Soft Computing (2024) 28:6695–6708
https://doi.org/10.1007/s00500-023-09477-y

APPLICATION OF SOFT COMPUTING

Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children

Fatma M. Talaat[1,2,3] · Zainab H. Ali[4,5] · Reham R. Mostafa[6,7] · Nora El-Rashidy[1]

Accepted: 14 November 2023 / Published online: 4 January 2024

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that is characterized by abnormalities in the brain, leading to difficulties in social interaction and communication, as well as learning and attention. Early diagnosis of ASD is challenging as it mainly relies on detecting abnormalities in brain function, which may not be evident in the early stages of the disorder. Facial expression analysis has shown promise as an alternative and efficient solution for early diagnosis of ASD, as children with ASD often exhibit distinctive patterns that differentiate them from typically developing children. Assistive technology has emerged as a crucial tool in improving the quality of life for individuals with ASD. In this study, we developed a real-time emotion identification system to detect the emotions of autistic children in case of pain or anger. The emotion recognition system consists of three stages: face identification, facial feature extraction, and feature categorization. The proposed system can detect six facial emotions: anger, fear, joy, natural, sadness, and surprise. To achieve high-performance accuracy in classifying the input image efficiently, we proposed a deep convolutional neural network (DCNN) architecture for facial expression recognition. An autoencoder was used for feature extraction and feature selection, and a pre-trained model (ResNet, MobileNet, and Xception) was applied due to the size of the dataset. The Xception model achieved the highest performance, with an accuracy of 0.9523%, sensitivity of 0.932, specificity of 0.9421, and AUC of 0.9134%. The proposed emotion detection framework leverages fog and IoT technologies to reduce latency for real-time detection with fast response and location awareness. Using fog computing is particularly useful when dealing with big data. Our study demonstrates the potential of using facial expression analysis and deep learning algorithms for real-time emotion recognition in autistic children, providing medical experts and families with a valuable tool for improving the quality of life for individuals with ASD.

Keywords Emotion recognition · Assistive technology · Autism · IoT · Deep learning

1 Introduction

This section discusses some important issues such as autism spectrum disorder, Autism-related issues with emotion recognition, and assistive technology.

1.1 Autism spectrum disorder

A neurological disorder called autism spectrum disorder (ASD) affects behavior and communication. Although Kanner was the first to recognize it in 1943 ((Kanner 1968)), our knowledge of ASD has greatly increased in terms of diagnosis and treatment. The initial indicators of this developmental syndrome can be seen in a child's early years, even though it can manifest at any age. The DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) (Australia 2015) states that a person with autism exhibits ineffective behavior, and social, and communication abilities. Even though the impairment is permanent, therapy and assistance can help a person accomplish some things more effectively. Some people with autism have problems sleeping and exhibit obnoxious behavior. Because certain symptoms in adults may overlap with other intellectual illnesses, like Attention Deficit Hyperactivity Disorder (ADHD), It is simpler to diagnose ASD in kids than in adults.

They have a consistent demeanor and exhibit no desire to interact with others. Based on the individual's

Extended author information available on the last page of the article symptoms, autism is a disorder with a wide range of symptoms. It might be modest to really severe, which is why the word "spectrum" is included in the disorder's name (Maenner et al. 2021). An autistic individual with a severe disorder is mostly nonverbal or has difficulties speaking. They have a hard time interpreting their feelings and communicating them to others. As a result, a person with Autism has difficulty executing everyday tasks. A real-time emotion identification system for autistic youngsters is a vital issue to detect their emotions to help them in case of pain or anger.

1.2 Autism-related issues with emotional recognition

Empathy is referred to as the capacity for comprehending and reciprocating the feelings of another. Sympathy, on the other hand, is the ability to share comparable feelings with another individual. People with ASD may not be able to empathize or sympathize with others (Cheng et al. 2002). When someone is harmed, they may express gladness, or they may show no emotion at all. As a result of their inability to respond correctly to others' emotions, autistic people may appear emotionless. Several studies, however, have looked into if someone with autism can genuinely express their emotions to others.

Empathy requires a careful examination of another person's body language, speech, and facial expressions in order to comprehend and interpret their sentiments. People with ASD lack the necessary social skills associated with interpreting body language and reciprocating feelings, whereas youngsters learn to understand the facial expressions needed to demonstrate empathy by seeing and emulating people around them. In people with ASD, the majority of social skills needed to engage with others are significantly hampered.

ASD is linked to a specific social and emotional deficit that is characterized by cognitive, social semiotic, and social understanding deficits. Autism typically prevents a person from understanding another person's emotions and mental state through facial expressions or speech intonation. They may also have trouble anticipating other people's actions by analyzing their emotional conditions. Facial expressions are heavily used in emotion recognition studies. The ability to recognize emotions and distinguish between distinct facial expressions is normally developed from infancy (Dollion et al. 2022).

Children with ASD frequently neglect facial expressions (Howard et al. 2021). Additionally, children with autism perceive facial expressions inconsistently, suggesting that they lack the ability to recognize emotions (Banire et al. 2021). When interpreting emotions, multiple sensory processing is frequently necessary (Conner et al. 2020). The measurements of the face, body, and speech can all be used to interpret emotion. The capacity to divide attention and concentrate on pertinent facial information is necessary for the recognition of emotions; this sort of processing is largely subconscious.

1.3 Assistive technology's role in the lives of individuals with ASD

Any device or piece of gear that allows persons with ASD to perform tasks they previously couldn't is considered assistive technology. These technology devices make it easier for people with disabilities to complete daily duties. In recent years, technology that helps people with autism has advanced significantly. From basic to advanced, this technology is diverse (O'Neill and Gillespie 2014). The primary objective of assistive technology is to benefit those with special needs. Such facilities, schools, and the government might work together to develop therapeutic spaces that use technology. Most academics concur that it is crucial to select appropriate assistive technology for people with autism in a methodical manner, depending on the severity of the disease (Knight et al. 2013).

As a result, not every person should use every type of assistive technology. Each ASD sufferer has their own unique set of traits. It is evident that there is no set of assistive technologies that are universal.

Only experts are qualified to distinguish between the differences and offer the required assistance. Assistive technology uses everything from elementary to cutting-edge computing techniques (Aresti-Bartolome and Garcia-Zapirain 2014). It can be divided into three main categories: I basic assistive technology, which refers to pictorial cards used to facilitate communication between the student and the teacher; (ii) medium assistive technology, which refers to graphical representation systems; and (iii) advanced technology, which includes applications for human–computer interaction like robots and gadgets (Anwar et al. 2010).

A wide range of tools known as assistive technology is available to help people with autism overcome (Brumfitt 1993) their functional limitations. Another technological tool used to enhance communication for people with ASD is augmented and alternative communication (AAC), which is a plan of action that could allow a nonverbal person to interact with others (Auyeung et al. 2013). Additionally, the ways for instructing such extraordinary people to study in order to enhance their life can be improved through the use of computer-based adaptive learning. This media could consist of software, hardware, or a combination of the two.

Dynamic assistive technology incorporates control apparatus, touch displays, and augmented and virtual reality applications, among other advanced technological computer gadgets that have evolved. These technologies can be used for both diagnosis and treatment. Pictures and images have piqued the interest of people with ASD (Charlop-Christy et al. 2002). They have proven to be efficient visual learners, and pictorial cards have been shown to be a successful teaching tool for children to learn how to perform daily tasks.

The aim of this study is to develop a real-time emotion recognition system based on DL CNN model. Emotion detection using face images is a challenging task. The consistency, quantity, and caliber of the photos used to train the model have a big effect on how well it works. From the images used in training, the model should be able to distinguish between different emotions. It was a challenge due to several reasons including (i) face images have different characteristics, for example, some of them have shorter faces, border faces, wider images, smallmouth, etc. (ii) some faces may indicate emotions that are different from their actual emotions. (iii) sad emotions may sometimes overlap with anger emotions and the same for joy and surprise. The results showed that the higher accuracy is for the natural class.

