International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimal Deep Convolutional Neural Network Based Face Detection and Emotion Recognition Model

Ambika G. N. *,1 , Yeresime Suresh

Submitted: 28/04/2023 **Revised**: 25/06/2023 **Accepted**: 07/07/2023

Abstract: Face detection and emotion recognition are two closely connected tasks in computer vision that include analysing facial images to identify faces and detect the emotions expressed by the individual. Face detection is the way of localizing and locating faces within image or video frames. The objective is to detect the presence and position of faces, by drawing bounding boxes around them. Facial emotion recognition (FER) aims to detect and classify the emotions expressed by individuals based on facial expressions. Typically, this task can be done after face detection, where the faces detected are analysed further for emotional cues. Emotion recognition can be advanced by means of classical deep learning (DL) or machine learning (ML) techniques. Contemporary research on emotion classification has accomplished grand performance over DL based approaches. This article introduces an Optimal Deep Convolutional Neural Network based Face Detection and Emotion Recognition model (ODCNN-FDER) technique. The aim of the ODCNN-FDER technique is to detect faces and identify the existence of different emotions in them. To achieve this, the ODCNN-FDER technique initially employs Multi-Task Cascaded Convolutional Neural Network (MCCNN) model. Next, the fusion based feature extraction process is involved using two DL models namely EfficientNetB3 and InceptionResNetV2. For emotion recognition, Convolutional Attention Gated Recurrent Neural Network (CAGRNN) model is used. Lastly, root mean square propagation (RMSProp) optimizer was exploited for the optimal hyperparameter tuning of the CAGRNN approach. The performance validation of the ODCNN-FDER methodology was tested on the FER-2013 database. The experimental values highlighted the improved face detection and FER results of the ODCNN-FDER technique over other models.

Keywords: *Computer vision, Deep learning, Face detection, Facial emotion recognition, RMSProp optimizer*

1. Introduction

Nowadays, face recognition and detection methods including facial expressions analysis are an efficient research field in the Computer Vision (CV) community [1]. Face detection is a computer technology that can give a digital video or image, identify the facial features, and determine the sizes and positions of human faces by excluding anything else like, bodies and buildings, trees existing in the video or image. This detection and localization of human faces is a pre-requisite for recognizing faces or analysing facial expressions, which is utilized in applications like Human Computer Interface (HCI), video monitoring, and image database management [2]. A machine can recognize and detect an individual's face utilizing a normal web camera. But the factors such as a blurry or contrasting image in shadows, absence of proper brightness or lighting of an image, and viewing a person from an angle can considerably increase the complication for detecting a face [3]. In the year of 1990s, face recognition has been a prominent field of investigation but still it is less dependent than face recognition and far away from being considered an effective technique of user authentication [4].

In recent decades, Emotion recognition is a technique that has been achieving plenty of interest with the growth of Artificial Intelligence (AI) techniques. This could be obtained by investigating the facial expressions or voice tone, body postures [5]. In this research, the main focus is on recognizing emotions by applying facial expressions. Gathering the facial expression of other individuals supports human communication by understanding the purposes of others [6]. The technique of Facial Emotion Recognition (FER) is a booming area of research, which has the utilizations like Human-Machine Interaction (HMI), computer animations, and different learning processes – comprehending the internal state of attention of the learners [7]. The method of vision sensor based FER has fascinated attention in existing research and has higher capability of FER recognition in real-time. In the vision based FER, the researchers mainly concentrated on seven basic expressions like disgust, happiness, fear, anger, surprise, sad, and neutral and it categorized the FER into 2 sub-categories namely traditional and DL based approaches [8]. The requirement for intelligent

[1] *Assistant Professor, CSE Dept, BMS Institute of Technology and Management, Bangalore-560064.*
ORCID: 0000-0001-5511-6893
[2] *Associate Professor, CSE Dept, Ballari Institute of Technology & Management, Ballari-583104.*
ORCID: 0000-0002-8372-3612
* *Corresponding Author Email: gnambika@outlook.com*

technologies for deciding a potential user has needs and desires and then selecting an effective action method has rocketed with the wide acceptance of intelligent technology in modern life. Moreover, CV and DL techniques are applied in almost all engineering domains and social spheres like manufacturing, medical imaging, speech and text recognition, and emotion recognition [9]. Despite the important achievement of traditional FER algorithm based on the extraction of manual features, in previous decades, researchers have changed their attention to the DL method because of its outstanding capability of automatic recognition [10].

