



Emotion Recognition from Facial Expressions Using a Genetic Algorithm to Feature Extraction

Laura Jalili, Jair Cervantes^(✉), Farid García-Lamont, and Adrián Trueba

Computer Sciences Department, UAEMEX (Autonomous University of Mexico State),
56259 Texcoco, Mexico

{lydominguezj, jcervantesc, fgarcial, atruebae}@uaemex.mx

Abstract. In this paper, we present a new method for emotion recognition from facial expressions. The proposed algorithm concentrates on only two specific areas (eyes and mouth), reducing features and descriptors and focusing only on these areas. The algorithm extracts characteristics from these two regions of the face and, in a subsequent process, eliminates the less significant characteristics or those that introduce noise into the classifier. The system allows obtaining a reduced set of features to improve the performance of the classification. In the experiments carried out, we obtain precisions of 99.56%. We evaluated the proposed algorithm on two benchmark datasets; we find that SVM consistently outperforms traditional machine learning techniques.

Keywords: Facial recognition · Vision system · Features selection

1 Introduction

The recognition of emotions is helpful for the identification of human behaviour. Understand human behaviour can help us to solve problems in several disciplines. Some of the applications are used in psychology, robotics, emotions in text, video games, intelligent environments, marketing [1], among many more. Last years, research in this area had increased, which has allowed the emergence of better techniques and the development of techniques to improve precision in systems to recognize emotions. Some research speaks of practical algorithms for improving precision, the aspects that influence are quality, and the extension of the data set, the preprocessing of the images, feature descriptors and the optimization of the parameters of the classifiers [2–4].

The most used techniques to improve precision are the feature selection stage and a fine selection of areas of interest. Several studies have concluded that feature selection intervenes in classification performance [5]. Feature selection techniques allow filtering the most discriminative characteristics and eliminating those with little discriminatory power or features that add noise to the classifier. The adequate feature selection stage has helped to increase the classification precision. Furthermore, the training and testing times can be reduced significantly. On the other hand, the identification of the face [6–8] and select areas of interest also reduces the processing and search for features in the entire image, concentrating only on small areas and reducing the processing significantly.

In this paper, we identify the six basic emotions according to Ekman [9]. In the proposal, two areas of interest of the face are selected: the eyes and the mouth. From these areas, we extracted textural features, LBP and HoG. Finally, a genetic algorithm is used as a selector of characteristics. Both techniques allow to significantly reduce processing by considerably reducing the work area and features used. In the experiments carried out, we used two benchmark data sets. In the characteristic extraction stage, we use the HoG, LBP, and Haralick textural descriptors as information to identify emotions, and a genetic algorithm is used to identify the most discriminative features in the data set. The present work makes it possible to focus the search for features on only two areas of the face and reduce the number of features used to only those with high discriminative power. This work's objective is to reduce the region of interest and then reduce features by selecting the most discriminating and finding out the combination of features that can differentiate between types of facial emotions.

The work is structured in the following sections: Sect. 2 presents state of the art in the area of selection of characteristics for the recognition of emotions, analyzing the different selection algorithms, descriptors, and classifiers for the identification of emotions, also, to raise the different applications that these systems have today. In Sect. 3, we show the preliminaries. Section 4 presents our proposed methodology. In Sect. 5, we perform an analysis of the results. Finally, in Sect. 6, we present the conclusions and future work.

2 State of the Art

Identifying emotions is a research topic of great interest due to many implementations in different areas. Identifying the moods in people allows obtaining essential data, such as likes, approvals, rejections, and interest towards a specific situation or objects. These data are valuable because they can be raised in a large number of applications and investigations. The emotion classifier's accuracy depends on the quality of the data sets, the preprocessing, feature extraction techniques, and parameters selected in classifiers. Working with the most discriminative features allows us to reduce the identification time and increase precision. Most of the investigations seek to eliminate redundant features from their dataset by using algorithms. Some of these are genetic algorithms [10], ant colony system (ACS), Cuckoo search, sequential direct selection (SFS) [11], PSO optimization, among others. As methods of selecting discriminative features. Some investigations perform combinations of metaheuristic search algorithms; for example, in [12], a non-dominated classification genetic algorithm II (NSGA-II) and a Cuckoo binary search for characteristics selection. The authors obtain subsets of the initial population using the Relief algorithm. In their results, the authors report an improvement in its times and an increase in its precision using SVM and treeBagger classification algorithms. However, many investigations are focused only on the genetic algorithm for the selection of the characteristics, in most of the investigations, they propose the extraction of characteristics using popular descriptors such as HoG, LBP, Transformed Fourier, Gabor, among others, and the classifiers most used in research are deep networks and SVMs. Research shows that the selection of characteristics with genetic algorithms decreases the error rate, improving identification [13].

