Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in e-ISSN: 3009-075X

Volume 3 Issue 3

Article Number: 240119

Optimizing Facial Expression Recognition with Biogeography-Based Feature Selection

Garima Sharma*[1] and Latika Singh[2]

[1]Department of Computer Science, The NorthCap University, Gurugram, Haryana, India 122017
[2]School of Engineering and Technology, Sushant University, Gurugram, Haryana, India 122003

Abstract

Facial expression recognition remains a challenging task in computer vision due to factors such as occlusion, variable lighting, and camera angles. Efficient extraction and selection of relevant features are crucial for accurate recognition. This paper introduces a novel metaheuristic-based approach for feature selection and classification using a Biogeography-Based Optimization (BBO) algorithm. The BBO algorithm optimizes the recognition accuracy of a Support Vector Machine (SVM) by selecting the most relevant features through cross-validation accuracy as the evaluation criterion. The proposed BBO-SVM model is tested on three public databases—JAFFE, MUG, and CK+—and demonstrates superior performance compared to traditional filter-based approaches. Notably, the model achieves a recognition accuracy of [insert specific accuracy], outperforming several existing methods. This study highlights the potential of BBO for enhancing facial expression recognition systems.

Keywords: Facial Expression Recognition System;, Biogeography-Based Optimization; Feature Selection; Support Vector Machine; Wrapper-Based Feature Selection

1 Introduction

The development of intelligent systems facilitating human-computer interaction has sparked significant interest in computer vision applications across various domains, including healthcare, agriculture, intrusion detection, and surveillance. Among these, automatic facial expression recognition systems have been a prominent area of research for several decades. The process of detecting expressions from digital facial images is typically divided into four stages: face detection, feature extraction, feature selection, and expression classification. Face detection is the initial and most crucial step, as the face conveys essential information about a person's emotions through various gestures, such as eye and lip movements. This step isolates the facial region from the image background, enabling subsequent processes to focus on relevant data. Following face detection, the feature extraction process retrieves pertinent information from the face, which is essential for accurate expression recognition. This paper primarily focuses on the feature selection process[1], examining the impact of various feature selection mechanisms available in the literature to extract an optimal subset of features, thereby enhancing overall system performance. Feature selection is a critical step, as robust expression classification relies on the identification of efficient and distinct features. By reducing redundant and irrelevant information, the feature selection process aims to represent input features more effectively, improving classifier recognition rates and reducing the dimensionality of the feature set. Feature selection methods are broadly categorized into three types: filter methods, wrapper methods, and embedded methods. Filter methods rank features based on intrinsic dataset properties without considering the classifier model, making them popular for high-dimensional datasets due to their scalability and low risk of overfitting.

Wrapper methods, on the other hand, use optimization techniques to select feature subsets based on classifier evaluation, taking into account feature dependencies. Embedded methods combine aspects of both filter and wrapper methods, utilizing

*Corresponding author: garimasharma@ncuindia.edu
Received: 26 March 2024; **Revised:** 15 May 2024; **Accepted**: 21 July 2024; **Published:** 04 August 2024
© 2024 Journal of Computers, Mechanical and Management.
This is an open access article and is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.
DOI: 10.57159/gadl.jcmm.3.3.240119.

a built-in approach for feature selection. In this study, we evaluate the performance of an expression recognition system after applying two filter-based feature selection methods. Additionally, we propose a novel Biogeography-Based Optimization and Support Vector Machine (BBO-SVM) approach for automatic facial expression recognition. Experiments are conducted on three publicly available facial expression databases—JAFFE, MUG, and CK+—with the proposed system achieving promising classification accuracies across all three datasets. The proposed work extends our previous research [2], which introduced a framework for the automatic recognition of facial expressions based on a fusion of deep and geometric features. In this study, we aim to maximize recognition accuracy by selecting the best-performing features from the original feature set using an evolutionary biogeography-based optimization algorithm. The key contributions of this paper are as follows:

- **Feature Extraction Process:** This paper employs a fusion of geometric and deep features for feature extraction. Geometric features include the x and y coordinates of 49 landmark points on the face, constructing a feature vector of 98 values. Deep features are extracted from images using a deep convolutional neural network, with 84 features obtained just before the softmax classification layer. These geometric and deep features are then concatenated to form a final feature vector containing 113 features.
- **BBO-Based Feature Selection and Optimization:** The generated feature set is optimized using the metaheuristic Biogeography-Based Optimization algorithm, which selects the best-performing features with the goal of maximizing classification accuracy.
- **Emotion Classification Using SVM:** Emotion classification is performed using an SVM model, which is integrated into the BBO algorithm as a fitness function. The BBO algorithm maximizes the classifier's cross-validation accuracy while selecting the optimal feature subset.

The classification results of the proposed BBO-SVM method are compared with two filter-based feature selection methods, namely mutual information and ANOVA. The results indicate that the proposed BBO-SVM approach outperforms these filter-based methods and achieves commendable recognition rates compared to other works in the literature. The remainder of the paper is organized as follows: Section 2 presents a literature survey of existing feature selection mechanisms, Section 3 details the databases used for system validation, Section 4 discusses the proposed methodology in detail, and Section 5 concludes the paper and outlines future research directions.

2 Related Works

Various feature selection methods have been studied and proposed in the literature. A study proposed in [3] utilized an iterative feature selection method based on Random Forest to reduce the appearance-based features extracted from face images. The results demonstrated a 22% improvement in performance with the use of this feature selection process. Another approach involved a hybrid of wrapper and filter feature selection mechanisms based on ant colony optimization, which outperformed many state-of-the-art methods and showed promising results [4]. Additionally, a wrapper-based method employing a genetic algorithm was introduced to reduce log-Gabor features, aiming to minimize classification error and achieve better results compared to filter-based methods [5]. Lajevardi and Hussain [6] presented an automatic facial expression recognition system that utilized a combination of filter and wrapper-based techniques to select the most informative features. The framework was validated through experiments on the CK+ and JAFFE databases, yielding desirable results. Similarly, a feature selection mechanism based on mutual information was employed to reduce primitive and complex features extracted from face images [7], with validation performed on the JAFFE dataset. Another study conducted feature selection on scale-invariant feature transform-based features using the Grey Wolf optimization metaheuristic algorithm, which were then used for emotion classification [8]. In the realm of social cognition, Zwick [9] examined the impact of reduced facial mimicry on expression recognition and depressive measures. An intelligent facial expression system was also proposed, which extracts facial features using modified local binary pattern descriptors and optimizes them with a variant of the Firefly algorithm [10]. Further research utilized spatial-temporal features, enhancing dynamic face information through a weight-based strategy for automatic facial expression recognition. Experiments on the CK+ and MMI face datasets demonstrated good recognition accuracies [11]. Moreover, a rough set theory-based self-learning feature reduction framework was introduced, achieving a high emotion classification rate by focusing on the most effective features in the mouth region [12]. In another study, a binary particle swarm optimization technique with an enhanced mutation operator was proposed to mitigate the premature convergence problem during feature selection. This wrapper approach employed a support vector machine for expression classification [13]. Lastly, Zhang et al. [14] presented a system that generates low-dimensional features by applying local fisher discriminant analysis (LFDA) on extracted local binary pattern (LBP) features, followed by classification using an SVM model built on the reduced feature set. The review suggests that feature selection is a critical step in developing an automatic facial expression recognition system, as it aids in selecting the most effective feature subset, thereby leading to higher classification accuracy and overall system performance. The remaining sections of this paper present a novel BBO-SVM feature selection approach, which identifies the optimal feature subset from deep and geometric features extracted from face images and performs classification using the SVM model. The proposed framework is evaluated on the datasets discussed in the next section.

3 Methods

3.1 Datasets Used

Three publicly available facial image databases were used in this study. These datasets are described in detail below:

3.1.1 JAFFE

The Japanese Female Facial Expression (JAFFE) database [15] contains 213 sample images from 10 female models of Japanese origin. The images are posed in seven universal facial expressions: angry, disgust, happiness, fear, sad, surprise, and neutral. All 213 images, which are in grayscale format, were used in this study for performing experiments. Figure 1 depicts sample images from the JAFFE dataset across six universal emotion classes.