This article introduces an Optimal Deep Convolutional Neural Network based Face Detection and Emotion Recognition model (ODCNN-FDER) technique. The aim of the ODCNN-FDER technique is to detect faces and identify the existence of different emotions in them. To achieve this, the ODCNN-FDER technique initially employs Multi-Task Cascaded Convolutional Neural Network (MCCNN) model. Next, the fusion based feature extraction process is involved using two DL models namely EfficientNetB3 and InceptionResNetV2. For emotion recognition, Convolutional Attention Gated Recurrent Neural Network (CAGRNN) model is used. Lastly, root mean square propagation (RMSProp) optimizer was exploited for the optimum hyperparameter tuning of the CAGRNN algorithm. The performance validation of the ODCNN-FDER method is tested on the FER-2013 dataset.

2. Related Works

Jain et al. [11] present a novel Squirrel Search Optimizer with DL Enabled Facial Emotion Recognition (SSO-DLFER) approach for Autonomous Vehicle Drivers. The presented SSO-DLFER system follows 2 main procedures such as emotion recognition and face detection. The RetinaNet approach was utilized at primary step of face recognition method. For emotion detection, the SSO-DLFER approach executed the NASNet Large feature extraction with GRU method as a classification. In order to enhance the emotion detection solution, the SSO-based hyper-parameter tuning process was executed. In [12], a DL- based structure was presented for human emotion recognition. The presented structure utilizes the Gabor filtering for feature extraction and CNN for classification.

Nasri et al. [13] present a FER method from static image dependent upon Xception CNN structure and K-fold cross-validation approach. The presented method can be enhanced by utilizing the fine-tuned system. The Xception approach pre-training on ImageNet dataset for object detection can be fine-tuned for recognizing 7 emotional states. Gao et al. [14] introduce an automatic optimizer structure utilizing binary coding method and GPSO with gradient penalties for selecting the design. Such SI optimizer methods can be employed but not widely utilized, and the present work concentrates on methods with set depth of networks. Shao and Qian [15] presented 3 new CNN approaches with various structures. A primary one is a shallow network termed Light-CNN that is fully CNN comprising 6 depthwise separable residual convolutional elements for resolving the problem of difficult topology and over-fitting. The secondary one is dual-branch CNN that extract typical LBP features and DL feature in parallel. The tertiary one is a pre-training CNN that is planned by TL system for overcoming the lack of trained instances.

In [16], a DL-based method was presented for detecting the facial expression of persons. The presented system comprises 2 parts. The previous one learns local features in face images utilizing a local gravitational force descriptor, but, in the end, the descriptor was provided as new DCNN approach. The presented DCNN contains 2 branches. A primary branch searches geometric features like curves, edges, and lines; but holistic features were extracted by secondary branch. Cui et al. [17] introduce an endwise Regional-Asymmetric CNN (RACNN) for emotion detection that comprises temporal, regional, and asymmetric feature extractions. Especially, continuous 1-D convolutional layers can be employed in temporal feature extraction for learning time–frequency representations. Afterward, the regional feature extraction comprises two 2-D convolutional layers for capturing regional data amongst physically nearby channels.

3. The Proposed Model

In this study, we have concentrated on the development of the ODCNN-FDER technique. The major intention of the ODCNN-FDER technique is to detect faces and identify the existence of different emotions in them. To achieve this, the ODCNN-FDER method follows two most important processes: MCNN based face detection and emotion recognition. Fig. 1 depicts the overall flow of ODCNN-FDER algorithm.

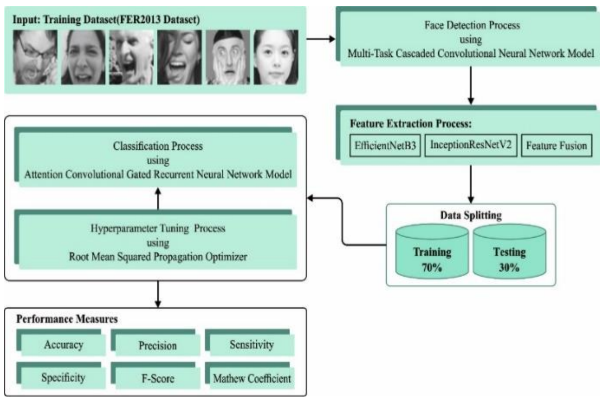

Fig. 1. Overall flow of ODCNN-FDER algorithm

3.1. Face Detection using MCCNN Model

Primarily, the MCCNN algorithm is used for automated face detection process. MCCNN is used as a solution for face alignment and detection [18]. The process comprises three phases of convolution network that are capable of recognizing faces and landmark locations like mouth, nose, and eyes. Initially, it exploits a shallow CNN to rapidly generate candidate window. Next, through more complex CNN it refines the candidate window. At last, it exploits a CNN to further refine the result and output the facial landmark position which is more complex than the others. At first, take the image and rescale it to dissimilar scales for building an image pyramid viz., input of three-stage cascaded networks.