Another essential aspect is the selection of the areas of interest of the face. The muscles of the face have different movements that together express emotions. The emotions

are shown in the eyelids when manifesting an eye movement, or an expression of the eyes performs the movement in the muscles around the eyes, lifting the eyebrows, mouth movements, frowning cheek lift, and chin tension [14]. The expression of emotions is transmitted mainly on the face. However, the movement of the eyebrows and nose is showed in the eye area. The nose's movement is shown in the mouth area, and finally, the movement in the chin and cheeks is dependent on the movement applied in the mouth. However, several investigations identify emotions by taking the person's entire face to extract features. Other investigations are based on masks with crucial points on the entire face to identify the relationship of movements and interpret the sets of movements associated with an emotion. In this paper, we use specific areas of the face that we consider discriminatory to identify emotions. In our case, we consider the mouth and eyes as our areas of interest to identify emotions.

3 Proposed Method

Each of the steps of a vision system has a significant impact on its performance; that is, the success of a machine vision system depends on the inter-dependent processes. This paper tries to solve the following problems: extracting information from images from regions of interest, improving performance by reducing the features, and automatically recognizing facial emotions. In this section, we describe in detail the proposed methodology. Figure 1 shows a flow chart of the proposed methodology. The first step detects human faces in the image, in the second step detects the mouth and eyes in the face. The initial detection of the face allows the system to improve the detection of regions of interest. In the third step, the features of the mouth and eyes regions are extracted then a genetic algorithm is used to select the best features. This step allows to eliminate features with low discriminative power and features that introduce noise in the classifier. The genetic algorithm allows getting the best features with the best discriminative power and obtaining the best combination of features. Finally, the vector of selected features is used for the classification of facial emotions. In this paper, we compare the results with several classifiers. Each step is explained in detail in the following subsections.

3.1 Detection of Areas of Interest

Preprocessing is an essential step in face recognition that consists of several techniques such as illumination normalization, alignment, resized and cropped images to assure that the location of the face in a picture is the same in all pictures and reduce the time of training. In this paper, we work with two areas of interest for the identification of emotions. Based on practice and research, we determine that these two areas give the movements of the face. When we are moving the nose, immediately there are movements in the mouth. Similarly, the movement of the forehead or eyebrows is shown in the eye area. In addition, emotions are expressed directly through macro expressions or micro-expressions located in these areas of the eyes and mouth. That is why we decided to focus only on these two areas of interest. Figure 2 shows how the system works in two steps in the proposed method. First, the system works with Haar features to improve the performance of the Viola-Jones algorithm, which performs the identification of the

