Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

CNN Based Efficient Facial Emotion Recognition System for Children

Nilofer Attar
Assistant Professor, Dept. of Computer Engineering, MKSSS'S Cummins College of Engineering for Women, Pune.
Research scholar, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune
Email : kittadnilofer@gmail.com

Dr. Shilpa Paygude
Professor, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune
Email : shilpa.paygude@mitwpu.edu.in

Abstract
An efficient and robust technique for facial expression analysis for the different human face image input is carried out. During the analysis with every dataset, it is analysed through the Convolution Neural network (CNN) method. In order to analyse the facial data in terms of emotions, at this stage of research it is considered with the three different emotions such as Happy, Sad and Anger. For the analysis of facial data for the very rarely occurring disorder autism disorder in the age group around 10 years termed as children age group. As an initial step of analysis of autism disorder emotion expressed through face is a significant symptom hence in the present research is carried out using the freely available Kaggle database to achieve the better accuracy and stability in the obtained results by the proposed technique of methodology. To show the proposed technique is better than the existing algorithms, two different database are considered and cross correlation among the database in terms of its efficiency are considered.

Keywords: Emotion, Database, CNN, Face Expression, autism disorder.

1. Introduction

A neurodevelopment disorder that can affect the ability of a person to change his behaviour, socialize, and communicate can be termed as autism spectrum disorder (ASD). A general symptom of ASD is they Lack of facial expression. At the very early stage of childhood, these patients should be taught to enact necessary emotions in their face. A methodology that helps to detect human emotion, artificially is called facial emotion recognition [1].

The ASD is normally associated with interaction disorder and social communication shortage. In such patients, it is clearly seen to have less emotions on their face. The autism patients will normally have a less emotions on their face. Hence, it is commonly difficult to observe the emotions what actually they feel. This is termed as Facial Emotion Recognition (FER). The facial emotion recognition is a function of brain. The disorder in the. Network distribution of brain leads to shortage of FER. Some theory sees the deflates of FER Rised because of more complex interaction within the brain network.

A cortical connectivity has been observed in a typical ASD disorder. A neural network is required to integrate the information in between the essential components of cognition and alert connectivity. To the core diagnostic

character of ASD, the abounded mechanism can be used to deny the characteristics of ASD [2]. One of the characteristic of this ASD is FER.

Till date, there are majority of investigation carried out based on the coherency of individual diagnosed with ASD[3-4]. The patients were analysed in the resting conditions. The conclusions were made based on the atypically connectivity caused in the ASD. The patients suffering form ASD are always hindered during the tasks, like participants groups and measures [5]. The presence of atypical connectivity of the ASD is found to be in the theta band, this is one of the major takeaways in the survey. Atypical connectivity is also found in the synchrony and coherence conditions in some of the basic investigations. Again the issue was found in the theta band (the range is between the 4 to 8 Hertz) [6-8]. Some of the basic emotions like fear, disgust, surprise, sadness, anger, fear, happy all these have been synchranised in the weeker part of the theta and delta signals. It is common to have all these emotions in the adults and adolescents [6]. The same set of week emotions are found in the ASD adults in the theta signals. The adults barely have neutral, angry and happy faces [7]. In the ASD children, the Reduced right frontal theta modulation is majorly observed, and is kind of static emotion the ASD kids have [8].

The patients are also observed with the beta and alpha frequencies, but the consistency of these frequencies is low [6-7]. The earlier investigations of the stimuli will have not been provided with the valid assessment of Ecologically FER autistic adults. There is a trajectory development of the FER in the ASD patients [9]. The ASD adults can have a more socialistic life when they develop complex emotions like jealous, guilt and intimacy etcetera.

In this paper, the convolution neural net worth is used to detect The emotions using patient's facial expression. The subjects are considered to be of children less than 10 years of age. Three different emotions are analysed using the proposed algorithm. The emotions are Happy, Sad and Anger[10].

