Distributed and Parallel Databases
https://doi.org/10.1007/s10619-021-07361-y

FIGS-DEAF: an novel implementation of hybrid deep learning algorithm to predict autism spectrum disorders using facial fused gait features

A. Saranya[1,2] · **R. Anandan**[1]

Accepted: 14 August 2021

Abstract

Autism spectrum disorder (A.S.D.) is considered a heterogeneous mental disorder, which is notoriously difficult to identify for a better diagnosis, especially among children. The current diagnosis methodology is purely based on the behavioural observation of symptoms prone to misdiagnosis. Several hybrid methods were explored, which also needs its improvisation in better prediction and diagnosis to move this field towards intelligent and accurate diagnosis. The main objective of this research paper was to develop the new diagnosis software which integrates the novel fuzzy hybrid deep convolutional neural networks and fusion of facial expressions and human gaits based on input video sequences. The algorithm has been trained and validated with the different video datasets such as Kaggle FER2013 and Karolinska Directed Emotional Faces (KDEF) datasets with real-time test scenarios, and various parameters such as accuracy, recall and F1-score were evaluated. Our proposed deep learning model outperforms another state-of-the-art method with an increase in prediction accuracy up to 30% with maximum accuracy of 95%. The model presented in this paper yields more advantages in terms of prediction time and accuracy also. However, the speech therapists, teachers, caretakers, and parents can use the software as a technological tool when working with children with A.S.D.

Keywords Autism spectrum disorder · Fuzzy · Hybrid deep convolutional · Fusion · Facial expressions · Human gaits

✉ A. Saranya
aksaranya@gmail.com

R. Anandan
anandan.se@velsuniv.ac.in

[1] Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies, Chennai, India

[2] Department of Computational Intelligence, S.R.M. Institute of Science & Technology, Chennai, India

Published online: 24 August 2021

1 Introduction

Autism spectrum disorder (A.S.D.) is one of the spectrums of neurodevelopment disorder characterized by creating complexities in social interaction and communication by its repetitive and stereotyped behaviours [1]. In addition, A.S.D. is characterized by a variety of physiological and behavioural features spanning sensory, neuromorphological and motor never deficits [2]. Many types of research have been conducted for the detection of A.S.D. in children [3]. But initial prediction for A.S.D. can help children receive timely behavioural therapy, which can improve their daily functioning, decrease symptom severity and optimize long-term outcomes [4, 5]. Various researches have been conducted from infants to grown adults [6] for early detection of A.S.D. using facial expressions [7], emotions identification [8, 9], and eye-tracking mechanism [7, 10].

As mentioned above, for children with A.S.D., face processing is challenging. It has been argued that children with A.S.D. to understand facial expression is impaired [11], and this difficulty may account for other problems that they demonstrate during social interactions [12]. Research studies discussed in [13], the accuracy and latency of emotion recognition were evaluated in children with A.S.D. and typically developing children while viewing videos of faces transitioning from a neutral expression to one of the six basic emotions [14]. Children with A.S.D. were slower in emotion recognition and selectively made more errors in detecting anger. Several other studies, including [15] and [16], showed impairment for children with A.S.D. in classifying and understanding facial expressions compared to healthy children of the same age. At this juncture, the prediction of A.S.D. using human gait signals [17–19] is gaining its importance due to its accurate differentiating between normal and abnormal children. Though it has gained its limelight, gait signals are used only for recording the locomotives of A.S.D. However, children in which the abnormalities can be identified by their walking styles, however, with the advent of machine learning [20] and deep learning algorithms [21], the A.S.D. Prediction with the techniques as mentioned above has reached its new height in the early diagnosis of A.S.D.

As one of the deep learning models convolutional neural networks (CNN), CNN plays a vital role in video analytics and computer vision systems. CNN utilizes the temporal character of data for establishing the non-linear simulation of relationships in input data [22]. In addition, CNN can effectively simulate the spatial dependence between the data [23]. CNN has two crucial parts. One is the feature extraction layer, in which the input of each neuron is connected to the local area block of the previous layer. The complete feature vector can be obtained by extracting the features of each local area block. The other is the feature mapping layer. As can be seen from the previous layer, each computing layer of the network would be composed of multiple feature maps. An activation function is added to the convolutional network of the feature mapping structure to ensure that the feature map has a constant displacement. In addition, the weight sharing method is used in the mapping layer to reduce model parameters.

Even though CNN plays a vital role in video analytics, employing CNN for predicting the A.S.D. seems to be a little complex in training because of the imbalance problems in A.S.D. Dataset. Meanwhile, fine-tuning a pre-trained model also yields poor performance. Therefore, this paper proposes integrating fuzzy fused extreme learning training [24, 25] to learn CNN features and accurately predicting and classifying A.S.D. using scalable datasets (both for large and even small datasets). This proposed methodology The contribution of this research paper is as follows.

(1) The paper discusses the novel creation of datasets for A.S.D. prediction. The new datasets are created by the fusion of facial emotions and human gait sequences. As discussed above, the single inputs are used to detect the A.S.D., leading to overfitting and imbalance problems. The proposed fused multimodal feature extraction has created the larger sensitive datasets that can be used for accurate classification, analysis, and even prediction.
(2) Secondly, the paper proposes the new learning model DEAF (deep extreme adaptive fuzzy) learning algorithms with two-layer sandwitched CNN feature extractor and integrated with fuzzy extreme learning machines (E.L.M.s) for proper training and prediction A.S.D. To the best of our knowledge, this proposed hybrid learning model, along with the multimodal feature extraction, would be the first methodology to solve the darker sides of A.S.D. prediction mechanism.

The organization of the paper is as follows.

Section 2 discusses the different works of more than one author. Section 3's thoughts about CNN, E.L.M. and fuzzy methods are addressed. The suggested models, multimodal feature extraction, CNN feature extractors, and fuzzy fused E.L.M. training are mentioned in Sect. 4. In Sect. 5, several trails of inputs, experiments, and datasets are described. The paper concludes with its future scope in Sect. 6.

2 Related works

Li et al. [26] developed a classification algorithm for autism detection. This paper portrays a new method based on rankings produced in support vector machine (SVM) modelling combined with high dimensional model representation (HDMR) sensitivity analysis. The novel approach finds and ranks the important causal metabolites in FOCM/TS pathways and then determines their independent and correlated motion patterns. Such information is precious now not solely for presenting a foundation for a pathological interpretation but additionally for potentially offering an early, dependable prognosis ideally leading to a subsequent complete therapy of A.S.D. With solely tens of SVM model runs, the new technique can perceive the combos of the most necessary metabolites in the FOCM/TS pathways that lead to A.S.D. Previous efforts to find these metabolites required hundreds of thousands of model runs with the same data.

Likewise, Bi et al. [27] developed multiple SVM classifiers for A.S.D. detection. Single SVM detection accuracy is less efficient for large data's. Random SVM

classification technique is developed to detect the abnormalities of A.S.D. in maximizing the accuracy to overcome this. Global decision-making points are detected randomly in the SVM cluster, which segregates the typical controls and A.S.D. independently. The performance of the proposed algorithm is evaluated using ABIDE dataset. Alarifi and Young [28] created various machine-learning classifiers such as support vector, multilayer perceptron, and random forest algorithms to detect A.S.D. for children and adults. These classifiers utilize social skills, repetitive behaviours, speech and non-verbal communication to train and test purpose.

