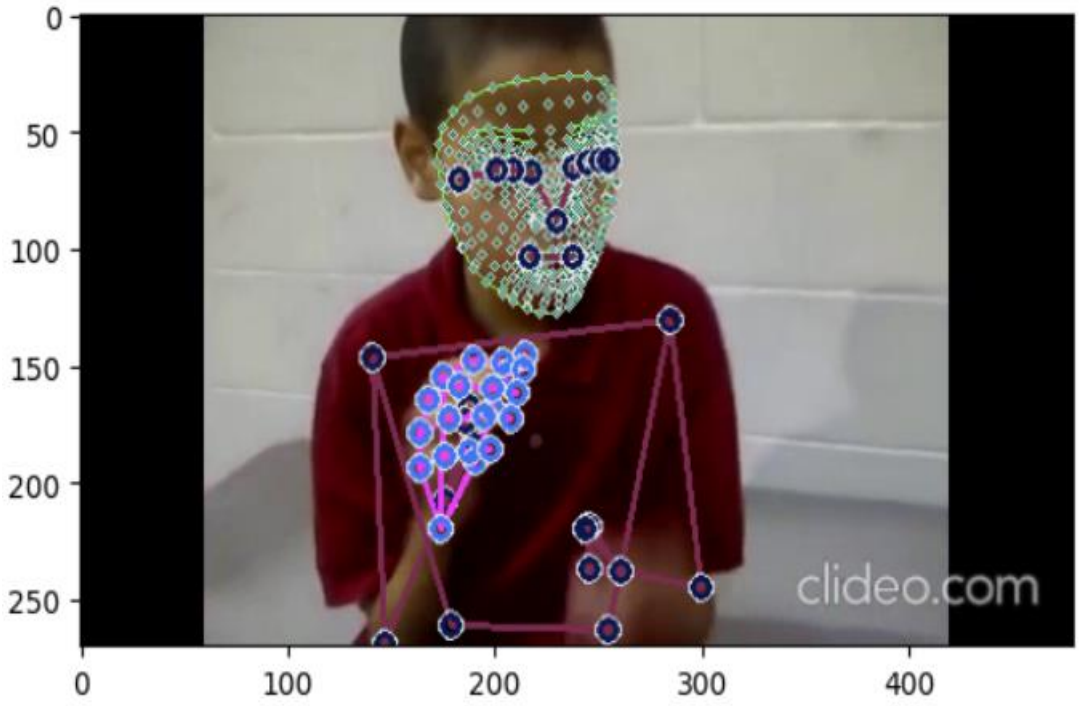

Figure :4 Visualization of Landmark points of the body

[50]:

	x	y	z
0	0.485183	0.373146	-0.014683
1	0.486839	0.325866	-0.040460
2	0.482950	0.338218	-0.017612
3	0.475040	0.282518	-0.038534
4	0.486828	0.312836	-0.044822
...	...	...	...
463	0.491132	0.233048	-0.002279
464	0.488311	0.238226	-0.007927
465	0.487360	0.242183	-0.013174
466	0.525187	0.212117	0.010914
467	0.528617	0.204004	0.011885

468 rows × 3 columns

Figure :5 Coordinates of Facial landmark points

[48]:

	x	y	z
0	0.362987	0.817125	1.574293e-07
1	0.399303	0.708579	9.842556e-04
2	0.416496	0.629553	-4.282036e-03
3	0.431118	0.573952	-1.292917e-02
4	0.449992	0.543840	-2.129436e-02
5	0.366381	0.575365	-5.570281e-03
6	0.397552	0.548853	-2.477447e-02
7	0.425556	0.553188	-4.092061e-02
8	0.446926	0.563511	-5.141919e-02
9	0.351219	0.613658	-1.678469e-02
10	0.381686	0.590764	-3.525076e-02
11	0.415211	0.593547	-4.602389e-02

Figure :6 Coordinates of right-hand landmark points

Emotion prediction

At the time of predicting the emotion, the model gives the probabilities of each emotion in each frame. As the children are autistic in nature, it will be beneficial to consider the top two emotions with highest probabilities.

