A Multi-Population FA for Automatic Facial Emotion Recognition

Kamlesh Mistry
Dept. of Computer & Information Sciences
Northumbria University
Newcastle, UK
k.mistry@northumbria.ac.uk

Baqar Rizvi
Dept. of Computer & Information Sciences
Northumbria University
Newcastle, UK
baqar.rizvi@northumbria.ac.uk

Chris Rook
Dept. of Computer & Information Sciences
Northumbria University
Newcastle, UK
c.rook@northumbria.ac.uk

Sadaf Iqbal
Transitional and Clinical Institute-Faculty of Medical
Newcastle University
Newcastle, UK
s.iqbal3@newcastle.ac.uk

Li Zhang
Dept. of Computer & Information Sciences
Northumbria University
Newcastle, UK
li.zhang@northumbria.ac.uk

Colin Paul Joy
Dept. of Computer Science & Information System
Teesside University
Middlesbrough, UK
c.joy@tees.ac.uk

Abstract — **Automatic facial emotion recognition system is popular in various domains such as health care, surveillance and human-robot interaction. In this paper we present a novel multi-population FA for automatic facial emotion recognition. The overall system is equipped with horizontal vertical neighborhood local binary patterns (hvnLBP) for feature extraction, a novel multi-population FA for feature selection and diverse classifiers for emotion recognition. First, we extract features using hvnLBP, which are robust to illumination changes, scaling and rotation variations. Then, a novel FA variant is proposed to further select most important and emotion specific features. These selected features are used as input to the classifier to further classify seven basic emotions. The proposed system is evaluated with multiple facial expression datasets and also compared with other state-of-the-art models.**

Keywords—feature optimization, facial expression recognition, local binary pattern

I. INTRODUCTION

Facial expression recognition plays an important role in computer vision and human computer interaction (HCI). Its a widely known fact the that facial expressions are a topic of interest in applications such as healthcare [1], video games [2], surveillance systems [3], and humanoid robots [4,5]. However, recognising real-time facial emotions has proven to be very difficult due to facial pose variations, illumination changes, occlusion and scale variations. Another major difficulty is that everybody's way of expressing facial emotion is different, which leads to false predictions. Another difficult task is to accurately identify emotion specific facial features, which can aid the system to discriminate between the facial emotions.

In order to deal with above mentioned challenges, an optimal, robust and accurate facial emotion analysis system is required. Literature shows that many researchers have worked on mode-specific and parametric feature extraction models to overcome the above-mentioned challenges. Yet, most of the models find it difficult to overcome all the challenges while preserving the high quality features and low computational complexity. This paper aims to deal with such challenges while producing robust and optimized discriminative facial representations to benefit real-time facial emotion recognition. According to [6], evolutionary algorithms (EA) show powerful global and local search capabilities and are widely applied for feature selection applications. The widely used EA algorithms are genetic algorithm (GA) [7], Particle swarm optimisation (PSO) [8], differential evolutions (DE) [9], and firefly optimisation algorithm (FA) [10]. These EA based feature selection models come with their own flaws e.g. PSO and FA tend to converge prematurely and therefore are at risk of local stagnation. As a result, in this paper a novel multi-population FA is proposed which employs information sharing concept to further diversify the population.

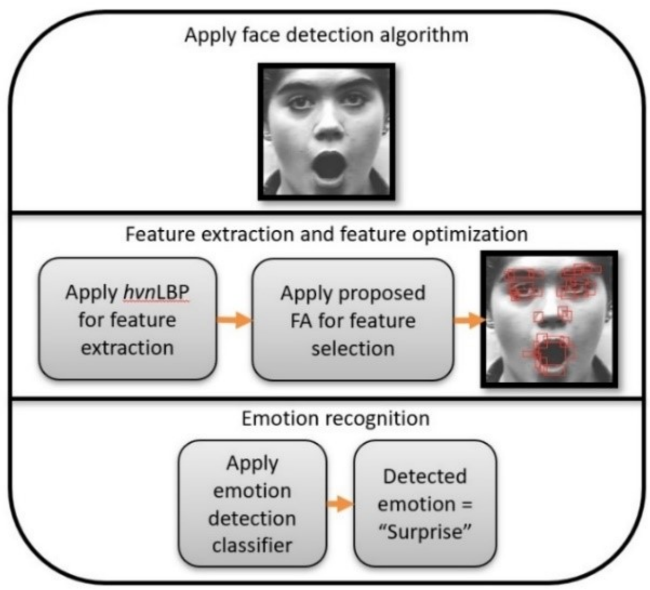

Fig 1. System architecture of the proposed system

The overall system shown in figure 1 consists of three steps, namely feature extraction, feature selection and emotion classification with the following major contributions:

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

- A multi-population FA with information sharing concept is proposed in order to avoid the premature convergence and local stagnation problems. It allows the system to separate the non-emotion features from the emotion specific features.
- Our proposed system is evaluated with five datasets CK+ [11], MMI [12], JAFFE [13], Bosphorus 3D [14] and BU-3DFE [15] databases. It outperforms state-of-the-art EA based algorithms, and other facial expression recognition methods reported in the literature significantly.