Figure 1: Sample images from the JAFFE dataset

3.1.2 MUG

The Multimedia Understanding Group (MUG) database [16] contains images of 86 subjects posing in seven different facial expressions. Out of these, images of 52 subjects (including 52 men and 35 women) are publicly available. The subjects, aged between 20 to 35 years, are of Caucasian origin. This study used 712 randomly selected images from this dataset, covering the seven universal emotion classes. Figure 2 shows some sample images from the MUG dataset.

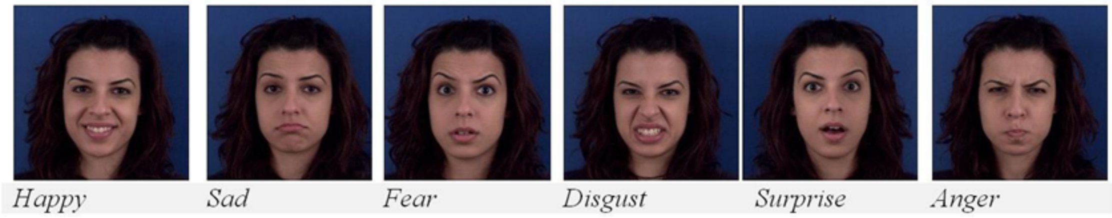

Figure 2: Sample images from the MUG dataset

3.1.3 CK+

The extended Cohn-Kanade (CK+) database [17], released in 2010, contains 593 video sequences collected from 123 subjects aged between 18 to 50 years, from various origins and genders. For this study, 325 images of 118 subjects were selected from the video sequences in the dataset to perform the experiments. Sample images from the CK+ dataset are depicted in Figure 3.

Figure 3: Sample images from the CK+ dataset

The number of samples and subjects from each dataset used in this study are summarized in Table 1.

The proposed BBO-SVM feature selection and classification framework for facial expression recognition operates on the deep and geometric features [2] extracted from the input face images. The complete process of automatic facial expression recognition can be divided into four stages, as depicted in Figure 4.

Table 1: Details of the datasets used

Dataset	No. of Samples	No. of Subjects
JAFFE	213	10
MUG	712	52
CK+	325	118

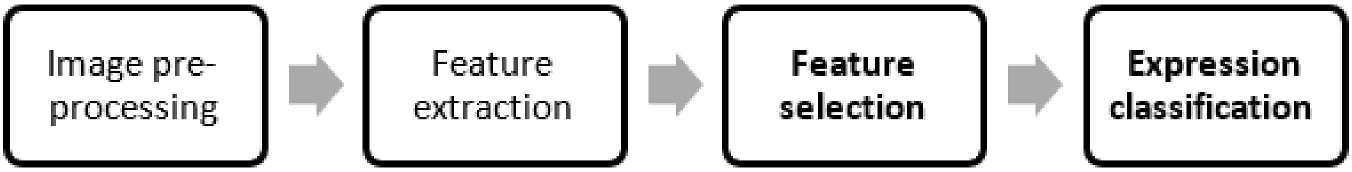

Figure 4: Different phases of the facial expression recognition system

3.2 Pre-processing and Feature Extraction

The pre-processing of the input images is the first step in building an automatic facial expression recognition system. This step is crucial for concealing unwanted details or enhancing essential details in the image. In this study, face detection is performed on the input image using the Viola-Jones algorithm [18] as part of the pre-processing step, since the face contains all the important information required for detecting expressions. After the face is extracted, a low-pass Gaussian filter is applied to remove noise from the image. Additionally, the illumination conditions of the image are enhanced using the contrast histogram equalization technique to capture sharp details such as edges and face contours. The pre-processed image is then used for extracting relevant features. Two kinds of features are used in this work: 1) Geometric features, and 2) Deep features.

