Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.023445
Article

Tech Science Press

Facial Action Coding and Hybrid Deep Learning Architectures for Autism Detection

A. Saranya[1,*] and R. Anandan[2]

[1]Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai, Tamilnadu, 603203, India
[2]Department of CSE, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamilnadu, 603203, India
*Corresponding Author: A. Saranya. Email: saranyajournalssaro1@gmail.com

Received: 08 September 2021; Accepted: 18 November 2021

Abstract: Hereditary Autism Spectrum Disorder (ASD) is a neuron disorder that affects a person's ability for communication, interaction, and also behaviors. Diagnostics of autism are available throughout all stages of life, from infancy through adolescence and adulthood. Facial Emotions detection is considered to be the most parameter for the detection of Autismdisorders among the different categories of people. Propelled with a machine and deep learning algorithms, detection of autism disorder using facial emotions has reached a new dimension and has even been considered as the precautionary warning system for caregivers. Since Facial emotions are limited to only seven expressions, detection of ASD using facial emotions needs improvisation in terms of accurate detection and diagnosis. In this paper, we empirically relate the facial emotions to the ASD using the Facial Action Coding Systems (FACS) in which the different features are extracted by the FACS systems. For feature extraction, DEEPFACENET uses the FACS integrated Convolutional Neural Network (FACS-CNN) and hybrid Deep Learning of LSTM (Long Short-Term Memory) for the classification and detection of autism spectrum disorders (ASD). The experimentation is carried out using AFFECTNET databases and validated using Kaggle Autistic facial datasets (KAFD-2020). The Multi-Layer Perceptron (48.67%), Convolutional neural networks (67.75%), and Long ShortTerm Memory (71.56), the suggested model showed a considerable increase in recognition rate (92%), from this proposed model prove its superiority in detecting autistic facial emotions among children effectively.

Keywords: Autism spectrum disorder; facial emotions; facial action coding systems; convolutional neural networks; LSTM

1 Introduction

Autism is a hereditary illness that manifests itself in aberrant behaviors, nonverbal communication, and repeated speaking. With all of these traits, Autism can be classified as ASD is a condition that affects people of all ages. Furthermore, the Center for Disease Control and Prevention has said that the rate of autism is expanding exponentially and that it now affects 2% of the population in the United States. Autism can be recognized in children as young as two to three years old when they change their behaviors and isolate themselves from normal life. Unnoticed children may result in a meltdown crisis, which is the most deadly. As a result, early recognition of ASD is necessary to address this issue among children.

Machine learning algorithms are being used to predict ASD using many input techniques, including Gene Expressions [1], Behavioral observation [2], metabolic control rates [3], and home movies [4]. More study was done on the application of machine and deep learning algorithms to detect autism before it is too late [5]. In impoverished countries like India, where clinical resources are limited, this study is especially useful for children with ASD in evaluating and treating them more rapidly [6]. T SVM [7], Random Support Forest [8] and Logistic Regression [9] have been used for the early identification of ASD in children.

Because of their anomalous unique alterations, a facial recognition system is the best feasible technique to diagnose a patient. Early signs of Autism can be seen in children's faces, which are characterized by unusually large upper faces and wide-set eyes. Although facial expressions can convey autism symptoms, researchers continue to experience a significant issue in distinguishing between normal and autistic faces.

Early detection of ASD via facial expressions has taken on a new level because of the above-mentioned machine and deep learning techniques. To be effective, detection must be improvised in terms of both accuracy and complexity. CNN and Long Short-Term Memory (LSTM) [10] are combined in this research to create unique hybrid DEEP FACE networks that can identify autism (LSTM). Furthermore, a Facial Action Coding Systems (FACS) based CNN (FACS-CNN) has been developed to extract distinct aspects from normal and autism-affected youngsters.

The following is the research's contribution:

1. A unique FACS-trained CNN has been presented to improve the rate of Autism Spectrum Disorder detection among children.
2. It has been proposed that LSTM be included in planned CNN networks. This integration in DEEPFACENETS must improve recognition rates not just for static but also for dynamic image sets.
3. The suggested technique is evaluated utilizing AFFECTNET datasets and validated using Kaggle Autistic Facial Datasets-2020, which has been compared to previous ASD detection models.

There are three parts to the work that is: Various authors have proposed various models of intelligent learning, which are discussed in Section 2. Convolutional Neural Networks, Long Short Term Memory (LSTM), and the suggested DEEPFACENETS operating concept are discussed in Section 3. Descriptions of datasets, experimental setup, and results with comparative analysis are discussed in Section 4. Summing it all up with a future perspective, the paper concludes in Section 5.

2 Related Works

DadangEman provides an overview of ASD using controlled learning and ML (Machine Learning calculation). According to the site's analysis, the objects were in the same condition as indicated in this post. The articles were acquired from web databases, and 16 exploration articles satisfied the investigation's requirements. Based on the results, the support vector machine is the most widely used calculation in the writing focus in this study, accounting for 68.75%. With the use of machine learning on behalf of ASD, it should be able to speed up and increase the precision of their analytical decisions. By enhancing the precision and speed of conclusion to speed up the process of making judgments in the management of patients as soon as feasible so that patients receive faster treatment, ML can distinguish a condition, like diagnosing someone with a chemical imbalance [11].

Madison Beary uses Deep Learning to train a DL model that can classify children as either healthy or psychologically unstable with 94.6 percent accuracy. Social abilities, dull activities, and verbal and nonverbal correspondence are all battlegrounds for mentally imbalanced patients. Despite the fact that the virus is regarded as genetic, the most significant rates of precise conclusions occur when the child is tested on social characteristics and facial features.