Stage1: Proposal Network (P-Net)

A full convolution network (FCN) is the first stage. The only distinction between a FCN and a CNN is that a FCN doesn't apply dense layer as part of the architecture. This P-Net was employed for achieving candidate windows and their bounding box regression vector. Bounding box regression is a traditional way of predicting the localization of boxes once the objective is to detect an object of some predefined classes. Some refinement can be done after attaining the bounding box vector, for combining overlapping regions. After refinement, each candidate window is used to decrease candidate counts.

Stage2: Refine Network (R-Net)

At the last stage, each candidate from the P-Net was given into the R-Net. Further, the R-Net decreases the volume of candidates, performs calibration with bounding box regression, and exploits non-maximum suppression (NMS) to combine overlapping candidates. The R-Net output if the input is a face or not, is a ten-element vector for the localization of facial landmarks, and a four element vector that is bounding box for faces.

Stage3: Output Network (O-Net)

This phase could not be unlike the R-Net, however, this Output Network purposes to define the face in further detail and output the 5 facial landmarks' place for mouth, eyes, and nose.

3.2. Process involved in FER Technique

In the second phase, the FER process takes place using three subprocesses namely fusion based feature extraction, CAGRNN based classification, and RMSProp based parameter tuning.

3.2.1. Fusion based Feature Extraction

In this stage, the fusion based feature extraction procedure takes place using two DL models namely EfficientNetB3 and InceptionResNetV2. EfficientNet presents an exceptional method for scaling NN by improving precision, depth, and width [19]. The CNN method scaling model revolves around the application of compounded coefficients that uniformly scale-up the dimension, width, and depth of the network. Furthermore, EfficientNet is a kind of DL algorithm that is derived in baseline model introduction by using NN search. The basic component of EfficientNetB0 is a mobile inverted bottleneck convolution (MBConv) that is somewhat adapted due to the more of special block called a squeeze-and-excitation optimization block. Therefore, every MBConv block relies on shortcut connection and depth wise convolution layers between the blocks. EfficientNetB3 is a CNN model designed by researcher workers at Google that attained remarkable performance on the ImageNet classification tasks. The architecture belongs to the family of EfficientNet model that is developed to accomplish higher performance while being computationally effective. It exploits the synthesis of convolutional layers with diverse kernel sizes along with squeeze-and-excitation elements that selectively increase relevant features.

Inception-ResNetv2 is a deep CNN (DCNN) architecture which integrates the Inception model from the Inception network and the residual connection from the ResNet model [20]. It was developed as an improvement to the original Inception and ResNet models to enhance their accuracy and performance. The basic concept behind Inception-ResNetv2 is to resolve the problems of training deep neural networks. The deep network tends to suffer from gradient vanishing problems, where the gradient becomes very smaller during backpropagation, which makes them challenging for the network to learn efficiently. Residual connection, as presented in ResNet, lessens these problems by the network for learning residual mappings rather than direct mapping. Inception ResNetv2 integrates the Inception model, which is developed to capture multiscale features by applying dissimilar filter sizes (1x1, 3x3, 5x5) in similar branches and concatenating their output. This allows the network to capture local and global data at dissimilar scales. The residual connection skips more than one layer and directly feeds the input to the output, which makes it easier for the network to learn residual data.

3.2.2. CAGRNN based Classification

For emotion detection and classification process, the CAGRNN model is exploited. Recurrent neural network (RNN) exploits historic data instead of the present data for classification. In addition, a bi-directional RNN (BRNN) architecture was introduced, which uses past and present data [21]. Hence, two RNNs are employed to implement the forward and reverse functions. The output was connected with the same output layer for recording the feature sequence. An additional bi-directional GRU

(BiGRU) model was proposed based on the BRNN model, which replaces the hidden state of BRNN with the single GRU memory units. Now, the combination of these two BiGRU models with attention mechanisms is considered an AGRNN. Fig. 2 showcases the framework of CAGRNN method.