Ganapathi Raju [29] has suggested the application of supervised machine learning algorithms to A.S.D. datasets. The methodology involves the reduction of the datasets by removing the outliner values. Additionally, the author has developed the XGBoost classifier and gradient boosting classifiers to evaluate the datasets in which 97.1% accuracy has been achieved. Furthermore, the author has also established that more balanced datasets increase classifier accuracy in predicting and categorizing the diseases mentioned above. Omar et al. [30] has developed a successful autism forecast model by consolidating Random Forest-CART (Classification and Regression Trees) and Random Forest-ID3 (Iterative Dichotomiser 3). The proposed model was assessed with the AQ-10 dataset and 250 real-time datasets gathered from individuals with and without authentic qualities. In this methodology, prediction accuracy and reduced false rates were achieved compared with traditional machine learning algorithms.

Hyde et al. [31] gives an extensive survey of 45 papers using regulated A.I. in A.S.D., including calculations for classification and text analysis. The objective of the article is to recognize and portray managed ML slants in A.S.D. literature and inform and guide researchers keen on extending the collection of clinically, computationally, and measurably stable methodologies for mining A.S.D. information.

Thabtah [32] concentrated on machine learning techniques to handle A.S.D. as a classification problem and fundamentally examined their points of interest and drawbacks. Besides, this work demonstrated the significant advances required to improve insightful diagnostic approaches dependent on A.I. by replacing the handmade principles inside the A.S.D. Screening instruments with a prescient model. Ultimately, the proposed work featured the earnestness of refreshing A.S.D. Clinical screening devices to reflect changes proposed in the DSM-5 manual. The scattering of the DSM-5 requested an adjustment in how the indicative calculation coded inside the A.S.D. The screening apparatus acts during the time spent arranging cases. There is a need to reconsider questions or highlights inside the A.S.D. demonstrative instruments to satisfy the new criteria of the DSM-5. It requires mapping the new A.S.D. criteria to the highlights or properties utilized in the clinical conclusion instrument, just as assessing how the analytic calculation works.

Bram van cave Bekerom use A.I. to decide many conditions that together demonstrate to be prescient of A.S.D. The proposed method is extraordinary to physicians, helping them identify A.S.D. at a lot before the stage. It would be done through writing surveys, information investigation, and assessment. Anticipating if a youngster has A.S.D. demonstrated conceivable by utilizing formative deferral, learning inability and discourse, or other language issues as properties and incorporate physical movement, untimely birth, and birth weight to improve the accuracy. Utilizing

the 1-away technique, it was additionally conceivable to foresee the seriousness of A.S.D. sensibly. The 1-away strategy improved the accuracy from 54.1 to 90.2%, which is a significant increment. The seriousness is dependent on the contribution from simply the caretakers of the kids, prompts the requirement for further research in this issue [33].

Hyde et al. [31] developed an ANN for autism detection for ABIDE datasets in. The proposed neural network includes input layers with hidden neurons. Auto encoders were utilized for feature selection and trained using M.L.P. Multilayer perceptron trained with fine-tuned features for A.S.D. detection, which achieved only 70% of accuracy at runtime.

3 Preliminary background

This section discusses the methodology of CNNs, E.L.M.s.

3.1 Convolutional neural networks

In recent years, CNNs is the type of deep learning model and belongs to the artificial neural networks (ANN), which finds its applications in image processing and video analytics. The structure of the traditional CNN is shown in Fig. 1. There are five layers in the CNN model. The input layer consists of a matrix of the normalized patterns and feature maps used to connect inputs with its previous layers. The features obtained by the convolutional layer are used as the inputs to pooling layers. All the neurons in one feature map share the same kernel and connecting weights (known as the sharing weights in [34]).

The unique structural characteristics of the CNN model are listed as follows.

(1) It has three unique structural characteristics such as local sensing domain, weight sharing and down sampling.
(2) Weight sharing is integrated, which reduces the training parameters of the networks and the number of training samples.

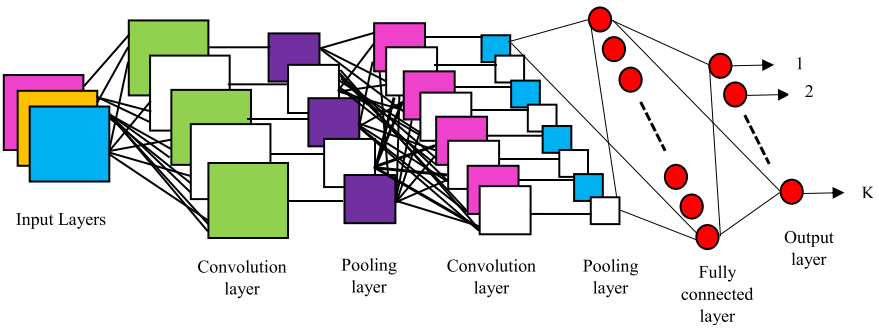

Fig. 1 Block diagram for the traditional convolutional neural networks

(3) The implementation of down sampling in the model reduces the noises capability and feature dimension of images.

The CNN model is divided into the input layer, hidden layer, and output layer. There are two hidden layers: convolution layer (extracting feature) and down sampling layer (selecting the optimization feature). After CNN feature extraction, these feature maps are trained by fully connected layers backed with the back-propagation neural networks.

3.2 Extreme learning machines

E.L.M.s proposed by Huang and Chen [35], in which the network utilizes the single hidden layer, high speed and accuracy and preparing velocity, great speculation/exactness, and universal function approximation capabilities [35, 36]. In this system, the' L' neurons in the hidden layer must work with a vastly differentiable activation function (for instance, the sigmoid function). However, that of the output layer is straight. In E.L.M., hidden layers do not need to tune mandatorily. In E.L.M., the hidden layer compulsorily need not be tuned. Loads of the hidden layer are arbitrarily appointed (counting the bias loads). It isn't the situation that hidden nodes are irrelevant. However, they need not be tuned, and the hidden neuron parameters can be haphazardly produced even in advance. That is, before taking care of the training set data. For a single-hidden layer E.L.M., the system yield is given by Eq. (1)

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta, \tag{1}$$

where x input.

β output weight vector and it is given as follows as

$$\beta = [\beta_1, \beta_2, \ldots, \beta_L]^T. \tag{2}$$

$H(x)$ output hidden layer which is given by the following equation

$$h(x) = [h_1(x), h_2(x), \ldots, h_L(x). \tag{3}$$

To determine output vector O, which is called the target vector, the hidden layers [h(x)] are represented by Eq. (4)

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix}. \tag{4}$$

The basic implementation of the E.L.M. uses the minimal non-linear least square methods and βcan be determined in Eq.(5) can be determined in Eq. (5)

$$\beta' = H^* O = H^T (HH^T)^{-1} O, \tag{5}$$

where H * inverse of H known as Moore–Penrose generalized inverse.

After rearranging, the Eq. (5) is given as

$$\beta' = H^T \left(\frac{1}{C} HH^T \right)^{-1} O. \tag{6}$$

Hence the output function can be determined by using Eq. (6) in Eq. (1)

$$f_L(x) = h(x)\beta = h(x)H^T \left(\frac{1}{C} HH^T \right)^{-1} O. \tag{7}$$

E.L.M. uses the kernel function to yield good accuracy for better performance. The significant advantages of the E.L.M. are minimal training error and better approximation. Since E.L.M. uses the auto-tuning of the weight biases and non-zero activation functions, E.L.M. finds its applications in classification and prediction values. The detailed description of E.L.M.'s equations can be found in [35, 36].

