Facial expression recognition based on active region of interest using deep learning and parallelism

Mohammad Alamgir Hossain and Basem Assiri

Department of COMPUTER SCIENCE, College of Computer Science & Information Technology, Jazan University, Jazan, Kingdom of Saudi Arabia

ABSTRACT

The automatic facial expression tracking method has become an emergent topic during the last few decades. It is a challenging problem that impacts many fields such as virtual reality, security surveillance, driver safety, homeland security, human-computer interaction, medical applications. A remarkable cost-efficiency can be achieved by considering some areas of a face. These areas are termed Active Regions of Interest (AROIs). This work proposes a facial expression recognition framework that investigates five types of facial expressions, namely neutral, happiness, fear, surprise, and disgust. Firstly, a pose estimation method is incorporated and to go along with an approach to rotate the face to achieve a normalized pose. Secondly, the whole face-image is segmented into four classes and eight regions. Thirdly, only four AROIs are identified from the segmented regions. The four AROIs are the nose-tip, right eye, left eye, and lips respectively. Fourthly, an info-image-data-mask database is maintained for classification and it is used to store records of images. This database is the mixture of all the images that are gained after introducing a ten-fold cross-validation technique using the Convolutional Neural Network. Correlations of variances and standard deviations are computed based on identified images. To minimize the required processing time in both training and testing the data set, a parallelism technique is introduced, in which each region of the AROIs is classified individually and all of them run in parallel. Fifthly, a decision-tree-level synthesis-based framework is proposed to coordinate the results of parallel classification, which helps to improve the recognition accuracy. Finally, experimentation on both independent and synthesis databases is voted for calculating the performance of the proposed technique. By incorporating the proposed synthesis method, we gain 94.499%, 95.439%, and 98.26% accuracy with the CK+ image sets and 92.463%, 93.318%, and 94.423% with the JAFFE image sets. The overall accuracy is 95.27% in recognition. We gain 2.8% higher accuracy by introducing a decision-level synthesis method. Moreover, with the incorporation of parallelism, processing time speeds up three times faster. This accuracy proves the robustness of the proposed scheme.

Subjects Human-Computer Interaction, Computer Vision, Distributed and Parallel Computing, Graphics, Visual Analytics

Keywords Optimized searching algorithm, Pose estimation, Rotation and correction, Parallelism, Deep learning, Active region of interest, Facial expression recognition, Convolutional neural network
INTRODUCTION

Facial expressions are an important way for humans to express emotions, and they play a noteworthy role in our day-to-day lives. Facial expression recognition (FER) is a multifaceted problem that has motivated people during the last few decades. The applications of FER are found in virtual reality, security surveillance (Hossain & Sanyal, 2016), driver safety, homeland security, human–computer interaction (HCI) (Jiwen, Gang & Jie, 2017), health care, medical applications, eyes’ disease diagnosis (Hossain, Samanta & Goutam, 2012), and many more. There are various types of applications based on face images in our daily life so FER has attracted many researchers to explore image processing. Research scholars propose various innovative approaches to recognize facial expression (Jun et al., 2018) and its authentication (Mohammad & Assiri, 2020) in the recent past. However, there is still lacking such a FER system enriched with a higher percentage of recognition accuracy and lesser execution time (Hossain & Assiri, 2020a). There is the scope to incorporate new ideas and better techniques so that higher recognition accuracy can be achieved in a minimal time (Daikhong et al., 2021). There are challenges in FER because in recognizing any face, several factors are interrelated.

The researchers usually use normal digital images to recognize human faces for many reasons, e.g., the cost of cameras, ease of understanding, and simplicity of implementation. This recognition scheme is divided into the following major phases:

- Processing of the facial image
- Training and testing image representation
- Classification of the desired expression

**Processing of the facial image:** This part is vital to complete the whole process because it includes all the processes mentioned in the proposed workflow diagram (Fig. 1), namely: read an image, pre-processing, face image and point of interest (POI) (Daikhong et al., 2021; An et al., 2019) detection, pose estimation (Hossain & Assiri, 2020b; Yongbin et al., 2019), image segmentation and AROI detection. A FER system is entirely dependent on the actual pixels of that face image. An optimized searching algorithm is incorporated to complete faster tracking. This algorithm is explained later and is called Algorithm-2.

**Training and testing image representation:** This part is concerned with preparing images for testing in the system. Here images are fed in for testing and training. Based on successful training, trained images are achieved and subsequently, the required features will be extracted. A ten-fold cross-validation technique is proposed for image set splitting for training and testing.

**Classification of the desired expression:** The prime goal of this paper remains in classifying the desired expressions out of five kinds (neutral, happiness, fear, surprise, and disgust). In this work, the Convolutional Neural Network (CNN) (Ran et al., 2019; Galea & Farrugia, 2018; Jae & Bumshik, 2020; Ferrari et al., 2018; Sun et al., 2018; Mollahosseini, Chan & Mahoor, 2016; Shen, Wang & Liu, 2013; Yahia, Mohammad & Hossam, 2020; AlBdairi, Zhu & Mohammed, 2020; Almeida et al., 2020; Boyina, 2020; Rajan et al., 2020) is incorporated to classify the expression from the trained data. To classify the desired
expressions, features are extracted first from the trained image sets. We introduced CNN to classify three AROIs in parallel, so there will be three outputs. We made this classification because it is necessary for experimentation. We need to conduct experiments on both training and testing data sets. In this experiment, we proposed the decision-tree-level synthesis (Salmam, Madani & Kissi, 2016) method, which is illustrated in ‘Classification by decision-tree-level synthesis’.

This paper has a few prime contributions that are as follows.
1. The entire face is not used. Only four AROIs are chosen from the entire face to classify expressions.
2. A search technique called OAROI is proposed to search the similarity of the AROIs.
3. We propose a decision-tree-level synthesis-based framework to achieve improved accuracy in expression and its classification.
4. Parallel processing is applied to reduce execution time.

The remaining part of our manuscript is organized in the following way: Section 2 illustrates few “Related Works”. In Section 3, the “Proposed Methodology for Expression Recognition” depicts the designed frameworks for expression recognition. Sub-section 3.1 describes the “Data Processing”, Sub-section 3.2 illustrates “Rotation Using the Spatial Normalization Method”. Sub-section 3.3 explains the “Searching for the Optimized Active Region of Interest”. In Sub-section 3.4, “Parallel Face and Facial AROI Detection” is explained. In Sub-section 3.5, the “Classification by Decision-tree-level Synthesis” is introduced to achieve improved accuracy in expression classification. The experimentation
and discussion are explained in Section 4. Lastly, we came to a conclusion that shows the robustness of our work. In this regard, we made the necessary comparison based on our experimental results and the results of other researchers. However, we have given the future direction of our work.

RELATED WORK

Human face tracking has become a promising area in the last few decades. To search for a few parts from an entire face for training is an emergent idea. Based on a few parts it is possible to gain all the required features. This unit overviews to extract features by incorporating appearance-based facial expression (Hossain & Sanyal, 2013) and emotion tracking techniques (Hossain, Zogan & Sanyal, 2018). To track facial expression accurately, it is necessary to use a better facial recognition technique. Due to the rising importance of accurate face recognition techniques scholars have investigated and proposed various techniques in the current time. Hossain & Assiri, (2020a) proposed a method for tracking and verification at minimal cost by using a pipeline computation method. They incorporated the famous Sinusoidal Head Model (SHM) and applied adaptive PDF (APDF) to estimate pose. The completion of the FER methods depends on active facial landmark points. A high-resolution camera is required to capture the images. The traditional emotion recognition schemes (Hossain, Zogan & Sanyal, 2018; Zhang et al., 2020) have some limitations because they follow self-measurement methods. Such schemes are not fit for quick surveillance. Self-reporting systems are incapable of classifying emotion quickly. Since emotion is associated with physical appearance, some investigators (Hossain, Zogan & Sanyal, 2018; Hossain, Samanta & Sanyal, 2012a) have used the physical signs for emotion recognition. A few proposed models track people based on the lips (Hossain & Sanyal, 2014), mouth, and eyebrows. There are some methods proposed by researchers to detect facial expressions by using various methods. Authors have suggested the FACS (facial action coding system) (Clark et al., 2020) to identify facial action. They presented face motions by using the Action Units (AUs) from the facial landmark-point. Correct recognition of AUs is essential when it is measured employing FACS, although it is hard to identify all the types of AUs. The scholars (Clark et al., 2020; Hossain & Goutam, 2012) have classified FER by applying FACS, geometric algorithm, and region-based systems (Hossain & Goutam, 2012) and as well as appearance-based approaches to extract features (Happy & Routray, 2015; Lajevardi & Hussain, 2021). In a geometric-based system (Lajevardi & Hussain, 2021), the locations or regions of a face (like eyes, eyebrows, mouth, and lips) are extracted from features-vector that identifies facial geometry. A geometry-based approach should attain alike performance to that of an appearance-based scheme. Appearance-based approaches (Hossain & Goutam, 2012; Happy & Routray, 2015; Lajevardi & Hussain, 2021) use all features related to facial appearance. A few investigators (Happy & Routray, 2015; Lajevardi & Hussain, 2021) have separated facial regions into a few blocks so that feature extractions become easier. Improved accuracy is attained using these techniques (Lajevardi & Hussain, 2021). It has been shown that facial features are directly related to expressions. These expressions have a direct correlation with the mouth, eyes, lips, nose, etc. Their
poses, positions, and shapes are required for classifying facial expressions. These works are inspiring. These lead us to find out the concept of active regions of interest (AROI) on a human face. Zhang & Wang (2020) proposed face recognition algorithms for fatigue detection by using video image information.