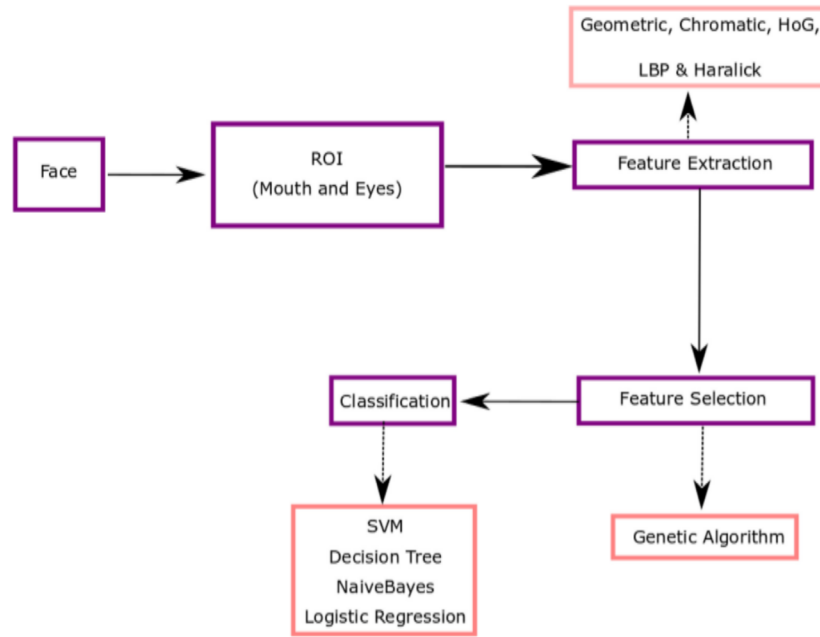


Fig. 1. Proposed Methodology

face. Then we used a cascade-type classifier trained with positive images of the eyes and mouths to automatically select the regions of interest. In this step, HoG and LBP type characteristics are extracted for the training of the cascade classifier. Once we identify the areas of interest, we cut and set aside working only in these areas.

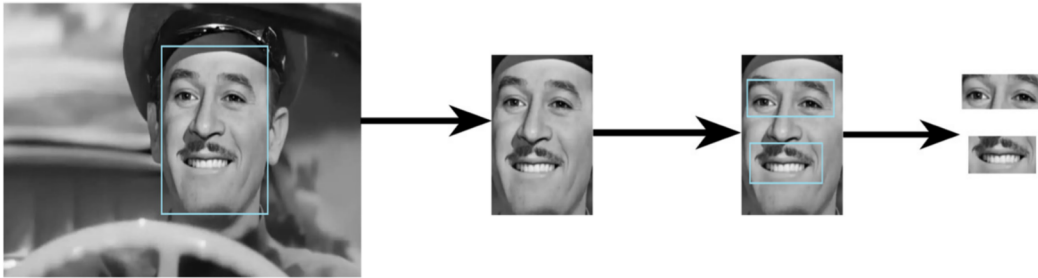


Fig. 2. Selection of regions of interest

3.2 Features Extractors

For each area of interest, we extract Haralick, HoG, and LBP features. We show the arrangement of our feature vectors in Fig. 3, and we carry out a concatenation of our characteristic vectors of each area of interest.

Haralick Texture Features. The Haralick textural features are textural patterns such as roughness, the softness of the area, smooth or thick, wavy, or irregular. These characteristics allow us to identify an object within an image. These characteristics help us

identify similar patterns in the object of interest, creating a relationship between shades in grayscale. The texture is obtained according to the consistent distribution of patterns and colours of an object. Haralick's method uses a co-occurrence matrix of gray levels using adjacent pixels. Texture analysis is performed by obtaining the frequency of the levels of pairs in the pixels with different distances and directions. Haralick obtains 14 texture descriptors.

LBP Features. The LBP descriptor is one of the most used techniques for pattern recognition due to its performance; even if there are somewhat complicated environments, LBP is invariant to lighting and translations. LBP works with monotonic changes in the scale of grays. This characteristic allows us to identify the same object or texture even though it has a darker or lighter tonality.

LBP descriptor works in a block size of 3×3 . The center of the block is used as a threshold for the remaining pixels in the block and encoded the computed threshold value into a decimal value as follows:

$$LBP = \sum_{i=0}^{P-1} s(n_i - G_c)2^i \quad (1)$$

$$s(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Where G_c represents the value of the center pixel in the block. P is the number of pixels in the block, n_i represents the i th neighboring pixel. Finally, a histogram of size 2^P is obtained from all the LBP code.

HoG Features. HoG descriptor is a technique for the extraction of shape features. Although it is also used for texture features, the algorithm works with the structure or shape of objects. HoG generally performs processes to obtain edges of objects. HoG gets the edges of the objects in the image. It calculates the magnitude and direction according to the intensity value of each pixel of a small spatial region referred as "cell". Finally, the algorithm obtains a histogram of the magnitude, depending on the direction angle of each pixel. These values allow us to distinguish different local changes around the pixels with the contrast change and the local shape.

Let L be an intensity (grayscale) function describing the image. The image is divided into cells of size $n \times n$ pixels, and the orientation $\theta_{x,y}$ of the gradient in each pixel is computed as follows:

$$\theta_{x,y} = \tan^{-1} \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \quad (3)$$

The orientations θ_i^j , $i = 1, \dots, n^2$ in the cell j are quantized and accumulated into a m -bins histogram.