2. Literature Survey

Wardana, et. al. [11], tells how the face of a person changes when he hears the music. In short, the influence of music on human life or on the facial expression is thoroughly studied in this paper. As we know, there are different interests for music. Based on the lyrics, the emotion of the person will change. The author explains music and emotions are too very relative things that one human being can have. The author collects different facial expression data sets to examine the changes. Daughter in his methodology, he uses python programming language to develop the framework. In the framework, it consists of a simple music player played one's favorite song. The person reaction is recorded. And it is fed to the python program. The methodology consists of facial recognition that recognizes the emotion of the human being. In this it consists of two mark the first is facial landmark detection, and the second one is action classification. Based on these predictions, the output result is predicted. The author finally finds out the effective bar emotion recognition. The authors conclude that the mouth movement can be utilized to detect the emotion. The movement of the mouth have a higher consultation in this algorithm if. The movement of the mouth is very rare or very low, then it is classified under different categories like ASD. In the research another parameter was added to interest as the quality of song.

Cai. Et.al. [12], make use of Identity free Condition Generative Adversal network, this is used identify the FE. The proposed algorithm was introduced to reduce the cost identity related facial attribute The identity related facial attributes are gender, race and the age. Different emotions where Raghavan as an important to the developed algorithm. The emotions chosen to train and test the model were discussed to come happy. Sad, surprise, angry and fear. All these six emotions were given with the different age, race and gender. The database is selected of about 1200. After resting and training the model, the results were obtained where strong, average, and less emotion range. The author was successfully able to classify the 6 emotions of a human face based on the level of emotion expressed. The performance of proposed algorithm have a very good result when compared to Existing state of art technologies.

Chen et. al., Proposed To identify the facial expression, the CNN algorithm can be used. The author uses the architecture of deep learning to extract the essential features of a facial expression. This method is completely

better than the existing traditional methods, like SVM algorithm and principal component analysis. The proposed paper user cnn and identify the facial expression of different types. The author assures to identify seven different facial expression in his study The new model of CNN is Will it rain to analyse the expressions of a human being. In the proposed methodology consists of different layers of convolutional layers, where the input image is taken with the pre defined resolution . There are multiple convolution layers and pulling layers in the recognition model. The algorithm consists of feature extraction that helps in identifying the exact emotion of the face. The author used 327 video sequence acted by 118 participants. All these facial expressions were identified in accurate manner. The author also says that the database also consisted of 230 image from Japanese female subjects. The Japanese were considered as a different race and for different phase structure. The author also included some children images to identify the same . The proposed algorithm gave a very good result when compared to the existing technologies and state of what methods. The algorithm was also faster and less cost consuming.

Chen, Wang et. al. [13], Proposes a deep learning based facial expression recognition system The author says the existing system failed to extract the representative features of the face. Hence the author proposes facial motion prayer network that helps to extract the representative features of the face. The authors says the movement official muscles is also very important to recognise the facial expression. In the proposed approach, the author uses classification network and prayer fusion network. These two are the basic building blocks of emotional recognition algorithm. The colour images are converted to grey and convoluted using facial mask motion generator. Finally, the obtained data is used to train the developed model. the algorithm shows promising results for different emotions And the results are better when compared to existing methodologies .

Xu et, al.[14], Proposed an algorithm using CNN and local binary model to develop the emotions of a given image. The proposed approach is applied on high dimensional feature images and is used to extract the features with the rotation in variance .The proposed method consists of convolution neural network structure. This network is defined as emotion net model which has three superposition structures of convolution that helps to extract different features of the face. The author uses LBP operator to rotate the image and to perform the convolution First stop the combined LBP and CNN algorithm has different levels where the Images are tested and trained with proposed algorithm. Around 1500 images were used to train and test the algorithm. The result show the proposed algorithm has a great ability to identify different emotions of the person. The author conducted the experiment at different lighting conditions of different faces, The result shows a significant improvement in the level of analysing different emotions using this algorithm.

3. Methodology

A new structure of CNN is proposed in this paper to recognise the facial expressions. The cargo database is used to carry out the experiments The face images has to undergo pre-processing to remove the non-expression features. This also helps in face image cropping and face detection. The OpenCV library is used to detect the face and also to crop. The proposed convolution neural network now finds its chance to extract the features using deep learning framework. The proposed training model uses classification of each special image. It is try to implement to recognise three different expressions from the face of ASD kids, The expressions are: Happy, sad and anger [15].

