Journal of Current Science and Technology, May-August 2023 Vol. 13 No. 3, pp. 517- 532
Copyright ©2018-2023, Rangsit University ISSN 2630-0656 (Online)

Cite this article: Kumari, R., & Wasim, J. (2023, May). A deep learning approach for human facial expression recognition using Residual Network – 101. *Journal of Current Science and Technology, 13*(3), 517-532. https://doi.org/10.59796/jcst.V13N3.2023.2152

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

A Deep Learning Approach for Human Facial Expression Recognition using Residual Network – 101

Ranjana Kumari[1*] and Javed Wasim[2]

[1]Department ECE, Institute of technology,
Mangalayatan University Aligarh, Beswan, Uttar Pradesh 202145, India
[2]Department of Computer Engineering & Applications,
Institute of Engineering & Technology, Mangalayatan University, Beswan, Uttar Pradesh 202145, India

[*]Corresponding Author; E-mail: ranjnaguddi@gmail.com

Received 20 June, 2022; Revised 3 January, 2023; Accepted: 24 January, 2023;
Published online 30 August, 2023

Abstract

Emotion recognition is a dynamic process that focuses on a person's emotional state, which implies that the emotions associated with each individual's activities are unique. Human emotion analysis and recognition have been popular study areas among computer vision researchers. High dimensionality, execution time, and cost are the main difficulties in human emotion detection. To deal with these issues, the proposed model aims to design a human emotion recognition model using Residual Networks-101 (ResNet-101). A Convolutional Neural Network (CNN) design called ResNet-101 solves the vanishing gradient issue and makes it possible to build networks with thousands of convolutional layers that outperform networks with fewer layers. An image dataset was used for this emotion recognition. Then, this image dataset was subjected to preprocessing to resize the image and eliminate the noise contents present in the images. After preprocessing, the image was given to the classifier to recognize the emotions effectively. Here, ResNet-101 was used for the classification of six classes. The experimental results demonstrate that ResNet-101 models outperform the most recent techniques for emotion recognition. The proposed model was executed in MATLAB software and carried out several performance metrics. The proposed architecture attained better performance in terms of accuracy 92% and error with 0.08 and other performances like 92% of precision, 85% of specificity and 98% of sensitivity so on, and this shows the effectiveness of the proposed model to existing approaches such as LeNet, AlexNet and VGG. In comparison to current techniques, the suggested model provides improved recognition accuracy for low intensity or mild emotional expressions.

Keywords: *convolutional neural network; human emotion recognition; image resizing; noise removal; residual networks-101*

1. Introduction

Human facial expression identification is a process of classifying the reactions via body movements, verbal expression, facial expression and multiple physical signals measurements. Human emotions are undeniably important in the development of modern technology (Hassan et al., 2019). Automated tutoring systems, driver warning systems, mental health monitoring, smart environments, human-computer interaction, and image and video retrieval are the emotion analysis and detection systems recently (Pal, Mukhopadhyay, & Suryadevara, 2021). Moreover, emotion recognition is used by psychiatrists and psychologists to identify different mental health disorders (Al Machot et al., 2019). Many scholars and researchers developed various strategies and algorithms for recognizing emotions from facial features and speech signals during the last few decades. Because of its complexity, it remains a difficult subject in the area of physiology, computer vision, psychology and artificial intelligence (Gupta, Chopda, & Pachori, 2018; Zhang, Zheng, Cui, Zong, & Li, 2018). Facial expressions are the most important factor in recognizing human emotion. Due to the sensitivity of external factors such as dynamic head movements and lighting conditions, it is challenging to identify human mood using facial expression features.

For instance, Patients' emotional and physical conditions can be monitored in real-time in a healthcare system with an emotion recognition module, and appropriate therapy can be administered appropriately (Hossain, & Muhammad, 2019). Several researchers have shown an interest in recognizing human facial features such as age, gender, and emotions over the last ten years (Amrani & Jiang, 2017). This is because several applications of this difficult topic are offered, including security, human-computer interaction, and the recognition of human emotions, among others (Tzirakis, Trigeorgis, Nicolaou, Schuller, & Zafeiriou, 2017). Humans also communicate their emotions through facial expressions as part of communication (Amrani et al., 2017). Humans can easily understand emotions and facial expressions, but emotion recognition is a difficult challenge for machines (Chen et al., 2018). A concept for improving the subject-dependent and subject-independent human emotion recognition systems is required based on the prior information (Abdullah, 2019). Most approaches to analyzing facial expressions and recognizing emotions use deep learning, a type of machine learning. However, its performance is dependent on the amount of the data (Wang, Phillips, Dong, & Zhang, 2018). Many strategies and procedures have been used to classify emotions in recent years, but developing an automated system is still difficult (Wattana, Janpong, & Supichayanggoon, 2018).

Deep learning cannot be implemented since the size of facial expression datasets is still insufficient (Chen et al., 2019). Main intention of this research used deep learning approach to recognise emotions and demonstrate how data preprocessing affects deep learning performance. The preprocessing stage involves image resizing and noise removal. This preprocessing strategy was quite helpful in improving deep learning performance. In this research, the human emotion recognition model was attained using ResNet 101. The image data was fed into the input of the classifier for human emotion prediction. The emotions were classified into seven classes such as neutral, disgust, sadness, happiness, anger, fear and surprise. Finally, the accuracy of the proposed model and the existing models were compared with and without features. The major contribution of this paper is given below.

- Deep learning based Residual network 101 presents a powerful model for recognizing human emotions.
- Image preprocessing techniques were used to enhance accuracy, reduce process complexity, and improve deep learning performance.
- A ResNet 101 classifier that was tested and trained using the supplied emotion dataset carries out an efficient recognition process.
- The experimental analysis was performed by comparing with the lately suggested detection methods.

The upcoming portion of the research paper is structured as follows, and this paper includes a description of related works in section 2 and section 3 briefly explains the proposed methodology with a network diagram. Section 4 presents experimental research, and section 5 concludes the entire paper.

1.1 Literature Review

In this section, several research related to human emotion recognition utilizing various techniques are reviewed as follows.

Salama, El-Khoribi, Shoman and Shalaby (2021) developed a new multimodal framework for human emotion detection. This designed model was based on deep learning architecture of 3D Convolutional Neural Network (3D-CNN) for retrieving spatiotemporal attributes from video data and electroencephalogram (EEG) signals of human faces. Thus, fusion prediction was achieved by combining data augmentation and learning techniques. Fusion of multi-modalities using score fusion methods. Convolutional neural networks were used to design a facial emotion identification system model presented by Mehendale (2020). The facial emotion recognition utilizing the convolutional neural networks (FERC) model consists of two phases of convolutional neural network (CNN). The initial stage was responsible for eliminating the background noise from the image (Mohammed & Abdulazeez, 2021). Then, the facial attribute vector extraction was performed in the second stage. In this way, the FERC model predicts the five dissimilar kinds of facial emotions utilising an expressional vector (EV). Jiang et al. (2020) designed a probabilistic and integrated learning (PIL) classification model for handling high-level human face expression identification challenges. Initially, a unique integrated learning topology was designed to mimic the human thought process. This was built to provide the necessary foundation for understanding complex human emotions. Furthermore, by computing the confidence interval of the classification probability, a PIL-based classification model was developed to adapt to the fuzziness in facial expression categorization caused by emotional uncertainty. Additionally, the classification probability introduces three new analysis methods: the emotional tube, emotional sensitivity, and emotional decision preference. Nannapaneni and Chatterjee (2021) presented human facial expressions and emotion recognition utilizing a new machine learning algorithm. The designed model comprises three phases: preprocessing, feature extraction using the histogram of gradients, and emotion classification with the K-nearest neighbor classifier. Said and Barr (2021) developed a Face-Sensitive Convolutional Neural Network approach for detecting human emotional expressions. This approach was designed to predict expressions on grand scale pictures, which were subsequently evaluated for face landmarks to predict facial expression detection. There were two phases included in the FS-CNN model, the first stage was patch cropping, and another phase was CNN. The first stage was responsible for discovering high dimensional face images and image cropping was also performed. Then, face expression detection was done in the second stage, utilizing the convolutional neural network. For improving identification performance, two channel EEG signals and ocular modality of multimodal approach had presented by Ngai, Xie, Zou and Chou (2022). In addition, the face modality utilizing facial depths as well as facial pictures also adapts a convolutional neural network.