Patients have a common example of specific facial deformations, allowing professionals to analyze simply a picture of the child to determine whether or not the child has the sickness. While distinct tactics and models are used for face inquiry and chemical imbalance order on their own, this work links these two ideas, allowing for a less expensive, more effective strategy. To accomplish highlight extraction and picture layout, this structure profound learning model employs MobileNet and two thick layers. The model is created and tested using 3,014 photographs that are evenly distributed between children with and without a chemical imbalance. 90% of the information is used for planning, and 10% is used for testing. Because of its precision, this technique is capable of detecting mental instability with only one image. There may also be different ailments that can be diagnosed in this way [12].

During a Meltdown emergency, Salma KammounJarraya provided another approach to deal with distinct micro articulations for mentally unbalanced kids' compound sensations. This research was used to find out how to create a smart "Sensor Autistic Meltdown" to ensure the safety of mentally unstable children. This method was tested on a collection of recordings captured using a Kinect camera, and it was found to have several flaws. The quantitative and subjective findings demonstrated how this method produces a powerful model. Since then, this structure has used highlights selection strategies such as channel approaches RelifF calculation and Information Gain Algorithm, Wrapper Methods, and NCA strategy to investigate the significance of the supplied highlights. It was decided to analyze the effects of highlight selection strategy by using a few order computations such as FF and CFF, as well as RNN (Recurrent Neural Network) and the LSTM. RNN Classifier (85.5%) and Information Gain highlight selection technique have been shown to improve order precision with this strategy. Using only five features, the proposed technique is superior [13].

RBM is used to extract highlights from the fMRI data and SVM is used to discriminate ASD subjects from solid controls in Md Rishad Ahmed's model. As a first step, this structure accomplishes certain standardization and cut time revision operations. This method was tested on a dataset that included 105 typical control and 79 ASD participants from the prominent ABIDE database. The results reveal that using network search cross-approval, the suggested approach works very well in grouping ASD and achieves 80% accuracy. The discoveries likewise show that the blend of RBM (Restricted Boltzmann Machine) and SVM (Support Vector Machine) strategies might be applied as a future apparatus to analyze ASD [14].

The exploratory outcomes from the suggested work execution assessment are discussed, taking into account each Autism Patient and the feeling names separately. In comparison to other classifiers, the proposed study has shown test results that can accurately identify feelings. This work accomplishes a 6% preferable precision for the Proposed Model over the Support Vector machine and 8% more exactness than back Propagation calculation [15].

Marco Leo proposed an innovative system to explore how both Autism and TD children construct looks computationally. The proposed pipeline was used to evaluate the ability to create key sentiments. This ability was evaluated while it was beginning to be acquired in normally developing children and when it was a lack in ASD children. The pipeline is precise (more than existing approaches), quick, and objective in assessing both the strength of looks and how much each facial part is related to the look, according to mathematical results. The reference standard comprised of the manual comments made by a group of clinicians. A limit of the current investigation is the example size [16–18].

The Internet of Things edge figure and multi-headed 1-dimensional convolutional neural organization (1D-CNN) model developed by Ramendra Pathak allows for continuous monitoring and categorization of facial expressions into happy, weeping, and resting categories. For example, IoT edge gadgets govern computing capacity on a local basis, which promotes data security and decreases inertia, as well as lowering data transfer speed costs. The proposed approach is presented and contrasted with the ML models in terms of exactness, review, and f1-score.

The proposed approach beats all ML models in all classes (glad, crying, and resting) [19].

Lakshmi Praveena proposed that streamlined DL (Deep Learning) techniques be used to predict ASD in children aged 1 to 10 years. The proposed model is tested on a dataset obtained from the chemical imbalance guardians' Facebook group, as well as on a dataset gathered from ASD children and typical children's facial photo datasets gathered from Kaggle datasets. Enhancement approaches, dropout, clump standardization, and boundary updating are used to apply Convolutional Neural Networks (CNN) to extricated face milestones. To accurately predict ASD children, the most significant six types of feelings are examined. The proposed approach appears to be more viable and stable. The approach can also be used to differentiate distinct highlights and characteristics in chemical imbalance children's facial photographs, such as activity units, enthusiasm, and valence [20–22].

As Omar Shahid explains, information-driven movement assessment can be used to identify chemical imbalance robotically. Structures such as gloomy conduct, abnormal step designs, and surprising visual saliency were used to test cutting-edge information-driven approaches for ASD recognition through active investigation. Additionally, this article examined the results of various ML and DL computations in the context of recognizing ASD and TDs. In addition, this work has briefly discussed possible challenges with expected arrangements, available assets, and optimal test setups. Discoveries show this innovation is superior to the conventional clinical examination of ASD diagnosis procedures, which normally takes a long time with little assurance that the help will be accessible to the general public. As a result, robotized identification's precision and flexibility may be limited by a few limitations.

The research of Mohamed A. Saleh focused on cutting-edge ML and DL and their application in feeling recognition. It also presented a second mental imbalance restoration treatment paradigm based on the DL. Both the DL methodology and a humanoid robot are employed in training youngsters with ASD to recognize sensations. Open CV and Caffe-based Deep Learning (DL) models are applied to a GPU using a Convolutional Neural Network (CNN) model. Humanoid-assisted restoration treatment for mentally imbalanced children in the proposed framework is coordinated with DL articulation discovery. Children with ASD can benefit from robot-based social treatment if they have access to a reliable dataset and the proposed DL model. Although new deep learning advancements for look acknowledgment calculations could lead to enhanced discovery precision, it's evident that only a large and dependable specialized dataset can offer sufficient results.

