Signal, Image and Video Processing (2024) 18:3643–3660
https://doi.org/10.1007/s11760-024-03029-z

ORIGINAL PAPER

Emotion recognition of the driver based on KLT algorithm and ShuffleNet V2

Faiyaz Ahmad[1] · U. Hariharan[2] · N. Muthukumaran[3] · Aleem Ali[4] · Shivi Sharma[5]

Received: 29 December 2023 / Revised: 12 January 2024 / Accepted: 15 January 2024 / Published online: 22 February 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract
Emotional monitoring was essential in the development of sophisticated automobiles with advanced driver assistance systems (ADAS) to ensure safety and monitor potential collision trends for evaluating the driver's mental state. Factors affecting driver emotional identification include posture changes, illumination, and occlusions. Existing emotion recognition using CNN ResNet101 has low sensitivity, high false-positive rate, and error. To overcome these challenges, this paper proposes Driver Emotion Recognition (DER) using ShuffleNet V2 to effectively recognize emotions and determine the driver's mental condition. Initially, the facial images from different peoples are collected as a dataset and pre-processed using image resizing, Gaussian filter, median filter, histogram equalization and wiener filter for removing noise and enhancing the image quality. For segmentation and feature extraction of face images from a variety of datasets such as CK_Plus, FER_2013, TFEID, KMU_FED, and KDEF, the Region of Interest (ROI) and Kanade-Lucas-Tomasi (KLT) algorithm are used, which segments the face images based on region. Then, the ShuffleNet V2 classification is used to categorize emotions into six unique expressions such as happy, surprise, sad, fear, anger, disgust, and neutral. The performance of the proposed model is assessed by comparing it to that existing models. The proposed approach achieved an accuracy rate of 0.98% in CK_Plus, 0.97% in FER_2013, 0.97% in TFEID, 0.99% in KMU_FED, and 0.99% in KDEF. In comparison to other existing techniques, the proposed technique performs better. In order to determine the different emotions, the created model is the best choice.

Keywords ShuffleNet V2 · Human emotion recognition · Human wellbeing · Multi-objective · Kanade–Lucas–Tomasi (KLT)

1 Introduction

Emotions of the driver have a significant influence on comfort and safety when driving [1]. One of the main factors reducing driving safety in the 20–50 million non-fatal injuries and 1.24 million fatal road traffic accidents that occur globally is drivers' incapacity to regulate their emotions. The swift advancement of intelligent automobiles necessitates the amalgamation of driver-automation communication and cooperation to augment driving comfort wherever driver sentiment is a crucial condition [2]. Understanding the feelings of the driver is essential for enhancing comfort and safety when operating intelligent cars.

Driver Emotion Recognition (DER) technology is to evaluate the active state of a driver using their facial expression. The DER system is about improving the Human Machine Interface (HMI) in vehicles. Numerous applications, including social security, mental health monitoring, safe driving, and health care, have made use of emotion identification [3]. By detecting and raising a driver's awareness of their emotions, the DER system may be utilized to manage their emotional state. So, the effective recognition of driver emotion is very much important to develop a better DER system.

Several techniques are used in the literature to evaluate a driver's emotional condition [4]. A soft computing tool is

✉ Faiyaz Ahmad
ahmad.faiyaz@gmail.com

1 Department of Computer Engineering, Jamia Millia Islamia, New Delhi 110025, India

2 Department of CSE, Apex Institute of Technology, Chandigarh University, Ajitgarh, Punjab 140413, India

3 Centre for Computational Imaging and Machine Vision, Department of ECE, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu 641202, India

4 Department of CSE, UIE Chandigarh University, Mohali, Punjab 140310, India

5 Department of CSE, Jain University, Bengaluru, Karnataka 560069, India

used to identify the mood and facial motion for a driving assistant technology that uses Fuzzy Rule-Based Systems (FBS). The analysis of facial gesture variations is used to define the fuzzy rules [5]. DER system is also designed by utilizing Local Binary Pattern (LBP) for texture base assessment and face recognition. LBP is efficient and very simple. Oriented Fast and Rotated BRIEF (ORB) is an additional DER system approach for data analysis. With the Support Vector Machine (SVM) acting as the classifier, ORB serves as a quick and reliable feature detector that utilizes Binary Robust Independent Elementary Features (BRIEF) as a face feature detector [6]. ECG (Electrocardiogram) is also one of the very efficient tools to detect and recognize a person's emotion. DER system process is also based on EEG data in a method of Self-Assessment Manikin (SAM) [7]. When a person is exposed to a certain stimulus, the density and frequency of the stimulus are utilized to determine the subject's emotional state [9]. We can recognize the driver's emotion by these signals.

A camera-based DER system, a camera is installed within the vehicle. It monitors the driver's face continuously. This camera-based DER system uses a face detection algorithm to detect the driver's face in the natural setting [8]. Under realistic driving situations, despite varying lighting conditions, occlusions, and other passengers in the vehicle, accurate identification of the driver's face is required. The major problem faced by the camera-based driver monitoring system [9]. DER system helps to alert the driver by raising the alarm. So it is necessary that the future generation vehicles should have the extra safety feature to alert the driver about the driver's emotional state not only for road safety but also human-wellbeing.

The proposed method's primary contributions are listed below:

- Driver's emotion recognition is detected using a deep learning-based ShuffleNet V2 and KLT Algorithm-based feature extraction method.
- Face image Dataset are collected and pre-processed using image resizing, Gaussian filter, median filter, histogram equalization and wiener filter for filtering the noise and enhancing contrast of the image.
- Pre-processed images are segmented using Region of Interest (ROI) for segmenting the unwanted portions from the facial images to reduce the complexity of the model.
- Segmented portions of the face images are extracted for features using the Kanade-Lucas-Tomasi (KLT) algorithm for training the model with the facial distinct features.
- ShuffleNet V2 classifier is used to categorize the emotions of the drivers into six distinct expressions such as happy, surprise, sad, fear, anger, disgust, and neutral.

The remaining portion of the study is composed of the literature review pertaining to the DER system is presented in Chapter 2. The proposed design and technique for the proposed portion are included in Chapter 3. Portion 4 contains the proposed method's outcome and discussion section. The final section is included in portion 5.

2 Literature review

This section summarizes research in the field of recognition of emotion. Based on their previous and existing activities, people often show emotions such as happy, neutrality, sadness, disgust, surprise, fear, and anger. Below is a study and review of a few of the current detection methods.

Du et al. [10] introduced the Convolution Bidirectional Long Short-term Memory Neural Network (CBLNN), a novel deep learning architecture for identifying the driver's emotional state. CBLNN was used to recognize emotions easily and accurately in real-time. The process of CBLNN was based on CNN to evaluate the face shape. CBLNN had better accuracy in identifying anger, sadness, happiness and neutrality. However, CBLNN accuracy in detecting fear is significantly lower.

Wang et al. [11] established a combination of several electrocardiogram (ECG) features to identify the driver's emotional state. Three components make up the ECG signals such as waveform, nonlinear characteristics, and time–frequency interval. To determine if a motorist was relaxed or nervous, an emotion detection technique was used while driving. ECG had a 91.34% accuracy rate for identifying the driver's calm and a 92.89% accuracy rate for identifying tension. But, the processing of ECG data requires the combination of multiple evidence fusion in nonlinear analysis.

Xia et al. [12] recommended a unique method for recognizing cross-dataset transfer driver expressions in shared projection subspace (GD-LS-SS) by utilizing global discriminative and local structural knowledge. GD-LS-SS makes use of the data's local geometrical structure by using graph topology knowledge. The kernel-based GD-LA-SS is designed by using the advantages of the kernel trick to investigate the kernel projection in order to increase identification accuracy and handle the nonlinear cross-dataset transfer. However, the GD-LA-SS had several challenges like how to eliminate unfavorable transfer in the current method, how to choose the most important aspects presented in the face pictures, and how to transfer the key facial images attachment to SD.

Hu et al. [13] presented a novel deep learning architecture that combines a 3D conditional generative adversarial network with a two-level attention bidirectional long short-term memory network (3DcGAN-TLABiLSTM). On public NTHU-DDD dataset, test the 3DcGAN-TLABiLSTM framework. Because of the significant intra-class variations

in head position, face expression, and lighting conditions, it is still difficult to diagnose fatigue.

Jeong et al. [14] designed a lightweight multilayer random forest (LMRF) model, a deep model formed by non-neural network-based layer-by-layer random forests. Even with fewer hyper-parameters, LMRF achieves performance comparable to DNN, and it runs quicker on a CPU. But LMRF had a performance degradation problem when more than three numbers of layers were presented.

Shojaeilangari et al. [15] presented an Extreme Space Learning (ESL) approach to identify the human face's inherent emotions. ESL students were able to learn a vocabulary and a nonlinear classification model at the same time. When dealing with noisy signals and faulty data acquired in natural situations, to achieve accurate classification, ELM combines the resemblance capability of a sparse representation along with the discriminative strength of the Extreme Learning Machine (ELM). A higher computational expenses for extracting and categorizing features, as well as the need to optimize a large number of parameters, are also disadvantages.

Kim et al. [16] created a line-segment feature analysis (LFA), convolutional recurrent neural network (CRNN) model for face sentiment analysis, and a streaming based on images PingPong256 (PP2) method. Real-time pictures gathered by image devices were secured by encryption and decryption using the PP2 algorithm. The LFA-CRNN paradigm, on the other hand, is unable to leverage miniaturization technologies like mobile edge computing systems.