A few researchers (Hossain & Sanyal, 2012a; Alamir & Sanyal, 2012c; Hossain, Samanta & Sanyal, 2012c) recognize expressions like smiles, panic, and other expressions. They incorporated a haar-like method in detecting face and used the AdaBoost method to select features. This method had also been incorporated into the Open Source Computer Vision (OpenCV) Library (Fan et al., 2012). An OpenCV technique uses API tool to identify activity from a desktop computer in a real-time application. So the research scholars have used OpenCV to recognize facial image. Face alignment is an essential scheme to achieve better performance in FER. The objective of an alignment is restricted to a predefined AROI (Sun et al., 2018) from a face image. Scholars (Liu et al., 2016) proposed how to identify facial landmarks based on shape indexed features. A few companies, like Microsoft, Face++ have applied their state-of-the-art techniques to identify facial landmarks. They used their API tools and techniques in identifying face landmark points.

In general, it is found in FER schemes that the whole area of a face is taken into account. But if the whole face is divided into few regions then there must have few benefits. In the article by Liu et al. (2016) it was shown that a whole face-image is separated into 6x7 identical patches. Then, the LBP (Local Binary Patterns) features are pulled out from these patches. The SVM (Support Vector Machine) classifier is applied to classify facial expressions. There are some drawbacks to this method in fixing regions’ sizes and their positions. Scholars proposed the AdaBoost (Geng et al., 2019) method to get rid of this problem and they used LBP features in their work. They achieved two features AdaBoost and LBP. They used sub-window in their application and scaled the face as per the window size. Liu et al. (2016) separated the face into 8x8 and suggested multi-task sparse (MTS) technique to find the active areas. They pinpointed active areas within the nose, mouth, eyes, and lip regions. We have incorporated AROIs in this paper and proposed a searching algorithm called OAROI. Hossain et al. propose the Gabor-wavelet (Hossain & Sanyal, 2012b) and genetic algorithm (GA) (Hossain, Samanta & Sanyal, 2012b) for feature extraction and classification. They showed that these techniques are efficient than that of a geometric-based method. These kinds of features are expensive to extract in terms of computation. However, the proposed LBP as a low-cost scheme to extract dissimilar features from a facial expression. LBP is broadly used to identify facial expressions (Alam & Jianu, 2017) due to its performance.

During the last couple of years, it is noticed that mutation technique (Hossain & Sanyal, 2012b) and GA is applied in FER. Liu et al. (2016) addressed Boosted Deep Belief Network (BDBN) to identify face. The BDBN had a small number of weak classifiers and for each of them is applied for an expression. They conducted their experiment on both CK+ (Mollahosseini, Chan & Mahoor, 2016; Shen, Wang & Liu, 2013) and JAFFE (Mollahosseini, Chan & Mahoor, 2016) databases and attained 96.7% and 91.8% accuracy in recognition. This accuracy is very high as compared to other methods but the BDBN needs more time to train its data sets. The expected time required for training data was almost 8 days.
Ding, Zhou & Chellappa (2017) proposed the FaceNet2ExpNet method and they designed a two-folded scheme. Initially, the face-net is applied to systematize FER at the pre-training stage. However, they completed the testing of the data set in a connected layer. Zeng et al. (2018) proposed FER classification by using DSAE (deep sparse auto encode). They used three categories of classifiers namely LBP, gray value, and HOG.

In recent times Manda, Bhaskare & Muthuganapathy (2021) used the deep learning (Zeng et al., 2018; Rajeev, Vishal & Rama, 2019; Guosheng et al., 2018; Nada & Heyam, 2020) method to classify engineering models by using CNN (Lopes et al., 2017). They used huge datasets in their experiment and applied GPUs technique to minimize the computation time. It is addressed in Geng et al. (2019) that if ANN (artificial neural network) (Suri, Vatsa & Singh, 2020) is applied with CNN (Lucey et al., 2010; Mollahosseini, Chan & Mahoor, 2020; Yahia, Mohammad & Hossam, 2020; Almeida et al., 2020; Rajan et al., 2020) for detecting face then more accuracy can be gained. These detection methods are associated with correctness and recall rates, so they can be evaluated and applied in object detection. It is found that among the experiment results that the best one is the ANN method in terms of accuracy. Apart from the aforementioned algorithms, there are a few approaches like CNN to identify people by face recognition where the influences of poses, viewpoints, and illuminations have direct effects by tracking. The CNN is one kind of deep learning technique (Manda, Bhaskare & Muthuganapathy, 2021; Samadi, Akbarizadeh & Kaabi, 2019; Carletti et al., 2020; Shiqing et al., 2019; Rajeev, Vishal & Rama, 2019; Shen, Wang & Liu, 2013; Boyina, 2020) that can obtain features from used training data. It achieves features with a combination of convolutional and its sub-layers. This is surveyed by using a set of fully connected layers. In recent years, CNN is widely being used because of its proficient performance (Yahia, Mohammad & Hossam, 2020; Rajan et al., 2020). With the incorporation of this technique better performance is achieved and they have shown it in performance matrix (Sun et al., 2018; Yahia, Mohammad & Hossam, 2020) form. A few studies had been completed based on deep learning using CNN (Manda, Bhaskare & Muthuganapathy, 2021; Carletti et al., 2020; Shiqing et al., 2019; Sun et al., 2018; AlBdairi, Zhu & Mohammed, 2020; Saranya et al., 2020) and Machine Vision (MV) because MV is an appropriate method for achieving a high-performance rate. It plays a significant role together with Artificial Intelligence (AI) and Machine Intelligence (MI). Many scholars showed their investigational results based on deep learning (Jin, Cruz & Gonçalves, 2020; Jiwen, Junlin & Yap-Peng, 2017; Rajeev, Vishal & Rama, 2019; AlBdairi, Zhu & Mohammed, 2020; Saranya et al., 2020) and MV (Moghaddam, Akbarizadeh & Kaabi, 2019; Tirandaz & Akbarizadeh, 2016; Hossain, Samanta & Sanyal, 2013) techniques. Lopes et al. (2017) suggested a new approach to recognize facial expression by merging CNN with pre-processing like sample generation, correction of rotation, normalization of intensity, and so on. In this process, they attained 96.76% accuracy on the CK+ database with reduced time. The proposed method deals with various things so that it will be (i) faster in terms of access and process, (ii) have better performance in recognition. The proposed CNN achieves high accuracy in recognition with minimal training time. Apart from that, AROI is proposed to classify and identify the active regions for expected expressions as an alternative for the whole face region. Happy & Routray (2015) designated 19 active
landmarks over the face. They extracted LBP features from nineteen active landmarks but only four landmarks from top were studied for classification. AlBdairi, Zhu & Mohammed (2020) have proposed that their CNN-based deep learning method has achieved 96.9% accuracy in the verification of the human face. Saranya et al. (2020) have proposed that their proposed CNN- and LSTM-based method has achieved facial expression recognition with 81.60%, 56.68%, and 95.21% accuracy by using MMI, SFEW, and their data set respectively.

Liu et al. (2014) and Moghaddam, Akbarizadeh & Kaabi (2019) have tried to speed up the processing time by applying various techniques (Almeida et al., 2020) in their research, but few of them became successful. Here processing time is reduced to a greater extent, although the measurement is done in floating-point, not in the exact value. Geng et al. (2019) used parallelism in their experiments. During application, they applied both parallelism and pipeline in their application.

**PROPOSED METHODOLOGY FOR EXPRESSION RECOGNITION**

The proposed FER methodology is explained in sub-section 3.1. And it is elaborated through sub-methods. However, the other sub-sections are entitled “Rotation using the Spatial Normalization Method” (sub-section 3.2), “OAROI searching” (sub-section 3.3), “Parallel and facial AROI detection” (sub-section 3.4), and “Classification by the decision-tree-level synthesis” (sub-section 3.5).

**Proposed methodology in general**

In completing data processing, the following three phases are needed to work together. The three phases are (a) pre-processing of data; (b) training of classified data, and (c) testing of classified data. Figure 1 shows the flow diagram and illustrates the details of all the used steps. During the recognition process, these are followed. Phase-1 begins by accepting an input image into the system and it ends by detecting AROI. In phase-2, OAROI is detected first from the identified AROIs and then the training data sets are prepared for completing the testing in phase-3. Three operations are carried out in phase-3. The operations are (i) testing based on the OAROI data set, (ii) completion of the feature extraction training data being trained, and finally, (iii) classification of FER is completed.

**a. Pre-processing of data**

In pre-processing of data, images are being fed as input into the system and the required filtering techniques are applied. We applied weighted vector directional filter (WVDF) (Hossain, Samanta & Sanyal, 2012c) and SPAM filter (Shiqing et al., 2019). The mentioned filters are applied to minimize noises from the image. Moreover, pose estimation and rotation of an image for normalization are done as per requirement. To detect accurate facial expressions, pose estimation is an important aspect. Thus, the famous Adaptive Sinusoidal Head Model (ASHM) (Mohammad & Assiri, 2020) is applied, and it will be explained in detail in ‘Rotation using the Spatial Normalization Method’.
Now, it is necessary to detect landmarks points in the face image. These landmark points are known as POIs and Regions of Interest (ROIs) as shown in Fig. 2. Figure 2 depicts the AROIs. Four AROIs are taken to experiment out results. We gained the maximum accuracy in recognition from these four AROIs.

Then a segmentation method is applied to divide a face image into 4 classes and 8 regions as shown in Fig. 3. The classes are the forehead, eyes, nose, and mouth areas, where every class is divided vertically into two halves.

The face image is segmenting an image into four classes and eight regions. We have chosen AROI from those segmented regions. These AROIs are nose-tip (A), left eye (B), right eye (C), and lip (D) as shown in Fig. 2.

**b. Training of classified data**

Now, an optimized searching algorithm is applied called OAROI. This scheme is illustrated in 'Optimized Active Region of Interest (OAROI) searching'. To prepare data for training, as stated in the second phase of the proposed system, the following methods are incorporated.

a. An algorithm is applied, Algorithm 2, that will detect the OAROI and it will measure how much area will be considered for the most optimized area with the best result.

b. The data set is prepared for training data.

c. The training data set is ready and stored in the pre-defined memory storage location.
c. Testing of classified data

The testing data set is classified and segregated to prepare the trained data set. In this way, the classification is completed. However, the AROI-based CNN method is incorporated to gain the features for classification. The three regions left eye, right eye, and lip are considered. These three AROIs are trained and classified in parallel using CNN (explained in 3.4). The final classification is accomplished based on decision level synthesis method (explained in Classification by decision-tree-level synthesis’).

