



# Intelligent deep learning based ethnicity recognition and classification using facial images



Gurram Sunitha <sup>a</sup>, K. Geetha <sup>b</sup>, S. Neelakandan <sup>c,\*</sup>, Aditya Kumar Singh Pundir <sup>d</sup>, S. Hemalatha <sup>e</sup>, Vinay Kumar <sup>f</sup>

<sup>a</sup> Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, India

<sup>b</sup> Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, India

<sup>c</sup> Department of CSE, R.M.K Engineering College, India

<sup>d</sup> Department of ECE, Arya College of Engineering and Information Technology, India

<sup>e</sup> Department of Computer Applications, Kongu Engineering college, Erode, India

<sup>f</sup> Department of Computer Engineering and Application, GLA University, Mathura., India

## ARTICLE INFO

### Article history:

Received 16 August 2021

Received in revised form 17 January 2022

Accepted 10 February 2022

Available online 15 February 2022

### Keywords:

Ethnicity recognition

Facial analysis

Deep learning

Facial landmarks

Image classification

Parameter tuning

## ABSTRACT

Recently, computer vision-based face image analysis has sparked considerable interest in a variety of applications such as surveillance, security, biometrics and so on. The goal of the facial analysis was to derive facial soft biometrics such as identification, gender, age, ethnicity, expression and so on. Among these, ethnicity recognition remains a hot study topic, a major aspect of society with profound linkages to a variety of environmental and social concerns. The introduction of machine learning (ML) and deep learning (DL) technologies has proven advantageous for effective ethnicity recognition and classification. In this regard, the IDL-ERCFI technique, which is based on intelligent DL, is designed in this paper. The IDL-ERCFI technique's purpose is to distinguish and classify ethnicity based on facial photos. The IDL-ERCFI technique uses face landmarks to align photos before sending them to the network. Furthermore, the proposed model employs an Exception network as a feature extractor. Because the retrieved features are high-dimensional, the feature reduction procedure employs the principal component analysis (PCA) technique, which is effective in overcoming the "curse of dimensionality." Furthermore, the ethnicity classification procedure is carried out using an optimal kernel extreme learning machine (KELM), with parameter tuning of the KELM model carried out using the glow worm swarm optimization (GSO) technique. A complete experimental analysis is carried out to demonstrate the superiority of the IDL-ERCFI technique over the other techniques.

© 2022 Published by Elsevier B.V.

## 1. Introduction

Facial analysis has recently gained popularity in the realm of computer vision. The human face has features that define a person's gender, identity, age, ethnicity and emotion [1]. The automated recognition of ethnic and racial groups has enormous ramifications in the fields of human computer interface (HCI), medicine, surveillance systems, visagisme, biometrics and so on. Race targeted pharmacogenomics and race-based medicine, for example, encourage the use of race data in the diagnosis and treatment of diverse disease problems that have specific reactions from the organisms of different races. Soft biometrics data (including race) could be embedded in video surveillance systems to improve the accuracy of person recognition and limit the possible matches [2]. Racial data could potentially be used in HCI and target advertising systems to provide users with ethnically appropriate facilities,

hence avoiding the possibility of being hurt by social limitations. Not only can the HCI system identify face traits in head-tilted conditions, but it can also reliably distinguish hand motions throughout the image. It is also resistant to crowded backgrounds and various clothing situations, efficiently obtaining hand regions and recognising hand motions with a trained neural network system.

Despite this, ethnicity detection, specifically the ability of a method for determining a person's belonging to one of the  $E = e_1, \dots, e_E$  ethnicity groups based on facial appearances observation such as morphology, skin colour and another explicit pattern, has not received equal consideration from researchers [3]. The interest in ethnicity detection is undoubtedly growing, given that novel datasets and methods [4,5] have recently been presented for improving the accuracy of real-time applications currently achieving an efficiency biased with ethnicity (gender classification, age estimation, face recognition and detection) or to provide a definite force to the application in forensics (ethnicity-based subjects recognition for the safety of public). Nonetheless, the researchers of a recent exhaustive study [6] state that the development of this

\* Corresponding author.

E-mail addresses: [gurramsunitha@gmail.com](mailto:gurramsunitha@gmail.com) (G. Sunitha), [geethak5@srmist.edu.in](mailto:geethak5@srmist.edu.in) (K. Geetha), [snksnk17@gmail.com](mailto:snksnk17@gmail.com) (S. Neelakandan), [vinay.kumar@gl.a.ac.in](mailto:vinay.kumar@gl.a.ac.in) (V. Kumar).

study is primarily hampered by the lack of ethnicity information; indeed, in this age of deep learning (DL), there is a need for having a large number of data accessible to efficiently train convolution neural networks.

There is presently no ethnicity dataset equivalent in size to the massive datasets available for other facial soft biometrics [7]. To thwart this monitoring, it was recently found that the convolutional neural network (CNN) is trained on the currently available dataset for ethnicity recognition, which has limited ability for generalisation across various test sets. The absence of ethnicity data is mostly owing to the difficulties inherent in their annotation and collection methods. Indeed, ethnicity is an extremely disputed topic. Unlike other biometrics, such as gender, it can be characterised subjectively rather than numerically, complicating the discovery of common differentiating characteristics. As a result, the categorisation is based on visual variations in the somatic face characteristics that humans are familiar with [8]. It suggests that automated annotation techniques including, for example, birthplace could not be identified; the ground truth of ethnicity group must be made automatically by humans and the consistency of ethnicity labels is strongly dependent on the capacity of the annotators.

Various ways to solve ethnicity detection and classification difficulties have been presented in recent years. Many of these algorithms are handcrafted, which performs insufficiently on gender and age predictions of unrestrained in the wild photos [9]. Such typical hand engineered systems are based on variations in dimension of face descriptors and facial features that are incapable of handling the many levels of variation assessed in this stimulating unrestricted imaging condition. The photographs in this category differ in lighting, look, noise and position, which may affect the capacity of this automatically recommended computer vision approach to accurately label the image's gender and age. Deep learning-based techniques have lately demonstrated encouraging efficiency in these sectors, particularly in the gender and age classifications of unfiltered face photos. The current work in gender and age classifications, as well as hopeful signals of progress in DL and CNN, presented a novel end-to-end DL-based classification method for ethnicity detection.

This paper develops an intelligent DL-based ethnicity recognition and classification technique utilising facial pictures, known as the IDL-ERCFI technique, to recognise and classify ethnicity based on face images. The IDL-ERCFI technique primarily uses face landmarks to align images prior to transferring them to the network. Furthermore, the proposed model employs an Xception network as a feature extractor and principal component analysis (PCA) as a feature reduction technique. Additionally, ethnicity categorisation is carried out using an ideal kernel extreme learning machine (KELM) methodology, with the KELM method's parameters optimised using the Glow-worm Swarm Optimization (GSO) technique. A detailed experimental examination is conducted to establish the IDL-ERCFI technique's superiority on a benchmark face image dataset.

The remainder of the paper is structured in the following manner. **Section 2** contains a review of the literature, while **Section 3** has an introduction to the suggested model. **Section 4** expands on the results analysis, and **Section 5** concludes.

## 2. Prior ethnicity recognition approaches

This section provides a thorough examination of present automated ethnicity recognition approaches. AlBdairi et al. [10] suggested fresh DL-CNN methodologies for developing a novel method for identifying a person's ethnicity based on visual traits. The fresh datasets for ethnics of people contain 3141 photos collected from two distinct peoples. As we all know, these are the principal image databases obtained for people's ethnicities and therefore datasets would exist for researchers to use. The unique strategy was linked to two sophisticated approaches, Inception V3 and VGG and the validation accuracy for each CNN approach was estimated.

Wong et al. [11] used a large-scale ML architecture to estimate ethnicity using a new collection of census and name locations. The binary and multi class classification ML pipeline was established with Census 1901. Regularised LR, C-support vector and NB classifier are all part of the ML method. The name feature includes substrings, double metaphones, the entire name string and many names entity patterns, whereas the location feature includes substrings of district, province and sub districts, as well as the entire location string. Vo et al. [12] investigated using a DL approach called RRF, which consists of IC, RR modules for face recognition and pre-processing (FD and P). This paper proposes two distinct modules for the RR model. The first module is RR, which employs a DCNN (the RR CNN module). The following module (the RR-VGG module) is a fine-tuning method for RR based on VGG, the well-known trained method for object recognition.

Sukumaran and Brindha [13] proposed a unique face recognition algorithm based on face shape features. The described method consists of two major stages: (a) detection and (b) feature extraction. Initially, the feature based on shape and facial colour is mined. The reduction dimension feature is then exposed to DBN, which identifies the race. Furthermore, in order to make the current method the most efficient in terms of predictions, the weights of DBN are finetuned utilising a novel hybrid methodology known as LMUDA. Ahmed et al. [14] discuss the difficulty of detecting race in four diverse racial groups: Indian, Caucasian, African and Asian. The freshly evolved BUPT Equalised Face dataset, which contains approximately 1.3 million photos in an unconstrained platform which achieves an advanced accuracy of 97%.