3.3 Feature Selection

The number of extracted features is enormous and, in some cases, can affect the classification accuracy by introducing noise into the classifier. We decided to apply a genetic

algorithm to select features to eliminate redundant characteristics with low discriminative power or not contribute to the classification precision. One of the most used feature selection techniques is the genetic algorithm. This is a stochastic method for function optimization inspired by the mechanics of natural genetics and species evolution. Genetic algorithms operate on a population of individuals to produce better approximations in each iteration.

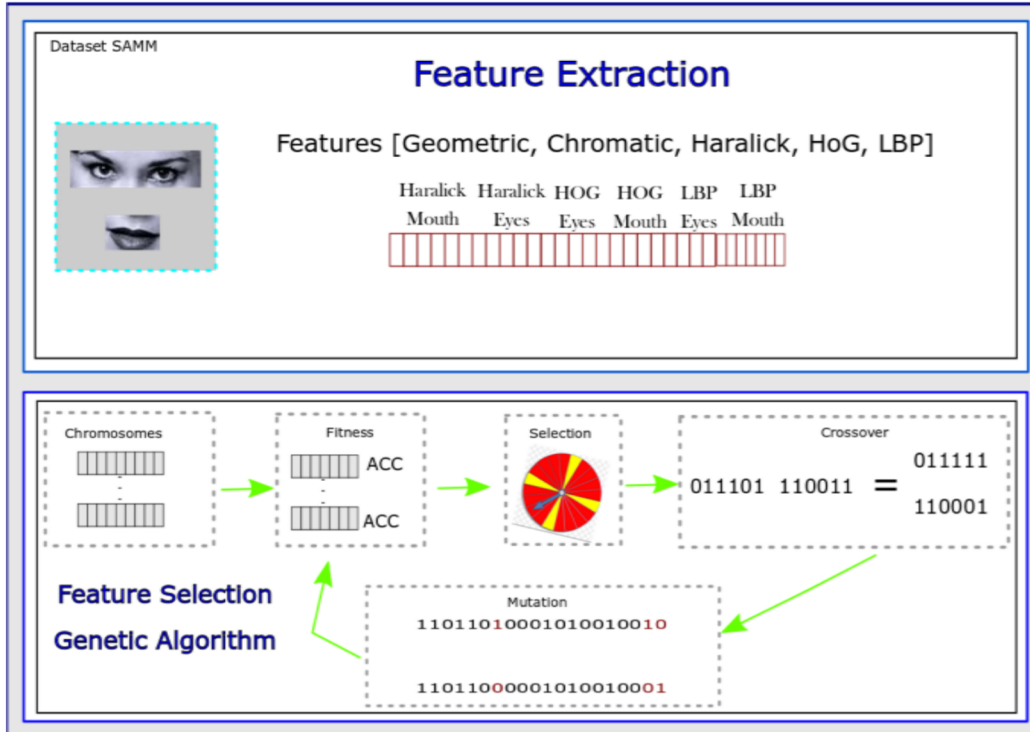


Fig. 3. Feature extraction and features selection of the proposed method.

For this stage, a basic genetic algorithm is implemented, using four different classification methods to obtain subsets of four features. In our problem, we use binary string encoding, in which a chromosome represents the inclusion/exclusion of the set of features. The chromosome is represented as bit strings of 0's and 1's. In the chromosome, 1 represents that the feature is selected and 0 otherwise. The algorithm starts with a random population of feature subsets (chromosomes). Each chromosome is evaluated, measuring its ability to predict a set of labels based on the accuracy. The algorithm replaces the initial population with a new one that includes features from different chromosomes that present a higher classification accuracy. This process is repeated until a number of generations are reached, or accuracy is reached. At each generation, a new population is created to improve the chromosome population by selecting individuals according to their fitness level and recombining them together using operators of selection, mutation, and crossover.

In our experiments, use binary encoding, elitism, size of population = 30, number of generations = 100, Crossover operator: 2-point crossover with probability =

0.7, mutation operator: Bit flip with probability = 0.02 and roulette wheel as selection operator.

