Multimedia Tools and Applications (2019) 78:25753–25779
https://doi.org/10.1007/s11042-019-07811-x

Feature selection for facial emotion recognition using late hill-climbing based memetic algorithm

Manosij Ghosh[1] · **Tuhin Kundu**[2] · **Dipayan Ghosh**[1] · **Ram Sarkar**[1]

Received: 16 August 2018 / Revised: 10 May 2019 / Accepted: 21 May 2019 /
Published online: 1 June 2019

Abstract
Facial Emotion Recognition (FER) is an important research domain which allows us to provide a better interactive environment between humans and computers. Some standard and popular features extracted from facial expression images include Uniform Local Binary Pattern (uLBP), Horizontal-Vertical Neighborhood Local Binary Pattern (hvnLBP), Gabor filters, Histogram of Oriented Gradients (HOG) and Pyramidal HOG (PHOG). However, these feature vectors may contain some features that are irrelevant or redundant in nature, thereby increasing the overall computational time as well as recognition error of a classification system. To counter this problem, we have proposed a new feature selection (FS) algorithm based on Late Hill Climbing and Memetic Algorithm (MA). A novel local search technique called Late Acceptance Hill Climbing through Redundancy and Relevancy (LAHCRR) has been used in this regard. It combines the concepts of Local Hill-Climbing and minimal-Redundancy Maximal-Relevance (mRMR) to form a more effective local search mechanism in MA. The algorithm is then evaluated on the said feature vectors extracted from the facial images of two popular FER datasets, namely RaFD and JAFFE. LAHCRR is used as local search in MA to form Late Hill Climbing based Memetic Algorithm (LHCMA). LHCMA is compared with state-of-the-art methods. The experimental outcomes show that the proposed FS algorithm reduces the feature dimension to a significant amount as well as increases the recognition accuracy as compared to other methods.

Keywords Feature Selection · Late Acceptance Hill Climbing · Memetic Algorithm · Facial Emotion Recognition · RAFD · JAFFE

1 Introduction

Facial expressions constitute an important part of human communication, which help in galvanizing social interactions and relationships. There are seven universal facial

Code https://github.com/ManosijGhosh/Late-Hill-Climbing-based-Memetic-Algorithm-LHCMA

✉ Manosij Ghosh
manosij1996@gmail.com

Extended author information available on the last page of the article

expressions [1], validated by the scientists, namely *Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise*. According to Action Units, the expressions depend on the different positions of the muscles on the face [48] for an individual as explained by Ekman and Rosenberg in Facial Action Coding System [14]. There are 46 Action Units in Facial Action Coding System and each of which is related to a clearly defined set of muscle movements of the face [23]. Facial Emotion Recognition (FER) allows for monitoring of emotional state of people helping in mental state identification as well as in detection of mental disorders. Therefore, FER system has the ability to make human and robot interaction much more effective and meaningful.

Since decades, facial emotion of human beings has been utilized in various applications like understanding criminal behavior, mental state identification for treatment of people in depression, anxiety, etc. It also has applications in marketing for recognizing clients' emotional state, achieving synthetic human expressions, and many more. Nowadays, robots interact with kids having autistic conditions by interpreting their expressions. Usage of emotion-aware mobile applications has been increasing due to their smart features and user acceptability. Therefore design of real-time and accurate FER systems is a necessity.

Developing a FER system is an interesting as well as challenging research problem. Over the years, there are numerous research articles have been published in this domain. From the literature survey, we can safely claim that more focus has been put on identifying features from the facial images. Development of different feature descriptors and application of different feature set combination are the common trends in this domain. Increment of accuracy, in many cases, has been attempted through use of feature vectors with higher dimensions which contain more information about the image. This sometimes leads to the disadvantage of generating redundant and/or irrelevant features. This problem of "curse of dimensionality" [5], incurs higher memory requirements and computational cost during training of the model. Here comes the necessity of a proper feature selection (FS) method to obtain a more accurate model for prediction [55]. It is capable of optimizing the number of features extracted by different means thereby omitting the redundant and/or irrelevant features, which in turn helps in increasing the predictive capabilities of the model. Our paper aims to deal with such challenges to produce effective, optimized and discriminative facial representations in order to devise a real-time FER system.

To this end a new FS algorithm called Late Hill Climbing based Memetic Algorithm (LHCMA) has been proposed. The proposed algorithm has been evaluated using various popular feature descriptors with features extracted from varying image sizes, from two datasets namely RaFD and JAFFE. To account for the presence of images of varying quality in real-life FER systems, dataset images were resized to 32×32, 48×48 and 64×64 pixels. A number of texture as well as frequency based feature descriptors are used such as Uniform Local Binary Pattern (uLBP), Horizontal-Vertical Neighborhood LBP (hvnLBP), Gabor filter, Histogram of Oriented Gradients (HOG) and Pyramidal HOG (PHOG). So, in total 30 feature sets are used to evaluate the FS capacity of our proposed algorithm – LHCMA.

Contributions of this paper:

- A novel local search technique called Late Acceptance Hill Climbing through Redundancy and Relevancy (LAHCRR) has been proposed.
- LAHCRR is used as local search in MA to form Late Hill Climbing based Memetic Algorithm (LHCMA).

- Popular features like Uniform Local Binary Pattern (uLBP), Horizontal-Vertical Neighborhood Local Binary Pattern (hvnLBP), Gabor filters, Histogram of Oriented Gradients (HOG) and Pyramidal HOG (PHOG) are extracted from images.
- Two popular FER datasets, namely RaFD and JAFFE are used for experiments.
- Image sizes of 32 × 32, 48 × 48 and 64 × 64 are taken to account for variation in image quality.
- Comparison of LHCMA with state-of-the-art methods proves it performs better.

2 Literature survey

Till date, many FER systems consisting of different feature descriptors and FS methods have been proposed in the literature. Of these, some of them are briefed here. Section 2.1 mainly focuses on the feature descriptors that have been used in recognition of facial expressions, while section 2.2 reports the existing works on FS methods in this field.

2.1 Feature extraction techniques

FER has been extensively investigated during past few decades, and a lot of methods focusing on the feature descriptors have been proposed. The authors of [1] have performed face detection by Viola Jones algorithm and then smoothed the invalid regions (like hair) and excluded regions near to ear. Sobel edge detector is applied for detecting edges and an improved LBP has been used to get the collection of histograms from all the sub-blocks. In [55], 59 Local Binary Pattern Histogram labels are estimated from each of the sub regions of facial image. Discriminative LBP Histogram bins are made to learn by adopting the AdaBoost classifier and the top 200 features are selected by spatial distribution. Finally, FER has been performed using the selected LBP Histogram bins with Support Vector Machine (SVM) classifier. Makhmudkhujaev et al. [40] have proposed a new feature descriptor called Local Prominent Directional Pattern. For FS they computed the occurrence histogram of the features that helps to identify frequent expression-related features.

Gradient Local Ternary Pattern proposed by Faisal Ahmed et al. in [2] uses local texture features of LTP or Local Ternary Pattern which is enhanced by the Gradient encoding scheme. Gradient Local Ternary Pattern histograms are computed from each of the regions in the image and are concatenated to obtain a spatially combined histogram. They have used SVM for classification. In [67], Zhao et al. have proposed that Gabor multi-orientation fusion histogram and spatial-temporal motion LBP give a better result when facial points and locations are tracked. SVM with RBF kernel is used for classification.

In [17], estimation of facial landmarks positions is done which is followed by an exhaustive search technique to get the local face region. Geometric features, LBP and Normalized Central Moments are extracted from cropped images of face and classified using SVM. In [8], the face is detected and ellipse fitting is performed on the face blob. SVM classifier is used on HOG features extracted from the registered faces using a coherent spatial reference. In [50], a geometrical fuzzy approach extracted using Action Units, appearance and geometrical parameters from eye and mouth regions. Datasets CK+, JAFFE and ISED are used. On the same lines authors in [4] have extracted normalized distance signature and normalized texture signatures of landmarks from Active Appearance Model based grids. Statistical measures and stability indices of the signatures are also measured. These features are tested on CK+, JAFFE, MMI and MUG datasets.