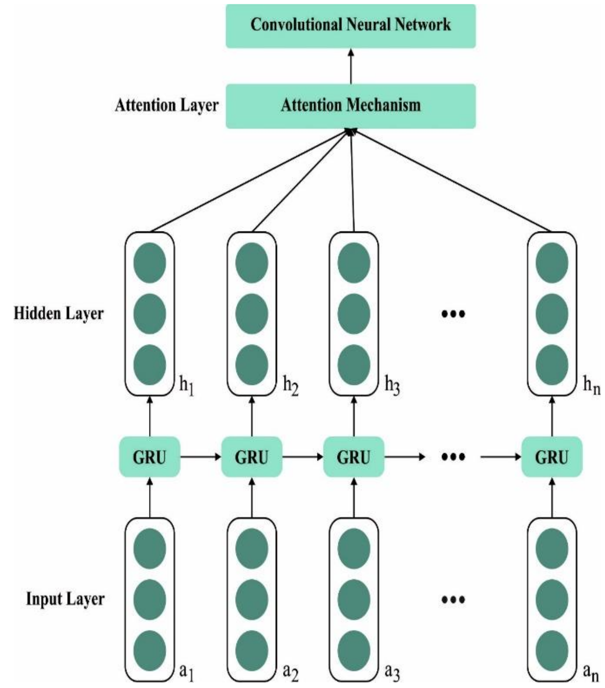

Fig. 2. Architecture of CAGRNN model

Assume an m-dimensional input dataset as $(y_1, y_2, \ldots., y_m)$. The hidden state in the BGRU generates an output H_{t_1} at t_1 time interval is given below;

$$\vec{H} = \sigma\left(w_{e_{\overrightarrow{yH}}} y_{t_1} + w_{e_{\overrightarrow{HH}}} \vec{H}_{t_1-1} + c_{\vec{H}}\right)$$
(1)

$$\overleftarrow{H}_{t_1} = \sigma\left(w_{e_{\overleftarrow{yH}}} y_{t_1} + w_{e_{\overleftarrow{HH}}} \overleftarrow{H}_{t_1-1} + c_{\overleftarrow{H}}\right)$$
(2)

$$H_{t_1} = \vec{H}_{t_1} \oplus \overleftarrow{H}_{t_1}$$
(3)

Wherew_e represents the weighted factor for two connected layers,c shows the bias vector, σ denotes the activation function,$\vec{H}_{t_1}$ and $\overleftarrow{H}_{t_1}$ represents the positive and negative outputs of GRU,$\oplus$ and denotes the bitwise operator. In the CAGRU, the convolutional layer was widely used for the feature maps or input sequences for capturing spatial data. Then, the attention module is used to calculate attention weight over the convolutional feature, which highlights the crucial region. The weighted feature was combined and fed into the GRU that sequentially process them to capture temporal dependency. The last output of the network can be utilized for dissimilar tasks namely regression, classification, or sequence generation.

3.2.3. Hyperparameter Tuning

At the final stage, the RMSProp optimizer is used. AGRNN is a hybrid network of BiGRU with the attention module. The main concept behind RMSprop is to keep the moving average of the squared gradient for all the weights [22]. In RMSProp learning rate is automatically adjusted and it selects a dissimilar learning rate for all the parameters. RMSProp splits the learning rate by the average of the exponential decay of squared gradient. The RMSprop optimizer attempts to restrict the oscillation in the vertical direction that consecutively assists in increasing the learning rate such that the model takes large step in the horizontal direction and converges faster. The calculation of RMSprop is shown in the following. The value of momentum is signified as β and is generally fixed as 0.9.

$$vdw = \beta \cdot vdw + (1-\beta) \cdot dw^2$$
(4)

$$vdb = \beta \cdot' vdw^{-\vdash(x-\beta)\cdot db^2}$$
(5)

$$W = W - \alpha \cdot \frac{dw}{\sqrt{vvdw + \epsilon}}$$
(6)

$$b = b - \alpha \cdot \frac{db}{\sqrt{vdb + \varepsilon}}$$
(7)

In backward propagation, dW and db are used for updating W and b parameters:

$$W = W - learning\ rate * dW$$
(8)

$$b = b - learning\ rate * db$$
(9)

Rather than applying dW and db autonomously for all the epochs, we take the exponentially weighted average of the square of dW and db in RMSprop.

$$S_{dw} = \beta^* S_{dw} + (1-\beta)^* dW^2$$
(10)

$$S_{db} = \beta^* S_{db} + (1-\beta)^* db^2$$
(11)

From the expression, beta β' is another hyperparameter and takes value from 0 to 1. The newly weighted average is generated by the weights, average of prior value, and present value square. The parameters are updated afterward computing exponentially weighted average.

$$W = W - learning\ rate * dW/sqrt(S_{dW}) \quad (12)$$

$$b = b - learning\ rate * db/sqrt(S_{db}) \quad (13)$$

If S_{dW} is relatively lesser, then we're dividing it by dW. If S_{db} is relatively larger then divide db with comparatively large number to delay the update on vertical dimension.