In Fig. 2, we show the features extraction stage of two areas of interest selected and the final arrangement of our characteristic vectors. In this figure, we also mention our feature selection process with the genetic algorithm to optimize our precision and execution time.

3.4 Classification Techniques

In this section we describe the classification techniques used in our experiments.

Support Vector Machines (SVM). The SVM classifier has turned out to be one of the best classifiers used in many applications, and according to recent research, the SVM generally provides good precision for the recognition of emotions. The basic principle of an SVM is to create an optimal hyperplane over a set of data with linearly separable classes.

Logistic Regression. Logistic regression is a predictive algorithm based on the statistics and probability of the data. Logistic regression allows us to work with binary classes using a sigmoid activation function. The logistic regression estimates possible relationships between the dependent and independent variables of the data and defines the possible classification of the data according to the predisposition of the curve of an activation function. When the curve tends to be infinite positive, the classification is 1; otherwise, it is 0.

Random Forest. Decision trees are predictive algorithms, and this algorithm allows us to create a model in the form of a tree with the data. It is based on data impurity metrics and data probabilities preset in a class. The tree is made up of internal nodes and leaf nodes, where the internal nodes are based on the impurity metric; a decision is made, iteratively constructing bifurcations until reaching the leaf nodes where the complete path is associated with a class. Decision trees work by calculating the probability of each class, which is used to obtain a profit from each node. There are pre-labelled data for each of the existing classes to later calculate the impurity of each of the characteristics or variables of the data, being the nodes exhaustive for decision-making and new bifurcations. Decision trees are based on recursion. In each node, data are divided by eliminating characteristics that were already used as results of the nodes.

Naïve Bayes. Naive Bayes is a fast and easy to implement classification technique. The principal advantage of Naïve Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification. On other hand, a naive Bayes classifier considers each of the features to contribute independently. It is a disadvantage because, in most real-life cases, the predictors are dependent, which impacts the classifier's performance. Despite their simplicity, naive Bayes classifiers have worked quite well in many complex real-world applications.

4 Experimental Results

In this section, the parameters selection technique is shown, also data normalization and experimental results obtained with the proposed method.

4.1 Data Sets

In the experiments, we use two data sets that are described as follows.

SAMM Database. The SAMM dataset was created at the University of Manchester is a dataset of actions and spontaneous micro-movements with demographic diversity and is based on the facial action coding system (FACS). This set contains 159 spontaneous micro-facial movements of 32 participants of 13 different ethnic groups, a total of 3634 images, and six emotions (Joy, Surprise, Anger, Disgust, Sadness, and Disgust) [15, 16].

SMIC Database. The SMIC dataset created by the University of Oulu is a spontaneous micro-expressions database, with 164 micro-expressions from 16 people. SMIC contains 1909 images and five emotions (Joy, Surprise, Disgust, Sadness, and Disgust) [17, 18, 19].

4.2 Feature Extraction and Selection

After of selection of the regions of interest, we extract features from the mouth and eyes images. Features are extracted one by one and concatenated into a single feature vector. The final feature vector T was stored in a $m \times 3824$ size array containing m images with 3824 features. The features are obtained using LBP, textural and HoG features as described above. All the extracted features were normalized with mean zero and standard deviation equal to 1. In this stage, we use a GA to select the best features. For each classifier, we select the best features based on accuracy. In all the experiments, we use binary encoding, elitism, population with 30 individuals, 100 generations, 2-point crossover with probability = 0.7, bit flip mutation with probability = 0.02 and roulette wheel selection operator.

4.3 Parameter Selection

In all used classifiers, optimal parameters were obtained by cross-validation and grid search. Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. On the other hand, grid search exhaustively searches all parameter combinations obtaining the best. For SVM, the regularization parameter C which induced the best average test error was picked from the discrete values $10^{-4}, \dots, 10^4$.

5 Results

In the experiments, all data sets were normalized, and cross-validation was used with $k = 10$. Table 1 shows the results obtained with 2 datasets. In the Table Acc represents the Accuracy, TP represents the true positives, FP represents the false positives, and Fm represents the f-measure metrics. For each classifier used, accuracies obtained with each individual set of characteristics are reported. The metric used to evaluate the classifier's performance was Accuracy, which is obtained from the classifier hits divided by the total of the data set.