The method proposed by [59] detects faces and crops them using AdaBoost face detection algorithm, from which the Weber Local Descriptor and HOG features are calculated. Then the chi-square distance is calculated using the weighted fused histograms of both the descriptors and the closest neighbor methodology is used for classification. The FER system proposed by Dhall et al. [11] uses the Pyramidal HOG (PHOG) and Local Phase Quantization (LPQ) to compute features. Key frames are selected using K-means clustering on normalized shape vectors derived from the constraint local model. Later for classification, SVM and *Large Margin Nearest Neighbor* are used. In a feature hybridization approach Wang et al. [60] extracted Scale-Invariant Feature Transform features along with features extracted from the different levels of a Convolutional Neural Network model. They used this feature extraction on CK+ dataset. Wang et al. in [61] have used stationary wavelet entropy as their feature descriptor and classified the features using a feed forward neural network trained with the help of Jaya algorithm.

2.2 Feature selection techniques

FS helps to improve recognition accuracy through the removal of redundant and irrelevant features while it also decreases the training time of classifier. Poor selection of features may lead to inaccurate recognition; hence it is essential that the information contained within the feature set is sufficient enough to predict the correct input class. In general, FS is done by three different approaches - filter [31, 44], wrapper [29, 41] and hybrid [18, 19, 24]. Filter methods perform FS using intrinsic data properties, while not using any learning algorithms. Wrapper methods require a learning algorithm and therefore display better performance but also take much higher computation time. Hybrid methods are the combination of wrapper and filter methods and outperform both of its counterparts.

Keeping the importance of FS in mind, researchers have paid attention to apply the same in the domain of FER. Wrapper method based on Genetic Algorithm (GA) [32] has been used to perform FS on log-Gabor features in FER. PHOG is extracted which is then refined using GA in [15]. FS using Linear Programming has been applied in [13], where a 216-dimensional feature vector generated using 18 Gabor filters from facial images. Classifiers such as SVM, Bayes and AdaBoost are used. In [33] the features - Gabor filters, log Gabor filter, LBP operator, higher-order local autocorrelation and higher-order local autocorrelation -like features are extracted and Mutual Information (MI), Maximization of Relevance and Minimization of Redundancy (mRMR) and GA are performed on them. Cascade of fixed filters and trainable non-linear 2-D filters, which are based on the biological mechanism of shunting inhibition is used to extract features and then FS is done using MI and class separation score [36]. A unique approach is used in [62], where a self-learning attribute reduction algorithm is used for FS. It is based on rough set and domain-oriented data-driven data mining 3DM. In [65], Gabor filtered image is taken for generating primitive features and then those are fed into Genetic Programming to generate composite features. Thereafter, fitness values are evaluated after running the SVM classifier. In a different approach, Grey Wolf Optimization is used in [57] to select the optimal feature subset as well as to find the weights in Neural Network using Grey Wolf Optimization.

Memetic Algorithm (MA) [68] is an improvement over GA due to the incorporation of a local search, which allows for a more rigorous exploitation ability. There have been limited instances in the past decade of research on FS in FER. Hence our objective is to highlight the importance and contribute a novel *hybrid FS method* based upon MA. Keeping in mind the fact that hybrid FS methods are the combination of both filter and wrapper methods, which also tend to outperform both individually, in this work, we have proposed a MA based hybrid FS method. Redundancy and relevancy are important criteria for estimating the importance of a feature. The local search used in MA as proposed in [68] is based on relevancy of a feature. Therefore, there is no way to check if the feature is redundant or not i.e. there is no check on the subset's discriminative ability. This shortcoming is addressed in our proposed algorithm through the incorporation of mRMR. Traditionally, mRMR uses Hill Climbing approach to accept the newly created solutions. However, this technique is ill equipped for overcoming local optima. This very problem is addressed in Late Acceptance Hill Climbing (LAHC), which has been incorporated in the proposed algorithm. Our algorithm has been evaluated on FER experiments using two popular datasets, namely Japanese Female Facial Expression (JAFFE) and Radboud Faces Database (RaFD) while generating the feature sets for the datasets using five well-known feature descriptors - uLBP, hvnLBP, Gabor filters, HOG and PHOG. The experimental results lead to the conclusion that our algorithm can successfully reduce the feature dimension to a significant extent as well as increase the expression recognition accuracy.

From the literature survey it can be concluded that FS is an important part of FER. Selection of subsets by considering redundancy and relevancy of features can lead to a better performance [16]. This has motivated us to develop a local search mechanism which can be included in MA to form LHCMA. The evaluation of our FS model is done on a number of FER based feature descriptors and compared with state-of-the-art FS algorithms.

3 Dataset and preprocessing

In this work, we have considered two standard datasets, JAFFE and RaFD, which include image samples taken in restricted laboratory environments. The datasets used, JAFFE and RaFD are described in sections 3.1 and 3.2 respectively. In section 3.3, the preprocessing done before feature extraction is described.

3.1 JAFFE

The JAFFE database [38] includes image samples collected from 10 female Japanese models having 213 image samples containing 7 facial expression classes (6 basic facial expressions which are angry, disgust, happy, fear, sad and surprise, and 1 neutral facial expression).

The original samples per facial expression class in the dataset are non-uniform. Hence to balance the number of images per class within the dataset, Gaussian white noise of constant mean and variance is added to 11 image samples within JAFFE database and included to the respective facial expression classes to make the total number of image samples per facial expression 32. This makes the total images in the dataset 224(=32 × 7). Sample images are provided in Fig. 1(a-g).

3.2 RaFD

RaFD [34] contains facial images of 67 models (Moroccan Dutch males, Caucasian males and females, and Caucasian children, both boys and girls) having 8 facial expression classes (angry, disgust, happy, fear, neutral, sad, surprise and contempt). Each facial expression of

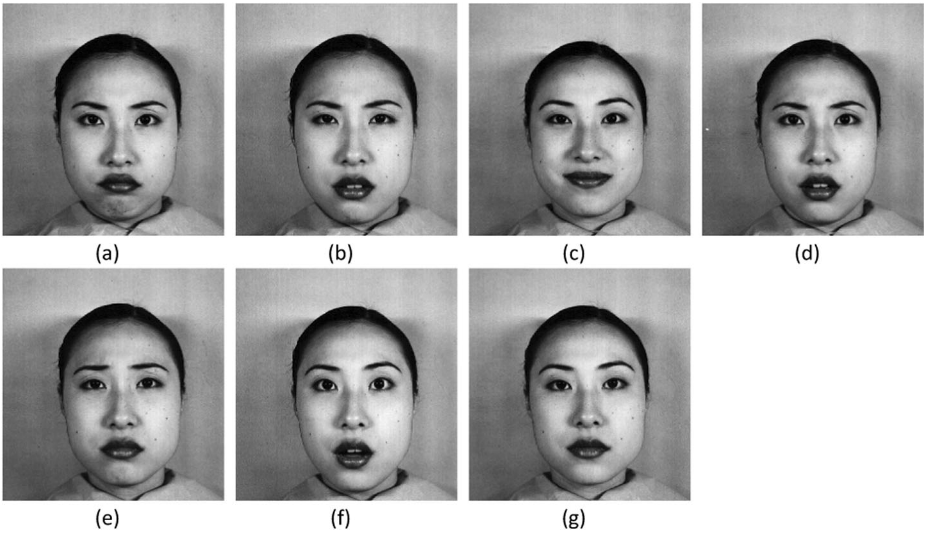

Fig. 1 Sample images of JAFFE dataset showing different emotions anger (**a**), disgust (**b**), happy (**c**), fear (**d**), sad (**e**), surprise (**f**) and neutral (**g**)

each model has been recorded in 3 gaze directions (left, right and frontal) and each image has been recorded simultaneously from 5 different camera angles. Each facial expression class contains 201 images from a single camera angle, hence total number of samples used for our facial expression classification model is $1608(= 201 \times 8)$ images. Some sample images are provided in Fig. 2(a-h).

3.3 Face detection and pre-processing

The fundamental challenge is to extract features which are relevant to any model and to remove or get rid of the unnecessary surface area inside every image sample excluding the facial area. Hence, the Viola Jones Haar cascade classifier [58] is used to detect the region of interest, that is, the region containing human facial features such as eyes, nose, lips, eyebrows and provide us with the coordinates of the said region along with its dimension.

To account for the variation in image quality in real-world scenarios, we conduct our experiments considering different pixel dimensions of the image such as 32×32, 48×48 and 64×64. In doing so, we have resized the images to each one of the three resolutions. We reiterate the fact that the image samples at this stage are already in grayscale format.