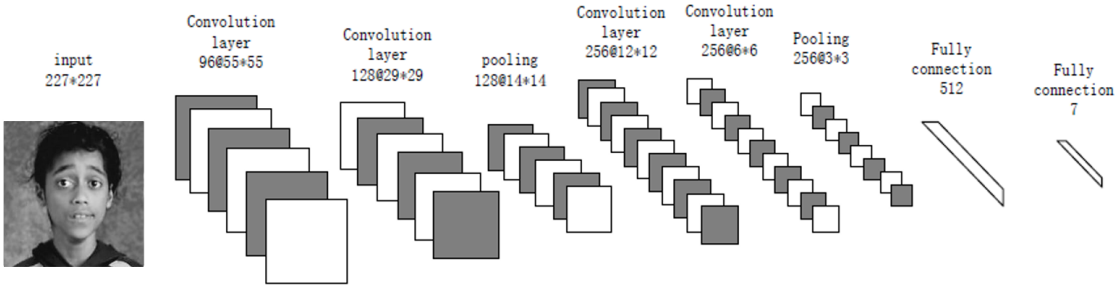

Figure 1: Facial Expression Recognition network Structure

The following are the major steps involved in the Emotion recognition of autism patients

A. CNN for Feature Extraction: the features of facial expressions are recognised and classified in the CNN network The network is designed to recognise the expressions with name as FENet [16]. Figure 1 shows the entire network that has four different convolution layers. There are two levels of cooling layers and a fully connected layer in the figure 1, it is clear that the very first layer is the convolution layer and this is the layer which is responsible for the feature extraction. In the very first layer, the kernel is trained using the convolution operation. The previous layer always consists of feature maps. And this creates a neuron formation in the current layer because of the activation function and the output . This creates a current convolution layer for the characteristic formation. The calculations of the convolution layer is given as below (as In equation (1) and (2)):

$$net_j^l = \sum_{i \in M_j} a_i^{l-1} \otimes \omega_{i,j}^l + \omega_b \quad (1)$$

$$a_{i,j}^l = F(net_{i,j}^l) \quad (2)$$

In the above expression, the weight of layer one is represented by net_j^l. The characteristic map is represented as a_i^{l-1}, With respect to the 1-1 layer. The Convolution kernel Matrix is represented by $\omega_{i,j}^l$. It is one of the connexion between the neurons. The offset terms belonging to j[th] future map is represented as ω_b. To improve the speed of network training, the value of ω_b is set to zero. At the same time. Other features and parameters are reduced. For the convolution layer 1 the feature graph is represented by $a_{i,j}^l$. The activation function is represented by F(). The total of ninety six philtres are used in the model each of dimension 11x11x4. All these layers are used to extract the low level edge features . As a first step, the images are cropped to a size of 227x227. After the first convolution, the image size is reduced to 55x55. After the very first convolution, the filter is applied to face image. The rectified linear unit (ReLU) is used to improve the nonlinearity of the network. Figure 2 represents the output After the very first convolution as in equation (3).

$$F(x) = max(0, x) \quad (3)$$

In the above equation, X is considered to be the input number of neurons. The vanishing gradient problem is eliminated using the ReLU activation function [17-18]. The vanishing gradient is a resultant of some other activation functions.

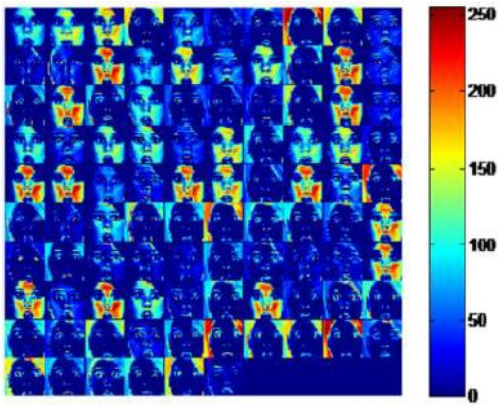

Figure 2: First Convolution Layer output.

