

Emotion Exploration in Autistic Children as an Early Biomarker through R-CNN

S. P. Abirami^{1,*}, G. Kousalya¹ and R. Karthick²

¹Coimbatore Institute of Technology, Coimbatore, India

²Steps Rehabilitation Center (I), Steps Groups, Coimbatore, India

*Corresponding Author: S. P. Abirami. Email: abirami.sp@cit.edu.in

Received: 20 January 2022; Accepted: 01 March 2022

Abstract: Autism Spectrum Disorder (ASD) is found to be a major concern among various occupational therapists. The foremost challenge of this neurodevelopmental disorder lies in the fact of analyzing and exploring various symptoms of the children at their early stage of development. Such early identification could prop up the therapists and clinicians to provide proper assistive support to make the children lead an independent life. Facial expressions and emotions perceived by the children could contribute to such early intervention of autism. In this regard, the paper implements in identifying basic facial expression and exploring their emotions upon a time-variant factor. The emotions are analyzed by incorporating the facial expression identified through Convolution Neural Network (CNN) using 68 landmark points plotted on the frontal face with a prediction network formed by Recurring Neural Network (RNN) known as the RCNN based Facial Expression Recommendation (FER) system. The paper adopts Recurring Convolution Neural Network (R-CNN) to take the advantage of increased accuracy and performance with decreased time complexity in predicting emotion as textual network analysis. The papers prove better accuracy in identifying the emotion in autistic children when compared over simple machine learning models built for such identifications contributing to autistic society.

Keywords: Autism spectrum disorder; facial expression; emotion; CNN; RNN

1 Introduction

In India, the rise of autism rate is witnessed over the years as specified by the Autism Society of India [1]. Though the reasons behind this autism seemed to be a still ongoing exploration, the identification along with basic biomarkers and proper guidance are the target of many researchers in India. Also, in India, there are no specific social, community, or population-based studies which may provide exact predominance of the autism spectrum. Most of the individual in our country with ASD are still not diagnosed properly and the majority lives a life of ignorance, deprivation, and non-acceptance nature of the disability [2].

Although the root cause and mechanism contributing to ASD could not be completely identified, it is apparent that the abnormal neurotransmission in brain regions hinders the child from normal behavior and



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

motor actions to an altered behavior [3]. This impairment of mental difficulties could be identified with regular periodic analysis of the child's communication behavior, social interaction, object identification, emotional sequence, linguistic ability, and knowledge adaptation rate [4]. These abilities of the child are observed between the ages of 10 months to 3 years. Owing to the special consideration and sometimes to an extent of negligence, the children fail in thinking of them beyond their capability which leads to psychological barriers that are expressed as emotions in them.

The ultimate responsibility in training autistic children lies with the high priority scale of early identification and exploring its characteristics. Additionally, analyzing the functional complexity such as low and high functioning autism and on least priority lies the appropriate coaching [5]. The clinical procedure has strongly proved that the children respond to the interaction based on their neurodevelopment overages. The clinical evaluation and analysis are through the regular dyads manner which is considered to be a time-consuming process. These clinical evaluations take place with input from the parents and caretakers. Also, many other identifications methods were adopted to evaluate the nature of autism in every child-like vocal analysis, identification of objects, interest factor analysis, etc [6].

With the advent of facial expression contributing towards the publishing of neural changes and responses to an event, the proposed paper builds the system for early identification of autism through expression and emotional analysis. The FER system analyses the various expressions perceived by the children and a comparison of varied expressions possessed would result in gaining insight into the pattern of expression and interest in every child. This active learning of the neural network models could be adopted for autism evaluation at a very early stage where the child tends to make expressions towards the mother and the family members between the ages of 4 to 8 months. This holds a pre-screening technique to the clinical evaluation that hints at the autistic characteristics and does not utterly confine into the autism spectrum.

Summary: The research paper deals in briefing the introduction and need for the proposal in Section 1; the state of art made towards the paper in Section 2; the proposed methodology for emotion identification in Section 3; followed by the result discussions and conclusion along with future scope in the consecutive sections.

2 State of Art

Autism being one of the notable neurodevelopmental disorders, the basic reasons, symptoms, and combination of varied characteristics of the spectrum is still an ongoing investigation. Though there are various methodologies evolved in exploring autism in children and adults, there has always been a high demand for early intervention. This is owing to the rapid increase in the development of intelligence systems. The basic characteristics are evaluated through defined and authorized checklists by clinical experts. This is further extended by many researchers such as incorporating facial expression, voice modulation, automatic movement tracking, brain image analysis, and gene expression analysis along with the basic evaluation. Those evaluations have proved that high functioning of autism spectrum disorder will have significantly impaired communication abilities in the social environment.

Among the various disabilities, expression and emotional discharge in high functioning autism depict a clear picture of the neurodevelopmental state of autistic people. Such exposure of expressions can either be in a contactless environment or can be under the influence of the object they face. It is this important to differentiate between the emotions possessed and the cause of emotion in autistic people [7].