4. Results and Discussion

In this section, the results of the ODCNN-FDER technique is tested on the FER2013 Dataset [23] which holds seven classes (Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Contempt). Fig. 3 demonstrates the extraction feature maps images.

Fig. 3. Extracted Feature Maps

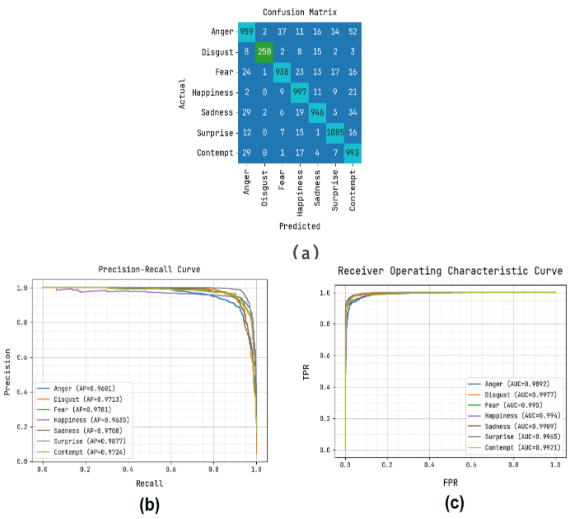

Fig. 4. Results on Training Dataset a) Confusion Matrix b) PR-Curve c) ROC

Fig. 4 reveals the classifier outcome of the ODCNN-FDER algorithm on training dataset. Fig. 4a describes the confusion matrix provided by the ODCNN-FDER approach. The result inferred that the ODCNN-FDER approach has detected and classified all 7 classes accurately.Followed by, Fig. 4b exposes the PR curve of the ODCNN-FDER method. The outcomes implied that the ODCNN-FDER algorithm has achieved higher PR outcomes on 7 classes. At last, Fig. 4c exemplifies the ROC curve of the ODCNN-FDER algorithm. The outcome outperformed that the ODCNN-FDER system has led to able outcomes with superior ROC values on 7 classes.

Fig. 5 depicts the classifier result of the ODCNN-FDER system on testing database. Fig. 5a showcases the confusion matrix offered by the ODCNN-FDER algorithm. The result stated that the ODCNN-FDER approach has recognized and classified all 7 classes accurately.Afterward, Fig. 5b reveals the PR outcome of the ODCNN-FDER system. The outcome depicts that the ODCNN-FDER algorithm has attained maximal PR results on 7 classes. Lastly, Fig. 5c demonstrates the ROC study of the ODCNN-FDER system. The outcome exhibited that the ODCNN-FDER approach has resulted in capable outcomes with higher ROC values on 7 class labels.

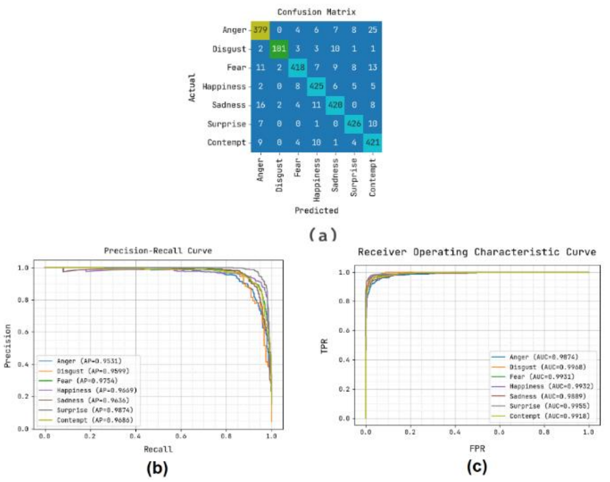

Fig. 5. Results on Testing Dataset a) Confusion Matrix b) PR-Curve c) ROC

In Table 1 and Fig. 6, the FER results of the ODCNN-FDER technique are investigated on TRS and TSS. The results indicate that the ODCNN-FDER technique reaches improved results on both TRS and TSS. On TRS, the ODCNN-FDER technique offers $accu_y$ of 95.29%, $prec_n$of 93.15%, $sens_y$ of 91.85%, $spec_y$ of 98.72%, F_{score} of 92.42%, and MCC of 91.19%.