3 Proposed Framework

The proposed learning model for ASD identification using facial expressions is shown in Fig. 1. The first layer is FACS-CNN in which the convolutional layers are used to extract the different features of faces such as head, nose, eyebrows, chin, etc.,

3.1 System Overview

The facial action coding engine is used to calculate the facial features using different thresholds for effective categorization of ASD faces and ASD faces. These FACS based values are then used to train the Long ShortTerm Memory (LSTM) for ASD detection. The working mechanism of each layer is discussed in the preceding section.

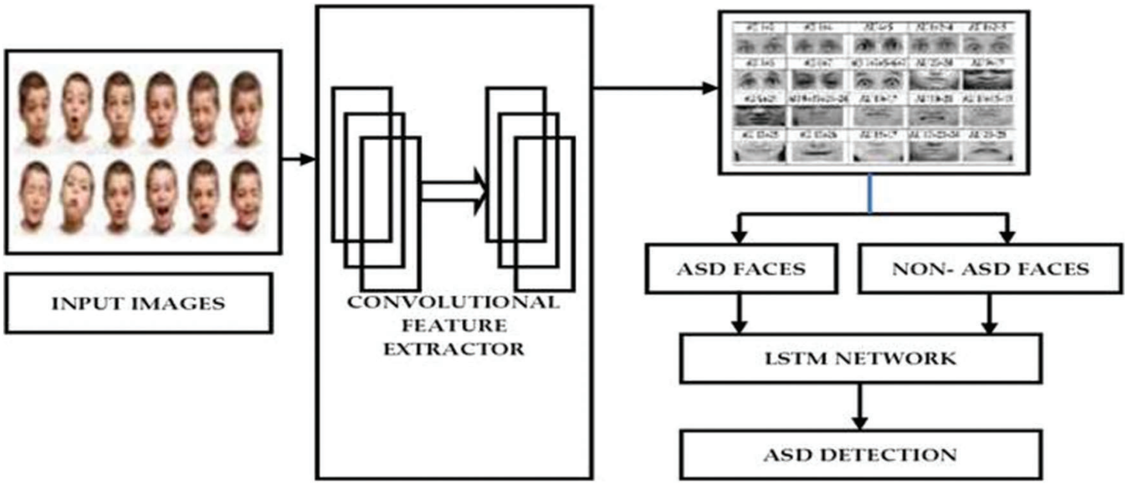

Figure 1: Proposed architecture for the deep face networks

3.2 Deep Face Nets-Its Working Model

An overview of convolutional networks is presented in this section (CNN). Deepfacenets with LSTM.

Convolutional neural networks (CNN): CNNs are a sort of deep learning model, but they also fall under the umbrella of Artificial Neural Networks (ANNs), which are used in image processing [23] and video analytics. There are five layers in the CNN model. The input layer consist of a matrix of the normalized patterns and feature maps are used to connect inputs with its previous layers. The features obtained by the convolutional layer are used as the inputs to pooling layers. These features are used for training the network to obtain the required output. The CNN network has been demonstrated in Fig. 2.

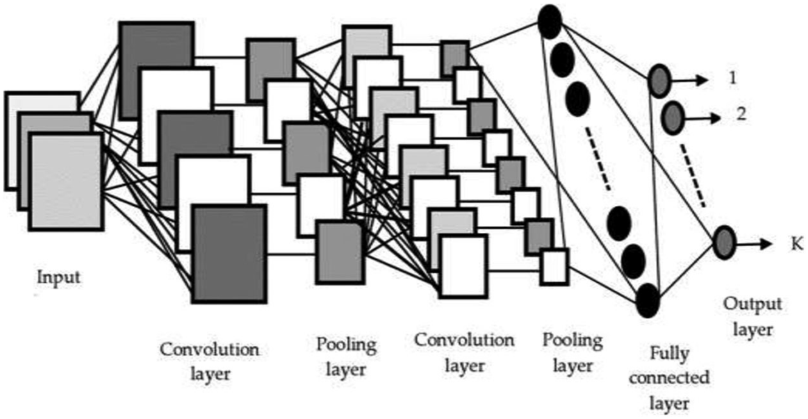

Figure 2: Convolutional neural networks-a general overview

3.3 Long Short Term Memory-LSTM

A Long Short-Term Memory network is a popular learning model utilized for several applications due to its flexibility in memory and more appropriate for a huge database. The LSTM network has been demonstrated in Fig. 3.

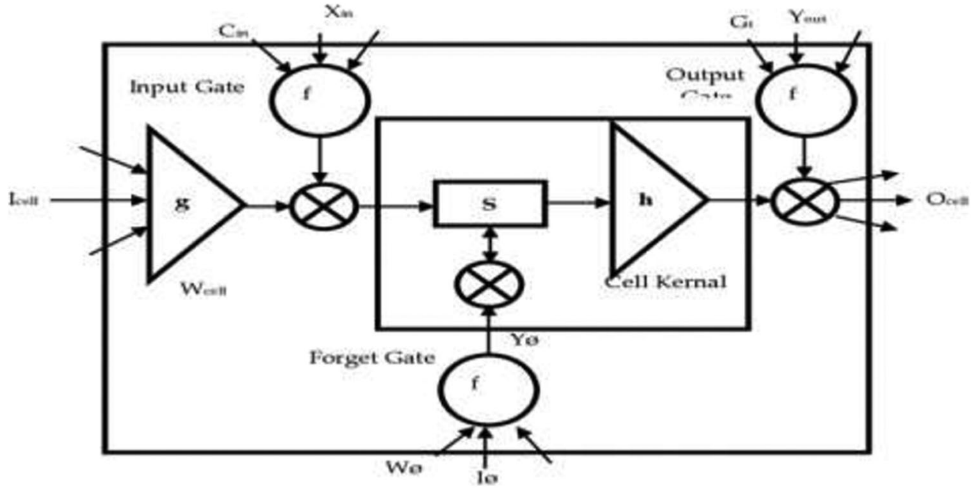

Figure 3: LSTM structure

In a hybrid learning model, there is an LSTM and a Whale optimizer, respectively. Four separate blocks are used to construct an LST. These blocks are named the Input Gate I, Forget Gate F, Cell Input C, and Output Gate (O.G). Generally, LSTM is a memory-based neural network which to remembers the values after every iteration. Let xt, the unseen layer output is 'ht' and its former output is 'ht−1', the cell input state is 'Ct', the cell output state is 'Gt' and its former state is 'Gt-1', the three gates' states are j_t, T_f and T0.