Mohan et al. [17] suggested a deep convolution neural network (DCNN). The first branch looks at geometric elements, including lines, edges, and curves, whereas the second branch looks at holistic traits. DCNN methodology outperforms all the cutting edge techniques across all datasets. A GF-based edge descriptor is utilized to get a small amount of local features in order to use the DCNN model.

Cui et al. [18] introduced a unique multi-task neural network called Multi-EmoNet, which can repair noisy images and classify human face emotions under various settings. When compared to baseline networks, Multi-EmoNet obtains a significantly greater degree of classification results on images with different levels of light. A generic design that can be used for any noisy picture classification issue is the multi-task network.

Madupu et al. [19] developed a Convolutional Neural Network (CNN) based automated face emotion categorization method using the Speeded Up Robust Features (SURF) feature. This approach had an accuracy rate of 91%. However, the dataset sample size for this approach was just 200.

Based on the above-revealed article, several significant challenges has been raised for face emotion detection. The accuracy of CBLNN in detecting fear is significantly lower [10]. The processing of ECG data necessitates the integration of multiple evidence fusion in nonlinear analysis [11]. The GD-LA-SS faced challenges in avoiding unfavorable transfer, selecting crucial face picture aspects, and transferring key facial image attachments to SD [12]. Diagnosis of weariness is still difficult since there are large intraclass differences in head posture, expression, and illumination [13]. LMRF experienced performance degradation when presented with more than three layers [14]. The disadvantages of this method include higher computational expenses for feature extraction and categorization, as well as the need to optimize numerous parameters [15]. The dataset sample size for this approach was limited to 200 [19].

3 Proposed methodology

Most of the road accidents are caused due to the unpleasant emotions of the driver. To overcome such disasters, the Emotion Recognition of the Driver based on KLT Algorithm and ShuffleNet V2 has been designed. In the proposed work Driver Emotion Recognition (DER) system is developed to avoid this type of accident. Initially images from the datasets such as CK_plus, FER_2013, TFEID, KMUFED and KDEF are considered as input. These dataset consist of several regions of human face with different emotions based on both the genders. At first, the image dataset are preprocessed using the Histogram equalization, Wiener filter, 2D Gaussian filter, and 2D median filter for image enhancement and noise reduction to attain better performance on recognition. Subsequently, the preprocessed images are segmented based on the ROI (Region of Interest) segmentation algorithm, where a rectangular-shaped ROI (Region of Interest) is placed over the facial images to remove the unwanted portions of the facial image based on the interested regions. These segmented facial regions are extracted for features by using the KLT algorithm to attain the scattered feature points which have enough texture on facial images. Finally, the extracted features form the KLT algorithm is given to shuffle net V2 for training and recognizing the driver emotions. This ShuffleNet V2 classifier identifies six distinct expressions such as happy, surprise, sad, fear, anger, disgust, and neutral. The below Fig. 1 explains the entire process of the proposed DER system.

3.1 Preprocessing

Image preprocessing has become an important phase in any face image processing for scientific purposes. Preprocessing goal is to improve the picture quality and increase its characteristics for future preparation. The general preprocessing includes Noise reduction, Color Normalization, Histogram Equalization and Edge detection. In this proposed method,

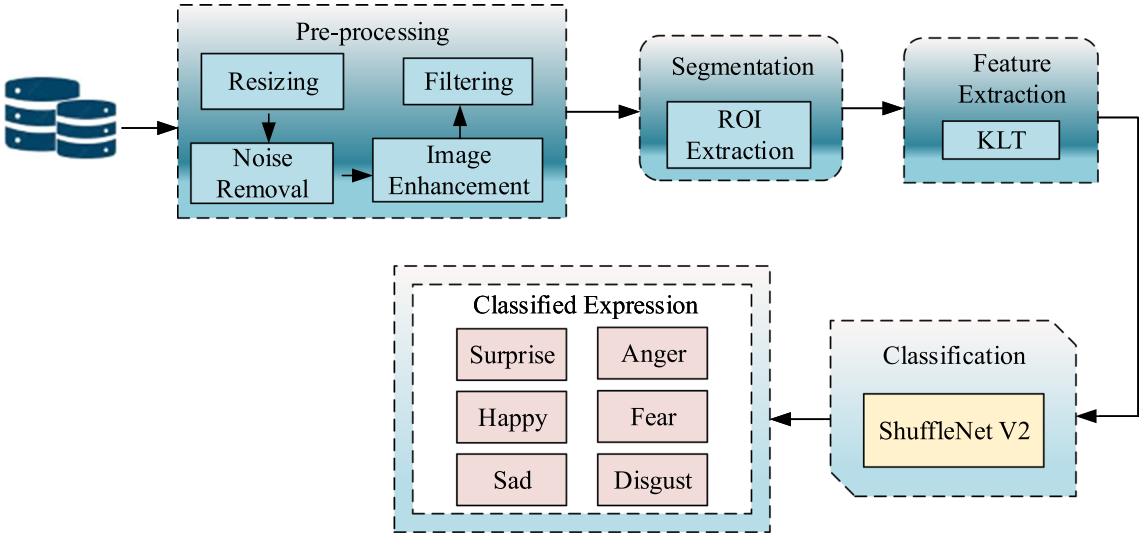


Fig. 1 Structure of the proposed approach

image preprocessing can resize the image, remove the noise in the image and do the enhancement the image.

3.1.1 Step 1: Resizing

Images may be resized without developing any sections removed by using this technique. To improve or reduce the total number of pixels in the image, image resizing is important. Pixel data is altered when an image is resized. Avoid resampling an image that has been resized. Make simple sizing adjustments; do not alter the image's data content. A picture of 700 by 700 pixels, for instance, gets shrunk to 256 by 256 pixels. The scaled texture of the original image is seen below in Fig. 2.

3.1.2 Step 2: Noise removal

Technique of eliminating or decreasing noise from a picture is known as noise reduction. By smoothing the entire image and leaving the region around the contrast limits, it can reduce the appearance of noise. A 2D Median filter and 2D Gaussian filter are the two techniques used in this proposed system for the noise removal process. These techniques can remove most of the noise present in the image.

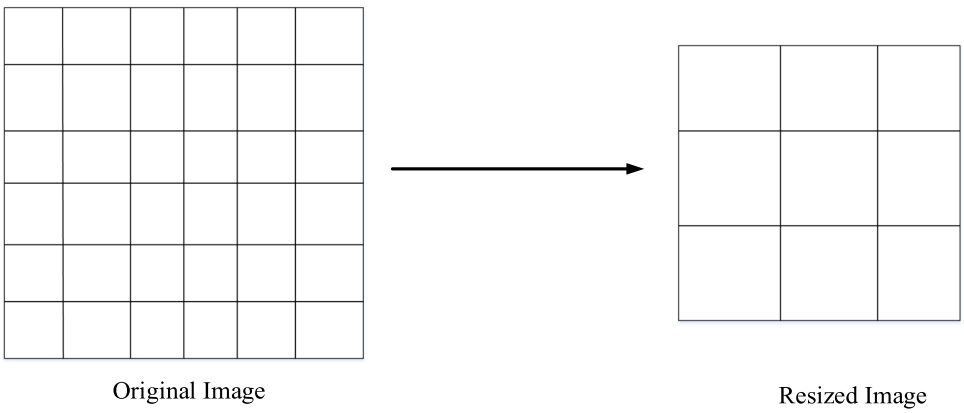
2D Gaussian Filter: Gaussian Filter serves as a 2D convolutional filter to remove noise and smoothen the image. Its impulse response is a Gaussian filter. This proposed method uses a 2D Gaussian filter to remove noise from the given set of images. Image noise is mostly smoothed by using the Gaussian filter [20]. The following Eq. (1) represents the Gaussian filter,

$$\widehat{f}(x,\ y) = f(x,\ y) * g(x,\ y) \tag{1}$$

where, the 2D input image is denoted by f, the 2D output image is denoted by $\widehat{f}$.

2D Median Filter: Visual noise is removed using a nonlinear digital filtering method known as the median filter. Because it may sometimes preserve edges while eliminating

Fig. 2 Resized Diagram



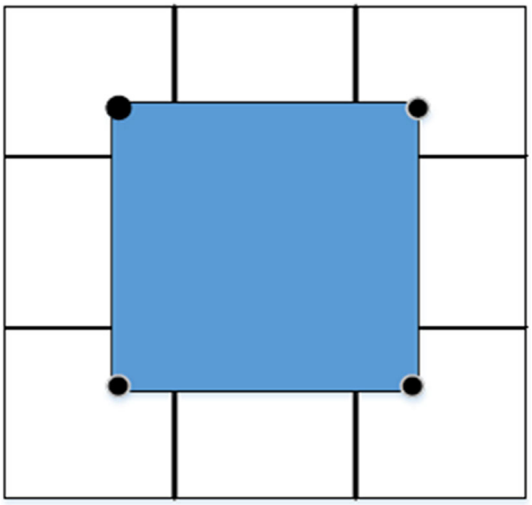


Fig. 3 Image Segmentation

noise, it is often used in digital image processing. The proposed method employs a 2D median filter as an additional method to remove noise from the picture collection. It works by gradually substituting the median of neighboring pixels for each value in the image, pixel by pixel. Because the neighbor's pattern moves pixel by pixel throughout the image, it is referred to as a window. The pixel image's median value may be found using Eq. (2) [21].

$$MF(i,\ j) = \text{Median}(x1,\ x2,\ \ldots.,\ x8,\ x9) \tag{2}$$

where, MF (i, j) is the median values of neighbor pixels and (i, j) represents the pixel coordinates.