3.2.1 Geometric Features

The geometric features are extracted from the face region in the form of x and y coordinates of various facial points. A total of 49 landmark points are selected on the face region using the iPar-CLR method [19], forming a feature vector with 98 coordinate values, as shown in Figure 5. In addition to these features, 15 distinct distance values between the facial points are calculated using the Euclidean distance and concatenated with the 98-dimensional feature vector, resulting in a final feature vector of size 113.

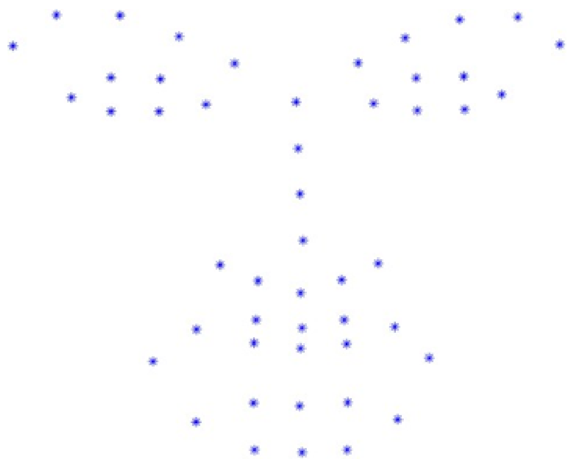

Figure 5: Facial landmark points used for feature vector construction.

3.2.2 Deep Features

Deep features are highly effective in detecting facial expressions due to the rich information they capture and the intricate details they provide. In this study, deep features are extracted from a 6-layer convolutional neural network (CNN) built from scratch [2]. The architecture of the network used for deep feature extraction is shown in Figure 6. These features are extracted from the network after the fully connected layers and just before the final classification layer, resulting in a feature vector with 84 deep features. These geometric and deep features are combined to form the final feature vector. To select the best-performing features and maximize the recognition rate of the system, feature selection is performed after the feature extraction process.

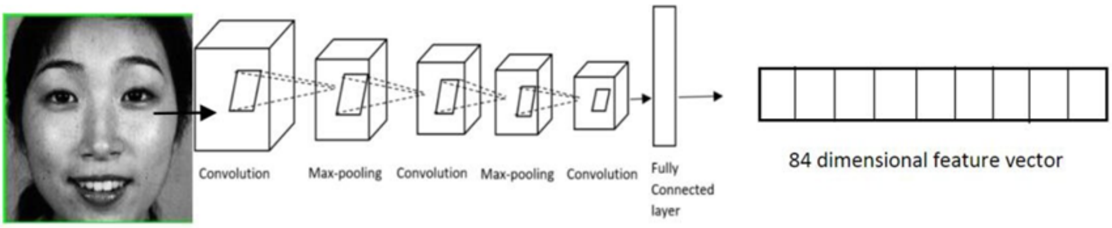

Figure 6: Deep features extraction process.

3.3 Feature Selection

Feature selection, or attribute selection, is the process of identifying the most contributing features while eliminating the least contributing ones. The goal is to reduce the dimensionality of the feature set while maximizing the performance of the classifier trained on these features. This paper proposes a wrapper-based feature selection approach that employs a Biogeography-Based Optimization (BBO) technique for optimal feature subset selection. The selection is based on the cross-validation accuracy of the Support Vector Machine (SVM), used as the fitness function to evaluate the performance of the subset on the classifier. The proposed approach iteratively performs feature optimization to maximize the expression recognition accuracy of the SVM. In addition, some filter-based methods were used to select the most appropriate features from the original feature vector. The performance of these filter-based methods is compared with the proposed wrapper-based BBO-SVM method. This section discusses the different feature selection mechanisms used in this paper.

3.3.1 Mutual Information

Mutual information feature selection [20] is based on information gain theory. Information gain is calculated by determining the amount of information gained by selecting or not selecting a particular feature in the feature set. Mutual information between two variables A and B can be defined by the Eq. 1:

$$I(A,B) = \int\int p(a,b)\log\left(\frac{p(a,b)}{p(a)p(b)}\right) dadb \tag{1}$$

where $p(a,b)$ represents the joint probability function of A and B, and $p(a)$ and $p(b)$ are the marginal distribution functions. Mutual information is a measure of similarity between $p(a,b)$ and the product of the marginal distributions $p(a)$ and $p(b)$. If the variables are completely independent, the mutual information will be zero, indicating that $p(a,b) = p(a)p(b)$. The feature selection mechanism aims to maximize this mutual information between the target variable b and the selected set of features.