The rest of the paper is organized as follows. Section 2 presents literature review. Section 3 introduces the facial emotion recognition system including a hvnLBP for feature extraction, a novel multi-population FA for feature selection and emotion recognition. Section 4 presents an evaluation of the proposed system in comparison with other related research. Section 5 draws conclusions and identifies future directions of research.

II. Related Works

A. *EA based Feature selection*

In literature, several state-of-the-art FA variants can be found which overcomes the problem of premature convergence associated with the original FA model. Some of the variants are FA with neighborhood attraction (NaFA) [16], FA using a Logistic map as the attractiveness coefficient (denoted as CFA1) [17], opposition and dimensional FA (ODFA) [18], a modified FA (MFA) [19], FA with a variable step size (VSSFA) [20], FA with random attraction (RaFA) [21], FA with a Gauss map as the attractiveness coefficient (denoted as CFA2) [22], a hybrid multi-objective FA (HMOFA) [23], and SA incorporated with both FA (SFA) and FA with Levy flights (LSFA) [24]. Table 1 summarizes and lists the key characteristics of the above-mentioned FA variants and also the FA model proposed by us in this research.

FA algorithms have been used in literature for feature selection. Zhang et al, 2016 [25] worked on a hybrid moth-firefly algorithm for facial feature selection and expression recognition. Their work made use of the spiral operation of Moth-Flame Optimization applied to FA to identify features related to seven basic facial expressions. Jothi et al, 2016 [26] proposed a hybrid feature selection FA model by integrating the Tolerance Rough Set into the FA. The authors used the algorithm for image classification of brain tumors, magnetic resonance imaging (MRI) data. The modified FA algorithm was used to identify the most significant characteristics from segmented MRI images and then perform tumor classification. The authors in the work by Kora [27] worked on electrocardiogram (ECG) signals to detect Bundle Branch Block (BBB) using hybrid FA-based feature selection. Their model utilised the personal and global best solutions of Particle Swarm Optimization (PSO) along with the attractiveness search mechanism of FA for BBB pattern recognition. Zhang, 2018 [28] used a modified FA model to identity the most optimal topology for ensemble classifier/regressor construction. They made use of both the attraction and evading mechanisms and used FA algorithm to construct the most optimal ensemble models. Using their algorithm, they removed redundant base classifiers/regressors without compromising classification accuracy. Su, 2017 [29] proposed FA for parameter tuning and optimal band selection for the Extreme Learning Machine in hyperspectral image classification, while FA integrated with probability distributions was used for facial feature selection in the work by Mistry et al, 2017 [30]. Kazem, 2013 [31] used a regression forecasting model for stock market price prediction, where CFA1 was used to optimize hyper-parameters of Support Vector Regression (SVR). PSO and GA-based feature selection methods have also been applied for acute lymphoblastic leukemia classification [32], skin cancer detection [33] and arousal/valence regression for bodily expression recognition [34].

B. *Emotion Recognition*

Various types of conventional approaches have been studied for automatic FER systems. The basic principle of these approaches is to detect the face region and extract the geometric features, appearance features, or a hybrid of geometric and appearance features on the target face. In geometric feature calculation, the relationship between different facial components is used to construct a feature vector for training [35,36]. Ghimire and Lee [36] proposed an approach using two types of geometric features calculated using the position and angle of 52 facial landmark points. Firstly, they calculated the angle and Euclidean distance between each pair of landmarks within a frame. Then secondly, they subtracted the distance and angles of the frame from the corresponding distance and angles in the first frame of the video sequence. Classification was done using the 2 methods namely 1) Multi-class AdaBoost with dynamic time warping, or 2) using a SVM on the boosted feature vectors. The appearance features are usually extracted from the global face region [37] or different face regions making use of different types of information [38,39]. Global features were calculated by Happy et al. [37] using local binary pattern (LBP) histogram of different block sizes from a global face region. Then classification was done to identify various facial expressions using principal component analysis (PCA). Since this method was implemented in real time, the recognition accuracy was poorer because the feature vector could not reflect local variations of the facial components. Different face regions have different levels of importance. For example, the eyes and mouth contain more information than the forehead and cheek, so global feature extraction can pose some limitations. Ghimire et al. [40] extracted appearance features region wise by dividing the entire face region into domain-specific local regions. Important local regions were determined using an incremental search approach, and hence resulted in reduced feature dimensions and improvement in the recognition accuracy.