The remaining work is organized as follows. In Sect. 2, some of the recent related work in emotion recognition techniques is presented. Section 3 presents the problem definition. Experimental evaluation is provided in Sect. 4. And in Sect. 5, we conclude this work.

2 Literature review

A facial identification system that not only recognizes faces in a picture but also infers the type of emotion from facial features has long been investigated by academics and tech experts. Bledsoe (Rashidan et al. 2021); (Banire et al. 2021) documented some of the first research on automatic facial detection for the US Department of Defense in 1960. Since then, the software has been developed specifically for the Department of Defense, although little information about the product is given to the general public. Kanade (Ahmed et al. 2022) developed the first fully effective autonomous facial recognition system. By discriminating between features retrieved by a machine and those derived by humans, this system was able to measure 16 distinct facial features.

One of the studies (Baron-Cohen et al. 2009) demonstrates how teaching emotions to autistic children through the use of visual clues. To teach kids about various emotions and monitor their development, they produced several movies and games. Another study (Cheng et al. 2002) provided a web application that gives these gifted children a platform to interact with a simulated model. A human–computer interaction interface was also created as a result of the "AURORA" project, which employed a robot and permitted interaction between the kid and the robot (Goldsmith and LeBlanc 2004). Another study (Dautenhahn and Werry 2004) confirms the human–computer connection by demonstrating a series of brief films depicting various emotional states of children with exceptional needs. The writers of (Robins et al. 2009) look into the various possibilities for using robots as therapeutic instruments.

Despite several studies on teaching emotions to autistic children, several obstacles remain. An autistic individual has a hard time deciphering emotions from facial expressions. In a study (Kaliouby and Robinson 2005), the application of giving an emotional hearing aid to special-needs individuals was proven. The facial action coding system (FACS), introduced by Ekman and Friesian (Donato et al. 1999), is one of the approaches used to recognize facial expressions. Depending on the facial muscular activity, the facial action coding system depicts several sorts of facial expressions.

This method allows for the quantitative measurement and recording of facial expressions. Facial recognition systems have advanced significantly in tandem with advances in real-time machine learning techniques. The study (Pantic and Rothkrantz 2000) presents a detailed review of current automatic facial recognition technologies and applications. According to the research, the majority of existing systems recognize either the six fundamental facial expressions or different forms of facial expressions.

When it comes to recognizing emotions, emotional intelligence is crucial. Understanding emotions entails biological and physical processes as well as the ability to recognize other people's feelings (Staff et al. 2022). By watching facial expressions and somatic changes and converting these documented changes to their physiological presentation, an individual can reliably predict emotions. According to Darwin's research, the process of detecting emotions is thought to entail numerous models of behavior, resulting in a thorough classification of 40 emotional states (Magdin et al. 2019). On the other hand, the majority of studies on facial attribute stratification, on the other hand, refer to Ekman's classification of six primary emotions (Batty and Taylor 2003). Joy, sadness, surprise, fear, disgust, and anger are the six fundamental emotions. However, six other fundamental emotions were later added to the neutral expression.

The ease with which various groupings of emotions may be identified is a big advantage of using this approach. A variety of computer-based technologies have been developed to better read human attitudes and feelings in order to improve the user experience (Leony et al. 2013). To anticipate meaningful human facial expressions, they mostly use cameras or webcams. With moderate accuracy, one can deduce the emotions of another person who is facing the camera or webcam. Meanwhile, several machine-learning and image-processing experiments have shown that face traits and eye-gazing behaviors may be used to identify human moods (Lakshminarayanan et al. 2017). The Facial Action Coding Method (FACS) is a classification system for facial impressions based on facial affectation. It was first proposed in 1978 by Ekman and Friesen, and it was modified in 2002 by Hager (Lakshminarayanan et al. 2017).

Several studies depend on face images. For example (Wells et al. 2016), R. Sadik et al. utilize transfer learning in emotion recognition. The proposed model is implemented using CNN to develop the MobileNet model. The experimental result showed the recognition model achieved 89%, and 87% in terms of accuracy and F1 score, respectively. The same in (Ahmed et al. 2022), they utilized three pre-trained models include (MobileNet, Xception, and Inception V3) to detect autism based on facial features. Given the accuracy of 95%, 94%, and 89% for MobileNet, Xception and Inception, respectively. In another study (Akter et al. 2021), T. Akter et al. utilized an improved version of MobileNet V1.

The enhanced version has augmented additional layers to increase performance including batch normalization to normalize output, average pooling to recenter and rescale the input, and two fully connected layers before the output layer. The improved model achieved 90.67% in terms of classification accuracy. Other studies analyze the facial image of autistic children for other purposes. For example, in (Banire et al. 2021), build DL to recognize attention from the facial analysis. They achieved performance of 88.89% and 53.1% in terms of ACC and AUC, respectively.

Regarding children with Autism. Various studies used DL and CNN to diagnose Autism based on facial analysis. For example in (Beary et al. 2020), M. Beary et al., introduce DL model to classify children as either normal or potentially autistic. Authors utilized a pre-trained model (MobileNet) and achieved an accuracy reached 94.6%. In (Nagy et al. 2021), E. Nagy et al. compare the accuracy of the response of six emotions include (neutral, sad, disgust, anger, fear, and surprise) for normal and Autism children under non-timed and timed conditions. The result showed children with Autism are less accurate in identifying surprise and anger when compared to normal children. More extensive studies that detail emotion recognition among Autism could be found in the following reviews (Rashidan et al. 2021); Harms et al. (Harms et al. 2010).

2.1 Problem definition

This section proposes a real-time emotion identification system for autistic youngsters. Face identification, facial feature extraction, and feature categorization are the three stages of emotion recognition. A total of six facial emotions are detected by the propound system: anger, fear, joy, natural, sadness, and surprise. This research presents a Deep Convolutional Neural Network (DCNN) architecture for facial expression recognition. The suggested architecture outperforms earlier convolutional neural network-based algorithms and does not require any hand-crafted feature extraction.

2.2 Proposed emotion detection framework

The proposed emotion detection framework is based on three layers which are: (i) Cloud Layer, (ii) Fog Layer, and (iii) IoT Layer. The two main layers are (i) IoT and (ii) Fog. In IoT, there is the proposed assistant mobile application which is used to capture an image of the child while using the smart device and then sends this face image to the fog layer. In the Fog layer, there is a controller which is the fog server (FS) responsible for getting this face image and detecting the emotion depending on the proposed DL technique. In the Fog layer, there is also a database (DB) containing the pre-trained dataset. After detecting the emotion, the fog server sends an alert message to the parent device in case of detected emotion is anger, fear, sadness, or surprise.

The main controlling and managing is implemented at the controller at the fog layer to reduce the latency for real-time detection with fast response and to be a location awareness. The overall proposed framework is shown in Fig. 1.