Greco et al. [15] recommended the VMER datasets, which contain over 3 million facial images annotated with four ethnicity classifications: African American, East Asian, Caucasian Latin, and Asian Indian. The final annotation is accomplished by a process that entails the opinions of three persons of diverse ethnic origins in order to avoid the prejudice associated with well-documented race effects. Additionally, they provide an in-depth evaluation of many prominent deep network frameworks, including ResNet-50, VGG-Face, VGG-16, and MobileNet v2. Al-Humaidan et al. [16] give an Arab dataset labelled appropriately for Arab subethnic groupings, which they subsequently classify using DL algorithms. Three Arab image datasets were created: GCC, Egyptian, and Levant. The two modes of learning were combined to overcome the obstacles. Initially, supervised DL (classification) was utilised, with a CNN pre-trained technique chosen because to the CNN model's superior performance on the CV classification test. Then there's unsupervised data collection (deep clustering). Unsupervised learning is used to evaluate a model's ability to categorise ethnicity.

Khan et al. [17] created a method for categorising races based on an existing face segmentation architecture. A technique for segmenting faces has been developed utilising a DCNN methodology. They trained the DCNN by labelling face photographs according to seven separate categories: skin, nose, eyes, hair, back, mouth, and brows. The segmentation result was generated using the DCNN approach described in the previous phase. We employ a probabilistic classification strategy and construct PMs for all semantic classes. Matkowski and Kong [18] examine unrestricted gender and ethnicity categorisation. Gender and ethnicity labels are supplied and collected for an open-source database of hand photographs taken from the Internet. The 5 DL approach has been optimised and calculated in scenarios including ethnicity and gender categorisation using palmar 1) palmprint, 2) full hand, and 3) segmented hand pictures.

Kalinga [19] investigates the issue of ethnicity bias in the current facial detection system and proposes a novel way for taking ethnicity into account during the module training phase. The goal is to investigate the ethnicity bias problem in the current advanced facial identification method Facenet and to propose a novel picture election approach for minimising the effects of ethnicity bias without modifying the NN framework. Facenet's work is used to modify the image selection

technique, which initiates the module's training phase. As the input data to the training module is fed in batch by batch, the batch forming method is updated so that each batch represents the overall ethnicity distributions of the training dataset. Following that, the trained technique is comprehensively computed by classification task, clustering attempt by related to the base module. Osman et al. [20] investigate the accuracy of a texture-based ethnicity detection approach from facial models in four ethnic classes. The presented technique had several stages, including FS, classification and face detection. The identified face is then used with three displayed face blocks, namely 11, 12 and 22. A GLCM under different face blocks was used in the FE technique. Later, the final step was completed utilising several classification methods such as NB, KNN, RF and MLP.

### 3. Materials and methods

#### 3.1. System architecture

In this paper, ethnicity detection is viewed as a classification problem, which is addressed by the construction of the IDL-ERCFI technique. Fig. 1 depicts the overall system architecture of the IDL-ERCFI approach. Pre-processing, feature extraction, feature reduction, classification and parameter tuning are the five distinct procedures. To ensure compatibility, the input facial photographs are pre-processed at the start. Following that, the Xception model extracts the features and the dimensionality is reduced using the PCA technique. Furthermore, the KELM model is employed for classification and the parameter tuning procedure is carried out using the GSO algorithm. The following describes how these procedures function in detail.

#### 3.2. Pre-processing

Since the images exist in distinct poses, pre-processing is needed to guarantee uniformity and thereby improves the model performance. It includes five different subprocesses namely

- Face Recognition and Alignment,
- Resize images,
- RGB to grayscale transformation,
- Mean centering and
- Normalisation into the range [0, 1].

The images perform a 2D alignment process and the steps involved in it are provided in the following:

- a. The facial landmarks are identified by the use of dlib. The angle done by line passing over the eyes with a parallel line is determined.
- b. The rotation of the images takes place over the centre of the eyes using the angle determined.
- c. The target picture is cropped around the face, resulting in the eyes performing at a distance of 32% of width and 38% of height from the top. The generated picture was then scaled to 40\*40 pixels and converted to grayscale.
- d. The face with maximum yaw is considerably clipped on the bottom regions when the previous steps are performed. For correcting them, the chin is also included in the frame.

**Algorithm 1.** Steps involved in pre-processing.

```

Input : image and target size
Output : Aligned image pre-Process (image, target size)
Start:
    Determine 67 facial landmark
    A = angle develop the straight-lines pass eyes by the horizontal
    Rotation of images takes place using A over the eye centre
    Crop the face changing the scale
    Resizing of cropped image take place
End
Step 6: Display outcome of rotation, cropping and resizing processes

```

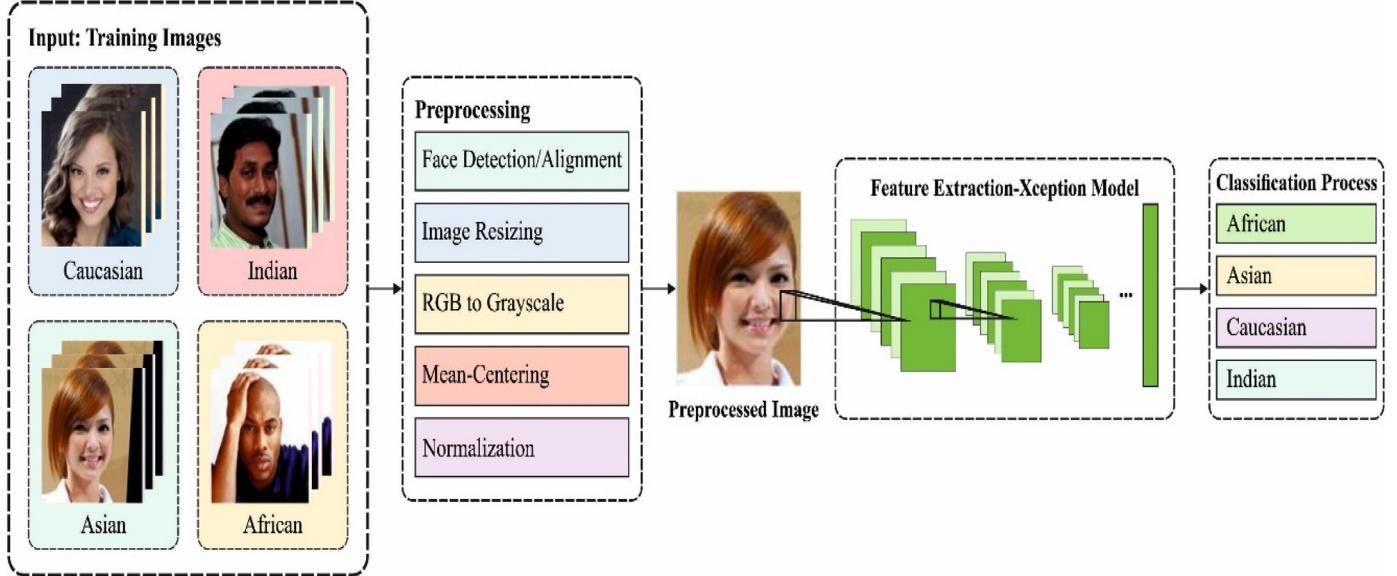
#### 3.3. Feature extractor: Xception model

The Xception model is applied to extract feature from the applied pre-processed facial photos. Recently, DL technology has become a well-known application derived from an ML method with an increase in multilayer FFNNs. The number of layers in classic NN is constrained in applications of constrained hardware features due to learned parameters and relationships between the layers requiring highest calculation time. With the development of a higher-end system, it may be able to teach deep methods using multilevel NN.

The DL approach is derived from CNN and is used in a wide-ranging range of applications such as voice analysis, object prediction, machine learning and image processing. CNN is also a multilayer NN [21]. Furthermore, CNN benefits from FE, which reduces the pre-processing stage to a more extent. As a result, it is insufficient for doing a

pre-study to detect visual features. Convolution, Input, Fully Connected (FC), Pooling, Dropout, Classification and Relu layers make up the CNN. This layer's performance is determined farther down.

- The input layers are the initial layer of CNN. Now, data is providing to these systems without pre-processing. Now, input size varies from pretrained DL structure is employed.
- Convolutional layers are depending upon the CNN approach i.e., employed to extract the feature maps through pixel metrics of image. As well, it has relied on the circulation of certain filters. Accordingly, a new image matrix achievement has been made. The filter has various sizes such as  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , or  $11 \times 11$ .
- The ReLU layers are located behind the Conv. layers. The crucial characteristics that distinguish the function from the activation function, such as a hyperbolic tangent and sinus, have resulted in a compelling conclusion. This function is typically used for non-linear conversion procedures.



**Fig. 1.** Overall process of IDL-ERCFI model.

- The pooling layers are determined after the Conv. and ReLU algorithms have been applied. It is mostly used to reduce the size of the input and prevent the scheme from memorising. These filters consider the max/average pooling function on image matrices that consider a certain feature. Following that, maximal pooling is used because it shown the best function.
- FC layer occurs afterward Pooling, Conv and ReLU layers. The neuron in these layers is connected to the region of present layer.
- Dropout layers are employed to eliminate the network memorises. The efficiency of this method is calculated using the random removal of some nodes.
- The classification layers are created following the FC Layer, when the classification operation was carried out. The likelihood value in classification refers to how close the class is to being selected. The softmax function is used to achieve this metric.