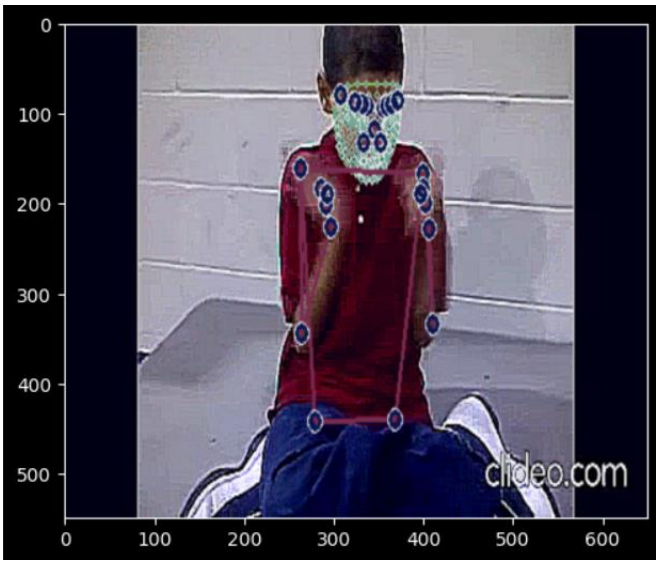

Figure :7 Intermediate frame of the predicted video

The subsequent image is structured as a tabular representation, with each row corresponds to a distinct frame. The 'frame' column serves as the index of the alternate frames in the predicted video, while the subsequent columns, such as 'anger,' 'fear,' 'happy,' 'neutral,' 'sad,' and 'surprise,' contain predicted probabilities for the respective emotion categories. Additionally, the 'MAX1' and 'MAX2' columns identify the two emotion categories with the highest scores for each frame

[143]:

	frame	anger	fear	happy	neutral	sad	surprise	MAX1	MAX2
0	1	0.041474	0.008115	0.560596	0.378209	0.006541	0.005065	happy	neutral
1	3	0.042845	0.007582	0.709878	0.228023	0.006488	0.005186	happy	neutral
2	5	0.044415	0.006955	0.775919	0.161300	0.006388	0.005022	happy	neutral
3	7	0.035339	0.004838	0.841645	0.109401	0.004687	0.004090	happy	neutral
4	9	0.024637	0.002761	0.897648	0.069230	0.002939	0.002786	happy	neutral
5	11	0.017624	0.001512	0.937569	0.039794	0.001752	0.001751	happy	neutral
6	13	0.013091	0.001008	0.960224	0.022941	0.001334	0.001402	happy	neutral
7	15	0.009323	0.000429	0.972836	0.015998	0.000687	0.000727	happy	neutral
8	17	0.005105	0.000247	0.985514	0.008156	0.000449	0.000529	happy	neutral
9	19	0.003677	0.000153	0.989486	0.005964	0.000327	0.000394	happy	neutral
10	21	0.002734	0.000064	0.992323	0.004396	0.000233	0.000251	happy	neutral
11	23	0.001703	0.000022	0.995287	0.002728	0.000143	0.000117	happy	neutral
12	25	0.001374	0.000012	0.996244	0.002190	0.000100	0.000081	happy	neutral
13	27	0.000889	0.000004	0.997746	0.001273	0.000050	0.000038	happy	neutral
14	29	0.000637	0.000001	0.998568	0.000747	0.000028	0.000018	happy	neutral
15	31	0.001852	0.000012	0.991608	0.006118	0.000226	0.000185	happy	neutral
16	33	0.004384	0.000094	0.988859	0.005852	0.000543	0.000269	happy	neutral

Figure :8 Probabilities of emotions in each frame

Performance comparison between reward function

Now, the performance of the prediction without reward function and with different reward functions. The graphical representation of the training loss and training accuracy for all the functions given below