The formation of LSTM resembles that both "Gt and ht" are communicated to the next neural network in RNN. LSTM merges the output of the previous unit with the current input state in which the output and forget gates are used to update the memory. To calculate Gt and ht, we use the following equations.

$$I.G: \quad j_t = \theta(G_l^i.\ O_t + G_h^i.e_{t-1} + s_i) \tag{1}$$

$$F.G: \quad T_f = \theta(G_l^f.O_t + G_h^f.e_{t-1} + s_f) \tag{2}$$

$$O.G: \quad T_o = \theta\ (G_l^0.O_t + G_h^o.e_{t-1} + s_o) \tag{3}$$

$$C.I: \quad \widetilde{T_C} = tanh(G_l^C.O_t + G_h^C.e_{t-1} + s_C) \tag{4}$$

where G_l^0, G_l^f, G_l^i, G_l^C represents the weight matrices between input gates and output layers and G_h^i, G_h^f, G_h^o, G_h^C denotes the weight conditions generated between hidden and input layers. The "s_i, s_f, s_o, s_c are the bias vectors and tanh is considered to be hyperbolic function". The cell output state is calculated and it is given as follows as

$$T_C = k_t * \ \widetilde{T_C} + T_f * T_{t-1} \tag{5}$$

$$e_t = T_o * tanh(T_C) \tag{6}$$

The final output score is obtained using the above equation.

3.4 Proposed DeepFace Nets

So that we may get better results from the model, we try to merge CNN with FACS Feature maps and LST models. Initial training is done with a large number of static images that show both autistic and non-autistic facial emotions. The last layer is then removed and flattened into single-dimensional feature maps, which are then compared with Facial Action Coding Systems (FACS) to form highly accurate feature maps.

Fig. 4 shows the complete working of the proposed learning model. The functioning of the proposed FACS-CNN is detailed as follows:

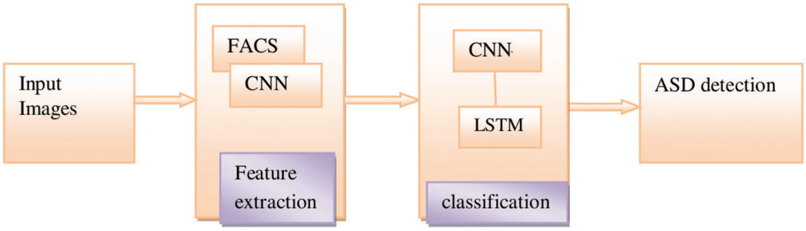

Figure 4: Proposed framework for deep face nets

3.5 Algorithm for FACS

FACS is a face expression index, however, it does not provide any biomechanical information on muscle activation levels. Even though muscle activation is not part of FACS, the key muscles involved in face expression have been included for the reader's convenience. The fundamental activities of individual muscles or groups of muscles are referred to as action units (AUs). Action descriptors (ADs) are a type of unitary movement that might involve numerous muscle groups (for example, a forward-thrusting movement of the mouth). The muscle foundation for these motions hasn't been identified, and specific behaviors haven't been defined as precisely as they have for AUs.

The FACS-based graph's algorithm shows how the characteristics are computed. Similarly, the algorithm computes two feature elements for each segment, and this process is repeated for all of the FACS-based graph segments to compute all of the features.

Algorithm: Feature computation with FACS

```
Data: $\varphi_{i_1,i_{2,}}$ where i→ ∈{1, 2, 3….$N$}
Result: $f_k$, where k∈ 1, 2, 3, …$M$
For k← 1 $to\ Mby1\ do$
F(k) = $\sqrt{(\varphi_{x,i_1} - \varphi_{x,i_2})^2 + (\varphi_{y,i_1} - \varphi_{y,i_2})^2}$, k← $k+1$
F(k) = $\frac{\varphi_{x,i_1} - \varphi_{x,i_2}}{\varphi_{y,i_1} - \varphi_{y,i_2}}$, k← $k+1$
End
```

3.6 FACS-CNN Feature Extractor

For better recognition, time-series features of normal and autistic children are calculated using the FACIAL Action Coding Systems (FACS). Originally developed by Carl_Herman Hjorstjo and later adopted by Paul Ekman, FACS is a system that is used to classify human expressions based on the movement of individual facial muscles and finds its significant role in categorizing the expression of emotions. The proposed architecture employs the FACS engine to distinguish the difference between autistic and non-autistic emotions. To extract the time series FACS features, facial expressions of autistic children under ASD is shown in Fig. 5

In FACS, specific action units for autistic facial emotions (AU) are used to code the anatomically possible facial expressions. When it comes to muscular contractions and relaxations, FACS defines AUs in. They can be employed in high-level decision-making processes such as face emotion recognition because they are independent of any interpretation.