Table 1. FER outcome of ODCNN-FDER approach on TRS and TSS

Metrics	***Training Set***	***Testing Set***
Accuracy	95.29	94.74
Precision	93.15	92.27
Sensitivity	91.85	90.88
Specificity	98.72	98.60
F-Score	92.42	91.48
Mathew Coefficient	91.19	90.14

Also, on TSS, the ODCNN-FDER approach offers $accu_y$ of 94.74%, $prec_n$ of 92.27%, $sens_y$ of 90.88%, $spec_y$ of 98.60%, F_{score} of 91.48%, and MCC of 90.14%.

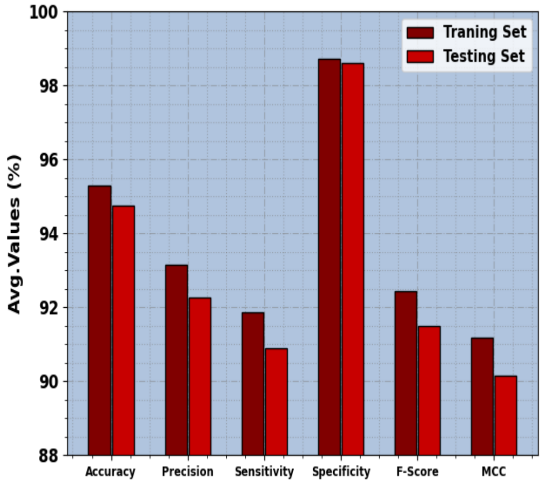

Fig. 6. Average outcome of ODCNN-FDER approach on TRS and TSS

Fig. 7 displaying the training accuracy TR_accu_y and VL_accu_y of the ODCNN-FDER method. The TL_accu_y is defined by the evaluation of the ODCNN-FDER technique on TR dataset whereas the VL_accu_y is computed by assessing the performance on a separate testing database. The results display that TR_accu_y and VL_accu_y maximum with an increase in epochs. Accordingly, the outcome of the ODCNN-FDER method gets improved on the TR and TS dataset with a rise in count of epochs.

In Fig. 8, the TR_loss and VR_loss curve of the ODCNN-FDER algorithm is exposed. The TR_loss demonstrates the error among the predictive solution and original values on the TR data. The VR_loss represents the measure of the efficiency of the ODCNN-FDER technique on individual validation data. The outcomes implied that the TR_loss and VR_loss tend to reduce with increasing epochs. It portrayed the enhanced performance of the ODCNN-FDER technique and its capability to generate accurate classification. The reduced value of TR_loss and VR_loss demonstrates the greater performance of the ODCNN-FDER system on capturing patterns and relationships.

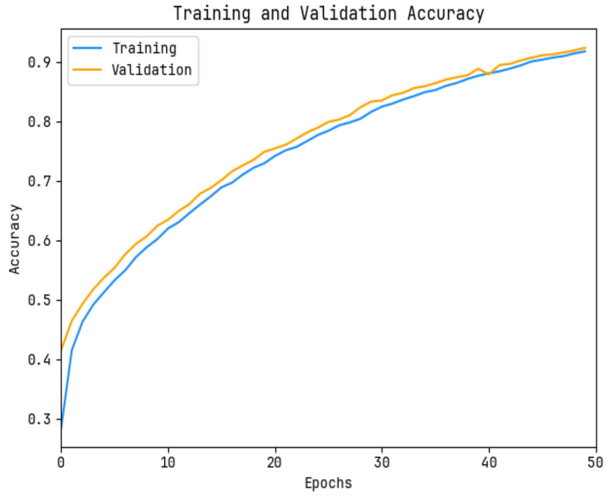

Fig. 7. $Accu_y$ curve of the ODCNN-FDER approach

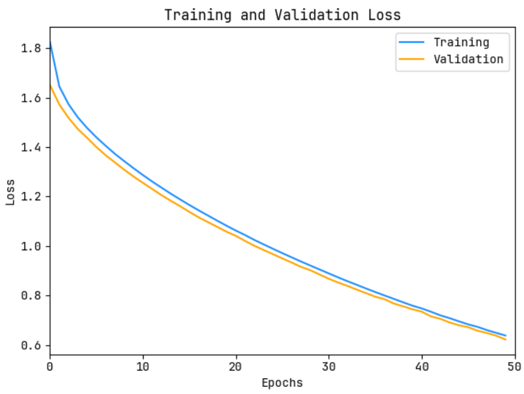

Fig. 8. Loss curve of the ODCNN-FDER approach

In Table 2 and Fig. 9, the FER results of the ODCNN-FDER technique are compared with recent DL models [24]. The experimental results imply that the Mini-Xception and CNN-transfer learning models accomplish worse performance. At the same time, three layer-CNN, six-layer-CNN, and InceptionResNetv2 models exhibit slightly enhanced results.