Recently, with the development of big data and the improvement of hardware technology, many algorithms based on deep learning have been researched. The field of FER is also being influenced by these advancements. More robust and efficient feature recognition techniques have been proposed that learn the extracted facial features automatically. In this section, we introduce the CNN-based FER algorithms. Lopes et al. [41] proposed a representative facial expression algorithm utilising CNN based deep learning. The authors used the data argumentation process to resolve the scarcity of the FER dataset and their approach was robust to facial emotions and changes such as rotation and transportation. In this algorithm, except for the parts with unnecessary elements around the face, the AUs are cropped into blocks at the center of the action unit. CNN then classified the emotions into six to seven categories. Although Deep learning requires large datasets for learning but in such algorithms, the problem of small datasets is solved by argumentation methods, and FER research based on CNN is gaining popularity.

III. The Proposed System

In this section, we introduce the proposed facial emotion recognition system. The overall system consists of three key steps, i.e. A hvnLBP-based feature extraction, proposed FA based feature selection and emotion recognition. Each step is introduced in detail in the following sub-sections.

A. Feature Extraction

In feature extraction process, pre-processing is applied to reduce image noise. A histogram equalization method is initially used to improve the contrast of an input image. Then, a bilateral filter is applied to reduce image noise, while preserving the edges. We subsequently apply Viola and Jone's face detection algorithm provided in the openCV package to detect the face region of the input image. The detected face is further processed using *proposed LBP* in order to extract robust features.

1) The horizontal vertical neighbourhood LBP

Ojala et al. [42] proposed the conventional LBP which thresholds each of the 3x3 neighbouring pixels with a centre pixel value. The conventional LBP was further extended to use various numbers of circular neighbouring pixels [17]. The LBP operator $LBP_{p,r}$ can produce 2^p different binary patterns, where p denotes the number of neighborhood pixels and r denotes the radius of the circular pattern. The equation for calculating the $LBP_{p,r}$ operator can be given as follows:

$$LBP_{p,r} = \sum_{p=0}^{p-1} S(g_p - g_c)2^p \, , S(x) = \{^{1 \; if \; x \geq 0}_{0} \tag{1}$$

where g_p denotes the neighborhood pixel at location p and g_c is the center pixel. The important information such as edges, corners, spot and flat area can be detected using the LBP [43]. The conventional LBP is robust to illumination and scaling variations but fails to deal with rotation variations [44]. Whereas, the gradient images contain enhanced edge information and are more stable than raw pixel intensities, which can benefit to deal with rotation and illumination variations.

In order to improve the feature extraction quality in terms of low contrast ratio, [45] proposed horizontal vertical neighborhood LBP (hvnLBP). The hvnLBP operator can be calculated by using the following equation:

$$\begin{aligned} hvnLBP_{p,r} = \{ &\mathrm{S}(\max(p_0, p_1, p_2)) \, , \mathrm{S}(\max(p_7, p_3)), \\ &\mathrm{S}(\max(p_6, p_5, p_4)) \, , S(\max(p_0, p_7, p_6)), \\ &\mathrm{S}(\max(p_1, p_5)) \, , S(\max(p_2, p_3, p_4)) \} \end{aligned} \tag{2}$$

where p_i denotes the pixel intensity of neighborhood pixels at the i^{th} location, r is the radius, and S denotes the comparison operation, as follows.

$$S(\max(p_j, p_k, p_m)) = \begin{cases} 1 & if \; maximum \\ 0 & if \; not \; maximum \end{cases} \tag{3}$$

where, p_j, p_k, and p_m represent the neighborhood pixels in a row or column. Note that p_k is removed if it represents the center pixel. The Figure 2 shows the output results generated using hvnLBP in comparison with conventional LBP. In comparison to conventional LBP, the proposed extended $hvnLBP$ operator captures more discriminative contrast information and can achieve better face representation. However, the final feature vector generated by hvn-LBP operator is high dimensional and will need a longer processing time in real-time applications. To minimize the high dimensionality problem, we propose a multi-population FA to identify the most discriminative features related to each facial expression.

In this paper, the resolution of detected face image is 75x75 pixels and after applying the proposed *LBP* operator, the face image is divided into 25x25 (i.e. 625) sub-regions with the size of each sub-region being 3x3. In comparison to conventional LBP and its variants, the hvnLBP operator captures more discriminative contrast information such as corners and edges amongst neighbourhoods. The feature extracted by proposed LBP are further used as input features to train and test the various classifiers.

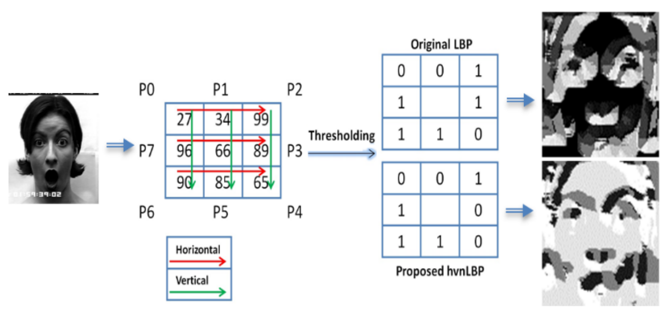

Fig 2. Comparison of hvnLBP with the conventional LBP [45].