The DL-based Xception framework is utilised in this work to extract characteristics from a facial image. The Xception approach is identical to inception, except that the inceptions are replaced by depth wise separate Conv. layers. Specifically, the architecture of Xceptions is built on a linear stack of depth wise independent convolutional layers with a linear residual attached. Pointwise and depth wise layers are used in this method. In the depth wise layer, a spatial convolutional occurs manually in the channel of input data and in the pointwise layer, a  $1 \times 1$  convolution layer maps the results of new channel space in the depth wise convolutions applications. The framework of the Xception model is depicted in Fig. 2.

### 3.4. Feature reduction: PCA approach

The retrieved characteristics may be high dimensional, necessitating the reduction process. The PCA is a well-known dimension reduction technique since it resolves the “curse of dimensionality” without data loss. The arithmetic expression for the PCA approach is supplied sequentially.

#### 3.4.1. Mean

It can be a rate of central tendency. The arithmetic form for mean is written in Eq. (1). At this point,  $Q$  represents the arbitrary number and size of instances are referred to as  $l$ .

$$\text{Mean}(\bar{Q}) = \frac{1}{l} \sum_{i=1}^l Q_i \quad (1)$$

#### 3.4.2. Standard deviation (SD)

The average distance among the mean as well as point at that data has been estimated by squaring them. It can be mathematical determined in Eq. (2).

$$SD = \sqrt{\frac{1}{l} \sum_{i=1}^l (Q_i - \bar{Q})^2} \quad (2)$$

#### 3.4.3. Covariance

It defines the quantity of differences in dimensional from mean. The mathematical equation for covariance is written in Eq. (3).

$$\text{Cov}(Q, R) = \frac{1}{l} \sum_{i=1}^l (Q_i - \bar{Q})(R_i - \bar{R}) \quad (3)$$

When the extracted feature is decreased with its dimensional, it can be normalised and is employed for classification.

### 3.5. Image classification: KELM model

The KELM model, which obtains the reduced features from the PCA, can handle the ethnicity recognition problem, which is considered a classification challenge. ELM is identified as a new learning strategy for training “generalised” SLFN, i.e., capable of achieving the largest generalisation function while learning at a rapid pace on challenging situations. ELM, unlike CNN, does not require standard variable updates. The basic ELM is divided into two steps: ELM parameter solving and ELM feature mapping [22]. Initially, a hidden representation is created from actual input data via non-linear feature mapping. Different activation functions are used in the ELM feature mapping stage depending on the ELM standard. Following that, the output weight parameter is addressed using an MP generalised converse and low norm least square solution using a standard linear technique without a learning round. It can be employed that  $N$  different training instances were applied,  $\{X, Y\} = \{(x_i, y_j)\}_{i=1}^N$ , where  $x_i \in \mathbb{R}^d$  denotes a  $d$ -dimension input vector

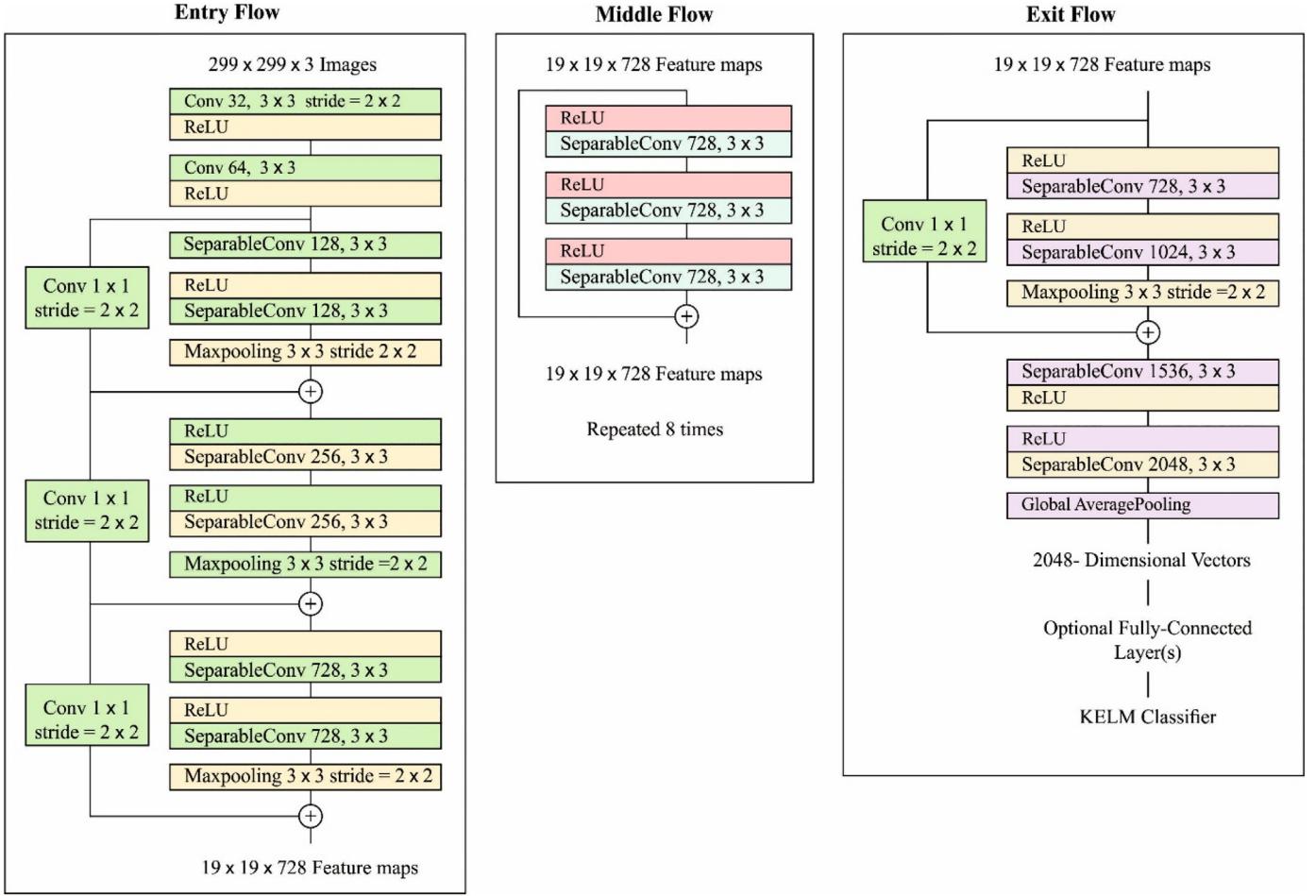


Fig. 2. Architecture of Xception model.

and  $y_j \in \mathbb{R}^c$  signifies a  $c$ -dimension target vector. The ELM method and  $Q$  hidden node is shown in Eq. (4):

$$y_i = \sum_{j=1}^Q \beta_j h_j(\omega_j \cdot x_j + b_j), i = 1, 2, \dots, N, \quad (4)$$

Whereas  $\omega_j \in \mathbb{R}^d$  refers to the input weight vector,  $b_j$  implies the hidden layer bias and  $\beta_j \in \mathbb{R}^c$  represents the last weight vector for  $j$ th hidden node.  $h_j(\cdot)$  indicates the last value of  $j$ th hidden node. In the event of generalisation, Eq. (4) is demonstrated by

$$H\beta = Y, \quad (5)$$

In which  $H$  implies the hidden layer result matrix and  $\beta$  denotes the resulting weight matrix.

Because the ELM has numerous advantages, numerous ELM modifications were created to report the function of sensing. The optimal role of ELM on HSI is then used in real-world applications. The kernel learning model is created using ML approaches to detect the usual relationships between labels and features. When compared to generic applications, the kernel learning approach performs best in terms of speeding up the non-linear association between labels and features. In HSI, non-linear relationships between pixels and ground cover are common. The KELM architecture is offered as a spectral spatial classification solution for tackling this challenge. Following that, KELM is widely used as a classifier method to forecast ground cover for each pixel. Fig. The KELM method's network framework is depicted in Fig. 3. In ELM, present mapping  $h(x_i^{ss})$  are not known to the user. Therefore, Mercer's state was employed as a kernel matrix for ELM by,

$$\Omega_{ELM} = HH^T, \quad (6)$$

In which  $i$ th row and  $r$ th column units represents  $\Omega_{ELMi,r} = h(x_i^{ss}) \cdot h(x_r^{ss}) = K(x_i^{ss}, x_r^{ss})$ . In the event of  $i$ th row vector in  $\Omega_{ELM}$ ,  $\Omega_{ELMi} = [h(x_i^{ss}), \dots, h(x_N^{ss})]$ . Hence, the summarisation of KELM is given by,

$$f(x_i^{ss}) = h(x_i^{ss})H^T \left( \frac{I}{C} + HH^T \right)^{-1} Y = \begin{bmatrix} K(x_i^{ss}, x_1^{ss}) \\ \vdots \\ K(x_i^{ss}, x_N^{ss}) \end{bmatrix}^T \left( \frac{I}{C} + \Omega_{ELM} \right)^{-1} Y. \quad (7)$$

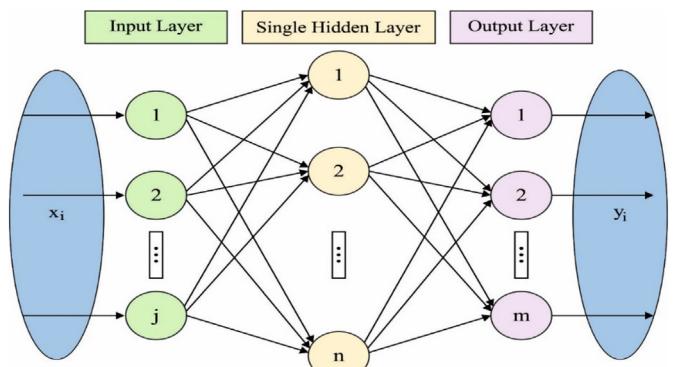


Fig. 3. Structure of KELM.

