Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 860 – 874

doi:10.3906/elk-2003-75

An evolutionary-based image classification approach through facial attributes

Selim YILMAZ, Cemil ZALLUHOĞLU*
Department of Computer Engineering, Engineering Faculty, Hacettepe University, Ankara, Turkey

Received: 13.03.2020 • **Accepted/Published Online:** 12.10.2020 • **Final Version:** 30.03.2021

Abstract: With the recent developments in technology, there has been a significant increase in the studies on analysis of human faces. Through automatic analysis of faces, it is possible to know the gender, emotional state, and even the identity of people from an image. Of them, identity or face recognition has became the most important task which has been studied for a long time now as it is crucial to take measurements for public security, credit card verification, criminal identification, and the like. In this study, we have proposed an evolutionary-based framework that relies on genetic programming algorithm to evolve a binary- and multilabel image classifier program for gender classification, facial expression recognition, and face recognition tasks. The performance of the evolved program has been compared with that of convolutional neural network, one of the most popular deep learning algorithms. The comparative results show that the proposed framework outperformed the competitor algorithm. Therefore, it has been introduced to the research community as a new binary- and multilabel image classifier framework.

Key words: Evolutionary-based algorithms, genetic programming, facial expression classification, gender classification, face recognition, multilabel image classification, convolutional neural network

1. Introduction

Analyzing human faces is a problem that has been frequently studied in many areas. For example, image processing, computer vision, pattern recognition, neural networks, and psychology. A significant increase has been observed in studies in this field in recent years due to security reasons in public places and authentication problems in digital environment over time. Face analysis studies have primarily targeted many objectives which include face recognition [1], face detection [2], face verification [3], facial expression recognition [4], and gender classification [5]. Among them, gender identification has a potential to be utilized in a wide range of real-world applications such as security surveillance and customer statistics collection in supermarkets, restaurants, shopping malls, and building entrances [6]. Facial expression recognition, however, has been used in many applications to achieve effective results such as user profile creation, customer satisfaction studies for publishing, and web services [7]. As for face recognition, known as the most challenging problem in the computer vision literature, it has primarily been used to identify people through their faces. Public security (e.g., credit card verification, criminal recognition), person verification, Internet communication (access to important systems, access control), and human-machine interactions are only a part of the domains where face recognition applications take part.

The concept of survival-of-the-fittest that originated from the Darwinian evolutionary theory [8] has led to the emergence of many evolutionary-based algorithms such as genetic algorithm (GA) [9], genetic programming

*Correspondence: cemil@cs.hacettepe.edu.tr

(GP) [10, 11], and grammatical evolution [12]. Among them, GP is a well-known machine learning algorithm that was first proposed by Cramer [10] and later developed by Koza [11]. The very simple yet effective search framework of GP has made it so popular in the literature that it has been applied to solve a wide range of machine learning problems including malware detection [13], voice detection [14], and software testing [15]. However, in spite of the great range of adoption of evolutionary-based methods in different areas, the application of such methods in the domain of image classification is still immature and hence needs further efforts.

In this study, we have studied the learning capability of evolutionary-based GP algorithm on binary- and multilabel image classification tasks. Speaking concretely, it has been used as a learning method to evolve classifier programs for gender classification, facial expression recognition, and face recognition tasks. The performance of the proposed framework has been compared with a recent, most-popular deep learning algorithm, convolutional neural network (CNN). The comparative experimental findings obtained from the proposed framework are very promising. GP has performed better or at least had a comparative performance on the tasks explored in this study.

The paper is organized as follows: Section 2 surveys the existing evolutionary-based approaches on the image-classification problems. The contributions have been highlighted in this section. The GP algorithm has been introduced in detail in Section 3. The proposed framework including feature extraction and model evolution has been clearly explained in Section 4. Section 5 provides the experimental environment and the findings along with a discussion on the GP's performance. Finally, Section 6 concludes the paper.

2. Related work

As stated earlier, the exploration of evolutionary-based approaches in the domain of image classification is very recent. There are only a few evolutionary-based approaches proposed to deal with image classification tasks in the literature so far.

A GP-based framework was proposed in [16] for the face recognition task. The proposed framework *i*) analyzes three different features (i.e. nose width, [mouth] width, distance between eyes), *ii*) estimates a face length according to the evolved model, and finally *iii*) returns an identified face where the difference between the estimated and the actual face length is smallest. Hence, the problem was regarded as a typical regression problem. It is reported that the proposed framework could reach up to a recognition rate of 98%. However, it is a must for the proposed framework to train a model separately for every person in the database, which shows that it is too experimentally costly and even not applicable for a very large database. In addition, only basic arithmetic operations were explored, which limits the learning capability of the GP. This can be regarded as another drawback of the framework.

Another 'experimentally exhaustive' work that relies on the GP was proposed for face recognition task in [17, 18]. Similar to [16], a separate model for every class, i.e. person, was evolved through the proposed GP-based framework. Instead of the predefined features like nose width in [16], the principal component analysis (PCA) was first employed to extract/reduce the features which were then fed into the GP for the model evolution. The low detection rates ranging from 63.5% to 67.5% of the GP-based framework was increased when it was hybridized with a leveraging method. This work was expanded in [19], where the same framework as [17, 18] was used. In the expanded study, the contribution of different PCA's such as two-dimensional PCA, multilinear PCA with respect to the detection rate was examined. The result showed that PCA's variants deteriorated the performance of the framework.

Another GP-driven framework was proposed in [20] targeting a multilabel image classification task. The main goal of the framework was to discriminate the different three shapes (circles, triangles, and rectangles) as well as four different grass-blades from each other. Similar to [16], only arithmetic operations were fed into the GP as a function set, but different from that work, the problem here was regarded as a classification problem. The results proved the efficacy of GP on the multilabel image classification problem. However, using only a few classes in the experiments prevented the proposed framework from showing its performance on a task with a larger class set (e.g., ≥ 10), which is an obvious drawback of the study.