Rotation using the spatial normalization method

It is found that a few images are askew due to the motion of the camera. This would have affected the accuracy of recognition. To get rid of this situation and improve the approximation, we apply a head-pose estimation algorithm, called ASHM (Hossain & Assiri, 2020a). It has been shown (Hossain & Assiri, 2020a) that with ASHM, better results are achieved. We applied an adaptive sinusoidal head model (ASHM) to recognize and correct the head pose. We also applied PDF (Probability Density Function) along with ASHM. The Pose Matrix Pm was constructed first from the frame. The pose vector has a Rotation Matrix (Rm) and Transformations Matrix (Tm) for each frame, f. Each pose
vector has an Rm and a Tm. \(\text{i.e.,}\)

\[
\begin{align*}
P_m &= \begin{bmatrix} R_m & T_m \\ T_m^T & 1 \end{bmatrix} \begin{bmatrix} P_{m1} \\ P_{m2} \end{bmatrix} \begin{bmatrix} P_{m3} \\ P_{m4} \end{bmatrix} \\
&= \begin{bmatrix} P_{m1} \\ P_{m2} \\ P_{m3} \\ P_{m4} \end{bmatrix} \\
&= \begin{bmatrix} P_m^1 \\ P_m^2 \\ P_m^3 \\ P_m^4 \end{bmatrix}
\end{align*}
\tag{1}
\]

where \(R_m \in \mathbb{R}^{8 \times 8}\) and \(T_m \in \mathbb{R}^{8 \times 1}\) are vectors after transformations by Tm. \(P_m^1, P_m^2, P_m^3, \text{and } P_m^4\) are pose vectors.

The ASHM is applied to all frames starting from frame 1. In this way, by accessing all the frames, accurate features are attained from the trained images. We applied the LKT (Lucas Kanade Tracker) algorithm in the process of features extraction. These features have a straight association to the segmented classes. The identified AROIs are directly associated with head motion. However, it is necessary to apply the appropriate projection technique depending on the model of the camera. So a connotation can be established between a 3-D and a 2-D image plane.

Figure 4 illustrates an association among the 3-D points (indicated by \(a = (x, y, z)^{T_m}\) on the sinusoidal area) and its prediction point (indicated by \(b = (p, q)^{T_m}\) on the image plane. Here \((p & q)\) are evaluated following the Eqs. (2) and (3).

\[
p = d \frac{x}{z}
\tag{2}
\]

\[
q = d \frac{y}{z}.
\tag{3}
\]

where \(d\) is the foretold distance from camera. Meanwhile 2-D features ensure different inferences in constructing pose-related information, two factors have a direct relation to feature extraction:

1. Intensity of features.
2. Applied tracking method in 3-D form.

Moreover, to overcome this situation, a rotation correction algorithm is incorporated. The positions of that part of AROIs maybe eyes, maybe lips, or both.

Figure 5 shows the original image without any alteration. Figure 6 shows the image after rotation of 3.1% required for normalization. Figure 7 shows the connection between two eye-lids. They are associated by a red line over the horizontal axis. It is used to achieve the adjusted pose. Figure 8 shows the image after rotation & normalization. Indeed, the angle in the middle of the two lines must be corrected. We applied the rotation and normalization method to gain the proper pose. In this process, 3.1% rotation is required to connect two eyelids. It is as shown through Figs. 5–8 respectively. The process of rotation has been demonstrated in the following paragraph.

Let \((x_1, y_1)\) and \((x_2, y_2)\) be left-eye and right-eye centers. Now let us calculate the angle \(\theta\) by applying the equation no 1. So the image needs to be rotated to the extent of \(\theta\).

\[
\theta = \tan\left(\frac{y_2 - y_1}{x_2 - x_1}\right).
\tag{4}
\]

Equation (4) is for rotation by the angle \(\theta\). The face-image size may vary after rotation. In this situation, a difficulty arises in classifying FER with increased accuracy. During
rotation, if there exists a huge background part of the face area then a problem may arise. Although, the background part has less significance in recognizing facial expressions. As a consequence, we used the Spatial Normalization Method (SNM) to normalize the AROI areas of the face. The SNM is carried out based on a central point (the nose tip) of the face image. Let the length of the facial AROI be $\infty$ and the width is $\beta$. In this literature, it is set that $\infty = \beta = 120$ pixels. Figure 6 shows the association between both left eyes of the face image and rotation is applied as required. After the SNM, facial POIs and the AROIs are spotted once more.

Figure 9 explains how the AROIs are being scaled. Here ellipse shape is considered. We considered geometric shape because it becomes easy to scale all the required parts of a human face. In this manuscript, three active regions (AROIs) have been taken for our experiment.

**Optimized active region of interest (OAROI) searching**

The AROIs are detected based on four classes of segregation. The researchers (Hossain & Assiri, 2020a; Clark et al., 2020; Lopes et al., 2017) proposed that few regions are not active at all. Scholars (Lopes et al., 2017) separated a face area into 8*8 patches. They select active regions from 64 patches. These active patches are found within the eyes, nose, and mouth regions. It is a patch-based method only. The identified AROIs are shown in Fig. 9. The strong features are found in the AROIs. The strong features is considered by its feature weight because they yield more trustworthy information. The feature weights are extracted
from nose tip, left eye, right eye, and lip regions and they are named FWA, FWB, FWC, and FWD respectively. The achieved feature-weights are as follows:

- Nose-tip FWA (24, 25, 26, 28, 29) = 0.0286943
- Left Eye FWB (30, 31, 32, 33, 34, 35, 36, 37) = 0.0129583
- Right Eye FWC(39,40,41,42,43,44,45,46,47) = 0.0173181
- Lip FWD(48,49,50,51,52,53,54,55) = 0.0319725

In the experiment, similarity amongst the similar AROIs of dissimilar expressions differs based on their sizes. During the choice of AROI, an optimized technique is applied so that those active AROIs should be chosen where maximum similarity is matched with the sizes of the regions. A searching method is applied so that the Optimized AROIs are found
directly. In doing this, the hash distance (Lopes et al., 2017) algorithm is applied. There is a well-established rule concerning hash distance. There is a thumb rule that if the hash distance is smaller then similarity should be greater. The algorithm-1 is used to estimate the hash distance between two images.

Algorithm 1: Measurement of Hash Distance
Input: 2 images
Output: Measured Hash Distance between 2 images
a. Find AROIs (Nose-tip, Left-eye, Right-eye and Lip regions)
b. Choose Central-point (Nose-tip)
c. Resize images to 120*80
d. Convert image from RGB to gray.

e. Compute Mean value (MV) from gray image

f. Set POI = 1; if MV >= average, otherwise POI = 0.

g. Output will be the Hamming Distance based on a new gray value (0 or 1).

The following symbols are introduced to illustrate the optimized searching method. Let $I = \{1, 2, 3, 4, \ldots, 7\}$ be the no of subjects in (CK + & JAFFE) image data sets. Here $[J = \{1, 2, 3, 4, 5\}]$ are five expressions. (i.e., 1 if for Normal; 2 is for Happy; 3 is for Fear; 4 is for Surprise; and 5 is for Disgust). Set $E_{ij}, i \in I, j \in J$. This implies that the ith subject with jth expression.

Here $[C = \{\text{Left-eye, Right-eye, Lip regions}\}]$ denotes the classes of AROIs. $L_c$ is equal to $\{1,2,\ldots, \text{bound}_c\}$ where $c \in C$ which is taken as radius of AROIs. Its size is $c$. After
measuring size we estimate the boundary as named bound$. There are four directions $D = \{1, 2, 3, 4\}$ identified as 1 is on left, 2 is on right, 3 is on top, and 4 is on bottom. Here $\text{Disc}_{ij}^{cd}$ $\text{Disc}_{ij}^{cd}$ as equal to the distance amongst the midpoint of $c$ on the direction $D$ over $E_{ij}$. For simplification of $\text{Disc}_{ij}^{cd}$ we took $c$ as left eye distance as shown in Fig. 10. Here $c$ (left eye distance) is the distance amongst the midpoint of left eye and around POIs. The $\text{Disc}_{ij}^{2}$, $\text{Disc}_{ij}^{3}$, and $\text{Disc}_{ij}^{4}$ are the distances amongst the midpoint of left eye’s left-edge, right-edge, top-edge and, bottom-edge, respectively. Here $c$ (right eye distance) is the distance amongst the midpoint of right eye and around POIs. The $\text{Disc}_{ij}^{2}$, $\text{Disc}_{ij}^{3}$ and $\text{Disc}_{ij}^{4}$ are the distances amongst the midpoint of right eye’s left-edge, right-edge, top-edge, and bottom-edge, respectively. In case of mouth it is measured by $\text{Disc}_{ij}^{cd}$ $d \in D$ and the
distances amongst the midpoint of mouth and its 4 edges are measured by $E_{ij}$. The bound $c$ is measured by Eq. (2).

$$\text{bound}_c = \min_{i,j,d}(\text{Disc}_{ij}^{cd}), i \in I, j \in J, d \in D. \quad (5)$$

Let $R_{ij}^{cl}$, $c \in C, l \in L_c$ be the region size which is equal to $l$. Here $\text{Hash}_{ij}^{cl}$, $c \in C, l \in L_c$ is the summation of distance amongst all $c$ regions of subject $i$ with a size of $l$. It is measured by Eq. (3).

$$\text{Hash}_{ij}^{cl} = \sum_{x \neq y} \text{Algorithm1}(R_{ix}^{cl}, R_{jy}^{cl}). \quad (6)$$
Here $x$ & $y \in J$. In order to match the $c$ regions with $i$ and size is $l$, we incorporated the following equation no. 7.

$$\text{Sim}_i^c = \frac{1}{\text{Hash}_i^c + 1}, c \in C, l \in L_c.$$  

(7)

Let $\text{Sim}_i^c$ be the average of $\text{Sim}_i^c$ among the 4 subjects. Where $\text{Sim}_i^c$ $\text{Sim}_l^c$ is the matching among $c$ regions with size is equal to $l$. In lieu of individually $c \in C$, we took the size $l$ to have maximum $\text{Sim}_i^c$ in our experiment.