B. Feature Selection

1) Conventional Firefly Algorithm

The firefly algorithm was first introduced by Yang [10] which is inspired by the natural behavior of fireflies. The conventional FA runs on three important rules:

1. All fireflies are gender blind.
2. The fireflies with higher illumination will attract the less illuminated firefly. While the highest

illuminated firefly will move randomly as no brighter firefly exists.

3. The quality of the solution depends on the firefly's illumination intensity.

The search mechanism used in FA shows better performance when compared with other metaheuristic algorithms such as PSO and GA [45]. In FA, the illumination intensity and attractiveness vary depending on the distance between two fireflies. The illumination intensity variation is presented in Eq. (4).

$$I = I_0 e^{-\gamma r} \tag{4}$$

Where I_0 represents the original illumination intensity at $r = 0$ and γ represents the fixed illumination absorption coefficient. $\beta(r)$ represents the attractiveness factor, which is directly proportional to the illumination intensity and can be defined as follows:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{5}$$

Where the initial attractiveness is represented by β_0 with $r = 0$. Furthermore, Eq. (6) is used to measure the distance between fireflies i and j.

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2} \tag{6}$$

The position of two fireflies is represented as x_i and x_j, where $x_{i,k}$ and $x_{j,k}$ represents the kth dimension of the position x_i and x_j.

Originally, FA uses Gaussian distribution equation to conduct randomization but Yang [10] applied Levy flight to further enhance the search performance. The firefly movement equation using levy flight is as follows:

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_j - x_i) + \alpha\, sign[rand - \tfrac{1}{2}] \oplus Levy \tag{7}$$

Where the second part of the equation represent the attraction movement and third part shows the randomization using Levy flight.

In FA and its variant, γ is very important while determining the attractiveness and the convergence speed of the algorithm. For Example, if $\gamma = 0$ then the attractiveness and illumination intensity remain unchanged, which means FA will mimic the search behavior presented in PSO algorithm.

2) *Proposed Firefly Algorithm*

The conventional FA only employs one set of population to explore the search space. The single set of population restricts the search in only one direction or one set of features. The application of facial feature can lead to multiple clusters of features, which makes conventional FA less useful in this scenario.

In this paper, we propose a new variant of FA which employs two populations in order to cluster the features. The proposed FA starts by generating two primary populations with the same population size. A different criterion is used to evaluate the fitness of population 1 and population 2. In this application the population 1 is evaluated against features corresponding to the specific emotion and the population 2 is evaluated against features corresponding to non-emotion.

This process will generate a two set of features after every iteration. However, this approach leads to running two FA in parallel independently which leads to inheriting the local stagnation and population diversity problem from the conventional FA.

In order to further diversify the populations, we have introduced an information exchange system. This system selects the population of individuals with record of higher fitness values to share the search information with other population members. Let us consider that the two populations P_1 and P_2 are initialized randomly. P_1 is set to search for positive features i.e. emotion specific features, P_2 is set to search for negative features i.e. non-emotion specific features, respectively.

Algorithm 1: The Pseudo Code of the Proposed FA Variant

Step-1 Initialize two sets of the population.
Step-2 Evaluate the fitness value for each individual using the **Equation 8**. With separate criteria for each population.
Step-3 While (satisfying termination criteria)
  Perform standard FA steps on $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ as follows:
    Move fireflies using **Equation 6**.
    Evaluate and update illumination intensity using **Equation 4**.
  Rank fireflies according to their fitness values and find the current global best.
  If (iterations%50 ==0)
    Compare $\boldsymbol{P_1}$ with $\boldsymbol{P_2}$.
    Evaluate fitness for each individual in $\boldsymbol{P_1}$ with fitness criteria of the $\boldsymbol{P_2}$ or vise-versa.
    Swap individuals from $\boldsymbol{P_1}$ if their fitness is higher than top 5 individuals from $\boldsymbol{P_2}$ or vise-versa.
  End If
  Re-evaluate the fitness value for each individual in P_1 and P_2.
End While
Step-4 End

The FA is set to run 500 iterations over both populations with different search criteria. Les us consider that after 50 (selected based on experiments) iterations the member of P_1 will have higher fitness values then each individual from P_1 will be compared with the individuals from P_2. This comparison will be based on the correlation between each population members. The population with the highest correlating members will be selected for information sharing. The fitness of individuals from P_1 will be evaluated using the search criteria used by P_2. If any member of P_1 can achieve higher or similar fitness than the top five members from P_2, then that individual will be swapped by weakest member from P_2. This mechanism shows a significant improvement in diversifying the population and reducing the risk of local stagnation. The fitness function used to evaluate each individual is given as follows:

$$F(x) = w_a * acc_x + w_f * (number\_feature_x)^{-1} \tag{8}$$

where w_a and w_f are two predefined constant weights for acc_x (classification accuracy) and $number\_feature_x$ (the number of features), respectively. In this paper, $w_a = 0.8$ and $w_f = 0.2$, with $w_a > w_f$ to represent the fact that classification accuracy is more important than the number of selected features. The pseudo code is illustrated in algorithm 1.