Different from the aforementioned frameworks where GP was generally applied as a single-objective modelling algorithm representing the detection rate, multiobjective GP was also explored to avoid imbalance problem which may falsify the true capability of the evolved model when the detection rate or accuracy is used as a performance metric. To avoid that, multiobjective GP (MOGP) was proposed in [21] such that there were two objectives of the evolved model. While one of the objectives represents a detection rate obtained from the majority class, the other represents a detection rate obtained from the minority class. As in [16], the problem was regarded as a regression problem and the image object was classified according to the outcome of the GP model. The proposed framework is compatible with only binary-label image classification and thus should be expanded for multilabel image classification, which is the main disadvantage of the study.

Another MOGP-based classification framework was also proposed in [22]. Different from the other evolutionary-based frameworks, the main role here was to evolve a feature extraction model that is best for the image classification problem to be handled. While the data inputs (i.e. image objects) were given as the leaf nodes; filtering, pooling, and concatenation were fed into the GP as function set. Hence, the outcome of the MOGP could be a combination of these feature extraction approaches. The objectives were set to detection rate and tree depth which ensured that the evolved framework has a moderate computational complexity without sacrificing the detection/classification performance.

In addition to the GP-based approaches, genetic algorithm has also been applied for different image-classification tasks. Among them, a GA-based framework targeting only the face recognition task was proposed in [23]. Similar to [16], PCA was used for feature reduction while GA was used as a search method to optimize the weight vector of a support vector machine (SVM) classifier. The proposed framework yielded a detection rate of around 92%. In another study [24], an ensemble method that combines five different pretrained classifiers weighted by GA algorithm was proposed. The proposed approach targeted only the gender classification task. According to the results reported in [24], individual classifiers could reach up to a detection rate of 92%, whereas the proposed ensemble method could reach up to a detection rate of 94%. This shows that the contribution of the proposed approach is very limited when considering the cost of the application of GA for optimizing the classifier's weights.

As can be seen here, the use of evolutionary-based approaches is very limited and most of the existing methods proposed in the literature have serious drawbacks. In this study, GP has also been explored for both binary- and multilabel image classification tasks, which is a must considering the deficiencies of the studies proposed in the literature so far. The motivation and the contribution of this study are multifold, which differentiate this study from the existing studies. They can be outlined as follows:

- This is the first study that explored the efficacy of GP-based approach with a much wider range of function set (i.e. arithmetic and logical operators, mathematical functions, and the like) and feature set.

- The proposed framework is the first framework that targets both binary- and multilabel class image classification tasks at the same time. Hence, it can be adaptable for both problem types without requiring an extra effort. In addition, it is an efficient framework that does not require a separate model evolution for every class like in [16–18] as opposed to some existing approaches; rather, it evolves a generic model that can discriminate every class from another in a provided database.

- It is the first time, with this proposed study, that a very recently published database has been taken into consideration for the performance evaluation on three different tasks, i.e. gender classification, facial expression recognition, and face recognition tasks. In addition, this is the first study that assesses the learning capability of GP on the task having larger classes. The experiments have been conducted in an extensive environment where the performance of the GP algorithm has been compared for the first time with the CNN algorithm which is well-exploited on image-classification tasks. Therefore, this is the first study that reports the strength of the GP algorithm over CNN with respect to the detection rate on different image classification tasks.

3. Genetic programming

Tvolutionary-based GP algorithm is known as one of the most popular machine learning algorithms in the literature. Similar to the existing machine learning (ML) techniques, GP relies on the data as well to construct a hypothesis. However, GP is based on a population in which a number of solutions take part, which differentiates it from many of the ML techniques in the literature such as SVMs and decision trees. In GP, each solution (the term 'individual' is interchangeably used in this paper) in the population has a tree-like data structure.

Each of the GP individuals is regarded as a program that corresponds to a candidate hypothesis or a model towards an ML problem at hand. Hence, the main goal of GP is to evolve a fittest and a robust program. To evolve a program, a GP population is first created in a random manner (i.e. a number of individuals is generated), then generic operators (i.e. reproduction, crossover, and mutation) are applied to the population. In reproduction, a pair of individuals is first selected based on selection operators such as tournament-selection or roulette-wheel selection. Here the selection operator probabilistically chooses the fitter individuals from the population to be reproduced without completely ignoring the not-so-good individuals which contribute to the global search in the algorithm. In crossover, a node from each of the parent tree is randomly determined as crossover points. Then, two new offspring individuals are generated by replacing the subtrees rooted at the determined crossover points. In mutation, however, a mutation point of the generated offspring is determined in a way similar to the crossover and the subtree rooted there is replaced by a randomly generated new tree. After the generic operations, the fitness values that indicate how well the generated offspring could identify the problem at hand is evaluated. While the fittest ones are allowed to survive for the next generation, the poorest individuals die out. The evolution process continues until the termination criterion that is drawn by a designer is satisfied. The general steps of the GP algorithm is given in Algorithm 1.

```
Initialize population;
repeat
    Evaluate the fitness of each individual;
    Rank the population according to fitness values;
    Apply generic operators and reproduce new population;
until termination criterion is satisfied;
return best-of-run individual
```

Algorithm 1: General steps of the genetic programming algorithm.

4. The proposed framework

There are three consecutive phases in the proposed framework: *i*) feature reduction, *ii*) evolution of a classifier program through the GP algorithm, and *iii*) evaluation of the evolved best programs. In the first phase, the features of given image files have been reduced; hence, a dimensionally eliminated feature vector dataset has been constructed. Even if this phase brings about an extra computation effort to the proposed framework, it would be too exhaustive to train the classifier through GP in a case where the raw image features had been used as features. That is because GP tree would require a depth of log_n considering n (size of the features in GP tree) would be very large even if it was balanced. Another advantage of this phase is to avoid evolution of the GP classifier on the unimportant regions of the images. In the training/evolution phase, GP has been applied on the image dataset processed in the first phase. Here k-fold cross-validation approach has been adopted so that the evolved best classifier could have a satisfying performance on all over the dataset by not missing any of the objects belonging to a minor class. To do that, the whole dataset has been split into 5 chunks (i.e. $k = 5$). While 4 of these have been used in the evolution of the GP classifier, the remainder has been used in performance evaluation, which constitutes the final phase of the proposed framework. The general framework of this study is demonstrated in Figure 1. Each of the outlined phases has been explained in the following sections in detail.