4 Feature extraction

Feature descriptor is a representation of any image that simplifies the image by extracting useful information and disregarding the extraneous information. Several feature descriptors have been developed over the years such as Speeded-Up Robust Features, Scale-Invariant Feature Transform, HOG, LBP, which have found numerous utilities in applications of computer vision.

In our model implementation, we have taken a few feature descriptors namely grid based uLBP, hvnLBP, Gabor filter-based descriptor, HOG and PHOG on samples taken from the

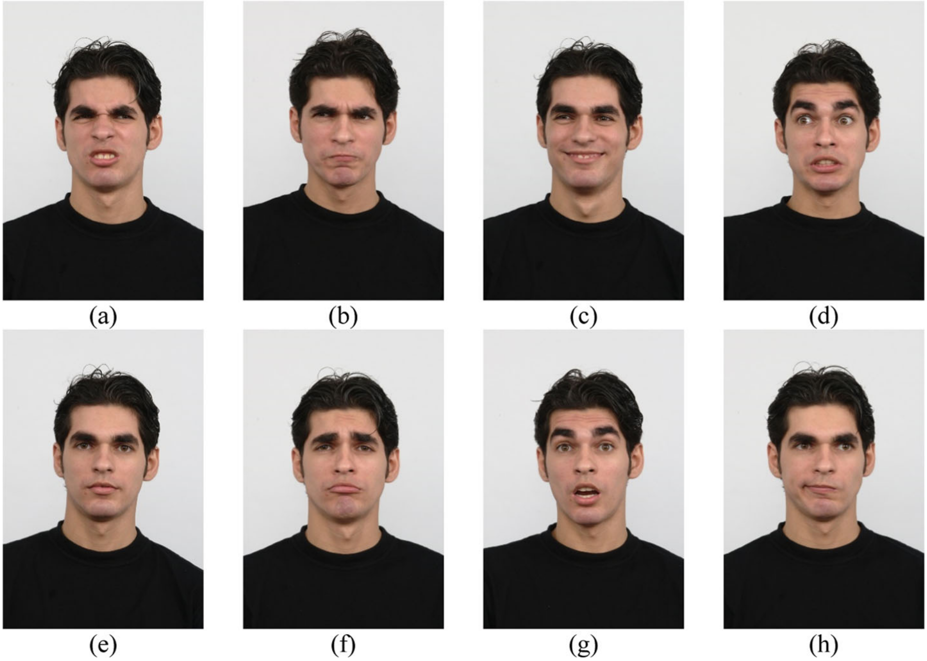

Fig. 2 Sample images of RaFD dataset showing different emotions anger (**a**), disgust (**b**), happy (**c**), fear (**d**), neutral (**e**), sad (**f**), surprise (**g**) and contempt (**h**)

JAFFE and RaFD datasets. In this section, we discuss the feature extraction processes applied on the facial image databases. In total five feature descriptors have been used and they are described in sections 4.1–4.5.

4.1 Uniform local binary pattern

Ojala et al. [46] have introduced a texture descriptor called LBP. It is very powerful as a discriminative texture operator. The operator used here takes a 3 × 3 window of pixels and the center pixel is the threshold value for each window as the radius taken is 1. A comparison of gray values is done for pixel (x_{center}, y_{center}) with its neighborhood pixels. The resulting LBP is expressed as Equation (1).

$$LBP(x_{center}, y_{center}) = \sum_{k=0}^{7} new(i_k - i_{center})2^k \tag{1}$$

Where k runs over the 8 neighbors of the central pixel, i_{center} and i_k are the gray values of the central pixel and the surrounding pixels respectively, and new(x) is 1 if $x \geq 0$ and 0 otherwise.

The top left corner is taken as the 7^{th} bit and the bits are considered in a clockwise fashion until it reaches to the 0^{th} bit. The resultant 8 bit binary number is formed which is then converted to its equivalent decimal number. This decimal number replaces the threshold value [45]. From Fig. 3, we can justify that the resultant pixel values with 51, 3, 11, and 65 are lesser than that of the center pixel value 77. Hence, the resultant pixels in that cell are 0. The rest pixel values like 97, 101, 212, and 255 are greater than that of the center pixel value 77 henceforth, the resultant pixels in those cells would set to 1. In clockwise direction, the binary string forms the decimal number 116, which would be the new center pixel in the LBP image as shown in Fig. 3.

By this, we would be able to generate a 256-bin histogram of the LBP labels computed over a region. The histogram of the labelled image $Image_{label}(x, y)$ using LBP operator can be defined as Equation (2)

$$Hist_n = \sum_{x,y} I(Image_{label}(x, y) = i), \quad i = 0, \ldots, L{-}1 \tag{2}$$

Here L is the number of different labels produced by the LBP operator and has a maximum value of 255. $I(A)$ is 1 if A is *true* and 0 otherwise. This is the basic LBP. Henceforth, LBP is extended by uLBP. LBP is said to be 'uniform' if it contains at most two $0-1$ or $1-0$ bit transition when viewed in a clockwise direction. It is observed that uniform pattern accounts for nearly 90% of all patterns in the $(8, 1)$ neighborhood and for about 70% in $(16, 2)$ neighborhood in texture images [66].

This uLBP has been used with several combinations to bring out the best features for FER. In [23, 26, 54], we can get an overview of the importance of uLBP for estimating proper textures from faces for recognizing facial expressions. The resized images are divided into non-overlapping 4×4 sub-blocks from which 944 uLBP extracts features.

4.2 Horizontal - vertical neighborhood local binary pattern

Mistry et al. [43] have proposed a new texture descriptor called hvnLBP for FER. It is very powerful for facial expression analysis as it improves the discriminative abilities of LBP. It obtains better contrast information among the neighborhood pixels such as edges and corners. In order to produce the discriminative facial representation, it is integrated with the 2-D Gabor filter. A total of 16 magnitude pictures of various wavelength and orientation from Gabor filter are obtained at the beginning of the process upon which hvnLBP operator is applied. Similar to uLBP, the operator used here takes a 3×3 window of pixels and the center pixel is the threshold value for each window as the radius taken is 1. A comparison of gray values is done for pixel (x, y) with its neighborhood pixels. From Figure 4, we can understand that the eight neighborhood pixels in *hvnLBP* are represented as L= $\{l_0, l_1, l_2, l_3, l_4, l_5, l_6, l_7\}$. The resulting hvnLBP is expressed as Equation (3 a-b).

$$\begin{aligned}\mathrm{hvnLBP}_{pixel,radius}(\mathrm{x,y}) = \Big\{ & C(\max(l_0, l_1, l_2)), C(\max(l_7, l_3)), \\ & C\,(\max(l_6, l_5, l_4)), C(\max(l_0, l_7, l_6)), \\ & C(\max(l_1, l_5)), C(\max(l_2, l_3, l_4)) \Big\}\end{aligned} \tag{3a}$$

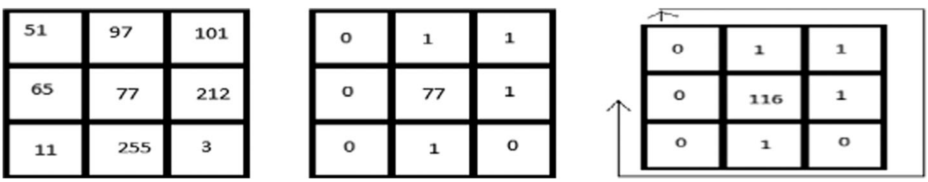

Fig. 3 Clockwise direction gives $(01110100)_2 = (116)_{10}$ as the center pixel after LBP computation

Where the comparison operation is denoted by C.

$$C\left(\max\left(l_j, l_k, l_m\right)\right) = \begin{cases} 1 & \textit{if maximum} \\ 0 & \textit{otherwise} \end{cases} \tag{3b}$$

Note: l_k is absent when the horizontal or the vertical neighborhood taken has no center pixels included.

Similar to uLBP, the clockwise direction is taken, then resultant decimal number would be 39, which would be the new center pixel in the hvnLBP image shown in Figure 4.

By this we would be able to generate a 256 bin histogram of the hvnLBP labels computed over a region. Similar to uLBP, the histogram of the labelled image $Image_{label}(x, y)$ using hvnLBP operator can be defined as Equation (2).