3.1.3 Step 3: Image enhancement

Image enhancement is a method that allows the user to emphasize certain aspects of an image while reducing or eliminating any unwanted information. For example, removing noise and adjusting levels to emphasize the characteristics of an image. In this proposed approach, image enhancement is utilized to enhance the given image dataset. Histogram equalization and Wiener Filter are the two techniques used for image enhancement in this proposed method.

Histogram Equalization; Histogram Equalization is a computer image processing method used to increase the image's brightness and smooth the image. This is achieved by increasing the image's range. This technique frequently raises all images' overall contrast when the user data is represented by near contrast values. This raises the contrast level in areas that have low local contrast. So, this Histogram Equalization technique is used in this proposed method to improve image quality. Equation (3) represents the resultant of the

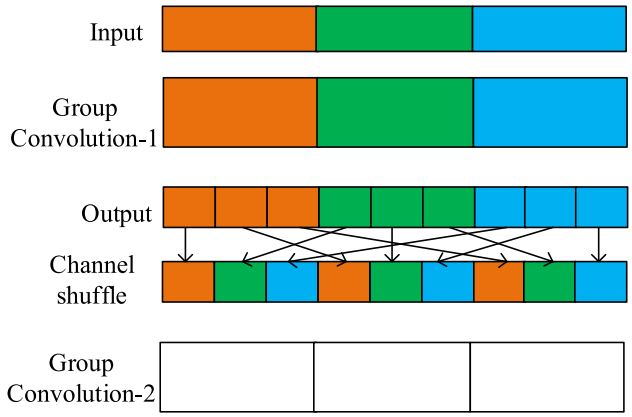


Fig. 4 Group convolution with channel shuffling

enhanced image using histogram equalization [22].

$$EI(i,\ j) = T_f(x) = \left\{T_f(x(i,\ j))/\forall x(i,\ j) \in X\right\} \tag{3}$$

where, i and j are the coordinates of the image, T_f(x) is the transformation function and EI (i, j) represents the enhanced image.

Wiener Filtering: Wiener filtering technique is used for image restoring. Gaussian filter used for Noise removal in this proposed method may lightly blur the image. So the Wiener Filter can balance the image quality. Using this approach, if noise is present in this system, it is assumed to be adaptive white Gaussian noise. Wiener filtering actually requires deep knowledge about the original picture and power spectra of the noise. The Wiener Filter estimates the actual image linearly. Wiener filter suppresses the Gaussian Noise existing in the picture [23] according to the following Eq. (4),

$$J(i,\ j) = m + \frac{i^2 - \sigma^2}{\sigma^2}(I(i,\ j) - m) \tag{4}$$

where, i, j stand for the input image's row and column, J stands for the output image's intensity, I stands for the input image's intensity, and σ^2 represents the input noise variance.

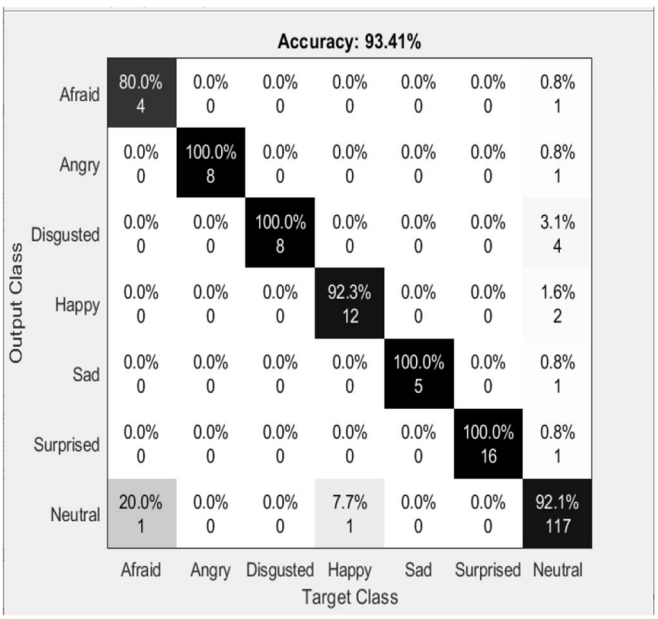
3.2 ROI extraction in segmentation

Digital image is divided into many fragments using the segmentation procedure. By using this technique, the depiction of an image may be made more meaningful or simpler. This segmentation procedure yields a collection of images that together cover the full image. In this proposed method, rectangular shape ROI is extracted during the segmentation process.

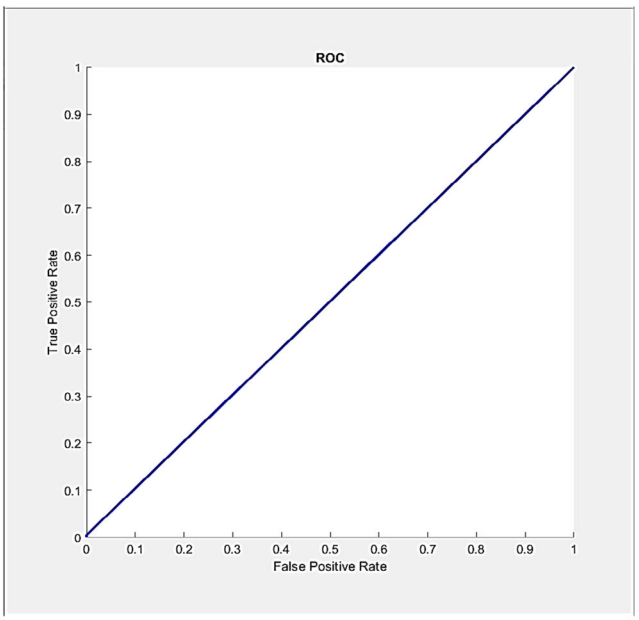
An image's designated area intended subsequent processing or analysis is called a Region of Interest (ROI). ROI will be extracted using this proposed approach in a rectangular form. A rectangular extraction of the specified region with

Table 1 CK_Plus dataset's pre-processing and feature extraction.

Input image	Resized image	Gaussian filter	Median filter	Histogram equalization	Wiener filter	KLT feature



(a)



(b)

Fig. 5 Confusion matrix plot and ROC plot for CK_Plus dataset

Table 2 Pre-processing and Feature extraction in FER_2013 dataset.

Input image	Resized image	Gaussian filter	Median filter	Histogram equalization	Wiener filter	KLT feature

improved precision is made. The primary region required for subsequent procedures is this extracted region. It is mostly employed in the image portioning process during segmentation. The segmented image is shown in the following Fig. 3. Rectangular ROI extraction is used to extract the necessary targeted region from the image in a rectangular form. Figure 3 shows the extraction of the rectangular shape ROI during the segmentation process.

3.3 Feature extraction using KLT algorithm

A step in the dimensionality reduction method that breaks down enormous amounts of raw data into smaller pieces is called feature extraction. The image's attributes are described or the feature group is mentioned in this feature extraction method. These characteristics accurately and distinctively define the actual data collecting process, and they are simple to implement. In this proposed method Kanade–Lucas—Tomasi (KLT) face feature is utilized to obtain the features of the face image [24].

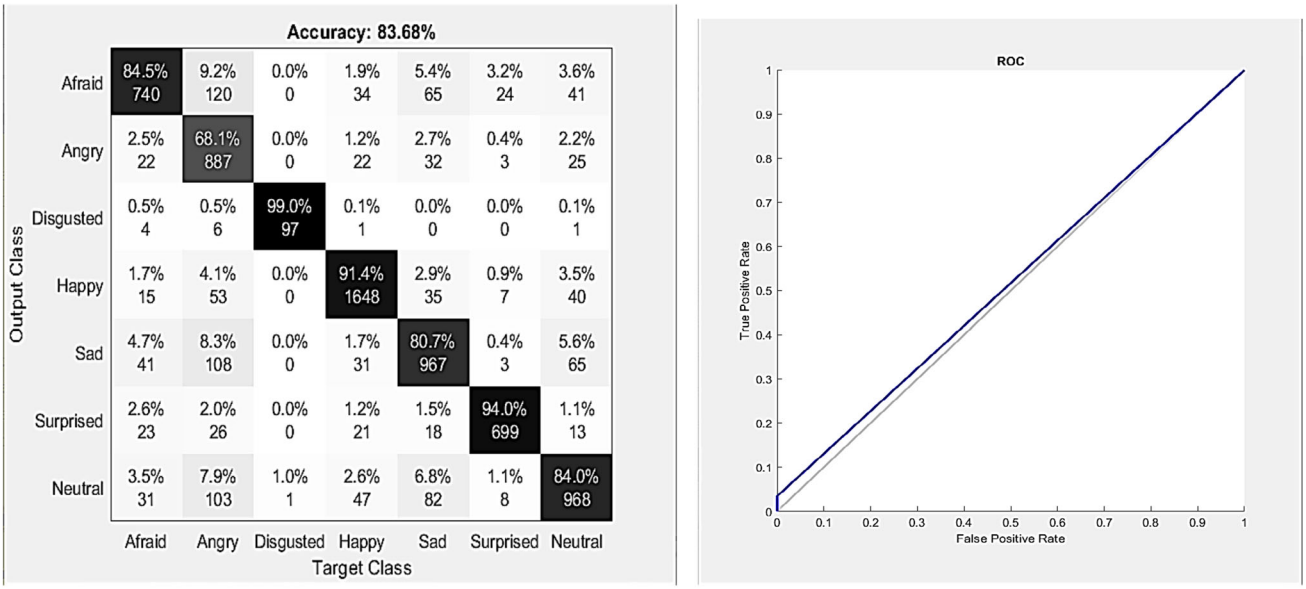


Fig. 6 Confusion matrix plot and ROC plot for FER_2013 dataset

KLT algorithm is used for tracking human faces or features from a captured frame. First, determine the displacement of tracked points that have moved from one frame to the next. The movement of a human face may be easily computed using this displacement determination, and then the feature points of a human face can be tracked. The aim of the KLT is to consider the intensity information of the pixels. Equation (5) represents the KLT algorithm calculation.