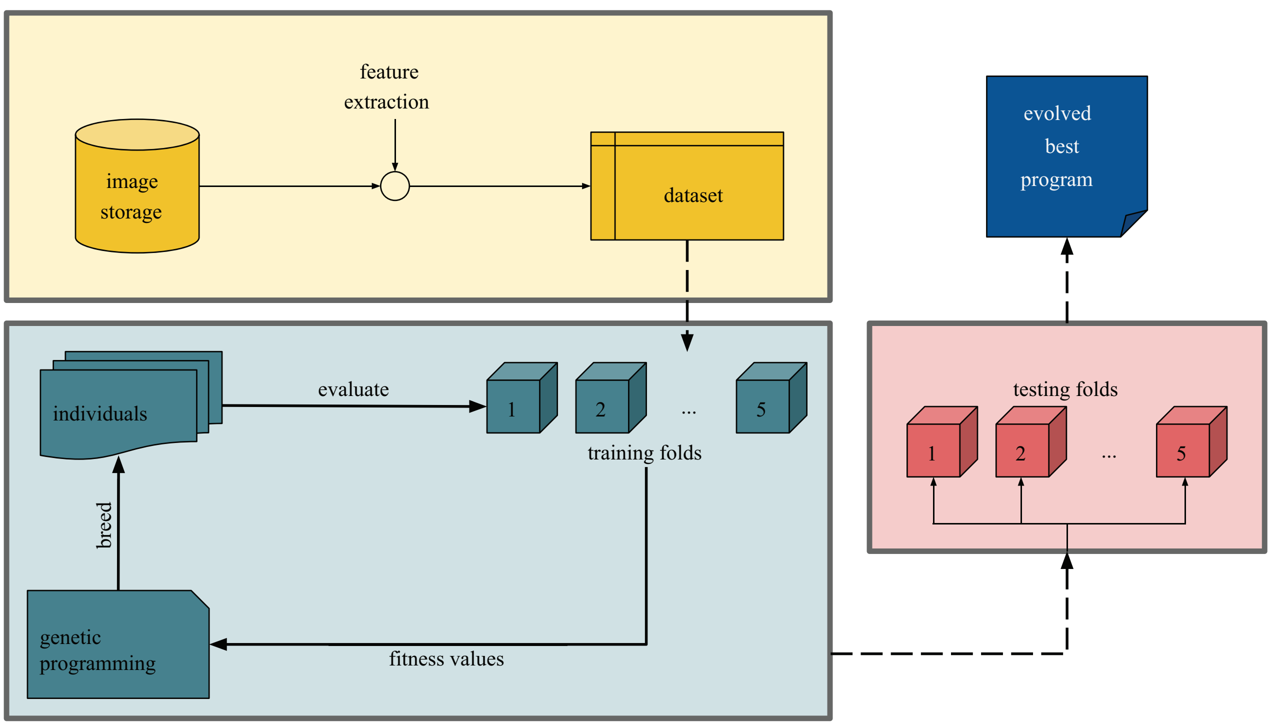

Figure 1. General framework of the proposed GP-based image classification task. The yellow region shows preprocessing processes, while green and red regions show, respectively, training and testing phases that yield the evolved best model.

4.1. Feature extraction and reduction

Feature extraction has been performed on each region, which are the bounding boxes of the faces for use in the GP algorithm. In this study, region covariance descriptors have been used as features. Region covariance descriptor [25] has widely been used for encoding the texture information within an image region. In this study, we have used this descriptor as a 9-dimensional vector that is constructed for each pixel. The descriptor consists of the image intensities, x-y coordinates, first-order partial derivative, second-order partial derivative, and the edge orientation. More formally, the feature vector f is identified as:

$$f(x,y)=\begin{bmatrix} x & y & I(x,y) & |\mathbf{I}_x| & |\mathbf{I}_y| & |\mathbf{I}_{xx}| & |\mathbf{I}_{yy}| & |\mathbf{I}_{xy}| & tan^{-1}(\frac{\mathbf{I}_y}{\mathbf{I}_x}) \end{bmatrix}^T \tag{1}$$

where (x y) denotes the x and y coordinates of the pixel, I(x,y) denotes the intensity of the pixel, $|\mathbf{I}_x|$ and $|\mathbf{I}_y|$ are defined with $|\frac{\partial I(x,y)}{\partial x}|$ and $|\frac{\partial I(x,y)}{\partial y}|$, respectively. They represent the first-order partial derivatives in horizontal and vertical directions, respectively. $|\mathbf{I}_{xx}|$, $|\mathbf{I}_{yy}|$, and $|\mathbf{I}_{xy}|$ are defined with $|\frac{\partial^2 I(x,y)}{\partial x^2}|$, $|\frac{\partial^2 I(x,y)}{\partial y^2}|$, and summation of $|\mathbf{I}_{xx}|$ and $|\mathbf{I}_{yy}|$. They are the second-order partial derivatives in horizontal, vertical, and mixed directions. The last component, however, represents the edge orientation.

Given this 9-dimensional feature vector, an image region can be represented with a 9×9 covariance matrix C of the feature points f_i that fall into that region, such that:

$$C = \frac{1}{N-1} \quad \sum_{i=1}^{N}(f_i - \mu_f)^*(f_i - \mu_f)^T \tag{2}$$

where μ is the mean of the feature vectors f_i, N is the total number of feature points, and * denotes the complex conjugation.

While the covariance matrix provides important discriminative information about a region, one important drawback of using this covariance matrix as is that the covariance matrices lie in Riemannian manifold and they cannot be compared using regular similarity measures. The similarity measures defined in [25] for their comparison are time-consuming, which is prohibitive for use in person counting problem, because of the excessive amount of processing required for each frame.

To overcome these limitations, an alternative method has been proposed in [26], which makes use of the fact that every covariance matrix has a unique Cholesky decomposition, and use this decomposition to define a feature vector that is based on the covariance matrices and, at the same time, lie on Euclidean space. Their proposed region descriptor called sigma set is defined as $S = \{s_i\}$ and each Sigma point s_i is calculated as

$$s_i = \{\beta L_1, \ldots, \beta L_d, (-\beta L_1), \ldots, (-\beta L_d)\} \tag{3}$$

where L_i is the i-th column of L which is the lower triangular matrix obtained via Cholesky decomposition of the covariance matrix $C = LL^T$ [26]. β is a scalar, which is taken as $\beta = \sqrt{2}$ in our experiments. In order to efficiently use the discriminative properties of region covariances, we use the sigma set as the first part of our feature vector.