### 3.6. Parameter tuning using GSO algorithm

The GSO technique is used to modify the KELM model's parameters suitably in order to improve its classification performance. GSO is a metaheuristic technique created by Krishnanand and Ghose [23] that requires fewer parameters and achieves a higher degree of convergence. It can be used to solve a variety of optimization issues, such as pattern recognition and combinatorial optimization. The GSO algorithm is powered by glow-worm behaviour, which is based on the process of locating flaring adjacent glow worms in the search zone [29]. In the GSO algorithm, glow worms are attracted to each other based on their brightness. Throughout the search zone, all glow worms indicate the solutions to the target function and transfer a small amount of luciferin. Luciferin levels are related to the agents' fitness at their current location. The brighter one indicates an appropriate location (is an optimal solution). With a probabilistic system, all agents may be lured to neighbours whose luciferin intensities in the local decision domain are greater than their own and later travel toward them. The densities of a glow worm's neighbours affect the decision radius and thus the size of the local decision domain: if the neighbours' densities are lower, the local decision domain will expand to accommodate a greater number of neighbours; if the neighbours' densities are higher, the local decision domain will contract to allow the swarm to split into small groups [30]. The preceding stages are continued until the algorithm satisfies the desired outcome. The majority of individuals now collect brighter glow worms. The GSO is composed of five basic stages: neighbourhood selection, luciferin update, decision radius update, movement, and moving probability.

#### 3.6.1. Luciferin-update phase

It is based on fitness and past luciferin levels, and also includes the guidelines [24].

$$l_i(t+1) = (1-\rho)l_i(t) + \gamma \text{Fitness}(x_i(t+1)). \quad (8)$$

Now,  $l_i(t)$  is the luciferin values of glowworms  $i$  at  $t$ , time  $\rho$  indicates the luciferin decay constants,  $\gamma$  denotes the luciferin improvements;  $x_i(t+1) \in R^M$  implies the position of glow worm  $i$  at  $t+1$  and Fitness ( $x_i(t+1)$ ) denotes the amount of fitness at glowworms  $i$  position at time  $t+1$ .

#### 3.6.2. Neighbourhood-select phase

Neighbour's  $N_i(t)$  of glowworms  $i$  at  $t$  time contains the lighter one also it is expressed as follows:

$$N_i(t) = \left\{ j : d_{ij}(t) < r_d^i(t); l_i(t) < l_j(t) \right\}. \quad (9)$$

Now,  $d_{ij}(t)$  represent the Euclidean distances  $i$  and  $j$  at time  $t$  and  $r_d^i(t)$

#### 3.6.3. Moving probability-computer phase

The glowworms exploit the probability rules for shifting toward other glowworms possessing superior luciferin level [31]. The probability  $P_{ij}(t)$  of glowworm  $i$  moves toward its neighbour,  $j$  could be given by:

$$P_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum_{p \in N_i(t)} l_p(t) - l_i(t)}. \quad (10)$$

#### 3.6.4. Movement phase

Assume glowworm  $i$  select glowworm  $j \in N_i(t)$  using  $P_{ij}(t)$ ; the distinct time method of the motion of glowworms  $i$  is below in Eq. (11):

$$x_j(t+1) = x_j(t) + \text{step} \left( \frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|} \right). \quad (11)$$

where  $\|\cdot\|$  is the Euclidean norm operator and step represents the step size.

#### 3.6.5. Decision radius update

At the time of update phase, the assessment radius of glowworms  $i$  can be characterised using Eq. (12):

$$r_d^i(t+1) = \min \left\{ r_{sy}, \maxm \left\{ 0, r_d^i(t) + \beta(n_t - |N_i(t)|) \right\} \right\}. \quad (12)$$

Now,  $\beta$  indicates a constant,  $r_{sy}$  is the radius of  $i$  and  $n_t$  represents a variable for monitoring the neighbours. The process description of the GSO technique is provided in Algorithm 2.

**Algorithm-2.** GSO pseudocode.

```

Initialize:  $m$  dimensions
Initialize:  $n$  glowworms
Let us Consider  $s$ : step size
Let us Assume  $x_i(t)$ : position of glowworm  $i$  at time instant  $t$ 
Place agents randomly
deploy - agents - randomly;
for  $i = 1$  to  $n$  do  $\ell_i(0) = \ell_0$ 
 $r_d^i(0) = r_0$ 
assume maximum_number of iterations= max_iter;
set  $t = 1$ ;
while ( $t \leq \text{max\_iter}$ ) do:
{
    for every glowworm  $i$  do:
         $\ell_i(t) = (1 - \rho)l_i(t - 1) + \gamma \text{Fitness}(x_i(t))$ ;
    for every glowworm  $i$  do:
    {
         $N_i(t) = \{j : d_{ij}(t) < r_d^i(t); \ell_i(t) < \ell_j(t)\}$ ;
        for every glowworm  $j \in N_i(t)$  do:
             $p_{ij}(t) = \frac{\ell_j(t) - \ell_i(t)}{\sum_{p \in N_i(t)} \ell_p(t) - \ell_i(t)}$ ;
             $j = \text{choose\_glowworm}(\vec{p})$ ;
             $x_i(t+1) = x_i(t) + \text{step} \left( \frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|} \right)$ 
             $r_d^i(t+1) = \min \left\{ r_{sy}, \maxm \left\{ 0, r_d^i(t) + \beta(n_t - |N_i(t)|) \right\} \right\}$ ;
    }
     $t \leftarrow t + 1$ ;
}
}

```



Fig. 4. Sample images BUPT-GLOBALFACE.

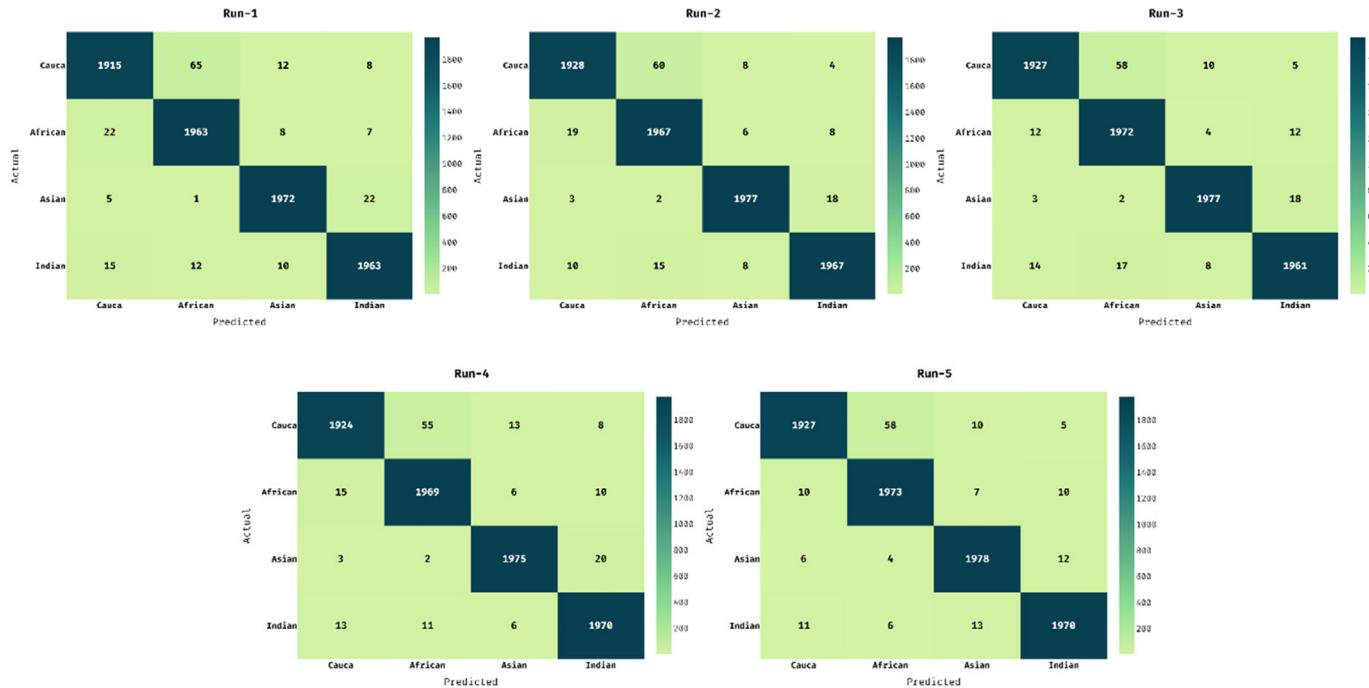


Fig. 5. Confusion matrix of proposed IDL-ERCFI model.

**Table 1**

Result analysis of proposed IDL-ERCFI model in terms of different measures.