Finally, we show the effectiveness of the proposed method using feature selection and with the entire dataset. Table 1, shows the results obtained with all the features and using only the features selected by the genetic algorithm. The first column describes the results obtained with all features (AF) and the reduced features with the GA (ReF). In the Table, SVM represents the results obtained with Support Vector Machines, in our experiments, we use the Sequential Minimal Optimization algorithm, RF represents the results obtained with Random Forrest, Bayes (Naïve Bayes), and LR represents the results with logistic regression.

In the Table, the maximal predictive accuracy to the SAMM dataset is obtained with SVM classifier 99.56% using the GA to reduce the features, and 98.72% with all the features. On the other hand, the maximal predictive accuracy to SMIC dataset is also obtained with SVM classifier 98.14% using the GA to reduce the features, and 97.68% with all the features.

From the experimental results, one can see that eliminating features in the dataset can significantly improve the performance of the classifiers and make the algorithm fast. Moreover, using individuals with all the features and elitism in the GA, guarantees that the GA improves the performance. The table shows that the results improve less than 1 percentage point when we select features. Only in the SMIC dataset using the Bayesian classifier, the improvement is 6% points.

Figures 4 and 5 show the confusion matrices obtained with the different classifiers (SVM, Bayes, random forest and regression logistic) for SAMM and SMIC datasets. Figure 4 shows that the best-predicted emotions with SVM are dislike and sadness with 100% and the worst predicted is surprise with 99.1%. However, each classifier has its difficulties in predicting some emotions. Figure 5 shows that best-predicted emotions are again obtained using SVM, and it is sadness with 100%, and the worst predicted is happiness with 96.9%. On the other hand, the worst results are obtained with Bayes classifier with the anger emotion as the best classified with 96.5% and sadness emotion as the worst predicted with 71.7%. Figure 5 shows that the worst predicted emotions are again obtained using the Bayes classifier with the SMIC dataset.

Experimental results presented in this section indicate that the proposed method can be employed for face recognition. Compared to the same method with all the features. The proposed method attained the highest classification accuracy using the GA on SAMM and SMIC datasets. In our experiments, we can observe that the use of elitism permits us to improve the accuracy. Due to the above, in the population generation, we generate an individual with all the features. GA starts with the individual with all features. In the worst case, at least the fitness of this individual will be preserved.

Table 1. Performance of the two datasets with different features.

SAMM		SMIC											
	Subset	Acc	TP	FP	Recall	Fm	ROC	Acc	TP	FP	Recall	Fm	ROC
ReF	SVM	99.559	0.996	0.001	0.996	0.996	1.0	98.14	0.981	0.006	0.98	0.981	0.99
	Bayes	86.075	0.861	0.029	0.861	0.859	0.97	78.29	0.783	0.063	0.78	0.784	0.93
	RF	98.87	0.990	0.003	0.990	0.990	1.0	93.61	0.936	0.021	0.93	0.936	0.99
AF	LR	99.312	0.993	0.002	0.993	0.993	1.0	95.87	0.959	0.013	0.95	0.959	0.99
	SVM	98.721	0.987	0.003	0.987	0.986	0.99	97.68	0.977	0.007	0.97	0.977	0.99
	Bayes	85.828	0.858	0.029	0.858	0.856	0.95	72.76	0.728	0.073	0.73	0.729	0.89
	RF	98.514	0.985	0.004	0.985	0.985	1.0	93.11	0.931	0.023	0.93	0.931	0.99
	LR	98.348	0.983	0.005	0.983	0.981	0.99	95.32	0.953	0.014	0.95	0.953	0.99