Each set of sigma consists of a 81-dimensional vector, and the regions in the images are defined by these vectors which are sparse vectors. These vectors are transformed into dense vectors for use in genetic programming applications. The size of the vectors has been reduced to 41 dimensions as a result of the transformation process.

4.2. Evolution of classifier program

As stated earlier, GP has been employed as a learning algorithm to evolve a classifier program. Hence, every individual in GP has been regarded as a candidate classifier program; among them, the fittest one is returned by GP as a classifier for the tasks handled. Of different approaches representing the classification problems in GP such as binary decomposition, and static range selection, dynamic range selection, the class enumeration approach has been adopted here in the formation of a GP tree due to the reported limitations of other approaches [27]. An example to the class enumeration approach for a binary-label classification problem is given in Figure 2a. This approach enforces the root node of the GP tree to be IF statement that evaluates the condition at the left most subtree whether it satisfies the condition (T) or not (F). In the case where the condition is satisfied, the return value is class 1 (C1); otherwise, it is class 2 (C2). The GP tree representation used in the framework of this study and the corresponding program are given in Figure 2b and Algorithm **??**, respectively. In this study; class labels, all the features extracted from the images, and a constant value ranging from 0 to 1 have been used as terminal items. Therefore, these items have been constrained in the GP algorithm to be leaf nodes in the GP tree. As for nonterminal items, arithmetic and logical operators as well as mathematical functions have been used.

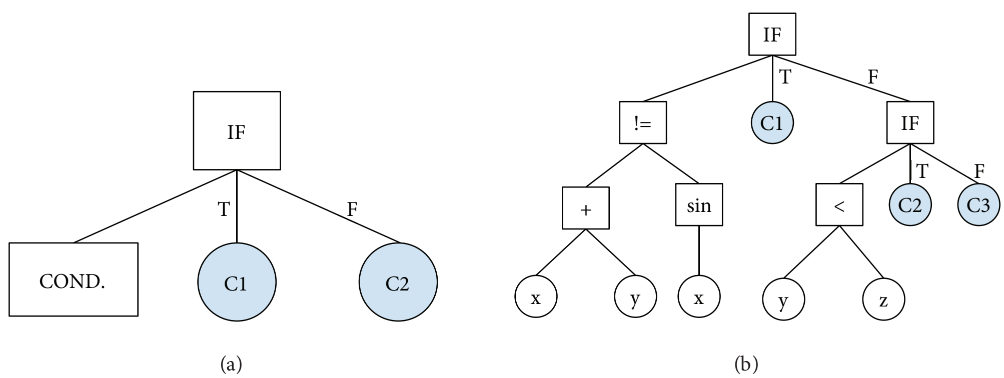

Figure 2. An example of the class enumeration representation (a) and the GP tree representation in the proposed framework (b).

Classification error percentage (CEP) has been employed as a measurement criteria to reveal how well a solution in GP can classify the objects in the dataset. The formulation of CEP is given in Equation 4.

$$CEP = \frac{|\{\forall X \in S_i : \nexists E(X) = S_i,\ S_i \subset D\}|}{|\{X : X \in D\}|} \tag{4}$$

where X stands for an object belonging to a class i $(\in \{1, 2, \ldots, S\})$, $E(X)$ represents a class estimated by candidate classifier program on the object X, and D represents the dataset. To be concrete, CEP returns the percentage of misclassified objects set from all the objects in the dataset. GP has been run 10 times per fold (50 runs in total) because of the stochastic nature of the evolutionary-based algorithms. Java-based evolutionary computation toolkit (ECJ)[1] has been used for the GP implementation. The parameter setting of the GP algorithm including the terminal and nonterminal items is outlined in Table 1. The parameters not listed in this table have been set to the default values used in ECJ.

[1]ECJ (2017). A Java-based Evolutionary Computation Research System [online]. Website https://www.cs.gmu.edu/ eclab/projects/ecj [accessed 00 10 2020].

Table 1. GP parameters and their values.

Parameters	Value
Nonterminals	+, -, *, /, sin, cos, log, ln, sqrt, abs, exp, ceil, floor, max, min, pow, mod, $<$, $\leq$, $>$, $\geq$,== ,! =, and, or
Terminals	class labels, features in Sec. 4.1, and rnd(0,1)
Generations	1,000
Population size	100 (fraction of elite individuals = 10%)
Crossover probability	0.8
Mutation probability	0.2
Selection strategy	Tournament selection (Tournament size: 7)
Initial depth of the GP tree	2 to 6
Maximum depth of the GP tree	17

5. Experiment

5.1. Database

CAS-PEAL (Chinese Academy of Science - Pose, Expression, Accessories, and Lighting) [28] and FERET (Facial Recognition Technology) [29] face databases have been employed in this study. A detailed overview of these databases have been outlined in the following sections.

5.1.1. CAS-PEAL database

CAS-PEAL database contains 99,594 images from 1040 individuals (595 males and 445 females). A total of nine cameras are mounted horizontally on an arc arm to simultaneously capture images across different poses.

Each person has three pose types and seven azimuths of the camera. Thus, a total of 21 images are obtained for each person with different pose variations. The poses are “looking down”, “looking into the camera”, and “looking up”. The azimuth values are –22°, –45°, –67°, 00°, 22°, 45°, and 67°. Example pose illustrations are given in Figure 3.

In addition to the neutral expression, some subjects were asked to smile, to frown, to be surprised, to close eyes, and to open mouth in construction of the database. For each expression, nine images of the subject under different poses were obtained using the nine cameras. In this study, we have obtained all the expressions from a single node because our goal is to recognize the facial expressions. In addition, both to avoid the high computational complexity and to ensure interpretability of the classifier program, we have used 950 images from 50 people by splitting them into 10 different groups such that each group has five people in the face recognition task. In this database, there are features such as pose, gender, and expression, as well as features such as lighting, accessory, background, time, and distance variations.