4.3 Gabor filter

Gabor filters are used due to its properties of optimal joint spatial or the spatial frequency localization or stimulation of receptive fields of simple cells in the visual cortex [10, 22, 27], while being governed by the "Uncertainty principle" [30]. The advantages of Gabor filters include their invariance to rotation, scale and translation while showing robustness against photometric disturbances, such as image noise and illumination changes [28, 37, 42, 56]. Two dimensional Gabor filter involves the sinusoidally modulated Gaussian kernel function in the spatial domain depicted by Equation (4 a-c) [21] where the Gabor filter based features are directly extracted from the grey level images.

$$G(x, y) = \frac{f^2}{\pi\gamma\eta}\, exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \times exp\left(j.2\pi f x' + \phi\right) \tag{4a}$$

where

$$x' = x\cos\phi + y\sin\phi \tag{4b}$$

$$y' = -x\sin\phi + y\cos\phi \tag{4c}$$

ϕ is the phase offset, σ is the standard deviation of the Gaussian expression, γ specifies the Gabor function ellipticity and denotes the spatial aspect ratio, θ denotes the orientation of normal to parallel stripes of the Gabor function and f is the frequency of the sinusoid.

In the Gabor model, we have taken up 5 different scales and 8 different orientations which make up 40 different Gabor filters. The facial image samples are convolved using multiple spatial resolution and orientation set of the two-dimensional Gabor filter bank. For example,

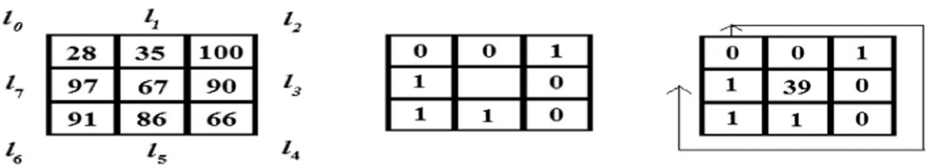

Fig. 4 Clockwise direction gives $(00100111)_2 = (39)_{10}$ as the center pixel after evaluation

given an image sample $I(x, y)$ and Gabor filter kernel $\Psi_{u,\,v}(x, y)$, characterization of the image sample $O_{u,\,v}(x, y)$ can be denoted as in Equation (5) [47].

$$O_{u,v}(x,y) = I(x,y).\Psi_{u,v}(x,y) \tag{5}$$

Hence the feature vector generated by the Gabor filter, for example, for a 32 × 32 pixel image is 40960 (refer Table 1) which is later reduced by down sampling by a factor of 8, while being normalized to zero mean and unit variance. Similarly, image samples of size 48 × 48 and 64 × 64 are processed using the Gabor filter-based descriptor, and the calculations for their feature vector size are given in Table 1.

4.4 Histogram of oriented gradients

Many problems such as object identification, pedestrian identification and human detection have found practical usage of the HOG descriptor proposed by Dalal and Triggs [9]. HOG of an image is calculated using the orientation information and the gradient information corresponding to the first derivative of the image [59]. Gradient of an image $I(x, y)$ about any arbitrary pixel point is a vector that can be defined as in Equation (6).

$$\nabla I(x,y) = \left[G_x, G_y\right]^T = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right] \tag{6}$$

where G_x is the gradient of the image in X direction and G_y is the gradient of the image in Y direction with the direction and magnitude of the gradient denoted by Equation (7) and Equation (8) respectively.

$$\theta(x,y) = arctan\frac{G_y}{G_x} \tag{7}$$

$$|\nabla I(x,y)| = \sqrt{G_x^2 + G_y^2} \tag{8}$$

The image is divided into various cells. Many adjacent cells in an image patch make up a block where the block frames are overlapping in nature. The HOG descriptor is applied to each block and the gradient and orientation information are extracted. Orientations relating to the same cell are quantized and integrated through histogram bins which are then sorted and arranged into a final histogram. N_{HOG}, which is the number of features generated using the HOG descriptor and can be calculated using Equation (9) where B_s denotes block size, B_i is the

Table 1 Calculation of feature dimension for feature sets generated using Gabor filter bank comprising 5 scales and 8 orientations with down sampling by factor of 8

Image dimension (in pixels)	Feature dimension	Feature dimension after down sampling
32 × 32	40960 (=40 × 32 × 32)	$640\left(= \frac{40960}{8\times8)}\right)$
48 × 48	92160 (=40 × 48 × 48)	$1440\left(= \frac{92160}{8\times8}\right)$
64 × 64	163840 (=40 × 64 × 64)	$2560\left(= \frac{163840}{8\times8}\ \right)$

number of block present in the taken image sample and N_b which is the number of histogram bins used.

$$N_{HOG} = B_s \times B_i \times N_b \tag{9}$$

In our experiment, we have a cell size of 8 × 8 with block size being 2 × 2 with 50% overlap while using 9 histogram bins to integrate orientation information into the feature space.

4.5 Pyramidal HOG

PHOG feature descriptor [6] is widely used for object recognition and it computes the spatial pyramid representation of the HOG descriptor. It generates the local shape of the input image and preserves its spatial information by segmenting it into different levels and applying the HOG descriptor to each of these levels, while also stratifying the layers.

The edge contours of every region are extracted using the Canny edge detector. By doubling the divisions in every axis direction, a sequence of minute spatial grids is formed for every region. The grid has 2^l cells for each resolution level $L = l$ along every dimension. A Sobel mask of 3 × 3 is then used for computing the orientation gradients on the edge contours. At this stage, the process of gradient binning takes place as in the HOG descriptor, where the edge orientations relating to the same cell are quantized and integrated in N histogram bins which are sorted and concatenated into a single sequence within the same level. The orientation for binning is performed using either $[0 - 180\ \textit{degree}]$ range (where the contrast sign is neglected) or the $[0 - 360\ \textit{degree}]$ range, where all the orientation gradients are used [11, 35].

In our work, we have used the PHOG descriptor to get edge contours from segmented regions to capture facial features and generate feature sets for the images of the JAFFE and RaFD datasets for all three image dimensions, 32 × 32, 48 × 48 and 64 × 64. The feature sets generated using PHOG descriptor having $L = 3$ number of pyramids, $N = 8$ number of histogram bins and orientation range as $[0 - 360\ \textit{degree}]$. The dimension of the generated feature set is calculated by $N * \sum_{l \in L} 4^l$, hence resulting into a feature vector of size 680 for the parameters used in our experiment.

5 Proposed work

Memetic Algorithm or MA improves upon traditional Genetic Algorithm or GA [20] to build a hybrid method. The genetic operations of crossover and mutation are accompanied with local search. GA, though widely used, suffers from poor local search capabilities. This inability is addressed in MA through the introduction of a local search technique. In MA, the population is usually randomly or heuristically initialized, after which the fitness of each individual is improved using genetic operations and local search. This makes local search an important component of MA. In this work, an improvement in the local search of MA has been proposed.

Local search can be of three types – *improvement first*, *greedy strategy* and *sequential strategy*. Local search in MA is then used to improve the population in each iteration. The features are ranked offline using a filter method like ReliefF [31, 63]. In the greedy approach, random pair of numbers (k, d) is generated such that both k and d are less than $n/20$, where n is the feature dimension. The top ranked k features (according to ReliefF rankings) that are not included in the chromosome are inserted and d least ranked features from the chromosomes are removed. The generated chromosome is compared with the original one and if the chromosome is a better performer then we substitute the original in the population. Both the other two approaches are computationally more expensive than greedy. In this work, we try to improve the greedy search by taking into consideration a more comprehensive evaluation of new subsets using a filter method to achieve less computational complexity (compared to using a classifier). Moreover, a more unique scheme for accepting a solution is adopted to give chromosomes a better chance to improve.