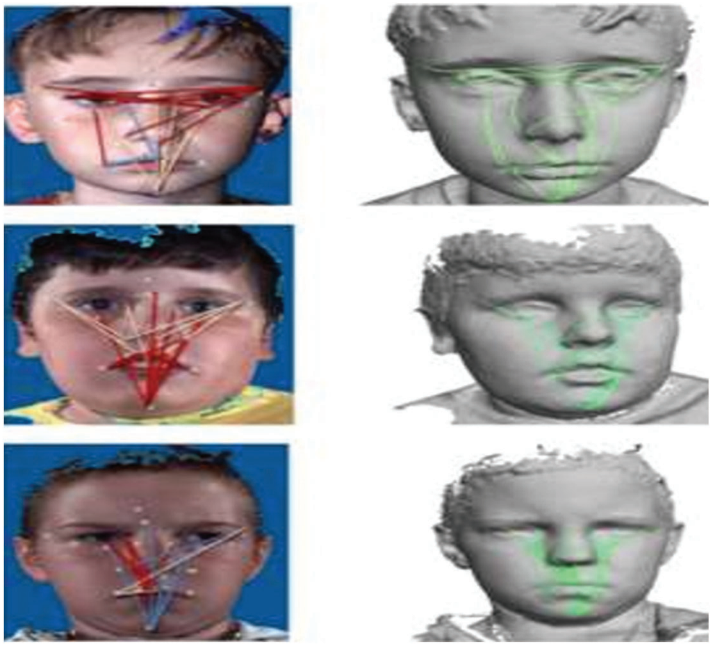

Figure 5: Different autistic facial emotions for the children with early ASD

The proposed CNN architecture consists of 10 convolutional layers, as well as batch normalization (normalizes input channel concerning to the mini-batch made), FACS (performs threshold operation and returns 0 if the value is less than 0), average pooling (calculates the average of each patch), and global average pooling layers. Softmax, on the other hand, calculates error and ends with a classification layer. As you progress through the network, the number of filters grows. A dropout layer with a probability of 0.5 means that some neurons are randomly dropped out. The parameters listed in Tab. 1 are trained in the first layer of CNN, and the data is pooled before being sent to the fully connected layer for feature extraction for classification.

Table 1: Feature extraction methodology using FACS mechanism

Permanent features (Left and right)			Other features
rbinner-motion of inner brow	rbouter-motion of outer brow	reheight-height of eye	Dbrow-brows distance
rbinner = $\frac{bi-bi_0}{bi_0}$ If rbinner > 0, Inner brow up	rbouter = $\frac{bo-bo_0}{bo_0}$ If rbouter > 0, Outer brow up	reheight = $\frac{(h1+h2)-(h1_o-h2_o)}{(h1_o+h2_o)}$ If reheight > 0, Increase in eye height	Dbrow = $\frac{D-D_0}{D_0}$ If dbrow < 0 Two brows drawn together
rtop-eye top lid motion	rbtm-eye bottom lid motion	rcheek-cheek motion	Wleft/right-crows-feet wrinkles
rtop = $\frac{h1-h1_0}{h1_0}$ If rtop > 0, Eye top lid	rbtm = $\frac{h2-h2_0}{h2_0}$ If rtop > 0, Eye bottom lid	rcheek = $\frac{c-c_0}{c_0}$ If rcheek > 0, Move up of cheek	If Wleft/right = 1, Presence of left/right crow's feet wrinkle

Tab. 1 illustrates the different features to be extracted from the different faces using the different mathematical expressions. With the help of mathematical expressions, new AU units have been formulated for autistic and normal facial emotions. The significance of using FACS in determining ASD will increase the recognition rate in ASD detection. Tab. 2 Illustrates the different AU units for both normal and autistic facial emotions.

Table 2: Different AU units formed with the normal and autistic facial emotions

Sl. no	Normal facial emotions AU units*	Autistic facial emotions (AU)#
1	Happiness	Autistic Happiness (AH) = 1
2	Anger	Autistic Anger (AA) = 2
3	Disgust	Autistic Disgust (AD) = 3
4	Fear	Autistic Fear (AF) = 4
5	Anxiety	Autistic Anxiety (AAA) = 5
6	Surprise	Autistic Surprise (AS) = 6
7	Sadness	Autistic Sadness (ASA) = 7

Note: *-Normal facial emotions use the AU codes which are mentioned in the FACS manual
#-Newly formed AU codes to distinguish the autistic facial emotions.

The newly formed AU codes are used for labeling the data feature which is used to distinguish effectively between the normal and autistic facial emotions. To detect the ASD using the FACS features, convolutional layers are formed for an input image as 128 × 128 facial images and provides the outputs of either, this phase of training accepts the input images as 128 × 128 facial images and provides the outputs such as segmented eyebrows, foreheads, chin, nose, etc. The filter layers are maintained at 2 × 2 and pooling layers are also maintained at 3 × 3. There is a normalizing layer after each activation layer in order to improve accuracy. As a result, the Dropout is kept at 0.4 to prevent the training model from being too well-suited. Consequentially, in the convolution-pooling group, there are four layers: Convolution and Activation; Batch Normalization; Pooling and Dropout. The hyperparameters used for the training the facial emotions are tabulated in Tab. 3.