2.3 Cache replacement strategy using fuzzy

While there are a huge number of captured images that will be sent to the fog every second, there will be a problem in the space of the cache memory of the fog server. Hence, we should use an efficient cache replacement technique to free the memory from the old captured images after a specific time. The fog server contains a table containing data about each captured image as shown in Table 1 such as (i) Image ID: which is a sequence number, (ii) Arrival time, (iii) Current Time, (iv) TTL: Time To Live which is calculated from the arrival time and current time. (v) Priority depends on the location of the capturing of the image. The priority is high if the captured image is taken at a remote location from the location of the parent. The priority is low if the captured image was taken at the same location as the

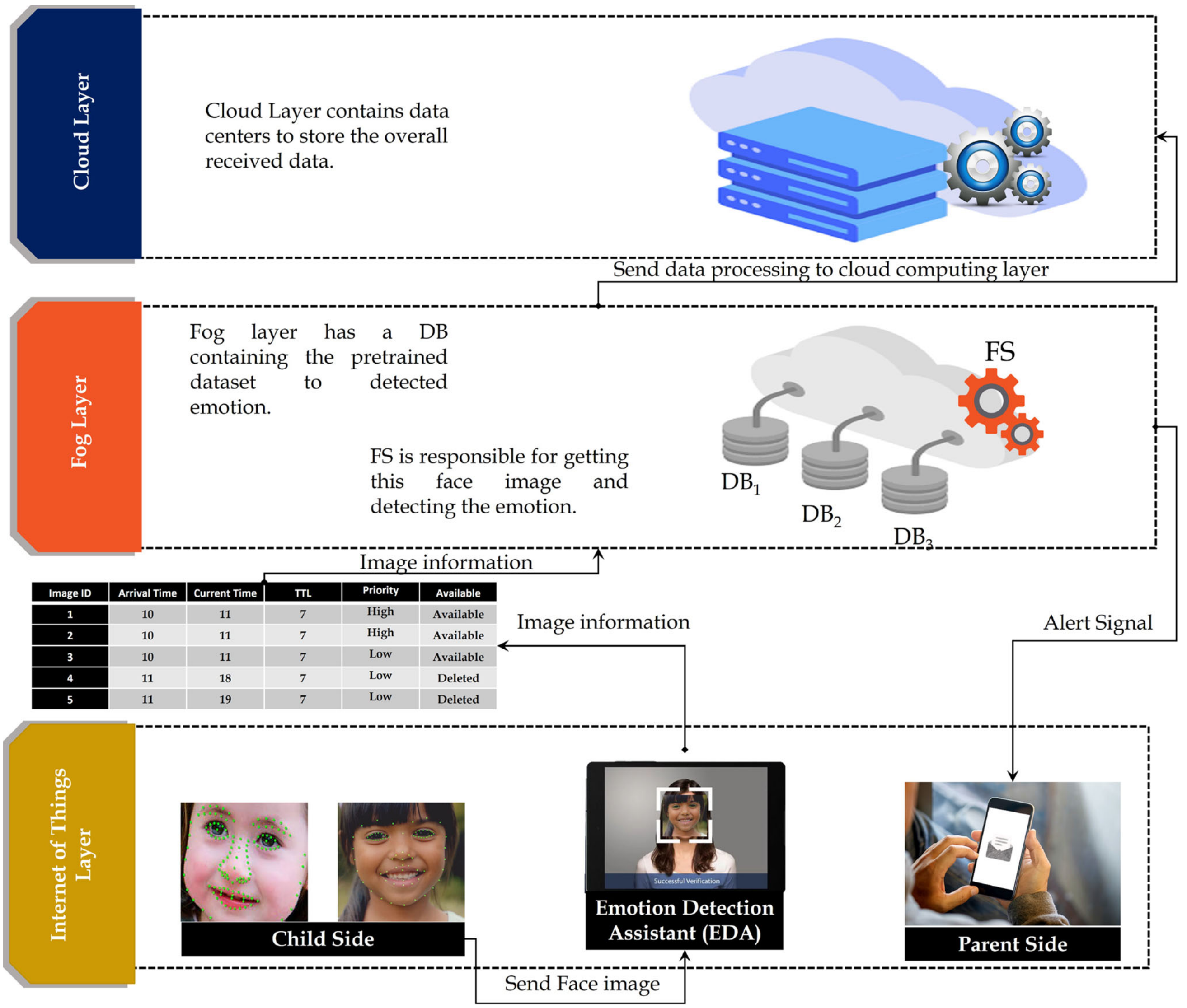

Cloud Layer
Cloud Layer contains data centers to store the overall received data.
Send data processing to cloud computing layer
Fog Layer
Fog layer has a DB containing the pretrained dataset to detected emotion.
FS is responsible for getting this face image and detecting the emotion.
FS
DB$_1$
DB$_2$
DB$_3$
Image information
Image information
Alert Signal
Internet of Things Layer
Child Side
Successful Verification
Emotion Detection Assistant (EDA)
Parent Side
Send Face image

Image ID	Arrival Time	Current Time	TTL	Priority	Available
1	10	11	7	High	Available
2	10	11	7	High	Available
3	10	11	7	Low	Available
4	11	18	7	Low	Deleted
5	11	19	7	Low	Deleted

Fig. 1 Proposed emotion detection framework

Table 1 Images information table

Image ID	Arrival time	Current time	TTL	Priority	Available
1	10	11	7	High	Available
2	10	11	7	High	Available
3	10	11	7	Low	Available
4	11	18	7	Low	Deleted
5	11	19	7	Low	Deleted

parent. (vi) Available: to decide whether the image is available or deleted.

Fuzzy logic is used to determine if an image should be kept around for a while longer or should be removed. In contrast to computationally exact systems, the reasoning process is frequently straightforward, saving processing power (Ranjan and Prasad (Ranjan and Prasad 2018)). Particularly for real-time systems, this is a really intriguing aspect. Typically, fuzzy approaches require less time to design than traditional ones.

The following consecutive steps are used to carry out the fuzzy inference process: (i) Input fuzzification, (ii) Applying fuzzy rules, and (iii) Defuzzification. The fuzzy procedure depicted in Fig. 2 illustrates those processes.

Each image is ranked based on the following three characteristics: (i) Arrival time (AT), (ii) Time To Live (TTL), which are determined by the arrival time and the current time, (iii) Priority (P): This factor is determined by where the photograph was taken. Each photo receives a Ranking (R) value from the Fuzzy by taking into account its three preset attributes (AT, TTL, and P). The fuzzy process takes into account each of those characteristics.

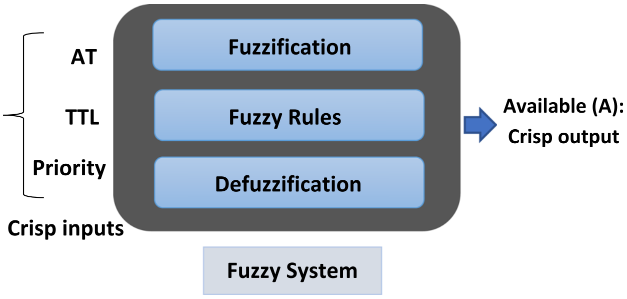


Fig. 2 The steps of the fuzzy process

The ranking is quickly and accurately determined by a fuzzy algorithm. Insofar as they are less sensitive to shifting settings and incorrect or forgotten rules, fuzzy algorithms are frequently robust.

i. Arrival Time (AT): has values: Early, Medium, or Late.
ii. Time To Live (TTL): Small, Medium, or Large.
iii. Priority (P): Low, Medium, or High.

The rating value is R1, meaning that the photo is significant and will require additional time (15 min). If a photo has a rating value of R2, it has a low priority and can be deleted to make room for another.

i. ***Fuzzified inputs***

The considered three features for each image are fuzzified. The Fuzzified Arrival Time (FAT), Fuzzified Time to Live (FTTL), and Fuzzified Priority (FP) is used to predict the value of R. They are calculated as shown in Eqs. (1), (2), and (3).

$$FAT = \{(\mathrm{AT},\mu\mathrm{FAT(AT)})/AT \in T\} \quad \mu\mathrm{FAT(AT)} \ \in [0,1] \tag{1}$$

where, AT = {Early, Medium, Late}, T = [0,100]. FAT: Fuzzified Arrival Time.

$$\mathrm{FTTL} = \{\ (\mathrm{TTL},\mu\mathrm{FTTL(TTL)})/\mathrm{TTL} \in \mathrm{TL}\} \quad \mu\mathrm{FTTL(TTL)} \ \in [0,1] \tag{2}$$

where, TTL = {Small, Medium, Large}, TL = [0,100]. FTTL: Fuzzified Time To Live.

$$\mathrm{FP} = \{\ (\mathrm{P},\mu\mathrm{FP(P)})/\mathrm{P} \in \mathrm{p}\} \quad \mu\mathrm{FP(P)} \ \in [0,1] \tag{3}$$

where, P = {Low, Medium, High}, p = [0,100]. FP: Fuzzified Priority.

ii. ***Applying regulations***

The rules in this stage explain the connection between the output and the specified input variables (AT, TTL, and P) (R). The fuzzy language rules are founded on IF–THEN statements like these:

If TTL is small and P is low and AT is early THEN R is R2
If P is high THEN R is R1

iii. ***Crips values***

It will be chosen whether to delete the image or keep it for longer than was originally determined based on the fuzzy output value (R).