5.1.2. FERET database

FERET database has been employed for only face recognition because image records are not categorized according to the gender or expression. It contains 1762 images from 1010 individuals. There are 5–11 images for each individual in the dataset. The images in this dataset are obtained with different pose variations. The frontal images have 0 degree of pose angle. The left and right profiles have –90 and +90 degrees, respectively. The left and right quarter profiles, and left and right half profiles have –22.5, +22.5, –67.5, and +67.5 degrees of

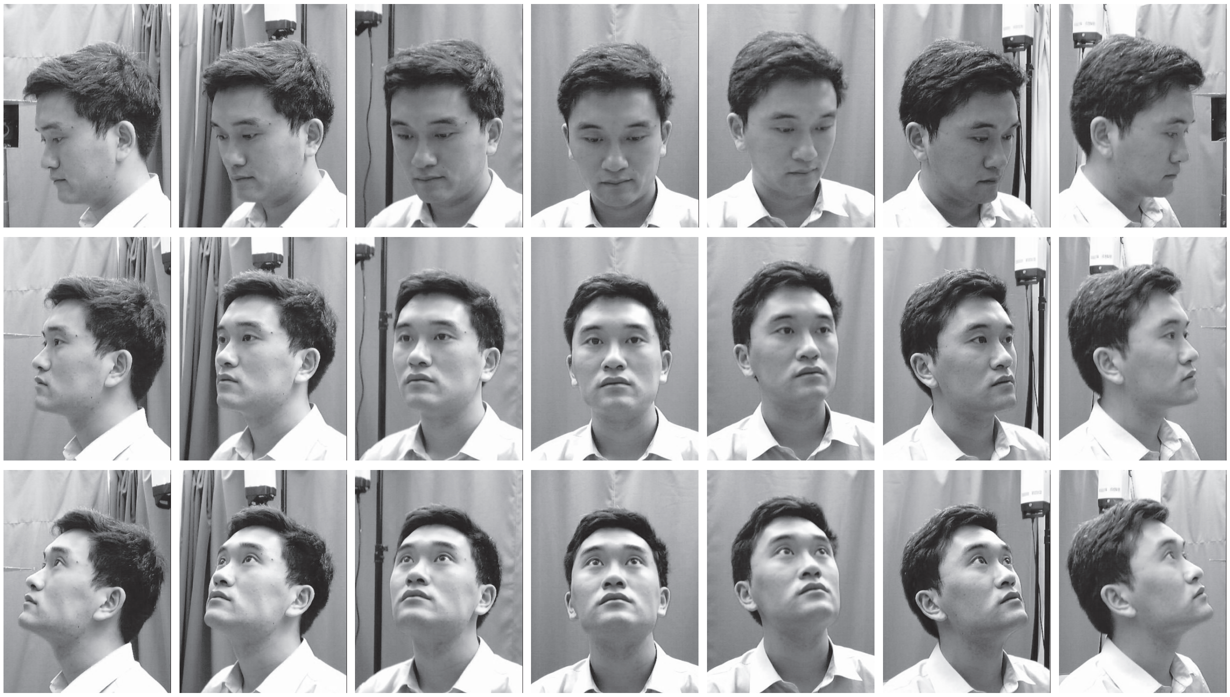

Figure 3. Example frame pose illustrations from CAS-PEAL face database. The images in the first row were obtained from the "looking down" pose, the second line from the "looking into the camera" pose, and those in the last row from the "looking up" pose. The images in all rows are at azimuths from –22° to +67°, from left to right.

pose angle respectively. In our study, we used all the pose variations available to recognize faces in the dataset.

5.2. Results

The performance of the proposed framework has been evaluated here on the unseen part of the database at the end of the evolution. In performance evaluation, the best classifier (GP program) evolved for each fold in the training phase has been taken into consideration. As in the training phase, Equation 4 has also been used for performance assessment of the best GP programs in the testing phase.

In order to analyze the classification success of the proposed framework in a comparative manner, we have applied a CNN algorithm as stated earlier. In this study, we have investigated three different CNN architectures. These are AlexNet [30], DenseNet161 [31], and VGG16 [32]. AlexNet architecture is one of the first deep learning architectures. This architecture has five stacked convolutional layers and three fully-connected layers. DenseNet161 architecture, however, contains dense blocks which are formed only by convolution layers and transition layers which are formed by convolution layers and pooling layers. VGG16 architecture is a feed-forward CNN model. VGG16 architecture is a widely used architecture for different problems such as object detection, image localization, segmentation, classification, and video classification. The experimental results showed that the VGG architecture gives better results than others. For this reason, we decided to use the VGG architecture in comparisons in our study.

For the sake of a fair comparison; *i*) the same database with the same fold split settings has been used, *ii*) Equation 4 has been used as a loss function for CNN as well, *iii*) the other experimental conditions used for GP (i.e. iteration, run time, etc.) have also been used in the application of CNN. The comparative mean

CEP values obtained from each of the testing folds have been outlined separately for each task (i.e. gender classification, facial expression recognition, and face recognition) in Table 2.

Table 2. Comparative mean error percentage values obtained from gender classification, facial expression recognition, and face recognition tasks.

Methods	Tasks			
	Gender	Expression	Face recog. (CAS-PEAL)	Face recog. (FERET)
GP	0.0000	0.0095	0.0590	0.0291
CNN	0.0080	0.3760	0.0418	0.0375

In addition to the classification performance, we have also measured the overall testing performance of the evolved best classifier program at particular midsteps in order both to reveal the classification performance on each testing fold at the earlier evolution steps and to find out if the evolved program has a problem of over-/underfitting problem. To do that, the best program obtainable after every 100 generations (i.e. at 0th, 100th, ..., 1000th generation) has been taken under testing environment. The overall classification convergence performance obtained from gender classification, facial expression recognition, and face recognition tasks for CAS-PEAL and FERET databases have been presented in Figure 4. The obtained findings have separately been discussed in the following sections.

5.3. Discussion

5.3.1. Gender classification

There are two classes in the gender classification task, which are ‘female’ and ‘male’, so it can be regarded as a binary-label image classification problem. The main role of the evolved GP problem here is to evaluate the interested features of the images with respect to its fittest tree structure and then to separate them as either female or male with a high accuracy.

From the findings obtained from the gender classification task (Table 2), it can be concluded that the GP has successfully classified all the testing objects into where they actually belong to. The CNN, however, could perform with an accuracy of 99.20%. It reveals that the proposed framework is capable of classifying the genders with a slightly better performance than the CNN on overall.