Table 3: Hyperparameters for FACS-CNN were used for training the model using the FACS features

Layers	Output layer	Filter size/stride length
Input layer	128 × 128 × 10	16(2 × 2)/1
Convolution1	128 × 128 × 10	16(2 × 2)/1
Max-pooling	64 × 64 × 5	—
Convolution 2	64 × 64 × 5	16(2 × 2)/1
Max-pooling	32 × 32 × 5	———
Convolution-3	32 × 32 × 5	8(2 × 2)/1
Max-pooling	16 × 16 × 5	———
Convolution 3	16 × 16 × 5	8(2 × 2)/1
Max-pooling layer	8 × 8 × 2	——
Activation (layers)	ReLU	———
Optimizer	Adam	———

These feature maps are also imported into LSTM as time series sequences of data. The last layer of the CNN is removed and then connected to LSTM for further processing. For accurate detection of Facial Autistic Emotion Detection, the input uses an image from FACS features, which makes the proposed algorithm rectify the problem of detection error that occurs due to the different facial points from similar subjects. Accordingly, the proposed architecture uses CNN with high recognition rate FACS as a feature extractor. In dynamic time-sensitive images, the most important extracted features can be inputted into the LSTM through the FACS-CNN feature extractor. This can enhance the recognition success ratio when the different categories of autistic and non-autistic images are given.

4 Dataset Descriptions

The performance of the new model is proved by the results of cross-validation comparing current models using the leave-one-out method. AFFECTNET datasets and Kaggle Facial Autistic Datasets-2019 are used in the tests. There are 450,000 human-annotated and 500,000 automatically annotated photos of various sizes in the AFFECTNET databases. They are described as neutral, pleased, afraid, disgusted, angry, and contemptible, among other emotions mention that because the collection is significantly unbalanced, we randomly chose about 50,000 photos with basic emotions.

Also, we have validated the proposed model using the Kaggle Facial Autistic Datasets-2020 which consist of the *1667 facial images of Autistic children in 224 × 224 × 3, jpg format and 1667 facial* images of normal children also in 224 × 224 × 3, jpg format.

4.1 Performance Metrics

The proposed learning models have been implemented using Tensorflow-Keras API which runs on an i7 CPU with 2TB HDD, 4GB NVDIA Geoforce GPU and 8 GB RAM. The following metrics were used for calculating the performance of the proposed algorithm is mentioned below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

where,

TN-True classified negative samples, *TP*-True positive samples, *FP*-False classified positives, *FN*-False classified negatives. To verify the performance of the proposed model using the above parameter metrics, we have used the leave-out cross-validation process to evaluate the proposed model. Then we have compared with the other existing learning models such as. For the evaluation, the settings, structures and number of parameters of the proposed model are tabulated in Tab. 4.

Tab. 5 represents the performance metrics for the proposed networks using the different number of epochs at the learning rate is 0.001. It is found from the Table, validation loss is equal to 0.001 between the training and testing process which proves the proposed model has good performance in detecting the six facial emotions from the normal faces. Also, the proposed model has been validated with the Kaggle Facial Autistic Datasets (KFAD-2020) to determine the ASD from facial emotions.

Table 4: Hyperparameters tuned for the deep face nets models used for facial feature extraction

Layers	Output layer	Filter size/stride length
Input layer	128 × 128 × 10	16(2 × 2)/1
Convolution 1	128 × 128 × 10	16(2 × 2)/1
Max-pooling	64 × 64 × 5	—
Convolution 2	64 × 64 × 5	16(2 × 2)/1
Max-pooling	32 × 32 × 5	———
Convolution-3	32 × 32 × 5	8(2 × 2)/1
Max-pooling	16 × 16 × 5	———
Convolution 3,	16 × 16 × 5	8(2 × 2)/1
MaxPooling layer	8 × 8 × 2	——
Activation (layers)	ReLU	———
Optimizer, No of dense layers	Adam, 06	———, 700
LSTM cells	256	590080, 131483

Table 5: Performance metrics for the proposed algorithms using affect net datasets to detect the normal facial emotions

No of epochs	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
	AFFECTNET datasets	Accuracy	Precision	Recall	Accuracy	Precision	Recall
20		91.0%	91.0%	91.0%	90.89%	91.0%	90.80%
40		92.2%	91.4%	91.2%	92.15%	91.35%	91.3%
60		92.32%	91.45%	91.25%	92.29%	91.43%	91.20%
80		92.40%	91.5%	91.4%	92.35%	91.29%	91.30%
100		92.50%	91.5%	91.4%	92.45%	91.45%	91.35%
120		92.50%	91.5%	91.4%	92.45%	91.45%	91.35%
140		92.5%	91.5%	91.4%	92.45%	91.45%	91.35%

Figs. 6 and 7 show the performance metrics of the Proposed DEEPFACENET networks to detect the normal facial emotions and autistic facial emotions using Kaggle Facial Autistic Datasets (KFAD-2020). From Fig. 8, it is found that the proposed model has shown 92% accuracy in detecting the normal faces and 92.5% in detecting the autistic facial emotions from the epochs of 80-100. Moreover, to prove the superiority of the proposed algorithm, we have calculated leave-out cross-validation values and compared them with other existing learning models such as Multi-Layer Perceptron. Convolutional Networks (CNN) and Long Short Term Memory (LSTM) with the same settings and parameters which has been adopted for the proposed model. Tab. 6 shows the settings for the other learning models which is used for the comprehensive analysis, moreover, the hyperparameters tuned for the convolutional neural networks (CNN) models used for facial feature extraction are tabulated in Tab. 7.