We have incorporated a searching method in detail is shown in Algorithm 2. During our experiment, we took 35 images from the CK + and 10 JAFFE databases, i.e., $E_{ij}, i \in I, j \in J$. To get proper output, the image sizes are kept the same ($640 \times 800$) for all kinds of images.
Algorithm 2: Searching for the OAROI

Input: 45 (35 + 10) images selected from CK++ and JAFFE database
Output: the sizes of OAROI

a. for c in C
b. Compute bound_i.
c. for i in I
d. for l in L
e. Compute Hash^c_i
f. Compute Sim^c_i

g. Compute the MV of Sim^c_i amongst 7 subjects and acquire Sim^c_i.
h. In drawing the picture of Sim^c_i, let l be start value of X-axis and Sim^c_i be value of Y-axis.
i. Catch the size l from max value of Sim^c_i and record the output.

Based on the Algorithm 2 the required picture will be following the step h. This is shown in Fig. 11. The sizes for the right-eye, left-eye, and lip regions are set to 160 × 120, 160 × 120, and 180 × 80 pixels. We found that the obtained size is half of the edge length of the OAROIs. As an outcome, the edge sizes of 3 kinds of OAROIs have been set to 140 × 80, 140 × 80, and 160 × 60 pixels, respectively. The OAROIs are shown in Fig. 9. It is found almost all active POIs are around the nose regions. For this reason, the nose region is kept aside of our experiment.

Parallel face and facial AROI detection

One of the emergent objectives of this research is to speed up the process of facial expression with improved recognition accuracy. As stated earlier (shown in Fig. 1) the whole process is into three phases: Data processing, training, and testing of classified data. The data processing phase is executed in sequential order. In the first phase, images are recognized and then the following are carried out namely-pre-processing, POI detection, head-pose adjustment using rotation and ASHM method, image segmentation and AROI detection. Lastly, OAROI is identified based on our proposed searching method. The usage of OAROIs rather than the entire face helps us reduce data processing time and speed up processing of the first phase. The last two phases run in parallel since each ORAOI (namely the left-eye, right-eye, and lip region) is identified separately in parallel. This technique reduces the execution time for these two phases by 64% (approx) as shown in the experimental results. As per experimental calculation, only 1% of the total execution time is needed to complete the first phase and the last two phases took 99% (approx) of the total execution time. In applying parallelism, the system is three times faster. In calculating the required time in parallel processing we applied the famous formula of Amdahl’s Law (Gustafson, 1988) as mentioned below through Eq. (8).

\[
\text{Speedup} = \frac{1}{\left(1 - P\right) + P/N}.
\]  

(8)

Where P is the part that runs in parallel that is 0.99 (in our system), and N is the number of processors i.e., 3. We use only three processors for simplification, as our system has 3
OAROIs and each one of them runs separately. Therefore, our parallel model is 2.94% faster, which is almost three times faster. We incorporated a Decision-Tree-Level Synthesis scheme that helps us to gain enhanced accuracy in recognition.

**Classification by decision-tree-level synthesis**

An emergent facial expression recognition technique demands a very high accuracy rate for recognition. In our study, the AROI-based CNN method is incorporated to extract features of the FER and their classification. Three OAROIs are chosen namely left-eye, right-eye, and lip regions. These three OAROIs are trained and classified in parallel using CNN. The final classification result is achieved which is remarkably higher by using the decision tree level synthesis technique. The schematic diagram (Fig. 12) depicts the applied technique. The sizes of the optimized active regions (i.e., the inputs of CNN) are dissimilar for each images. To make the sizes the same, we completed the following actions.

a. Optimized active regions of interest are resized to 120*80 pixels.

b. The CNN-based convolution layers and its connected sub-layers were applied to input images to extract features-weight and features.

c. Lastly, the execution is completed based on fully connected layer.

The feature weights have been applied to the fully-connected-layers to achieve the desired expressions. The One-hot (1) code is applied to detect desired expression. In our
In the last phase, the final expression recognition result is achieved based on the decision-level synthesis method. This decision-level synthesis method is carried out based on the following:

\[
\text{Final Output} = \begin{cases} 
  j, & \text{if } 2 \text{ or more CNNs classify the expression} \\
  0, & \text{otherwise}
\end{cases}
\]

where \( j \) for all \( j \in J \). Here CNN result over left_eye, right_eye & lip regions, otherwise 0

In order to classify any expression we need two or more classifiers as stated in the above expression. Here a class is indicated by \( j \) for all \( j \in J \). The final outcome will be measured based on the \( j \) only. On the other hand, it is revealed that the desired classification is
completed based on 3 CNNs. We considered 5 expressions (namely-Normal, Smiley, Fear, Surprise, and Disgust) in our experiment. If two classified expressions are “Normal” then the synthesis result will be “Normal”. If the results of the three CNNs are different then we have to choose the CNN result based on the highest accuracy. This has been demonstrated in the “Experimental Section”. Figure 12 shows the schematic flow diagram of the feature weight that we already calculated earlier and have shown mathematically through the decision-level-synthesis method as illustrated above.

EXPERIMENTAL EVALUATION

In these five expressions, five expressions are addressed. Based on the segmented images, an optimized AROI is introduced. The face areas/regions explored are the nose-tip, left-eye, right-eye, and lip regions. These four regions are the selected AROIs. The dimension of an active region is directly correlated with facial expression identification. An optimized searching algorithm is introduced to search for the appropriate dimension of active regions. An AROI of an appropriate size is called an OAROI. The eight divided segments are processed in parallel. A decision-tree-level-synthesis framework is built to classify FER. Among the 4 OAROIs, 3 (left-eye, right-eye, and lips) regions are chosen for our experiment. They are trained using CNN. The final classified expressions are achieved based on gained results. Our experiment is conducted based on the Extended Cohn-Kanade(CK+) (Kanade, Cohn & Tian, 2000; Lucey et al., 2010) and the JAFFE image sets which are freely available at http://web.mit.edu/emeyers/www/face_databases.html.

By applying parallelism the processing time has increased three times faster as compared to that of a sequential method. Since each region is processed individually, the results of these regions verify the accuracy of each other. This result confirms that the proposed technique achieves much-improved accuracy compared to earlier works (Salmam, Madani & Kissi, 2016; Mollahosseini, Chan & Mahoor, 2016; AlBdairi, Zhu & Mohammed, 2020; Almeida et al., 2020; Rajan et al., 2020).

During performance evaluation, we carried out both quantitative and qualitative performances. In our experiment, we used two datasets, CK+ and JAFFE, and their experimental results in expression recognition are shown in Tables 1 and 2.

In the experiment, the nose-tip region is the central point of the AROI. So, the detected value for a normal expression was kept fixed for the corresponding image set for all types of expression (Normal, Happiness, Fear, Surprise, and Disgust).

Used databases

The CK+ i.e., Extended Cohn-Kanade database contains 327 male and female images, and each of them has seven types of expression. Out of them, we use five (i.e., Normal, Happiness, Fear, Surprise, and Disgust) types in our experiment. To get the proper expression as a sample image, we select one from the last four frames in this experiment. The JAFFE image database comprises 213 facial expressions from 10 females and each of them has seven types of expression. However, we select five, i.e., Normal, Happiness, Fear, Surprise, and Disgust. In the JAFFE database, it is found that for each expression (subject),
Table 1  Performance detection accuracy of AROI with normal expression based on CK+ images.