The children with autism will show varied expressions based on the event and routine that they experience, Van Eylen et al. (2018) [6]. These abilities are difficult to perceive and their ambiguous expression is compared to the regular developing toddlers' counterparts to a greater extent as discussed

by Guha et al. (2016). The author employs a new computational intelligence method that analyses facial expression to support autism [8]. Though the characteristics and the diagnostic probability of autism varying between genders are based on sensory symptoms, the basic characteristics of autism remain the same at the early developmental stage as discussed by Duvekot et al. (2017) [3].

Classification of facial expressions of the children was focused when the target object possesses anger and happy emotion, Whitaker et al. (2017). Exploring such emotions through facial detections using machine learning algorithms could result in the early identification of autism before the clinical analysis [9]. Though, to further improve the accuracy in classification and better reliability of the screening mechanism, a deeper feature identification and feature analysis should be involved as discussed by the author.

Also, the facial feature specification was proposed and upgraded by researchers that categorized the face into T-shaped structure extracting eyes, nose, and position in three feature dimensions, Wang & Adolphs (2017). The insights were analyzed based on the facial expression evaluated over the center axis of the face including eyes, nose and mouth. The major difficulty faced was the position of the face with respect to view the entire features of the face [10].

Similarly, the method that analyzes the geometric facial features to establish the face models of every individual was proposed, Gopalan et al. (2018). This geometric analysis was made with light illumination, orientation, and color scaling of the captured image [11]. This did not prove its best to the real-time applications, as the considered parameters are harder to achieve in real time applications with varied light rays enlightenment.

The facial expression analysis experimented with facial region detection using the Haar technique that covered the regions of the face like mouth and eyes, Bone et al. (2016). The face image was divided into two regions on the application of Sobel edge detection that classified the regions with facial features. Similarly, various other researchers have given reasonable accuracy for FER systems through techniques such as Histogram of Oriented Gradients (HOG), Support Vector Machine (SVM), random forest, basic neural network, etc [12,13].

Considering the advantage of the deep neural network, an attention convolutional network that works in identifying the facial expression, by classifying the regions of the face that contribute major towards the expression identification, Shervin et al. (2019) [14]. The basic CNN architecture was improved for its accuracy namely incremental boosting CNN, identity-aware CNN, CNN with varied hidden layers thus resulting in modified CNN structure for FER systems.

Likewise, there are various possibilities in inculcating advanced features selection and feature analysis for face detection in a practical real-time environment. The ultimate aim of the technical improvement is to produce a better solution until the false positive detections are rejected quickly during the early stages, Jain et al. (2018) and Jan et al. 2019 [15].

The proposed paper importantly bridges the gap identified in the early exploration of autism behavior as the developmental ability is so far measured using a biomarker on how accurate the autistic children identify the expressions of the human they face. This is to analyze the perseverance and processing capability of the child's brain to differentiate between expressions. This kind of evaluation was extended in the view of understanding objects by the children under static and dynamic considerations rather failed to measure the expression and emotion perceived by the children [16,17]. In view of the expression identification, emotion, and social behavior of the autistic child, the paper aims in building a model for such neural connectivity and articulating nature.

3 Proposed Methodology

With the dawn of the improvisation in the technological field relating to healthcare development in particular to autism, the proposed methodology aims in identifying facial expression and emotion along with the time-variant. The block diagram for the proposed architecture is shown in Fig. 1.

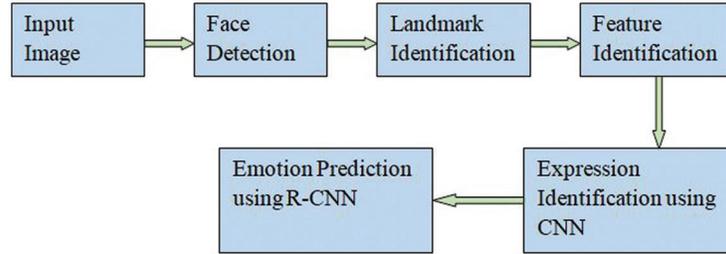


Figure 1: Basic architecture of the proposed RCNN-FER system

3.1 FER System Using CNN

The facial expression of the children is identified on the application of CNN. In reference to our previous work [18], in the proposed technique, the input images were pre-processed in resizing them to a 64 x 64 dimension image. The images were grouped based on the various classes of expressions in their labeled order. The input image is then processed to identify the 68 facial landmark points distributed over the frontal face. The nose points are kept static to analyze the angle of facial position. The neural network is then trained in an order under each category of expressions to identify the important features in the facial structure. The convolution neural network is trained for both ASD positive (ASD^{+ve}) and ASD negative (ASD^{-ve}) to check for the training correctness.