Figure 6: Facial emotion in the AFFETNET datasets used for training the proposed framework

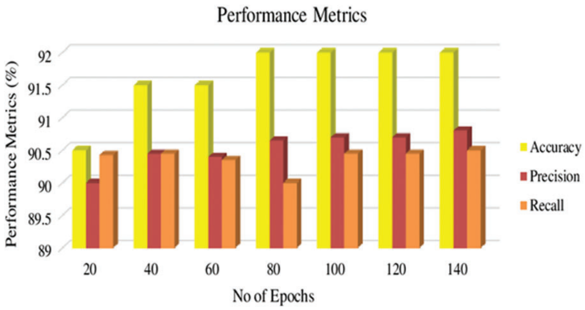

Figure 7: Performance metrics for the proposed deep face net network for detecting again the normal faces using the kaggle facial autistic datasets (KFAD-2020)

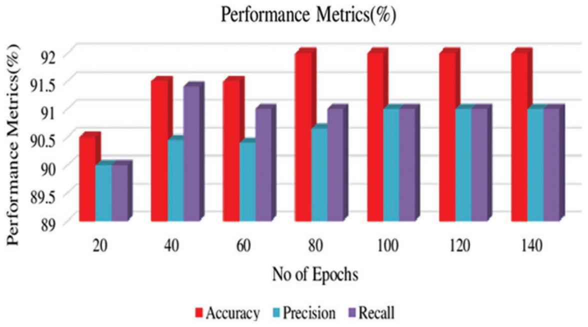

Figure 8: Performance metrics for the proposed deep face net network for detecting again the autism faces using the kaggle facial autistic datasets (KFAD-2020)

Table 6: Hyper parameters tuned for the multi-layer perceptron (MLP) models used for facial feature extraction

Layers	Settings	Parameters
Input layer	128 × 128 × 10	
Dense layer-1	128 nodes	20971648
Activation function	Sigmoid	
Dense layer-2	128 nodes	
Activation function	Sigmoid	16512
Dense layer-3	128 nodes	
Activation function	Sigmoid	16512
Dense layer-4	128 nodes	774
Activation function	Softmax	

Table 7: Hyperparameters tuned for the convolutional neural networks (CNN) models used for facial feature extraction

Layers	Output layer	Filter size/stride length
Input layer	128 × 128 × 10	16(2 × 2)/1
Convolution 1	128 × 128 × 10	16(2 × 2)/1
Max-pooling	64 × 64 × 5	—
Convolution 2	64 × 64 × 5	16(2 × 2)/1
Max-pooling	32 × 32 × 5	———
Convolution-3	32 × 32 × 5	8(2 × 2)/1
Max-pooling	16 × 16 × 5	———
Convolution 3	16 × 16 × 5	8(2 × 2)/1
Max-pooling layer	8 × 8 × 2	——
Activation (layers)	ReLU	———
Optimizer	Adam	———

Tab. 9 shows the leave out one validation process for the different models using the two different datasets such as AFFECTNETS and KFAD-2020. From Tab. 8, the model performance has exhibited the superior ability in detecting normal facial emotions and autistic facial emotions using the different datasets. The proposed model has exhibited 92.5% accuracy, the precision of 91.5% and recall of 91.5% whereas the MLP has exhibited the very low performance (50.67% accuracy, 49.67% precision and 48%, 67%) in detecting the normal facial emotions using AFFECTNET databases and very low performance in detecting autistic facial emotions (48.67% accuracy, 49.87% precision and 49.5%). The other models such as CNN (67.75% accuracy, 66.45% precision and 68.5% recall) and LSTM have shown moderate performance (71.56% accuracy, 66.50% precision and 65.60% recall) in detecting both normal facial emotions and autistic facial emotions using both datasets. The FACS based CNN-LSTM has proved a significant role in boosting the performance of the proposed model. Furthermore, the contingency table of the proposed model for detecting normal facial emotions and autistic emotions are illustrated in Tab. 10. All occurred numbers in true positive (TP), True negative (TN), False Negative (FN) and False Positive (FP) are completely addressed in the tables. For effective detection of autistic facial emotions, true positive and true negative values denote the accuracy of detection and the contingency table for detection of autistic and non-autistic facial emotions using KFAD-2020 datasets is tabulated in Tab. 11.

Table 8: Hyper parameters for the LSTM models used for facial feature extraction

LSTM cells	256	590080
LSTM cells	256	131483
No of dense layers	06	700

Table 9: Leave-one cross-validation process for the different algorithms using AFFECTNET datasets in the detection of normal facial emotions

Algorithm	Accuracy (%)	Precision (%)	Recall (%)
Multi-layer perceptron	50.67	49.67	48.67
CNN	68.45	67.35	68.45
LSTM	70.56	69.56	67.56
Proposed model	92.5	91.5	91.5

Table 10: Leave-one cross-validation process for the different algorithms using KAFD-2020 datasets in the detection of autistic facial emotions

Algorithm	Accuracy (%)	Precision (%)	Recall (%)
Multi-layer perceptron	48.67	49.87	49.67
CNN	67.75	66.45	68.55
LSTM	71.56	66.50	65.60
Proposed model	92.5	91.5	91.5

Determining reliability and performance is also demonstrated. False-positive and false negative readings, on the other hand, represent the effects of faulty detection. As can be shown in Tab. 11, the number of false positives and false negatives in the experimentation is relatively tiny.

Table 11: Contingency table for detection of autistic and non-autistic facial emotions using KFAD-2020 datasets

Truth values	Autistic facial emotions	Non-autistic facial emotions	Total
Autistic facial emotions	1534	133	1667
Non-autistic facial emotions	133	1534	1667

5 Conclusion

This study proposes a hybrid deep learning model to detect Autism spectrum disorder among children. This model employs the newly formed facial action coding systems (FACS) systems for an effective feature extraction and these features are used to train the new hybrid model of CNN-LSTM. The proposed DEEPFACENETS integrates the CNN-LSTM and exploits the advantages of both CNN and RNN. The experimentation is carried out using AFFECTNET databases and validated using Kaggle Autistic facial datasets (KAFD-2020). When compared to current models such as Multi-Layer Perceptron (48.67%), Convolutional neural networks (67.75%), and Long Short Term Memory (71.56), the suggested model showed a considerable increase in recognition rate (92%). Hence the proposed model proves its superiority in detecting autistic facial emotions among children effectively.