As the number of convolution layer increases, the feature dimension also increases rapidly. This will help to avoid the dimensions of disaster, But pooling layer are used to reduce the dimensions. The Max Pooling Layer helps to down-sample the performance. The feature graphs are unaffected by the down-sampling process. This helps in removal of unnecessary information and reduce the number of parameters from the feature map. The calculations of the above set is mentioned in the equations (4) and (5).

$$net_{i,j}^l = down(a_j^{l-1}) + \omega_b \quad (4)$$

$$a_j^l = F(net_j^l) \quad (5)$$

The pool layer l has the j[th] feature map, And it is represented by a_j^l. The offset term of down-sampling layer is represented by β. The offset term of the down-sampling layer is represented with ω_b. The down sampling function is represented as down(). The convolution layer is nothing but the full connexion layer. For a full connexion. Here, the input is always defined to be one dimensional array. This makes the kernel size and the original data size similar. The below equations represent the output of each neuron:

$$net_j^l = \sum_i \omega_{i,j}^l a_j^{l-1} + \omega_b^l \quad (6)$$

$$y = F(net_j^l) \quad (7)$$

In the above equations (6) and (7), net_j^l represent the output vector. The output vector is connected to fully connected layer l. The weight coefficient matrix is represented by ω_b^l. The offset term for the fully connected layer is represented by ω_b^l. The activation function is denoted as F().

B. Pre-processing image

The GPU weighs the CNN Library. OpenCV and C++ were used to implement the algorithm. The background information is removed during the process of reprocessing. The next step is to normalise the face image to the size of 227x227 pixels [19-20]. The three represents different samples after pre-processing of images. These images are the partial outputs during the training process The initial learning rate of the algorithm is obtained as, 0.01 And the attenuation is set to, 0 .0005. The result is also a super parameter set. The figure 3 shows different emotions in ASD patients.

Figure 3. Different Expressions captured. (Sad, Happy and Angry)

4. Kaggle Database

Kaggle is Believed to be the alternative community like Google LLC. This is an online community of ML and DS practitioners. It allows users to published the data sets. It helps many to access free data set according to their requirement. This database is built to explore and build model website data science environment. This website was first launched in 2010 . It was only confined to machine learning competitions, and also offered a public data platform.

5. Results and Discussion

In this exercise, the stochastic gradient descent method, is used for the training of the model. The formulae to update the weight is as follow:

$$V_{t+1} = 0.01V_t - 0.0005\nabla L(\omega_t) \quad (8)$$

$$W_t = W_t + V_{t+1} \quad (9)$$

In the above equation, number of iteration is represented by 't'. The momentum is represented by 'V'. The error of the objective functions back propagation is represented as W_t. At the initial stage, the mean value is set to zero. The standard deviation is made as 0.01 for Gaussian distribution in random initialization. The basic parameters for every training will remain the same. In the actual training process, the weight is adjusted manually. The accuracy rate does not change after certain iterations. The weight of learning will also reduce by one by 10^{th} of its original value. The convolution neural network structure is the basis of the developed algorithm. The developed algorithm has an improved training speed and improved model accuracy. The table 1 shows the comparison of a different networks with the number of iterations. The addition of BN layer has improved the results significantly. The Figure 4, showing the Plot of different accuracies vs the network methods.

Table 1: The accuracy comparison of three different network structures

	Network		
Number of iterations	FENet with BN	Alexnet[19]	FENet
5000	0.9741	0.9158	0.8525
9000	0.98	0.9263	0.8528
18000	0.981	0.9368	0.8736
25000	0.9815	0.9438	0.8722
35000	0.9815	0.9468	0.8735