3.3.2 ANOVA F-test Feature Selection

The Analysis of Variance (ANOVA) feature selection method [21] is a statistical measure in which feature selection is performed by analyzing the responses of the feature variables based on certain conditions. Two key measures are used in this method: p-score and F-score. The F-score determines the ratio between the mean variance values of two samples.

The p-score is the probability of discarding the F-score. Features are ranked based on their importance, and those that do not show significant differences in the target classes are discarded. A predefined number of higher-ranked features are selected to be part of the final feature set for emotion classification.

3.3.3 Biogeography-Based Optimization (BBO)

The BBO algorithm belongs to the class of metaheuristic algorithms and was first introduced by Dan Simon [22] in 2008. It is based on the natural process of biogeography, i.e., the study of the distribution of biological species. Species or populations are usually distributed within a geographical area with defined boundaries known as habitats. Each habitat has a Habitat Suitability Index (HSI), representing the desirability of a habitat for accommodating life. Factors such as temperature, rainfall, and vegetation patterns contribute to this HSI and are termed as Suitability Index Variables (SIVs). Populations tend to be higher in high-HSI habitats due to better conditions. However, due to high population density and competition, some native species migrate to neighboring habitats. This migration is termed emigration, and thus, emigration rates are high for high-HSI habitats. Conversely, the migration of species into high-HSI habitats is low due to the already abundant population, resulting in low immigration rates. In low-HSI habitats, immigration rates are higher compared to emigration rates. The reason for the high immigration rate is not the appropriateness of the habitat but its ability to accommodate more species. As varied species occupy the low-HSI habitats, the HSI tends to improve. However, if the HSI does not improve due to a lack of resources, some species may become extinct, creating opportunities for new species to move into the habitat. In the feature selection problem, each feature in the original feature vector is assigned a value considered as the SIV, constructing a real number array. Each candidate solution of the population can be thought of as a habitat, and the goodness of the candidate solution, or fitness function, is analogous to the HSI. The HSI of any habitat h can be represented as a function of SIVs, as given by the Eq. 2:

$$HSI(h) = f(SIV_i) \tag{2}$$

A candidate solution is a subset of selected features or SIVs. This indicates that a solution with a good HSI value shares the features or SIVs with the weaker solutions [23]. This migration is influenced by emigration and immigration rates, denoted by μ and λ, respectively. Figure 7 depicts the simple linear migration model, where E is the maximum emigration rate and I is the maximum immigration rate. The values of both E and I are set to 1.

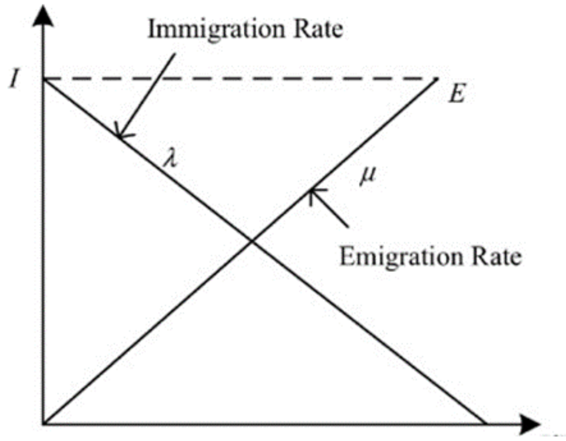

Figure 7: BBO migration model

In the proposed work, the best-performing features from the original feature set obtained during the feature extraction step are determined using the BBO algorithm. The SVM learning model is deployed, and its cross-validation accuracy is used to evaluate the performance of the selected feature subset, as shown in Figure 8.