6 Future Scope

In future, multi-modal feature extractions can be employed along with the optimized or fused learning models to increase the detection rate for ASD-affected children belong to different age groups.

Acknowledgement: The author with a deep sense of gratitude would thank the supervisor for his guidance and constant support rendered during this research.

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] D. H. Oh, I. B. Kim, S. H. Kim and D. H. Ahn, “Predicting autism spectrum disorder using blood-based gene expression signatures and machine learning,” *Clinical Psychopharmacology and Neuroscience*, vol. 15, no. 1, pp. 47, 2017.

[2] M. Duda, R. Ma, N. Haber and D. P. Wall, “Use of machine learning for behavioral distinction of autism and ADHD,” *Translational Psychiatry*, vol. 6, no. 2, pp. 732–732, 2016.

[3] G. Li, O. Lee and H. Rabitz, “High efficiency classification of children with autism spectrum disorder,” *PLoS One*, vol. 13, no. 2, pp. 0192867, 2018.

[4] Q. Tariq, S. L. Fleming, J. N. Schwartz, K. Dunlap, C. Corbin *et al.,* “Detecting developmental delay and autism through machine learning models using home videos of bangladeshi children: Development and validation study,” *Journal of Medical Internet Research*, vol. 21, no. 4, pp. 13822, 2019.

[5] Q. Tariq, J. Daniels, J. N. Schwartz, P. Washington, H. Kalantarian *et al.,* “Mobile detection of autism through machine learning on home video: A development and prospective validation study,” *PLoS Medicine*, vol. 15, no. 11, pp. 1002705, 2018.

[6] M. J. Maenner, M. Y. Allsopp, K. V. N. Braun, D. L. Christensen and L. A. Schieve, "Development of a machine learning algorithm for the surveillance of autism spectrum disorder," *PLoS One*, vol. 11, no. 12, pp. 0168224.

[7] B. Li, A. Sharma, J. Meng, S. Purushwalkam and E. Gowen, "Applying machine learning to identify autistic adults using imitation: An exploratory study," *PLoS One*, vol. 12, no. 8, pp. 0182652, 2017.

[8] X. A. Bi, Y. Wang, Q. Shu, Q. Sun and Q. xu, "Classification of autism spectrum disorder using random support vector machine cluster," *Frontiers in Genetics*, vol. 9, pp. 18, 2018.

[9] F. Thabtah, N. Abdelhamid and D. Peebles, "A machine learning autism classification based on logistic regression analysis," *Health Information Science and Systems*, vol. 7, no. 1, pp. 1–11, 2019.

[10] Y. Qian, W. Zhou, J. Yan, W. Li and L. Han, "Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery," *Remote Sensing*, vol. 7, no. 1, pp. 153–168, 2015.

[11] D. Eman and A. W. Emanuel, "Machine learning classifiers for autism spectrum disorder: A review," in *Proc. ITISEE*, Yogyakarta, Indonesia, pp. 255–260, 2019.

[12] A. Appathurai and P. Deepa, "Design for reliablity: A novel counter matrix code for FPGA based quality applications.," in *Proc. Asia Symp. on Quality Electronic Design*, IEEE, Kula Lumpur, Malaysia, pp. 56–61, 2015.

[13] S. K. Jarraya, M. Masmoudi and M. Hammami, "Compound emotion recognition of autistic children during meltdown crisis based on deep spatio-temporal analysis of facial geometric features," *IEEE Access*, vol. 8, pp. 69311–69326, 2020.

[14] M. R. Ahmed, M. S. Ahammed, S. Niu and Y. Zhang, "Deep learning approached features for ASD classification using SVM," in *Proc. IEEE Int. Conf. on Artificial Intelligence and Information Systems*, Dalian, China, pp. 287–290, 2020.

[15] A. Sivasangari, P. Ajitha, I. Rajkumar and S. Poonguzhali, "Emotion recognition system for autism disordered people," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, pp. 1–7, 2019.

[16] M. Leo, P. Carcagnì, C. Distante, P. L. Mazzeo, P. Spagnolo *et al.,* "Computational analysis of deep visual data for quantifying facial expression production," *Applied Sciences*, vol. 9, no. 21, pp. 4542, 2019.

[17] R. Pathak and Y. Singh, "Real time baby facial expression recognition using deep learning and Iot edge computing," in *Proc. Int. Conf. on Computing, Communication and Security*, IEEE, Patna, India, pp. 1–6, 2020.

[18] B. Helfer, S. Boxhoorn, J. Songa, C. Steel, S. Maltezos *et al.,* "Emotion recognition and mind wandering in adults with attention deficit hyperactivity disorder or autism spectrum disorder," *Journal of Psychiatric Research*, vol. 134, pp. 89–96, 2021.

[19] A. S. Mohamed, N. Marbukhari and H. Habibah, "A deep learning approach in robot-assisted behavioral therapy for autistic children," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1.6, pp. 437–443, 2019.

[20] D. R. Ramji, C. A. Palagan, A. Nithya, A. Appathurai and E. J. Alex, "Soft computing based color image demosaicing for medical image processing," *Multimedia Tools and Applications*, vol. 79, no. 15, pp. 10047–10063, 2020.

[21] E. Friesen and P. Ekman, "Facial action coding system: A technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, pp. 5, 1978.

[22] V. S. Ramachandran, "Microexpression and macroexpression," *Encyclopedia of Human Behavior*, vol. 2, pp. 173–183, 2012.

[23] A. Mollahosseini, B. Hasani and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.