No. of Runs	Classes	Accuracy	Precision	Recall	F-Score
Run-1	Caucasian	98.41	97.85	95.75	96.79
	African	98.56	96.18	98.15	97.15
	Asian	99.28	98.50	98.60	98.55
	Indian	99.08	98.15	98.15	98.15
	<b>Average</b>	<b>98.83</b>	<b>97.67</b>	<b>97.66</b>	<b>97.66</b>
Run-2	Caucasian	98.70	98.37	96.40	97.37
	African	98.62	96.23	98.35	97.28
	Asian	99.44	98.90	98.85	98.87
	Indian	99.21	98.50	98.35	98.42
	<b>Average</b>	<b>98.99</b>	<b>98.00</b>	<b>97.99</b>	<b>97.99</b>
Run-3	Caucasian	98.72	98.52	96.35	97.42
	African	98.69	96.24	98.60	97.41
	Asian	99.44	98.90	98.85	98.87
	Indian	99.08	98.25	98.05	98.15
	<b>Average</b>	<b>98.98</b>	<b>97.98</b>	<b>97.96</b>	<b>97.96</b>
Run-4	Caucasian	98.66	98.41	96.20	97.29
	African	98.76	96.66	98.45	97.55
	Asian	99.38	98.75	98.75	98.75
	Indian	99.15	98.11	98.50	98.30
	<b>Average</b>	<b>98.99</b>	<b>97.98</b>	<b>97.98</b>	<b>97.97</b>
Run-5	Caucasian	98.75	98.62	96.35	97.47
	African	98.81	96.67	98.65	97.65
	Asian	99.35	98.51	98.90	98.70
	Indian	99.29	98.65	98.50	98.57
	<b>Average</b>	<b>99.05</b>	<b>98.11</b>	<b>98.10</b>	<b>98.10</b>

The fitness function was determined as  $1 - CA_{validation}$  of 10-fold cross-validation technique in the trained dataset that is demonstrated as Eqs. (13) and (14). Also, the solution with superior  $CA_{validation}$  is lesser smaller fitness value.

$$Fitness = 1 - CA_{validation} \quad (13)$$

$$CA_{validation} = 1 - \frac{1}{10} \sum_{i=1}^{10} \left| \frac{y_c}{y_c + y_f} \right| \times 100 \quad (14)$$

Where  $y_c$  and  $y_f$  indicates the count of true and false classification outcome respectively.

#### 4. Performance validation

The suggested model's performance is assessed using the Python 3.6.5 tool on the BUPT-GLOBALFACE dataset [25–27]. The dataset used to construct the approach, BUPT Equalised Face, was just released. It is expected to be built at Beijing University of Posts and Telecommunications are made public in August 2019. The dataset contains 1.3 million photos from four different races. All of the races have roughly 320 K+ photos in equal amounts. Images are distributed to a total of 29,000 celebrities, with 7000 going to each race [32]. Images are in an unrestricted natural environment with a variety of faces, ages and stances ranging

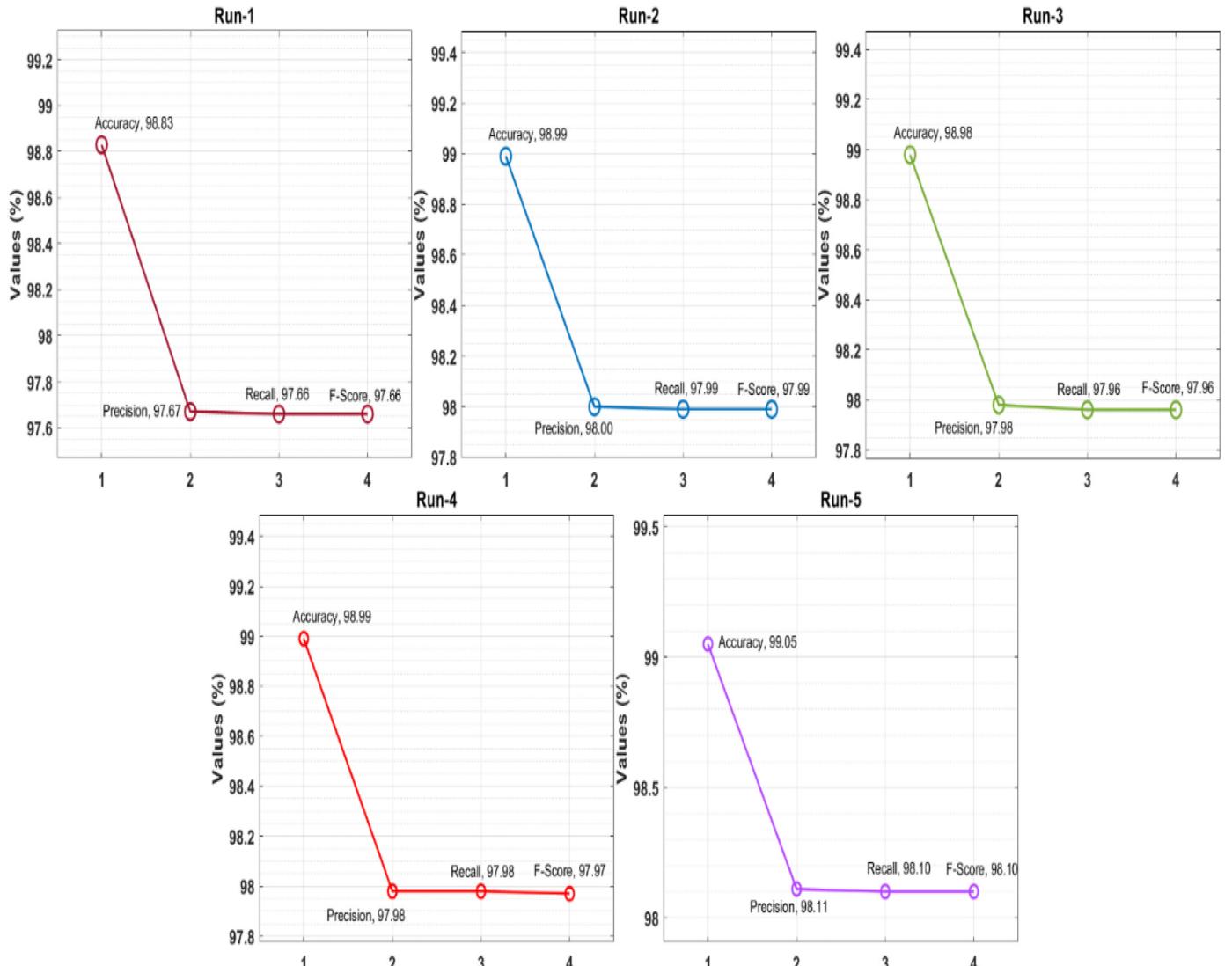
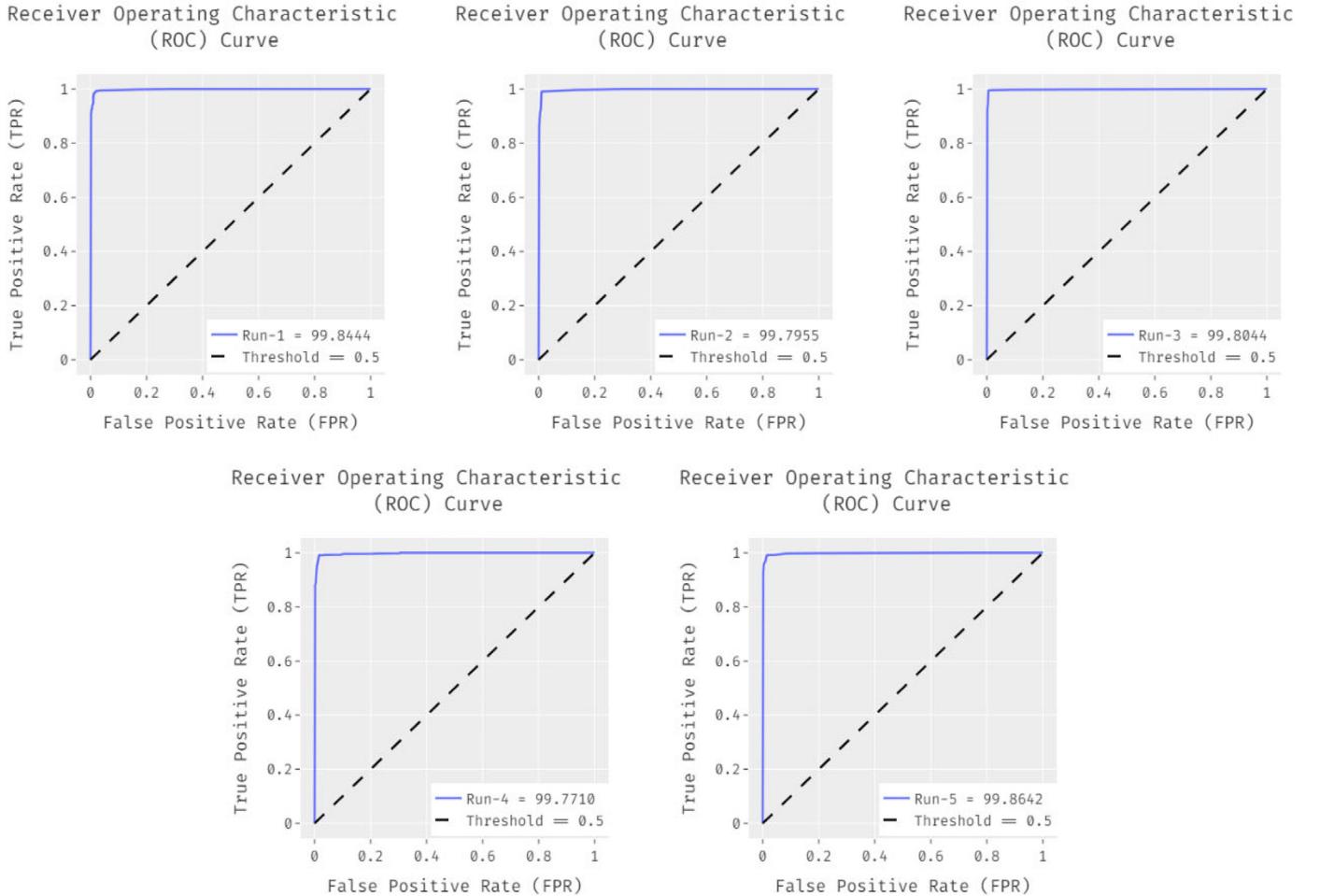