4 Proposed DEAF classifiers

This section discusses the system overview, feature extraction using CNN, a fusion of fuzzy features, and finally trained with the E.L.M.s.

4.1 System overview

The proposed classifier used for this model is inspired by the family of deep learning algorithms and machine learning training models, especially single feedforward networks mechanism. Since the model has to be trained with different input features, improvisation of DCNN is required in terms of scalability and high performance. The design idea is also heavily inspired by [37] in layers' formation, output size and training methodology. Also motivated by traditional deep learning networks [38], the proposed classifier has replaced feedforward training and fuzzy classifiers. The architecture of the proposed classifier is shown below in Fig. 2. In phase one, the proposed model takes the input facial images of 48×48 pixels and processes them through several convolutions, max-pooling and single feedforward layers providing the phase -1 output of either seven classes of emotions such as Anger, Disgust, Fear, happiness, sadness, surprise and neutral. In similar fashions, the second layer of the proposed model is trained with the different gait video sequences to provide the spatio-temporal outputs.

4.2 DEAF models for facial feature extraction

In the first phase, the proposed DEAF models have been applied for facial emotion extraction. As mentioned in Sect. 4.3, this phase of training accepts the input images as 48X48 facial images. It provides the outputs of seven classes of emotional categorization as Anger, Disgust, Happiness, Sadness, Surprise and Neutral. All convolution layers have a filter size of 3×3, and all the pooling layers have a

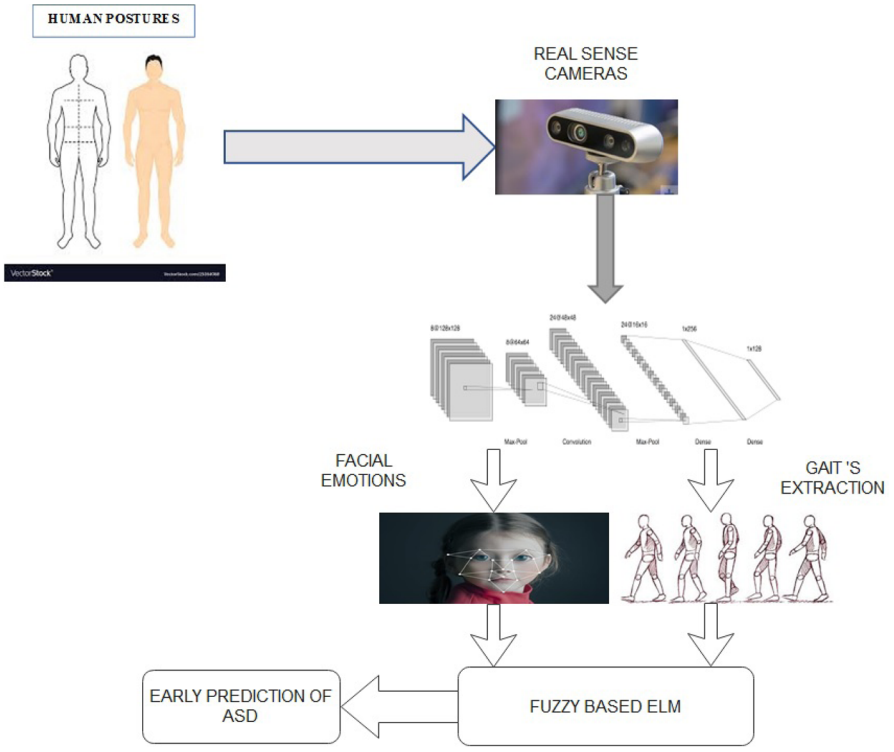

Fig. 2 Working architecture of proposed hybrid deep learning framework

size of 2×2. These are hyper parameters and tuned during the training process to determine the optimized models. The activation layer is included, followed by the normalization after every activation layer to gain higher accuracy. The Dropout layer is usually used after training layers. However, after performing run trials, it has been seen that using the Dropout after each Pooling layer helps to reduce the overfitting of the model. Therefore, the convolution-pooling group consists of Convolution, Activation, Batch Normalization, Pooling, and Dropout layers, respectively. The hyper parameters used for the training the facial emotions are tabulated in Table 1. The facial image datasets such as KDEF and FER2013 datasets were trained, and E.L.M.s were employed to classify facial emotions effectively. Figure 3 shows the DEAF models employed for facial emotions.

4.3 DEAF models for gait feature extraction

4.3.1 Gaits features

In this phase, human gait video sequences are considered as the inputs for training the proposed models. Gaits can be termed the transformation of brain activity of muscle contraction patterns resulting in a walking sequence. It is a command

Table 1 Hyper parameters tuned for the CNN-DEAF models used for feature extraction

Layers	Output layer	Filter-pool layer
Input layer	48×48×1	3×3
Convolution1	48×48×32	3×3
Max-pooling	24×24×32	2×2
Convolution 2	24×24×64	3×3
Max-pooling	12×12×64	2×2
Convolution-3	12×12×64	3×3
Max-pooling	6×6×128	2×2
Fully connected	06	–
Classifier	Softmax	07
Activation	ReLU	–

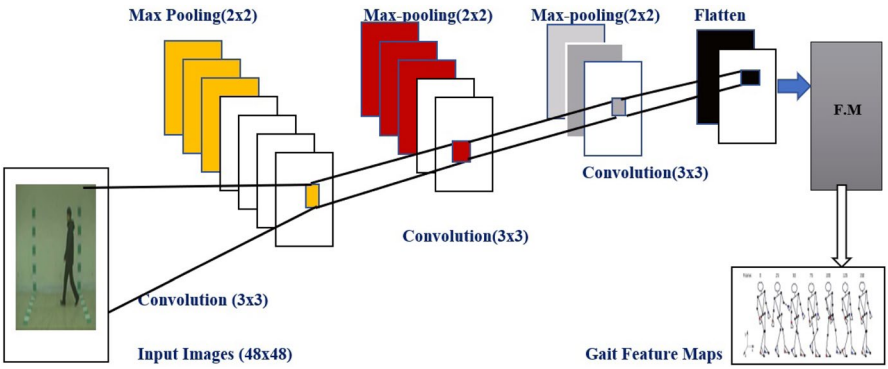

Fig. 3 Block diagram for CNN gait features using CNN-DEAF models

chain generated in the brain and transmitted through the spinal cord to activate the lower neural centre, which would trigger the muscular contraction patterns assisted by sensory feedbacks from joints, muscles and other receptors for controlling the movements.

Gait is a sequence of periodic events characterized as repetitive cycles of each foot [39], and gaits events that are typically employed are tabulated in Table 2.

With regards to the above gait events mentioned in the Table 2 following spatio-temporal gaits features are needed for an effective health care diagnosis (especially diagnosis of different levels of A.S.D.) which are presented below.

Frames/gait cycle it is the measure of the number of frames per gait cycle. This measurement gives the average walking speed of the person. This measurement plays an essential role in classifying the different persons in terms of physio neuro disorders. The frames per gait cycle are measured as the difference between the frame of the heel strike (H_F) and the frame of toe-off (T_F).

Table 2 Gaits events recorded for complete human walking mechanism

Sl. nos	Gait events	Description
01	– Heel strike or initial contact	Starts the moment the foot touches the ground, and it is the initial double-limb support interval
02	Loading response or foot flat	Single support interval following the double support interval
03	– Mid-stance	Ingle support interval between opposite toe-off and heel-off
04	Terminal stance or heel-off	The heel rises in preparation for the opposite swing
05	Pre-swing	This is the second double-limb support interval
06	Terminal swing	This is the last interval of the gait cycle and the end of the swing phase
07	Mid-swing	This interval begins with the toe-off into single support and starting to swing

Swing ratio it is the measurement of the swing of the hands during walking. The ratio of the maximum torsal width to the minimum torsal width gives an estimation of the swing ratio. The ratio is calculated for all such instances, and the final result is obtained by averaging the values.