Assuming the image was tracked at time t and the next image at time t + T.

$$I(x,\ y,\ t+T) = I(x-X,\ y-Y,\ t) \tag{5}$$

where, x and y are the variables of the first image. Using this, face landmarks will be spotted in this proposed method.

3.4 ShuffleNet v2

An essential first stage in the tiredness recognition process is facial feature point detection. In this work, ShuffleNet V2K16 classifier to categorize various face expression kinds. ShuffleNet v2 represents an enhanced version of ShuffleNet v1 that utilizes channel shuffling with four objectives for design. It works more complexly and accurately than ShuffleNet v1 and MobileNet v2. By working on relevant channel groups, grouped convolution lowers processing costs; however, it lessens the expressive potential of the output characteristics since it restricts the information flow across channel groups. By guaranteeing that feature maps share information without requiring more processing, it is ensured that input and output are associated through the channel shuffling approach. Using channel shuffling in group convolution is seen in Fig. 4 [25].

ShuffleNet V2 was a network design that splits the input feature map into two branches, each with half as many channels, in order to minimize MAC usage. The left branch stays unaltered while the right branch conducts three convolutions with a step size of one. Both regular and deep separable convolutions are used to carry out the convolutions. Following convolution, information is transferred between groups via channel shuffle, features are integrated, channel numbers are added, and the two branches are concatenated. Every channel is combined into one. The concatenation process expands the network, improves feature extraction, and doubles the number of channels without increasing the FLOPs. Information sharing between many channels is made possible by mixing and washing the same channel. The network's processing burden is lessened by this method.

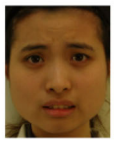
Channel separation and depthwise convolution are the two building blocks that ShuffleNet v2 uses to divide input features into two halves.

In order to reduce the cost of memory access, the convolutional layer should retain an equal amount of inputs and outputs for each of the feature channels. One x one convolution, for instance, has an output channel called c_o, FLOPs of B, and an input feature size of $c_i x h x w$.

$$B = hwc_ic_o \tag{6}$$

$$MAC = hw(c_i + c_o) + c_ic_o \tag{7}$$

Table 3 TFEID dataset's pre-processing and extraction of features.

Input image	Resized image	Gaussian filter	Median filter	Histogram equalization	Wiener filter	KLT feature
						

The mean inequality states that when B is kept fixed, the following holds:

$$MAC \geq 2\sqrt{hwB} + \frac{B}{hw} \tag{8}$$

Inequality sign holds when $c_i = c_o$, indicating that the maximum MAC consumption is achieved.

Since there are g groups in group convolution, it is advisable to utilize less of it to prevent memory access costs from rising.

$$B = \frac{hwc_i c_o}{g} \tag{9}$$

$$MAC = hw(c_i + c_o) + \frac{c_i c_o}{g} \tag{10}$$

$$= hwc_i + \frac{Bg}{c_i} + \frac{B}{hw} \tag{11}$$

When group count (g) and floating-point operations (B) are increased, MAC rises accordingly.

Reducing network branching through inception design has an impact on the computer's capacity for parallel computation. Speed is impacted by the numerous multi-branch structures in the network architecture. Significant MAC usage is achieved despite short FLOPs by reducing tensor operations such as ReLU activation function and feature summation operations.

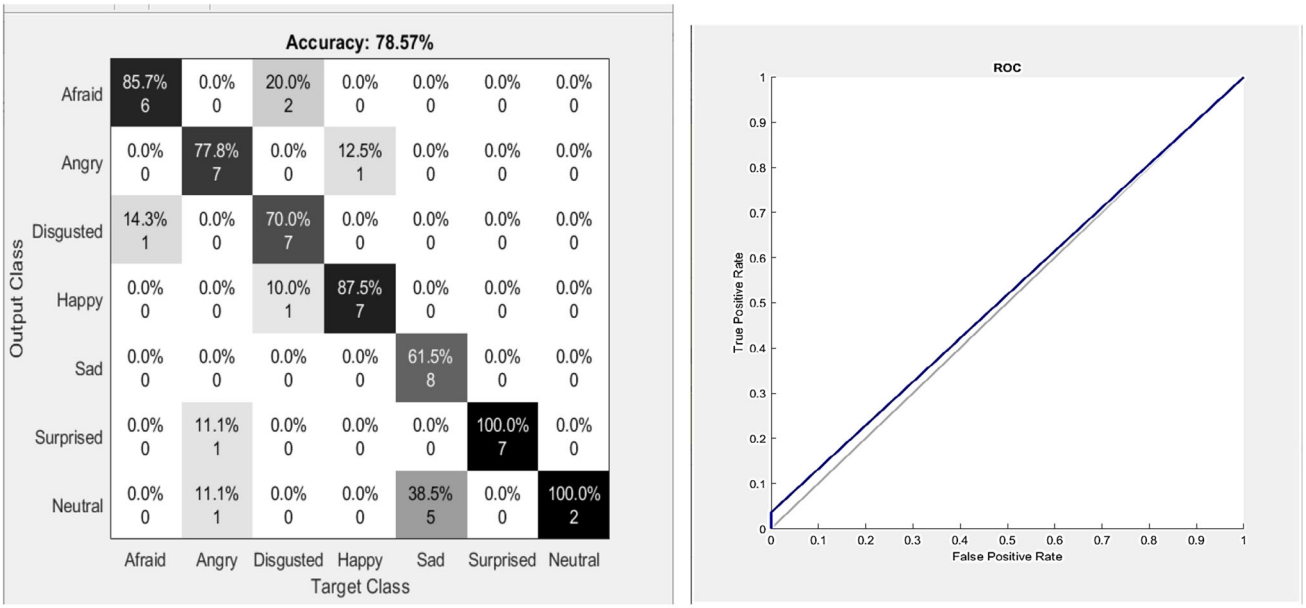
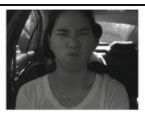
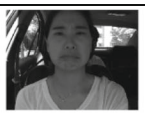


Fig. 7 Confusion matrix plot and ROC plot for TFEID dataset

Table 4 Pre-processing and Feature extraction in KMU_FED dataset.

Input image	Resized image	Gaussian filter	Median filter	Histogram equalization	Wiener filter	KLT feature
						
						