| SL# | Images-set with normal expression | Nose-tip region (A) detection (%) | Left eye region (B) detection (%) | Right eye region (C) detection (%) | Lip region (D) detection (%) |
|-----|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------|
| 1   | Image 72                          | 91.1019109                       | 91.2735824                       | 92.46843956                      | 91.19301281                 |
| 2   | Image 73                          | 92.8138211                       | 90.7120214                       | 94.20602842                      | 92.90663492                 |
| 3   | Image 74                          | 93.6106171                       | 93.7688752                       | 95.01477636                      | 93.70422772                 |
| 4   | Image 75                          | 93.6454519                       | 90.1541641                       | 95.05013368                      | 93.73909735                 |
| 5   | Image 76                          | 89.2206315                       | 89.0417528                       | 90.55894097                      | 89.30985213                 |
| 6   | Image 77                          | 92.0991071                       | 92.1111191                       | 93.48059371                      | 92.19120621                 |
| 7   | Image 78                          | 91.1621711                       | 91.1922011                       | 92.52960367                      | 91.25333327                 |
| 8   | Image 80                          | 88.1097217                       | 89.3309417                       | 89.4316753                       | 88.19783142                 |
| 9   | Image 81                          | 92.9349421                       | 91.8238321                       | 94.32896623                      | 93.02787704                 |
| 10  | Image 82                          | 89.1865975                       | 90.0020121                       | 90.52439646                      | 89.275841                   |
| 11  | Image 85                          | 90.6016111                       | 90.3113211                       | 91.96063527                      | 90.69221271                 |
| 12  | Image 87                          | 93.2652721                       | 93.1934003                       | 94.66425118                      | 93.35853737                 |
| 13  | Image 88                          | 91.2229318                       | 90.4141211                       | 92.59127578                      | 91.31415473                 |
| 14  | Image 91                          | 92.7427501                       | 93.3761829                       | 94.13389135                      | 92.83549285                 |
| 15  | Image 92                          | 89.8216319                       | 88.2253372                       | 91.16895638                      | 89.91145353                 |
| 16  | Image 95                          | 92.4109186                       | 90.2525962                       | 93.79708238                      | 92.50332952                 |
| 17  | Image 96                          | 92.2938016                       | 91.2046135                       | 93.67820862                      | 92.3860954                  |
| 18  | Image 100                         | 91.3924011                       | 90.3498596                       | 92.76328712                      | 91.4837935                  |
| 19  | Image 102                         | 93.2948016                       | 91.2046135                       | 94.6922362                       | 93.3880964                  |
| 20  | Image 106                         | 91.0923013                       | 90.3653753                       | 92.45868582                      | 91.1839396                  |
| 21  | Image 107                         | 90.2613712                       | 89.5112218                       | 91.61529177                      | 90.35163255                 |
| 22  | Image 109                         | 91.3004092                       | 90.2636735                       | 92.66991534                      | 91.39170961                 |
| 23  | Image 110                         | 93.2696765                       | 93.9670732                       | 94.66872165                      | 93.36294618                 |
| 24  | Image 111                         | 90.2222321                       | 90.8094187                       | 91.57566558                      | 90.3124533                  |
| 25  | Image 113                         | 90.1131231                       | 90.2914012                       | 91.46481995                      | 90.20326262                 |
| 26  | Image 114                         | 93.2696765                       | 92.4050127                       | 94.66872165                      | 93.36294618                 |
| 27  | Image 115                         | 92.1046126                       | 92.1932011                       | 93.48618179                      | 92.19671721                 |
| 28  | Image 117                         | 92.4140217                       | 91.3894982                       | 93.80023203                      | 92.50643572                 |
| 29  | Image 119                         | 91.9839921                       | 89.1932041                       | 93.36375198                      | 92.07597609                 |
| 30  | Image 122                         | 92.9339411                       | 90.2931029                       | 94.32795022                      | 93.02687504                 |
| 31  | Image 127                         | 93.1661731                       | 90.3643741                       | 94.5636657                       | 93.25933927                 |
| 32  | Image 128                         | 92.8139212                       | 89.2905013                       | 94.20613002                      | 92.90673512                 |
| 33  | Image 130                         | 91.2766855                       | 92.5040116                       | 92.64583578                      | 91.36796219                 |
| 34  | Image 132                         | 90.3653751                       | 88.9993104                       | 91.72085573                      | 90.45574048                 |
| 35  | Image 138                         | 93.2389458                       | 92.236415                        | 94.63752999                      | 93.33218475                 |
| **Average** | | **91.73593**                       | **90.91469816**                    | **93.11196895**                     | **91.82766593** |
Table 2 Performance detection accuracy of AROI with normal expression based on JAFFE images.

| SL# | Images-set with normal expression | Nose-tip (A) detection (%) | Left eye (B) detection (%) | Right eye (C) detection (%) | Lip (D) detection (%) |
|-----|-----------------------------------|-----------------------------|-----------------------------|-----------------------------|----------------------|
| 1   | Image 1                           | 86.490246                   | 86.70647162                 | 89.21468875                 | 90.413856            |
| 2   | Image 2                           | 88.097834                   | 88.31087589                 | 90.87291577                 | 90.81164             |
| 3   | Image 3                           | 88.846074                   | 89.06818919                 | 91.64472533                 | 90.345665            |
| 4   | Image 4                           | 89.818786                   | 90.04333297                 | 92.64807776                 | 90.3528835           |
| 5   | Image 5                           | 83.78361                    | 83.99306903                 | 86.42279372                 | 84.7715694           |
| 6   | Image 6                           | 88.366674                   | 88.58759069                 | 91.15022423                 | 89.9210442           |
| 7   | Image 7                           | 87.486834                   | 87.70555109                 | 90.24266927                 | 85.457184            |
| 8   | Image 8                           | 82.740398                   | 82.947249                   | 85.34672054                 | 86.767776            |
| 9   | Image 9                           | 89.151574                   | 89.37445294                 | 91.95984858                 | 89.173006            |
| 10  | Image 10                          | 84.69165                    | 84.90337913                 | 87.35943698                 | 89.17004             |
|     | Average                           | 86.947368                   | 87.16473642                 | 89.68621009                 | 88.71846641         |

In this study, we took the images detectable naturally for our experimental work. In deep learning techniques, huge training data sets are needed to achieve better performance. The total volume of available data from both databases was not enough for CNN based training under the deep learning method. To yield more data, firstly the original image is flipped to horizontal. Secondly, histogram equalization is conducted on both flipped and original images. Lastly, they have been converted to gray and the pose estimation algorithm (PEA) is also applied. By applying the above scheme the total volume of data has increased by 8 times. Moreover, we applied the segmentation technique that has divided each image into four classes. In this way, the total volume of images became 32 times that of the original one. The OAROIs of all the images are saved in the info-image-data-mask database for training and testing. In addition, we also took the image sets separately as samples of two databases and trained them to classify them based on Happy & Routray (2015) method. Our experimental result and the time of execution are better as compared to other researchers.

Tables 1 and 2 illustrate the performance of execution for normal expressions of two databases CK+ and JAFFE. Table 3 shows the average execution accuracy of 5 types of expression based on CK+ images. The last row of each table shows the average accuracy of each OAROI. Table 3 illustrates the performance of recognition for a selected image set of 35 images taken from the CK+ image database and their execution time in milliseconds in the sequential method. Table 4 demonstrates performance in recognition based on the CK+ image measured in millisecond. Table 5 shows the performance of recognition and their execution time in milliseconds by applying parallelism. The accuracy exceeds 91% in all cases and the average accuracy of the whole classification process reaches 96.23%.

Table 6 illustrates the average execution accuracy of Normal, Happiness, Fear, Surprise, and Disgust for JAFFE images and the accuracy of the proposed synthesis method (our model) in the process of OAROI detection. The last row of each table shows the average
Table 3 Average performance accuracy in detection (%) of facial expression of AROI of region (A, B, C & D) based on CK+ images.

| Facial expression of CK+ | Nose-tip region (A) detection (%) | Left eye region (B) detection (%) | Right eye region (C) detection (%) | LIP region (D) detection (%) | Proposed OAROI detection (%) |
|-------------------------|------------------------------------|----------------------------------|-----------------------------------|-------------------------------|-----------------------------|
| Normal                  | 91.73593                           | 90.91469816                     | 93.11196895                      | 91.82766593                  | 95.43976817                 |
| Smiley                  | 91.73593                           | 89.44253175                     | 94.94668755                      | 97.32035474                  | 97.32035474                 |
| Panic                   | 91.73593                           | 92.10287372                     | 93.21196895                      | 96.10153999                  | 96.10153999                 |
| Surprise                | 91.73593                           | 92.19460965                     | 95.86404685                      | 98.26064802                  | 98.26064802                 |
| Disgust                 | 91.73593                           | 89.99294733                     | 91.36898628                      | 94.02932825                  | 94.02932825                 |

The accuracy exceeds 88% in all cases and the average accuracy of the whole classification process reaches 96.23%.

Table 7 illustrates the performance of recognition for the selected image set of 10 images taken from the JAFFE image database and their execution time in milliseconds in the sequential method has been shown. In Table 8, with the same images, the performance of recognition and their execution time in milliseconds in the parallel method has been shown.

Moreover, Table 8 compares the ET in milliseconds for our proposed synthesis method (parallel execution) with the sequential one for the JAFFE image set. It shows the ET for our proposed parallel method reduces to 36.536% than that of a sequential method.

**Used CNN structure and tenfold validation**

The proposed CNN structure is shown in Fig. 13. In Fig. 12 it is shown that the system receives gray images as input with a fixed size of 120*80 pixels. Three CNN layers are offered along with AROIs. Here for each of them is interlinked via a sub-sampling layer. A relationship is built up by kernel sizes in CNN layers. The measurement of sub-sampling layers are 5 × 5 and 2 × 2, correspondingly. Every sub-sampling layer should be multiple of 2. In Fig. 13, it is being seen that the size after 1st convolution became 120*80*32. A similar method is applied to the rest of the following layers. There are 1,024 neurons in total. Each of two are fully connected. In this way, CNN helps us to gain better results. The performance is achieved based on the voting score of five dissimilar classified expressions. The output (shown in Fig. 13) is determined based on the number of dissimilar images. The proposed CNN structure (as shown in Fig. 13) illustrates how convolution works with the sub-sampling layer and fully connected layer.

**Experiments on ten-fold cross-validation**

We applied ten-fold cross-validation to calculate the performance of the proposed synthesis method and it is compared with its original CNN. Tables 9 and 10, Figs. 14 and 15 show the performance accuracy of each AROI. This performance is achieved based the proposed ten-fold cross-validation that we applied on two image datasets. Here we applied our proposed synthesis method. By using the CK+ image set the performance based on three CNNs (on left-eye, right-eye, and lip region) are 92.19%, 93.11%, and 95.86% respectively, whereas by using the JAFFE image set the performances are gained 90.207%, 91.042%, and 92.120% respectively. Our achievement is excellent as compared to other
Table 4  Performance matrix of selected CK+ image set based on execution time (ET) in milliseconds (MS).

| Image sets | Normal face images ET in MS in SM | Smiley face images ET in MS in SM | Panic face images ET in MS in SM | Surprise face images ET in MS in SM | Disgust face images ET in MS in SM |
|------------|-----------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|-----------------------------------|
|            | Right eye ET | Left eye ET | Lip ET | Right eye ET | Left eye ET | Lip ET | Right eye ET | Left eye ET | Lip ET | Right eye ET | Left eye ET | Lip ET |
| Total ET   | 390.647      | 384.45      | 421.996 | 386.74      | 380.6       | 451.54 | 396.45      | 390.23      | 426.94 | 402.37      | 395.982      | 451.54 |
| Average ET | 11.1613      | 10.984      | 12.057  | 11.05       | 10.87       | 12.901 | 11.327      | 11.149      | 12.198 | 11.496      | 11.3138      | 12.901 |