In addition, face modality should be facial photos or facial depth and convolutional neural network which can represent the spatiotemporal information from modality information for emotion recognition. The common arousal valence model was used for predicting emotions. Furthermore, using facial depth outperformed using facial photographs. The designed approach of emotion recognition has a lot of promise for use in a variety of educational settings. Furthermore, using facial depth outperformed using facial photographs. The designed approach of emotion recognition has a lot of promise for use in a variety of educational settings. Kittipongdaja and Siriborvornratanaku (2022) created an autonomous kidney segmentation using 2.5D ResUNet and 2.5D DenseUNet, in order to effectively extract intra-slice and inter-slice characteristics. In two separate training contexts, this method was trained and verified using the public data set from the Kidney Tumor Segmentation (KiTS19) challenge. Liu et al. (2022) developed a model for lung parenchymal segmentation in chest CT which was based on ResUnet. The residual learning unit was introduced in order to communicate low-level information and skip connections built on the U-Net architecture were used to improve the connection across layers.

Arabian et al. (2021) presented an effective model for image preprocessing through neural networks for designing effective classification models. This model was developed utilizing the standard "AlexNet" network structure and transfer learning, using three distinct ways for picture inputs. After training on a new randomly selected training set from the Oulu-CASIA database, the accuracy of a constructed model was tested with a randomly selected validation set, and visualizations of anticipated areas of significance were assessed.

Do et al. (2021) developed a deep learning method for distinguishing between good and negative emotional states. The model consists of three phases such as the gathering of the EEG dataset, de-noise the signal of EEG data in the preprocessing stage, and finally, the classification done using AlexNet with VGG-16 for deep learning. The Emotive Epoc+ 14 channel portable and wearable EEG equipment was then used to collect EEG data from 28 individual volunteers ranging in age from 21 to 28. This model recorded signals for each individual for a total of 20 minutes using four different video games as stimuli (2 positive and 2 negative labelled games). Amrani et al. (2022) have described a reliable feature extraction technique for classifying targets in SAR images by adaptively merging useful features from several CNN layers. The targets were first finetuned to be detected by the YOLOv4 network from appropriate MF SAR target images. Then, to decrease speckle noise an extremely deep CNN was trained entirely from scratch on MSTAR for recognizing and acquiring moving and stationary targets. Additionally, CNN gets deeper by adopting small-size convolution filters to minimize the number of parameters in each layer to reduce the computing cost. Based on a visual saliency model, an efficient feature extraction and classification technique was provided by Amrani et al. (2018). First, a graph based visual saliency model that was SAR-oriented was presented. Second, features like gabor and histogram of directed gradients are retrieved from processed SAR pictures using our saliency model's capacity to highlight the most important locations. Third, discrimination correlation analysis algorithm was employed for characteristic fusion in order to have more discriminative characteristics. The Mahalanobis distance based radial basis function kernel was utilized to develop a two level directed acyclic graph support vector metric learning approach that effortlessly exploits a two level DAG. This method accentuates significant qualities while reducing the influence of unnecessary information. Mehendale (2020) created a cutting-edge method for recognizing facial emotions using Convolutional Neural Networks. The FERC's two part CNN focuses on retrieving face feature vectors while the first portion of CNN eliminates the backdrop from the image. Expressional vector in the FERC model was used to recognize the five various varieties of usual facial expressions. Supervisory information was gathered from the 10,000 image database that was saved (154 persons). A technique for extracting features by employing deep residual network ResNet-50 which incorporates convolutional neural network for facial expression detection, was presented by Li and Lima (2021). By experimentally simulating the provided dataset, it can be shown this model outperforms the most widely used facial emotion recognition algorithms in terms of face emotion identification. A pre-trained DCNN model was adopted by substituting its dense upper layer(s) compatible with FER in Akhand et al. (2021), very Deep CNN modelling through transfer learning approach. The model was then tweaked with facial emotion information. Training of dense layer(s) was followed by tuning every pre-trained DCNN blocks in turn. The accuracy of FER has been improved over time because of this innovative pipeline technique.

According to the above-mentioned literature, many recognition systems' major flaw is a lack of data, which might make it difficult to generalize to new samples. The main problem with the upgraded model is the enormous number of internal network parameters and the high demands on the computation speed of hardware devices. These difficulties can be addressed by supplementing the data (increasing the number of samples) or reusing a model that has already been trained on a single task (transfer learning), as demonstrated in this study with the ResNet 101 network. Because, when analyzing among the other image preprocessing models, ResNet with many layers can be trained quickly, without raising the training error %. By applying identity mapping, ResNet assist in solving the vanishing gradient problem.

2. Objectives

The major objective of the proposed work is to develop a deep learning architecture for the human facial expression detection. Thus, a Convolutional Neural Network design known as ResNet-101 solves the vanishing gradient issue, enabling the construction of networks with thousands of convolutional layers that outperform shallower networks.

3. Methodology

3.1 Proposed methodology for emotion recognition

ResNet 101 for effective recognition of human facial expression is presented. Human expressions are one of the most important aspects of human contact. Emotion recognition is important for human health and the medical method that detects, analyses, and determines a person's medical status. Human physical parameters are translated (signals) into computing signals capable of displaying a variety of human emotions in real-time. For example, lights in a room can alter according to a person's mood, or teaching methods can be improved by recognizing human facial expressions. The proposed model addressed a number of issues in the development of scalable, reliable, efficient and human emotion recognition system with a variety of applications and services. Due to its superior performance, CNN has been extensively applied in the field of image recognition. Facial expression identification technique based on CNN model is suggested in this research. The activation function is at the center of CNN model's complicated hierarchical structure since it has the nonlinear properties that give the deep neural network its true artificial intelligence.

Figure 1 illustrates the proposed architecture of the human emotion recognition model using CNN. The model consists of three phases such as data gathering, preprocessing and classification for emotion recognition. The initial phase is responsible for gathering the image dataset. The next phase is preprocessing. This stage contains image resizing, noise removal using median filter and contrast enhancement. The final stage is image classification using CNN. The emotions were classified into six classes such as neutral, sadness, happiness, anger, fear and surprise. The step-by-step process involved in the proposed model is elaborately discussed as follows.

