Journal of Integrated
SCIENCE & TECHNOLOGY

Machine learning model for emotion detection and recognition using an enhanced Convolutional Neural Network

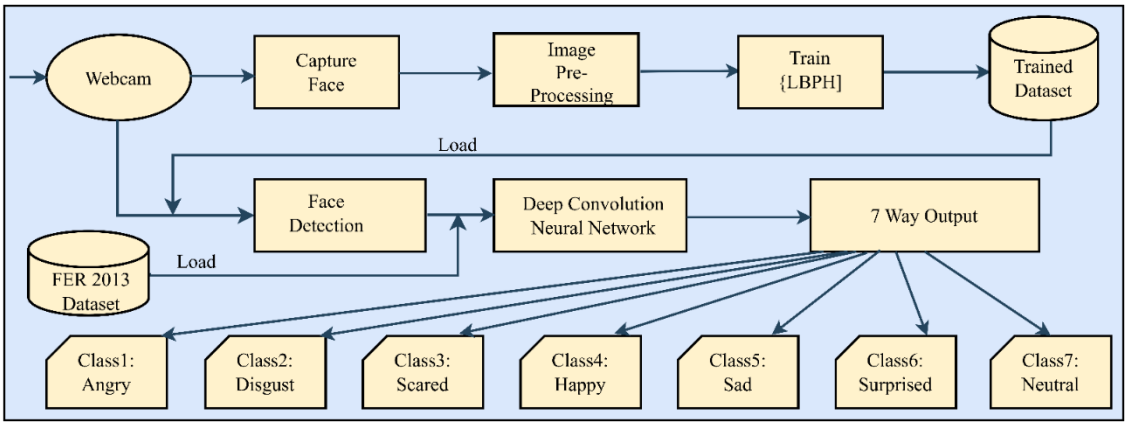
Sowmya BJ[1*], Meeradevi[2], Sini Anna Alex[3], Anita Kanavalli[1], Supreeth S[4], Shruthi G,[4] Rohith S[5]

[1]Department of Artificial Intelligence and Data Science, M S Ramaiah Institute of Technology, 560054, India. [2]Department of Artificial Intelligence and Machine Learning, M S Ramaiah Institute of Technology, 560054, India. [3]Department of CSE (AI & ML), M S Ramaiah Institute of Technology, 560054, India. [4]School of Computer Science and Engineering, REVA University, 560064, India. [5]Department of ECE, Nagarjuna College of Engineering & Technology, 562110, India.

Received on: 29-Sep-2023, Accepted and Published on: 27-Dec-2023

Article

ABSTRACT



Emotion expression recognition has been a challenging task in recent years due to large intra-class variation and persistent difficulty. Most studies fail on datasets with image variations and partial faces but work best on controlled datasets. Recent work using deep learning models has improved emotion recognition by developing mini-Xception based on Xception and Convolution Neural Network (CNN). This system can focus on important parts like the face, performing face recognition, and emotion classification simultaneously. A visualization method is used to distinguish between different emotions based on the classifier results. An experimental study on the FER-2013 dataset demonstrated that the mini-Xception algorithm successfully performed all tasks, including emotion recognition and classification, with an accuracy of approximately 95.60%.

Keywords: Emotion detection, Face recognition, Electric signal, Artificial Intelligence System

INTRODUCTION

Many situations require the use of human emotion detection when more security or knowledge of a person is required. It can be viewed as a continuation of face detection, in which case a two-phase security system that detects both facial and emotional activity may be needed. It may be ensured that the person being shot is only a 2-D depiction of them.[1]

Emotions have been the subject of numerous research investigations, however, there is no single definition of emotions in the literature. Emotion may be defined as the reflection or actualization of feeling. In addition to feeling, it may also be fake or actual.[2] For instance, while feelings cannot be felt completely and exactly, the experience of pain simply conveys the feeling. The inner esoteric interior condition of affairs is described by a feeling. Emotion has a crucial role in many study domains, including psychology, healthcare, biomedical engineering, and even neurology, and it has grown to be a vast research topic. Biomedical engineering has a strong interest in the research of emotion recognition. Recent work on this subject has centered on motion detection and automated computer-aided detection of psychological disorders.[2] Researchers have used a variety of tools to define emotions, such as multimodal, GSR, EEG, face expression, visual scanning actions, and galvanic skin response, etc. Over the years, deep learning has become more common and has made great strides in the arena of picture grading. One of the most commonly used and well-known methods for deep learning for

*Corresponding Author: Sowmya BJ
Email: researchrit1985@gmail.com

Cite as: *J. Integr. Sci. Technol., 2024, 12(4), 786.*
URN:NBN:sciencein.jist.2024.v12.786

image segmentation, recognition, and classification is Convolutional Neural Networks (CNNs).[3–8]

Among the most well-known ways for determining an individual's emotion using deep learning-based algorithms, Mini-Xception, a deep learning-based algorithm, was developed by us. The designed model's main objective is to accurately predict emotional states and automatically detect emotions. In this method, tagged face expression pictures out of the FER file are used to analyze the experimental outcomes. A created copy gets the pictures as a load and is trained to use them. The designed model decides which face expression is used after that.

Business promotions are a key area where emotion recognition is important. Most businesses rely on customer feedback for all their services and products to stay in business. An artificial intelligence algorithm can determine whether a user likes or dislikes a product or offer by identifying real feelings in a photo or video. This research has shown that the most frequent reason to identify someone is for their safety. It is feasible to use passwords, voice recognition, retina scanning, fingerprint matching, and other methods. To reduce risk, it is essential to ascertain someone's purpose. This is useful in high-risk locations where security breaches have recently occurred, such as airlines, concerts, and large public gatherings. The three main components of emotion detection are depicted in Figure 1. The following are the steps:

- Image Pre-processing
- Feature Extraction
- Feature Grading

Figure 1. Process of identifying human emotion

Fear, disgust, rejection, rage, surprise, sadness, happiness, and neural are all human emotions. These are truly minor emotions. It is difficult to identify them because even slight variations in facial muscle contortions produce distinct expressions.[9] Furthermore, various individuals may express the same emotion in different ways because emotions are very context-dependent. Even though attention may be drawn to features other than the facial that express emotions, such as the lips and optics, the question of how to extract and categorize these gestures is still important. These goals were achieved with the help of the synaptic and the expert systems.