Hossain and Assiri (2022), PeerJ Comput. Sci., DOI 10.7717/peerj-cs.894
Table 5  Comparing the Execution Time (ET) in milliseconds of parallel method and sequential method for CK+ image set.

|                       | Parallel ET for normal face images | Sequential ET for normal face Images | 34.2 |
|-----------------------|------------------------------------|-------------------------------------|------|
|                       | Parallel ET for smiley face images | 12.9                                | 34.8 |
|                       | Parallel ET for panic face images  | 12.19                               | 34.67|
|                       | Parallel ET for surprise face images| 12.9                                | 34.7 |
|                       | Parallel ET for disgust face images | 11.45                               | 34.38|
| Average ET for parallel Method | 12.299                             | Average ET for sequential method    | 34.158|

Table 6  Shows the execution time (ET) in milliseconds for all OAROI using JAFFE images considering the five targeted expressions.

| Facial expression of JAFFE | Nose-tip region (A) detection (%) | Left eye region (B) detection (%) | Right eye region (C) detection (%) | LIP region (D) detection (%) | Proposed OAROI detection (%) |
|----------------------------|------------------------------------|-----------------------------------|-----------------------------------|-------------------------------|-------------------------------|
| Normal                     | 86.947368                          | 87.16473642                      | 89.68621009                      | 88.71846641                  | 91.92836534                  |
| Smiley                     | 86.947368                          | 89.33842062                      | 90.03399956                      | 91.16431535                  | 93.44342323                  |
| Panic                      | 86.947368                          | 88.34722062                      | 88.79934694                      | 88.2950522                   | 91.01933061                  |
| Surprise                   | 86.947368                          | 90.2078943                       | 91.04258903                      | 92.1207364                   | 94.42375481                  |
| Disgust                    | 86.947368                          | 85.29536801                      | 85.51263116                      | 87.0778905                   | 89.25473378                  |

researchers (Zhang & Wang, 2020; Hossain & Sanyal, 2012b; Moghaddam, Akbarizadeh & Kaabi, 2019; Cheon & Kim, 2008; AlBdairi, Zhu & Mohammed, 2020). Moreover, our proposed synthesis method gained an excellent rate of higher accuracy, i.e., 94.499%, 95.439%, and 98.26% respectively in the CK+ set and 92.463%, 93.318%, and 94.423% respectively in the JAFFE image set. In contrast, the researchers (Happy & Routray, 2015) have conducted their experiments based on fused data from the CK+ and JAFFE databases, and their accuracies were 89.64% and 85.06% respectively, which are remarkably less than that of our results. Liu et al. (2016) achieved accuracies of 92.40% and 63.40% respectively. Zhong et al. (2012) used CSPL method and they used MNI and CK+ dataset with the accuracy of 73.53% and 89.89% respectively. Lee, Plataniotis & Ro (2014) applied SRC in their dataset MNI,JAFFE and CK+ and they achieved 93.81%,87.85% and 94.70% respectively. Mollahosseini, Chan & Mahoor (2016) achieved accuracies of 93.20% and 77.90% respectively. In their experimentation, they tried to locate active patches around the nose, mouth, and eye regions. Our achieved result is a robust one as compared to other researchers. Moreover, we can gain 2.8% higher accuracy by introducing a decision-level synthesis method. This confirms the effectiveness and robustness of our proposed method.

Discussion on the experimental results

The experiment was completed using OS-Window 7, Language- Python 3.6, Intel Core (TM) i7 with a CPU processing speed of 3.40 GHz with RAM size is 6 GB. It is a Laptop with 4 physical cores. The maximum amount of experimental time is spent for dataset training. We know that training a dataset in a deep learning method needs a longer time as compared to other techniques. Because training time is correlated with several
Table 7 Performance matrix of selected JAFFE image set based on execution time (ET).

| Image sets | Normal face images ET in MS in SM | Smiley face images ET in MS in SM | Panic face images ET in MS in SM | Surprise face images ET in MS in SM | Disgust face images ET in MS in SM |
|------------|----------------------------------|----------------------------------|----------------------------------|------------------------------------|-----------------------------------|
|            | Right eye ET                     | Left eye ET                      | Lip ET                           | Right eye ET                      | Left eye ET                      | Lip ET | Right eye ET | Left eye ET | Lip ET | Right eye ET | Left eye ET | Lip ET | Right eye ET | Left eye ET | Lip ET |
| Total ET   | 103.8401                         | 102.9574                         | 115.4043                         | 104.8785                          | 103.98697                        | 116.55834 | 106.9553 | 106.04612 | 123.4826 | 93.43609 | 92.66166 | 109.63409 |
| Average ET | 10.38401                         | 10.29574                         | 11.54043                         | 10.48785                          | 10.398697                        | 11.65834 | 10.69533 | 10.604612 | 12.34826 | 9.345609 | 9.266166 | 10.963409 |
Table 8  Comparing the Execution Time (ET) in milliseconds of parallel method and sequential method for JAFFE image set.

|                        | Parallel ET for normal face images | Sequential ET for normal face images |
|------------------------|-----------------------------------|---------------------------------------|
| Normal face images     | 11.54                             | 32.22                                 |
| Smiley face images     | 12.34                             | 32.82                                 |
| Panic face images      | 11.65                             | 32.54                                 |
| Surprise face images   | 12.34                             | 33.64                                 |
| Disgust face images    | 10.96                             | 29.57                                 |
| Average ET for parallel method | 11.75                             | Average ET for sequential method      |
|                        |                                   | 32.16                                 |

Figure 13  The structure of proposed CNN with three convolution, three sub-sampling layers and two fully connected layers.

Full-size DOI: 10.7717/peerjcs.894/fig-13

Table 9  Performance detection accuracy of AROI by CNN and proposed synthesis method based on CK+ image set.

| Method applied       | Left eye region (B) detection (%) | Right eye region (C) detection (%) | LIP region (D) detection (%) |
|----------------------|-----------------------------------|-----------------------------------|-------------------------------|
| AROI CNN             | 92.19460965                       | 93.11196895                       | 95.86404685                  |
| Proposed synthesis method | 94.49947489                      | 95.43976817                       | 98.26064802                  |

factors, like the volume of the data, the model of a structure, calculating the power, and so on.

As per the information shown in Table 11, it is clear that the CK+ database takes more time than the JAFFE database because there are more images in the CK+ database with complex conditions. In addition, the same reason is applicable for the training time for the Info-image-data-mask database as compared with the independent databases. Indeed, the Info-image-data-mask database includes both images of two databases. Moreover, ten-fold cross-validation is also applied in our experiment to increase the volume of data sets. As a result of which we have acquired a huge volume of data which is necessary for deep learning methods for better result.

We did not use other dataset like the NVIE (Nada & Heyam, 2020; Cheon & Kim, 2008; Wang & He, 2013; Shen, Wang & Liu, 2013; Rajan et al., 2020) or fused database because of two reasons. Firstly, most of the NVIE image database contains thermal
Table 10  Performance detection accuracy of AROI by CNN & proposed synthesis method based on JAFFE image set.

| Method applied          | Left eye region (B) detection (%) | Right eye region (C) detection (%) | LIP region (D) detection (%) |
|-------------------------|-----------------------------------|-----------------------------------|------------------------------|
| AROI CNN                | 90.2078943                        | 91.04258903                       | 92.1207364                   |
| Proposed synthesis method | 92.46309166                       | 93.31865376                       | 94.42375481                  |

Figure 14  Performance detection accuracy chart of each AROI by CNN and proposed synthesis method result in tenfold cross validation based on CK+ image set.

Figure 15  Performance detection accuracy chart of each AROI by CNN and proposed synthesis method result in tenfold cross validation based on JAFFE image set.
Table 11  Average training time required for two databases and the info-image-data-mask.

| Database                        | Required time |
|---------------------------------|---------------|
| CK+                             | 8 min 22 s    |
| JAFFE                           | 2 min 3 s     |
| Info-image-data-mask            | 40 min 7 s    |

Table 12  Performance result analysis of different researchers by applying different method based on NVIE database.

| Researchers                  | Data set       | Used method                              | Overall accuracy |
|------------------------------|----------------|------------------------------------------|------------------|
| Cheon & Kim (2008)           | NVIE           | AAM (Active appearance model) integrated + AAM | 66.24%           |
|                              |                | DAF integrated + DAF                     | 71.72%           |
| Wang & He (2013)             | USTC-NVIE      | AAM + Thermal statistical parameters KNN | 79.16%           |
|                              |                | Bayesian networks (BNs)                  | 91.52%           |
| Shen, Wang & Liu (2013)      | USTC-NVIE      | KNN                                      | 63%              |
|                              |                |                                          | 85.51%           |

images. Secondly, it is not very reliable as it is found from the performance analysis results of researchers (Cheon & Kim, 2008; Wang & He, 2013; Shen, Wang & Liu, 2013) that results gained from NVIE dataset differ depending on the method used. Their result analysis have been shown with the Table 12. The Table 12 illustrates that the achieved results varies depending on applied methods based on same database i.e., NVIE. Whereas the JAFFE and CK+ are very authentic and filtered. So we have chosen these two image sets in our application. Moreover, we concentrated on the required time for execution, so, for these two reasons, we applied the CK+ and JAFFE datasets.