Cadence it is measured in terms of no of steps /minute. This feature is calculated based on the total number of frames required for one step, and it depends on the camera's frame rate.

Velocity this feature is measured based on the distance covered by the body per unit time. Since user-defined velocity changes from time to time, average velocity is taken for the measurement. The average velocity is thus calculated by

$$\text{A.V.} = \text{S}_\text{L} \times \text{C}. \tag{8}$$

Step length is calculated based on the distance between the successive heel contact points of the opposite feet (S_L).

C cadence, which is calculated in the previous case.

Step length (S_L) this feature is measured by the step length of the subject from toe to toe. Step lengths are divided by the height of the subject to normalize the value step. The normalized step length is calculated by

$$\text{S}_\text{LN} = \text{Step length}/\text{H}, \tag{9}$$

where H is the height of the person.

Foot length (F_L) this feature measures the foot length, which is then calculated by the toe and heel distance (D_F). Again the normalized foot length is calculated by diving by the person's height.

$$\text{F}_\text{L} = \text{D}_\text{F}/\text{H}, \tag{10}$$

where H is the height of the person.

Cycle time (T_c) this feature calculates the time to complete one cycle by considering the number of camera frames in the gait cycle.

Mass location point (M.L.P.) it is the measurement of the relative position of the body under the centre. The normalized mass location is calculated by dividing the mass location point by the height of the person.

The proposed DEAF models are employed similarly, as mentioned in Sect. 4.4. Figure 3 illustrates the usage of proposed models for extracting the gait features.

Above mentioned gait features mentioned in the table have clear observational values, but extracting and differentiating among the gait features remains complex. The proposed CNN-DEAF models provide the model-free approach used to extract gait from video sequences using feature engineering, as proposed in [39, 40]. Here, proposed learning is utilized to automatically extract gait features from video sequences, which maximizes data variability and eliminates the dependence on handcrafting. The above CNN features are now used for training the fuzzy-E.L.M. in proposed hybrid deep learning frameworks.

4.3.2 DEAF-fuzzy E.L.M. training model

The figure shows the fuzzy fused E.L.M.s for training the CNN features. The proposed DEAF model replaces fully connected layers' w the new fuzzy-based E.L.M.s (FELM) to predict A.S.D. among the different categories of children. In the training phase, the four-layer architecture for E.L.M. is trained by randomly assigned hidden layer parameters and optimizing the associated weights between the hidden layer and output layer with the Inverse Moore pseudo matrix. Essentially, the fuzzy rule is formulated to establish the optimized relationship between each hidden node and weights of the output nodes. Compared with the other approaches, FELM can create fuzzy rules with greater interpretability and understand ability. It is because five equally distributed membership functions, i.e., deficient, low, medium, and high, are defined for each linguistic input variable. Besides, the "don't care" case for a linguistic input variable is considered to reduce the number of required linguistic input variables in fuzzy rules. Hence the operation of the proposed training model is equivalent to the fuzzy inference system. In this approach of making the stable E.L.M. for better training, first-order Takagi–Sugeno–Kang (TSK) type and traditional AND operators are used for forming the fuzzy rules.

The detailed training process of the FELM is described below.

Step 1 initialize the parameters of FELM. For each input variable xi, five Gaussian membership functions are defined as

$$g(xi, \mu k, \phi) = \exp\left(-\frac{xi - (\mu k)^{0.5}}{2\phi^2}\right), \tag{11}$$

where $\mu k, \phi$ are the mean and standard deviation, respectively. Meanwhile, $\mu k$is fixed with the kth value of sequence (0.0, 0.25, 0.50, 0.75, 1.0) and ϕ is randomly assigned. All the inputs variables are normalized before feeding into the training network. Besides, fuzzy rules are employed for selecting the number of hidden nodes (L) and randomly assigning binary values to the three-dimensional

rule-combination matrix with the size C of N × L × 5 and the two-dimensional D.C. matrix with N × L where N is the number of features used.

Step 2 calculate all the firing degrees of all fuzzy rules for each training input and output data. For each (xi, yi), the membership firing angle is calculated by the modified MODPRO.

Step 3 combine rule firing degrees of all training data into a hidden layer output matrix as mentioned in Eq. (3).

Step 4 calculate the Output matrix as mentioned in Eq. (4).

Step 5 with the help of the membership functions mentioned in Eq. (7) and fuzzy triggering angles mentioned in Eq. (8), the corresponding output values 'y' from the group of input features. The final output functions are given by

The final output functions are given by the Eq. (6) is:

$$f_L(x) = Fuzzy(h(x)\beta) = Fuzzy\left(h(x)H^T\left(\frac{1}{C}HH^T\right)^{-1}O\right). \tag{12}$$

The complete pseudo-code for the proposed hybrid deep learning models is depicted below.

Sl. no	Pseudo Code for the Proposed Hybrid DEAF Models
1	Input Variables X1 = Facial Emotions = (×1, ×2, ×3,…, ×7)
2	Input variables Y1 = Human Gait Features = (y1, y2, y3, y4, y5…,y8)
3	Output: Multiple classifications
4	Initialization: Fuzzy Assigned Neurons and Bias weights for Learning
5	For i = 1 to N do
6	For i = 1 to M do
7	X(k) = conv (xi, Wi) where Wi is the intermediate layers
8	Z = Fullyconnected (X(k)
9	Z′ = cross entrophy(Z)
10	End
11	For u = 1 to d
12	Y(k) = conv(yi, Wj)
13	Y′ = fullyconnected(Y(k))
14	Z″ = cross entropy(Y′)
15	End
16	V = Fuzzy-ELM(Z′, Z″) using the Eq. (11)
17	If V = = 0 && V < 0.5
18	/*No ASD is Found*/
19	Else if V = = 0.5 && V < = 1
20	/* Stage-1 ASD is found*/
21	Else if V = = 1
22	/* Abnormal Stage of ASD is found*/
23	End
24	End
25	End

Sl. no	Pseudo Code for the Proposed Hybrid DEAF Models
26	End

5 Experimentation setup

To investigate the performance of the deep learning models and the different feature selection, train the models with different scenarios such as Facial Features, Gait features, and combined. Moreover, testing was carried out with 20% testing data and also validated in the real-time scenario. The Intel Real Sense Camera captures the real-time video images with 562×762 resolution and a universal serial bus (USB) to interface with software developed to capture and analyse the different sets of video. The software was developed using Python 3.8 with Keras API (Tensor flow as Backend), which runs on Intel I5 CPU, 8 GB RAM, 2 TB HDD and 2 GB NVIDIA GPU. Figure 4 shows the experimentation setup carried out in the real-time scenario.