The local search technique of MA generally uses an offline univariate ranking of features to determine quality of a feature. This approach implies that each feature is evaluated individually with respect to the class labels alone. Not only does this constraint the search ability of the local search but also lacks the ability to check for feature redundancy i.e. evaluate the feature with respect to other selected or non-selected features in the feature subset. Many feature rankings [31] suffer from this problem. The evaluation of the feature subset using a filter method is also widely done, of which mRMR [49] is one of the most popular techniques [12, 51–53]. mRMR works by maximizing the Relevancy – *Rel* (Mutual Information or MI [25] between the class and the features) and minimizing the Redundancy – *Red* (average MI between the features selected). MI is calculated using Equation (10). *Rel* and *Red* are calculated using Equations (11) and (12) respectively. x_i denotes a feature vector, c is the class lables and S denotes a feature subset. The value of MI with respect to class and the other features are calculated offline and stored, making it a one-time computation.

$$MI(A;B) = \sum_{b\in x_i}\sum_{a\in x_j} P(b,a) log P(b,a)/(P(a)*P(b)) \tag{10}$$

$$Rel(S) = \frac{1}{|S|}\sum_{i=1}^{|S|} MI(x_i,c) \tag{11}$$

$$Red(S) = \frac{1}{|S|^2}\sum_{i=1}^{|S|}\sum_{j=1}^{|S|} MI\left(x_i,x_j\right) \tag{12}$$

The mRMR works using Equation (13) to find *Value _ mrmr*(*S*) which denotes the goodness of the subset (*S*) and we try to maximize it.

$$Value\_mrmr(S) = Rel(S) - Red(S) \tag{13}$$

In FS using mRMR [12, 51–53] generally mimics a Hill-climbing [3] approach where a new feature subset is accepted only if the accuracy increases. Instead the use of Late Hill Climbing Approach (LAHC) [7] is proposed here because of its ability to include a lower performing subset in order to overcome a local optima. In LAHC, an improved solution is immediately accepted while L_H number of worse solutions is stored to allow for acceptance of slightly poor solutions, so as to improve them into better ones. LAHC works by mutating the chromosome and checking if it allows for accepting the mutated chromosome. Using LAHC we try to optimize the value of *Value _ mrmr*(*S*). Here, mRMR is used to check the goodness of the chromosomes. As mRMR is a filter method, the computation cost of using it in LAHC is far lower than a local search approach which requires the use of a classifier. This combination is named Late

Acceptance Hill Climbing through Redundancy Relevancy (LAHCRR). The algorithm for local search is provided in *Algorithm 1 - LAHCRR*. *iter* is the number of iterations for which the local search is allowed to continue. The new subset generated from the *LAHCRR* is evaluated using the classifier and the set (chromosome) is substituted in the population with the new set if it is deemed to be better.

$\boldsymbol{Algorithm\ 1 - LAHCRR}$

$start$
$//\ S\ is\ the\ set\ of\ features$
$value\_array = array\ of\ size\ L_H$
$all\ elements\ of\ value\_array = Value\_mrmr(S)$
$for\ i = 1\ to\ iter$
 $S' = Mutation(S)$
 $temp = Value\_mrmr(S')$
 $if(Value\_mrmr(S) < temp)$
  $S = S'$
 $end\ if$
 $pos = i\ mod\ L_H$
 $if(value\_array[pos] > temp)$
  $S = S'$
 $end\ if$
 $if(value\_array[pos] < temp)$
  $value\_array[pos] = temp$
 $end\ if$
$end\ for$
end

It has already been said that local search in MA is a crucial part for exploitation. The use of LAHCRR enhances this ability of MA instead of adopting the static ranking based local search in [68]. The use of LAHCRR enables a comprehensive search of the feature space as it allows for subset evaluation instead of feature evaluation. The steps followed in Late Hill Climbing based Memetic Algorithm or LHCMA are similar to that of MA. First a random population is created and the chromosomes in the population are evaluated to determine their fitness. In an iterative process, each chromosome undergoes local search (by LAHCRR). Thereafter, selection from the population is done using a roulette wheel. On the selected chromosomes uniform crossover followed by uniform mutation is done.

Population of LHCMA is created by forming chromosomes which are binary encoded strings. Chromosomes represent a selected feature subset. '1' at the i^{th} position denotes that the i^{th} feature is selected and '0' implies otherwise. The population denotes the current set of solutions and the initial population of chromosomes is generated randomly.

LHCMA creates children through crossover and mutation. Here, two parents are selected using roulette wheel. The roulette wheel is built using the accuracies of the feature subsets therefore parents with higher accuracy have more probability of passing on their features to the next generation. Uniform crossover (*Algorithm 2 - Crossover*) is performed on the parents to form two children, with a probability of $p = 0.5$. On the children uniform mutation (*Algorithm 3 - Mutation*) is performed thereafter. Mutation is done with a probability of $q = 0.03$. The children are then compared to the least fit of the parents and if the children are deemed better than them, then they substitute the chromosomes in the population. n is the size of the feature vector.

$\boldsymbol{Algorithm\ 2 - Crossover}$

$start$
    $//\ the\ parents\ selected\ using\ roulette\ wheel\ are - g^1\ and\ g^2$
    $for\ i = 1\ to\ n$
        $if\ (rand() < p)$
            $exchange\ the\ values\ of\ g_i^1\ and\ g_i^2$
        $end\ if$
    $end\ for$
$stop$

$\boldsymbol{Algorithm\ 3 - Mutation}$

$start$
    $the\ child\ is\ g$
    $for\ i = 1\ to\ n$
        $if\ (rand() < q)$
            $g_i = 1 - g_i$
            $//\ g_i\ is\ the\ i^{th}\ feature\ in\ g$
        $end\ if$
    $end\ for$
$stop$

The fact that a chromosome is better than another is determined using a comparator scheme, where we compare the accuracy of the two chromosomes. We set a tolerance value of $\delta = 1$. If the difference is more than that, the chromosome with higher accuracy is deemed to be better, otherwise Equation (14) determines which chromosome is better (higher value denotes better chromosome). This allows us to retain feature sets which possess better accuracy while also trying to optimize the number of features. The weight of accuracy is taken as 4 to make it more important than feature dimension (which is given a weight of 1).

$$value = accuracy{*}4 + ratio\ of\ used\ features\ to\ total \tag{14}$$

The time complexity of LHCMA is proportional to the size of the population (*popCount*) and the number of iterations (*iter*). Since a filter method – mRMR has been used for evaluating the new subsets in LAHCRR, the time complexity of the search is negligible to the using a classifier for the same. The computation requirement of LHCMA is $O(popCount * iter * classifier)$. The variable *classifier* represents the time complexity of the classifier used for determining accuracy of the feature subset. As there is no dependency between our algorithm and classifier, any classifier can be chosen based on experimental requirements.

The model used for performing emotion recognition is to first process the data given using the steps defined in section 3.3. Then we extract features using the features described in section 4 from the training set. Any of the 5 described features can be extracted. Thereafter, using the extracted features we perform FS to determine the optimal set of features. Features selected in the FS step is then extracted from the test set. A classifier is trained using the train data (using only selected features) and then the test data is used to test the classification ability of the selected features. A block diagram of our work is provided in Figure 5.

6 Experimental Results and Discussion

As mentioned earlier, the proposed FS method is applied on images of two standard FER databases namely JAFFE and RaFD. To get the optimized feature vector, we have considered five feature descriptors namely uLBP, hvnLBP, Gabor filter, HOG and PHOG. Moreover, all five feature descriptors are extracted from images considering three different dimensions of the images, that are 32×32, 48×48 and 64×64. So, in total there are 30 ($=5 * 3 * 2$) feature sets which have been extracted and tested.

Parameter selection is done through experimentation i.e. parameters values for which the maximal results are obtained are chosen in our work. The various parameters selected and their corresponding values are given in Table 2.