The computational classification and categorization methods are depicted as being very useful. Attributes are the best key segments of any expert system strategy. For algorithms like Support Vector Machines, explore how attributes are extracted and updated in this work.[1] There will be comparisons between the algorithms and feature extracts used in various articles. The person feeling data may be used to evaluate categorization algorithms' strengths and nature, and how well they function across various sources of data. Facial recognition techniques are often used on the video or image frame before extracting features for emotional identification. The emotion detection steps are listed below:

- The Pre-processing of data sets
- Face recognition
- Extraction of features
- Classification based on characteristics

Facial Emotion Recognition

FER is typically categorized into four stages. The first stage identifying a face in an image and drawing a rectangle around it. Finding landmarks inside the face area comes next. The third stage is to clip the dimensional and transitory characteristics of the face components. Finally, a Feature Extraction (FE) classifier and attribute removal are used to generate a recognition result. Facial markers comprise the tip of the nose, the corners of the lips, and the ends of the brows. Features comprise the local texture of a landmark and the alignment of two visible signs. The dimensional and transitory attributes of a face are eliminated using pattern classifiers, and one of the face classifications is used to determine circulation.

Due to the ability to perform end-to-end learning using face-to-face physics-based modeling and other preprocessing techniques, DL-based FER solutions significantly reduce the amount of training required.[10] To use a CNN, a feature map is produced by convolutional filtering of the input picture. Following that, complete connection layers are loaded with the output of the FE classifier, which identifies the face circulation as fit into a group. The Facial Emotion Recognition 2013 (FER 2013) dataset was utilized to train this design (Fatima et al., 2021). For a Kaggle competition, this open-source file was produced for work and made accessible to the general public. 35,000 48 x 48 color face images with unique mood tags are used in something. There are five emotions used: dread, sadness, anger, and neutrality.

Organization of CNN

The basic CNN blueprint, with several elements that are simple to comprehend and relate to the designed CNN model. A basic CNN is made up of load, secret, and output layers. The load covering is where the data centers are located at CNN. It then goes through various secret measures before addressing the last layer. An output layer represents the prediction made by the network. The output of the network is compared to the real labels to detect any loss or error.

The network's hidden layers are the fundamental building blocks for data transformation. Each layer can be broken into its four actions: layer function, pooling, normalizing, and activation. The following layers help compensate for the architecture of the CNN. The CNN framework from the viewpoint of our work has also been considered.

- Convolution layer
- ReLu layer
- Pooling loop
- Fully connected layer
- Softmax
- Batch normalization

The main objectives of the work are:

- The objective of emotion recognition is to identify the emotions of a human.

- The purpose of carrying out this exploration is to accurately classify seven main emotions: happiness, surprise, anger, disgust, neutrality, and fear.
- The purpose of this research is to analyze the outcome of models in terms of precision in each class.
- An emotion can be captured either from a face or from a .csv file.
- ML may be used to deliver FER solutions that are cheap, reliable, and computation-intensive.

People often express their emotions on their faces during social encounters to demonstrate their characteristics and feelings. This research's primary objective is to derive feelings to which pictures with a single-face expression relate. Feelings can be recognized that are specifically split into the categorization of fundamental feelings and the classification of composite feelings because of the difficulty of reading a human face. The key challenge for the work and scope is to focus on classifying the seven fundamental feelings as happiness, sadness, surprise, neutrality, disgust, anger, and fear.

LITERATURE REVIEW

Visualization is a technique used to enhance an objective image or extract useful data from it. Nonverbal communication takes the form of facial expressions, and it is important to recognize these emotions on the face. A technology-based monitoring system for elderly people that detects emotions from video images is proposed, which includes video analysis technology to enable real-time monitoring of elders' living conditions. In the case of an emergency, the system will send a message to their relatives and children. It explains that face identification has been around for centuries, but emotion detection is essential for modern AI systems. It requires a range of algorithms for feature extraction, and expert system techniques can be used to complete the task. The study examined less system learning algorithms and attribute removal methods to aid in the more accurate detection of human emotion.

Deep learning is used to recognize human emotions through facial expressions, using the Kaggle FER2013 dataset to experiment with and train a deep convolutional network. This work has been implemented in a real-time system with great success.[11] This research used the Haar-Cascade Classifier and CNNs to classify facial emotion. Results showed that the CNN architecture gained MSE and accuracy values based on epoch variety, increasing the value, decreasing the MSE rate, and increasing the accuracy value. It has been demonstrated that the CNN algorithm is effective in recognizing facial expressions. Mehta et.al.[12] discussed automated face recognition due to the growing appeal for conducting biometric frameworks and human–machine interaction.

This study used Gabor filters, histogram-oriented gradients, a local binary array, a support vector machine, a random forest, and the nearest neighbour method to extract features, categorize emotions, and estimate the depth of such feelings in a directory. The results showed that this study can be used for actual-time behavioral of eye contact and depth notice.

The studies network has been interested in the capability to make use of human face emotion identity (FER). Deep Neural Nets, especially Convolutional Neural Networks, are utilized to collect emotional facts from high-resolution images. The application makes use of a Deep CNN model for growing a highly accurate FER device that uses transfer learning techniques. A novel pipe method was introduced to increase FER accuracy.

Emotion recognition is a hot topic in science, used in robotics view and correlated with robotic connections. This paper proposes an actual-time path for implementing feeling detection in the robotic view function using the Media Pipe facial grid method and Principal Component Analysis.[13] Mellouk and Handouzi (2020)[22] describe the automated feeling based on face explanation, which is used in various areas such as sentient interactions, wellness, and security. This paper provides a view of previous efforts on deep learning-based automatic facial emotion recognition. Jaiswal et.al.[14] projected a deep structure design on CNNs for feeling depiction of snapshots, which was evaluated by the usage of sets: Face Emotion explanation task and the Japanese girl face Emotion dataset. The designed system produced 70.14 and 98.65 percent, respectively.

Liliana et.al.[10] used a deep Convolutional Neural Network to detect facial expressions. They use a regularized technique called "dropout" in the CNN fully connected loops to reduce overfitting. The expanded Cohn Kanade dataset is used in the research, and the mean certainty rate of the system has increased up to 2%. The basic feeling set has been successfully categorized by the system, showing that it is active in emotion recognition. Fatima et.al.[2] have reported a mini-Xception based on Xception and a CNN to improve emotion expression recognition. They developed a system of seeing that uses the mini-Xception platform to perform face analysis and feeling categorization. The method can effectually complete all tasks, including emotion detection and classification, with an accuracy of about 95.60%, according to the exploratory investigation on the FER-2013 dataset.