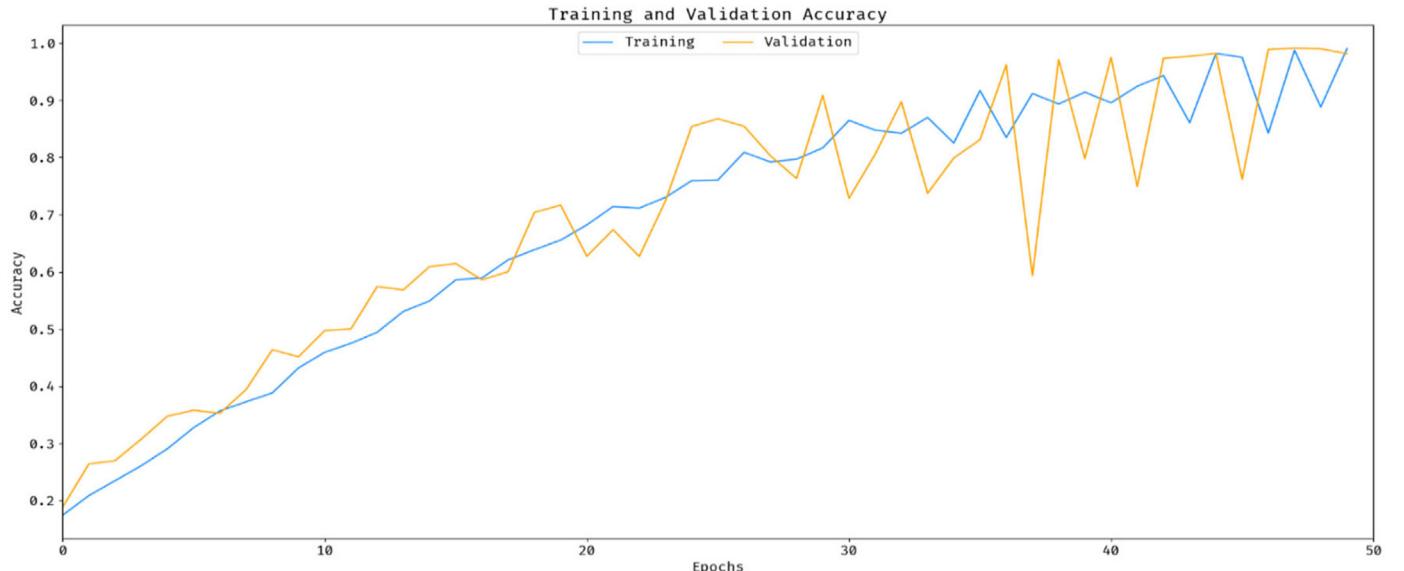
Fig. 6. Result analysis of IDL-ERCFI model under different runs.

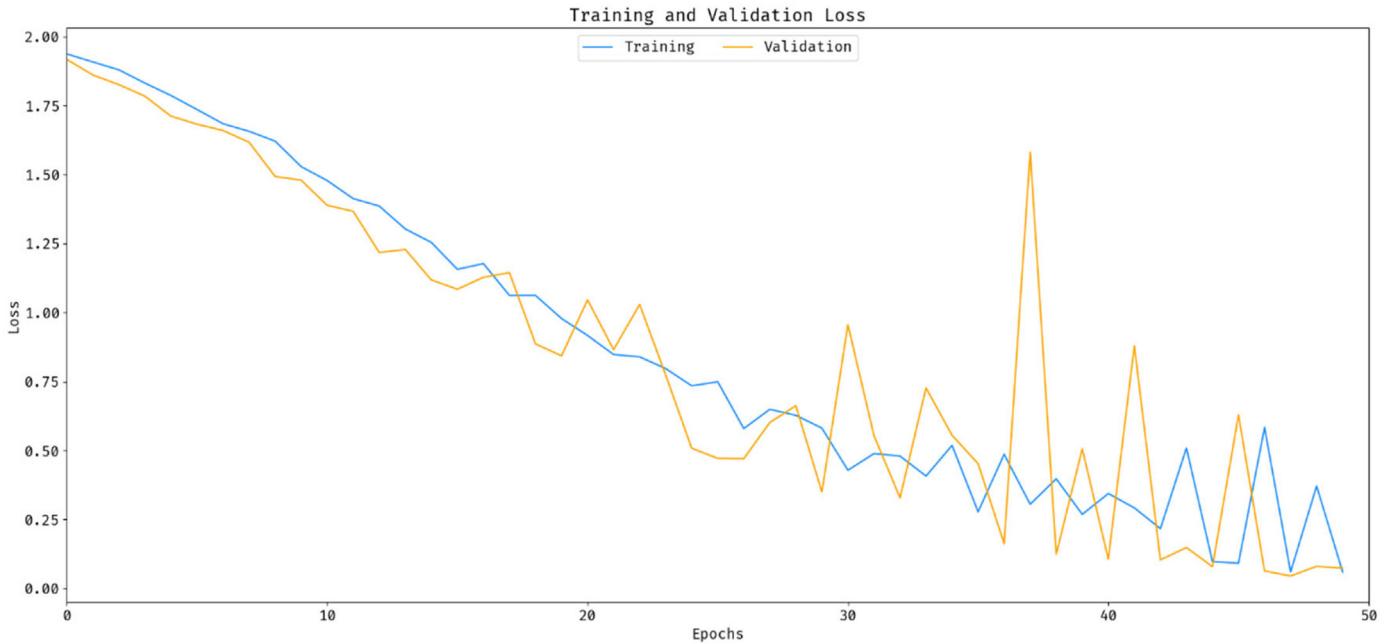
**Fig. 7.** ROC analysis of IDL-ERCFI model.

from modest to superior degrees of yaw and roll angles. The dataset has been down sampled in order to fit to the platform's capability. A total of 400,000 images were chosen for training, with 100,000 images each class, but only 32,000 images were used for validation, with 8000 images

per class [33]. The exam set contains 8000 photographs, 2000 of which are in each class. The sample photographs are shown in Fig. 4.

The set of confusion matrices obtained by the IDL-ERCFI technique on the classification of different ethnicities is shown in Fig. 5. The IDL-

**Fig. 8.** Training and validation accuracy analysis of IDL-ERCFI method.



**Fig. 9.** Training and validation loss analysis of IDL-ERCFI model.

ERCFI approach recognised 1915 occurrences as Caucasian, 1961 instances as African, 1972 instances as Asian and 1963 instances as Indian on execution run-1. Furthermore, on execution run-2, the IDL-ERCFI technique identified 1928 occurrences as Caucasian, 1967 instances as African, 1977 instances as Asian and 1967 instances as Indian. Furthermore, the IDL-ERCFI approach has well-known 1927 instances into Caucasian, 1972 instances into African, 1977 instances into Asian and 1961 instances into Indian on execution run-3. Furthermore, on run-4, the IDL-ERCFI method recognised 1924 instances as Caucasian, 1969 instances as African, 1975 instances as Asian and 1970 instances as Indian [34]. Furthermore, on run-5, the IDL-ERCFI algorithm recognised 1927 instances as Caucasian, 1973 instances as African, 1978 instances as Asian and 1970 instances as Indian.