The overall training and testing classification performances captured at the midgenerations (Figure 4) suggest that GP is able to evolve a program that can correctly classify all the objects before reaching the first half of the total iterations. In addition to that, GP does not have an overfitting problem, which proves that the evolved classifier program is generic and can further be applied on more unseen data for gender classification.

5.3.2. Facial expression recognition

Contrary to the gender classification, there are more than two classes in the facial expression recognition task. That is why it can be regarded as a multilabel image classification problem. Eyes closed, frowning, laughing, mouth open, and surprised are the facial expression variations which represent the classes in this problem. The goal of the GP here is to first recognize the facial expression from a given image then to accurately classify it.

The comparative mean CEP values obtained from the facial expression recognition task given in Table 2 reveal that the GP is capable of 99.05% accuracy for all the testing objects on overall whereas the CNN has shown a classification performance with an overall accuracy of 62.40%. As in the gender classification task,

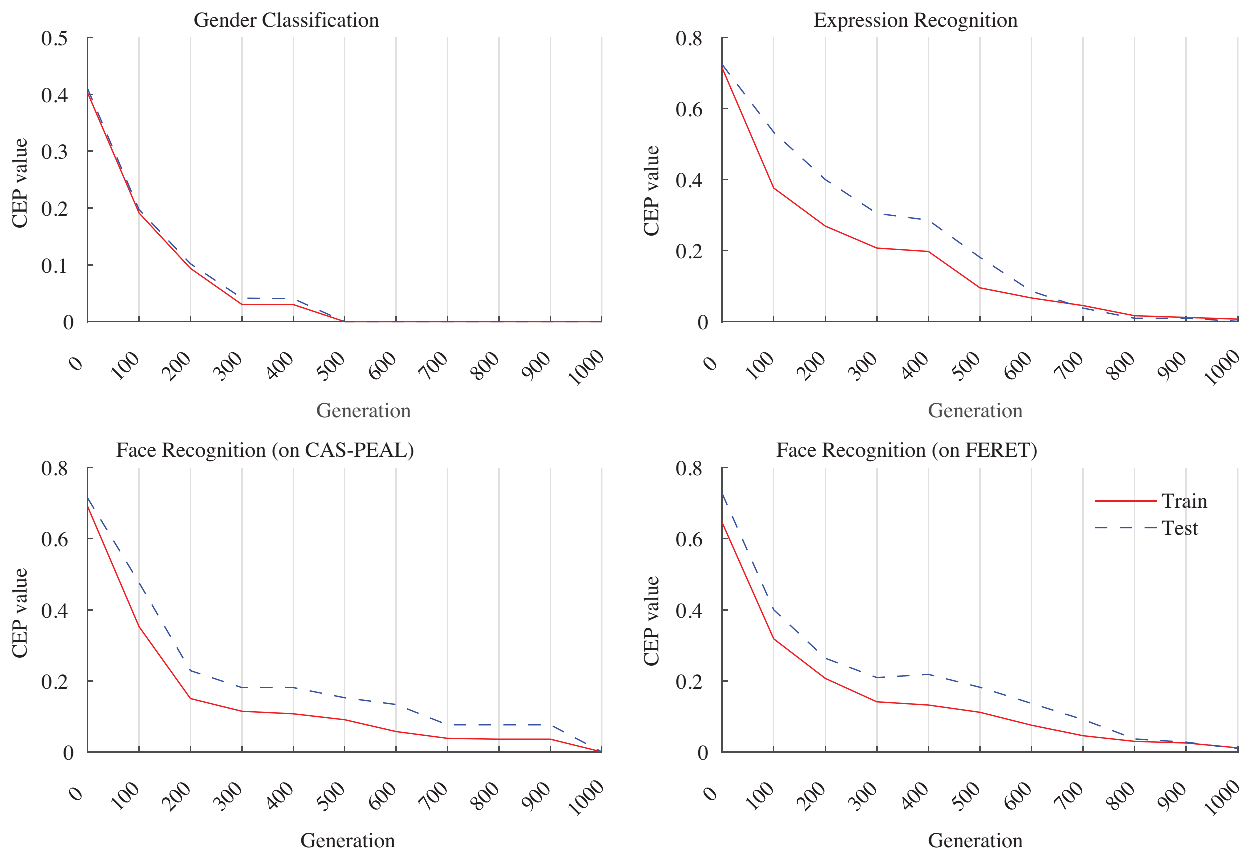

Figure 4. The overall convergence performances of the proposed framework measured throughout evolution on gender classification, expression recognition, and face recognition tasks (on both CAS-PEAL and FERET dataset).

the findings here also suggest that the GP has shown an overwhelmingly better performance than the CNN on overall. At first glance, it seems the overall classification performance of the evolved programs slightly degrades when compared to that of the gender classification task. However, that is not surprising because handling multilabel image classification would not be as easy as binary-label image classification.

The comparative overall convergence curves obtained at the midgenerations on facial expression recognition task show that, contrary to the observation on gender classification task, the GP has required at least 75% of the total evolution time so that it can evolve a generic classification program due to the problem's complexity. In brief, the curves in the figure suggest that setting 1000 generations for the evolution of classifier program would be fine to avoid under- and overfitting problems for this task.

5.3.3. Face recognition

Recognition of a face from a given image is regarded as the toughest task in comparison to both gender classification and facial expression recognition tasks. The classifier program evolved through GP should learn a generic hypothesis that is able to identify the characteristics from whole part of the faces that change from one to another. That is why the evolution of the classification model becomes too complex when compared to other tasks such as facial expression recognition where the model exploits the interest region of the image (e.g., the mouth to label as smiling).

The comparative mean CEP values obtained from the CAS-PEAL database (Table 2) conclude that the GP has shown a comparable performance with the CNN, but with an inferior classification accuracy of 94.10%. The CNN has shown an overall classification performance of 95.82% on this task. As for the findings obtained from FERET database. the GP has shown a better performance than the CNN with a classification accuracy of 97.09%. The CNN, however, has shown a classification accuracy of 96.25%. The comparative curves obtained from both the training and testing phases given in Figure 4 suggest that the overall classification rates approach zero in both training and testing phases just before training through GP is terminated.