Human emotions are spontaneous mental states of sentiments, and the research of facial recognition is challenging. Human emotions are spontaneous mental states of sentiments, and the research of facial recognition is challenging. Machine learning and neural networks are used to identify emotions, and features are extracted from images by using CNN. CNN Model is 80% accurate for the four emotions, and 72% accurate for the five emotions. CNN Pattern 2 is 79% accurate for the four and 72% accurate for the five.[15] Khare et.al.[16] used Deep learning with Convolutional Neural Networks (CNNs) to provide input images of human facial expressions for pre-trained models to be trained on datasets. The traditional approach can become corrupt due to changes in light and object location, making feature engineering difficult.

Sentiment analysis (SA) is used to assess the author's sentiment,[17] a variety of fields, such as forecasting, agriculture, psychology, the judiciary, social media, and the stock market. This work looks into various neural network-based methods for sentiment analysis, including lexicon- and ontology-based SA and machine learning. A customized SVM for machine learning is used to detect 6 distinct feelings using 68-point facial landmarks and video and system learning.[18] Alhalaseh & Alasasfeh, 2020)[17] studied how to build an automated system to identify emotions using brain signals. Four algorithms were used to classify emotional states: nave Bayes, K-Nearest Neighbour (K-NN), CNN, and

Decision Tree (DT). The effectiveness of the suggested tasks was evaluated using metrics like accuracy, specificity, and sensitivity.

Badrulhisham et.al[20] reviewed how emotions are expressed through body language, vocal inflection, and face expression. This study created a mobile-system sentiment recognition application that can identify feelings using facial expressions in real-time using Convolutional Neural Network (CNN) and MobileNet algorithm. [21], projected a real-time emotion detection approach using CNN and pre-processing techniques to improve model performance. The authors in [22] provide a narrative method for detecting facial emotions using convolutional neural networks (FERC) to extract face feature vectors and identify five varieties of normal face reactions. Supervisory information is obtained using a 10,000 image database, resulting in 96 % accuracy. Face emotion detection is used to analyze facial expressions of sadness, happiness, surprise, anger, and fear to determine a woman's emotional state in [23]. Machine learning algorithms are used to estimate sentiment using transformed photos.

DESIGN AND IMPLEMENTATION

The diagram step for developing software involves creating the specifications, goals to achieve, and so on which are subject to constraints. In the design process, have set the structure for the units that are integrated together to form the final system. The design description allows programmers to easily program modules and integrate them by following the design as depicted in Figure 2.

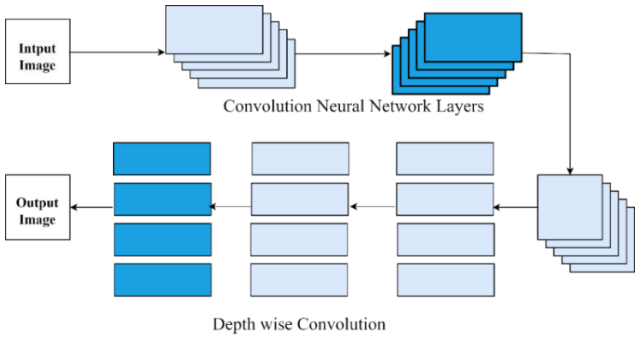


Figure 2. The architecture of Designed Model A

The designed architecture consists of four residual depth-wise separable convolutions, with a batch normalization operation and ReLUs activation function come after each convolution. To provide a prediction, the last layer applies a soft-max activation function and global average pooling.

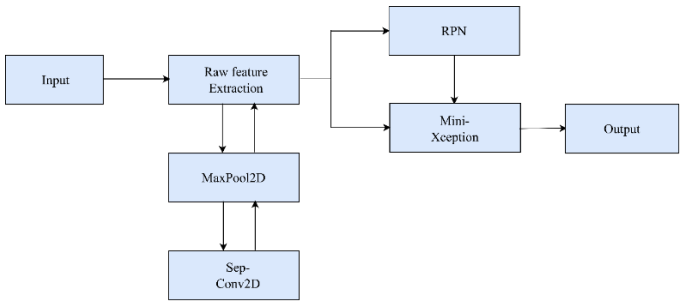


Figure 3. The architecture of designed Model B

There are above 60,000 parameters in this architecture. Fig. 2 and 3 depict the entire architecture, also referred to as mini-Xception. Residual modules modify the desired mapping between two subsequent layers so that the learned features become the difference between the original feature map and the desired features. Consequently, the desired features H(x) are modified to solve an easier learning problem F(X) such that:

$$H(x) = F(x) + x \qquad (1)$$

The following is a description of each component:

- Input:

The input layer is the source for the whole CNN. It normally constitutes the picture's pixel matrix in neural networks used for image processing. Must provide the CNN a 4D array as input. Thus, input data is in the form of height, width and size, where the first dimension denotes the batch size of the picture and the other three denote the height, breadth, and depth of the image, respectively.

- Pooling:

The pooling layer reduces the spatial size of a dispersed feature. The extent of work necessary to examine the data and retrieve the key, rotation and position-invariant structures is condensed as a result. Pooling can be divided into two types: maximum pooling and average pooling. The mean of matching values is the result produced by the pooling layer and max pooling outputs the maximum integer using the area of the photographs covered by the kernel. An outcome applies maximum and median pools in picture.[15]

- Max Pooling:

The term "max pooling" refers to a pooling operation that selects an item from the feature map region that the filter covers. The output of the max-pooling layer would thus be a feature map that includes the most noticeable features from the previous feature map.

Maxpooling down trials the input along its spatial dimensions by taking the extreme value for each input channel over an input window of the pool size. Each dimension of the window is moved one step at a time. The following is the pseudo-code:

tf.keras.layers.MaxPooling2D(pool_size= (2, 2), strides=None, padding="valid", data_format=None, **kwargs)

- Sep-conv2D:

This layer creates a convolution kernel which is also combined with the layer input to produce a matrix of output. Convolution matrix or masks are used in image processing to blur, sharpen, emboss, detect edges, and execute other tasks by convolutioning a kernel and an image. Convolutional Neural Networks (CNNs) may be used to classify data using image frames & and learn features. There are several types of CNNs. Depth-wise separable CNNs are one type of CNN.