A level-one feature—subject-out cross-validation was used in the experiments. One feature was left out in each cross-validation run as testing data, and testing performance was calculated. The process was repeated 50 times, and results were combined, which are used for evaluation. The parameters such as accuracy, specificity and sensitivity were calculated by using the following mathematical expressions.

$$\text{Accuracy} = \frac{D.R}{\text{T}}, \tag{13}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{TN}}, \tag{14}$$

$$\text{Recall} = \frac{\text{TN}}{\text{TP} + \text{TN}}. \tag{15}$$

TP and TN represent true positive and true negative values, and DR and T represents the number of detected results and total number of iterations. The process of evaluation has been carried out in four different phases. The proposed algorithm has been trained with facial features using the Facial Emotions datasets and

Fig. 4 Experimental setup for the real data set collection to evaluate the proposed framework

spatiotemporal gait features using Human Gaits Databases in the first phase. Various parameters such as accuracy, sensitivity, and specificity were calculated. This phase also evaluates the performance of the proposed learning models in real-time scenarios using the conditions mentioned above. The second phase demonstrates the performance of proposed learning models and the fused features from the different subjects to predict A.S.D. The comparative analysis using the different learning models in predicting the A.S.D. has been demonstrated in the third phase.

5.1 Experiments on facial emotion datasets

In the phase-I evaluation, datasets such as FERI datasets and KDEF datasets were used to train facial emotions. In contrast, the CASIA datasets were used for training the human gait sequences. In this evaluation, a real-time scenario was also used for training and testing the proposed model. Table 3 demonstrates the performance of the proposed learning models in terms of identifying the region of convergence (RoC).

Tables 3 and 4 illustrate the performance of the proposed learning models with the different training datasets of predicting facial emotions. From all the tables, it is clear that the prediction accuracy of proposed models has reached its maximum value of 89%, 84% for KDEF and FERI datasets at 100 hidden neurons. To validate the proposed learning models suitable for real-time scenarios, it has been experimented on 50 subjects ranging from the ages of 5 to 30. Table 5 demonstrates the performance of proposed models for the real-time subjects.

Table 5 shows apparent accuracy of prediction for real-time subjects is found to be 88% for 100 hidden neurons, and its matches with the performance of KDEF datasets 2013. Moreover, the results are compared the proposed hybrid CNN models with other exiting, deep learning models, as mentioned in [39–45].

Figures 5, 6 and 7 show the comparative analysis of different parameters such as accuracy, sensitivity, and specificity for the different hybrid learning models. The proposed hybrid DEAF models have outperformed the other learning models. Hence, the proposed model finds it suitable for detecting facial emotions, which can be considered an essential feature for predicting A.S.D.

5.2 Human gait databases

This evaluation investigates the proposed learning models' performance with two categories of datasets: CASIA datasets (http://www.cbsr.ia.ac.cn/users/szheng) and Real-time gait databases. For the real-time scenario, above mentioned experimentation has been carried out to determine the performance of proposed learning models, and performance metrics of the proposed model is presented in tables.

From Tables 6 and 7, the proposed model's accuracy is 90%, 88.5% for CASIA-A and real-time datasets, respectively. Moreover, stable accuracy is found from 100 to 200 neurons as in the previous case. These findings have given the way of determining the stability of the network for different datasets, which forms the basis of predicting the A.S.D. (Figs. 8, 9, 10).

Table 3 Performance metrics for the proposed algorithms using KDEF datasets 2013to detect the facial emotions

No. of hidden training networks	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
20	KDEF datasets	89.0	84.0	85.0	88.5	88.3	88.0
40		87.5	86.5	84.5	86.5	85.4	85.0
60		88.5	87.0	85.0	84.0	83.0	82.5
80		88.5	88.22	85.0	84.0	82.5	81.0
100		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**
120		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**
140		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**
160		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**
180		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**
200		**89.0**	**85.5**	**84.0**	**87.5**	**86.5**	**84.5**

Bold values indicate the final stage, after the 100th iteration that the values are constant

Table 4 Performance metrics for the proposed algorithms using FERI datasets to detect the facial emotions

No of hidden training networks	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
20	FERI datasets	86.5	84.2	83.5	82.0	81.5	80.5
40		85.0	84.5	82.5	83.0	81.5	81.25
60		84.5	81.5	82.0	81.5	80.5	80.0
80		83.5	82.0	81.0	84.0	81.0	78.5
100		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**
120		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**
140		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**
160		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**
180		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**
200		**84.0**	**83.5**	**82.5**	**83.0**	**82.5**	**81.5**

Bold values indicate the final stage, after the 100th iteration that the values are constant

Table 5 Performance metrics for the proposed algorithms using real-time datasets to detect the real-time facial emotions

No of hidden training networks	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
20	Real-time datasets	88.5	83.5.0	85.5	87.0	86.53	86.0
40		88.5	83.05	84.0	86.5	85.55	85.5
60		87.5	83.0	82.55	85.0	84.50	82.0
80		87.0	85.0	84.5	86.6	83.5	80.0
100		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**
120		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**
140		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**
160		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**
180		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**
200		**88.0**	**86.0**	**85.5**	**87.5**	**85.5**	**84.5**

Bold values indicate the final stage, after the 100th iteration that the values are constant

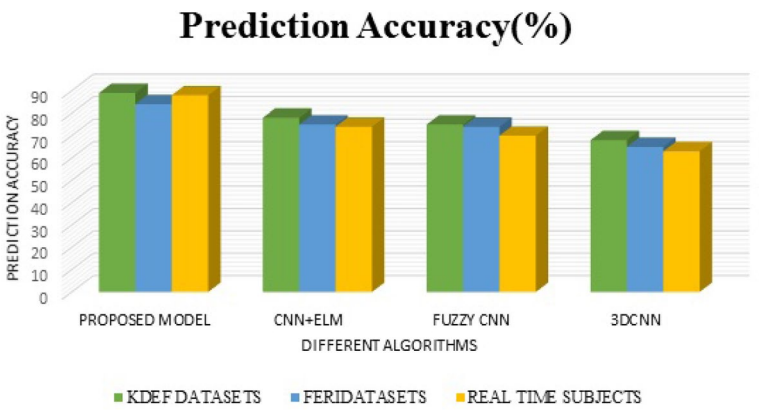

Fig. 5 Comparative analysis of accuracy between the proposed algorithms with the other existing algorithms in detecting the facial emotions

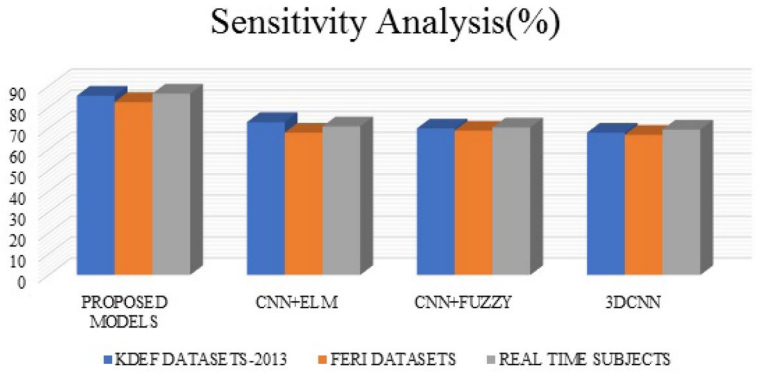

Fig. 6 Comparative analysis of sensitivity between the proposed algorithms with the other existing algorithms in detecting the facial emotions

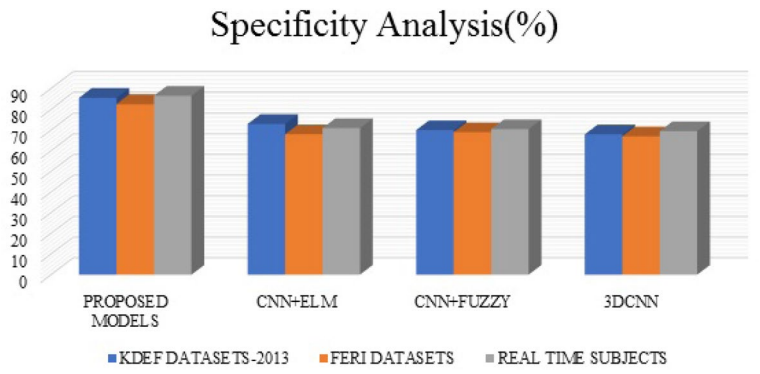

Fig. 7 Comparative analysis of specificity between the proposed algorithms with the other existing algorithms in detecting the facial emotions

As discussed in the previous section, the addition of fuzzy fused E.L.M. in the CNN training has outperformed all other algorithms for detecting human gait sequences. From all the above experimentation, it has been proved that proposed hybrid models work with comparatively high accuracy for single modal inputs.