Table 2. Comparative outcome of ODCNN-FDER algorithm with recent DL systems

Methods	Accuracy	Precision	Sensitivity	Specificity	F-Score
Mini-Xception Model	66.00	78.77	75.19	72.44	80.84
CNN-Transfer Learning	72.00	69.78	74.26	74.37	80.27
Three Layer-CNN Model	88.60	87.61	90.08	90.36	90.51
Six Layer-CNN Model	86.78	89.75	88.87	88.37	89.81
InceptionResNetV2	87.69	88.41	89.76	90.49	89.73
FERVCB-DL	89.20	88.22	88.77	88.13	90.23

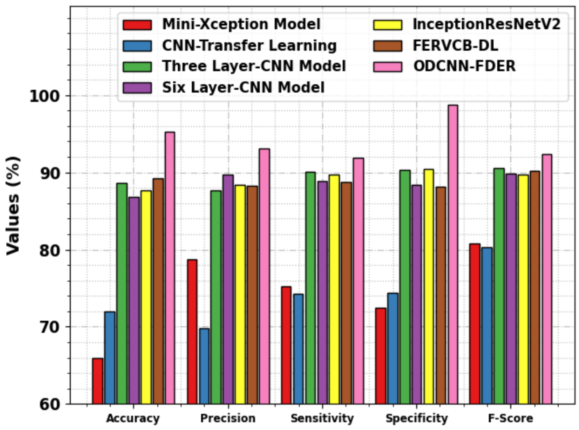

Fig. 9. Comparative outcome of ODCNN-FDER algorithm with recent DL systems

Although the FERVCB-DL model reaches considerable performance with $accu_y$ of 89.20%, $prec_n$ of 88.22%, $sens_y$ of 88.77%, $spec_y$ of 88.13%, and F_{score} of 90.23%, the ODCNN-FDER technique gains maximum performance with $accu_y$ of 95.29%, $prec_n$ of 93.15%, $sens_y$ of 91.85%, $spec_y$ of 98.72%, and F_{score} of 92.42%. These results confirmed the enhanced performance of the ODCNN-FDER technique on face detection and FER processes.

5. Conclusion

In this study, we have concentrated on the development of the ODCNN-FDER technique. The major intention of the ODCNN-FDER technique is to detect faces and identify the existence of different emotions in them. To achieve this, the ODCNN-FDER method follows two most important processes: face detection and emotion recognition. In the primary phase, the MCCNN approach was applied for the face detection method. Next, in the second phase, the FER process takes place using three subprocesses namely fusion based feature extraction, CAGRNN based classification, and RMSProp based parameter tuning. In this work, the RMSProp optimizer is exploited for the optimum hyperparameter tuning of the CAGRNN algorithm. The performance validation of the ODCNN-FDER technique was tested on the FER-2013 database. The experimental values highlighted the improved face detection and FER results of the ODCNN-FDER technique over other models. In future, the ODCNN-FDER technique can be extended to the design of metaheuristic optimizers.

Author contributions

Ambika G.N: Conceptualization, Methodology, Software, Field study, Visualization, Investigation, Writing-Reviewing and Editing

Yeresime Suresh: Conceptualization, Methodology, Data curation, Writing-Original draft preparation, Software, Validation., Field study.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models,". *Multimedia Tools and Applications*, *82*(8), pp.11365-11394, 2023.

[2] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *Electronics*, *10*(9), p.1036, 2021.

[3] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, L. "Learning affective video features for facial expression recognition via hybrid deep learning," *IEEE Access*, *7*, pp.32297-32304, 2019.

[4] D. Wu, X. Han, Z. Yang, and R. Wang, "Exploiting transfer learning for emotion recognition under cloud-edge-client collaborations," *IEEE Journal on Selected Areas in Communications*, *39*(2), pp.479-

490, 2020.

[5] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," In *2020 international conference for emerging technology (INCET)* (pp. 1-5). IEEE, 2020.