C. Emotion Classification

In this paper, we have developed an automatic facial emotion recognition system which detects seven annotated emotions (i.e. anger, happiness, sadness, surprise, disgust, fear, and neutral). In order to achieve this, we employ diverse classifiers NN, multi-class SVM, and the SVM-based and NN-based ensembles classifiers for emotion classification. The selected features generated by the proposed FA are used as inputs to the classifiers. The input layer nodes for NN is set to number of features generated by the proposed FA and has one hidden layer and one output layer with seven nodes indicating each emotion category. For SVM, grid-search is applied to optimise the parameter settings. The same NN and SVM parameters are used as the base classifier to for within each ensemble. Both ensembles employ three base classifiers and a weighted majority voting combination method to produce final classification.

We have trained the system using 250 images from the CK+ and 175 images from CK+, MMI, JAFFE, Bosphorus 3D and BU-3DFE database for testing. Overall, a system with SVM-based ensembles achieve the best accuracy when tested with images from the above mentioned five databases.

IV. Evaluation

In this section we evaluate our system in two steps, first, we show the comparison between proposed FA and state-of-the-art evolutionary algorithms without hvnLBP features. Then the proposed FA-based feature selection algorithm is evaluated against state-of-the-art evolutionary algorithms (EA) with hvnLBP features. Single and ensemble classifiers such as NN, SVM, and NN-based and SVM-based ensembles, are used for all the evaluation experiments.

Table 1. The average classification accuracy over 30 runs for all the feature selection algorithms in combination with NN classifier

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	85.5	80.5	82.6	84.9	83.2
PSO	86.1	81.2	81.5	83.5	84.3
FA	89.0	84.7	86.1	87.7	85.5
LSFA	88.2	82.6	83.5	85.1	87
VSSFA	85.5	82.5	84.7	84.0	85.5
ODFA	90.5	86.6	89.8	88.5	86.2
RaFA	84.5	83.7	85.8	86.9	87.5
Proposed FA	93.9	90.0	91.2	90.5	89.5

The first experiment conducted uses 250 from CK+ for training and 175 images from five datasets for testing, respectively. Table 1 shows the results of the proposed FA against state-of-the-art evolutionary algorithms (i.e. GA, PSO, FA, LSFA, VSSFA, ODFA, and RaFA). In order to conduct the comparison between the proposed FA and other EA, Table 1, 2, 3, and 4 presents the results obtained using only feature selection and selected classifiers without any feature extraction method.

As shown in Table 1, 2, 3, and 4, proposed FA outperforms all the other seven EA algorithms when tested on all the datasets with a significant margin. Overall, the results achieved by SVM-Ensemble system are the highest compared to other classifiers for all the datasets.

Table 2. The average classification accuracy over 30 runs for all the feature selection algorithms in combination with SVM classifier

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	88.5	80.6	81.4	79.3	82.7
PSO	86.6	85.6	85.9	84.5	83.1
FA	91.6	86.9	88.6	83.8	83.5
LSFA	89.9	87.9	86	86.3	86.1
VSSFA	89.4	91.1	91.2	84.3	81.6
ODFA	93	88.3	89.8	83.1	85.5
RaFA	90.1	90.4	88.9	84.8	88.5
Proposed FA	93.7	91.3	91.4	88.6	89

Table 3. The average classification accuracy over 30 runs for all the feature selection algorithms in combination with NN-Ensemble classifier

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	87.6	88.5	83.3	82.3	84.5
PSO	87.7	86	81.1	85.3	81.1
FA	91.1	91.5	85.1	89.0	86.1
LSFA	89.0	91.2	87.4	87.0	86.9
VSSFA	89.6	88.3	88.3	90.2	88.6
ODFA	93.9	93.7	91.3	89.6	90.3
RaFA	89.5	90.2	88.4	89.1	88.6
Proposed FA	**94.3**	**93.7**	**92.5**	**91.4**	**91.2**

Table 4. The average classification accuracy over 30 runs for all the feature selection algorithms in combination with SVM-Ensemble classifier