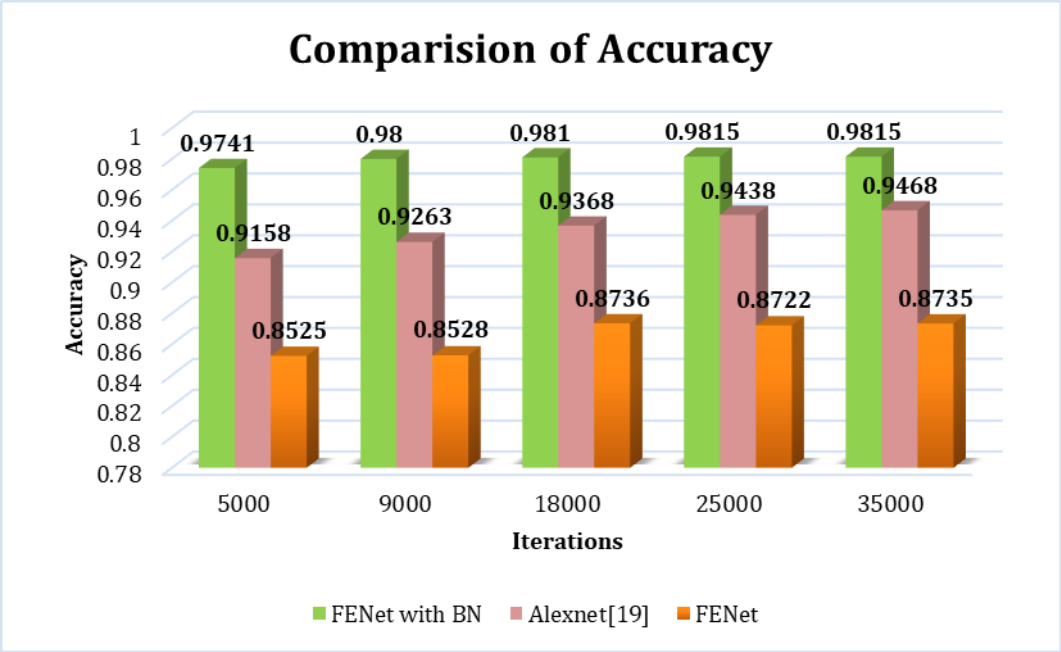

Figure 4: The Plot of different accuracies vs the network methods.

From the above expressed graph (figure 4), it is clear that, as the number of iterations increased, the accuracy also increased. There is a gradual increase in the accuracy with respect to the number of iterations. For the faster convergence, the BN layer addition gives a better result.

To evaluate the proposed methodology in still more accurate manner. Another database set is selected and is also tested on the algorithm. The cross database experiment was conducted to prove the efficiency of proposed algorithm. The JAFFE database was considered for the next part of the experiment. The Table 2, shows the cross training of the model. figure 5, showing the comparative plot of the same. The inference can be given that, the same database test and train has better results than the others.

Table 2: Cross Database Training and Testing

Train and Train	Kaggle and Kaggle	Kaggle and JAFFE	JAFFE and Kaggle
accuracy	0.991597	0.831172	0.87735

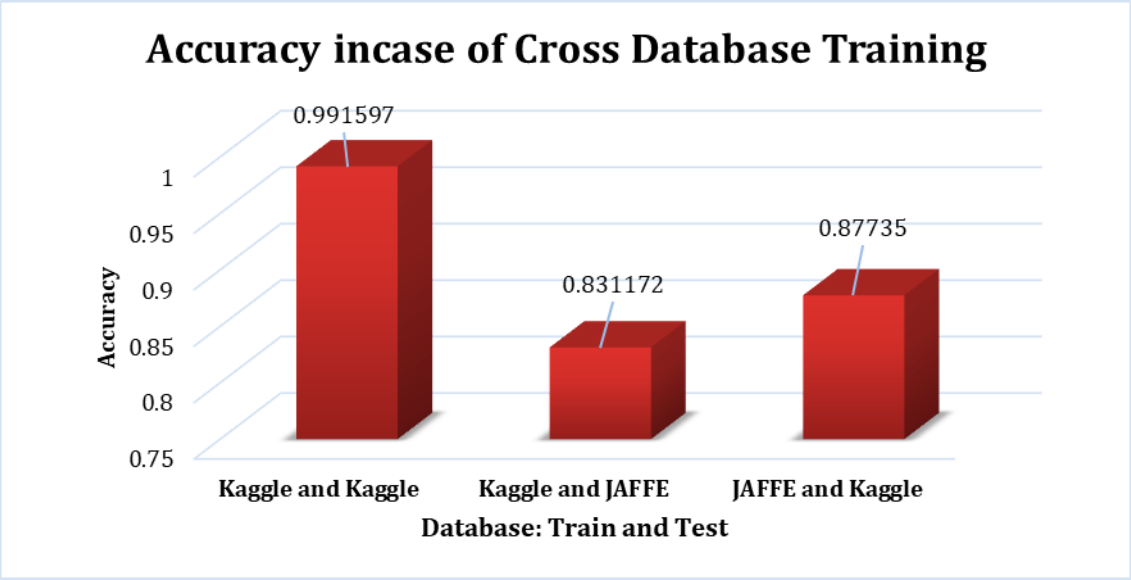

Figure 5: showing the comparative plot of Cross Database Training

The Proposed model is compared with other existing models, the models are also compared in terms of the number of layers. The comparison table is given in Table 3. The figure 6, shows the comparison chart.