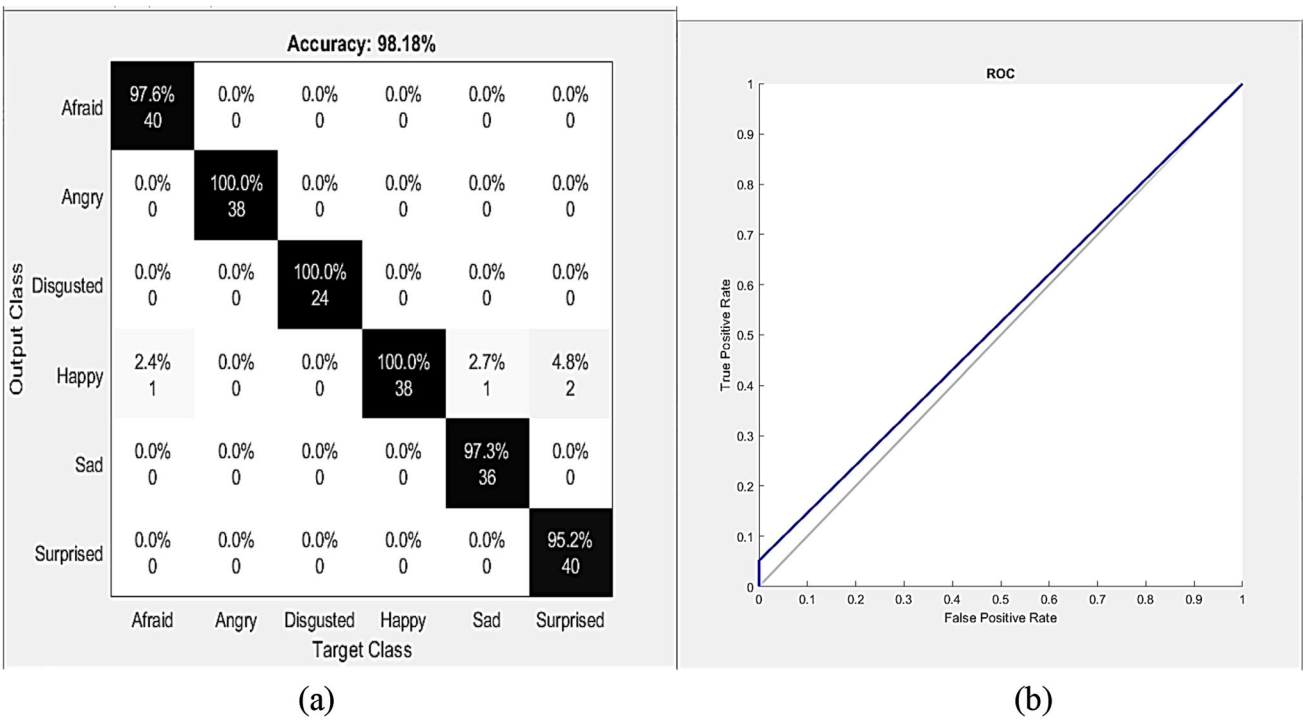


Fig. 8 Confusion matrix plot and ROC plot for KMU_FED dataset

4 Result and discussion

All vehicle users need to be aware of emotion recognition because road safety and human well-being are mainly dependent on the current emotional state of the drivers. The proposed DER system discovers the emotional mind of the driver, and if the driver is in unpleasant emotion, it makes an alarm to alert the driver. The ShuffleNet V2 classifier is used in the DER system classifies the different types of emotions. Matlab 2020b, with 16 GB RAM, an Nvidia GeForce GTX 1650 GPU, and an Intel Core i5 CPU, has been used to stimulate the suggested DER system using CNN classifier. CNN identifies the specific emotional state along with classifies the image's input data.

4.1 Dataset description

CK_plus [26], FER_2013 [27], TFEID [28], KMUFED [29], KDEF [30] are five datasets used in this proposed method. CK_Plus is a complete set of action units and emotion-based expressions. FER_2013 data is a 48 × 48 pixel face picture in grayscale. The Karolinska Directed Emotional Faces (KDEF) dataset is a set of human facial expressions.

Therefore, the data is initially, an image collected and it was preprocessed to eliminate noise in the dataset. The preprocessed data were then segmented to extract an area of an image that was used for further process. After segmentation, the features of the image were extracted, then the image data was fed into CNN classification. ShuffleNet V2 classifies the images, and the emotional state was determined. The preprocessed images of the Ck_plus dataset is mentioned in Table 1.

Probability can be corrected with the use of receiver operating characteristics (ROC). A single classifier's true positive rate is determined and plotted against the false positive rate to create the ROC curve. Figure 5 illustrates the Confusion matrix plot and ROC plot for CK_Plus dataset. An excellent classifier is represented by a value of 1, while a poor classifier is represented by a value of 0.5 on the ROC curve. The proposed method plots the true positive rate and false positive rate for the dataset CK Plus on a ROC graph. The proposed approach provides greater performance since the ROC curve hits 1. CK_Plus's ROC curve is displayed in Fig. 5a.

Examining the effectiveness of the categorization approach is done with the confusion matrix. Confusion matrix displaying the CK_Plus dataset's accuracy rate. There are seven distinct classifications in the dataset: fearful, furious, disgusted, joyful, depressed, shocked, and neutral. In class 0, accuracy was 80.0%; in class 1, accuracy was 100.0%; in class 2, accuracy was 92.3%; in class 4, accuracy was 100.0%; and in class 6, accuracy was 92.1%. At 93.41%, the dataset CK_Plus has an overall accuracy rating. As shown in Fig. 5b, the confusion matrix used the CK_Plus dataset.

The FER_2013 picture dataset is the second dataset utilized in this proposed methodology. The picture dataset is

Table 5 Pre-processing and Feature extraction in KDEF dataset

Input image	Resized image	Gaussian filter	Median filter	Histogram equalization	Wiener filter	KLT feature

first scaled, after which it undergoes processes such as Wiener filtering, histogram equalization, median filtering, and Gaussian filtering. The KLT feature method is then used to extract the image features. Table 2 indicates the way the FER_2013 dataset was preprocessed and features were extracted.

Plotting true positive and false positive values is provided by the ROC curve for the FER_2013 dataset. With a ROC score of 1, the proposed method performs better for this dataset. According to the FER_2013 dataset, the ROC curve is shown in Fig. 6a. 84.5, 68.1, 99.0, 91.4, 80.7, 94.0, and 84.0% are the accuracy rates for the seven distinct classes comprising the entire FER_2013 dataset. This dataset has an overall accuracy percentage of 83.68%. The FER_2013 dataset's confusion matrix is displayed in Fig. 6b.

TFEID image dataset is a subsequent dataset. The imagery dataset first performs resizing of an image, followed by the use of Gaussian, median, Wiener, Histogram equalization, and KLT feature extraction processes. Table 3 shows the steps of pre-processing and extraction of features in the TFEID dataset.

For the TFEID dataset, the true positive rate and false positive rate are displayed on a ROC curve. With this dataset, the proposed approach's ROC value is 1, indicating its great performance. Figure 7a displays the ROC curve based on the TFEID dataset. Seven distinct datasets are included within the TFEID dataset according to the proposed methodology. 85.7, 77.8, 70.0, 87.5, 61.5, 100.0, and 100.0% are the accuracy rates for the seven classes. The TFEID dataset's confusion matrix is displayed in Fig. 7b.

KMU_FED is an additional image dataset that undergoes a resizing procedure before undergoing the application of Gaussian, median, histogram equalization, and Wiener

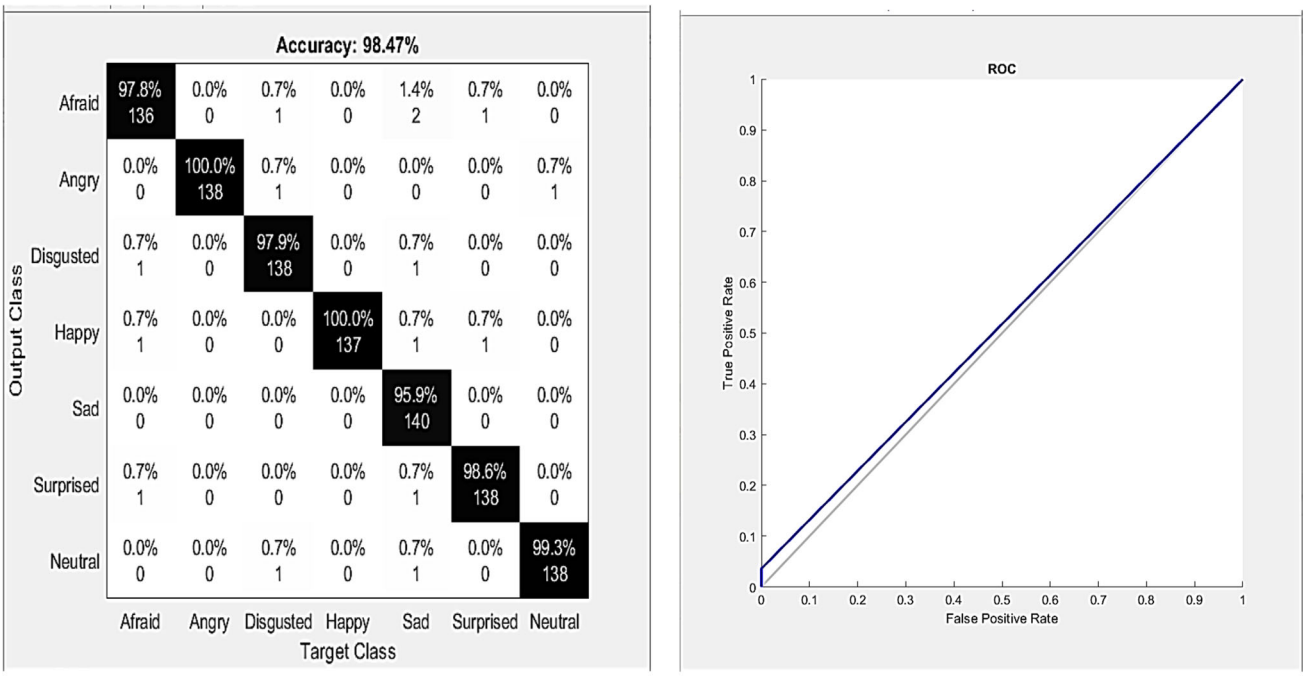


Fig. 9 Confusion matrix plot and ROC plot for KDEF dataset

filters. To extract the image's features, KLT extracting features is utilized. The KMU_FED dataset's pre-processing and extraction of features are displayed in Table 4.

For the KMU FED dataset, the true positive rate and false positive rate are displayed on the ROC curve. Proposed approach's ROC value for this dataset is 1, indicating that it performs well. Figure 8a displays the ROC curve with the dataset KMU_FED. Fearful, angry, disgusted, pleased, sad, shocked, and neutral are among the seven classes represented by the dataset KMU_FED. For class 0, accuracy was 97.6%; for class 1, accuracy was 100.0%; for class 2, accuracy was 100.0%; for class 3, accuracy was 92.3%; for class 5, accuracy was 95.2%. Dataset KMU_FED has an overall accuracy rating of 98.18%. Figure 8b displays a KMU_FED dataset's confusion matrix.