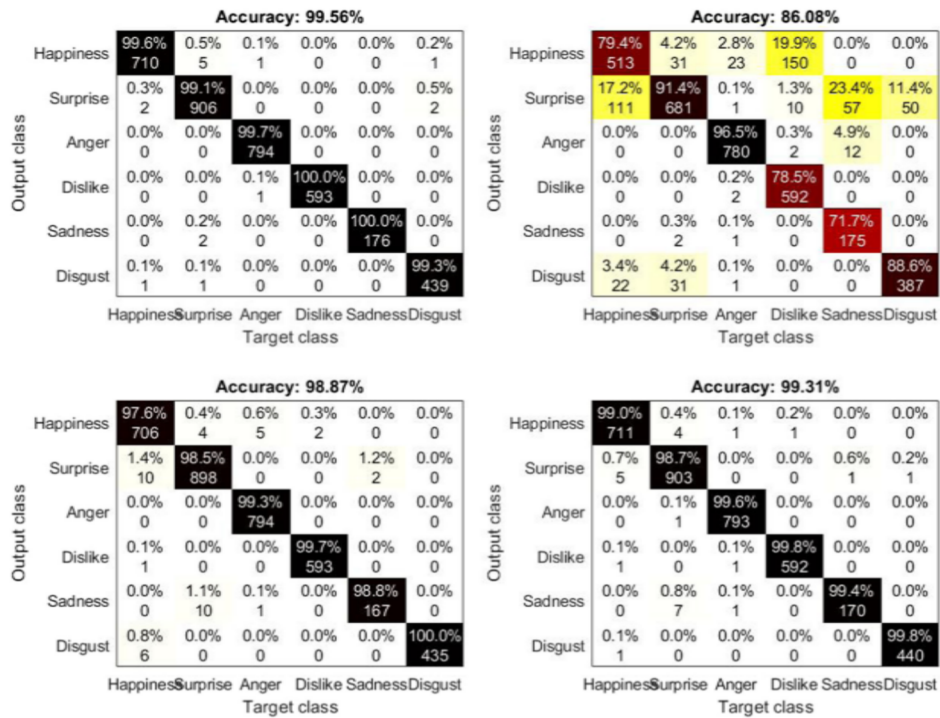


Fig. 4. Confusion matrices of the classifiers with SAMM dataset using reduced features with GA

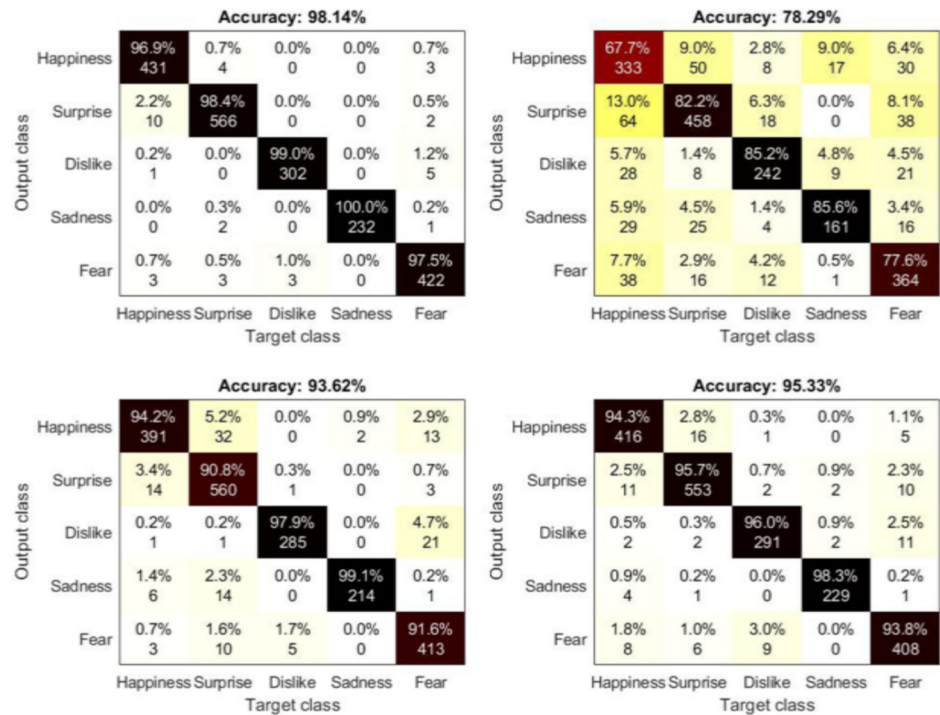


Fig. 5. Confusion matrices of the classifiers with SMIC dataset using reduced features with GA

6 Conclusions

In this paper, a novel system for emotion recognition is proposed. The proposed algorithm selects the region of interest in two stages. First, identify the face region in the image and then identify the mouth and eyes as principal regions of interest. The algorithm extract features from two regions in the face and uses a GA to reduce the search area and the features. The proposed system is evaluated using two benchmark datasets. In the experimental results, we can see that selecting regions of interest helps the system improve performance.