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	85.5	88.8	86.6	83.0	86.5
PSO	89.2	89.5	89.3	84.6	88.2
FA	91	90.4	88.5	86.6	89.4
LSFA	89.5	91.1	89.6	86.2	90.9
VSSFA	93.4	92.9	92.4	87.5	91.5
ODFA	91.5	93.9	93.9	89.7	92.3
RaFA	90.5	92.4	93.4	89.5	92.0
Proposed FA	**95.5**	**94.8**	**92.8**	**92**	**91.5**

In the second set of experiment, we conduct the comparison between the proposed FA and other EA by combining them with feature extraction and selected classifiers. These results are presented in Table 5, 6, 7, and 8. Moreover, the proposed FA continued to outperform all the selected EA with a significant margin. But in this set of experiment NN-Ensemble achieves the highest results compared to other classifiers respectively. In order to further demonstrate the efficiency of the proposed system, we have compared the best results of the proposed system with the existing research work. These results are presented in Table 9 for all the selected datasets.

TABLE 5. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE FEATURE SELECTION ALGORITHMS IN COMBINATION WITH NN CLASSIFIER

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	86.8	85.4	85.5	83.2	83.5
PSO	87.5	85.7	85.3	84.6	83.1
FA	87.4	85.0	86.5	85.6	85.4
LSFA	88.1	86.6	87.6	85.9	86.4
VSSFA	88.9	87.4	89.4	86.3	85.7
ODFA	90.9	89.7	90.9	87.1	88.5
RaFA	92.4	90.3	90.4	88.5	87.2
Proposed FA	**95.8**	**93.5**	**93.3**	**91**	**90.7**

TABLE 6. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE FEATURE SELECTION ALGORITHMS IN COMBINATION WITH SVM CLASSIFIER

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	85.6	84.4	85.3	84.0	83.8
PSO	86	85.7	85.6	84.1	85.2
FA	86.2	86.7	85.7	85.6	84.9
LSFA	87.8	86.6	88.5	85.2	82.6
VSSFA	88.9	87.9	90.4	88.5	85.2
ODFA	91.4	90.6	89.9	88.7	88.5
RaFA	91.2	90.2	92.7	89.0	86.6
Proposed FA	**95.2**	**93.4**	**93.6**	**92.5**	**91.6**

TABLE 7. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE FEATURE SELECTION ALGORITHMS IN COMBINATION WITH NN-ENSEMBLE CLASSIFIER

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	88.0	87.9	87.0	85.0	86.9
PSO	89.2	88.5	88.2	86.6	86.2
FA	90.0	89.2	88.8	87.9	89.0
LSFA	90.7	90.0	89.5	88.3	89.5
VSSFA	92.4	91.4	90.7	89.5	90.4
ODFA	93.5	93.2	92.4	91.9	90.7
RaFA	94.9	94.4	93.3	91.50	92.1
Proposed FA	**98.1**	**96.0**	**95.1**	**94.15**	**93.6**

TABLE 8. THE AVERAGE CLASSIFICATION ACCURACY OVER 30 RUNS FOR ALL THE FEATURE SELECTION ALGORITHMS IN COMBINATION WITH NN-ENSEMBLE CLASSIFIER

	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
GA	89.1	88.7	87.7	87.4	86.7
PSO	90.1	89.1	87.4	86.5	85.6
FA	90.4	89.9	89.2	88.2	87.3
LSFA	91.3	90.7	90.4	89.5	89.4
VSSFA	93.7	92.9	91.8	90.4	89.9
ODFA	95.6	95.2	93.6	92.3	92.0
RaFA	96.7	96.6	95.3	93.2	92.5
Proposed FA	**99.4**	**98.7**	**97.1**	**95.2**	**94.5**

As indicated in Table 8, the proposed system outperforms all other related research when using CK+ for training and all five datasets for testing. The proposed FA outperforms other system even without the combination of feature extraction algorithm. This demonstrates the efficiency and robustness of the proposed system.

TABLE 9. COMPARISON WITH RELATED RESEARCH FOR DIFFERENT DATASETS

Methods	Methodology	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
Shan et al. [44]	Boosted LBP+SVM	91.40	86.9	81	-	-
Elaiwat et al. [46]	Spatio-temporal RBM based model	95.66	81.63	-	-	-
Zhong et al. [47]	CSPL	89.89	73.53	-	-	-
Derkach and Sukno [48]	Graph Laplacian	-	-	-	81.5	-
Jan and Meng [49]	ULBP+LPQ+EOH+83P+FD	-	-	-	88.32	79.46
This work	Proposed System without hvnLBP	95.5	94.8	92.8	92	91.5
This work	**Proposed System with hvnLBP**	**99.4**	**98.7**	**97.1**	**95.2**	**94.5**