3.1.1 Source of images

The images were taken from a dataset published in Kaggle (ARES, 2013). Declaring of no copyright shows at https://creativecommons.org/publicdomain/zero/1.0/

3.1.2 Pre-processing

After data gathering, the image dataset was subjected to preprocessing. In image processing, this phase was considered significant for minimizing the complexity during classification. Because the images contain varying dimensions and a lot of noise before preprocessing, some preprocessing techniques were used to improve the raw image quality. In this research, two preprocessing strategies were used: image resizing and noise removal. A brief explanation of these strategies is provided in this section.

3.1.2.1 Image Resizing

Image resizing is the process of resizing the source of an input image. The original input image size was scaled to 256x256 pixels. This was because the 256x256 picture size makes deep learning algorithms like CNN easier to employ (Perumal, & Velmurugan, 2018). Bi-cubic interpolation offers a sharper result at the edge than previous approaches like bilinear interpolation. Equation (1) represents the mathematical term used to represent image resizing.

$$p(x,y)=\sum_{i=0}^{3}\sum_{j=0}^{3} b_{ij}x^{i}y^{j} \qquad (1)$$

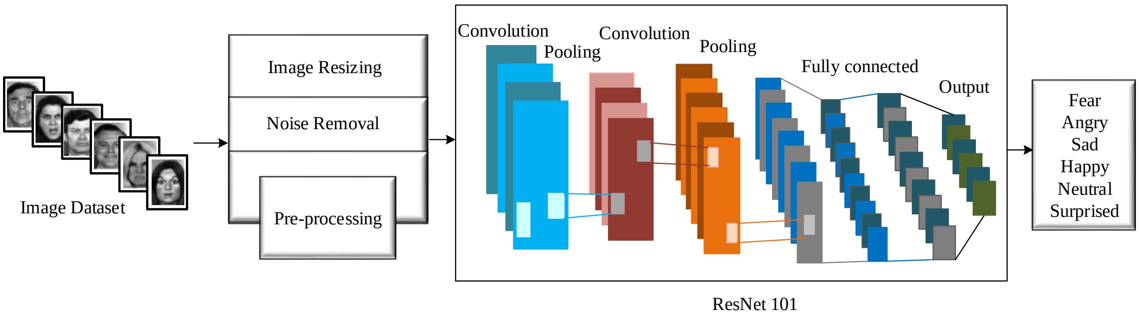


Figure 1 Architecture of human emotion recognition model using ResNet 101

From equation (1), x denotes the width in pixels, y signifies the height in pixels and b_{ij} denotes the squared area in the image. The noise is removed using a filtering algorithm using the scaled image as an input.

3.1.2.2 Noise Removal

The image was scaled to a 256x256 matrix in this research, and then a median filter is used to remove the image's unwanted noisy contents (Maheswari, & Radha, 2010). The median filter was applied to the input picture N_i which effectively removes background noises.

$$F_i(a,b)=\text{median}\{N_i(a,b)\} \qquad (2)$$

Here, (a, b) indicates the current position of an image N_i and $F_i(a, b)$ represents the filtered image. The preprocessed image was then given into ResNet 101 for classification.

3.1.3 Classification of human emotion recognition using ResNet-101

The classification problems were resolved using Residual Networks (ResNet) structure which has a big impact on computer vision problems. To prevent gradients from becoming zero following chain rule application, the ResNet network employs residual connections in this way the gradients can flow directly. Normally, the ResNet-101 entirely comprises 104 convolutional layers. It also completely consists of 33 layers, with 29 of these blocks directly using the output of the previous block. At the conclusion of each block, the summing operator was applied to get the input for subsequent blocks which were delivered as residual connections. These residuals served as the initial operand of summing operator by (Bhatti et al., 2021). The outcome of the previous block was fed into convolution layer with filter size of 1x1 and stride of 1, followed by a batch normalization layer that normalizes the output, and the result was fed into the summing operator at that block's output. ResNet 101 is one of ImageNet's most complex network structures for image classification and object detection. Inside CNN, many layers were typically interconnected and trained to do various tasks in this manner. The network learns multiple tiers of characteristics at the end of each layer.

The layers in ResNet had the same number of filters for the same output feature map size, and the number of filters was doubled if the characteristic map size was cut in half to keep the temporal complexity of every layer constant. It does direct down sampling by convolutional layers with a two-step stride. This ResNet was completed when a SoftMax and a global average pooling layer were triggered in a fully connected layer. Residual learning was removal of learned input characteristics from that layer. ResNet does this by utilizing shortcut links for individual pairs of 33 filters, instantly interconnecting k^{th} layer's input to the $(k+x)^{th}$ layer's output (Bharati, Podder, Mondal, & Prasath, 2021). By repeating activations from the prior layer until the layer below it had learnt its weights, the purpose of layer skipping was to avoid the issue of fading gradients. Weights adjust to amplify the layer adjacent to the current one when training the network, as well as to mute the layer before it. Simple deep convolutional neural networks have been proven to be more difficult to train than this network. It also deals with the issue of declining precision. ResNet-101 is a modified version of the 50-layer ResNet with a 101-layer Residual Network. ResNet-101 is a 101-layer Residual Network that is a modified version of the 50-layer ResNet. Here, 80% of preprocessed data was initially given into the ResNet-101 to train the model. After training the model, it was tested with the remaining 20% of the data. During testing, the model produces better outcomes. The classification model produces six classes such as neutral, sadness, happiness, anger, fear and surprise.

4. Results and discussion

The proposed deep learning approach of the human emotion recognition model was tested on Mat Lab software. Data collection, preprocessing and classification for emotion recognition were the three stages of the model. The image dataset was gathered during the first stage. After data gathering, the raw data were subjected to preprocessing which performs image scaling, noise removal using a median filter and contrast enhancement were all included in this stage. The preprocessed images were fed into the input of the classification process in the last step. Here, the ResNet101 model has applied this process and the convolutional layers were responsible for extracting the effective image features. The fully connected layer finally classified the image features into six different human emotions including neutral, sadness, happiness, anger, fear, and surprise. The experimental analysis of this proposed framework's deep learning approach considered several parameters. The considered parameter was MCC, Precision, Specificity, Kappa, F1_Score, Error, FPR, Sensitivity and accuracy. These parameters were evaluated for the proposed work to prove effective and accurate emotion prediction. The existing machine learning techniques considered for this comparison investigation, such as LeNet, AlexNet and VGG. These conventional approaches were compared with ResNet-101 and the proposed framework using the ResNet-101 technique for accurate emotion detection.

4.1 Dataset description

The proposed model was evaluated and tested on human emotion image dataset, and it was gathered from (ARES, 2013). The data consists of 48x48 pixel portraits of faces in grayscale. Each face is roughly in the same place and occupies a similar amount of area due to the automatic registration of the faces. The dataset entirely contains 958 images divided into train and test datasets. Images are categorized based on the emotion shown in facial expressions like surprise, neutral, sadness, happiness, fear, anger, and disgust. The dataset has 200 images as surprise, 254 images as neutral, 152 images as sadness, 102 images related to happiness, 100 images related to anger and 150 as disgust images. The proposed ResNet 101 model was trained with 80% of images and once trained the model that was tested with the rest of the 20% of images. After that, the tested model detected six different classes as, surprise, neutral, sadness, happiness, anger and disgust. The following Figure 3 shows the sample facial emotion dataset.