5.3 Phase-II evaluation

This evaluation phase presents the proposed learning model's performance and the fused features (multimodal) using proposed hybrid deep learning models in predicting the A.S.D. Nearly 50 subjects were taken for this evaluation.

Table 8 illustrates the performance analysis of the proposed DEAF model with the different features in predicting the A.S.D. The fused features, along with the hybrid models, has produced a good performance in predicting A.S.D. Furthermore, proposed model results are validated with the actual reports from the autism coach and trainers.

Figures 11, 12, 13 and 14 shows the validation of the proposed DEAF models for 10 iterations (10-cross validation) along with the actual values in which the error (RMSE) has been calculated. The RMSE error is less than 0.00101, making the proposed DEAF models for an adequate prediction of the A.S.D. Moreover, to demonstrate the superiority of the proposed models, all the results are compared with the iteration of different learning models in detecting A.S.D. (Table 9).

From the above table, it is clear that the adoption of fuzzy fused E.L.M. training on CNN feature vector has a considerable increase in performance in detecting the A.S.D. Nearly 50% decrease in RMSE and highest prediction accuracy of 96.5% compared with other hybrid deep learning models. These extensive experiments show that the proposed model has strong feasibility in predicting the A.S.D. among the different age groups, and the extensive experiments have demonstrated it.

6 Conclusion

Fuzzy modelling has many advantages over the other conventional methods, such as uncertainties and less sensitivity to varying dynamics of non-linear systems. E.L.M., on the other hand, has high accuracy and less computation time. Besides, CNN has high-level feature abstraction, and it is suitable for extensive data analysis and learning. This research takes advantage of FELM and CNN, and a new hybrid DEAF model has been proposed. The proposed model has been tested under various categories of inputs such as facial emotions, human gait video sequences and even fused multimodal features maps. The results show that fused feature maps integrated with hybrid FELM have driven CNN has produced the excellent performance of 96.5% prediction accuracy and RMSE error of 0.0010. With this evaluation, the proposed DEAF model has proved its fitness in predicting the A.S.D. among the different age groups. The proposed algorithm can be improvised by using bio-inspired fuzzy optimizers to reduce the dimensionality and increase the classification and prediction performance.

Table 6 Performance metrics for the proposed algorithms using CASIA-A datasets to detect the human gait movements

No of hidden training networks	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
20	CASIA-A datasets	89.0	88.5	87.5	88.5	87.5	86.0
40		90.5	89.5	88.5	89.5	85.5	85.0
60		89.5	88.45	88.50	88.5	87.0	87.0
80		89.5	88.40	88.30	88.0	85.0	86.0
100		**90.0**	**89.0**	**88.50**	**90.0**	**88.5**	**88.0**
120		**90.0**	**89.5**	**88.5**	**90.0**	**88.5**	**88.0**
140		**90.0**	**89.0**	**88.5**	**90.0**	**88.5**	**88.0**
160		**90.0**	**89.5**	**88.5**	**90.0**	**88.5**	**88.0**
180		**90.0**	**89.5**	**88.5**	**90.0**	**88.5**	**88.0**
200		**90.0**	**89.0**	**88.50**	**90.0**	**88.5**	**88.0**

Bold values indicate the final stage, after the 100th iteration that the values are constant

Table 7 Performance metrics for the proposed algorithms using real-time datasets to detect the human gait movements

No of hidden training networks	Dataset details	Performance metrics					
		Training metrics			Testing metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
20	Real-time data	88.0	87.5	85.0	88.0	86.5	84.5
40		88.0	86.5	83.5	88.5	85.5	85.0
60		89.5	88.45	88.50	88.5	87.0	87.0
80		89.5	88.40	88.30	88.0	85.0	86.0
100		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**
120		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**
140		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**
160		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**
180		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**
200		**88.5**	**85.5**	**84.0**	**88.0**	**85.0**	**83.0**

Bold values indicate the final stage, after the 100th iteration that the values are constant

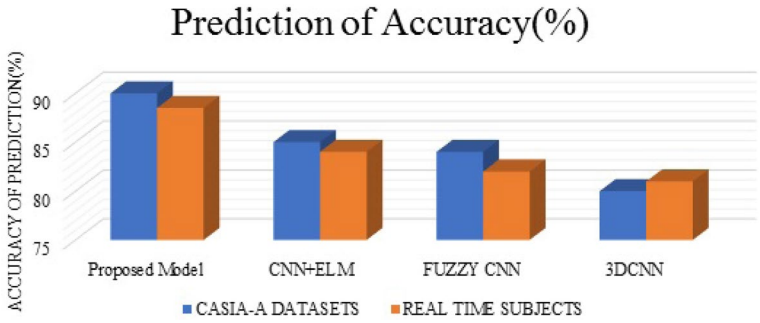

Fig. 8 Comparative analysis for accuracy between the proposed algorithms with the other existing algorithms in detecting the human gait features

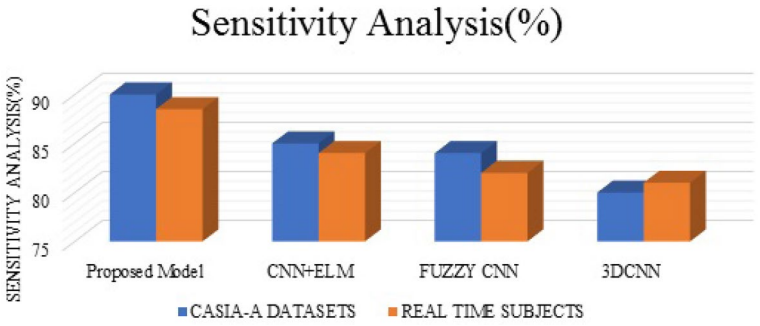

Fig. 9 Comparative analysis for sensitivity between the proposed algorithms with the other existing algorithms detecting the human gaits features

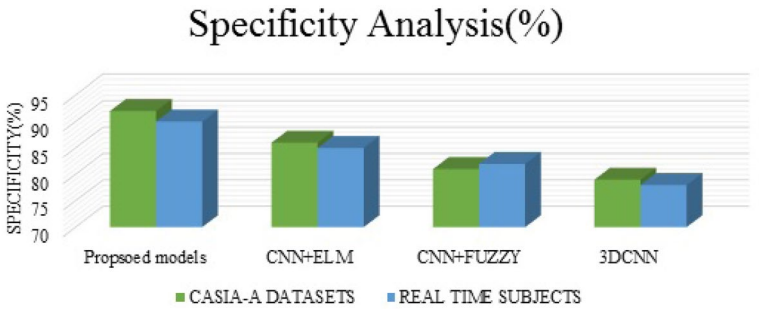

Fig. 10 Comparative analysis for specificity between the proposed algorithms with the other existing algorithms detecting the human gaits features