Table 3. Comparative accuracy obtained by the GP and different strategies on various dataset/tasks.

Dataset/task	Method	Accr.	Dataset/task	Method	Accr.
CAS-PEAL (Gender)	Leng and Wang [33]	89.00	CAS-PEAL (Expression)	Bai et al. [34]	87.57
	Haider et al. [35]	97.30		Maturana et al. [36]	97.00
	Proposed study	100.00		Proposed study	99.05
CAS-PEAL (Face Recog.)	Liu et al.[37]	84.20	FERET	Mi et al. [38]	78.10
	Pang et al.[39]	85.33		Liu et al. [40]	79.00
	Zui and Liu [23]	92.38		Plichoski et al. [41]	80.50
	Sharma [42]	92.57		Alahmadi et al. [43]	96.09
	Umer et al. [44]	98.20		Umer et al. [44]	97.76
	Proposed study	94.10		Proposed study	97.09

To sum up, the GP has shown a better or at least highly competitive performance compared to the CNN algorithm on gender classification, facial expression recognition, and face recognition problems. Among these problems, the most obvious performance difference has been observed in facial expression recognition task. In this task, the GP has shown prominently better classification performance than the CNN algorithm. The gender classification and face recognition, however, have become tasks in the experiments where the GP and the CNN have shown very similar performances. Among them, the GP and the CNN have shown better classification performance on gender classification and face recognition tasks, respectively.

5.3.4. Literature comparison and efficacy

In this section, the comparison of the proposed method with the studies in the literature has been given for all tasks and then the efficacy of the GP and CNN methods has been evaluated here.

Compared to the existing studies in the literature, it can be observed that the proposed GP method achieves better results than the other studies in both gender classification and facial expression recognition tasks of the CAS-PEAL dataset. The comparative results are given in the upper block of Table 3.

As for the bottom block of the table, a comparison of the proposed method with the studies on both CAS-PEAL and FERET datasets in face recognition task is given. While the proposed method produces better results than the most existing studies, it becomes comparable to the study of Umer et al. [44].

In addition to the effectiveness, the efficacy of the proposed GP-based approach has also been investigated in this study. For this purpose, training and testing times taken by the GP and CNN have been measured. The comparative times are provided in Table 4. According to the results, the time taken by training and testing through GP is considerably less than that taken by CNN in all the tasks. However, it is important to note here that the GP requires a preprocessing step so that the dataset is prepared to be fed into the GP, whereas CNNs are able to learn from a raw dataset.

Table 4. Comparative running times (in s) measured in training and testing phases.

Approaches	Tasks (train/test)		
	Gender	Expression	Face recognition
CNN	322.589/0.8932	1844.4436/3.719	187.443/0.284
GP	24.212/2.26E-2	36.192/1E-3	5.007/4E-4

6. Conclusion

In this study, an evolutionary computation-based framework has been proposed to handle binary- and multilabel image classification tasks, i.e. gender classification, facial expression recognition, and face recognition. The tasks studied in this paper are very crucial in our daily lives. While gender classification is critical for many commercial applications supporting human-computer interaction, facial expression recognition and face recognition are very important in the public security, accessing the digital world, human-computer intelligent interaction (with artificial intelligence), and the like. The GP algorithm has been applied in the proposed framework to evolve a classifier program for every task handled. The performance of the proposed framework has been compared with that of the CNN algorithm. The achieved experimental findings show that the proposed framework can yield a detection rate of 100%, 99.05%, 94.10%, and 97.09% on gender classification, facial expression recognition, face recognition (CAS-PEAL), and face recognition (FERET) tasks, respectively. This suggests that the GP-based approach can perform a classification ability that is better or at least very competitive compared to the CNN algorithm in such tough tasks. In addition to these tasks, there are several tasks in the published CAS-PEAL database such as person recognition with various lighting conditions, matching people from photos taken at different times, and analyzing the effect of having accessories. Considering the high capability of the GP-based framework on binary- and multilabel image classification, it is worth stressing here that each of the these tasks can further be studied through GP, which may be considered a future direction of this study for the interested researchers.

References

[1] Parkhi OM, Vedaldi A, Zisserman A. Deep Face Recognition. In: British Machine Vision Conference; London, UK; 2015. pp. 1-20.

[2] Paul V, Michael J. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference On Computer Vision and Pattern Recognition; Kauai, USA; 2001. pp. I-511.

[3] Yaniv T, Ming Y, Marc'Aurelio R, Lior W. Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; San Diego, USA; 2005. pp. 1701-1708.

[4] Kaihao Z, Yongzhen H, Yong D, Liang W. Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Transactions on Image Processing 2017; 26 (9): 4193-4203. doi: 10.1109/TIP.2017.2689999.

[5] Gil L, Tal H. Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; Boston, USA; 2015. pp. 34-42.

[6] Liangliang C, Mert D, Yun F, Thomas HS. Gender recognition from body. In: Proceedings of the 16th ACM International Conference on Multimedia; New York, USA; 2008. pp. 725-728.

[7] Irene K, Ioannis P. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Transactions on Image Processing 2006; 16 (1): 172-187. doi: 10.1109/TIP.2006.884954

[8] Darwin C. On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life. London: John Murray, 1859.

[9] John H. Adaptation in Natural and Artificial Systems. Cambridge, MA, USA: MIT Press, 1992.

[10] Nichael LC. A Representation for The adaptive generation of simple sequential programs. In: Proceedings of the First International Conference on Genetic Algorithms; New York, USA; 1985. pp. 183-187.

[11] John RK. Genetic Programming: On the Programming of Computers by means of Natural Selection. Cambridge, MA, USA: MIT press, 1992.

[12] Ryan C, Collins JJ, Neill MO. Grammatical evolution: evolving programs for an arbitrary language. In: Banzhaf W, Poli R, Schoenauer M, Fogarty TC (editors). Genetic Programming. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 83-96.

[13] Le TA, Chu TH, Nguyen QU, Nguyen XH. Malware detection using genetic programming. In: 7th IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA); Hanoi, Vietnam; 2014. pp. 1-6.