Depth-wise separable convolutional networks' effectiveness over simple CNNs is discussed in this article along with the design and operation of these networks. Suppose that the input data has the

dimensions: Df x Df x M, where M is the number of routes and Df x Df can be the size of the image (3 for an RGB image). Suppose there are N filters or kernels of size Dk x Dk x M. The output size if a standard convolution action is carried out will be Dp x Dp x N.

The size of filter = Dk x Dk x M is the number of multiplications in one convolution operation. The total number of multiplications is N x Dp x Dp x since there are N filters and each filter slides vertically and horizontally Dp times (Multiplications per convolution).

Thus, for a standard convolution operation, the total number of multiplications = N x Dp^2 x Dk^2 x M. Separable convolution consists of two different convolution layers such as:

- Depth-wise convolutions:

Unlike standard CNNs, where convolution is applied to all M channels at once, the depth-wise operation only applies convolution to one channel at a time. The filters/kernels used here will thus be Dk x Dk x 1. Given load information as M channels, M where refinement is required. The outcome will be Dp x Dp x M in size.

- Point-wise Convolutions:

Point-wise processing employs A 1 x 1 convolution operation on ~~the~~ M channels. Hence, the filter size for this operation will be 1 × 1 x M. The output size is Dp x Dp x N when N of these filters are used.

- Region Proposal Network(RPN):

An RPN is a fully convolutional network that consecutively predicts object limits and objectness scores at each location. Fast R-CNN leverages the fully trained RPN to produce outstanding region recommendations for detection.

- Mini-Xception:

A pre-trained convolution model is famous for its cutting-edge performance in several applications, such as object detection and image classification. It was trained on the Image Net dataset. Following the suggested architecture's training, the trained model was assessed in real-time using the below pseudocode.

Step 1. Load the FER-2013 dataset
Step 2: Partition the dataset into training and testing
Step 3: Apply the pre-processing techniques
Step 4: Build the model using the Mini-Xception algorithm
Step 5: Test data is given to Mini-Xception for classification
Step 6: Calculate the Accuracy of Classification of Emotions

- Output:

The receptive field of a convolution combines all the pixels into a single value. A 4D array is another output from the CNN. The three other picture dimensions may alter depending on the filter, kernel size, and padding settings that are used, but the sample size remains the same as the batch size of the input images.

Users can submit the input data to a front end and then click the predict button. Mini-Xception, a deep learning-based algorithm, is used. The model for machine learning gets the training data as a csv file. The user's data is processed before it is sent to the model as input. The model suggests the output with different emotions shown to the user. Modules used in the models are as follows:

- Face capturing module
- Pre-processing module
- Training module
- Face recognition module
- Expression recognition module

Face Capturing Module: Users are now capturing individual faces for the next processing. For this, used a webcam or an external webcam [10]. Without taking the image first, the procedure could be finished, and without doing so, it is impossible to determine the emotions.

Pre-processing Module: will process the captured images after the photos are taken. The colour photos have been converted to grayscale to create the grayscale photos.

Training module: It will be necessary to prepare a dataset in this step, which will be made up of a binary array of the taken images. The images will be saved in a.yml file that contains the collected face data. Since the YML file is compressed, it can process the collected photos more quickly.

Face Recognition Module: Training the host system using the collected facial data is the latest stage in face detection procedures. The subject's face is photographed using computer system's web camera, which records 60 different images of the subject's face [24]. In this section, examines how to use the LBP algorithm to identify faces. The term "local binary pattern histogram" is an acronym. Using previously saved NAME and face ID saved earlier, the faces in the database will be placed.

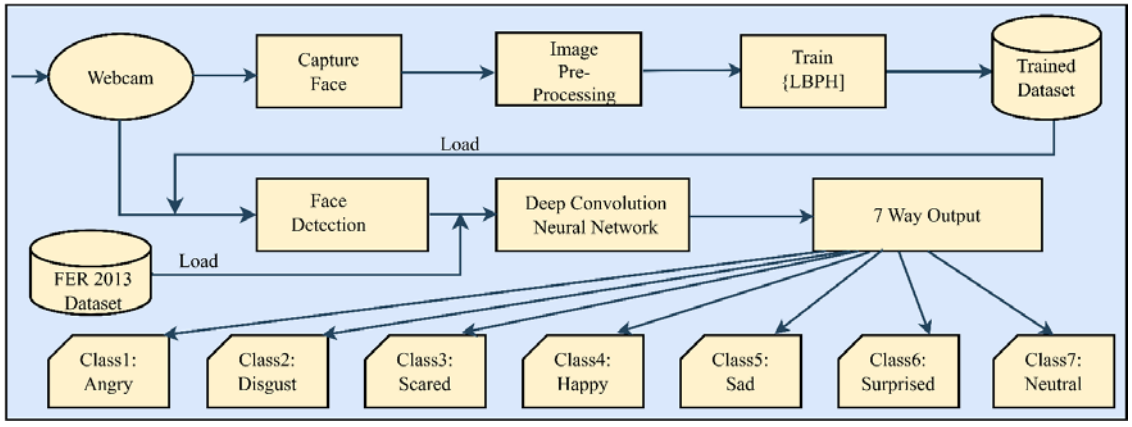


Figure 4. Designed methodology for facial emotion detection

Face Expression Recognition Module: Software for recognizing facial expressions uses biometric data to recognize emotions in human faces [14]. It has the potential to offer an unfiltered, impartial emotional reaction or data because it collects and analyses information from images.

By using software, a webcam is utilized to capture, recognize, and record a person's facial expressions. It is possible to acquire a rectangular frame on the face area in the camera by using the Viola Jones method, the LBPH Face Recognizer algorithm, and the Haar cascade frontal face dataset the rectangular frame on the face area in the camera can be acquired. The face region is distinguished from a non-facial non-facial region in this manner. Before ~~to~~ being saved in a folder labelled with the subject's ID and name, captured person faces are pre-processed. The trained dataset for these pictures is saved as Trainer.yml in the Trainer folder after being trained using the LBPH method. The face on a video camera is matched with the face in the dataset during the Face Detection process that uses a trained dataset. A person's ID and name will be displayed on the screen if their face matches one in the trained dataset. Convolutional neural networks and the FER2013 database are used to perform the classification on the obtained face [2]. Based on the individual's characteristics, facial expression shows the possibility of achieving the highest expression. A facial expression from a possible seven is shown with the subject's identifiable picture. The complete process is depicted in Figure 4.