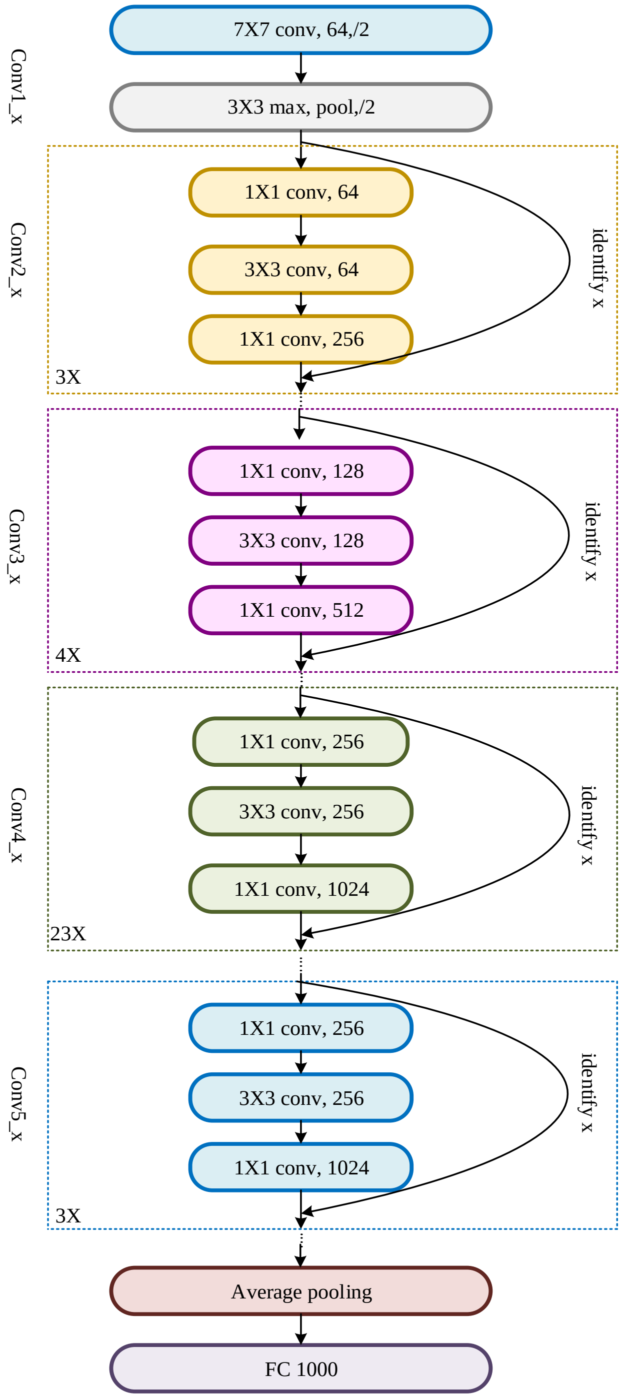


Figure 2 ResNet 101 Layer Diagram

Table 1 shows the confusion matrix attained for the developed approach. Here, there are six different classes were detected through the proposed ResNet model. Class 0 was specified as surprise that correctly predicted 200 data. Class 1 which was specified as neutral that correctly predicted 200 data and 54 data were wrongly predicted as surprise and 10 data were wrongly predicted as sadness. Class 2 was correctly predicted as 100 data and 2 data was wrongly predicted as happiness. Class 3 was correctly predicted as 100 and 2 was wrongly predicted as neutral. Class 4 was correctly predicted as 76 and wrongly predicted as 40. Class 5 was correctly predicted as 100 data.

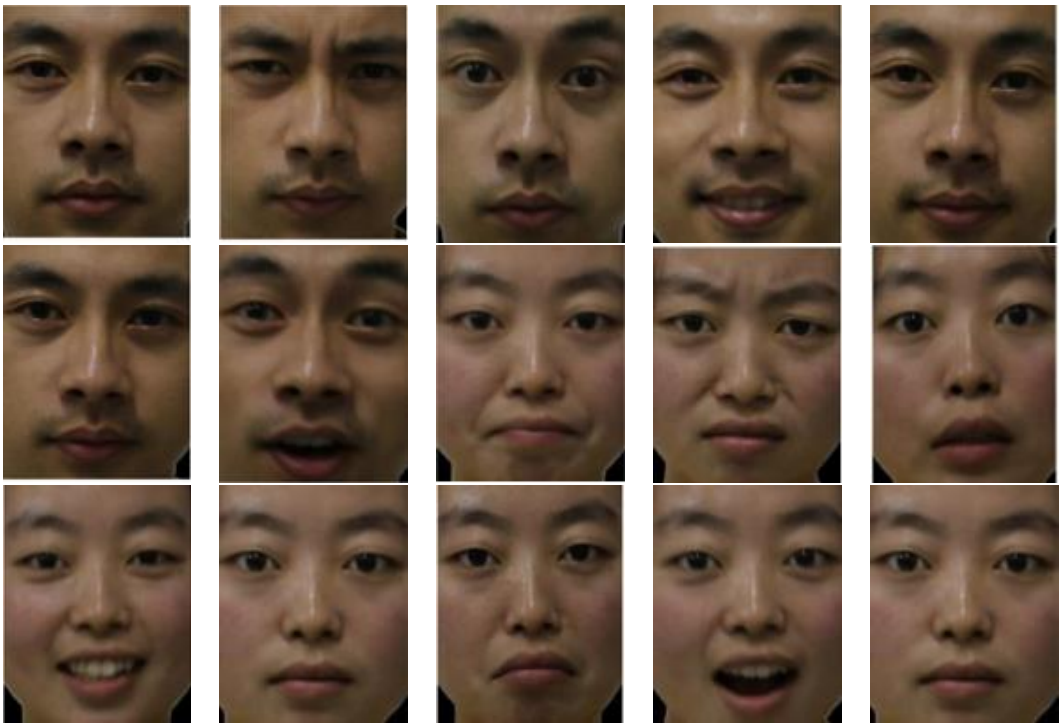
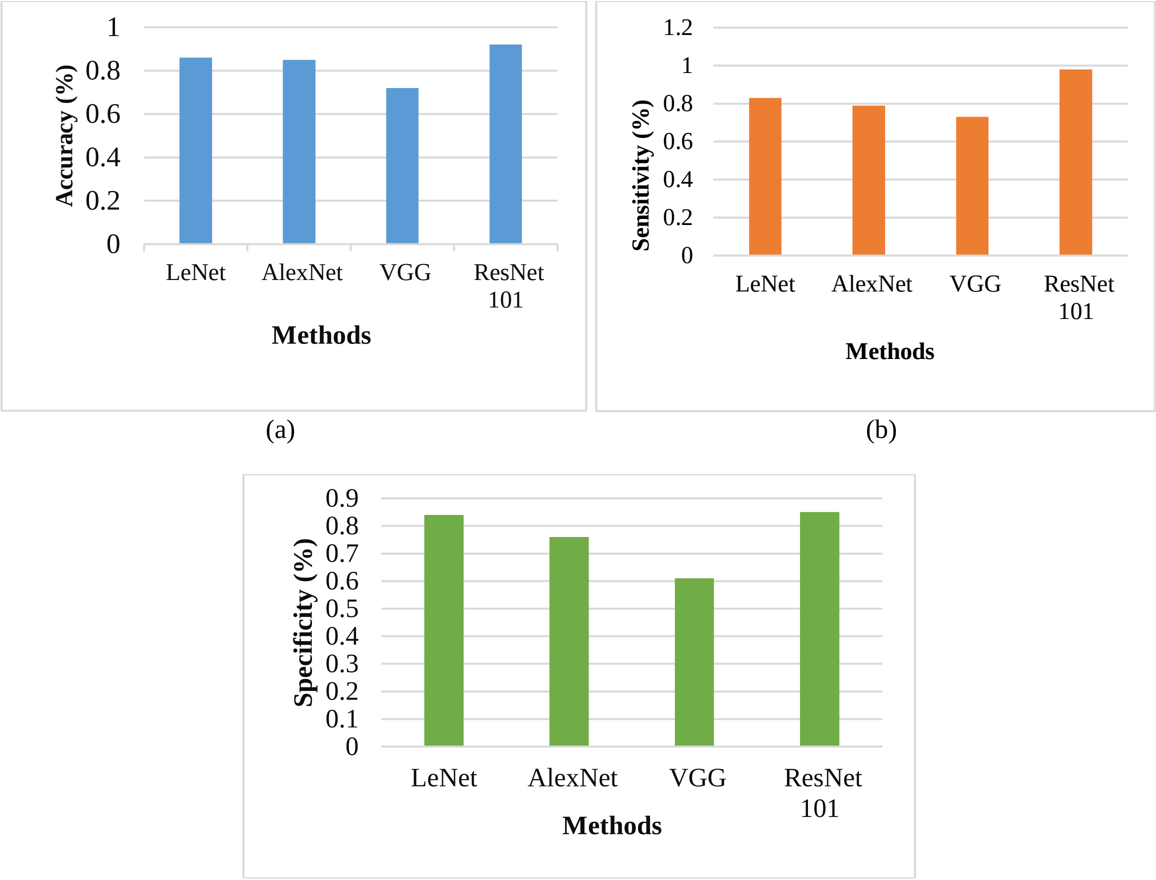