Due to a diverse experiment setting, it is not easy to make an impartial evaluation of the proficiency of dissimilar methods. AlBdairi, Zhu & Mohammed (2020) have proposed that their CNN-based deep learning method has achieved 96.9% accuracy in the verification of the human face. Rajan et al. (2020) have proposed their CNN and LSTM-based methods where they have used MMI, SFEW, and their dataset. They achieved the accuracy of 81.60%, 56.68%, and 95.21% respectively. Hossain, Samanta & Sanyal (2012c) trained their data for CNN and it took about 20 min. They used the CK+ database in their experiment. Their hardware configuration was an Intel Core i7 CPU of 3.4 GHz. Liu et al. (2016) trained their data by using eight-fold cross-validation. Their hardware configuration was 6 core with a processing speed of 2.4 GHz. Their experiment took about 8 days to complete. The comparison of performance and its result analysis with different researchers are shown in a tabular form. It is shown by the following Table 13.
| Researchers                | Method used                                      | Accuracy based on used datasets |
|----------------------------|-------------------------------------------------|---------------------------------|
|                            |                                                 | JAFFE | CK+  | MNI  | SFEW | OWN  |
| **Happy & Routray (2015)** | SFP                                             | 85.06% | 89.64% |      |      |      |
| **Liu et al. (2016)**     | AUDN, Eight-fold cross-validation                | 63.40% | 92.40% |      |      |      |
| **Zhong et al. (2012)**   | CSPL                                            |       | 89.89% | 73.53% |      |      |
| **Lee, Plataniotis & Ro (2014)** | SRC                                              | 87.85% | 94.70% | 93.81% |      |      |
| **Kanade, Cohn & Tian (2000)** | DNN                                              |       | 93.20% | 77.90% |      |      |
| **Shen, Wang & Liu (2013)** | CNN and LSTM-based methods, ten-fold cross-validation |     |       |      |      |      |
| Our applied method         | CNN by applying on                               |       |      |      |      |      |
|                            | 1. left-eye                                      | 90.207% |      |      |      |
|                            | 2. right-eye                                     | 91.042% |      |      |      |
|                            | 3. lip region                                    | 92.120% |      |      |      |
| Our proposed method        | Decision-level synthesis method and ten-fold cross-validation | 94.499% | 95.439% |      |      |
|                            | by applying on                                   |       |      |      |      |      |
|                            | 1. left-eye                                      |      | 95.439% |      |      |
|                            | 2. right-eye                                     |      | 98.260% |      |      |
|                            | 3. lip region                                    |      |      |      |      |      |
| **Our Achievement**        | **We gain 2.8% higher accuracy by introducing a decision-level synthesis method and ten-fold cross-validation** |      |      |      |      |      |
CONCLUSIONS

In our investigation, a sophisticated and proficient facial expression recognition scheme is introduced for classifying five different types of expressions. In our experiment, we considered only four active regions namely nose-tip, right-eye, left-eye, and lip regions. A search method called the Optimized active Region of Interest (OAROI) is introduced for faster searching of active regions. Among the four AROIs, three regions (right eye, left eye, and lip) are chosen for experimentation. The required data sets are trained by using AROI-based CNN to classify expressions. A pose estimation algorithm is also introduced. An approach “rotate and correct” based on the exact pose is also incorporated. A decision-level synthesis method is applied to gain higher recognition accuracy. Experimentation on both separate databases and info-image-data-mask databases is voted for to assess the performance of the proposed technique.

To increase the volume of the image set for better recognition accuracy, we proposed a ten-fold-cross validation method where CNN is used. By using the CK+ image set we achieve an accuracy of 92.19%, 93.11%, and 95.86% respectively for the left-eye, right-eye, and lip region, whereas by using the JAFFE dataset we achieve the accuracies of 90.207%, 91.042%, and 92.120% respectively. This achievement is excellent as compared to other researchers (Happy & Routray, 2015; Liu et al., 2016; Mollahosseini, Chan & Mahoor, 2016; AlBdairi, Zhu & Mohammed, 2020; Rajan et al., 2020). Moreover, we introduce a parallel processing technique that has accelerated the system’s performance to a greater extend. The AROIs are processed in parallel, so processing time is almost three times faster than that of a sequential method. By introducing our proposed method we achieved 95.27% in recognition. Our proposed synthesis method helped us to gain excellent accuracy, i.e., 94.499%, 95.439%, and 98.26% respectively in CK+ and 92.463%, 93.318%, and 94.423% respectively in the JAFFE image set. However, we gain 2.8% higher accuracy by introducing a decision-level synthesis method. This confirms the effectiveness and robustness of the synthesis method. In the future, a few other datasets will be used to check our results. In that work, AI will be incorporated along with more than ten-fold cross-validation.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding
This work is supported by the Deanship of Scientific Research, Jazan University, Kingdom of Saudi Arabia. This work is done to complete the funded project that is approved by the Deanship of Scientific Research, Jazan University, KSA. The project Grant No is Waed-20-40. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures
The following grant information was disclosed by the authors: Deanship of Scientific Research, Jazan University, Kingdom of Saudi Arabia.
Competing Interests
The authors declare there are no competing interests.

Author Contributions
• Mohammad Alamgir Hossain conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
• Basem Assiri performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:

The code are available in the Supplemental Files. The CK+ Image Set is available at Access to the JAFFE Images can be requested at: https://zenodo.org/record/3451524/accessrequest.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.894#supplemental-information.

REFERENCES

Alam SS, Jianu R. 2017. Analyzing eye-tracking information in visualization and data space: from where on the screen to what on the screen. The IEEE Transactions on Emerging Topics in Computational Intelligence 23(5):1492–1405 DOI 10.1109/TVCG.2016.2535340.

AlBdairi AJA, Zhu X, Mohammed A. 2020. Identifying ethnics of people through face recognition: a deep cnn approach. Hindawi Scientific Programming 2020:6385281 DOI 10.1155/2020/6385281.

Almeida WR, Andalo FA, Padilha R, Gabriel B, William DR, Torres S, Jacques W, Anderson R. 2020. Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor aware loss function. PLOS ONE 15(9):e0238058 DOI 10.1155/2020/6385281.

An Z, Deng W, Hu J, Zhong Y, Zhao Y. 2019. Adaptive pose alignment for pose-invariant face recognition. IEEE Access 7:14653–14670 DOI 10.1109/ACCESS.2019.2894162.

Boyina SR. 2020. Accurate leukocoria predictor based on deep VGG-net CNN technique. The IET Image Processing 14(10):2241–2248 DOI 10.1049/iet-ipr.2018.6656.

Carletti V, Greco A, Percannella G, Vento M. 2020. Age from faces in the deep learning revolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(9):2113–2132 DOI 10.1109/TPAMI.2019.2910522.

Cheon Y, Kim D. 2008. A natural facial expression recognition using differential-AAM and k-NNS, multimedia, 2008. In: ISM 2008. Tenth IEEE international symposium on 15-17 December 2008. Piscataway: IEEE, DOI 10.1109/ISM.2008.121.
Clark EA, Kessinger J, Duncan SE, Bell MA, Lahne J, Gallagher DL, O’Keefe SF. 2020. The facial action coding system for characterization of human affective response to consumer product-based stimuli: a systematic review. Frontiers in Psychology 11:920 DOI 10.3389/fpsyg.2020.00920.

Daihong J, Hu Y, Lei D, Jin P. 2021. Facial expression recognition based on attention mechanism. Hindawi Scientific Programming 2021:6624251 DOI 10.1155/2021/6624251.

Ding H, Zhou SK, Chellappa R. 2017. Facenet2exprnet: regularizing a deep face recognition net for expression recognition. In: 12th IEEE, ICAFGR, FG. Piscataway: IEEE, 118–126 DOI 10.1109/FG.2017.23.

Fan X, Zhang F, Wang H, Lu X. 2012. The system of face detection based on opencv. In: 24th Chinese control and decision conference (CCDC). 648–519 DOI 10.1109/CCDC.2012.6242980.

Ferrari C, Lisanti G, Berretti S, Del Bimbo A. 2018. Investigating nuisances in DCNN-based face recognition. IEEE Transactions on Image Processing 27(11):5638–5651 DOI 10.1109/TIP.2018.2861359.

Galea C, Farrugia RA. 2018. Matching software-generated sketches to face photographs with a very deep CNN, morphed faces, and transfer learning. IEEE Transactions on Information Forensics and Security 13(6):1421–1431 DOI 10.1109/TIFS.2017.2788002.

Geng T, Tianqi W, Wu C, Yang C. 2019. LP-BNN: ultra-low-latency BNN inference with layer parallelism. Vol. 2160. IEEE DOI 10.1109/ASAP.2019.00-43.

Gustafson JL. 1988. Reevaluating Amdahl’s law. Communications of the ACM 31(5):532–533 DOI 10.1145/42411.42415.

Happy S, Routrav A. 2015. Automatic facial expression recognition using features of salient facial patches. IEEE Transactions on Affective Computing 6(1):1–12 DOI 10.1109/TAFFC.2014.2386334.

Hossain MA, Assiri B. 2020a. An enhanced eye tracking approach using pipeline computation. The Arabian Journal for Science and Engineering 45(4):1–14 DOI 10.1007/s13369-019-04322-7.

Hossain MA, Assiri B. 2020b. Facial emotion verification by infrared image. In: International conference on emerging smart computing and informatics (ESCI-2020). Piscataway: IEEE, DOI 10.1109/ESCI48226.2020.9167616.

Hossain MA, Goutam S. 2012. Object tracking podium on region covariance for recognition and classification. International Journal of Emerging Technologies in Computational and Applied Sciences 2(1):68–73.

Hossain MA, Samanta D, Goutam S. 2012. Eye diseases detection based on covariance. International Journal of Computer Science and Information Technology & Security 2(2):376–379.

Hossain MA, Samanta D, Sanyal G. 2012c. Statistical approach for extraction of panic expression. In: CICN-2012@ IEEE. Piscataway: IEEE, 420–424 DOI 10.1109/CICN.2012.189.
Hossain MA, Samanta D, Sanyal G. 2012d. A novel approach for panic-face extraction based on mutation. In: *IEEE, international conference on advanced communication, control and computing technologies*. Piscataway: IEEE, 473–477 DOI 10.1109/ICACCT.2012.6320825.

Hossain MA, Samanta D, Sanyal G. 2012a. Extraction of panic expression depending on lip detection. In: *2012 international conference on computing sciences ICCS-2012*. Piscataway: IEEE, 137–141 DOI 10.1109/ICCS.2012.35.