Figure 5 depicts that in the training phase, the system learned a set of weights for the network by receiving training data comprising grayscale images of faces with signals of the respective emotion~~s~~. An image with a face was used as input in this phase and then subjected to intensity normalization.

Figure 6 depicts that the convolutional network is trained using the normalized images. The use of a validation dataset is made to select the final best set of weights from a set of training performed with examples presented in a different order, ensuring that the training performance is not dependent on the order in which the examples are presented.

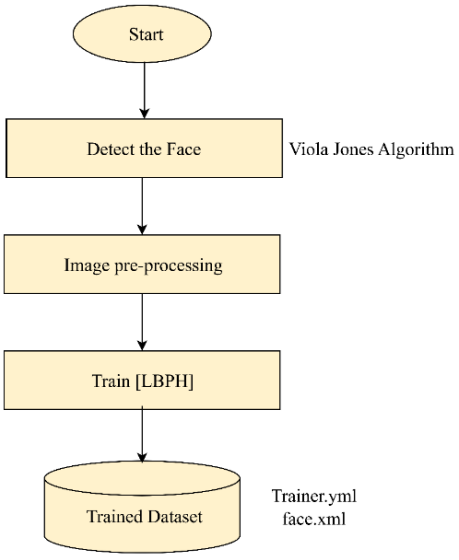


Figure 5. Flow of the Training Model

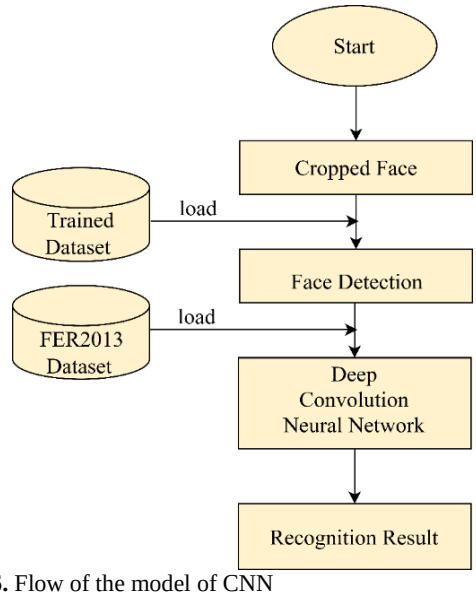


Figure 6. Flow of the model of CNN

The outcome of the training phase is a set of weights that perform well with the training data. When the grayscale image of a face is fed during the test, the system generates the projected expression using the final network weights discovered during training. The program generated a single number that represents one of the seven fundamental expressions.

Results and Discussions

The designed model consists of a machine learning architecture that is operated using a graphical user interface associated with the Windows application. The application is launched using the Visual Studio code with Anaconda Navigator (Anaconda 3). The FER2013 data from the Kaggle competition on FER2013 was the data source used for the application. The framework for detecting facial expressions is integrated using the database. The database comprises of 35,887 images total, which are split into 28,709 pictures of trains and 3589 tests. For the final test, the dataset also includes 3589 more private test images. The transcript pattern of the FER2013 data is represented in Figure 7.

for validating

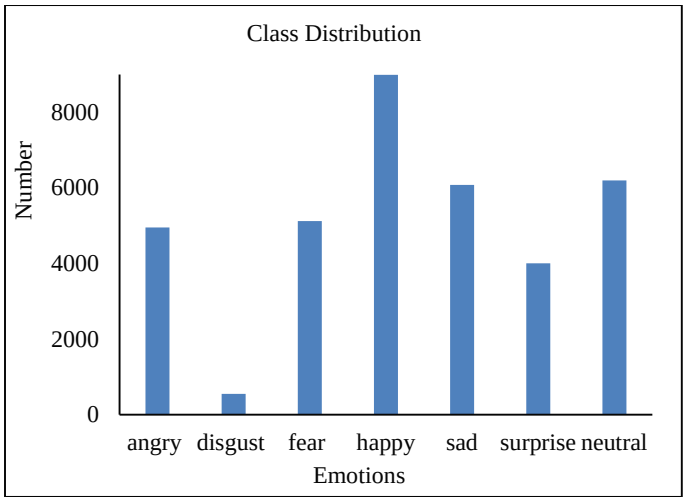


Figure 7. Shows expression disposition using FER 2013 data

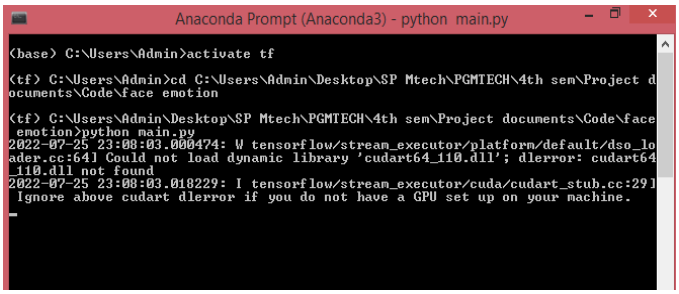


Figure 8. Running the main file

Open the command prompt from the search tab, activate tensor flow (tf), then run the main file as shown in Figure 8. All the python libraries were installed.

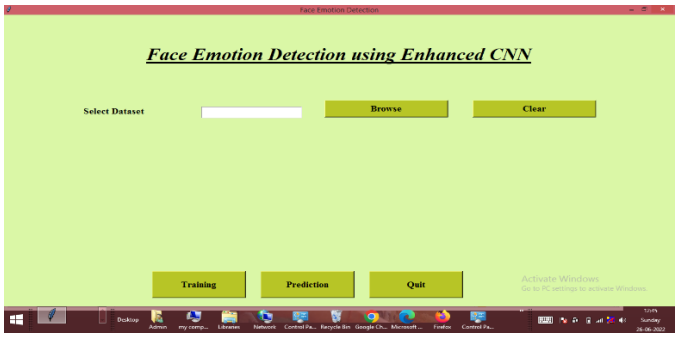


Figure 9. GUI window

Once the main file is run, the user interface window opens as shown in Figure 9. Using a graphical user window, can upload the csv dataset.