Figure 3 Sample facial emotion dataset

Source: Dataset published in Kaggle (ARES., 2013)

Table 1 Confusion matrix attained for the proposed model

Total tested data =958		**Predicted data**					
		Surprise	**Neutral**	**Sadness**	**happiness**	**Disgust**	**Anger**
Actual data	Surprise	200	0	0	0	0	0
	Neutral	54	200	10	0	0	0
	Sadness	0	0	110	40	0	0
	happiness	2	0	0	100	0	0
	Disgust	0	4	0	0	76	40
	Anger	0	0	0	0	0	100



(a)

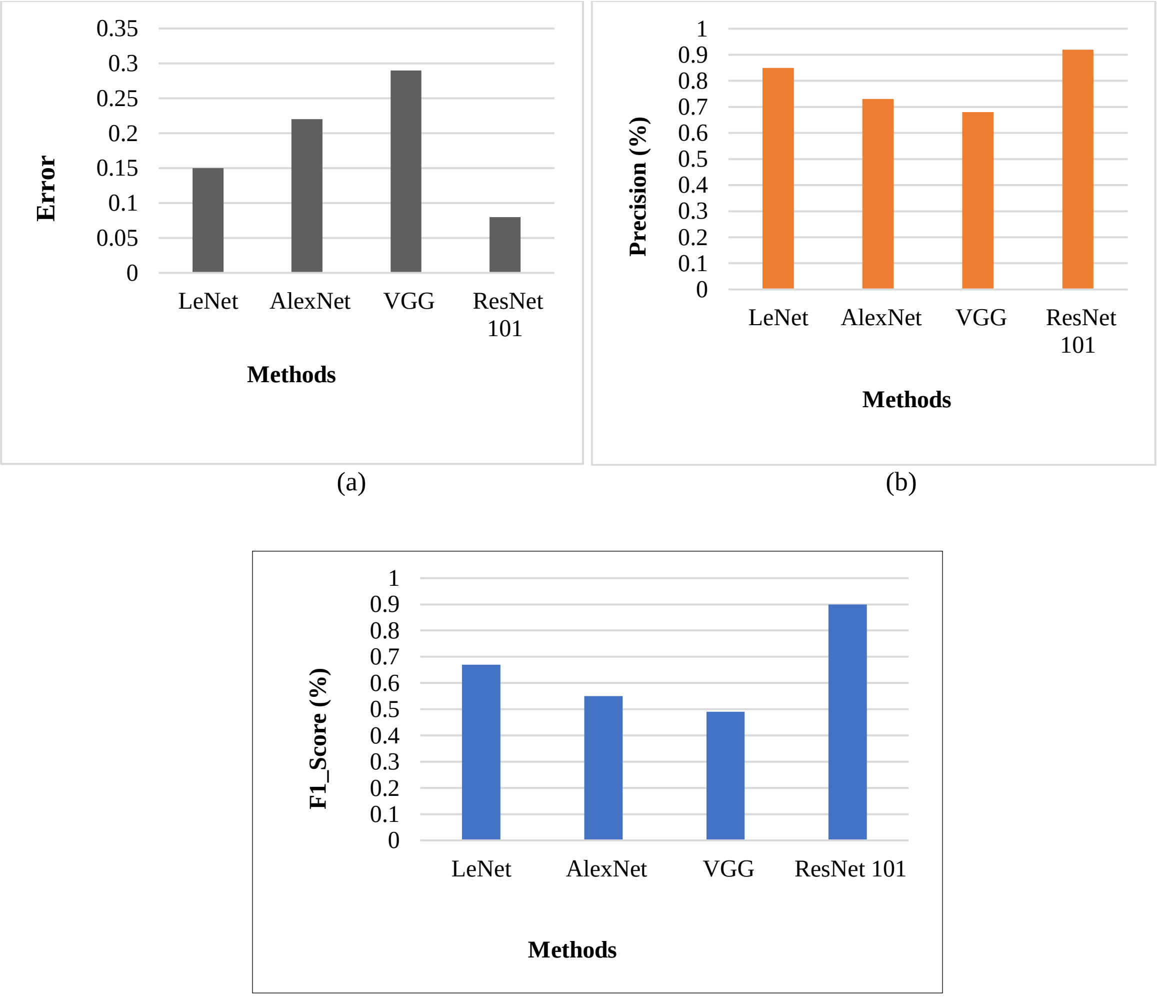
(b)

(c)

Figure 4 Comparison of (a) accuracy (b) sensitivity (c) specificity

Figure 4 (a) compares the accuracy of proposed and current deep learning algorithms. The proposed method's accuracy was determined to be 92%, which is higher than the other three approaches. LeNet has 86% accuracy, AlexNet has 85% accuracy, and VGG has 72% accuracy. Compared to other current approaches our proposed model attained high accuracy. Figure 4 (b) illustrates the comparison of sensitivity between the proposed and current deep learning approaches. The proposed model achieved higher sensitivity than other contemporary methods. The sensitivity value for the developed approach is 98%, which surpasses the three existing methods. LeNet has 83% of sensitivity, AlexNet has 79% of sensitivity, and VGG has 73% of sensitivity. Figure 4 (c) displays the specificity comparison between the proposed method and current deep learning algorithms. The proposed model achieved higher specificity than the existing methods. The specificity for the proposed method is measured at 85%, which surpasses the three other existing methods. LeNet has a specificity of 84%, AlexNet has 76%, and VGG has 61%.