The first convolution layer of the proposed CNN-FER include 32 3 x 3 filters with a stride of size 1. This hidden layer does includes batch normalization but with no max-pooling technique. The second convolution layer of CNN-FER, has 64 3 x 3 filters, with the stride of size 1, including batch normalization and max-pooling with a filter size 2 x 2. The Fully Connected layer has a hidden layer with 512 neurons and a Sigmoid activation function as specified in Eq. (1) owing to the multi labeled classification.

$$f(x) = \begin{cases} \frac{1}{1 + e^{(-x)}}; & \text{if } x > 0 \\ 0; & \text{if } x \leq 0 \end{cases} \quad (1)$$

In the proposed CNN-FER system, every image with facial detection having plotted the 68 landmark points is fed as input. The system based on landmark identification computes the points with x and y coordinates indicated as (l_x, l_y) . In general, with 68 landmark points the convolution layer will have (l_{1x}, l_{1y}) to (l_{68x}, l_{68y}) neurons defined where the landmark l specified as 1 to 68 are the facial landmark points pointed on the detected face. These landmarks are annotated and are labeled from each expression varying in the order of anger, disgust, fear, happy, sadness, neutral, sleep in the implementation. Each point in the landmark is profiled four major vectors for combinatorial analysis such as x-axis, y-axis, angular distance from a head central point and angle of a landmark [19]. Since autistic kids do not fall under the emotion category of contempt and anticipation like non autistic people, those emotions are not considered. Rather, autistic kids tend to avoid humans and other uninterested objects. This makes them avoid eye contact and close their eye which is perceived as sleep in the proposed RCNN FER system.

3.2 Emotion Prediction through R-CNN

The Recurring Neural Network (RNN) differs in the neural network structure by taking strong advantage of the feedback input from the previous layer. The major advantage of combining CNN with RNN is to solve the problem better when the state of output is static. This is because of the property of RNN taking activated outputs from the learned classifier to converge into a static solution space in an accurate manner during the training process. The ultimate aim in combining RNN with CNN is to predict the emotion of children with ASD from expression involving time-variant. The emotion of the children is perceived in obtaining a continuous analysis of expression over stages. Thus by analyzing the expression over a period of time, the emotion for the next $(t+1)^{th}$ time-variant could be predicted and can be classified.

The basic architecture of integrating the CNN with RNN is depicted in Fig. 2. On examining the facial expression of the children using CNN with maximized accuracy, the input layer at $t-1$ of RNN takes the processed expression of the CNN. The RNN network is then fed with input expression evaluated over the time period (t) , $(t+1)$, up to $(t+n)$ based on the computational complexity evaluation.

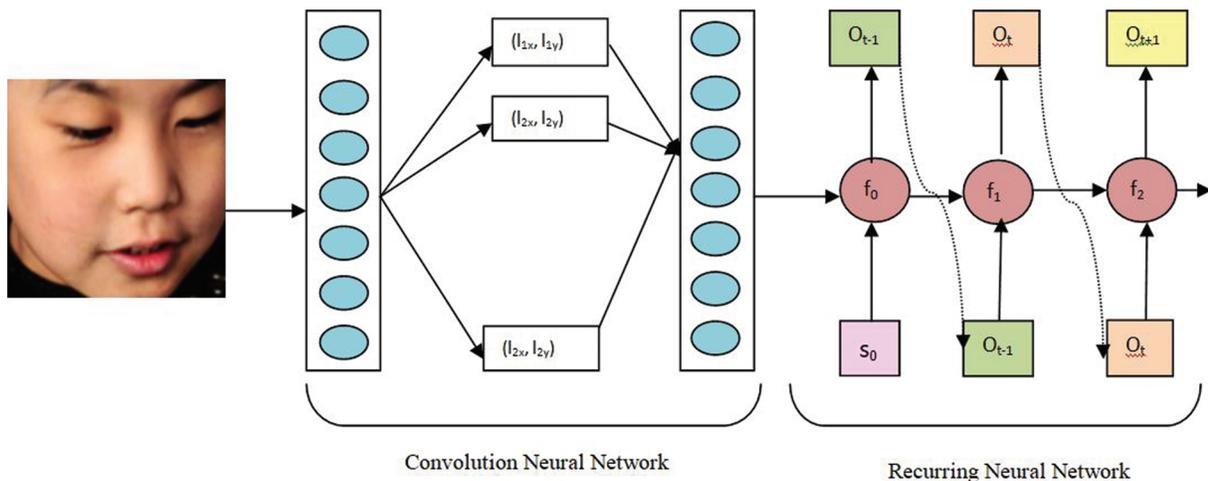


Figure 2: System architecture of the integrated CNN and RNN

The combined architecture of R-CNN evaluates the facial expression from the CNN network, using images captured over the frames. Consequently, this in turn is evaluated by RNN for emotion as a text processing system. This combined architecture leads to better identification of emotion over a specified period.