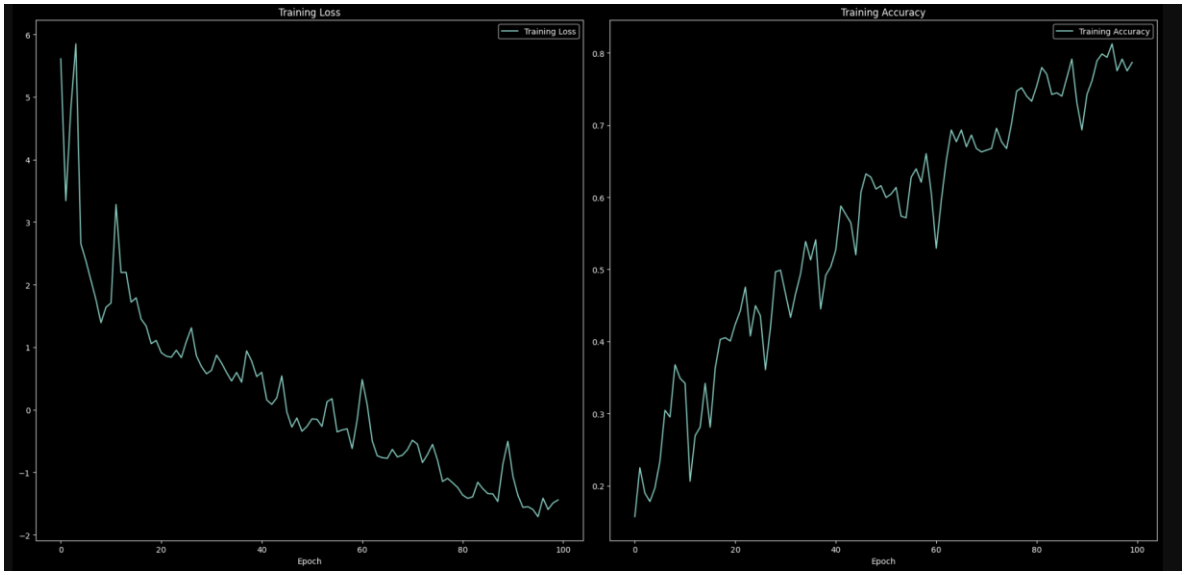

Figure :9 Training loss and training accuracy graph with reward function 1

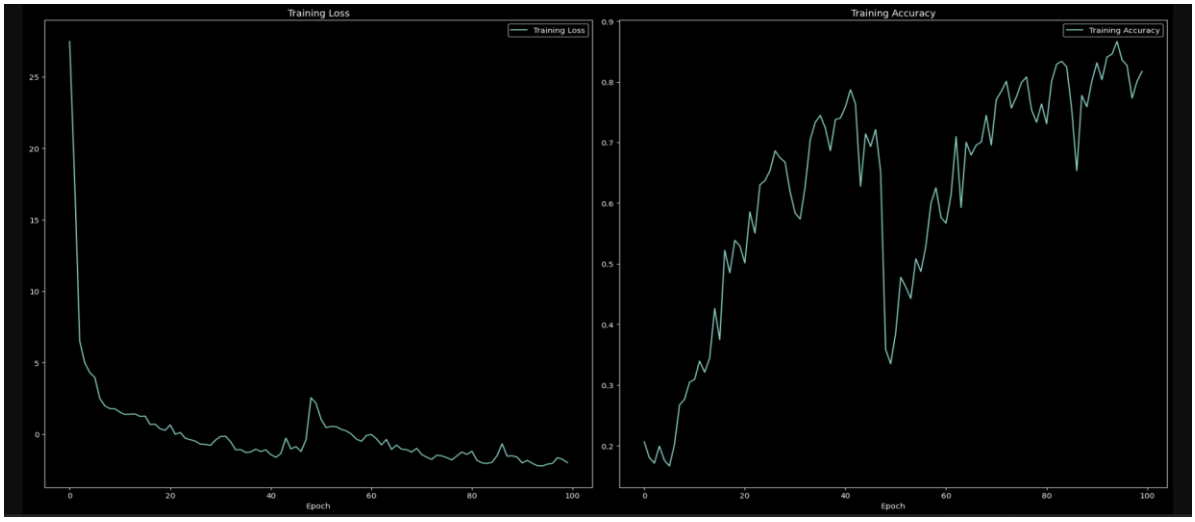

Figure :10 Training loss and training accuracy graph with reward function 2

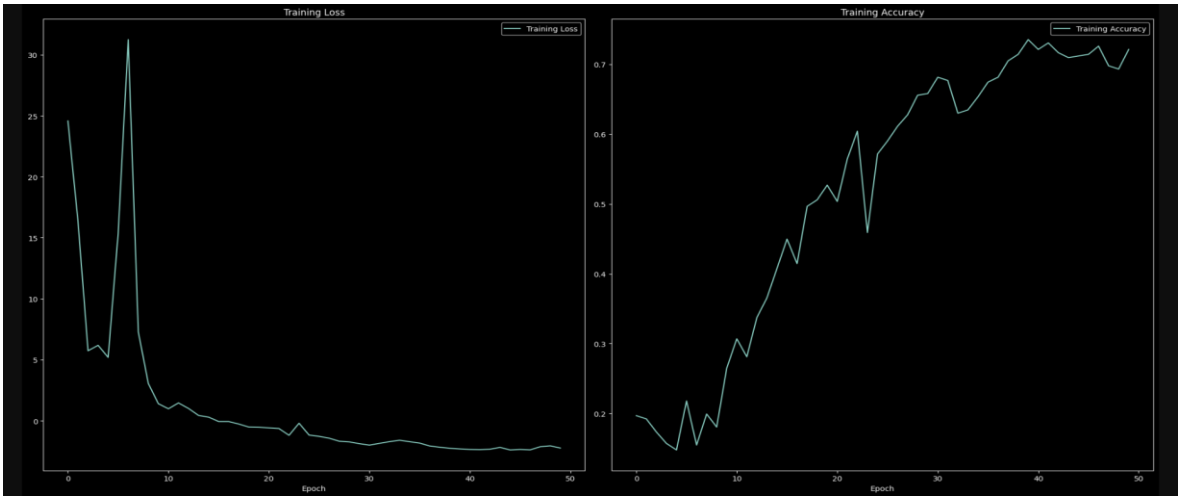

Figure :11 Training loss and training accuracy graph with reward function 3