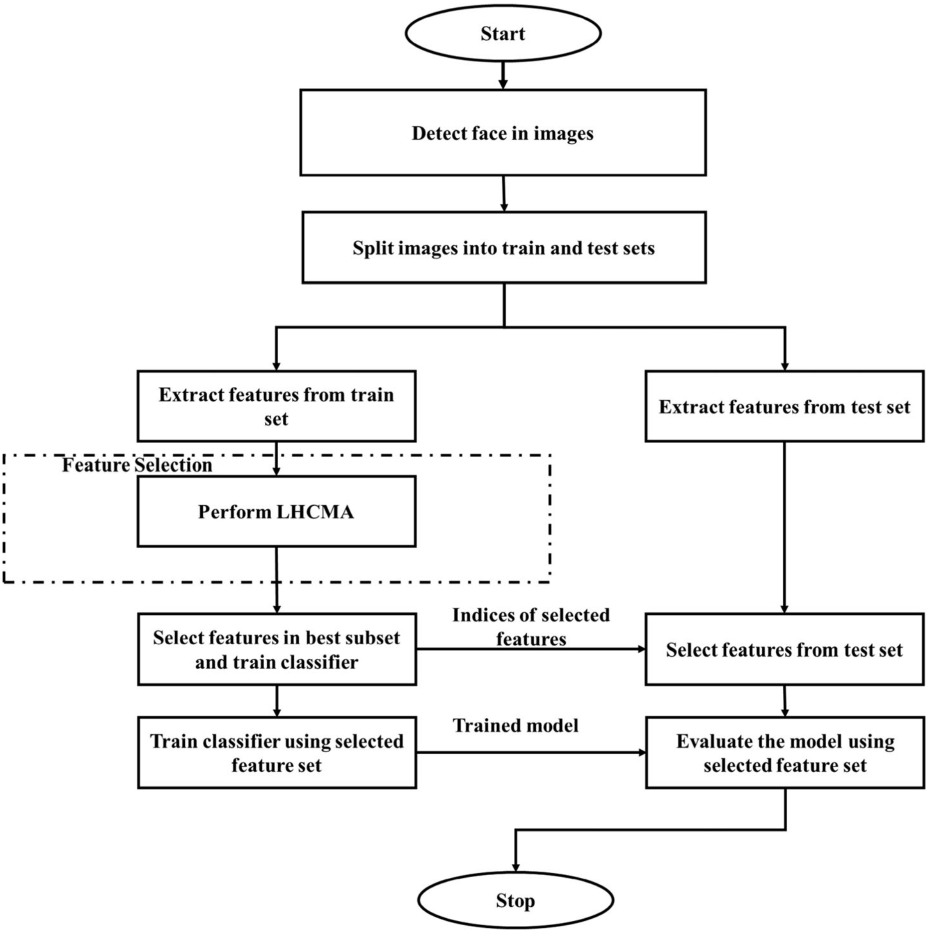

Fig. 5 Flowchart for working of our model for performing emotion recognition

Table 2 Value of parameters used in LHCMA

Parameter Name	Value
Probability of crossover	0.5
Probability of mutation	0.03
Maximum iterations	15
Population size	15

Accuracy of the chromosomes of Late Hill Climbing based Memetic Algorithm or LHCMA is observed using SMO (Sequential Minimal Optimization) classifier. Facial images portraying different expressions of each dataset are divided into train and test sets in the ratio of 2:1. The accuracies provided here are those obtained on the test set only. Summary of the results are provided in Table 3. Detailed results of the LHCMA based MA used as FS model on five feature vectors called uLBP, hvnLBP, Gabor, HOG and PHOG, are presented in Tables 4, 5, 6, 7 and 8 respectively. In comparisons the algorithm which has the highest accuracy is considered better and if the accuracies are equal then the algorithm which has a lower number of features is considered better. The best results are marked in bold. We have compared the results with that of no FS (i.e. on the entire feature vector), Simulated Annealing (SA), GA, basic MA, Mutation enhanced Binary Particle Swarm Optimization (ME-BPSO) [64] and Whale Optimization Algorithm – Crossover Mutation (WOA-CM) [39]. Both ME-BPSO and WAO-CM are recently published methods for FS. The results show that our method is far superior than the said methods as shown in the tables. The feature reduction capability of our model can be seen to be greater than that of its ancestors MA and also that of GA. As a summary, the average and standard deviation of the results for all 30 feature sets is provided in Table 3. In this table we can conclusively see that LHCMA has a better (highest accuracy average) as well as stable (lowest accuracy standard deviation) accuracy compared to the other methods.

Out of 30 cases, LHCMA performs best in 17 cases considering all algorithms. The second best results are of MA which is outperformed by LHCMA 20 times. Out of the 10 times LHCMA is outperformed by MA, it should be noted that the number of features is less in some cases while in rest there is a small difference in accuracy (less than 1%) in most cases. LHCMA outperforms GA in all but three cases. SA is completely outperformed by LHCMA. Comparison of LHCMA with other contemporary methods show that LHCMA is quite robust. WAO-CM outperforms LHCMA 6 times and ME-BPSO outperforms LHCMA only once. LHCMA, however, in all 30 cases has produced better results than the original feature vectors i.e. it is very much capable of removing redundant and/or irrelevant features. This shows LHCMA to be superior as compared to its ancestors MA as well as GA.

As seen in Table 4, compared to MA, LHCMA performs better in 66% of the cases. The maximum accuracy difference between MA and LHCMA where MA is better is only 1%

Table 3 Average and standard deviation values obtained without FS, SA, GA, MA, ME-BPSO, WAO-CM and LHCMA on all 30 feature sets

	Category	Initial	SA	GA	MA	ME-BPSO	WAO-CM	LHCMA
Number of features	Average	1652.53	**831.70**	1020.13	941.73	956.53	1181.23	898.87
	Std	1351.70	**674.91**	815.50	743.66	821.52	1156.50	729.08
Accuracy (in %)	Average	0.7408	0.6797	0.8195	0.8284	0.7827	0.7927	**0.8343**
	Std	0.1486	0.1700	0.1122	0.1098	0.1297	0.1316	**0.1048**

Table 4 Performance comparison of LHCMA with No FS, SA, GA, MA, ME-BPSO and WAO-CM for uLBP features

Dataset	Image Size	No FS		SA		GA		MA
		Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension
JAFFE	32X32	944	0.6234	481	0.3766	661	0.7143	**542**
	48X48		0.6234	487	0.4545	**678**	**0.7922**	672
	64X64		0.5974	493	0.4545	611	0.7013	703
RAFD	32X32		0.8358	485	0.6940	791	0.8638	716
	48X48		0.8862	441	0.7873	696	0.9216	696
	64X64		0.8638	487	0.7649	677	0.9011	**554**

Dataset	MA	ME-BPSO		WAO-CM		LHCMA	
	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy
JAFFE	**0.7662**	576	0.6493	911	0.6104	594	0.7662
	0.7662	472	0.6234	670	0.7403	574	0.7662
	0.7143	603	0.7234	869	0.6364	**570**	**0.7403**
RAFD	0.8675	620	0.7817	915	0.8228	**600**	**0.8713**
	0.9147	533	0.8489	**883**	**0.9272**	552	0.9161
	0.9142	638	0.8340	821	0.8601	555	0.9030

Table 5 Performance comparison of LHCMA with No FS, SA, GA, MA, ME-BPSO and WAO-CM for hvnLBP features

Dataset	Image Size	No FS		SA		GA		MA
		Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension
JAFFE	32X32	4096	0.5714	2059	0.5195	2613	0.7013	2232
	48X48		0.4675	2011	0.4286	2284	0.6104	2451
	64X64		0.4416	2132	0.3896	2254	0.5714	2208
RAFD	32X32		0.6642	2024	0.6119	2721	0.7015	**2081**
	48X48		0.7407	2049	0.681	2580	0.7500	2584
	64X64		0.694	2026	0.6418	2457	0.7276	2090

Dataset	MA	ME-BPSO		WAO-CM		LHCMA	
	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy
JAFFE	0.7013	2118	0.6104	3894	0.6104	**2235**	**0.7273**
	0.5844	1975	0.5065	2921	0.5195	**2158**	**0.6364**
	0.5844	2595	0.5065	1469	0.5455	**2060**	**0.5844**
RAFD	**0.7201**	2758	0.6698	3772	0.7071	2211	0.7034
	0.7649	2615	0.7257	**3489**	**0.7761**	2279	0.7519
	0.7444	2584	0.6922	**3914**	**0.7556**	2383	0.7369

Table 6 Performance comparison of LHCMA with No FS, SA, GA, MA, ME-BPSO and WAO-CM for Gabor filter-based features

Dataset	Image Size	No FS		SA		GA		MA
		Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension
JAFFE	32X32	640	0.6753	323	0.6623	375	0.7922	377
	48X48	1440	0.7273	704	0.6883	910	0.8052	836
	64X64	2560	0.7143	1293	0.7143	1541	0.8182	1408
RAFD	32X32	640	0.9049	320	0.8321	400	0.9328	429
	48X48	1440	0.9571	683	0.9160	851	0.9832	894
	64X64	2560	0.9851	1333	0.9590	1613	0.9832	1414

Dataset	MA	ME-BPSO		WAO-CM		LHCMA	
	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy
JAFFE	0.8442	301	0.8403	217	0.7922	**319**	**0.8442**
	0.8442	**701**	**0.8442**	630	0.8312	767	0.8312
	0.8312	1557	0.8312	**419**	**0.8961**	1428	0.8312
RAFD	0.9403	344	0.9236	557	0.9374	**337**	**0.9459**
	0.9888	770	0.9552	1074	0.9646	**758**	**0.9888**
	0.9875	1300	0.9813	1186	0.9701	**1271**	**0.9925**

Table 7 Performance comparison of LHCMA with No FS, SA, GA, MA, ME-BPSO and WAO-CM for HOG features