The architecture of unfolded recurring neural network sequence explains x_{t-1} , x_t and x_{t+1} as the input to the operating function, S_{t-1} , S_t and S_{t+1} respectively. O_{t-1} , O_t and O_{t+1} are the output generated from the input activated using x_i and W as weight from the feedback of the previous neuron. U and V are the weight vector for the hidden layer S_t and output layer O_t respectively. The state output of the neuron is computed as in Eq. (2) and the parameters are mapped to the referred Fig. 2.

$$S_t = f(U \times x_t + W \times S_{(t-1)}) \tag{2}$$

$$Output, O_t = ActivationFunction(V \times S_t) \tag{3}$$

The classification based on the recurring network is computed as in Eqs. (4) and (5). Where, Eq. (4) specifies the functional evaluation and Eq. (5) specifies the output of the current state involving source, function and activated feedback from the previous state.

$$f(t) = g_f(w_i x^{(t)} + w_f f(t-1) + b_f) \quad (4)$$

$$o(t) = g_o(w_o f(t) + b_o) \quad (5)$$

where x is the input to the function f at $(t-1)$, (t) , $(t+1)$; $f(t)$ is the function that is specified for every layer of RNN say f_0, f_1, \dots, f_n ; w_i are the weights acting upon the input x on each layer and S_0 is the source weight transferred to the initial input layer of RNN along with computed weight from the previous layer to the next layer specified by time (t) and $(t+1)$; $o(t)$ is the output of each layer at current state that is to be transferred to the next layer and b is the bias value.

The facial expression of the children with and without ASD is examined through a deeper analysis by a convolution neural network. Every frame that undergoes the neural network will result in the high prioritized facial expression apparently identified in the child. The expression is saved as the feedback input from time $(t-1)$ to the RNN network that takes the expression of the current frame at a time (t) . The RNN analyses the expression from the frames over a specified period to trigger the change in expression possessed by the children. The combination of expression from various frames over the network identifies the emotion perceived by the children.

The integrated R-CNN for emotional analysis takes the expression with the maximized probabilistic ratio for the $(t-1)$ roll as the weight of the function to $f(t)$ as specified in Eq. (4). The (t^{th}) layer of the R-CNN computes the expression of the (t^{th}) frame and the activating function is guaranteed to operate with the source input of (t^{th}) frame and the feedback from the $(t-1)^{th}$ frame. The layer 1 of the R-CNN depicted in Fig. 3 integrates the CNN with RNN by exhibiting the feedback input commonly termed as a weight factor to the next functional computation. The RNN figured out in Fig. 3 consists of one hidden layer with max-pooling function activated by Softmax function as in Eq. (6).

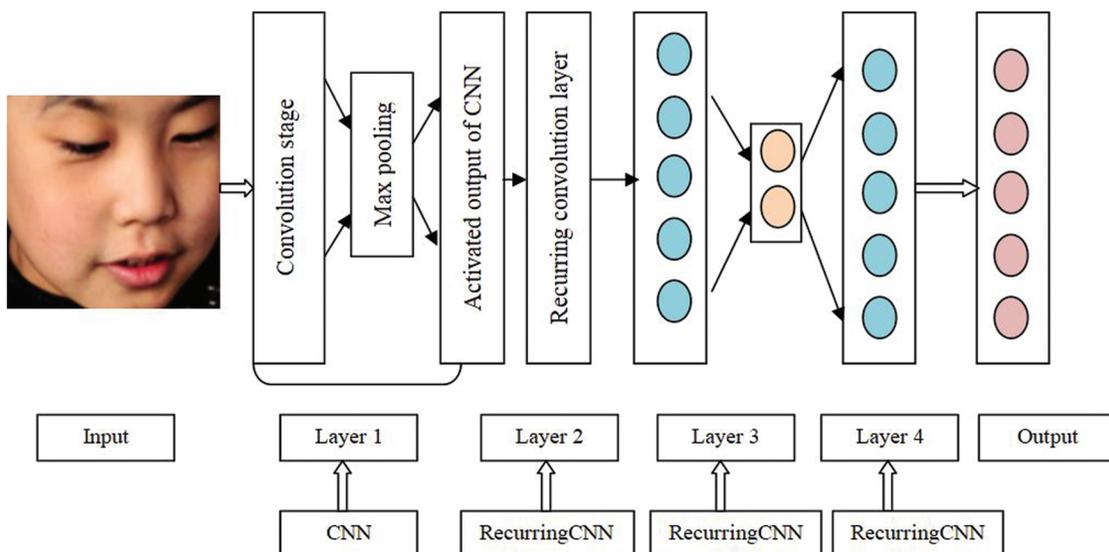


Figure 3: Working of R-CNN for emotion analysis in children

$$f(x) = \begin{cases} \frac{e^{f_i(x)}}{\sum_{j=1}^n e^{f_j(x)}}; & \text{if } x > 0 \\ 0; & \text{if } x \leq 0 \end{cases} \quad (6)$$

The input to any layered function in R-CNN at time (t) is the expression classified through simple implemented CNN. This is because any frame at $(t - 1)^{th}$ period undergoes a deeper analysis through CNN. The recurring networks pass the maximum probabilistic value from the function towards the next layer of R-CNN. The change in the output expression over the time period (t) is specified to be the emotion in the children and the change of emotion could also be inferred from the intermediate results of the layer. The output of the network at time t is denoted as in Eq. (7).

$$x_{(t+1)} = \sigma[f(x_t) + g(x_t)] \quad (7)$$

where, $f(x_t)$ and $g(x_t)$ are the weighted functional inputs to the neural network. The output of the function σ is denoted as Eq. (8)

$$\sigma(x) = \begin{cases} x; & \text{for } -1 < x < 1 \\ -1; & \text{for } x \leq -1 \\ 1; & \text{for } x \geq 1 \end{cases} \quad (8)$$

The proposed system is used to analyze the change in emotion among children with and without autism. The system also focused to ensure the change within a time interval and the pattern of change in emotion among them.