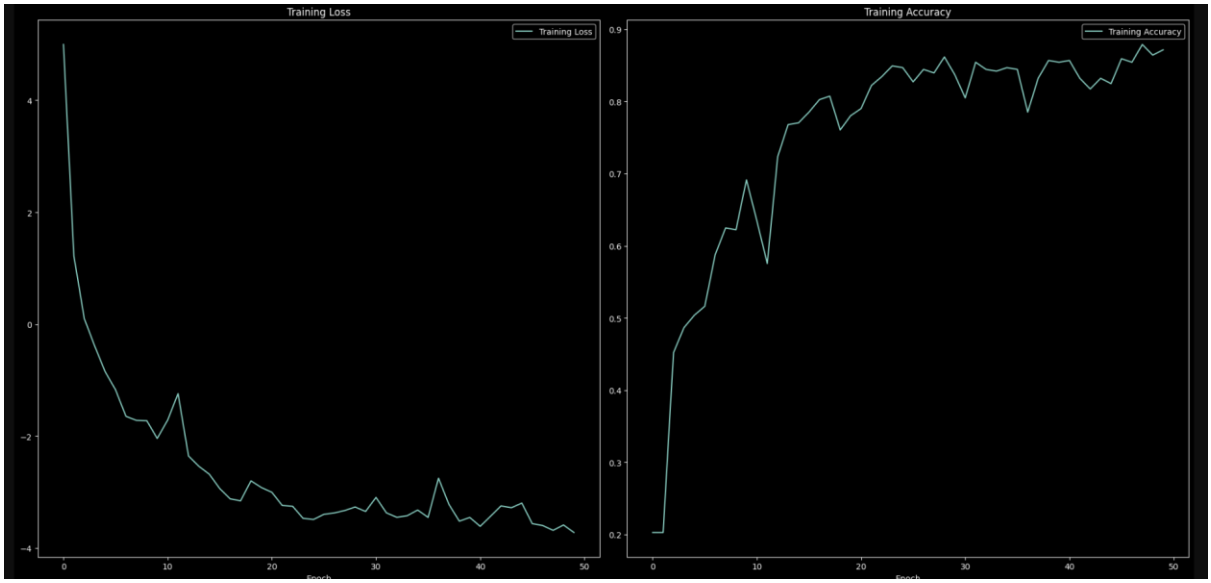

Figure :12 Training loss and training accuracy graph with reward function 4

Then these models with different reward functions are tested with the test data which give the different accuracies for each function, some function gives good accuracy and some functions yields lesser accuracy than the model without any reward function.