Figure 5 (a) compares the error rates of the proposed method with current algorithms. The proposed model yielded a lower error value than existing methods. The error for the proposed method is measured at 0.08, which is lower than the three other existing methods. LeNet has an error of 0.15, AlexNet has an error of 0.22, and VGG has an error of 0.29. Figure 5 (b) compares the precision of the proposed model to that of the currently available deep learning algorithms. The proposed model achieved a precision of 92%, which is more than the existing three methods. LeNet has 85% of precision, AlexNet has 73% of precision, and VGG has 68% of precision.



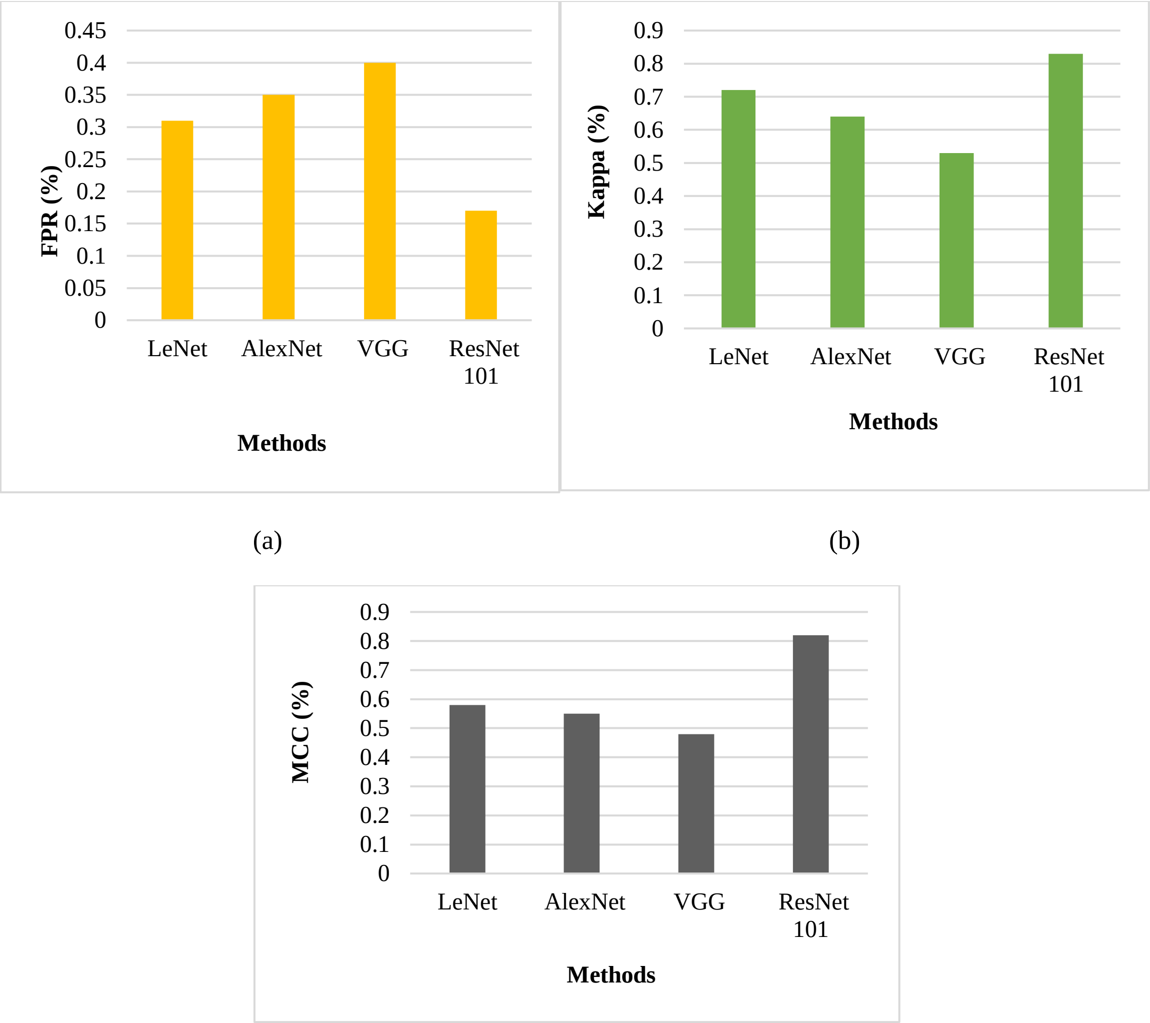
(a) (b)

(c)

Figure 5 Comparison of (a) error (b) precision (c) f1_score

Figure 5 (c) demonstrates the comparison of the F1 score between the proposed method and the currently used deep learning algorithms. The proposed model achieved a higher F1 score than other contemporary approaches. The F1 score for the proposed method is measured at 90%, surpassing the three other existing methods. LeNet has an F1 score of 67%, AlexNet has 55%, and VGG has 49%.

Figure 6 (a) compares the FPR for proposed and current deep learning algorithms. The proposed model attained less FPR value than the existing methods. The FPR for ResNet-101 is found to be 0.17, which is less than the existing three methods. LeNet has 0.31 of FPR, AlexNet has 0.35 of FPR and VGG has 0.40 of FPR. When compared to the other methods, the proposed ResNet-101 attained 0.14% less value than LeNet, 0.18% less value than AlexNet, 0.23% less value than VGG. This shows the effectiveness of the proposed technique. Figure 6 (b) compares the Kappa scores of developed and current deep learning algorithms. The proposed model attained high kappa value compared to other existing methods. Kappa value for the developed method is found to be 83%, which is more than the existing three methods. LeNet has 72% of kappa, AlexNet has 64% of kappa, and VGG has 53% of kappa.



(a) (b)

(c)

Figure 6 Comparison of (a) FPR (b) Kappa (c) MCC

Figure 6 (c) compares the MCC of the proposed method with existing deep learning algorithms. The ResNet-101 model achieved a higher MCC than other currently used methods. The MCC for the proposed method is found to be 82%, which is higher than the three existing methods. LeNet has an MCC of 58%, AlexNet has an MCC of 55%, and VGG has an MCC of 48%.

According to Table 2, which shows the comparison of the proposed and existing deep learning strategies. The proposed ResNet101 approach achieved 92% accuracy. DOG+CNN technique in author (Shin et al., 2016) attained 83.72%, Fused CNN technique in author (Alif et al., 2018) attained 83.72% of accuracy, Raw+CNN technique in author (Shin et al., 2016) attained 62.2% of accuracy, Gender+CNN technique in author (Dar et al., 2020) attained 94% of accuracy, CNN+kernel size and number of filters technique in author (Agrawal & Mittal, 2020) attained 65%, DCT+CNN technique in author (Shin et al., 2016) attained 56.09% of accuracy, multi-view DCNN technique in author (Alfakih et al., 2020) attained 72.27% of accuracy, Is+CNN technique in author (Shin et al., 2016) attained 62.16% of accuracy, CNN+pre-processing technique in author (Lopes et al., 2017) attained 82.10% of accuracy and Hist+CNN technique in author (Shin et al., 2016) attained 66.67% of accuracy.