Table 8 Performance of proposed hybrid DEAF models with single and multi-modal features

Age factors	Features used	Performance metrics		
		Accuracy (%)	Sensitivity (%)	Specificity (%)
Age 5–8	Face emotions	87.5	85.5	84.5
	Human gaits	88.5	85.5	84.0
	Fused features	95.4	93.5	94.0
Age 9–12	Face emotions	87.0	85.0	84.0
	Human gaits	88.0	85.0	84.5
	Fused features	96.0	94.5	94.5
Age 13–16	Face emotions	85.0	84.0	83.5
	Human gaits	86.5	85.5	84.5
	Fused features	95.5	94.5	94.0
Age 45–50	Face emotions	88.0	86.5	85
	Human gaits	88.5	87.0	84.5
	Fused features	96.5	94.5	95.0

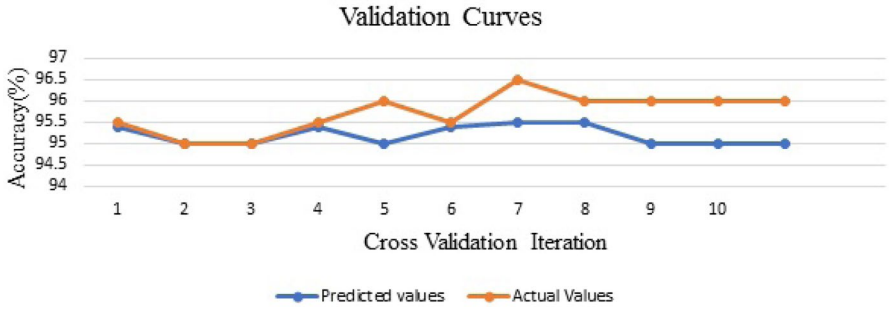

Fig. 11 Validation curves between the actual and predicted values for A.S.D. prediction from the age of 5 to 8

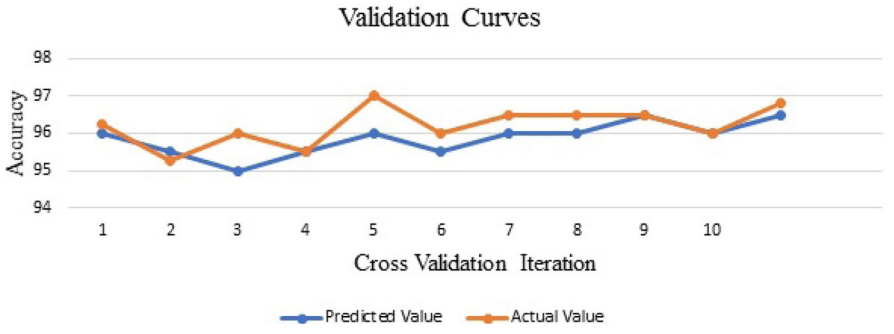

Fig. 12 Validation curves between the actual and predicted values for A.S.D. prediction from the age of 9 to 12

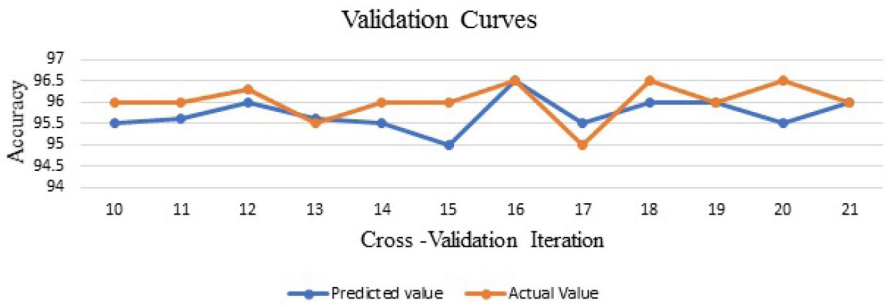

Fig. 13 Validation curves between the actual and predicted values for A.S.D. prediction from the age of 13 to 16

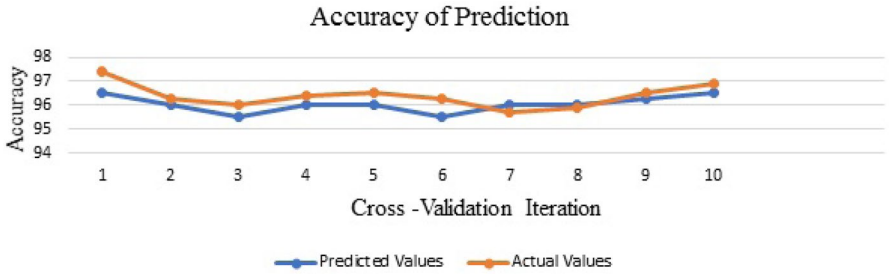

Fig. 14 Validation curves between the actual and predicted values for A.S.D. prediction from the age of 45 to 50

Table 9 Comparative analysis between the proposed models and other existing models used in A.S.D. detection

Sl. nos	Model details	Features used	Performance metrics	
			Accuracy (%)	RMSE error
01	3D CNN [4]	Human gaits	82	0.5
02	CNN + LSTM [40]	Face emotions	90	0.02
03	CNN + SVM [41]	Facial emotions	87	0.015
04	CNN + Fuzzy [42]	Facial emotions	88	0.032
05	LSTM models [43]	Human gaits	89	0.012
06	CNN + LSTM [44]	Facial emotions	91	0.015
07	CNN + ELM [45]	Facial emotions	91.5	0.0043
08	Proposed DEAF models	Multi-modal features	96.5	0.0010