Table 3: Accuracy of Different algorithms

Method	LBP+SVM(2009)	CLM+SVM(2011)	DCNN(2016)	Proposed Algorithm
Number of Layers	**6**	**6**	**7**	**7**
Accuracy	**89.10%**	**96%**	**96.02%**	**98.15%**

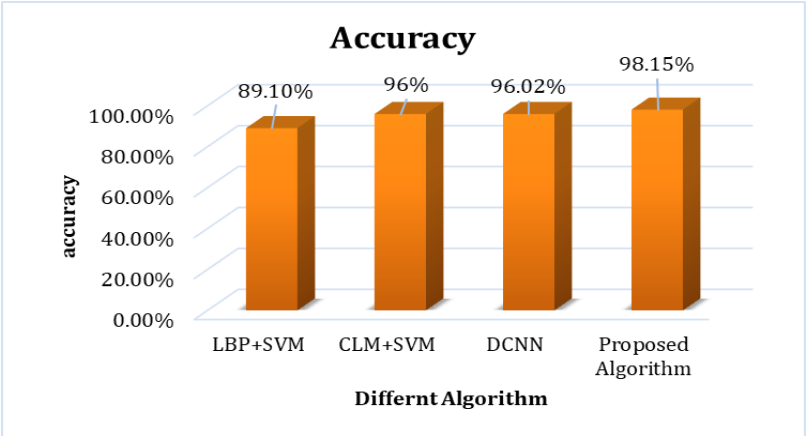

Figure 6: Accuracy plot of Proposed algorithm and others.

Facial expression recognition system

The facial expression recognition system is used in real time. It is used for everyday applications . The recognition system in this paper is proposed to construct the facial expression recognition model . The system workflow is shown in the figure y.

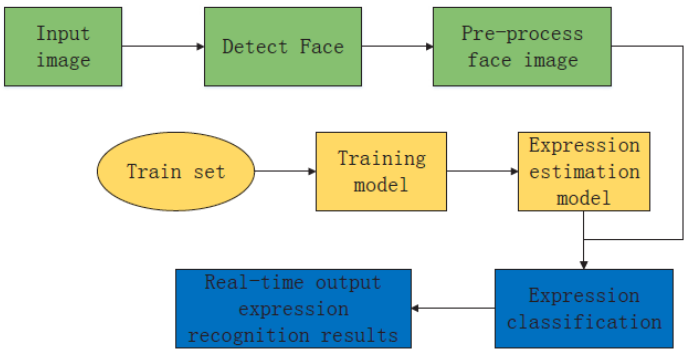

Figure 7: Face recognition system

As per the above figure 7, the pre-trained model should be loaded with the related configuration files. The Haar Classifier is used for face Detection. The area cropped is 227x227 pixels. The output recognition is instantaneously made. The results are shown below (in figure 8)

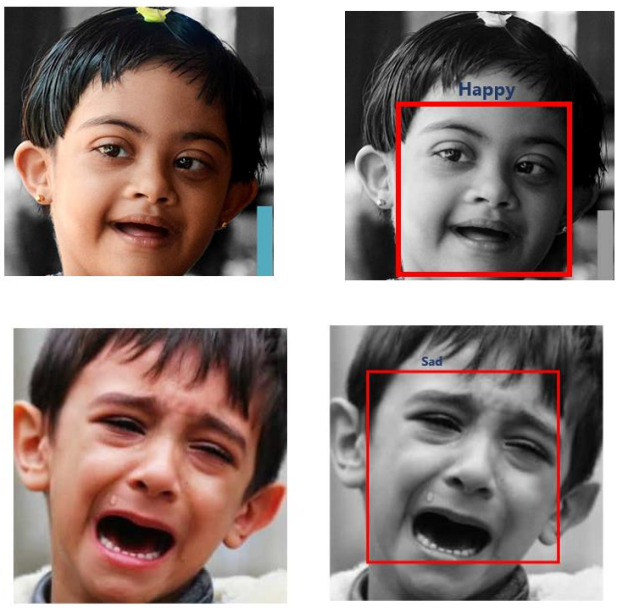

Figure 8: Identification of Happy and Sad emotions on the face of ASD kids.