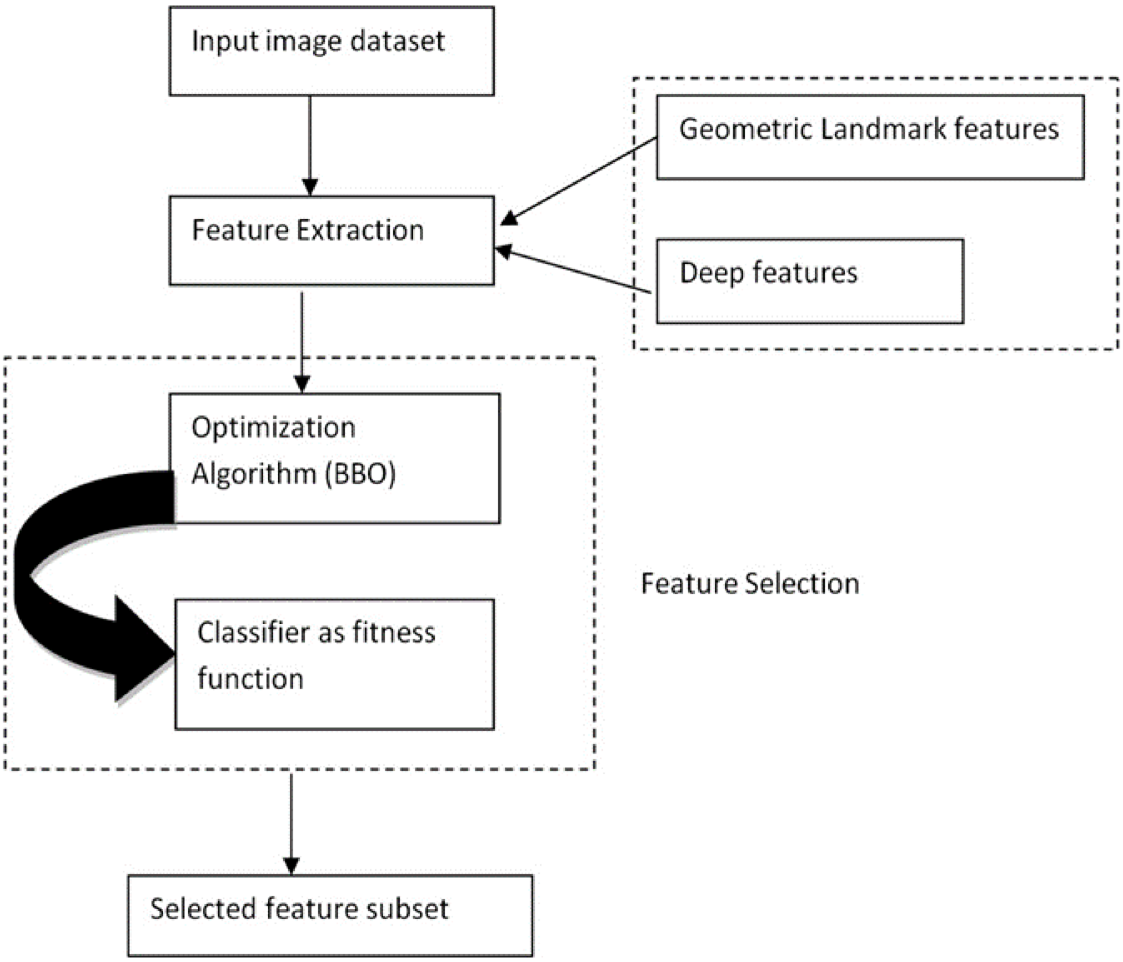

Figure 8: The proposed facial expression recognition system

3.4 Feature Selection Mechanism

The complete process of the feature selection mechanism proposed in this study is depicted in Figure 9. The detailed BBO-SVM algorithm for performing feature selection is presented in Algorithm 1 below.