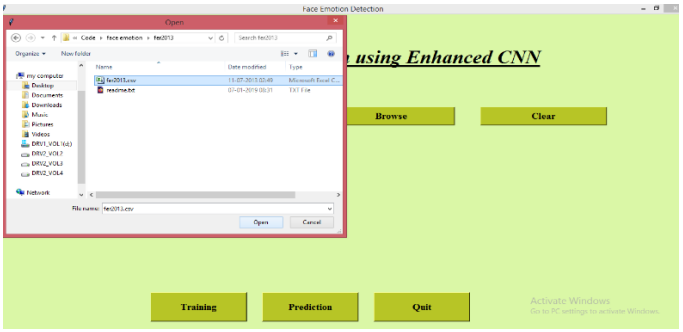


Figure 10. Upload csv dataset

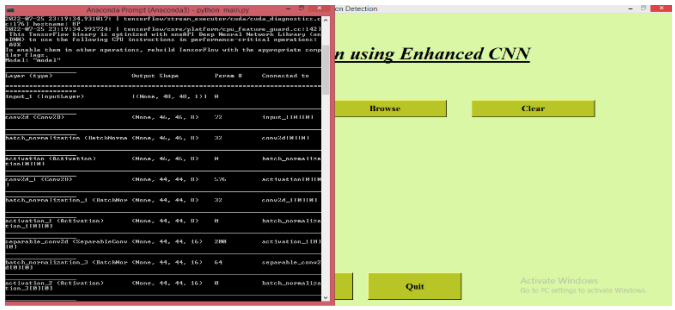


Figure 11. Train the parameters

Once the graphical user window appears, the dataset can be uploaded. To upload the dataset click on the browse button, then browse the dataset from the system and click on the open button as shown in Figure 10.

Once after clicking on the open button, need to click on the train button; the model trains the dataset as shown in Figure 11. There are a total of 110 epochs, each epoch consists of 897 steps. In this case, some parameters are trainable and some parameters are non-trainable.

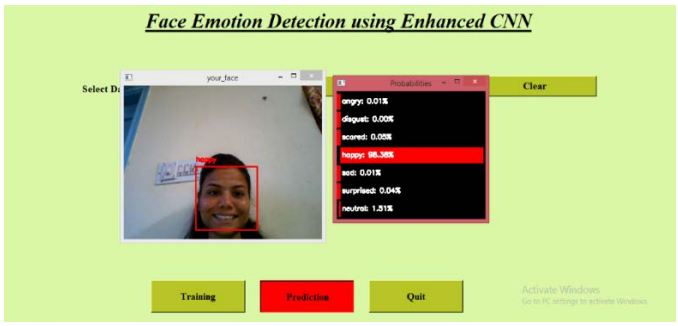


Figure 12. Real-time image processing

For real-time image processing, should run the main file, and the web camera will open. Once the camera is opened, the face must be shown to the camera and click on the predict button as shown in Figure 12. The accuracy of the happy emotion is shown in probabilities.

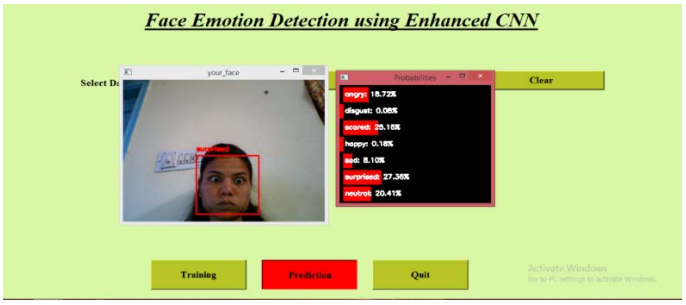


Figure 13. Surprise emotion

For real-time image processing, should run the main file, the web camera will open. Once the camera is opened, the face must be shown to the camera and click on the predict button as shown in Figure 13. The accuracy of the surprised emotion is shown in probabilities.

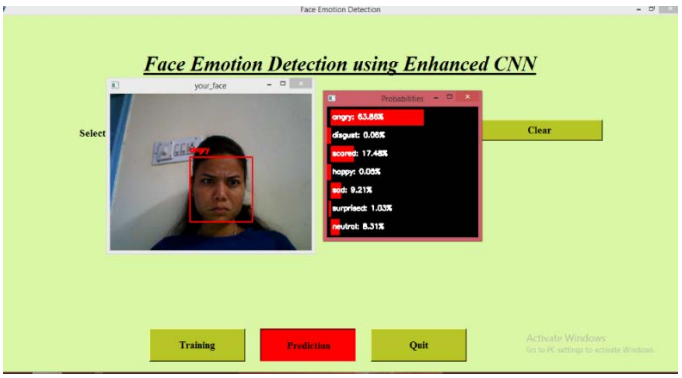


Figure 14. Angry emotion

For real-time image processing, should run the main file, the web camera will open. Once the camera is opened, the face must be shown to the camera and click on the predict button as shown in Figure 14. The accuracy of the angry emotion is shown in probabilities.

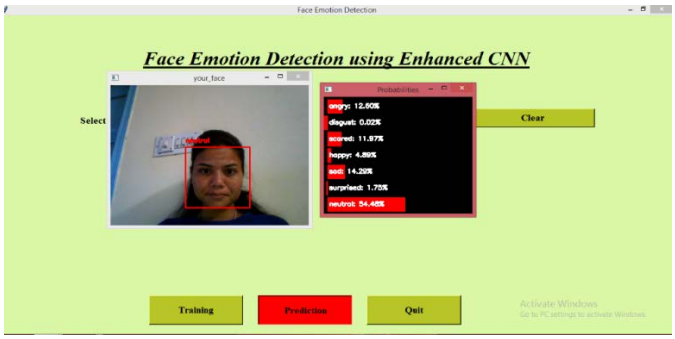


Figure 15. Neutral emotion

For real-time image processing, should run the main file, the web camera will open. Once the camera is opened, the face must be shown to the camera and click on the predict button as shown in Figure 15. The accuracy of the neutral emotion is shown in probabilities.

Comparison and Evaluation of Results

This research presents a precisely trained model for identifying driving distractions. The number of fatal accidents caused by driver mistakes or negligence has reached an all-time high in recent years. Drivers might be warned if they tend to become distracted to avoid accidents. Images of distracted drivers, such as those who are texting, changing radio stations, drinking, and/or engaging in other similar activities, are used as input to train the system. The best model for this job is chosen after this dataset has been used to train a variety of deep CNN algorithms.[25] The model proportionately detects a wide range of distracted drivers while eliminating the non-distracted drivers as distraction levels rise.