Conclusion

It is carried out an effective and resilient approach for facial expression analysis for various human face picture input. During the study, each dataset is examined using the Convolution Neural Network (CNN) approach. At this level of study, it is considered with the three different emotions such as Happy, Sad, and Anger in order to assess the face data in terms of emotions. For the study of face data for the extremely uncommon condition autistic disorder in the age group of 10 years, referred to as the children age group. As an early stage in the analysis of

autism disease, emotion displayed via the face is a crucial symptom; hence, the current study is carried out utilising the publicly accessible Kaggle database and Jaffee database to reach the desired results.

References

[1] Han, J. Y., Xu, F., and Chen, Z. B. 2018. A deep learning based interactive recognition method for telephone numbers. Journal of Bjing University of Aeronautics and Astronautics. 44(05): 1074--1080. DOI= 10.13700/j.bh.1001-5965.2017.0357

[2] Jin, H., Liu, D., Gao, L. L., and Song, L. 2016. An Analysis of The NMC Horizon Report: 2016 Higher Education Edition. Journal of Distance Education. 35(02):3--10. DOI= 10.15881/j.cnki.cn33-1304/g4.2016.02.001

[3] Xu, Z. G., Zhang, G. G., and Meng, X. Z. 2019. Learners' Emotion Recognition and Its Application Based on Deep Learning. e-Education Research. 40(02):87--94. DOI= 10.13811/j.cnki.eer.2019.02.011

[4] Chen, C. 2006. An Enhanced Speech Emotion Recognition System Based on Discourse Information. 449--456. DOI= 10.1007/11758501_62.

[5] Ekman, P., and Friesen, W. 1978. Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press Palo Alto. 12.

[6] Krizhevsky, A., Sutskever, I., and Hinton, G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. NIPS (Vol.25). Curran Associates Inc. DOI= 10.1145/3065386

[7] Li, J., and Lam, E. Y. 2015. Facial expression recognition using deep neural networks. 2015 IEEE International Conference on Imaging Systems and Techniques (IST). DOI= 10.1109/IST.2015.7294547

[8] Liu, P., Han, S., and Meng, Z. 2014. Facial Expression Recognition via a Boosted Deep Belief Network. IEEE Conference on Computer Vision & Pattern Recognition. Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop. DOI= 10.1109/CVPR.2014.233

[9] Ding, H., Zhou, S. K., and Chellappa, R. 2016. FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. In Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on. IEEE, 2017, pp. 118--126. DOI= arXiv:1609.06591

[10] Yang, H. Y., Ciftci, U., and Yin, L. j. 2018. Facial Expression Recognition by De-Expression Residue Learning. International Journal on Computer ence & Engineering. 2(5):2220--2224 vol.3. DOI= 10.1007/978-94-017-0295-9_7

[11] P.Burkert,et al. DeXpression:Deep Convolutional Neural Network for Expression Recognition.2015.

[12] G.Ali, et al. Boosted NNE collections for multicultural facial expression recognition. PatternRecognition. 2016, pp.14–27.

[13] P.Lucey,et al.The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on IEEE.2010, pp. 94–101.

[14] M.Lyons,et al. Coding facial expressions with gabor wavelets. Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition.1998,pp. 200–205.

[15] Krizhevsky et al. Imagenet classification with deep convolutional neural networks.In Advances in neural information processing systems, 2012, pp . 1097–1105.

[16] Anil Kumar C, Chethan Venkatesh, Lavanya Vaishnavi D A, Harish S, “Computer vision based Hand gesture recognition system”, Neuro Quantology, Jul 2022, Volume 20, Issue 7, Page 2859-2866, DOI: 10.14704/nq.2022.20.7.NQ33365

[17] Y.Jia,et al. Caffe: Convolutional architecture for fast feature embedding. Eprint Arxiv.2014,pp. 675-678.

[18] Anil Kumar C, Lavanya Vaishnavi D A, “Renesas Based Smart ROBOT using Embedded System”, International Journal of Research in Science, Commerce, Arts, Management and Technology, Volume 2, Issue 1, February 2021, ISSN (Online) 2582-8045

[19] S.E.Kahou,et al.Combining modality specific deep neural networks for emotion recognition in video. In Proceedings of the15th ACM on International conference on multimodal interaction.2013, pp. 543-550.

[20] S.Minchul,et al. Baseline CNN structure analysis for facial expression recognition. 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) 2016, pp.26-31.