KDEF dataset is the last dataset used in this proposed method. This dataset contains methods for image resizing, feature extraction using KLT, histogram equalization, Wiener filter, and Gaussian filter. Table 5 indicates the Pre-processing and feature extraction in the KDEF dataset.

In the proposed approach, the true positive rate and false positive rate are displayed on a ROC graph for the dataset KDEF. This proposed method's ROC curve becomes 1, indicating improved performance. The KDEF dataset's ROC curve is displayed in Fig. 9a. Seven distinct classes constitute Toward KDEF dataset; the accuracy rates for these classes are 95.9, 98.6, 99.3, 97.8, 100.0, and 97.9%. This dataset has an overall accuracy percentage of 98.47%. The KDEF dataset's confusion matrix is displayed in Fig. 9b.

The degree of similarity to the real value determines the accuracy. Figure 10a displays the accuracy of known methods using ResNet-101 and proposed approaches utilizing ShuffleNet V2 over five distinct datasets. The proposed method's accuracy rate based on the CK_Plus dataset is 0.94, whereas the existing method's accuracy rate was 0.89. Furthermore, the datasets KDEF achieves 0.99, TFEID achieves 0.79, FER_2013 achieves 0.99, and KMU_FED achieves 0.99 in the proposed approach, compared to 0.80, 0.94, 0.95, and 0.70 accuracy in the existing method. This demonstrates unequivocally the superiority of the proposed methodology over the existing one. The proposed ShuffleNet V2 technique and the existing ResNet-101 approach's sensitivity are depicted in Fig. 10b. In the existing technique, the sensitivity of the various datasets was 0.80, 0.78, 0.90, 0.90, and 0.70, respectively. Sensitivity results for the FER_2013, TFEID, KDEF, KMU_FED, and CK_Plus datasets were 0.87, 0.85, 0.99, 0.99, and 0.79, respectively, according to the proposed technique. Therefore, the proposed approach was superior to the existing strategy. Figure 10c illustrates the specificity of the proposed strategy and the existing procedure. The various datasets provide 0.98, 0.97, 0.99, 0.99, and 0.97 specificity in the proposed technique, respectively; in the existing method, these datasets generate 0.92, 0.91, 0.88, 0.87, and 0.88 specificity. This demonstrates unequivocally how superior the proposed approach is above the existing one.

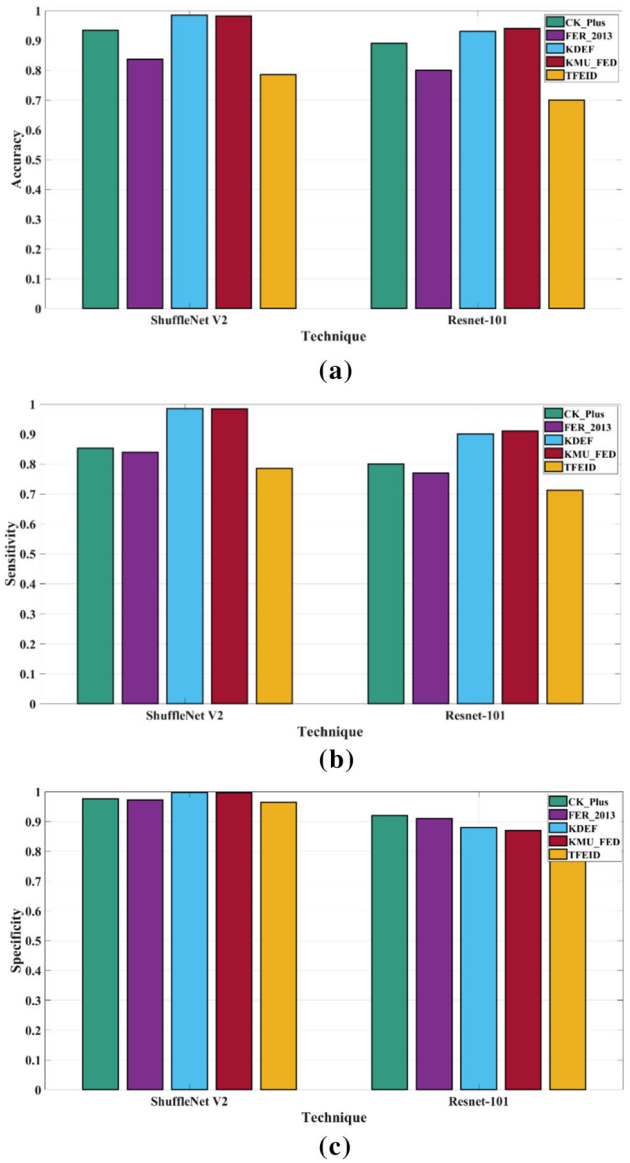


Fig. 10 Comparison of proposed ShuffleNet V2 and existing ResNet 101 approaches (**a**) Accuracy (**b**) Sensitivity (**c**) Specificity

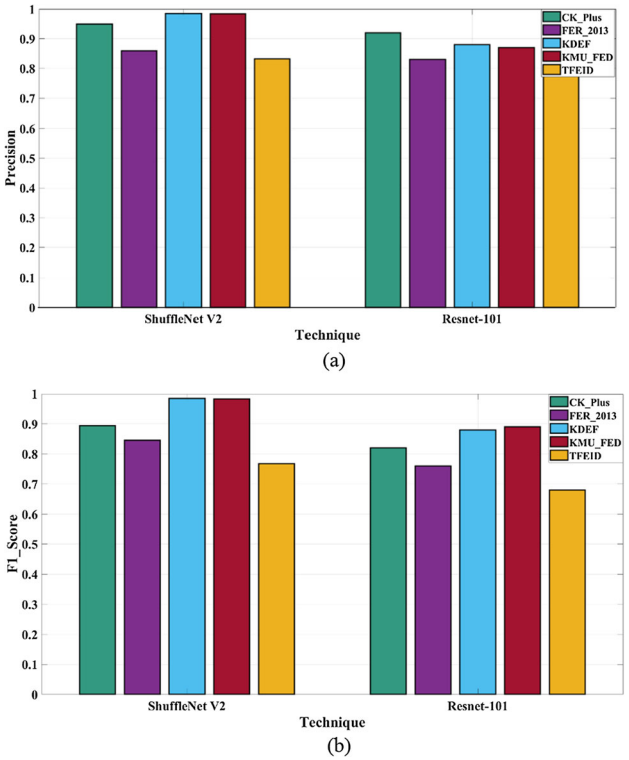


Fig. 11 Comparison of proposed ShuffleNet V2 and existing ResNet 101 approaches (**a**) Precision (**b**) F1_Score

Proposed ShuffleNet V2 and existing ResNet101 techniques' precision analyzes are displayed in Fig. 11a. For the datasets FER_2013, TFEID, KDEF, KMU_FED, and CK_Plus, the precision values in the proposed technique are, in order, 0.97, 0.88, 0.99, 0.99, and 0.84. For the same dataset, the existing method's precision values are 0.92, 0.82, 0.88, 0.88, and 0.81. This demonstrates that compared to the existing values, the proposed method yields a higher accuracy value. Next, F1_Score is examined and illustrated in Fig. 11b. In comparison to the existing method, which has an F1_Score value of 0.81, 0.78, 0.88, 0.89, and 0.69 for the various datasets, respectively, the proposed approach's F1_Score value is 0.9 in CK_Plus, 0.85 in FER_2013, 0.98 in KDEF and KMU_FED, and 0.77 in TFEID. This demonstrates the superiority of the proposed technique over the existing ones.

Figure 12a analyzes and illustrates FPR. The FPR value produced by ResNet 101, VGG-12, VGG-16 and ResNet 50 existing techniques are 0.08, 0.03, 0.09 and 0.025, respectively. But, the proposed model has 0.005 FPR value. The proposed approach gives less FPR compared to the existing approach, so this proposed method is comparatively better than the existing method. The proposed ShuffleNet V2 and existing ResNet101 methods' Kappa values are displayed in Fig. 12b. The Kappa value produced by ResNet 101, VGG-12, VGG-16 and ResNet 50 existing techniques are 0.83, 0.84, 0.81 and 0.79, respectively. But, the proposed model has 0.90 Kappa value. Then, Matthews's correlation coefficient (MCC) is sketched in Fig. 12c. Compared to the other four approaches, the proposed method's MCC is 0.9. According to the results, ResNet 101 had 0.83, VGG-12 had 0.82, VGG-12 had 0.88, and ResNet 50 had 0.84.