4 Dataset Descriptions

The data for the analysis of the proposed system is collected from countable children with and without autistic features and later the sample analysis was also made in analyzing videos of autistic children presented over the internet. As the video is generally the sequence of frames, they are processed sequentially. In order to satisfy the processing requirements and the capacity of computation, the dimensions of the input video are chosen to be 320 X 240. Further, the RGB frames are converted to grayscale where a grayscale image consists of only the intensity information of the image rather than the apparent colors. RGB vector is with three dimensional is converted into a gray-scaled vector is one dimensional with the inculcation of predefined function in OpenCV. After the data acquisition and preprocessing based on the computing capacity, the data are fed to the proposed evaluation system. Furthermore, the CNN system as an individual was evaluated over the still images of children with and without autism from Kaggle with a count of 304 and 390 images respectively. The accuracy of the system is proved better over the existing systems [6,7] which adds to the proven improvement for the RCNN FER system.

5 Results and Descriptions

Although the computational complexity in integrating CNN along with RNN for video data analysis is increased, the GPU processing capability for real-time video analysis could be employed. This could support better performance of the system with minimum computational and time complexity. The facial expression identified from the real-time analyzed ASD children using R-CNN resulted in an accuracy of 91% with an increase of 2% compared to CNN. The average probability of the children with various facial expressions is specified in Tab. 1 and the maximum perceived expression is found to be disgust, neutral, and happy in their

respective order. It is also observed that there was no change observed in the maximum perceived facial expression guaranteed from SVM and CNN classifiers rather influencing the classification.

Table 1: Probability of emotions identified using R-CNN

Facial expressions	ASD positive (Probability of expression)
Anger	0.036045
Disgust	0.277319
Fear	0.129893
Happy	0.144559
Sadness	0.141052
Neutral	0.20356
Sleep	0.070315

The emotion of the children with ASD is considered by capturing their facial expression along the period t , $t+10s$, $t+20s$, $t+30s$ respectively. The captured frames for expression analysis are sent through the combined R-CNN that evaluates the change in facial expression to be identified as emotional changes and the pattern of emotion in them. The rate of frames captured could be varied between 1 to n , and the implemented system captures an average amount of frames in a view of not missing any sequence of expression exposed by the children. The frames under the period are captured continuously at the rate of 30 frames per second. The frames are then dropped in a random manner so as to include frames of maximum variations. This also assures minimum computation complexity by reducing the number of frames to be analyzed. Thus by taking frames separated by period will support for emotional change in the child. [Tab. 2](#) depicts the probability value of the expression contributing to emotion under a specified time interval.

Table 2: Probability of various emotions observed during a range of time intervals identified using R-CNN

Facial expression	At time t	$t + 10s$	$t + 20s$	$t + 30s$
Anger	0.036045	0.036154	0.046045	0.043152
Disgust	0.277319	0.223589	0.207319	0.18456
Fear	0.129893	0.12785	0.129893	0.12389
Happy	0.144559	0.16243	0.154455	0.156711
Sadness	0.141052	0.13984	0.151052	0.149986
Neutral	0.20356	0.23487	0.25356	0.25876
Sleep	0.070315	0.07984	0.083145	0.102457

[Tab. 2](#) portrays the probability of facial expression observed from the autistic child during a specified time interval contributing to emotion pattern. The table also insights that the pattern of expression remains standard over a period of time and the change in the expression towards the period over the

variant is observed to shift from disgust to neutral. To a certain extent, the probability value got maximally biased oversleep and anger in an average. These patterns of emotion could be identified over a period of time to analyze the maximum emotion perceived by the child and also can be used to determine the pre-perceived expression before any stage (especially during aggression). These insights the rate of interest of the child over any specific object or under circumstances.

Fig. 4 depicts the emotion perceived by the children on an average during specified time duration. The figure depicts that there are no sudden falls and rise of emotion in children rather falling into the next highest and nearest emotional probabilistic value. The evaluated countable samples do not show many diverse emotions but there exist various cases that the subject responds vigorously with respect to any disliked objects. Such variations could be analyzed and can be used to control the child when the appropriate trigger is made. Every child’s behavior could be trained under all circumstances so as to ensure better accuracy and reliability of the system.