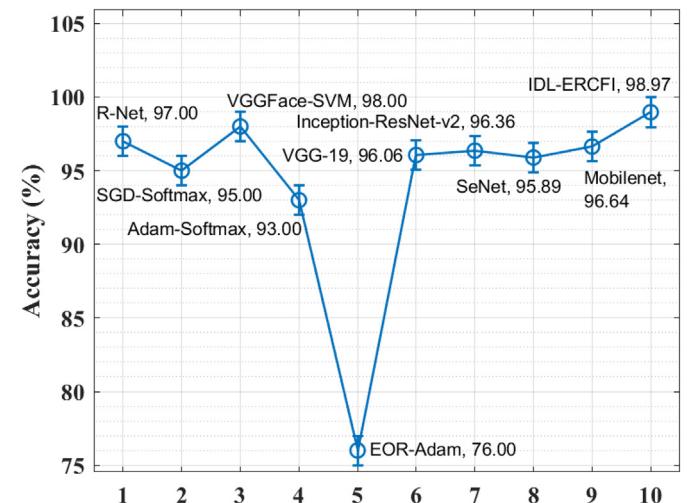
The classification performance of the IDL-ERCFI technique is shown in Table 1 and Fig. 6. The experimental results demonstrated that the IDL-ERCFI technique produced superior results in all of the runs tested. For example, in run-1, the IDL-ERCFI approach identified the instances with a maximum average accuracy of 98.83%, precision of 97.67%, recall of 97.66% and F-score of 97.66%. Furthermore, in run-2, the IDL-ERCFI method ranked the cases with the highest average accuracy of 98.99%, precision of 98%, recall of 97.99% and F-score of 97.99%. Furthermore, in run-3, the IDL-ERCFI algorithm identified the cases with a higher average accuracy of 98.98%, precision of 97.98%, recall of 97.96% and F-score of 97.96%. Furthermore, the IDL-ERCFI technique identified the

instances with a superior average accuracy of 98.99%, precision of 97.98%, recall of 97.98% and F-score of 97.97% in run-4. Finally, the IDL-ERCFI approach identified the cases with the highest average accuracy of 99.05%, precision of 98.11%, recall of 98.10% and F-score of 98.10% under run-5.

Fig. 7 depicts the performance of the IDL-ERCFI technique on ethnicity detection and classification under five different runs. The figure revealed that the IDL-ERCFI technique yielded effective results, with maximum ROC values of 99.8444, 99.7955, 99.9044, 99.7710 and 99.8642.

The training and validation accuracy graphs of the IDL-ERCFI technique on the applicable dataset are shown in Fig. 8. The figure showed that the IDL-ERCFI technique achieved higher training and validation accuracy levels. The validation accuracy is found to be greater than the training accuracy.

The training and validation loss graphs of the IDL-ERCFI technique on the applicable dataset are shown in Fig. 9. The graph demonstrated



**Fig. 10.** Accuracy analysis of IDL-ERCFI model with existing approaches.

**Table 2**

Result analysis of existing with proposed IDL-ERCFI model with respect to different measures.

Methods	Accuracy	Precision	Recall	F1-score
R-Net	97.00	97.00	97.00	97.00
SGD-Softmax	95.00	95.00	95.00	95.00
VGGFace-SVM	98.00	98.00	97.00	98.00
Adam-Softmax	93.00	93.00	93.00	93.00
EOR-Adam	76.00	76.00	76.00	76.00
VGG-19	96.06	96.06	96.06	96.06
Inception-ResNet-v2	96.36	96.36	96.36	96.36
SeNet	95.89	95.90	95.89	95.89
Mobilenet	96.64	96.64	96.64	96.64
IDL-ERCFI	98.97	97.95	97.94	97.94

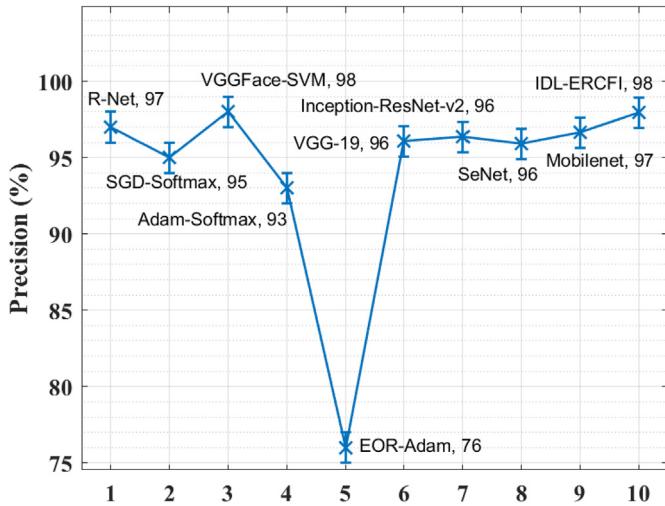


Fig. 11. Precision analysis of IDL-ERCFI model with existing approaches.

that the IDL-ERCFI method resulted in lower training and validation loss levels. Typically, the validation loss is set to be less than the training loss.

Finally, Table 2 [28] provides a quick comparison analysis of the IDL-ERCFI technique with existing techniques. Fig. 10 compares the accuracy of the IDL-ERCFI technique to that of existing techniques. According to the graph, the EOR-Adam technique produced poorer results with a lower accuracy of 76%, but the Adam-Softmax technique produced somewhat better results with a slightly higher accuracy of 93%. Simultaneously, the SGD-Softmax, SeNet, VGG-19, Inception-ResNet-v2 and Mobilenet models have improved to 95%, 95.89%, 96.06%, 96.36% and 96.64% accuracy. Furthermore, the R-Net and VGGFace-SVM approaches achieved near-optimal accuracy rates of 97 and 98%, respectively. The proposed IDL-ERCFI technique, on the other hand, produced better results, with an accuracy of 98.97%.

Fig. 11 depicts a precision comparison of the IDL-ERCFI approach with existing methods. The graph showed that the EOR-Adam approach produced the lowest results with a lower precision of 76%, but the Adam-Softmax method produced somewhat higher precision of 93%. Furthermore, the SGD-Softmax, SeNet, VGG-19, Inception-ResNet-v2 and Mobilenet approaches achieved somewhat higher precision of 95%, 95.90%, 96.06%, 96.36% and 96.64%, respectively. Furthermore, the R-Net and VGGFace-SVM approaches achieved near-optimal precision

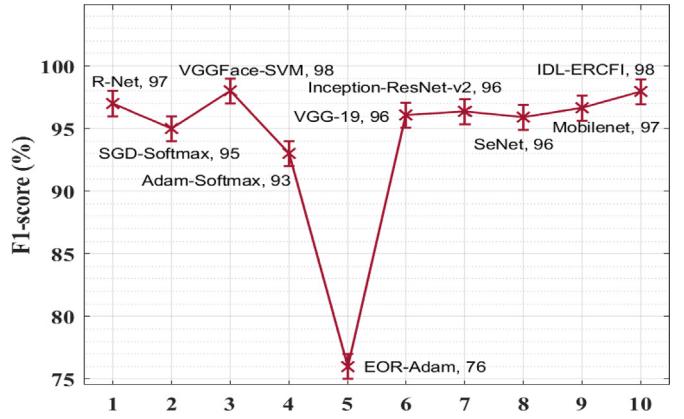


Fig. 13. F1-score analysis of IDL-ERCFI model with existing approaches.

values of 97% and 98%, respectively. Finally, the presented IDL-ERCFI approach yielded maximum efficiency with a precision of 98.97%.

Fig. 12 depicts a recall comparison of the IDL-ERCFI approach with existing methods. The figure outperformed that the EOR-Adam method yielded the lowest result with a recall of 76%, whereas the Adam-Softmax method yielded a little higher recall of 93%. Simultaneously, the SGD-Softmax, SeNet, VGG-19, Inception-ResNet-v2 and Mobilenet techniques achieved moderately higher recall of 95%, 95.89%, 96.06%, 96.36% and 96.64%, respectively. Furthermore, the R-Net and VGGFace-SVM approaches achieved near-perfect recall values of 97 and 98%, respectively. However, the proposed IDL-ERCFI methodology resulted in ideal performance, with a recall of 97.94%.

The F1-score comparison of the IDL-ERCFI method with previous approaches is shown in Fig. 13. According to the graph, the EOR-Adam method yielded the lowest F1-score of 76%, but the Adam-Softmax method yielded a slightly higher F1-score of %. Following that, the SGD-Softmax, SeNet, VGG-19, Inception-ResNet-v2 and Mobilenet methods achieved marginally closer F1-scores of 95%, 95.89%, 96.06%, 96.36% and 96.64%. Furthermore, the R-Net and VGGFace-SVM approaches achieved near-optimal F1-score values of 97 and 98%, respectively. Finally, the proposed IDL-ERCFI methodology produced ideal results, with an F1-score of 97.94%.

According to the tables and figures above, the IDL-ERCFI technique outperformed the other strategies and was discovered to be an excellent ethnicity recognition tool utilising facial analysis.

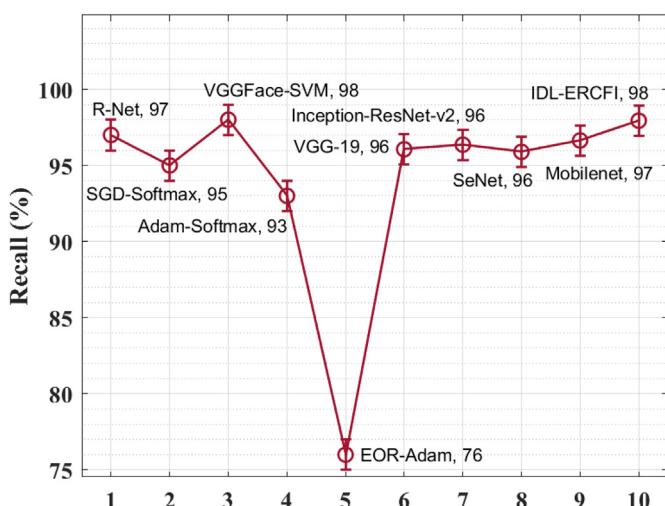


Fig. 12. Recall analysis of IDL-ERCFI model with existing approaches.