Table :1 Performance comparison based on the accuracy score

FUNCTION	ACCURACY
No Reward function	79.26
Reward Function 1	78.52
Reward Function 2	82.96
Reward Function 3	77.78
Reward Function 4	88.15

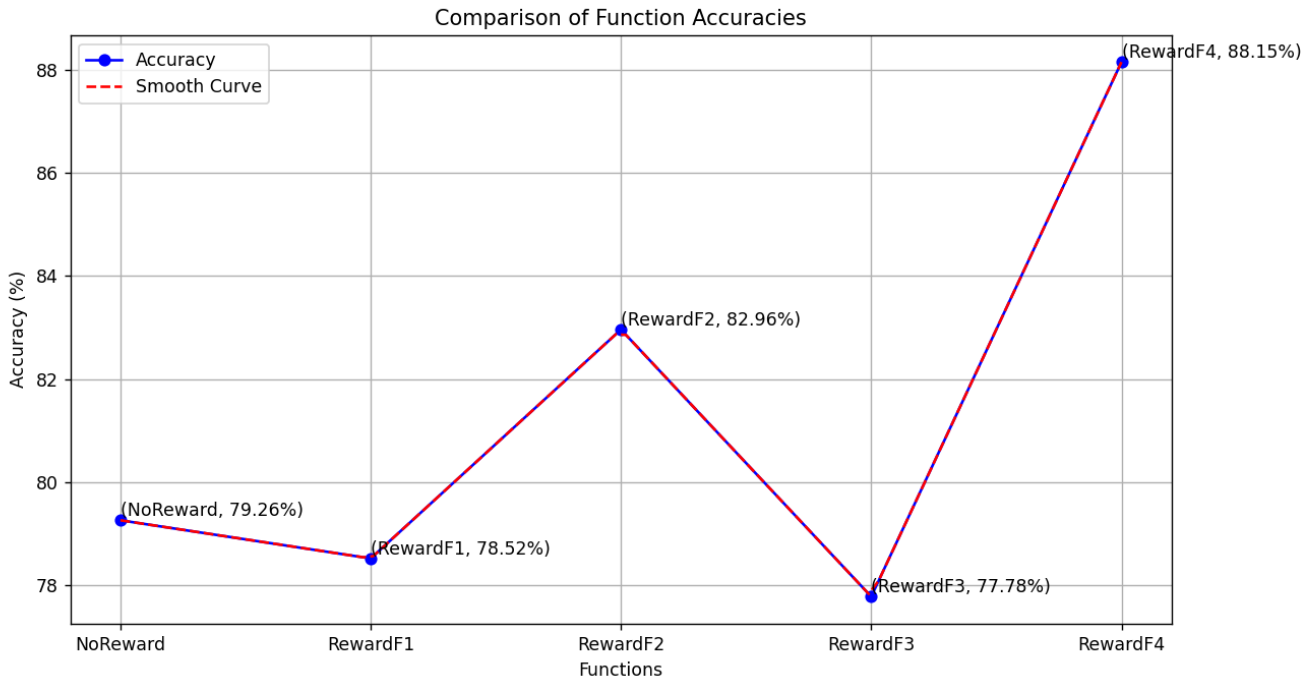

Figure :13 Graphical representation of performance of the different reward functions