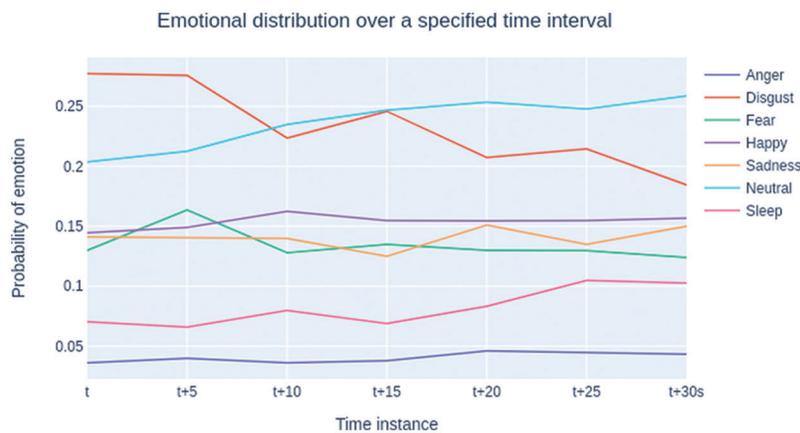


Figure 4: Probability of emotion observed using R-CNN over a time interval

Fig. 5 depicts the change in emotion pattern observed during time t and $(t + 30)$ sec in an autistic child. This insight the change in emotion over a period of time in an average and on further granular analysis patter of emotion in every autistic child could be analyzed to trigger and alert them falling aggressive. This could also be utilized by the parents and caretakers to analyze the duration of change in their child and keep them engaged excluding boredom.

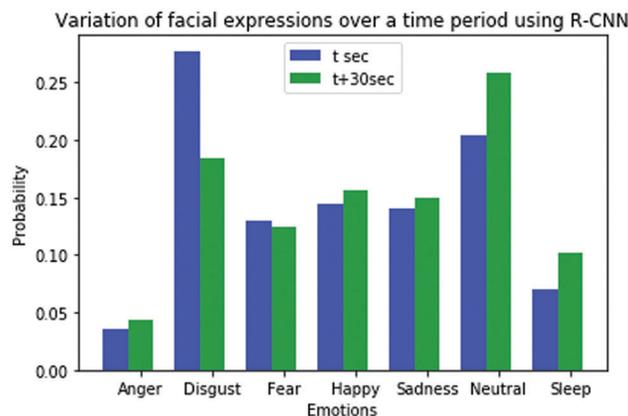


Figure 5: Rise and fall of emotion observed between the initial and end time-variant

Fig. 6 depicts the graphical representation of the change of emotion in children with and without autistic characteristics. The values are mapped to the plot on the execution of the RCNN system and observing the values of emotion for about 50 ticks. The minimum probability emotions could be eliminated and the maximum emotion perceived could be taken into consideration for further clinical analysis by clinicians and occupational therapists.

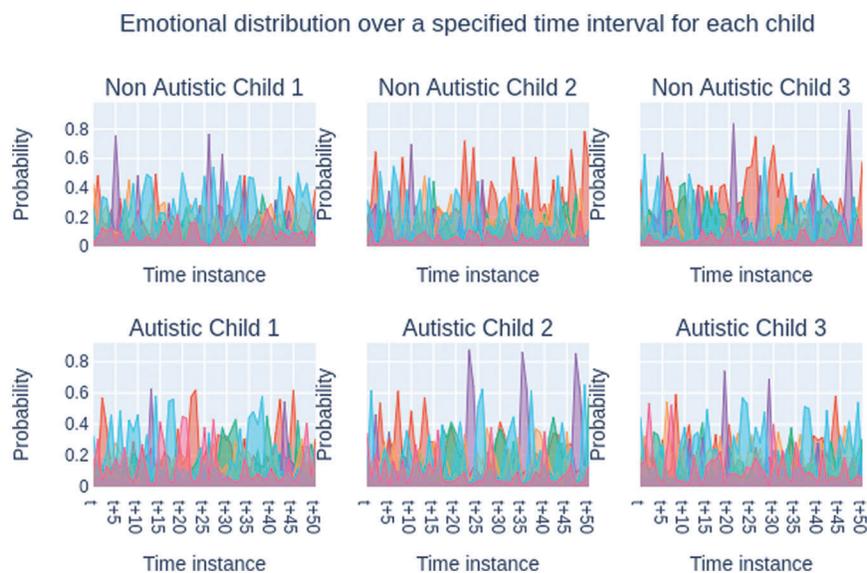


Figure 6: Probability distribution of emotion in the observed autistic and non-autistic children

Tab. 3 is an inference for the overall probabilistic variations among the seven identified facial expressions. The table insights that for every facial expression there could be minimal differentiated expression from the children. This inference helps in analyzing the performance of facial expression analysis and to provide appropriate advancements in detection. This implicates that ASD varies between “neutral”, “disgust” and “anger” as maximum distracted expressions while normal child shows variations revolving maximum “happiness” and “neutral”. Disgust in non autistic is observed only under maximum accuracy and not under any fall in probabilistic facial expression value.