V. CONCLUSION

In this paper, we have presented a novel variant of FA to enhance the performance of an automatic facial emotion recognition system. This system employs hvnLBP, proposed FA and diverse classifiers for recognizing seven facial emotions. The proposed FA can select the best features representing each facial emotion. It outperforms other state-of-the-art feature selection methods such as GA, PSO, FA, LSFA, VSSFA, ODFA, and RaFA, significantly. The proposed system achieves an average accuracy of 96.9% when evaluated with test images from five datasets. The system also shows promising performance for each dataset evaluation and achieves an accuracy of 94.5% for Bosphorus 3D, 95.2% for BU-3DFE, 97.1% for JAFFE, 98.7% for MMI, 99.4% for CK+, respectively. It also shows a promising performance when compared with other state-of-the-art related facial expression recognition researches. In future work, other hybrid or multi-objective feature selection models will also be explored for solving high dimensionality problems.

REFERENCES

[1] Zhang, L., Fielding, B., Kinghorn, P., and Mistry, K. 2016. A Vision Enriched Intelligent Agent with Image Description Generation. In *Proceedings of International Conference of Autonomous Agents and Multiagent Systems*. Singapore.

[2] G'Mussel, A.S. and Hewig, J. 2013. The value of a smile: Facial expression affects ultimatum-game responses. *Judgment and Decision Making*, 8 (3), 381-385.

[3] Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., and Movellan, J. 2008. Automated Drowsiness Detection for Improved

Driver Safety. In *Proceedings of the International Conference on Automotive Technologies*, 2008.

[4] Zhang, L., Jiang, M., Farid, D., and Hossain, A.M. 2013. Intelligent Facial Emotion Recognition and Semantic-based Topic Detection for a Humanoid Robot. *Expert Systems with Applications*, 40 (2013), pp. 5160-5168.

[5] Zhang, L., Mistry, K., Jiang, M., Neoh, S.C., and Hossain, M.A. 2015. Adaptive facial point detection and emotion recognition for a humanoid robot, *Computer Vision and Image Understanding* 140 (2015) 93-114.

[6] Xue B., Zhang M., and Browne W. N. 2013. Particle swarm optimization for feature selection in classification: A multi-objective approach. In *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671.

[7] Holland J.H. 1992. Genetic Algorithms. In Sci. Am. 267, 66-72.

[8] Eberhart R.C., Kennedy J. 1995. A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science, pp. 39–43.

[9] Storn R., Price K. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Opt. 11, 341–359.

[10] Yang X.S. 2020. Firefly algorithm, Levy flights and global optimization. In Research and Development in Intelligent Systems 26, 209–218.

[11] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

[12] Pantic, M., Valstar, M.F., Rademaker, R., Maat, L., Web-based database for facial expression analysis, in: Proceedings of IEEE Int'l Conf. Multimedia and Expo, Amsterdam, The Netherlands, 2005, pp. 317–321.

[13] Lyons M. J., Akemastu S., Kamachi M., Gyoba J. 1998. Coding Facial Expressions with Gabor Wavelets. In 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205.

[14] Savran A., Alyüz N., Dibeklioğlu H., Çeliktutan O., Gökberk B., Sankur B., and Akarun L. 2008. Bosphorus Database for 3D Face Analysis. In The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008), Roskilde University, Denmark.

[15] Yin L., Wei X., Sun Y., Wang J., and Rosato M. J. 2006. A 3D facial expression database for facial behavior research. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, pp. 211–216.

[16] H. Wang, W. Wang, X. Zhou, H. Sun, J. Zhao, X. Yu, Z. Cui, Firefly algorithm with neighborhood attraction, Information Sciences 382–383 (2017) (2017) 374–387.

[17] A. Kazem, E. Sharifi, F.K. Hussain, M. Saberic, O.K. Hussain, Support vector regression with chaos-based firefly algorithm for stock market price forecasting, Applied Soft Computing 13 (2) (2013) 947–958.

[18] O.P. Verma, D. Aggarwal, T. Patodi, Opposition and dimensional basedmodified firefly algorithm, Expert Systems with Applications 44 (2016) (2016) 168–176.

[19] L. He, S. Huang, (2017). Modified firefly algorithm based multilevel thresholding for colour image segmentation, Neurocomputing 240 (2017) 152–174.

[20] S.H. Yu, S.L. Zhu, Y. Ma, D.M. Mao, A variable step size firefly algorithm for numerical optimization, Applied Mathematics and Computation 263 (2015) 214–220.

[21] H. Wang, W.J. Wang, H. Sun, S. Rahnamayan, Firefly algorithm with random attraction, International Journal of Bio-Inspired Computation 8 (1) (2016) 33–41.

[22] A.H. Gandomi, X.S. Yang, S. Talatahari, A.H. Alavi, Firefly algorithm with chaos, Communications in Nonlinear Science and Numerical Simulation 18 (2013) 89–98.