The deep CNN algorithms can detect spatial and temporal dependencies in images with a minimal amount of preprocessing. Basic preprocessing techniques are still required to ensure that the dataset excludes irrelevant data. The RGB images are changed to a grayscale format, where a two-dimensional matrix structure serves as the representation for each image. Because of background noise from the car seats, thresholding of the images is required. Thresholding ensures that only the necessary portion(s) of the image is extracted, describing distracted driving. The main image-processing methods ensure the suitability of the final image and add to the diversity of the dataset. The Deep CNN architecture offers several image categorization models and methods. Three models were used: ResNet, Xception, and VGG16. The distracted driver dataset was used to train these models separately. To measure the effectiveness of these models, a variety of evaluation metrics were also employed. To choose the optimal model, evaluation metrics were utilized. For this purpose, the ResNet model was shown to be the most effective one for the successful classification of driver's distraction. Different assessment measures including precision, recall, and F1-score are used to assess the performance of system models. An overall accuracy is determined by using below formula:

$$AC = TP + TN \div TP + TN + FP + FN$$

Recall is given by

$$Recall = TP \div TP + FN$$

Precision is given by

$$Precision = TP \div TP + FP$$

Table 1 shows the performance comparison. The VGG-16 and ResNet-50 values are listed. The supplementary tower is excluded from the benchmark edition of Inception V3. The depth of a neural network and the outcome are not dependent on each other. Moreover, best results are not always produced by deeper neural networks except for VGG16-FACE, which was trained on face recognition. This probably led to the best result among the presented networks. Table 2 shows the overall comparison of the state-of-the-art techniques with the designed model.

Table 1. Accuracy results of neural networks retrained with FER2013

Neural network name	Layers	Accuracy %
GoogLeNet[26]	22	63.21
CaffeNet[27]	8	68
VGG16[28]	16	71.4
ResNet50[28]	152	73.8
Designed Model	36	95.60

Table 2. Comparison of the different techniques

Neural network name	Accuracy (%)	Precision (%)	Recall (%)
GoogLeNet[26]	63.21	62	62
CaffeNet[27]	68	67	66.2
VGG16[28]	71.4	81.9	79.4
ResNet50[28]	73.8	83.3	80.7
Designed Model	95.60	90	93

A confusion matrix is used to calculate accuracy. Quantitative analysis of the experimental outcomes was done by considering the following performance indicators, including accuracy (AC). This (FN) also makes use of the variables True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Uncertainty Matrix: This specific Table 3 style makes it possible to assess the effectiveness of an algorithm.

Table 3. Confusion matrix of the designed algorithm

Expression	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	63.85	0.06	0.0	0.05	6.21	1.03	8.30
Disgust	0.0	55.5	0.5	0.7	10.9	20.9	0.0
Fear	12.9	0.01	65.78	0.5	0.12	17.8	0.0
Happy	0.01	0.00	0.02	98.38	0.01	0.04	1.5
Sad	25.0	11.9	0.12	0.08	85.0	0.05	7.5
Surprise	18.7	0.08	17	0.18	8.10	28.8	2.41
Neutral	12.06	0.02	0.0	4.8	14.7	1.75	54.45

Accuracy and loss graph

The authors recognize that determining model fitting involves assessing the precision of training and validation. If there is a wide difference between the two, the algorithm is overfitting. The precision of the validation should be equal to, or less than, the precision of the preparation to be a superior design. The study is shown in Figure 16, where researchers improve the loops and eventually add some more convolution and different similar loops, making the wider and broader connection. An epoch advances; the trained accuracy is only slightly bigger than the validation accuracy. It seems that there is a loss of validation which may be lesser than the lack of preparation.

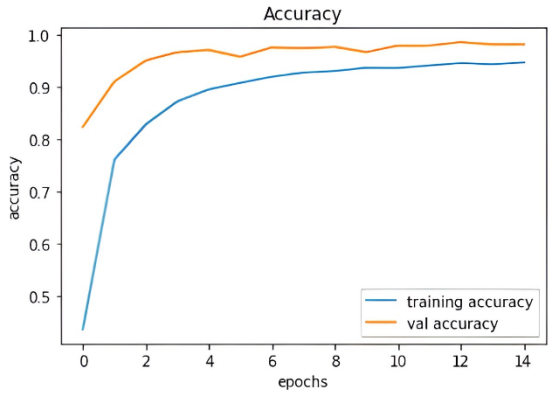


Figure 16. Graphical depiction of training & validation accuracy per epoch

Figure 16, Comparison of training failure and validation failure. This shows that as the epoch grows, the validating cost rises while the training loss reduces. Additionally, as the weights are changed, it is always expected that the validation data will decrease. Figure 17 may anticipate a lower value validation loss than training loss in this case as the epoch grows in a higher order, as the author had witnessed in the previous stage of the picture. The model therefore fits the training results well.

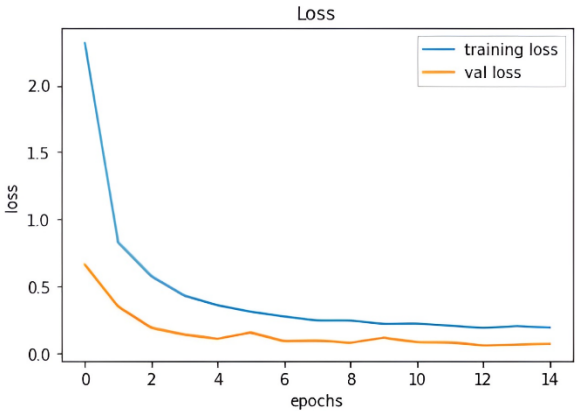


Figure 17. Graphical depiction of training and validation loss per epoch

CONCLUSIONS

Without personal contact with the topic or suspect, real-time detection of any suspicious activity is difficult. It might be hard to read someone's face in real-time. The majority of firms find it easy to understand employee behaviour and address some minor and major difficulties at an earlier phase due to small, portable gadgets. To do that, a framework that can be used in any firm to understand employee behaviour has been tested and recommended. For assessing the ability of facial expressions, facial expression representation is crucial. It may be seen as providing important features for characterizing the look, make-up, and movements of facial emotions. An increased version of the Xception design using unused networks for emotion circulation and identification implemented the Mini-Xception replica in this research. The performance of the FER-2013 file is better than the current technique. An effective system was created to recognize the seven ways in which sentiments can be conveyed, which are: disgust, fear, anger, happiness, sadness surprise, and neutral. In recent studies, deep learning models have been widely applied to propose an end-to-end method for expression identification. Even if it is a difficult task, recognizing emotion still needs a lot of development. Using Mini-Xception, accuracy for emotion expression and recognition was 95.60%, while recall rate and precision were 93% and 90%, respectively.