Figure 13a analyzes and illustrates the Negative Predictive Value (NPV). The NPV value produced by ResNet 101, VGG-12, VGG-16 and ResNet 50 existing techniques are 0.89, 0.87, 0.82 and 0.85, respectively. But, the proposed model has 0.91 NPV value. The proposed approach gives less NPV compared to the existing approach, so this proposed method is comparatively better than the existing method. The proposed ShuffleNet V2 and existing ResNet101methods'

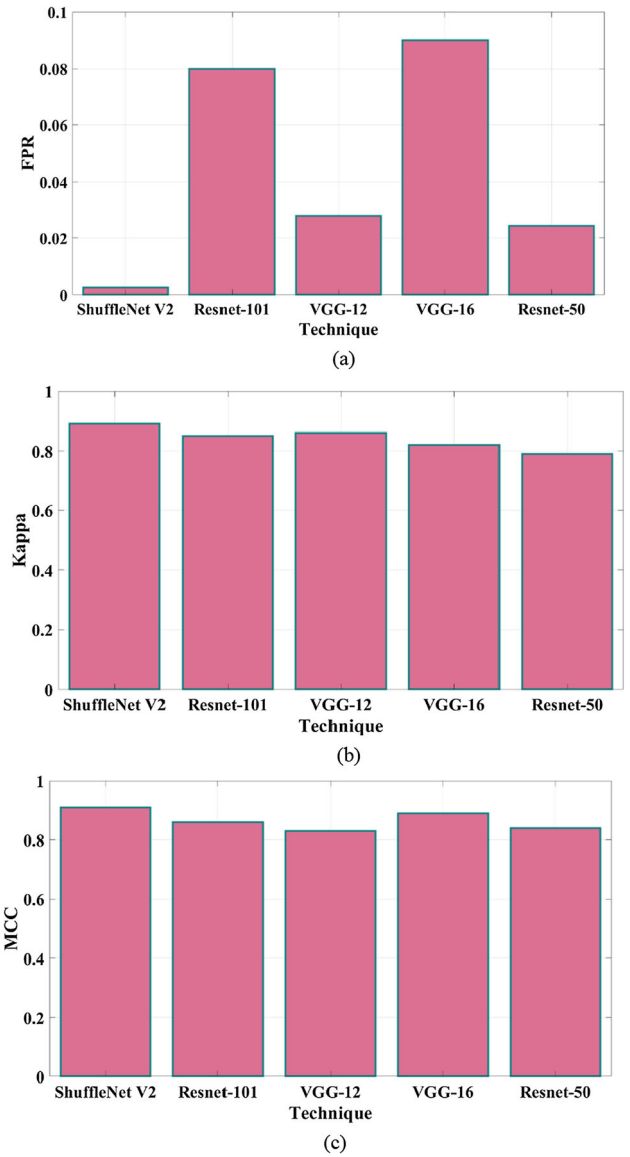


Fig. 12 Comparison of proposed Existing approaches (**a**) FPR (**b**) Kappa (c) MCC

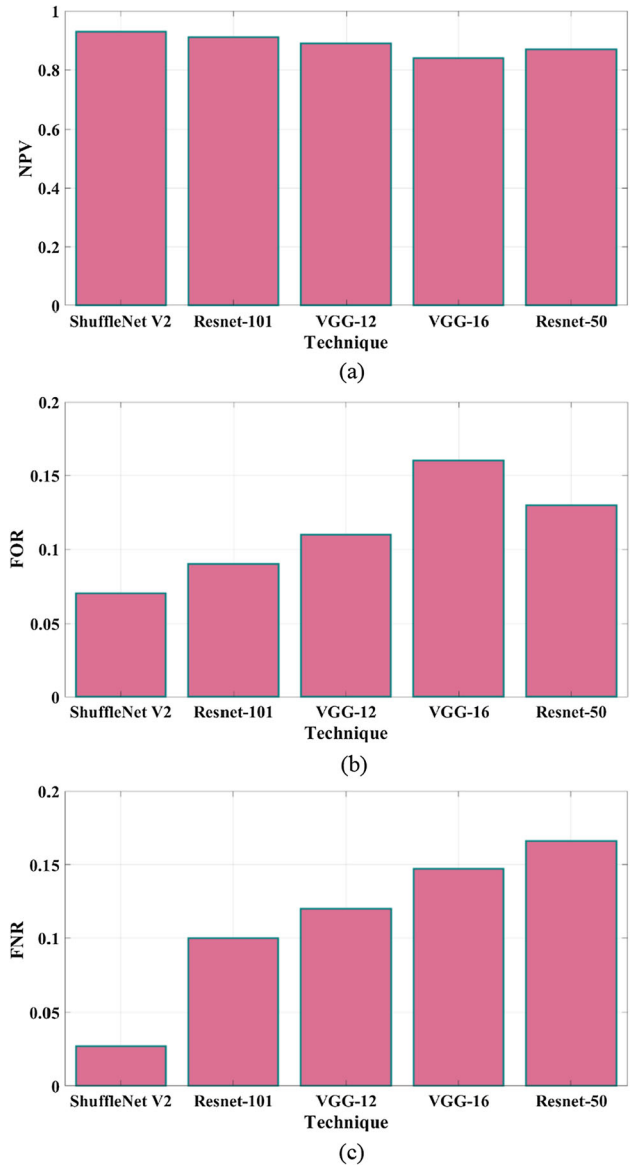


Fig. 13 Comparison of proposed and Existing approaches (**a**) NPV (**b**) FOR (c) FNR

False Omission Rate (FOR) values are displayed in Fig. 13b. The FOR value produced by ResNet 101, VGG-12, VGG-16 and ResNet 50 existing techniques are 0.08, 0.13, 0.16 and 0.13, respectively. But the proposed model has 0.07 FOR value. Then, False Negative Rate (FNR) is sketched in Fig. 13c. Compared to the other four approaches, the proposed method's FNR is 0.03. According to the results, ResNet 101 had 0.1, VGG-12 had 0.13, VGG-12 had 0.149, and ResNet 50 had 0.17.

Figure 14a analyzes and illustrates the False Discovery Rate (FDR). The FDR value produced by ResNet 101, VGG-12, VGG-16 and ResNet 50 existing techniques are 0.15, 0.07, 0.142 and 0.125, respectively. But, the proposed model has 0.05 FDR value. The proposed approach gives less FDR compared to the existing approach, so this proposed method is comparatively better than the existing method. The proposed ShuffleNet V2 and existing ResNet101 methods' Informedness values are displayed in Fig. 14b. Compared to the other four approaches, the proposed method's informedness is 0.9. According to the results, ResNet 101 had 0.7, VGG-12 had 0.82, VGG-12 had 0.73, and ResNet 50 had 0.83.

ShuffleNet V2 classifier is used in the proposed DER system to classify the driver's various emotional states. Compared to the existing method, the proposed ShuffleNet V2 methodology has higher accuracy, sensitivity, specificity, precision, F1_Score, Kappa, and Matthews's correlation coefficient (MCC), Negative Predictive Value (NPV), False

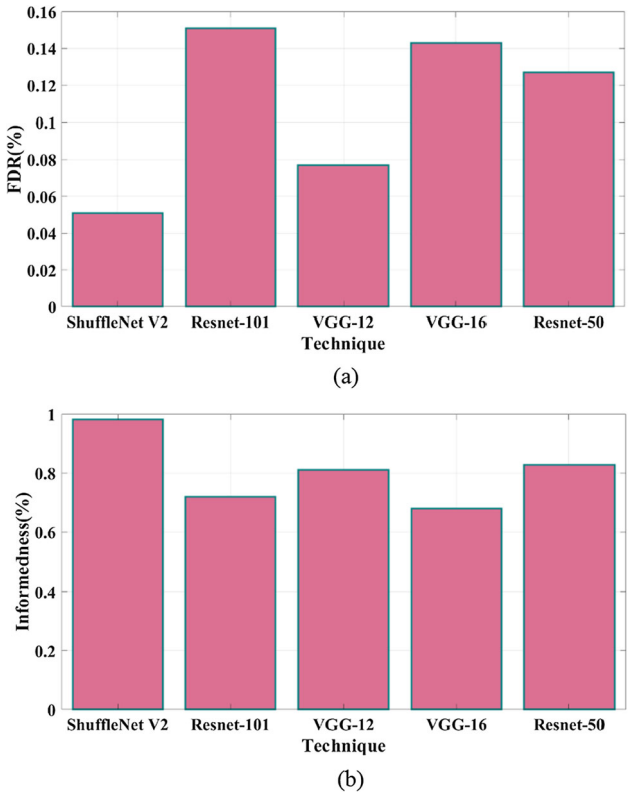


Fig. 14 Comparison of proposed and Existing approaches **a** FDR **b** Informedness

Omission Rate (FOR), False Negative Rate (FNR), False Discovery Rate (FDR), Informedness and lower False Positive Rate (FPR) errors. Based on the results, Driver Emotion Recognition using the ShuffleNet V2 classifier is the most appropriate method for identifying the drivers' emotional state at that moment.

Above mentioned Table 6 illustrates the state of the art methods for driver emotional recognition. The state of the art methods are CNN (InceptionV3-VGG16), MLCNN, MRE-CNN and DLBP-DCT. By evaluating these existing models with the CK_Plus dataset. The performance metrics such as accuracy, precision, sensitivity and F1_score of the model are attained. Based on these attained values, the performance metrics of the proposed model are compared with the existing state of the art methods that results the proposed model as a better driver recognition model than the existing models.