## 5. Conclusion

In this paper, ethnicity detection is viewed as a classification problem that is addressed by the design of the IDL-ERCFI technique. The overall system architecture of the IDL-ERCFI technique consists of five independent processes: pre-processing, Xception-based feature extraction, PCA-based feature reduction, KELM-based classification and GSO-based parameter adjustment. The features are extracted using the Xception model using the IDL-ERCFI technique and the dimensionality is reduced using the PCA technique. Furthermore, the KELM model is employed for classification and the GSO technique is used for parameter tweaking. The EOR-Adam method yielded the lowest F1-score of 76%, but the Adam-Softmax method yielded a slightly higher F1-score of %. Following that, the SGD-Softmax, SeNet, VGG-19, Inception-ResNet-v2 and Mobilenet methods achieved moderately closer F1-scores of 95%, 95.89%, 96.06%, 96.36% and 96.64%. Furthermore, the R-Net and VGG Face-SVM approaches achieved near ideal F1-score values of 97 and 98%, respectively. Finally, the proposed IDL-ERCFI methodology resulted in optimal performance with an F1-score of 97.94%. A rigorous experimental investigation is performed to demonstrate the superiority of the IDL-ERCFI technique on the benchmark facial picture dataset. The

results revealed that the IDL-ERCFI technique outperformed contemporary state-of-the-art techniques. The IDL-ERCFI technique can be extended in the future to create complex DL architectures using hyperparameter optimizers as part of a future extension.

## Author contribution

Both authors, S. Neelakandan and Gurram Sunitha, contributed to the conceptual and methodology of the research. Geetha, and Aditya Kumar Singh Pundir contributed acquisition of data, Testing and revision of the manuscript. Hemalatha and Vinay Kumar contributed for Validation, Testing and data visualization process of final outcomes.

## Funding

The authors received no specific funding for this Research

## Informed consent

Not applicable

## Consent for publication

Not applicable

## Declaration of Competing Interest

The authors declare that they have no conflict interests.

## References

- [1] A. Ahmed, K. Yu, W. Xu, Y. Gong, E. Xing, Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks, European Conference on Computer Vision, Springer 2008, pp. 69–82.
- [2] E.S. Madhan, S. Neelakandan, R. Annamalai, A novel approach for vehicle type classification and speed prediction using deep learning, *J. Comput. Theoret. Nano Sci.* 17 (5) (2020) 2237–2242 May 2020.
- [3] I. Anwar, N.U. Islam, Learned features are better for ethnicity classification, *Cybern. Inf. Technol.* 17 (3) (2017) 152–164.
- [4] H. Han, A.K. Jain, F. Wang, S. Shan, X. Chen, Heterogeneous face attribute estimation: a deep multi-task learning approach, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (11) (2017) 2597–2609.
- [5] S. Neelakandan, D. Paulraj, An automated learning model of conventional neural network based sentiment analysis on Twitter data, *J. Comput. Theoretical Nano Sci.* 17 (5) (2020) 2230–2236 May 2020.
- [6] S. Fu, H. He, Z.G. Hou, Learning race from face: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (12) (2014) 2483–2509.
- [7] L. Seidenari, A. Rozza, A. Del Bimbo, Real-time demographic profiling from face imagery with fisher vectors, *Mach. Vis. Appl.* 30 (2) (2019) 359–374.
- [8] P. Asha, L. Natrayan, B.T. Geetha, J. Rene Beulah, R. Sumathy, G. Varalakshmi, S., IoT enabled environmental toxicology for air pollution monitoring using AI techniques, *Environ. Res.* 205 (2022) 112574, <https://doi.org/10.1016/j.envres.2021.112574>.
- [9] T. Alafif, Z. Hailat, M. Aslan, X. Chen, On classifying facial races with partial occlusions and pose variations, IEEE International Conference on Machine Learning and Applications, 2017.
- [10] A.J.A. Albdairi, Z. Xiao, M. Alghaili, Identifying ethnics of people through face recognition: a deep CNN approach, *Sci. Program.* 2020 (2020).
- [11] K.O. Wong, O.R. Zaïane, F.G. Davis, Y. Yasui, A machine learning approach to predict ethnicity using personal name and census location in Canada, *PLoS One* 15 (11) (2020), e0241239.
- [12] T. Vo, T. Nguyen, C.T. Le, Race recognition using deep convolutional neural networks, *Symmetry* 10 (11) (2018) 564.
- [13] A. Sukumaran, T. Brindha, Nature-inspired hybrid deep learning for race detection by face shape features, *Int. J. Intell. Comput. Cybernet.* 13 (3) (2020) <https://doi.org/10.1108/IJICC-03-2020-0020>.
- [14] M.A. Ahmed, R.D. Choudhury, K. Kashyap, Race estimation with deep networks, *J. King Saud Univ. Comput. Inform. Sci.* (2020) <https://doi.org/10.1016/j.jksuci.2020.11.029>.
- [15] A. Greco, G. Percannella, M. Vento, V. Vigilante, Benchmarking deep network architectures for ethnicity recognition using a new large face dataset, *Mach. Vis. Appl.* 31 (7) (2020) 1–13.
- [16] N.A. Al-Humaidan, M. Prince, A classification of Arab ethnicity based on face image using deep learning approach, *IEEE Access* 9 (2021) 50755–50766.
- [17] K. Khan, J. Ali, I. Uddin, S. Khan, B.H. Roh, A Facial Feature Discovery Framework for Race Classification Using Deep Learning, arXiv preprint, 2021 arXiv:2104.02471.
- [18] W.M. Matkowski, A.W.K. Kong, Gender and ethnicity classification based on palmprint and palmar hand images from uncontrolled environment, 2020 IEEE International Joint Conference on Biometrics (IJCB), IEEE 2020, pp. 1–7.
- [19] A.T.N. Kalinga, Towards an Ethnic-Bias Free Approach for Facial Recognition (Doctoral Dissertation), 2021.
- [20] Rohit Anand, Harinder Singh, Interpretable filter based convolutional neural network (IF-CNN) for glucose prediction and classification using PD-SS algorithm, *Measurement* 183 (2021) <https://doi.org/10.1016/j.measurement.2021.109804>.
- [21] P.P. Mathai, C. Karthikeyan, et al., Deep learning based capsule neural network model for breast Cancer diagnosis using mammogram images, *Interdiscip. Sci. Comput. Life Sci.* (2021) <https://doi.org/10.1007/s12539-021-00467-y>.
- [22] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017, pp. 1251–1258.
- [23] H.L. Chen, G. Wang, C. Ma, Z.N. Cai, W.B. Liu, S.J. Wang, An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease, *Neurocomputing* 184 (2016) 131–144.
- [24] K.N. Krishnanand, D. Ghose, Detection of multiple source locations using a glow-worm metaphor with applications to collective robotics, *Proceedings of the IEEE Swarm Intelligence Symposium, IEEE Press June* 2005, pp. 84–91.
- [25] Z. Li, X. Huang, Glowworm swarm optimization and its application to blind signal separation, *Math. Probl. Eng.* 2016 (2016).
- [26] Mei Wang, Weihong Deng, Jian Hu, Xunqiang Tao, Yaohai Huang, Racial Faces in the Wild: Reducing Racial Bias by Information Maximization Adaptation Network, 2021 ICCV2019.
- [27] Mei Wang, Weihong Deng, Mitigating Bias in Face Recognition using Skewness-Aware Reinforcement Learning, 2021 CVPR2020.
- [28] C. Pretty Diana Cyril, J. Rene Beulah, Neelakandan Subramani, Prakash Mohan, A. Harshavardhan, D. Sivabalaselvamani, An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM, *Concurr. Eng. Res. Appl.* 29 (4) (2021) 386–395.
- [29] S.J. Rayen, J. Arunajasmine, Social media networks owing to disruptions for effective learning, *Proc. Comput. Sci.* 172 (2020) 145–151, <https://doi.org/10.1016/j.procs.2020.05.022>.
- [30] R.R. Bhukya, B.M. Hardas, Ch., T., et al., An automated word embedding with parameter tuned model for web crawling, *Intell. Automat. Soft Comput.* 32 (3) (2022) 1617–1632.
- [31] C. Al-Atrash, V.K. Nassa, B. Geetha, et al., Deep learning-based skin lesion diagnosis model using dermoscopic images, *Intell. Automat. Soft Comput.* 31 (1) (2022) 621–634.
- [32] D. Venu, A.V.R. Mayuri, Nilesh Shelke, an efficient low complexity compression based optimal homomorphic encryption for secure fiber optic communication, *Optik* 252 (2022) 168545 <https://doi.org/10.1016/j.ijleo.2021.168545>.
- [33] D.K. Jain, S.K.K. Sah Tyagi, S. Neelakandan, M. Prakash, L. Natrayan, Metaheuristic optimization-based resource allocation technique for cybertwin-driven 6G on IoE environment, *IEEE Transactions on Industrial Informatics*, 2021 <https://doi.org/10.1109/TII.2021.3138915>, (SCI, Scopus indexed, Impact Factor – 10.25).
- [34] A.S. Darabant, D. Borza, R. Danescu, Recognizing human races through machine learning, a multi-network, multi-features study, *Mathematics* 9 (2) (2021) 195.