[14] Estévez PA, Becerra YN, Boric N, Ramírez JA. Genetic programming-based voice activity detection. Electronics Letters 2005; 41 (20): 1141-1143. doi: 10.1049/el:20052475

[15] Le GC, Nguyen T, Forrest S, Weimer W. GenProg: a generic method for automatic software repair. IEEE Transactions on Software Engineering 2012; 38 (1): 54-72. doi: 10.1109/TSE.2011.104

[16] Ibrahem H, Nasef M, Mahmoud E. Genetic programming based face recognition. International Journal of Computer Applications 2013; 69 (27): 1-6. doi: 10.5120/12140-8187

[17] Bozorgtabar B, Noorian F, Rad GAR. A genetic programming approach to face recognition. In: IEEE GCC Conference and Exhibition; Dubai, UAE; 2011. pp. 194-197.

[18] Bozorgtabar B, Rad GAR. A genetic programming-PCA hybrid face recognition algorithm. Journal of Signal and Information Processing 2011; 2 (3): 170-174. doi: 10.4236/jsip.2011.23022

[19] Bozorgtabar B, Noorian F, Rad GAR. Comparison of different PCA-based face recognition algorithms using genetic programming. In: 5th International Symposium on Telecommunications; Kish Island, Iran; 2010. pp. 801-805.

[20] Li Y, Wang M, Cui L, Huang D. A new classification arithmetic for multi-image classification in genetic programming. In: International Conference on Machine Learning and Cybernetics; Hong Kong, China; 2007. pp. 1683-1687.

[21] Bhowan U, Zhang M, Johnston M. Genetic programming for image classification with unbalanced data. In: 24th International Conference Image and Vision Computing; New Zealand; 2009. pp. 316-321.

[22] Shao L, Liu L, Li X. Feature learning for image classification via multiobjective genetic programming. IEEE Transactions on Neural Networks and Learning Systems 2014; 25 (7): 1359-1371. doi: 10.1109/TNNLS.2013.2293418

[23] Zhi H, Liu S. Face recognition based on genetic algorithm. Journal of Visual Communication and Image Representation 2019; 58: 495-502. doi: 10.1016/j.jvcir.2018.12.012

[24] Mehmood Y, Ishtiaq M, Tariq M, Arfan JM. Classifier ensemble optimization for gender classification using genetic algorithm. In: International Conference on Information and Emerging Technologies; Karachi, Pakistan; 2010. pp. 1-5.

[25] Oncel T, Fatih P, Peter M. Region covariance: a fast descriptor for detection and classification. In: 9th European Conference on Computer Vision; Graz, Austria; 2006. pp. 589-600.

[26] Hong X, Chang H, Shan S, Chen X, Gao W. Sigma Set: A small second order statistical region descriptor. In: IEEE Conference on Computer Vision and Pattern Recognition; Florida, USA; 2009.

[27] Loveard T, Ciesielski V. Representing classification problems in genetic programming. In: Proceedings of the Congress on Evolutionary Computation; Seoul, Korea; 2001. pp. 1070-1077.

[28] Wen G, Bo C, Shiguang S, Xilin C, Delong Z et al. The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 2007; 38 (1): 149-161. doi: 10.1109/TSMCA.2007.909557

[29] Phillips PJ, Wechsler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing 1998; 16 (5): 295-306. doi: 10.1016/S0262-8856(97)00070-X

[30] Alex K, Sutskever I, Geoffrey EH. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (editors). Advances in Neural Information Processing Systems, USA: Curran Associates Inc., 2012, pp. 1097-1105.

[31] Huang G, Liu Z, Laurens VDM, Kilian QW. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Honolulu, USA; 2017. pp. 4700-4708.

[32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations; San Diego, USA; 2015. pp.1-20.

[33] XueMing L, Wang Y. Improving generalization for gender classification. In: 15th IEEE International Conference on Image Processing; Nanjing, China; 2008. pp. 1656-1659.

[34] Gang B, Wanhong J, Yang J. Facial expression recognition based on fusion features of LBP and gabor with LDA. In: 2nd International Congress on Image and Signal Processing; Tianjin, China; 2009. pp. 1-5.

[35] Zeeshan HK, Razzaq MK, Shehzad K, Tabassam T, Sohail J. Deepgender: real-time gender classification using deep learning for smartphones. Journal of Real-Time Image Processing 2019; 16 (1): 15-29. doi: 10.1007/s11554-017-0714-3

[36] Daniel M, Domingo M, Alvaro S. Learning discriminative local binary patterns for face recognition. In: 9th IEEE Conference on Automatic Face and Gesture Recognition; Santa Barbara, USA; 2011. pp. 470-475.

[37] Li L, Paul F, Guoying Z, Matti P, Dewen H. Extended local binary patterns for face recognition. Information Sciences 2016; 358: 56-72. doi: 10.1016/j.ins.2016.04.021

[38] Mi JX, Sun Y, Lu J, Kong H. Robust supervised sparse representation for face recognition. Cognitive Systems Research 2020; 62: 10-22. doi: 10.1016/j.cogsys.2020.02.001

[39] Meng P, Cheung YM, Binghui W, Risheng L. Robust heterogeneous discriminative analysis for face recognition with single sample per person. Pattern Recognition 2019; 89: 91-107. doi: 10.1016/j.patcog.2019.01.005

[40] Liu S, Li L, Jin M, Hou S, Peng Y. Optimized coefficient vector and sparse representation-based classification method for face recognition. IEEE Access 2020; 8: 8668-8674. doi: 10.1109/ACCESS.2019.2960928

[41] Plichoski GF, Chidambaram C, Parpinelli RS. A face recognition framework based on a pool of techniques and differential evolution. Information Sciences 2021; 543: 219-241. doi: 10.1016/j.ins.2020.06.054

[42] Poonam S. Face recognition under unconstrained environment for videos from internet. CSI Transactions on ICT 2020; 8 (2): 241-248. doi: 10.1007/s40012-020-00302-7

[43] Amani A, Muhammad H, Hatim A, Mansour Z. PCAPooL: unsupervised feature learning for face recognition using PCA, LBP, and pyramid pooling. Pattern Analysis and Applications 2020; 23 (2): 673-682. doi: 10.1007/s10044-019-00818-y

[44] Umer S, Dhara BC, Chanda B. Face recognition using fusion of feature learning techniques. Measurement 2019; 146: 43-54. doi: 10.1016/j.measurement.2019.06.008