According to the output value from fuzzy, it will be decided to delete the image or remain it for more extra time which is predetermined previously.

2.4 Proposed application

The proposed application can be implemented in any smart device such as smartphones, tablets, etc. When the application captures a photo of the kid, it can detect his feeling as shown in Fig. 3. The Emotion Detection Assistant (EDA) Application can be active in the background while the child uses another application. EDA is used to detect the emotion of the kid. If the detected emotion is natural or joy, then there is no problem. If the detected emotion is anger, fear, sadness, or surprise, it will send an alert signal to a connected application installed on the parent's device.

In normal, a child with autism has not had enough ability to express his feeling. Hence, EDA is very useful and essential to help his parent to be notified when he is not okay. The child with autism has no ability to ask for help. There are two main sides in this system as shown in Fig. 4: (i) the Child Side, and (ii) the Parent Side.

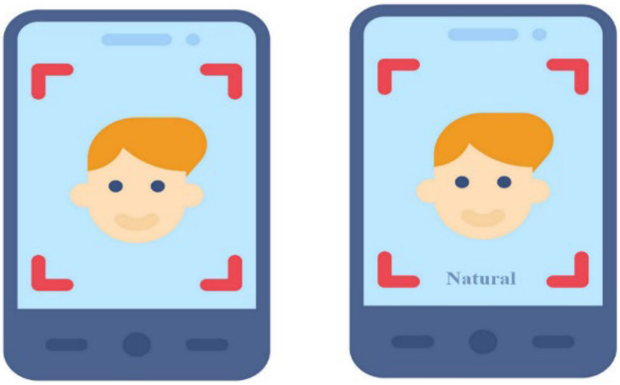


Fig. 3 Emotion Detection Assistant (EDA) application

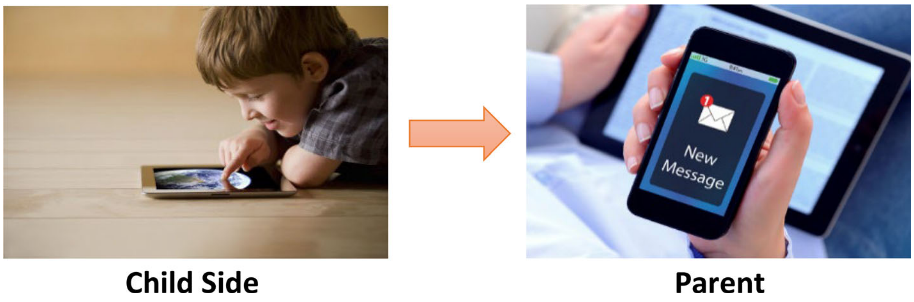


Fig. 4 Emotion Detection Assistant (EDA) system

2.5 Proposed DL framework

To improve the performance of the algorithm to classify the input image efficiently, the proposed algorithm contains an autoencoder for feature extraction and feature selection Fig. 5.

2.5.1 Autoencoder for feature extraction

Autoencoder is a type of unsupervised neural network that attempts to make a compressed representation of the input data. The design of the autoencoder restricts the architecture to a bottleneck from which it can reconstruct the image. It is utilized to reduce the dimension of the dataset when the relation between the independent and dependent datasets could be described using a nonlinear relationship. Autoencoders are considered the most promising tool for feature extraction that is used for various applications including self-driving cars, speech recognition, etc.

As shown in Fig. 6, autoencoder architecture consists of three main parts: the encoder, the bottleneck, and the decoder. First: the encoder tries to pick the most significant features to form the input data. While the decoder part tries

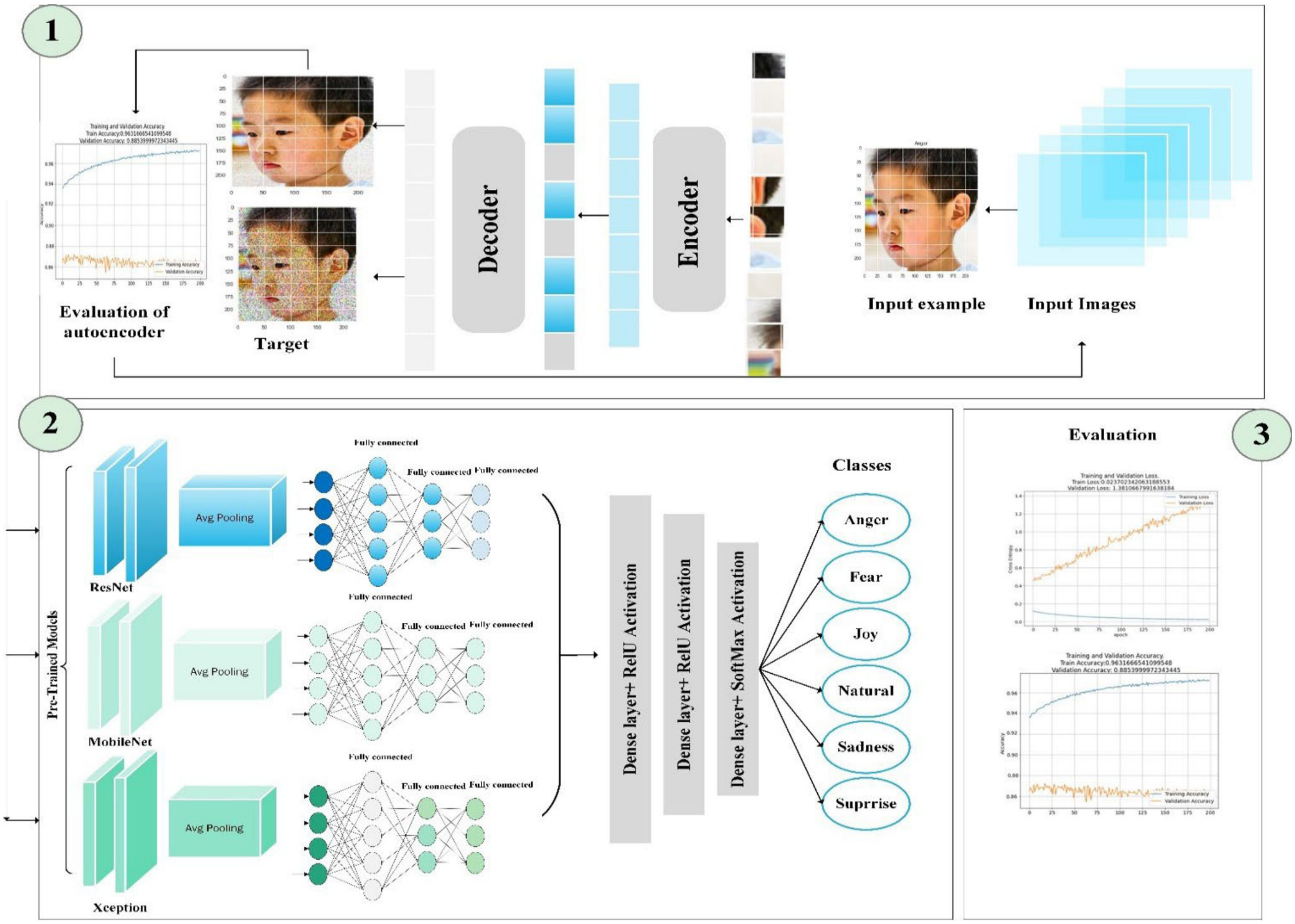

1
Training and Validation Accuracy
Train Accuracy:0.9631666541099548
Validation Accuracy: 0.8853999972343445
Evaluation of autoencoder
Target
Decoder
Encoder
Input example
Input Images
2
Pre-Trained Models
ResNet
MobileNet
Xception
Avg Pooling
Fully connected
Dense layer+ RelU Activation
Dense layer+ RelU Activation
Dense layer+ SoftMax Activation
Classes
Anger
Fear
Joy
Natural
Sadness
Suprrise
Evaluation
3
Training and Validation Loss
Train Loss:0.023702342063188553
Validation Loss: 1.3810667991638184
Training Loss
Validation Loss
Cross Entropy
epoch
Training and Validation Accuracy
Train Accuracy:0.9631666541099548
Validation Accuracy: 0.8853999972343445
Accuracy
Training Accuracy
Validation Accuracy