Table 3: Probability distribution across various facial expressions in ASD +ve and ASD –ve

Category	Facial expression	Anger	Disgust	Fear	Happy	Sadness	Neutral	Sleep
ASD +ve	Anger	0	0.15923	0.095921	0.020293	0.14268	0.13714	0.067018
	Disgust	0.15923	0	0.04258	0.09563	0.06866	0.04987	0.0824
	Fear	0.095921	0.04258	0	0.04562	0.041347	0.022793	0.0679
	Happy	0.020293	0.09563	0.04562	0	0.026975	0.168421	0.087725
	Sadness	0.14268	0.06866	0.041347	0.026975	0	0.021446	0.03925
	Neutral	0.13714	0.04987	0.022793	0.168421	0.021446	0	0.1247
	Sleep	0.067018	0.0824	0.0679	0.087725	0.03925	0.12474	0

(Continued)

Table 3 (continued)

Category	Facial expression	Anger	Disgust	Fear	Happy	Sadness	Neutral	Sleep
ASD -ve	Anger	0	0.18706	0.041775	0.176547	0.082745	0.126316	0.031966
	Disgust	0.18706	0	0.15813	0.00338	0.12716	0.03359	0.19794
	Fear	0.041775	0.15813	0	0.09475	0.020977	0.094541	0.06981
	Happy	0.176547	0.00338	0.09475	0	0.05998	0.04351	0.17456
	Sadness	0.082745	0.12716	0.020977	0.05998	0	0.04587	0.10078
	Neutral	0.126316	0.03359	0.094541	0.04351	0.04587	0	0.14735
	Sleep	0.031966	0.19794	0.06981	0.17456	0.10078	0.14735	0

5.1 Evaluation of the Model

On pre-processing the acquired data set, the number of images under analysis is reduced. In order to ensure the maximum accuracy of the proposed model, the contribution of the k-fold cross-validation technique is adopted. However, out of the entire dataset images, the FER system separated 80% of images under training dataset and 20% of the images. The 5 fold cross-validation process was incorporated such that, out of the sliced blocks of data, one slice was used for testing the data under every iteration. Similarly, the training data were used to test in either of the epochs. This cross-validation method act as a trial method to maximize the training dataset that results in increased accuracy.

Though the number of frames involved in training the ASD positive case stands to be minimum, the dataset is boosted by its volume on applying the k-fold method. Also, the exploration of facial expressions stands to be near accurate to the processing capacity. Through the analysis, it is clearly inferred that the majority of the facial expression shown by either group of children is computed through the summation of prioritized facial expressions shown by every individual.

5.2 Accuracy Comparison of the Proposed Models

The accuracy of the models are evaluated based upon the Eq. (9)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

The accuracy of emotion analysis in children with ASD is evaluated across the training and test data performance. The emotion analysis undergoes series of iterations in linear prediction from the feedback of the previous layer and the following accuracy was observed as in Fig. 7.

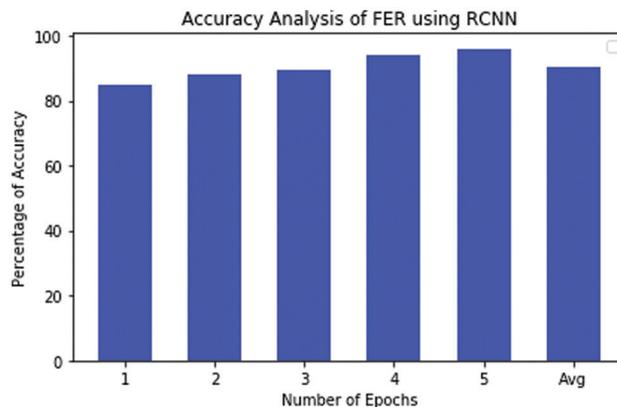


Figure 7: Accuracy analysis of the proposed classifier over epochs

On the first epoch, the accuracy factor is determined as 0.8462 and in the upcoming iterations of the linear classifier, the accuracy factors were determined as 0.8784, 0.8937, 0.937826, and 0.954157 respectively. On average, the accuracy of the analysis is found to be approximately 91% in the fifth epoch. The number of iterations is halted owing to the computational complexity involved in executing the neural network. The accuracy could be further improved by increasing the data set size and training the classifier with much more varied and deeper feature combinations.

The following Fig. 8 represents the accuracy comparison of the proposed models for FER systems namely SVM, CNN, and R-CNN at a specified time t . Though there is no change in the probability of the facial expression compared over the models, there is a strong advent to the influence of the facial expression identified within the ASD +ve and ASD -ve children. These models could be used as an early biomarker of autism identification and exploration of facial expression in them.