ABBREVIATIONS

CNN: Convolution Neural Network
FER: Facial Emotion Recognition
FE: Feature Extraction
SA: Sentiment analysis
KNN: K-Nearest Neighbour
DT: Decision Tree (DT)
RPN: Region Proposal Network
GUI: Graphical User Interface
AC: Accuracy (AC)
TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative

AUTHOR CONTRIBUTIONS

Sowmya B J and Meeradevi: Implemented and wrote the few text of the manuscript; Sini Anna Alex and Anita Kanavalli: Reviewed the article; Supreeth S and Shruthi G: were responsible for analyzing and building the algorithm and framework for the study.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGMENTS

The authors acknowledge the support from MSRIT, REVA University for the facilities provided to carry out the research.

CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

REFERENCES

1. P. Angusamy, S. Inba, K.S. Pavithra, M. Ameer Shathali, M. Athiparasakthi. Human Emotion Detection using Machine Learning Techniques. *SSRN Electron. J.* **2020**, 3591060.

2. S.A. Fatima, A. Kumar, S.S. Raoof. Real Time Emotion Detection of Humans Using Mini-Xception Algorithm. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, 1042 (1), 012027.
3. J. Bao, M. Ye. Head pose estimation based on robust convolutional neural network. *Cybern. Inf. Technol.* **2016**, 16 (Specialissue6), 133–145.
4. B.P. Babu, S.J. Narayanan. One-vs-All Convolutional Neural Networks for Synthetic Aperture Radar Target Recognition. *Cybern. Inf. Technol.* **2022**, 22 (3), 179–197.
5. A. Lazarov, C. Minchev. ISAR image recognition algorithm and neural network implementation. *Cybern. Inf. Technol.* **2017**, 17 (4), 183–189.
6. M.A. Ansari, D.K. Singh. ESAR, An Expert Shoplifting Activity Recognition System. *Cybern. Inf. Technol.* **2022**, 22 (1), 190–200.
7. H. V. Ramachandra, P. Chavan, S. Supreeth, et al. Secured Wireless Network Based on a Novel Dual Integrated Neural Network Architecture. *J. Electr. Comput. Eng.* **2023**, 2023, 1–11.
8. M. Pujar, M.R. Mundada, B.J. Sowmya, S. Supreeth, G. Shruthi. An Efficient Framework for Web Content Mining Systems Using Improved CD-PAM Clustering and the A-CNN Technique. *SN Comput. Sci.* **2023**, 4 (5), 692.
9. P.A. Riyantoko, Sugiarto, K.M. Hindrayani. Facial Emotion Detection Using Haar-Cascade Classifier and Convolutional Neural Networks. *J. Phys. Conf. Ser.* **2021**, 1844 (1).
10. D.Y. Liliana. Emotion recognition from facial expression using deep convolutional neural network. *J. Phys. Conf. Ser.* **2019**, 1193 (1).
11. I. Talegaonkar, K. Joshi, S. Valunj, R. Kohok, A. Kulkarni. Real Time Facial Expression Recognition using Deep Learning. *SSRN Electron. J.* **2019**, 1, 1–5.
12. D. Mehta, M.F.H. Siddiqui, A.Y. Javaid. Recognition of emotion intensities using machine learning algorithms: A comparative study. *Sensors (Switzerland)* **2019**, 19 (8), 1–24.
13. A.I. Siam, N.F. Soliman, A.D. Algarni, F.E. Abd El-Samie, A. Sedik. Deploying Machine Learning Techniques for Human Emotion Detection. *Comput. Intell. Neurosci.* **2022**, 2022.
14. A. Jaiswal, A. Krishnama Raju, S. Deb. Facial Emotion Detection Using Deep Learning. In *2020 International Conference for Emerging Technology (INCET)*; **2020**; pp 1–5.
15. P. Bagane, S. Vishal, R. Raj, T. Ganorkar. Facial Emotion Detection using Convolutional Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2022**, 13 (11), 168–173.
16. S.K. Khare, V. Bajaj. Time–Frequency Representation and Convolutional Neural Network-Based Emotion Recognition. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, 32 (7), 2901–2909.
17. P. Chakriswaran, D.R. Vincent, K. Srinivasan, et al. Emotion AI-Driven Sentiment Analysis: A Survey, Future Research Directions, and Open Issues. *Appl. Sci.* **2019**, 9 (24), 5462.
18. M. Healy, R. Donovan, P. Walsh, H. Zheng. A Machine Learning Emotion Detection Platform to Support Affective Well Being. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; **2018**; pp 2694–2700.
19. R. Alhalaseh, S. Alasasfeh. Machine-learning-based emotion recognition system using EEG signals. *Computers* **2020**, 9 (4), 1–15.
20. N.A.S. Badrulhisham, N.N.A. Mangshor. Emotion Recognition Using Convolutional Neural Network (CNN). *J. Phys. Conf. Ser.* **2021**, 1962 (1).
21. T. Gilligan, B. Akis. Emotion AI, Real-Time Emotion Detection using CNN. **2016**.
22. N. Mehendale. Facial emotion recognition using convolutional neural networks (FERC). *SN Appl. Sci.* **2020**, 2 (3), 1–8.
23. B. Hdioud, M. El, H. Tirari. Facial expression recognition of masked faces using deep learning. **2023**, 12 (2), 11591.
24. W. Mellouk, W. Handouzi. Facial emotion recognition using deep learning: Review and insights. *Procedia Comput. Sci.* **2020**, 175, 689–694.
25. B. Kabra, C. Nagar. Attention-Emotion-Embedding BiLSTM-GRU network based sentiment analysis. *J. Integr. Sci. Technol.* **2023**, 11 (4), 563.
26. P. Giannopoulos, I. Perikos, I. Hatzilygeroudis. Deep learning approaches for facial emotion recognition: A case study on FER-2013. *Smart Innov. Syst. Technol.* **2018**, 85, 1–16.
27. W. Mohamed Yassin, M. Faizal Abdollah, Z. Muslim, R. Ahmad, A. Ismail. An Emotion and Gender Detection Using Hybridized Convolutional 2D and Batch Norm Residual Network Learning. *ACM Int. Conf. Proceeding Ser.* **2021**, 79–84.
28. X. Zhao, X. Shi, S. Zhang. Facial Expression Recognition via Deep Learning. *IETE Tech. Rev.* **2015**, 32 (5), 347–355.