Fig. 5 The proposed DL framework

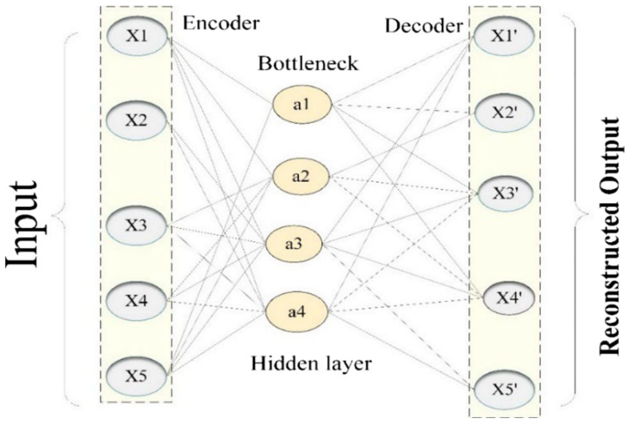


Fig. 6 Autoencoder architecture

to reconstruct the original inputs using the crucial feature. Autoencoders reduce the data dimension by keeping the features required to reconstruct the data. The output of the autoencoders is considered the same as inputs, but with some loss. Thus, sometimes, it is called lossy compression.

Considering X as a data sample n is the number of samples and m number of features and Y the encoder output (the reduced dimension of X). The decoder then tries to reconstruct X from Y. The main goal is to reduce the difference between the original input X and the reconstructed input $X\prime$.

Encoder function that maps between X and Y formulated as follows

$$Y = f(x) = S_F(WX + b_x), \tag{4}$$

where S_F is the activation function.

The decoder function maps the representation Y back to reconstruct image X

$$X\prime = g(Y) = S_g\left(W\prime Y + b_y\right), \tag{5}$$

where S_g is the decoder activation function (i.e. Sigmod, Softmax). Training the autoencoder mainly depends on finding the parameters of $(Wandb_x)$ that could minimize the recreation loss

$$\theta = min_\theta L(X, X\prime) = min_\theta(X, gF(x)) \tag{6}$$

2.5.2 Autoencoder for feature selection

In this section, a spare autoencoder is used for feature selection. The difference in sparse autoencoder is that it includes sparsity constraint on the hidden unit. This constraint is utilized to achieve a bottleneck by applying a penalty to the neurons. It ensures that the critical features are only activated. Thus, it forces the model to learn small and unique statistical features of the data. The average activation of the neuron in the hidden layer is calculated by the following equation.

$$\widehat{P}_j = \frac{1}{m}\sum_{i=1}^{m}\left[a_j^{(2)}(x)^{(i)}\right], \tag{7}$$

where $a_j^{(2)}(x)^{(i)}$ is the activation of neuron j in layer 2

$$\widehat{P}_j = P. \tag{8}$$

The sparsity in the model could be imposed using L1 regularization.

3 Implementation and evaluation

3.1 Used dataset

This paper uses a set of cleaned images for autistic children with different emotions (“Dataset link”. xxxx). The duplicated images and the stock images have been removed. Then dataset have been categorized into six facial emotions: anger, fear, joy, natural, sadness, and surprise. The six primary used emotions are shown in Fig. 7.

This paper used 758 images for training (Anger: 67, Fear:30, Joy: 350, Natural: 48, Sadness: 200, Surprise: 63) and 72 images for testing (Anger: 3, Fear:3, Joy: 42, Natural: 7, Sadness: 14, Surprise: 6).

3.2 Autoencoder for feature extraction and selection

The autoencoder model was created as a sequential model that sequentially adds one layer and deepens the network. Each layer feeds the output to the next layer. Starting with the input layer that takes the image dimension (width and height) and the code size. Then, the flatten layer is used to flatten the image matrix to a 1D array. The Dense layer tries to find the optimal features to ensure achieving the optimal output. L1 regularization utilized to achieve sparsity.

The decoder is also a sequential model that accepts the input from the encoder and then tries to reconstruct the input image in the form of one row. Then, it stacks through a dense layer and finally reshapes to construct the image. Figure 8 clarifies the used model for the autoencoder. The plot in Fig. 9 shows the learning curve of the autoencoder model. The curve clarifies that the model achieves a good fit in the reconstruction process.