Dataset	Image Size	No FS		SA		GA		MA
		Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension
JAFFE	32X32	324	0.7143	178	0.7013	195	0.8571	206
	48X48	900	0.7403	444	0.6753	530	0.8961	507
	64X64	1764	0.7143	887	0.5844	1097	0.8052	1105
RAFD	32X32	324	0.8843	143	0.8576	186	0.9216	167
	48X48	900	0.9422	538	0.9154	480	0.9701	390
	64X64	1764	0.9366	816	0.9266	867	0.9627	675

Dataset	MA	ME-BPSO		WAO-CM		LHCMA	
	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy
JAFFE	0.8312	195	0.8232	281	0.8103	**182**	**0.8701**
	0.8701	440	0.8532	371	0.8822	**482**	**0.8961**
	0.8312	923	0.8212	1446	0.8162	**1008**	**0.8312**
RAFD	0.9216	211	0.8507	295	0.8694	**160**	**0.9235**
	0.9701	530	0.9391	605	0.9347	**455**	**0.9720**
	0.9627	1039	0.9515	1041	0.9496	**800**	**0.9757**

Table 8 Performance comparison of LHCMA with No FS, SA, GA, MA, ME-BPSO and WAO-CM for PHOG features

Dataset	Image Size	No FS		SA		GA		MA
		Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension
JAFFE	32X32	680	0.5325	357	0.4675	419	0.6494	374
	48X48		0.6623	344	0.5844	**396**	**0.8182**	405
	64X64		0.5974	342	0.6234	423	0.7922	412
RAFD	32X32		0.7854	351	0.7519	489	0.8246	**416**
	48X48		0.8545	354	0.8489	398	0.9049	**344**
	64X64		0.8881	366	0.8787	411	0.9123	**364**

Dataset	MA	ME-BPSO		WAO-CM		LHCMA	
	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy	Feature Dimension	Accuracy
JAFFE	0.6883	361	0.6364	633	0.5974	**359**	**0.7273**
	0.8052	440	0.7143	441	0.6753	373	0.8052
	0.7922	334	0.7403	178	0.7143	**374**	**0.8182**
RAFD	**0.8414**	420	0.7892	491	0.8041	396	0.8321
	0.9198	391	0.8787	541	0.9011	331	0.9104
	0.9384	352	0.9067	503	0.9235	395	0.9310

while in the opposite case an improvement of nearly 2.5% is achieved. For Table 8 the percentage of times LHCMA outperforms MA decreases to 50% though still with an accuracy margin of less than 1%. In rest of the Tables LHCMA performs much better than MA. This conclusively shows that our local search mechanism LAHCRR performs quite well.

7 Conclusion

FER is an important research topic and use of FS in this domain not only brings about better performance but also enables us to create a FER system with real-time performance. While basic MA is a better FS algorithm than GA, the local search used in MA is constrained due to its inability to evaluate the presence of redundancy in a feature subset. Inclusion of mRMR to evaluate the feature sets in local search makes the search faster as well as more effective. The use of a filter mechanism (mRMR) in place of a classifier to evaluate a feature subset for redundancy, reduces the computational complexity considerably. Therefore, a larger number of feature subsets can be evaluated in this step. LHCMA on the other hand allows us to include a feature subset with lower performance, allowing us to overcome a local optimum and obtain better performance. Therefore, LHCMA's local search allows us to create a better FS with better exploitation ability. This can be seen form the experimental results that, our algorithm outperforms SA, GA, MA, ME-BPSO and WAO-CM. The experiments are performed on 2 facial expression datasets from which 5 features are extracted – uLBP, hvnLBP, Gabor, HOG and PHOG. To account for the presence of varying image qualities, 3 image sizes are used which are 32×32, 48×48 and 64×64. In total, 30 feature sets are used to assess the FS ability of LHCMA in which LHCMA outperforms the rest in 17 cases by all the other aforementioned algorithms. No single algorithm is outperforming LHCMA. In future, inclusion of other state-of-the-art classifiers might be considered. Application of LHCMA on other FER datasets or different pattern recognition problems may also be attempted.

References

1. Agada R, Yan J (2015) Edge based mean LBP for valence facial expression detection. Proc 2015 IEEE Int Conf Electr Comput Commun Technol ICECCT 2015. https://doi.org/10.1109/ICECCT.2015.7226014
2. Ahmed F, Hossain E (2013) Automated Facial Expression Recognition Using Gradient-Based Ternary Texture Patterns. Faisal Chinese J Eng 2013:8
3. Appleby JS, Blake DV, Newman EA (1961) Techniques for producing school timetables on a computer and their application to other scheduling problems. Comput J 3:237–245
4. Barman A, Dutta P (2019) Facial expression recognition using distance and texture signature relevant features. Appl Soft Comput
5. Belanche LA, González FF (2011) Review and Evaluation of Feature Selection Algorithms in Synthetic Problems
6. Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, pp 401–408
7. Burke EK, Bykov Y (2017) The late acceptance Hill-Climbing heuristic. Eur J Oper Res 258:70–78
8. Carcagnì P, Del Coco M, Leo M, Distante C (2015) Facial expression recognition and histograms of oriented gradients: a comprehensive study. Springerplus 4. https://doi.org/10.1186/s40064-015-1427-3
9. Dalai N, Triggs B, Rhone-Alps I, Montbonnot F (2005) Histograms of oriented gradients for human detection. IEEE Comput Soc Conf Comput Vis Pattern Recognition, 2005 CVPR 2005 1:886–893. https://doi.org/10.1109/CVPR.2005.177

10. Daugman JG (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A 2:1160. https://doi.org/10.1364/JOSAA.2.001160
11. Dhall A, Asthana A, Goecke R, Gedeon T (2011) Emotion recognition using PHOG and LPQ features. 2011 IEEE Int Conf Autom Face Gesture Recognit Work FG 2011 878–883. https://doi.org/10.1109/FG.2011.5771366
12. Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. J Bioinforma Comput Biol 3:185–205
13. Dyer CR (2003) Simultaneous feature selection and classifier training via linear programming: a case study for face expression recognition. 2003 IEEE Comput Soc Conf Comput Vis Pattern Recognition, 2003 Proceedings 1:I-346-I-352. https://doi.org/10.1109/CVPR.2003.1211374
14. Ekman P, Rosenberg EL (1997) What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)
15. Gharsalli S, Emile B, Laurent H, Desquesnes X (2016) Feature Selection for Emotion Recognition based on Random Forest. pp 511–517
16. Ghimatgar H, Kazemi K, Helfroush MS, Aarabi A (2018) An improved feature selection algorithm based on graph clustering and ant colony optimization. Knowledge-Based Syst 159:270–285
17. Ghimire D, Jeong S, Lee J, Park SH (2016) Facial expression recognition based on local region specific features and support vector machines. Multimed Tools Appl:1–19. https://doi.org/10.1007/s11042-016-3418-y
18. Ghosh M, Adhikary S, Ghosh KK et al (2019) Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods. Med Biol Eng Comput 57:159–176. https://doi.org/10.1007/s11517-018-1874-4
19. Ghosh M, Begum S, Sarkar R et al (2019) Recursive Memetic Algorithm for gene selection in microarray data. Expert Syst Appl 116:172–185. https://doi.org/10.1016/j.eswa.2018.06.057
20. Ghosh M, Malakar S, Bhowmik S, et al (2019) Feature Selection for Handwritten Word Recognition Using Memetic Algorithm. In: Advances in Intelligent Computing. Springer, pp 103–124
21. Haghighat M, Zonouz S, Abdel-Mottaleb M (2015) CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. Expert Syst Appl 42:7905–7916. https://doi.org/10.1016/j.eswa.2015.06.025
22. Hamamoto Y, Uchimura S, Watanabe M et al (1998) A Gabor filter-based method for recognizing handwritten numerals. Pattern Recogn 31:395–400. https://doi.org/10.1016/S0031-3203(97)00057-5
23. Happy SL, George A, Routray A (2012) A real time facial expression classification system using Local Binary Patterns. Intell Hum Comput Interact (IHCI), 2012 4th Int Conf 1–5. https://doi.org/10.1109/IHCI.2012.6481802
24. Hsu H-H, Hsieh C-W, Lu M-D (2011) Hybrid feature selection by combining filters and wrappers. Expert Syst Appl 38:8144–8150
25. Huang J, Cai Y, Xu X (2007) A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recogn Lett 28:1825–1844
26. Huang D, Shan C, Ardabilian M et al (2011) Facial image analysis based on local binary patterns : A survey. IEEE Trans Syst Man, Cybern Part C (Applications Rev) 41:765–781
27. Jain AK, Farrokhnia F (1991) Unsupervised texture segmentation using Gabor filters. Pattern Recogn 24:1167–1186. https://doi.org/10.1016/0031-3203(91)90143-S
28. Kamarainen JK, Kyrki V, Kälviäinen H (2006) Invariance properties of Gabor filter-based features - Overview and applications. IEEE Trans Image Process 15:1088–1099
29. Kashef S, Nezamabadi-pour H (2015) An advanced ACO algorithm for feature subset selection. Neurocomputing 147:271–279. https://doi.org/10.1016/j.neucom.2014.06.067
30. Kong WK, Zhang D, Li W (2003) Palmprint feature extraction using 2-D Gabor filters. Pattern Recogn 36:2339–2347. https://doi.org/10.1016/S0031-3203(03)00121-3
31. Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: European conference on machine learning. Springer, pp 171–182
32. Lajevardi S, Hussain Z (2009) Feature selection for facial expression recognition based on optimization algorithm. Nonlinear Dyn:182–185
33. Lajevardi SM, Hussain ZM (2012) Automatic facial expression recognition: Feature extraction and selection. Signal, Image Video Process 6:159–169. https://doi.org/10.1007/s11760-010-0177-5
34. Langner O, Dotsch R, Bijlstra G et al (2010) Presentation and validation of the radboud faces database. Cognit Emot 24:1377–1388. https://doi.org/10.1080/02699930903485076
35. Li Z, Imai J, Kaneko M (2009) Facial-component-based bag of words and phog descriptor for facial expression recognition. In: 2009 IEEE International Conference on Systems, Man and Cybernetics. IEEE, pp 1353–1358