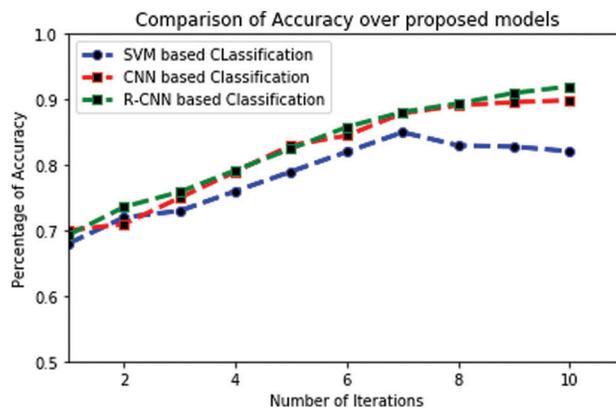


Figure 8: Comparison of accuracy on R-CNN over existing systems

6 Conclusion and Future Scope

The integrated R-CNN-based FEA system identifies the emotion of the children involving time variants using live video input. The system proves better improvement in expression analysis and also suggests the point of time when there exists a change in expression. The system indicates that the maximum expression perceived by autistic children is disgust, neutral, and anger. The fall of expression over time for autistic children is also into the perceived expression which supports the identification of repetitive behavior in autistic children. The system proves better in the accuracy of 91% when compared to the other existing models for facial expression evaluation. Also, the major contribution of this proposed system is to identify the autistic character in the child at an early stage by direct evaluation of the child's behaviour rather through questionnaires. This pattern of evaluated expression incurred in every child could also be inferred supporting early autism identification. The future scope of the paper lines in altering the family members prior to the emotional reaction of the children with proper accompanied IoT systems. This real-time intelligent surveillance can improve in the development stages of the child before falling into an aggression and hatred state.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] N. L. Romero, "A pilot study examining a computer-based intervention to improve recognition and understanding of emotions in young children with communication and social deficits," *Research in Developmental Disabilities*, vol. 65, no. Suppl.1, pp. 35–45, 2017.
- [2] D. Bone, S. L. Bishop, M. S. Goodwin, C. Lord and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion," *Journal of Child Psychology and Psychiatry*, vol. 57, no. 8, pp. 927–937, 2016.
- [3] J. Duvekot, L. Van Der Ende, F. C. Verhulst, G. Slappendel, A. Mara *et al.*, "Factors influencing the probability of a diagnosis of autism spectrum disorder in girls versus boys," *Journal of Autism*, vol. 21, no. 6, pp. 646–658, 2017.
- [4] S. D. Stagg, R. Slavny, C. Hand, A. Cardoso and P. Smith, "Does facial expressivity count? How typically developing children respond initially to children with autism?," *Journal of Autism*, vol. 18, no. 6, pp. 704–711, 2014.
- [5] M. Del Coco, M. Leo, P. Carcagni, P. Spagnolo, P. Luigi Mazzeo *et al.*, "A computer vision based approach for understanding emotional involvements in children with autism spectrum disorders," in *Proc. of the IEEE Int. Conf. on Computer Vision Workshops*, Venice, Italy, pp. 1401–1407, 2016.
- [6] L. Van Eylen, B. Boets, J. Steyaert, J. Wagemans and I. Noens, "Local and global visual processing in autism spectrum disorders: Influence of task and sample characteristics and relation to symptom severity," *Journal of Autism and Developmental Disorders*, vol. 48, no. 4, pp. 1359–1381, 2018.
- [7] D. P. Kennedy and R. Adolphs, "Perception of emotions from facial expressions in high-functioning adults with autism," *Journal of Neuropsychologia*, vol. 50, no. 14, pp. 3313–3319, 2018.
- [8] T. Guha, Z. Yang, R. B. Grossman and S. S. Narayanan, "A computational study of expressive facial dynamics in children with autism," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 14–20, 2016.
- [9] L. R. Whitaker, A. Simpson and D. Roberson, "Brief report: Is impaired classification of subtle facial expressions in children with autism spectrum disorders related to atypical emotion category boundaries," *Journal of Autism and Developmental Disorders*, vol. 47, no. 8, pp. 2628–2634, 2017.
- [10] S. Wang and R. Adolphs, "Reduced specificity in emotion judgment in people with autism spectrum disorder," *Journal of Neuropsychologia*, vol. 99, pp. 286–295, 2017.
- [11] N. P. Gopalan, S. Bellamkonda and V. S. Chaitanya, "Facial expression recognition using geometric landmark points and convolutional neural networks," in *Int. Conf. on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, pp. 1149–1153, 2018.
- [12] J. Guha, X. Yang, S. De Mello and J. Kautz, "Dynamic facial analysis: From bayesian filtering to recurrent neural network," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 1548–1557, 2017.
- [13] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognition Letters*, vol. 115, no. 2, pp. 101–106, 2018.
- [14] S. Minaee, M. Minaei and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Journal of Sensors*, vol. 21, no. 9, pp. 3046, 2021.
- [15] B. Jan, H. Farman, M. Khan, M. Imran, I. U. Islam *et al.*, "Deep learning in big data analytics: A comparative study," *Journal of Computers & Electrical Engineering*, vol. 75, pp. 275–287, 2019.
- [16] Y. Song and Y. Hakoda, "Selective impairment of basic emotion recognition in people with autism: Discrimination thresholds for recognition of facial expressions of varying intensities," *Journal of Autism and Developmental Disorders*, vol. 48, no. 6, pp. 1886–1894, 2018.
- [17] A. Prehn-Kristensen, A. Lorenzen, F. Grabe and L. Baving, "Negative emotional face perception is diminished on a very early level of processing in autism spectrum disorder," *Social Neuroscience*, vol. 12, no. 2, pp. 191–194, 2019.
- [18] S. P. Abirami, G. Kousalya and R. Karthick, "Varied expression analysis of children with ASD using multimodal deep learning technique," *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, vol. 21, no. 6, pp. 225–243, 2019.
- [19] S. P. Abirami, G. Kousalya and R. Karthick, "Identification and exploration of facial expression in children with ASD in a contact less environment," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 3, pp. 2033–2042, 2019.