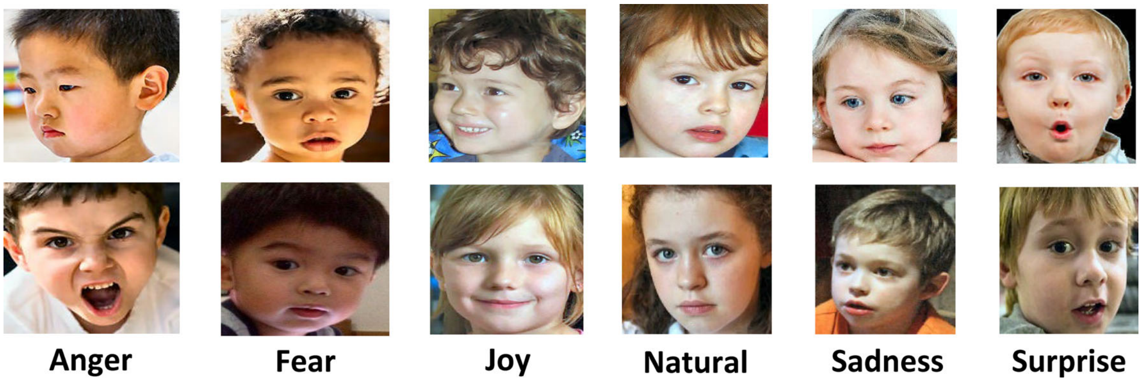


Fig. 7 The six primary used emotions

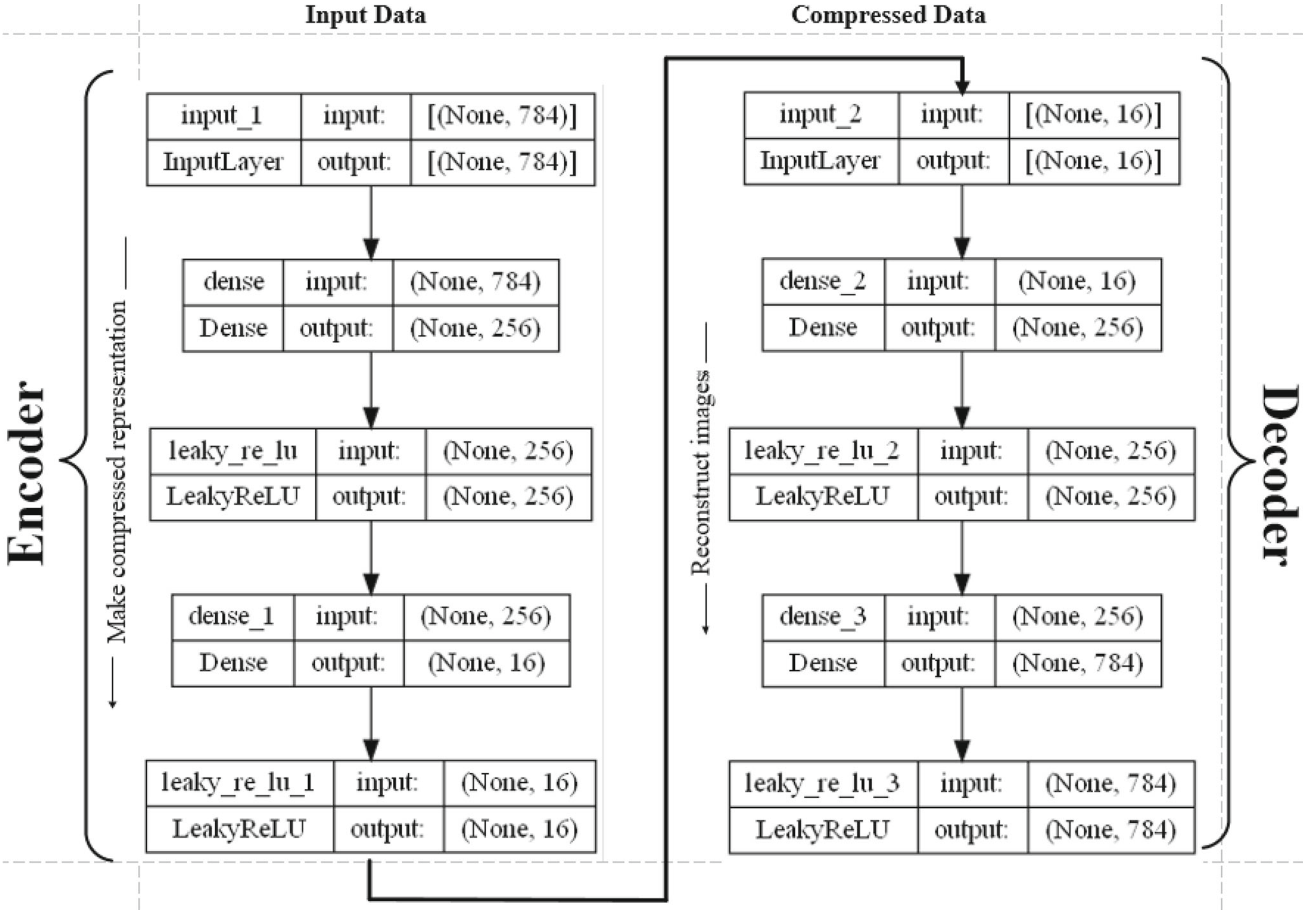

Input Data
Compressed Data
Encoder
Make compressed representation
input_1 | input: | [(None, 784)]
InputLayer | output: | [(None, 784)]
dense | input: | (None, 784)
Dense | output: | (None, 256)
leaky_re_lu | input: | (None, 256)
LeakyReLU | output: | (None, 256)
dense_1 | input: | (None, 256)
Dense | output: | (None, 16)
leaky_re_lu_1 | input: | (None, 16)
LeakyReLU | output: | (None, 16)
Reconstruct images
input_2 | input: | [(None, 16)]
InputLayer | output: | [(None, 16)]
dense_2 | input: | (None, 16)
Dense | output: | (None, 256)
leaky_re_lu_2 | input: | (None, 256)
LeakyReLU | output: | (None, 256)
dense_3 | input: | (None, 256)
Dense | output: | (None, 784)
leaky_re_lu_3 | input: | (None, 784)
LeakyReLU | output: | (None, 784)
Decoder

Fig. 8 Sparse autoencoder model architecture

3.3 The deep learning model for classification task

Several metrics are used to evaluate the model performance including the following (i) Accuracy, (ii) Precision: calculate the proportion of samples that have been accurately categorized compared to all samples. (iii) Recall: calculate the proportion of samples that are correctly categorized as members of the negative class. (iv) Cohen Kapa Cohen's kappa is a metric often used to assess the agreement between two parameters. It can also be used to evaluate a Classification model's effectiveness. where p_e is a measurement of the consistency between the model predictions and the actual class values as if it happened by random and p_0 is the model's overall accuracy.

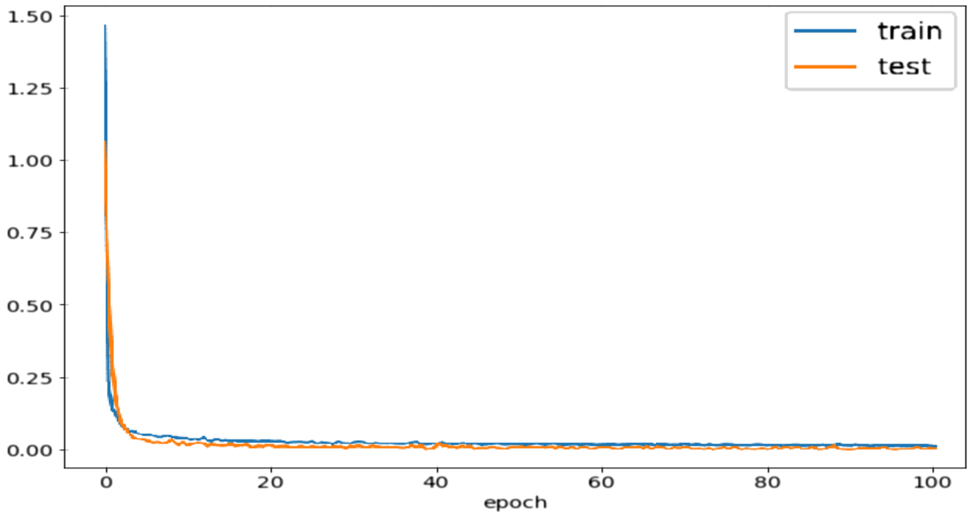


Fig. 9 Sparse autoencoder model results

(v) F1- score: calculate The harmonic mean of the precision and recall. It is considered an effective evaluation metric for unbalanced data. The area under the roc curve (AUC), is calculated based on the ROC curve. In medical images. The roc curve is preferable to accuracy. That is s because accuracy could not reflect the distribution of the prediction and which class has the height likelihood of estimation. Table 2 shows all metrics and the mathematical formulas to calculate them.

4 Results and discussion

In order to process the dataset efficiently, we employed data flow generators to divide the collected images into manageable batches, which were then fed into our proposed models: MobileNet, Xception, and ResNet. The training dataset consisted of a total of 1200 images, carefully curated to ensure diversity and representativeness, while the testing dataset comprised 220 images for evaluating the performance of the models. To provide transparency and reproducibility, we present the model hyperparameters used in our experiments in Table 3. These hyperparameters were fine-tuned through rigorous experimentation and empirical analysis. The performance evaluation of the models was conducted using various metrics, including accuracy and loss percentages, which were monitored during both the training and validation stages of the models.

Table 2 Evaluation metrics

Metric	Abbreviation	*Equation*
Accuracy	*ACC*	$\frac{tp+tn}{tp+fp+tn+fn}$
Precision	*P*	$\frac{tn}{tn+fp}$
Recall	*R*	$\frac{tn}{tn+fn}$
Cohen kapps	K	$K = \frac{p_0 - p_e}{1 - p_e}$
F1- score	F1	$\frac{2(P*R)}{P+R}$

These metrics are visualized in Figs. 10 and 11, respectively, providing insights into the model's convergence and generalization capabilities. Based on our comprehensive analysis of the results presented in Figs. 8, 9, and Table 2, we observed that the Xception model outperformed the other models in terms of accuracy and performance. It achieved an impressive accuracy of 95.23%, showcasing a substantial improvement of 7.3% compared to MobileNet and 1.8% compared to ResNet. Furthermore, the Xception model demonstrated an enhanced AUC value of 94.34, exhibiting a remarkable improvement of 3.3% over MobileNet and 2.7% over ResNet. These findings highlight the superior capabilities of the Xception model within the context of our experimental setup. The Xception architecture, with its advanced feature extraction and representation learning capabilities, proved to be highly effective in achieving higher accuracy and improved performance compared to the alternative models. This suggests that the Xception model is well-suited for the specific task at hand and holds great potential for similar image classification tasks.