5 Conclusion

Emotion-related human–machine systems are essential for intelligent automobiles, as driver emotions impact driving performance and contribute to traffic accidents. Detecting and recognizing driver emotions is emerging as a critical factor for improving the driver safety. The absence of real-scenario datasets hinders current research in on-road driver facial expression detection, which is crucial for automotive human–machine systems. In this proposed model, ShuffleNet V2 based Driver Emotion Recognition (DER) has been designed. Different types of emotions of the drivers are recognized and classified by this proposed DER system. The Image Resizing, noise removal, smoothening and improving the brightness of the image techniques for the datasets such as FER_2013, TFEID, KMU_FED, CK_Plus, and KDEF are performed in preprocessing. Then, the ROI based segmentation and KLT based texture feature extraction are processed for reducing the complexity and enhancing the recognition of the model. Different types of emotions are classified by the ShuffleNet V2 classifier. The accuracy rate obtained by the DER system with the ShuffleNet V2 classifier was much better than the existing model including ResNet 101, VGG-12, VGG-16 and ResNet 50. Accuracy, Precision, sensitivity, F1-score, specificity, False Positive Rate (FPR), Kappa, and MCC are some of the performance measures used to assess efficacy for this proposed model. The proposed model's achieved performance metrics values are 0.99, 0.99, 0.90, 0.89, 0.99, 0.005, 0.90, and 0.90. Thus, the proposed method may be a useful substitute for enhancing the existing techniques in recognizing the driver emotional. In future score, the designed driver recognition model can be able to design in an automotive industry as a safety and secure feature for futuristic vehicles. Likewise, this feature can also be implemented on other industries such as medical industry,

Table 6 Comparison of State-of-the-Art methods for driver emotional recognition

Performance metrics	State-of-the-Art Methods				
	CNN (InceptionV3-VGG16) [31] CK_Plus dataset	MLCNN [32]	MRE-CNN [33]	DLBP-DCT [34]	Proposed model
Accuracy	75	80	86	92	94
Precision	71	73.6	79.9	83.4	84
Sensitivity	68.7	71.3	75.9	79.5	79
F1-score	71.8	77	81	86	90

construction industry, chemical industry, petroleum industry and power engineering as a monitoring system for the workers in critical zone and highly secured zone. This emotional recognition model can also be used in various fields, such as human–computer interactions (HCI), medical health, Internet education, security monitoring, psychological analysis and the entertainment industry.

Author contributions Faiyaz Ahmad: corresponding author involved in conceptualization, methodology, writing, editing and reviewing. U. Hariharan: Modeling of Proposed Methodology and problem solving using the proposed technique. Dr. N. Muthukumaran: Results analysis and implementation. Dr. Aleem Ali: Manuscript writing and editing. Dr. Shivi Sharma : Results analysis and validation.

Funding There is no funding provided to prepare the manuscript.

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

Code availability No code is available for this manuscript.

Declarations

Conflict of interest There is no conflict of Interest between the authors regarding the manuscript preparation and submission.

References

1. Weber, M., Giacomin, J., Malizia, A., Skrypchuk, L., Gkatzidou, V., Mouzakitis, A.: Investigation of the dependency of the drivers' emotional experience on different road types and driving conditions. Transport. Res. F: Traffic Psychol. Behav. **65**, 107–120 (2019)
2. Karthick, S., Muthukumaran, N.: Deep regression network for single-image super-resolution based on down- and upsampling with RCA blocks. Natl. Acad. Sci. Lett. (2023). https://doi.org/10.1007/s40009-023-01353-5
3. Payalan, Y.F., Guvensan, M.A.: Toward next-generation vehicles featuring the vehicle intelligence. IEEE Trans. Intell. Transp. Syst. **21**(1), 30–47 (2019)
4. Seng, K.P., Ang, L.M., Ooi, C.S.: A combined rule-based & machine learning audio-visual emotion recognition approach. IEEE Trans. Affect. Comput. **9**(1), 3–13 (2016)
5. Karthick, S., Muthukumaran, N.: Deep regression network for the single image super resolution of multimedia text image. In: 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), pp. 394–399. IEEE (2023)
6. Kundu, T., Saravanan, C.: Advancements and recent trends in emotion recognition using facial image analysis and machine learning models. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), (2017) IEEE, pp. 1–6
7. Izquierdo-Reyes, J., Ramirez-Mendoza, R.A., Bustamante-Bello, M.R., Pons-Rovira, J.L., Gonzalez-Vargas, J.E.: Emotion recognition for semi-autonomous vehicles framework. Int. J. Interact. Des. Manuf. (IJIDeM) **12**, 1447–1454 (2018)
8. Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion recognition: a review. In 2011 IEEE 7th international colloquium on signal processing and its applications, (2011) IEEE, pp. 410–415
9. Kolli, A., Fasih, A., Al Machot, F., Kyamakya, K.: Non-intrusive car driver's emotion recognition using thermal camera. In Proceedings of the Joint INDS'11 & ISTET'11, (2011) IEEE, pp. 1–5
10. Du, G., Wang, Z., Gao, B., Mumtaz, S., Abualnaja, K.M., Du, C.: A convolution bidirectional long short-term memory neural network for driver emotion recognition. IEEE Trans. Intell. Transp. Syst. **22**(7), 4570–4578 (2020)
11. Wang, X., Guo, Y., Ban, J., Xu, Q., Bai, C., Liu, S.: Driver emotion recognition of multiple-ECG feature fusion based on BP network and D-S evidence. IET Intel. Transp. Syst. **14**(8), 815–824 (2020)
12. Xia, K., Gu, X., Chen, B.: Cross-dataset transfer driver expression recognition via global discriminative and local structure knowledge exploitation in shared projection subspace. IEEE Trans. Intell. Transp. Syst. **22**(3), 1765–1776 (2020)
13. Hu, Y., Lu, M., Xie, C., Lu, X.: Driver drowsiness recognition via 3D conditional GAN and two-level attention Bi-LSTM. IEEE Trans. Circ. Syst. Video Technol. **30**(12), 4755–4768 (2019)
14. Jeong, M., Nam, J., Ko, B.C.: Lightweight multilayer random forests for monitoring driver emotional status. IEEE Access **8**, 60344–60354 (2020)
15. Shojaeilangari, S., Yau, W.Y., Nandakumar, K., Li, J., Teoh, E.K.: Robust representation and recognition of facial emotions using extreme sparse learning. IEEE Trans. Image Process. **24**(7), 2140–2152 (2015)
16. Kim, C.M., Kim, K.H., Lee, Y.S., Chung, K., Park, R.C.: Real-time streaming image based PP2LFA-CRNN model for facial sentiment analysis. IEEE Access **8**, 199586–199602 (2020)
17. Mohan, K., Seal, A., Krejcar, O., Yazidi, A.: Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks. IEEE Trans. Instrum. Meas. **70**, 1–12 (2020)
18. Cui, Y., Ma, Y., Li, W., Bian, N., Li, G., Cao, D.: Multi-EmoNet: a novel multi-task neural network for driver emotion recognition. IFAC-Pap. OnLine **53**(5), 650–655 (2020)
19. Madupu, R.K., Kothapalli, C., Yarra, V., Harika, S., Basha, C. Z.: Automatic human emotion recognition system using facial expressions with convolution neural network. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA), (2020) IEEE, pp. 1179–1183
20. Mukherjee, D., Mukhopadhyay, S.: Fast hardware architecture for fixed-point 2D Gaussian filter. AEU-Int. J. Electron. Commun. **105**, 98–105 (2019)
21. Moghaddam, A.A., Rangarajan, L.: Enhancing radiographic images using two dimensional left median filter. In 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (2011) IEEE, pp. 204–206
22. Sree Vidya, B., Chandra, E.: Triangular fuzzy membership-contrast limited adaptive histogram equalization (TFM-CLAHE) for enhancement of multimodal biometric images. Wirel. Pers. Commun. **106**, 651–680 (2019)
23. Petkova, L., Draganov, I.: Noise adaptive Wiener filtering of images. In 2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) (2020) IEEE, pp. 177–180
24. Nam, T., Kim, S., Jung, D.: Hardware implementation of KLT tracker for real-time intruder detection and tracking using on-board camera. Int. J. Aeronaut. Space Sci. **20**, 300–314 (2019)
25. Guo, M., Du, Y.: Classification of thyroid ultrasound standard plane images using ResNet-18 networks. In 2019 IEEE 13th international conference on anti-counterfeiting, security, and identification (ASID) (2019) IEEE, pp. 324–328
26. Ashadullah Shawon: Kaggle: [https://www.kaggle.com/datasets/shawon10/ckplus]. Accessed on 06-08-2023 (2019)

27. Manas Sambare: Kaggle: [https://www.kaggle.com/datasets/msambare/fer2013/code]. Accessed on 06-08-2023 (2021)
28. Jonathan Oheix: Kaggle: [https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset]. Accessed on 06-08-2023 (2019)
29. KMU-FED. Available online: http://cvpr.kmu.ac.kr/KMU-FED.htm
30. Muhammad Nafian: kaggle: [https://www.kaggle.com/datasets/muhammadnafian/kdef-dataset]. Accessed on 06-08-2023 (2023)
31. Bolioli, A., Bosca, A., Damiano, R., Lieto, A., Striani, M.: A complementary account to emotion extraction and classification in cultural heritage based on the Plutchik's theory. In: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, New York, NY, United States, pp 374–382 (2022). https://doi.org/10.1145/3511047.3537659
32. Malakar, S., Ghosh, M., Bhowmik, S., Sarkar, R., Nasipuri, M.: A GA based hierarchical feature selection approach for handwritten word recognition. Neural Comput. Appl. **32**(7), 2533–2552 (2020)
33. Fan, Y., Li, V.O., Lam, J.C.: Facial expression recognition with deeply-supervised attention network. IEEE Trans. Affect. Comput. **13**(2), 1057–1071 (2020)
34. Bhattacharya, S.: A survey on: facial expression recognition using various deep learning techniques. Advanced computational paradigms and hybrid intelligent computing. Springer, Singapore, pp 619–631 (2022)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.