36. Li P, Phung SL, Bouzerdom A, Tivive FHC (2010) Feature selection for facial expression recognition. 2010 2nd Eur Work Vis Inf Process 35–40. https://doi.org/10.1109/EUVIP.2010.5699141
37. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans Image Process 11:467–476. https://doi.org/10.1109/TIP.2002.999679
38. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with Gabor wavelets. In: Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998. pp 200–205
39. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. Appl Soft Comput 62:441–453
40. Makhmudkhujaev F, Abdullah-Al-Wadud M, Bin IMT et al (2019) Facial expression recognition with local prominent directional pattern. Signal Process Image Commun
41. Malakar S, Ghosh M, Bhowmik S et al (2019) A GA based hierarchical feature selection approach for handwritten word recognition. Neural Comput & Applic:1–20. https://doi.org/10.1007/s00521-018-3937-8
42. Meshgini S, Aghagolzadeh A, Seyedarabi H (2013) Face recognition using Gabor-based direct linear discriminant analysis and support vector machine. Comput Electr Eng 39:727–745. https://doi.org/10.1016/j.compeleceng.2012.12.011
43. Mistry K, Zhang L, Neoh SC et al (2017) A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition. IEEE Trans Cybern 47:1496–1509
44. Mohamed NS, Zainudin S, Othman ZA (2017) Metaheuristic approach for an enhanced mRMR filter method for classification using drug response microarray data. Expert Syst Appl 90:224–231
45. Nava R, Cristóbal G, Escalante-Ramírez B (2011) Invariant texture analysis through Local Binary Patterns. Comput Vis pattern Recognit
46. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971–987. https://doi.org/10.1109/TPAMI.2002.1017623
47. Ou J, Bai X-B, Pei Y et al (2010) Automatic Facial Expression Recognition Using Gabor Filter and Expression Analysis. Second Int Conf Comput Model Simul 2010:215–218. https://doi.org/10.1109/ICCMS.2010.45
48. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22:1424–1445. https://doi.org/10.1109/34.895976
49. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27:1226–1238
50. Priya RV (2019) Emotion recognition from geometric fuzzy membership functions. Multimed Tools Appl:1–32
51. Radovic M, Ghalwash M, Filipovic N, Obradovic Z (2017) Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics 18:1–14. https://doi.org/10.1186/s12859-016-1423-9
52. Sakar CO, Kursun O, Gurgen F (2012) A feature selection method based on kernel canonical correlation analysis and the minimum Redundancy–Maximum Relevance filter method. Expert Syst Appl 39:3432–3437
53. Senawi A, Wei H-L, Billings SA (2017) A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking. Pattern Recogn 67:47–61
54. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image Vis Comput 27:803–816. https://doi.org/10.1016/j.imavis.2008.08.005
55. Shan C, Gritti T (2008) Learning Discriminative LBP-Histogram Bins for Facial Expression Recognition. Procedings Br Mach Vis Conf 2008:27.1–27.10. https://doi.org/10.5244/C.22.27
56. Shen L, Bai L, Fairhurst M (2007) Gabor wavelets and General Discriminant Analysis for face identification and verification. Image Vis Comput 25:553–563. https://doi.org/10.1016/j.imavis.2006.05.002
57. Sreedharan NPN, Ganesan B, Raveendran R et al (2018) Grey Wolf optimisation-based feature selection and classification for facial emotion recognition. IET Biometrics 7:490–499
58. Viola P, Jones MMJ (2004) Robust Real-Time Face Detection. Int J Comput Vis 57:137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
59. Wang X, Jin C, Liu W, et al (2013) Feature Fusion of HOG and WLD for Facial Expression Recognition. Syst Integr (SII), 2013 IEEE/SICE Int Symp 227–232. https://doi.org/10.1109/SII.2013.6776664
60. Wang F, Lv J, Ying G et al (2019) Facial expression recognition from image based on hybrid features understanding. J Vis Commun Image Represent 59:84–88
61. Wang S-H, Phillips P, Dong Z-C, Zhang Y-D (2018) Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. Neurocomputing 272:668–676
62. Wang G, Yang Y, Kong H (2009) Self-Learning facial emotional feature selection based on rough set theory. Math Probl Eng 2009. https://doi.org/10.1155/2009/802932

63. Wang Z, Zhang Y, Chen Z, et al (2016) Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image. In: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE, pp 755–758
64. Wei J, Zhang R, Yu Z et al (2017) A BPSO-SVM algorithm based on memory renewal and enhanced mutation mechanisms for feature selection. Appl Soft Comput J 58:176–192. https://doi.org/10.1016/j.asoc.2017.04.061
65. Yu J, Bhanu B (2006) Evolutionary feature synthesis for facial expression recognition. Pattern Recogn Lett 27:1289–1298. https://doi.org/10.1016/j.patrec.2005.07.026
66. Zhang S, Zhao X, Lei B (2012) Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans Signal Process 8:21–31
67. Zhao L, Wang Z, Zhang G (2017) Facial Expression Recognition from Video Sequences Based on Spatial-Temporal Motion Local Binary Pattern and Gabor Multiorientation Fusion Histogram. Math Probl Eng 2017. https://doi.org/10.1155/2017/7206041
68. Zhu Z, Ong Y-S, Dash M (2007) Wrapper–filter feature selection algorithm using a memetic framework. IEEE Trans Syst Man, Cybern Part B 37:70–76

Manosij Ghosh is presently an undergraduate student of Computer Science & Engineering department at Jadavpur University. His research interests comprise of Meta-Heuristic Algorithms, Artificial Intelligence and Feature Selection.

Tuhin Kundu is currently a final year undergraduate student pursuing Computer Science and Engineering at Jalpaiguri Government Engineering College. His areas of interest mainly include machine learning, computer vision and deep learning.

Dipayan Ghosh is currently pursuing Masters of Computer Applications at Jadavpur University, Kolkata (India). He received the B.Sc. (Hons.) degree in Computer Science from Ramakrishna Mission Residential College, Narendrapur (India). His current research interests include image processing, computer vision, pattern recognition and texture analysis.

Ram Sarkar received his B. Tech from University of Calcutta in 2003. He received his Master and PhD degrees from Jadavpur University in 2005 and 2012 respectively. He is a senior member of the IEEE.

Affiliations

Manosij Ghosh [1] · **Tuhin Kundu** [2] · **Dipayan Ghosh** [1] · **Ram Sarkar** [1]

Tuhin Kundu
tuhinkundu@outlook.com

Dipayan Ghosh
d6ghosh@gmail.com

Ram Sarkar
raamsarkar@gmail.com

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

[2] Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India