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders

(DSM-5R). American Psychiatric Association, Washington, DC (2013)
2. Kanner, L., et al.: Autistic disturbances of affective contact. Nerv. Child **2**(3), 217–250 (1943)
3. Carmona-Serrano, N., López-Belmonte, J., López-Núñez, J.-A., Moreno-Guerrero, A.-J.: Trends in autism research in the field of education in Web of Science: a bibliometric study. Brain Sci. **10**(12), 1018 (2020)
4. Goren, C.C., Sarty, M., Wu, P.Y.: Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics **56**(4), 544–549 (1975)
5. Dwyer, P., Saron, C.D., Rivera, S.M.: Identification of longitudinal sensory subtypes in typical development and autism spectrum development using growth mixture modelling. Res. Autism Spectr. Disord. **78**, 101645 (2020). https://doi.org/10.1016/j.rasd.2020.101645
6. Udayakumar, N.: Facial expression recognition system for autistic children in virtual reality environment. Int. J. Sci. Res. Publ. **6**(6), 613–622 (2016)
7. Black, M.H., Chen, N.T., Iyer, K.K., Lipp, O.V., Bolte, S., Falkmer, M., Tan, T., Girdler, S.: Mechanisms of facial emotion recognition in autism spectrum disorders: insights from eye tracking and electroencephalography. Neurosci. Biobehav. Rev. **80**, 488–515 (2017)
8. Anjana, R., Lavanya, M.: Facial emotions recognition system for autism. Int. J. Adv. Eng. Technol. **5**, 40–43 (2014)
9. Haque, M.I.U., Valles, D.: A facial expression recognition approach using DCNN for autistic children to identify emotions. In: IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 546–551
10. Crawford, H., Moss, J., Oliver, C., Elliott, N., Anderson, G.M., McCleery, J.P.: Visual preference for social stimuli in individuals with autism or neurodevelopmental disorders: an eye-tracking study. Mol. Autism **7**(1), 24 (2016)
11. Prem Kumar, K., Murugapriya, K., Varsha, M.R., Asmitha, R., Sureka, S.: Facial emotion recognition for autism children. J. Innov. Technol. Explor. Eng. **9**(7), 1274–1278 (2020)
12. Gepner, B., Deruelle, C., Grynfeltt, S.: Motion and emotion: a novel approach to the study of face processing by young autistic children. J. Autism Dev. Disord. **31**(1), 11 (2001)
13. Bal, E., Harden, E., Lamb, D., Van Hecke, A.V., Denver, J.W., Porges, S.W.: Emotion recognition in children with autism spectrum disorders: relations to eye gaze and autonomic state. J. Autism Dev. Disord. **40**(3), 358–370 (2010)
14. van 't Hof, M., Tisseur, C., van BerckelearOnnes, I., van Nieuwenhuyzen, A., Daniels, A.M., Deen, M., Hoek, H.W., Ester, W.A.: Age at autism spectrum disorder diagnosis: a systematic review and meta-analysis from 2012 to 2019. Autism SAGE J. **25**(4), 862–873. Article first published online: 19 Nov 2020; Issue published: 1 May 2021
15. Weeks, S.J., Hobson, R.P.: The salience of facial expression for autistic children. J. Child Psychol. Psychiatry **28**(1), 137–152 (1987)
16. Hobson, R.P.: The autistic child's appraisal of expressions of emotion: a further study. J. Child Psychol. Psychiatry **27**(5), 671–680 (1986)
17. Nobile, M., Perego, P., Piccinini, L., Mani, E., Rossi, A., Bellina, M., Molteni, M.: Further evidence of complex motor dysfunction in drug naive children with autism using automatic motion analysis of gait. Autism **15**(3), 263–283 (2011)
18. Mache, M.A., Todd, T.A.: Gross motor skills are related to postural stability and age in children with autism spectrum disorder. Res. Autism Spectr. Disord. **23**, 179–187 (2016)
19. Calhoun, M., Longworth, M., Chester, V.L.: Gait patterns in children with autism. Clin. Biomech. **26**(2), 200–206 (2011)
20. Yeung, S., Russakovsky, O., Mori, G., Fei-Fei, L.: End-to-end learning of action detection from frame glimpses in videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2678–2687
21. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. In: Advances in Neural Information Processing Systems, 2014, pp. 568–576
22. Mozo, A., Ordozgoiti, B., Gómez-Canaval, S.: 'Forecasting short-term data center network traffic load with convolutional neural networks.' PLoS ONE **13**(2), e0191939 (2018)
23. Zhang, C., Zhang, H., Yuan, D., Zhang, M.: 'Citywide cellular traffic prediction based on densely connected convolutional neural networks.' IEEE Commun. Lett. **22**(8), 1656–1659 (2018)
24. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: a new learning scheme of feedforward neural networks. In: Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004, vol. 2, pp. 985–990
25. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing **70**, 489–501 (2006)

26. Li, G., Lee, O., Rabitz, H.: High efficiency classification of children with autism spectrum disorder. PLoS ONE **13**(2), e0192867 (2018). https://doi.org/10.1371/journal.pone.0192867
27. Bi, X., Wang, Y., Shu, Q., Sun, Q., Xu, Q.: Classification of autism spectrum disorder using random support vector machine cluster. Front. Genet. **9**, 18 (2018). https://doi.org/10.3389/fgene.2018.00018
28. Alarifi, H.S., Young, G.S.: Using multiple machine learning algorithms to predict autism in children. In: International Conference on Artificial Intelligence, ICAI'18
29. Ganapathi Raju, N.V., Madhavi, K., Sravan Kumar, G., Vijender Reddy, G., Latha, K., Lakshmi Sushma, K.: Prognostication of autism spectrum disorder (A.S.D.) using supervised machine learning models. Int. J. Eng. Adv. Technol. **8**(4), 2249–8958 (2019)
30. Omar, K.S., Mondal, P., Khan, N.S., Rizvi, M.R.K., Islam, M.N.: A machine learning approach to predict autism spectrum disorder. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox' Bazar, Bangladesh, 2019, pp. 1–6. https://doi.org/10.1109/ECACE.2019.8679454
31. Hyde, K.K., Novack, M.N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D.R., Linstead, E.: Applications of supervised machine learning in autism spectrum disorder research: a review. Rev. J. Autism Dev. Disord. **6**, 128–146 (2019)
32. Thabtah, F.: Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment (2017). https://doi.org/10.1145/3107514.3107515
33. Gu, J., Wang, Z., Kuen, J., et al.: Recent advances in convolutional neural networks. Pattern Recognit. **77**, 354–377 (2018). View at: Publisher Site | Google Scholar
34. Huang, G., Huang, G.-B., Song, S., You, K.: Trends in extreme learning machines: a review. Neural Netw. **61**, 32–48 (2015)
35. Huang, G.-B., Chen, L.: Convex incremental extreme learning machine. Neurocomputing **70**(16–18), 3056–3062 (2007)
36. Xie, S., Hu, H.: Facial expression recognition with FRR-CNN. Electron. Lett. **53**(4), 235–237 (2017). View at: Publisher Site | Google Scholar
37. Li, J., Zhang, D., Zhang, J., et al.: Facial expression recognition with faster R-CNN. Procedia Comput. Sci. **107**(C), 135–140 (2017). View at: Publisher Site | Google Scholar
38. Whittle, M.W.: Whittle's Gait Analysis, 5th edn., p. 30. Elsevier, Amsterdam (2012)
39. Major, M.J., Raghavan, P., Gard, S.: Assessing a low-cost accelerometer-based technique to estimate spatial gait parameters of lower-limb prosthesis users. Prosthet. Orthot. Int. **40**(5), 643–648 (2015). https://doi.org/10.1177/0309364614568411
40. Camicioli, R., Howieson, D., Lehman, S., Kaye, J.: Talking while walking: the effect of a dual task in aging and Alzheimer's disease. Neurology **48**(4), 955–958 (1997). https://doi.org/10.1212/WNL.48.4.955
41. Llinas, J., Hall, D.L.: An introduction to multi-sensor data fusion. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), May 1998, vol. 6, pp. 537–540. https://doi.org/10.1109/ISCAS.1998.705329
42. Ding, Z., et al.: The real time gait phase detection based on long short-term memory. In: Proceedings of the IEEE 3rd International Conference on Data Science in Cyberspace (D.S.C.), June 2018, pp. 33–38. https://doi.org/10.1109/DSC.2018.00014
43. Vu, H.T.T., Gomez, F., Cherelle, P., Lefeber, D., Nowé, A., Vanderborght, B.: ED-FNN: a new deep learning algorithm to detect percentage of the gait cycle for powered prostheses. Sensors **18**(7), 2389 (2018). https://doi.org/10.3390/s18072389
44. Mazumder, O., Sankar, A., Kumar Lenka, K.P., Bhaumik, S.: Multichannel fusion based adaptive gait trajectory generation using wearable sensors. J. Intell. Robot. Syst. **86**(3–4), 335–351 (2016). https://doi.org/10.1007/s10846-016-0436-y
45. Mun, K.R., Song, G., Chun, S., Kim, J.: Gait estimation from anatomical foot parameters measured by a foot feature measurement system using a deep neural network model. Nature **8**, 9879 (2018). https://doi.org/10.1038/s41598-018-28222-2