Hossain MA, Samanta D, Sanyal G. 2012b. Automated smiley face extraction based on genetic algorithm. In: *SIPM-2012*. DOI 10.5121/csit.2012.35.

Hossain MA, Sanyal G. 2013. Human facial expression tracking and grading by histogram based approach. *Journal of Information Processing* 7(1):44–50.

Hossain MA, Sanyal G. 2012a. Extraction of panic expression from human face based on histogram approach. In: *ICIP 2012* Springer. 411–418 DOI 10.1007/978-3-642-31686-9_48.

Hossain MA, Sanyal G. 2012b. A novel approach to extract region from facial expression based on mutation. *UACE International Journal of Computer Science and Applications* 2(3):15–18 DOI 10.1109/RCAER.2016.7784056.

Hossain MA, Sanyal GA. 2013. A stochastic statistical approach for tracking human activity. *The International Journal of Information Technology, Modeling and Computing* 1(3):33–42 DOI 10.5121/iijitmc.2013.1304.

Hossain MA, Sanyal G. 2014. A novel stochastic tracking approach on human movement analysis. *International Journal of Computers and Applications* 86(18):36–40.

Hossain MA, Sanyal G. 2015. A new tactic to maintain privacy and safety of imagery information. *International Journal of Computer Applications* 110(5):6–12 DOI 10.5120/19310-0764.

Hossain MA, Sanyal G. 2016. Tracking humans based on interest point over span-space in multifarious situations. *International Journal of Software Engineering and Its Applications* 10(9):175–192.

Hossain MA, Zogan H, Sanyal G. 2018. Emotion tracking and grading based on sophisticated statistical approach. *International Journal of Advances in Electronics and Computer Science* 5(3):9–13.

Hu G, Peng X, Yang Y, Hospedales TM, Verbeek J. 2018. Learning deep face representations using small data. *IEEE Transactions on Image Processing* 27(1):293–303 DOI 10.1109/TIP.2017.2756450.

Jae YC, Bumshik L. 2020. Ensemble of deep convolutional neural networks with gabor face representations for face recognition. *IEEE Transactions on Image Processing* 29:3270–3328 DOI 10.1109/TIP.2019.2958404.

Jia S, Wang S, Hu C, Webster PJ, Li Z. 2021. Detection of genuine and posed facial expressions of emotion: databases and methods. *Frontiers in Psychology* 580287 DOI 10.3389/fpsyg.2020.580287.
Jin B, Cruz L, Gonçalves N. 2020. Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis. *IEEE Access* 8:123649–123661 DOI 10.1109/ACCESS.2020.3005687.

Jun K, Min C, Min J, Jinhua S, Jian H. 2018. Face recognition based on CSGF(2D) 2PCANet. *IEEE Access* 6:45153–45165 DOI 10.1109/ACCESS.2018.2865425.

Kanade T, Cohn JF, Tian Y. 2000. Comprehensive database for facial expression analysis. In: 4th IEEE ICA FGR, (FG’00). Piscataway: IEEE, 46–53 DOI 10.1109/AFGR.2000.840611.

Lu J, Wang G, Zhou J. 2017. Simultaneous feature and dictionary learning for image set based face recognition. *IEEE Transactions on Image Processing* 26(8):4042–4054 DOI 10.1109/TIP.2017.2713940.

Lu J, Hu J, Tan YP. 2017. Discriminative deep metric learning for face and kinship verification. *The IEEE Transactions on Image Processing* 26(9):4269–4282 DOI 10.1109/TIP.2017.2717505.

Lajevardi SM, Hussain ZM. 2021. Automatic facial expression recognition: feature extraction and selection. *Signal Image Video Process* 6(1):159–169 DOI 10.1007/s11760-010-0177-5.

Lee SH, Plataniotis KN, Ro YM. 2014. Intra-class variation reduction using training expression images for sparse representation based facial expression recognition. *IEEE Transactions on Affective Computing* 5(3):340–531 DOI 10.1109/TAFFC.2014.2346515.

Liu Y, Cao Y, Li Y, Liu M, Song R, Wang Y, Xu Z, Ma X. 2016. Facial expression recognition with pca and lbp features extracting from active facial patches. In: *IEEE 3rd RCAR, 2016*. Piscataway: IEEE, 368–373 DOI 10.1109/RCAR.2016.7784056.

Liu D, Bellotto N, Yue S. 2020. Deep spiking neural network for video-based disguise face recognition based on dynamic facial movements. *IEEE Transactions on Neural Networks and Learning Systems* 31(6):1843–1855 DOI 10.1109/TNNLS.2019.2927274.

Liu P, Han S, Meng Z, Tong Y. 2014. In: *Facial expression recognition via a boosted deep belief network*. Columbus DOI 10.1109/CVPR.2014.233.

Lopes AT, Aguiar ED, Souza AD, Oliveira-Santos T. 2017. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recognition* 61:610–628 DOI 10.1016/j.patcog.2016.07.026.

Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar , Matthews I. 2010. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: *IEEE CVPRW*. Piscataway: IEEE, 94–101.31 DOI 10.1109/CVPRW.2010.5543262.

Manda B, Bhaskare P, Muthuganapathy R. 2021. A convolutional neural network approach to the classification of engineering models. *IEEE Access* 9:22711–22723 DOI 10.1109/ACCESS.2021.3055826.

Moghaddam AE, Akbarizadeh G, Kaabi H. 2019. Automatic detection and segmentation of blood vessels and pulmonary nodules based on a line tracking method and generalized linear regression model. *Signal, Image and Video Processing* 13(3):457–464 DOI 10.1007/s11548-011-0638-5.
Mohammad AH, Assiri B. 2020. Emotion specific human face authentication based on infrared thermal image. In: ICCIS-2020. Piscataway: IEEE DOI 10.1109/ICCIS49240.2020.9257683.

Mollahosseini A, Chan D, Mahoor MH. 2016. In: Going deeper in facial expression recognition using deep neural networks, WACV. Lake Placid, 1–10 DOI 10.1109/WACV.2016.7477450.

Nada A, Heyam HAB. 2020. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits, MDPI. Sensor 20(19):5523–5539 DOI 10.3390/s20195523.

Rajan S, Chenniappan P, Devaraj S, Madian N. 2020. Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM. The IET Image Processing 14(7):1373–1381 DOI 10.1049/iet-ipr.2019.1188.

Rajeev R, Vishal MP, Rama C. 2019. HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(1):121–135 DOI 10.1109/TPAMI.2017.2781233.

Ran H, Xiang W, Zhenan S, Tieniu T. 2019. Wasserstein CNN: learning invariant features for NIR-VIS face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 41(7):1761–1773 DOI 10.1109/TPAMI.2018.2842770.

Said Y, Barr M, Ahmed HE. 2020. Design of a face recognition system based on Convolutional Neural Network (CNN). Engineering, Technology & Applied Science Research 10(3):5608–5612 DOI 10.1016/j.CAC48633.2019.8996236.

Salmam FZ, Madani A, Kissi M. 2016. Facial expression recognition using decision trees. In: 13th international conference on computer graphics, imaging and visualization (CGiV), 125–130 DOI 10.1109/CGiV.2016.33.

Samadi F, Akbarizadeh G, Kaabi H. 2019. Change detection in SAR images using deep belief network: a new training approach based on morphological images. IET Image Processing 13(12):2255–2264 DOI 10.1049/iet-ipr.2018.6248.

Shen P, Wang S, Liu Z. 2013. Facial expression recognition from infrared thermal videos. In: Intelligent autonomous systems 12. Berlin, Heidelberg: Springer, 323–333 DOI 10.1016/j.imavis.2011.07.002.

Sun A, Li Y, Huang Y, Li O, Lu G. 2018. Facial expression recognition using optimized active regions. Human-centric Computing and Information Sciences 8(33):1–24 DOI 10.1186/s13673-018-0156-3.

Suri A, Vatsa M, Singh R. 2020. A2-LINK: recognizing disguised faces via active learning and adversarial noise based inter-domain knowledge. IEEE Transactions on Biometrics, Behavior, and Identity Science 2(4):326–336 DOI 10.1109/TBIOM.2020.2998912.

Tirandaz Z, Akbarizadeh G. 2016. Unsupervised texture-based SAR image segmentation using spectral regression and Gabor filter bank. Journal of the Indian Society of Remote Sensing 44(2):177–186 DOI 10.1007/s12524-015-0490-0.

Wang S, He S. 2013. Spontaneous facial expression recognition by fusing thermal infrared and visible images. In: Intelligent autonomous systems 12. Berlin, Heidelberg: Springer, 263–272 DOI 10.1109/MLSP.2011.6064564.
Yongbin G, Naixue X, Wenjun Y, Hyo JL. 2019. Learning identity-aware face features across poses based on deep siamese networks. *IEEE Access* 7:105789–105799 DOI 10.1109/ACCESS.2019.2932760.

Zeng N, Zhang H, Song B, Liu W, Li Y, Dobaie AM. 2018. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* 273:643–649 DOI 10.1016/j.neucom.2017.08.043.

Zhang S, Pan X, Cui Y, Zhao X, Liu L. 2019. Learning affective video features for facial expression recognition via hybrid deep learning. *IEEE Access* 7:32297–32304.

Zhang L, Liu J, Zhang B, Zhang D, Zhu C. 2020. Deep cascade model-based face recognition: when deep-layered learning meets small data. *The IEEE Transactions on Image Processing* 29:1016–1029 DOI 10.1109/TIP.2019.2938307.

Zhang F, Wang F. 2020. Exercise fatigue detection algorithm based on video image information extraction. *IEEE Access* 8(11):199696–199709 DOI 10.1109/ACCESS.2020.3023648.

Zhong L, Liu Q, Yang P, Liu B, Huang J, Metaxas D. 2012. Learning active facial patches for expression analysis. In: *CVPR, 16-21 June 2012* DOI 10.1109/CVPR.2012.6247974.