The objective of our study is to develop a CNN model that could classify the emotions of ASD children based on facial expressions. A big challenge in analyzing facial expressions is the inability. To detect the crucial features, we utilized the sparse autoencoder model that includes an encoder and decoder model. The encoder transforms the input image into a feature vector and passes only the important features, then the decoder reconstructs the image from the encoder output. The output from the autoencoder is then used to classify facial emotions. Due to the size of

Table 3 Results of the pretrained models

Model	Acc	*P*	*R*	F1	*K*	AUC
MobileNet	0.8812 ± 0.0231	0.9138 ± 0.0281	0.9027 ± 0.0101	0.9002 ± 0.0256	0.813 ± 0.0264	0.892 ± 0.0312
ResNet	0.9143 ± 0.0338	0.9028 ± 0.0335	0.9311 ± 0.005	0.9162 ± 0.0227	0.8021 ± 0.0315	0.8625 ± 0.0103
Xception	0.9523 ± 0.0278	0.932 ± 0.0318	0.9421 ± 0.0134	0.9331 ± 0.0219	0.853 ± 0.0312	0.9134 ± 0.0121

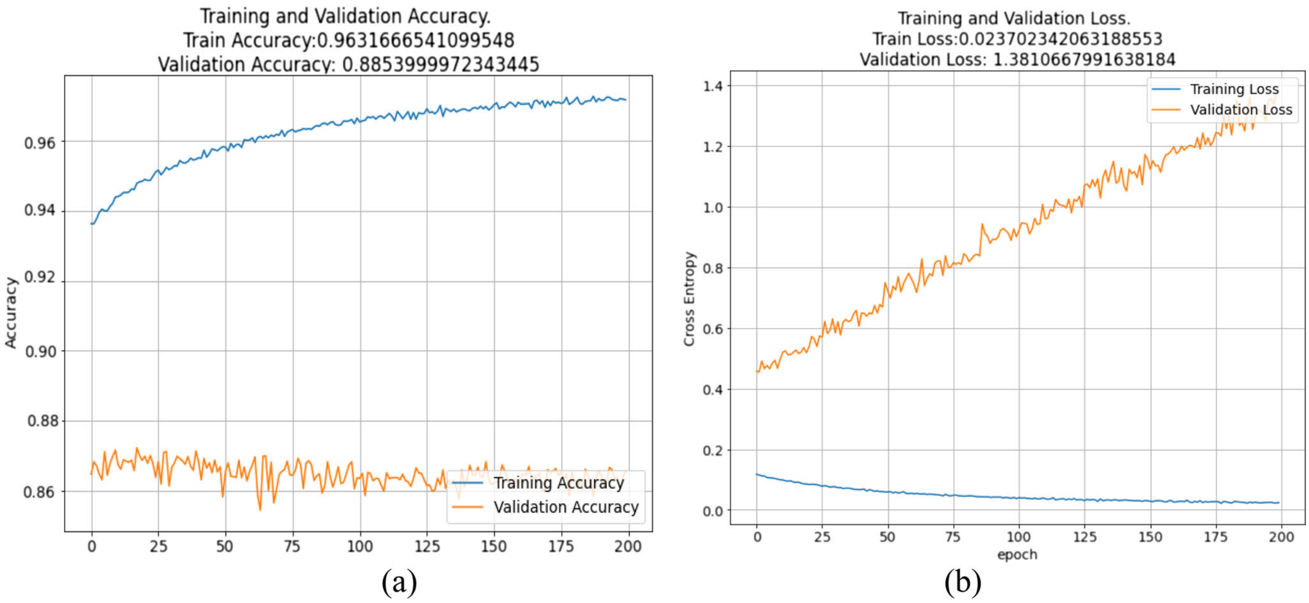
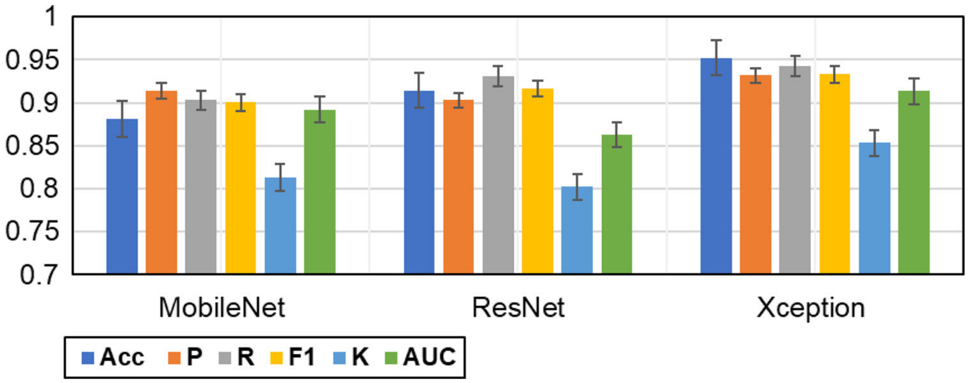


Fig. 10 Deep learning model results, **a** Training and validation accuracy, **b** training and validation loss

Fig. 11 Comparison of the pre-trained models



the used dataset, we choose to use a pre-trained model (ResNet, MobileNet, and Xception). Xception model achieved the highest performance (ACC = 0.9523%, sn = 0.932, *R* = 0.9421, and AUC = 0.9134%). To ensure the superiority of the exception model, we made the nymeni test to calculate the critical distance between the three models.

Using datasets like CIFAR and ImageNet in the pre-trained model improves the results by about (8% to 12%). Emotion detection using face images is a challenging task, the consistency, number, and quality of images used for training models have a significant impact on the model performance. From the images used in training, the model should be able to distinguish between different emotions. It was a challenge due to several reasons including (i) face images have different characteristics, for example, some of them have a shorter face, border face, wider image, or smallmouth, etc.; (ii) some faces may indicate emotions that are different from their actual emotions; (iii) sad emotions may sometimes overlap with anger emotions and the same for joy and surprise. The results showed that the higher accuracy is for the natural class.

5 Conclusion

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by abnormal brain development leading to difficulties in social interaction, communication, and learning. Medical experts face challenges in diagnosing ASD because diagnosis relies mainly on detecting abnormalities in brain function that may not appear in the early stages of the disorder. As an alternative to traditional diagnostic methods, facial expression analysis has shown promise in early detection of ASD, as children with ASD often exhibit distinctive patterns that differentiate them from typically developing children. Assistive technology has proven to be an effective tool in improving the quality of life for individuals with ASD. In this study, we developed a real-time emotion identification system for autistic children to detect their emotions in case of pain or anger. The emotion recognition system consists of three stages: face identification, facial feature extraction, and feature categorization. The proposed system can detect six facial emotions: anger, fear, joy, natural, sadness, and surprise. To enhance the performance of the algorithm in classifying the input image efficiently, we proposed a deep convolutional neural network (DCNN) architecture for facial expression recognition. An autoencoder was used for feature extraction and feature selection, and a pre-trained model (ResNet, MobileNet, and Xception) was applied due to the size of the dataset. The Xception model achieved the highest performance, with an accuracy of 0.9523%, sensitivity of 0.932, specificity of 0.9421, and AUC of 0.9134%. The proposed emotion detection framework leverages fog and IoT technologies to reduce latency for real-time detection with fast response and location awareness. Using fog computing is particularly useful when dealing with big data. Our study demonstrates the potential of using facial expression analysis and deep learning algorithms for real-time emotion recognition in autistic children, providing medical experts and families with a valuable tool for improving the quality of life for individuals with ASD. In the future, we will explore opportunities to expand our dataset in order to enhance the effectiveness and robustness of our models.

The list of abbreviations is shown in Table 4.

Table 4 List of abbreviations

Term	Abbreviations
AAC	Augmented and Alternative Communication
ASD	Autism Spectrum Disorder ASD
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
FACS	Facial Action Coding Method
FLED	Fisher Linear Discriminant analysis
PCA	Principle Component Analysis
SIFT	Scale-Invariant Feature Transform

Author contributions Single Author.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). The authors received no specific funding for this study.

Data availability FFFGG.

Declarations

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

Ethical approval There are no ethical conflicts.