Table 2 Comparison of accuracy among proposed and existing deep learning strategies

Methods	Author	Accuracy (%)
DOG+CNN	Shin, Kim and Kwon (2016)	58.96
Fused CNN	Alif et al. (2018)	83.72
Raw+CNN	Shin et al. (2016)	62.2
Gender+CNN	Dar, Javed, Bourouis, Hussein and Alshazly (2022)	94
CNN+kernel size and number of filters	Agrawal & Mittal (2020)	65
DCT+CNN	Shin et al. 2016	56.09
multi-view DCNN	Alfakih et al. (2020)	72.27
Is+CNN	Shin et al. 2016	62.16
CNN+preprocessing	Lopes, De Aguiar, De Souza and Oliveira-Santos (2017)	82.10
Hist+CNN	Shin et al. 2016	66.67
ResNet101+preprocessing	Proposed	92

Table 3 Comparison of accuracy utilizing dissimilar techniques

Methods	Expression	Accuracy (%)
DNN-Driven (Zhang et al., 2016)	5	80.10
C-CNN (Lopes et al., 2017)	6	91.64
PCRF (Bailly, & Dubuisson, 2017)	5	76.1
JFDNN (Jung, Lee, Yim, Park, & Kim, 2015)	5	72.5
HOG (Kumar, & Kumar, 2017)	6	89.70
Proposed ResNet101	6	92

When compared to other techniques used for human facial emotion recognition, the proposed ResNet-101 model achieved better accuracy, underscoring the effectiveness of the detection model.

Table 3 shows the comparison analysis of accuracy utilizing different techniques through BU-3DFE dataset. According to the analysis the proposed technique attained better accuracy than other methods. The proposed ResNet-101 attained 92% of accuracy which is 11.9% higher than DNN-Driven (Zhang et al., 2016), 0.36% less than C-CNN (Lopes et al., 2017), 12.9% less than PCRF (Bailly, & Dubuisson, 2017), 19.5% less than JFDNN (Jung et al., 2015) and 2.3% less than HOG (Kumar, & Kumar, 2017).

5. Conclusion

This research focuses on designing a model for recognising human emotion using ResNet-101. The proposed model consists of three phases such as data gathering, pre-processing and classification. The initial phase was data gathering, and the dataset includes human emotion-related images. The second phase was preprocessing, and the image dataset was subjected to preprocessing to resize the image as well as eliminate the noise contents present in the images. The ResNet 101 model was adopted to categorize facial expressions. After preprocessing the outcome was given as the input of the ResNet 101, here convolutional layers were responsible to extract the image features and then the fully connected layer categorized these image features into six different classes like, fear, anger, sad, happy, neutral and surprised. The performance analysis was performed by comparing the proposed technique, i.e. ResNet-101, with several convolution neural network structures such as LeNet, AlexNet and VGG. The experimental outcomes reveal that the applied methods executed with mostly better results considering the 92% of accuracy, 0.08 error, 92% of precision, 85% of specificity and 98% of sensitivity and so on using human emotion image dataset. This outcome reveals the proposed model attained better compared with several convolution neural networks. Although some characteristics and services like Microsoft Cognitive Services are becoming more popular, much work still has to be done to improve their effectiveness, accuracy, and usability. Emotion Recognition will therefore demand a lot more focus in the future. The model will eventually be enhanced to categorize primary and secondary emotions in real-time video and image data. The future scope is also focused on increasing the results' accuracy and packaging the model into a consumable service that can be used by a variety of collaborating systems.

6. Declarations

Funding: There is no funding provided to prepare the manuscript.

Conflict of Interest: The process of writing and the content of the article does not give grounds for raising the issue of a conflict of interest.

Ethical Approval: This article does not contain any studies of human participants or animals performed by any of the authors. The images are retrieved from a dataset in Kaggle (ARES. (2013) with no copywrite, and allow to copy, modify, distribute and perform without asking permission.

Data availability statement: If all data, models, and code generated or used during the study appear in the submitted article and no data need to be specifically requested.

Code availability: No code is available for this manuscript.

7. References

Abdullah, A. I. (2019). Facial Expression Identification System Using fisher linear discriminant analysis and K-Nearest Neighbor Methods. *ZANCO Journal of Pure and Applied Sciences*, *31*(2), 9-13. https://doi.org/10.21271/ZJPAS.31.2.2

Agrawal, A., & Mittal, N. (2020). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, *36*(2), 405-412. https://doi.org/10.1007/s00371-019-01630-9

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics*, *10*(9), Article 1036. https://doi.org/10.3390/electronics10091036

Al Machot, F., Elmachot, A., Ali, M., Al Machot, E., & Kyamakya, K. (2019). A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors. *Sensors*, *19*(7), Article 1659. https://doi.org/10.3390/s19071659

Alfakih, A., Yang, S., & Hu, T. (2020). *Multi-view cooperative deep convolutional network for facial recognition with small samples learning: Distributed Computing and Artificial Intelligence* [Conference presentation]. *16th International Conference* (pp. 207-216). Springer International Publishing. https://doi.org/10.1007/978-3-030-23887-2_24

Alif, M. M. F., Syafeeza, A. R., Marzuki, P., & Alisa, A. N. (2018). Fused convolutional neural network for facial expression recognition. *Proceedings of the Symposium on Electrical, Mechatronics and Applied Science* (SEMA’18) (pp. 73-74).

Amrani, M., & Jiang, F. (2017). Deep feature extraction and combination for synthetic aperture radar target classification. *Journal of Applied Remote Sensing*, *11*(4), 042616-042616. https://doi.org/10.1117/1.JRS.11.042616

Amrani, M., Bey, A., & Amamra, A. (2022). New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis. *International Journal of Remote Sensing*, *43*(15-16), 5800-5819. https://doi.org/10.1080/01431161.2021.1953719

Amrani, M., Jiang, F., Xu, Y., Liu, S., & Zhang, S. (2018). SAR-oriented visual saliency model and directed acyclic graph support vector metric based target classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(10), 3794-3810. https://doi.org/10.1109/JSTARS.2018.2866684

Amrani, M., Yang, K., Zhao, D., Fan, X., & Jiang, F. (2017, September 28-29). *An efficient feature selection for SAR target* *classification* [Conference presentation]. Advances in Multimedia Information Processing–PCM 2017: 18th Pacific-Rim Conference on Multimedia, Harbin, China, Revised Selected Papers, Part II. Springer, Cham.