[6] L. Schoneveld, A. Othmani, and H. Abdelkawy, "Leveraging recent advances in deep learning for audio-visual emotion recognition," *Pattern Recognition Letters*, *146*, pp.1-7, 2021.

[7] D. Canedo, and A. J. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences*, *9*(21), p.4678, 2019.

[8] D. Ruan, Y. Yan, S. Lai, Z. Chai, C. Shen, and H. Wang, "Feature decomposition and reconstruction learning for effective facial expression recognition," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 7660-7669, 2021.

[9] S. Minaee, M. Minaei, and A. Abdolrashidi, A, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, *21*(9), p.3046, 2021.

[10] T. S. Gunawan, A. Ashraf, B. S. Riza, E. V. Haryanto, R. Rosnelly, M. Kartiwi, and Z. Janin, "Development of video-based emotion recognition using deep learning with Google Colab," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *18*(5), pp.2463-2471, 2020.

[11] D. K. Jain, A. K. Dutta, E. Verdú, S. Alsubai, and A. R. W. Sait, "An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers," *Image and Vision Computing*, *133*, p.104659, 2023.

[12] M. M. T. Zadeh, M. Imani, and B. Majidi, "Fast facial emotion recognition using convolutional neural networks and Gabor filters," In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)* (pp. 577-581). IEEE, 2019.

[13] M. A. Nasri, M. A. Hmani, A. Mtibaa, D. Petrovska-Delacretaz, M. B. Slima, and A. B. Hamida, "Face emotion recognition from static image based on convolution neural networks," In *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE, 2020.

[14] Z. Gao, Y. Li, Y. Yang, X. Wang, N. Dong, and H. D. Chiang, "A GPSO-optimized convolutional neural networks for EEG-based emotion recognition," *Neurocomputing*, *380*, pp.225-235, 2020.

[15] J. Shao, and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, *355*, pp.82-92, 2019.

[16] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks," *IEEE Transactions on Instrumentation and Measurement*, *70*, pp.1-12, 2020.

[17] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowledge-Based Systems*, *205*, p.106243, 2020.

[18] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE signal processing letters*, 23(10), pp.1499-1503, 2016.

[19] S. A. Shah, G. M. Lakho, H. A. Keerio, M. N. Sattar, G. Hussain, M. Mehdi, R. B. Vistro, E. A. Mahmoud, and H. O. Elansary, "Application of Drone Surveillance for Advance Agriculture Monitoring by Android Application Using Convolution Neural Network," *Agronomy*, *13*(7), p.1764, 2023.

[20] R. Florez, F. Palomino-Quispe, R. J. Coaquira-Castillo, J. C. Herrera-Levano, T. Paixão, and A. B. Alvarez, "A CNN-Based Approach for Driver Drowsiness Detection by Real-Time Eye State Identification," *Applied Sciences*, *13*(13), p.7849, 2023.

[21] N. Parveen, P. Chakrabarti, B. T. Hung, and A. Shaik, "Twitter sentiment analysis using hybrid gated attention recurrent network," *Journal of Big Data*, *10*(1), pp.1-29, 2023.

[22] D. V. Babu, C. Karthikeyan, and A. Kumar, "Performance analysis of cost and accuracy for whale swarm and RMSprop optimizer," *In IOP Conference Series: Materials Science and Engineering*, Vol. 993, No. 1, p. 012080, 2020. IOP Publishing.

[23] https://www.kaggle.com/datasets/deadskull7/fer2013

[24] M. F. Alsharekh, "Facial Emotion Recognition in Verbal Communication Based on Deep Learning," *Sensors*, *22*(16), p.6105, 2022.

[25] Ms. Mohini Dadhe, Ms. Sneha Miskin. (2015). Optimized Wireless Stethoscope Using Butterworth Filter. International Journal of New Practices in Management and Engineering, 4(03), 01 - 05.

Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/37

[26] Banerjee, S. ., Chakraborty, S. ., & Mondal, A. C. . (2023). Machine Learning Based Crop Prediction on Region Wise Weather Data. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 145–153. https://doi.org/10.17762/ijritcc.v11i1.6084

[27] Joshi, K., Kumar, V., Sundaresan, V., Ashish Kumar Karanam, S., Dhabliya, D., Daniel Shadrach, F., & Ramachandra, A. C. (2022). Intelligent fusion approach for MRI and CT imaging using CNN with wavelet transform approach. Paper presented at the IEEE International Conference on Knowledge Engineering and Communication Systems, ICKES 2022, doi:10.1109/ICKECS56523.2022.10060322 Retrieved from www.scopus.com