References

Ahmed ZAT et al (2022) Facial features detection system to identify children with autism spectrum disorder: deep learning models. Comput Math MethOds Med 2022:3941049. https://doi.org/10.1155/2022/3941049

Akter T et al (2021) Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage. Brain Sci. https://doi.org/10.3390/brainsci11060734

Anwar A, Rahman M, Ferdous SM, Ahmed SI (2010) "Autism and Technology: An approach to new technology-based therapeutic tools A Computer Game-based Approach for Increasing Fluency in the Speech of Autistic Children," no. January 2010. doi: https://doi.org/10.1007/978-3-642-03893-8.

Aresti-Bartolome N, Garcia-Zapirain B (2014) Technologies as support tools for persons with autistic spectrum disorder: a systematic review. Int J Environ Res Public Health 11(8):7767–7802. https://doi.org/10.3390/ijerph110807767

Australia D, "Diagnostic criteria for Dementia," *Alzheimer's Aust.*, pp. 1–6, 2015, [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK56452/

Auyeung B, Baron-Cohen S (2013) Hormonal influences in typical development: implications for autism. In: Buxbaum JD (ed) San Diego. Academic Press, Elsevier, pp 215–232

Banire B, Al Thani D, Qaraqe M, Mansoor B (2021) Face-based attention recognition model for children with autism spectrum disorder. J Healthc Inform Res. 5(4):420–445. https://doi.org/10.1007/s41666-021-00101-y

Baron-Cohen S, Golan O, Ashwin E (2009) Can emotion recognition be taught to children with autism spectrum conditions? Philos Trans r Soc B Biol Sci 364(1535):3567–3574. https://doi.org/10.1098/rstb.2009.0191

Batty M, Taylor MJ (2003) Early processing of the six basic facial emotional expressions. Brain Res Cogn Brain Res 17(3):613–620. https://doi.org/10.1016/s0926-6410(03)00174-5

Beary M, Hadsell A, Messersmith R, Hosseini MP (2020) "Diagnosis of autism in children using facial analysis and deep learning," arXiv

Brumfitt S (1993) Clinical forum. Aphasiology 7(6):569–575. https://doi.org/10.1080/02687039308248631

Charlop-Christy MH, Carpenter M, Le L, LeBlanc LA, Kellet K (2002) Using the picture exchange communication system (Pecs) with children with autism: assessment of pecs acquisition, speech, social-communicative behavior, and problem behavior. J Appl Behav Anal 35(3):213–231. https://doi.org/10.1901/jaba.2002.35-213

Cheng L, Kimberly G, Orlich F (2002) "KidTalk: Online Therapy for Asperger's Syndrome,", [Online]. Available: https://pdfs.semanticscholar.org/186e/13195cb3f94dfeb8d978ed5317827ef08263.pdf

Conner CM, White SW, Scahill L, Mazefsky CA (2020) The role of emotion regulation and core autism symptoms in the experience of anxiety in autism. Autism 24(4):931–940. https://doi.org/10.1177/1362361320904217

Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy. Pragmat Cogn 12(1):1–35. https://doi.org/10.1075/pc.12.1.03dau

"Dataset link." https://www.kaggle.com/gpiosenka/autistic-children-data-set-traintestvalidate

Dollion N et al (2022) Emotion facial processing in children with autism spectrum disorder: a pilot study of the impact of service dogs. Front Psychol 13(May):1–13. https://doi.org/10.3389/fpsyg.2022.869452

Donato G, Bartlett MS, Hager JC, Ekman P, Sejnowski TJ (1999) Classifying facial actions. IEEE Trans Pattern Anal Mach Intell 21(10):974. https://doi.org/10.1109/34.799905

el Kaliouby R, Robinson P (2005) The emotional hearing aid: an assistive tool for children with Asperger syndrome. Univers Access Inf Soc 4(2):121–134. https://doi.org/10.1007/s10209-005-0119-0

Goldsmith TR, LeBlanc LA (2004) Use of technology in interventions for children with autism. J Early Intensive Behav Interv 1(2):166–178. https://doi.org/10.1037/h0100287

Harms MB, Martin A, Wallace GL (2010) Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychol Rev 20(3):290–322. https://doi.org/10.1007/s11065-010-9138-6

Howard K, Gibson J, Katsos N (2021) Parental perceptions and decisions regarding maintaining bilingualism in Autism. J Autism Dev Disord 51(1):179–192. https://doi.org/10.1007/s10803-020-04528-x

Kanner L (1968) Autistic disturbances of affective contact. Acta Paedopsychiatr 35(4):100–136

Knight V, McKissick BR, Saunders A (2013) A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. J Autism Dev Disord 43(11):2628–2648. https://doi.org/10.1007/s10803-013-1814-y

Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. Adv Neural Inf Process Sys 2017:6403–6414

Leony D, Merino P, Pardo A, Delgado-Kloos C (2013) Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2013.03.030

Maenner MJ et al (2021) Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring Network, 11 Sites, United States, 2018. MMWR Surv Summ 70(11):1–16. https://doi.org/10.15585/MMWR.SS7011A1

Magdin M, Benko L, Koprda Š (2019) A case study of facial emotion classification using affdex. Sensors 19(9):2140. https://doi.org/10.3390/s19092140

Nagy E, Prentice L, Wakeling T (2021) Atypical facial emotion recognition in children with autism spectrum disorders: exploratory analysis on the role of task demands. Perception 50(9):819–833. https://doi.org/10.1177/03010066211038154

O'Neill B, Gillespie A (2014) Assistive technology for cognition, no. January 2020.. doi: https://doi.org/10.4324/9781315779102-8

Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22(12):1424–1445. https://doi.org/10.1109/34.895976

Ranjan NM, Prasad RS (2018) LFNN: Lion fuzzy neural network-based evolutionary model for text classification using context and sense based features. Appl Soft Comput J 71:994–1008. https://doi.org/10.1016/j.asoc.2018.07.016

Rashidan MA et al (2021) Technology-assisted emotion recognition for Autism Spectrum Disorder (ASD) children: a systematic literature review. IEEE Access 9:33638–33653. https://doi.org/10.1109/ACCESS.2021.3060753

Robins B, Dautenhahn K, Dickerson P (2009) "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot." Sec Int Conf Adv Comput-Human Interact 2009:205–211. https://doi.org/10.1109/ACHI.2009.32

Staff AI, Luman M, van der Oord S, Bergwerff CE, van den Hoofdakker BJ, Oosterlaan J (2022) Facial emotion recognition impairment predicts social and emotional problems in children with (subthreshold) ADHD. Eur Child Adolesc Psychiatry 31(5):715–727. https://doi.org/10.1007/s00787-020-01709-y

Wells LJ, Gillespie SM, Rotshtein P (2016) Identification of emotional facial expressions: effects of expression, intensity, and sex on eye gaze. PLoS ONE 11(12):e0168307. https://doi.org/10.1371/journal.pone.0168307

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fatma M. Talaat[1,2,3] · Zainab H. Ali[4,5] · Reham R. Mostafa[6,7] · Nora El-Rashidy[1]

✉ Nora El-Rashidy
Noura.alrashidy@ai.kfs.edu.eg

Fatma M. Talaat
fatma.nada@ai.kfs.edu.eg

Zainab H. Ali
zainabhassan@ai.kfs.edu.eg

Reham R. Mostafa
reham_2006@mans.edu.eg

[1] Machine Learning and Information Retrieval Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt

[2] Faculty of Computer Science and Engineering, New Mansoura University, Gamasa 35712, Egypt

[3] Nile Higher Institute for Engineering and Technology, Mansoura, Egypt

[4] Embedded Network Systems and Technology Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt

[5] Department of Electronics and Computer Engineering, School of Engineering and Applied Sciences at Nile University, Giza, Egypt

[6] Research Institute of Sciences and Engineering (RISE), University of Sharjah, Sharjah 27272, United Arab Emirates

[7] Information Systems Department, Faculty of Computers and Information Sciences, Mansoura University, Mansoura 35516, Egypt