The performance evaluation of different reward functions shows different impacts on the accuracy of the model. In the absence of a specific reward function that only attention mechanism is implemented which achieves an accuracy of

79.26%. After introducing the first reward function, leads to a slightly lower accuracy of 78.52%, suggesting that this particular reward function might not significantly enhance the model's performance. However, second reward function shows a notable improvement with an accuracy of 82.96%, which indicates its positive impact on predictive capabilities on the test data. But the third reward function yields an accuracy of 77.78%, slightly below the baseline. The most promising result comes from Reward Function 4, which achieves the highest accuracy of 88.15%, making it the most effective in enhancing the model's overall performance. The selection of a proper reward function is important for optimizing the model's performance, and Reward Function 4 shows better performance in this comparative analysis.

VIII. CONCLUSION

In conclusion, this project aims to enhance the emotional well-being of autistic children by creating a specialized system for emotion detection. The use of GANs helps to overcome challenges like occluded faces to ensure that emotions can be recognized even when parts of the face are hidden. This system combines facial expression analysis, body movement tracking, and landmark detection to provide a comprehensive view of the child's emotional state. This not only aids in better understanding how autistic children experience emotions but also equips educators, therapists, and caregivers with valuable insights to provide effective emotional support.

IX. REFERENCES:

[1] Z. Shao, Y. Zhou, J. Cai, H. Zhu and R. Yao, "Facial Action Unit Detection via Adaptive Attention and Relation," in IEEE Transactions on Image Processing, vol. 32, pp. 3354-3366, 2023, doi: 10.1109/TIP.2023.327779

[2] G. Pons and D. Masip, "Multitask, Multilabel, and Multidomain Learning With Convolutional Networks for Emotion Recognition," in IEEE Transactions on Cybernetics, vol. 52, no. 6, pp. 4764-4771, June 2022, doi: 10.1109/TCYB.2020.3036935.

[3] Yao, L., Wan, Y., Ni, H. et al. Action unit classification for facial expression recognition using active learning and SVM. Multimed Tools Appl 80, 24287–24301 (2021).

[4] M. K. Tellamekala, Ö. Sümer, B. W. Schuller, E. André, T. Giesbrecht and M. Valstar, "Are 3D Face Shapes Expressive Enough for Recognising Continuous Emotions and Action Unit Intensities?," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2023.3280530.

[5] P. R. K. Babu et al., "Exploring Complexity of Facial Dynamics in Autism Spectrum Disorder," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 919-930, 1 April-June 2023, doi: 10.1109/TAFFC.2021.3113876.

[6] M. Talaat, Fatma. (2023). Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Computing and Applications. 35. 10.1007/s00521-023-08372-9.

[7] Geethu Miriam Jacob, Bjorn Stenger; Facial Action Unit Detection With Transformers, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 7680-7689

[8] N. Zhang, M. Ruan, S. Wang, L. Paul and X. Li, "Discriminative Few Shot Learning of Facial Dynamics in Interview Videos for Autism Trait Classification," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1110-1124, 1 April-June 2023, doi: 10.1109/TAFFC.2022.3178946

[9] Rajaram, Santhoshkumar & Geetha, M.. (2019). Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks. Procedia Computer Science. 152. 158-165. 10.1016/j.procs.2019.05.038.

[10] Hashemi J et al., “Computer vision analysis for quantification of autism risk behaviors,” IEEE Trans. Affect. Comput., vol. 12, no. 1, pp. 215–226, First Quarter 2021