Arabian, H., Wagner-Hartl, V., Chase, J. G., & Möller, K. (2021). Image Pre-processing Significance on Regions of Impact in a Trained Network for Facial Emotion Recognition. *IFAC-Papers OnLine, 54*(15), 299-303. https://doi.org/10.1016/j.ifacol.2021.10.272

ARES. (2013). Emotion Detection. *Kaggle.* Retrieved from https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer

Bailly, K., & Dubuisson, S. (2017). Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. *IEEE Transactions on Affective Computing*, *10*(2), 167-181. https://doi.org/10.1109/TAFFC.2017.2708106

Bharati, S., Podder, P., Mondal, M. R. H & Prasath, V. B. S. (2021) CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images. *International Journal of Hybrid Intelligent Systems Preprint*, *17*(1-2), 1-15. https://doi.org/10.3233/HIS-210008

Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., & Velastin, S. A. (2021). Facial expression recognition of instructor using deep features and extreme learning machine. *Computational Intelligence and Neuroscience*, *2021*, 1-17. https://doi.org/10.1155/2021/5570870

Chen, J. X., Zhang, P. W., Mao, Z. J., Huang, Y. F., Jiang, D. M., & Zhang, Y. N. (2019). Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access, 7,* 44317-44328. https://doi.org/10.1109/ACCESS.2019.2908285

Chen, L., Zhou, M., Su, W., Wu, M., She, J., & Hirota, K. (2018). Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. *Information Sciences*, *428*, 49-61. https://doi.org/10.1016/j.ins.2017.10.044

Dar, T., Javed, A., Bourouis, S., Hussein, H. S., & Alshazly, H. (2022). Efficient-SwishNet Based System for Facial Emotion Recognition. *IEEE Access*, *10*, 71311-71328. https://doi.org/10.1109/ACCESS.2022.3188730

Do, L. N., Yang, H. J., Nguyen, H. D., Kim, S. H., Lee, G. S., & Na, I. S. (2021). Deep neural network-based fusion model for emotion recognition using visual data. *The Journal of Supercomputing*, *71*, 10773–10790. https://doi.org/10.1007/s11227-021-03690-y

Gupta, V., Chopda, M. D., & Pachori, R. B. (2018). Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals. *IEEE Sensors Journal*, *19*(6), 2266-2274. https://doi.org/10.1109/JSEN.2018.2883497

Hassan, M. M., Alam, M. G. R., Uddin, M. Z., Huda, S., Almogren, A., & Fortino, G. (2019). Human emotion recognition using deep belief network architecture. *Information Fusion*, *51*, 10-18. https://doi.org/10.1016/j.inffus.2018.10.009

Hossain, M. S., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio–visual emotional big data. *Information Fusion*, *49*, 69-78. https://doi.org/10.1016/j.inffus.2018.09.008

Jiang, D., Wu, K., Chen, D., Tu, G., Zhou, T., Garg, A., & Gao, L. (2020). A probability and integrated learning based classification algorithm for high-level human emotion recognition problems. *Measurement*, *150,* Article 107049. https://doi.org/10.1016/j.measurement.2019.107049

Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. (2015). Joint fine-tuning in deep neural networks for facial expression recognition. *Proceedings of the IEEE international conference on computer vision* (pp. 2983-2991). https://doi.org/10.1109/ICCV.2015.341

Kittipongdaja, P., & Siriborvornratanakul, T. (2022). Automatic kidney segmentation using 2.5 D ResUNet and 2.5 D

DenseUNet for malignant potential analysis in complex renal cyst based on CT images. *EURASIP Journal on Image and Video Processing*, *2022*(1), 1-15. https://doi.org/10.1186/s13640-022-00581-x

Kumar, M., & Kumar, A. (2017). Decision making behaviour of elderly tribal people at household and community level in rural eastern Uttar Pradesh, India. *Asian Journal of Research in Social Sciences and Humanities*, *7*(7), 289-302. https://doi.org/10.5958/2249-7315.2017.00387.2

Li, B., & Lima, D. (2021). Facial expression recognition via ResNet-50. *International Journal of Cognitive Computing in Engineering*, *2*, 57-64. https://doi.org/10.1016/j.ijcce.2021.02.002

Liu, B., Ye, C., Yang, P., Miao, Z., Liu, R., & Chen, Y. (2022, February). *A Segmentation Model of Lung Parenchyma in Chest CT Based on ResUnet.*[Conference presentation]. *14th International Conference on Machine Learning and Computing (ICMLC)* (pp. 429-434). https://doi.org/10.1145/3529836.3529917

Lopes, A. T., De Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern recognition*, *61*, 610-628. https://doi.org/10.1016/j.patcog.2016.07.026

Maheswari, D., & Radha, V. (2010). Noise removal in compound image using median filter. *International Journal on Computer Science and Engineering*, *2*(04), 1359-1362.

Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), Article 446. https://doi.org/10.1007/s42452-020-2234-1

Mohammed, S. B., & Abdulazeez, A. M. (2021). Deep Convolution Neural Network for Facial Expression Recognition. *PalArch's Journal of Archaeology of Egypt/ Egyptology*, *18*(4), 3578-3586.

Nannapaneni, R., & Chatterjee, S. (2021). Human emotion recognition through facial expressions. *Machine Intelligence and Smart Systems: Proceedings of MISS 2020* (pp. 513-525). Singapore: Springer. https://doi.org/10.1007/978-981-33-4893-6_44

Ngai, W. K., Xie, H., Zou, D., & Chou, K. L. (2022). Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Information Fusion*, *77*, 107-117. https://doi.org/10.1016/j.inffus.2021.07.007

Pal, S., Mukhopadhyay, S., & Suryadevara, N. (2021). Development and progress in sensors and technologies for human emotion recognition. *Sensors*, *21*(16), Article 5554. https://doi.org/10.3390/s21165554

Perumal, S., & Velmurugan, T. (2018). Preprocessing by contrast enhancement techniques for medical images. *International Journal of Pure and Applied Mathematics*, *118*(18), 3681-3688.

Said, Y., & Barr, M. (2021). Human emotion recognition based on facial expressions via deep learning on high-resolution images. *Multimedia Tools and Applications*, *80*(16), 25241-25253. https://doi.org/10.1007/s11042-021-10918-9

Salama, E. S., El-Khoribi, R. A., Shoman, M. E., & Shalaby, M. A. W. (2021). A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition. *Egyptian Informatics Journal*, *22*(2), 167-176. https://doi.org/10.1016/j.eij.2020.07.005

Shin, M., Kim, M., & Kwon, D. S. (2016, August). *Baseline CNN structure analysis for facial expression recognition.* [Conference Presentation]. 25th IEEE international symposium on robot and human interactive communication (RO-MAN) (pp. 724-729) IEEE. https://doi.org/10.1109/ROMAN.2016.7745199

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of selected topics in signal processing*, *11*(8), 1301-1309.

https://doi.org/10.1109/JSTSP.2017.2764438

Wang, S. H., Phillips, P., Dong, Z. C., & Zhang, Y. D. (2018). Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing*, *272*, 668-676. https://doi.org/10.1016/j.neucom.2017.08.015

Wattana, A., Janpong, S., & Supichayanggoon, Y. (2018). Downdraft gasifier identification via neural networks. *Journal of Current Science and Technology*, *2*(8), 87-98. https://doi.org/10.1016/S0893-6080(18)30023-6

Zhang, T., Zheng, W., Cui, Z., Zong, Y., & Li, Y. (2018). Spatial–temporal recurrent neural network for emotion recognition. *IEEE transactions on cybernetics*, *49*(3), 839-847. https://doi.org/10.1109/TCYB.2017.2788081

Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., & Yan, K. (2016). A deep neural network-driven feature learning method for multi-view facial expression recognition. *IEEE Transactions on Multimedia*, *18*(12), 2528-2536. https://doi.org/10.1109/TMM.2016.2598092