[23] H. Wang,W.Wang, L. Cui, H. Sun, J. Zhao, Y.Wang, Y. Xue, A hybrid multi-objective firefly algorithm for big data optimisation, Applied Soft Computing (2017) 2017.

[24] L. Zhang, K. Mistry, S.C. Neoh, C.P. Lim, Intelligent facial emotion recognition using moth-firefly optimization, Knowledge-Based Systems 111 (2016) (2016) 248–267.

[25] G. Jothi, H.H. Inbarani, Hybrid tolerance rough set–firefly based supervised feature selection for MRI brain tumor image classification, Applied Soft Computing 46 (2016) 639–651.

[26] M. Alweshah, S. Abdullah, Hybridizing firefly algorithms with a probabilistic neural network for solving classification problems, Applied Soft Computing 35 (2015) 513–524.

[27] P. Kora, K.S.R. Krishna, Hybrid firefly and particle swarm optimization algorithm for the detection of bundle branch block, International Journal of the Cardiovascular Academy 2 (1) (2016) 44–48.

[28] L. Zhang,W. Srisukkham, S.C. Neoh, C.P. Lim, D. Pandit, Classifier ensemble reduction using a modified firefly algorithm: an empirical evaluation, Expert Systems with Applications 93 (2018) (2018) 395–422.

[29] H. Su, Y. Cai, Q. Du, Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (1) (2017) 309–320.

[30] K. Mistry, L. Zhang, G. Sexton, Y. Zeng, M. He, Facial expression recognition using firefly-based feature optimization, Proceedings of IEEE Congress on Evolutionary Computation 2017, pp. 1652–1658.

[31] W. Srisukkham, L. Zhang, S.C. Neoh, S. Todryk, C.P. Lim, Intelligent Leukaemia diagnosis with bare-bones PSO based feature optimization, Applied Soft Computing 56 (2017) (2017) 405–419.

[32] T.Y. Tan, L. Zhang, M. Jiang, An Intelligent Decision Support System for Skin Cancer Detection from Dermoscopic Images, Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNCFSKD), 2016.

[33] Y. Zhang, L. Zhang, S.C. Neoh, K. Mistry, A. Hossain, Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles, Expert Systems with Applications 42 (22) (2015) 8678–8697.

[34] Suk,M.; Prabhakaran, B. Real-time mobile facial expression recognition system—A case study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 24–27 June 2014; pp. 132–137.

[35] Ghimire, D.; Lee, J. Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. Sensors 2013, 13, 7714–7734. [CrossRef] [PubMed]

[36] Happy, S.L.; George, A.; Routray, A. A real time facial expression classification system using local binary patterns. In Proceedings of the 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, 27–29 December 2012; pp. 1–5.

[37] Siddiqi, M.H.; Ali, R.; Khan, A.M.; Park, Y.T.; Lee, S. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans. Image Proc. 2015, 24, 1386–1398. [CrossRef] [PubMed]

[38] Khan, R.A.; Meyer, A.; Konik, H.; Bouakaz, S. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit. Lett. 2013, 34, 1159–1168. [CrossRef]

[39] Ghimire, D.; Jeong, S.; Lee, J.; Park, S.H. Facial expression recognition based on local region specific features and support vector machines. Multimed. Tools Appl. 2017, 76, 7803–7821. [CrossRef]

[40] A. T. Lopes, E. de Aguiar, A. F. de Souza, and T. Oliveira-Santos, ''Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order,'' Pattern Recognit., vol. 61, pp. 610–628, Jan. 2017.

[41] Ojala, T., Pietikäinen, M., and Harwood, D. 1996. A comparative study of texture measures with classification based on featured distribution, *Pattern Recognition* 29 (1) (1996) 51–59.

[42] Pietikäinen, M., Mäenpää T., Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.

[43] Shan, C., Gong, S., and McOwan, P.W. 2009. Facial expression recognition based on local binary patterns: a comprehensive study, *Image Vis. Comput*. 27 (2009) 803–816.

[44] Mistry, K., Zhang, L., Neoh, S.C., Lim, C.P., and Fielding, B., (2016). A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. IEEE Transactions on Cybernetics. doi: 10.1109/TCYB.2016.2549639

[45] Elaiwat, S., Bennamoun, M., and Boussaid, F., A spatio-temporal RBM-based model for facial expression recognition, in: Pattern Recognition, 49 (2016), pp. 152-161.

[46] Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., Metaxas, D., Learning active facial patches for expression analysis, in: IEEE Conference on

Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2562–2569.

[47] Derkach, D., Sukno, F.M., Local shape spectrum analysis for 3D facial expression recognition, in: IEEE 12th International Conference on Automatic Face & Gesture Recognition, 2017.

[48] Jan, A., Meng, H., Automatic 3D facial expression recognition using geometric and texture feature fusion, in: 11th International Conference on Automatic Face & Gesture Recognition, 2015.