Moreover, elitism in the genetic algorithm for feature selection increased the precision of some of the classifiers. Another essential aspect that genetic algorithm gives us is to reduce the number of characteristics. This helps speed up computing times. In this paper, we work with 3824 features initially. However, we observed that the SAMM data set was reduced by 53.75%, and the SMIC data set was reduced by 52.05%. The performance of the developed model was tested using k fold cross-validation, and results revealed that the model is a good predictor.

Acknowledgments. This study was funded by the Research Secretariat of the Autonomous University of the State of Mexico with the research project 4996/2020/CIB.

References

1. Torres, E.P., Torres, E.A., Hernandez-Alvarez, M., Yoo, S.G.: Emotion recognition related to stock trading using machine learning algorithms with feature selection. *IEEE Access* **8**, 199719–199732 (2020)
2. Kauser, N., Sharma, J.: Automatic facial expression recognition: a survey based on feature extraction and classification techniques. In: 2016 International Conference on ICT in Business Industry & Government (ICTBIG) (2016)
3. Ozdemir, M.A., Elagoz, B., Soy, A.A., Akan, A.: Deep learning based facial emotion recognition system. In: 2020 Medical Technologies Congress (TIPTEKNO) (2020)
4. Neeru, N., Kaur, L.: Face recognition based on LBP and CS-LBP technique under different emotions. In: 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCC) (2015)
5. Belkov, D., Purtov, K., Kublanov, V.: Influence of different feature selection approaches on the performance of emotion recognition methods based on SVM. In: 2017 20th Conference of Open Innovations Association (FRUCT) (2017)
6. Li, B., Zheng, C.-H., Huang, D.-S.: Locally linear discriminant embedding: an efficient method for face recognition. *Pattern Recogn.* **41**, 3813–3821 (2008)
7. Zhao, Z.-Q., Huang, D.-S., Sun, B.-Y.: Human face recognition based on multi-features using neural networks committee. *Pattern Recogn. Lett.* **25**, 1351–1358 (2004)
8. Gui, J., Jia, W., Zhu, L., Wang, S.-L., Huang, D.-S.: Locality preserving discriminant projections for face and palmprint recognition. *Neurocomputing* **73**, 2696–2707 (2010)
9. Ekman, P.: Facial expression and emotion. *Am. Psychol.* **48**, 384–392 (1993)
10. Cervantes, J., Li, X., Yu, W.: Imbalanced data classification via support vector machines and genetic algorithms. *Connect. Sci.* **26**, 335–348 (2014)
11. Liogiene, T., Tamulevicius, G.: SFS feature selection technique for multistage emotion recognition. In: 2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE) (2015)

12. Yildirim, S., Kaya, Y., Kılıç, F.: A modified feature selection method based on metaheuristic algorithms for speech emotion recognition. *Appl. Acoust.* **173**, 107721 (2021)
13. Choudhary, D., Shukla, J.: Feature extraction and feature selection for emotion recognition using facial expression. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM) (2020)
14. Yuki, M., Maddux, W.W., Masuda, T.: Are the windows to the soul the same in the east and west? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J. Exp. Soc. Psychol.* **43**, 303–311 (2007)
15. Davison, A., Merghani, W., Yap, M.: Objective classes for micro-facial expression recognition. *J. Imaging* **4**, 119 (2018)
16. Yap, C.H., Kendrick, C., Yap, M.H.: SAMM long videos: a spontaneous facial micro- and macro-expressions dataset. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (2020)
17. Li, X., et al.: Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Trans. Affect. Comput.* **9**, 563–577 (2018)
18. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikainen, M.: A spontaneous micro-expression database: inducement, collection and baseline. In 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (2013)
19. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Recognising spontaneous facial micro-expressions. In: 2011 International Conference on Computer Vision (2011)
20. Jalili, L.D., Morales, A., Cervantes, J., Ruiz-Castilla, J.S.: Improving the performance of leaves identification by features selection with genetic algorithms. In: Figueroa-García, J.C., López-Santana, E.R., Ferro-Escobar, R. (eds.) WEA 2016. CCIS, vol. 657, pp. 103–114. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50880-1_10