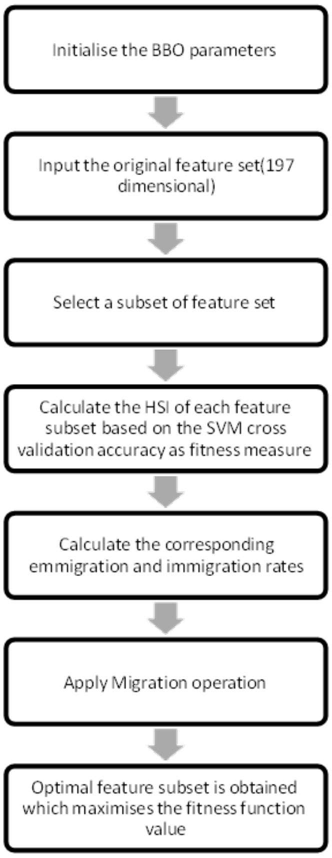

Figure 9: Proposed feature selection mechanism.

Migration operation [24] is performed so that the good habitats can share SIVs with the poor habitats. Emigrating habitats tend to share features with immigrating habitats. The value of the immigration rate is used as a basis to decide whether the habitat should be altered or not. The individual emigration and immigration rates for n habitats in the ith iteration can be calculated using the Eqns. 3 and 4:

$$\mu_i = \frac{Ei}{n} \tag{3}$$

$$\Lambda_i = I\left(1 - \frac{i}{n}\right) \tag{4}$$

A mutation operation is performed in the BBO algorithm to randomly alter the habitats using the mutation probability, which is a pre-determined value. The mutation rate for a solution s is given by the Eq 5:

$$\text{Mutation rate} = \text{Maximum mutation rate} \times \left(1 - \frac{P_s}{P_m}\right) \tag{5}$$

where P_s is the probability of the sth individual, and P_m is the maximum mutation probability.

The BBO optimization algorithm used in this work provides the optimal set of features based on the combination that maximizes the SVM cross-validation accuracy, making the emotion recognition process highly efficient and robust. The technique ensures that none of the features in the selected feature subset are redundant within a habitat, and the same number of features is selected for each candidate solution.

3.5 Expression Classification

The task of performing expression classification is an integral part of the feature selection mechanism discussed in the previous section. Since the wrapper-based approach is used for feature selection, the Support Vector Machine (SVM) classification model is embedded inside the BBO algorithm as the fitness function. SVM [25] is widely used for classification tasks and has consistently delivered strong classification results. Originally introduced by Vapnik [26] in statistical learning theory, SVM works by constructing a hyperplane that maximizes the distance between output classes.

4 Results and Discussion

The proposed system was evaluated on three publicly available datasets: JAFFE, MUG, and CK+. Facial expression classification was performed using the SVM classification model on each feature subset combination selected by the optimization

Algorithm 1 BBO-SVM Feature Selection Algorithm

Initialize the BBO parameters.
Generate a random population of initial habitats.
Calculate the fitness value or HSI for each habitat in the population. This fitness value corresponds to the cross-validation accuracy obtained by the SVM classification model for the selected feature subset combination.
for Generation = 1 to G **do**
    Calculate emigration rate μ and immigration rate Λ.
    Perform the migration operation to modify each elite habitat and recompute the HSI value.
    Perform the mutation operation to modify each non-elite habitat and recompute the HSI value.
    if maximum number of generations reached or optimal solution found **then**
        Terminate the loop.
    end if
end for
Output the selected feature subset and the maximum fitness value obtained.

algorithm, and the best feature subset—i.e., the combination yielding the maximum classification accuracy—was chosen. The initial feature vector, containing geometric and deep features, was passed through the BBO-SVM framework. The population size was set to 30, and the algorithm was run for 50 generations to obtain maximum recognition accuracy. The mutation probability was set to 0.05, and the emigration and immigration rates were set to 1, as shown in Table 2.

Table 2: Values of parameters used for the BBO algorithm

Parameter	**Value**
No. of generations (G)	50
Population size (N)	30
Max Emigration rate (E)	1
Max Immigration rate (I)	1
No. of elites retained	2
Mutation Probability (P_m)	0.05

Originally, SVM supports only binary or two-class classification. However, this work requires performing multiclass classification to classify the input into one of the seven emotion classes based on facial expressions. The one-versus-one strategy was used, constructing a multi-class classification model by building separate classifiers for each different pair of output labels. This leads to $\frac{N(N-1)}{2}$ classifiers being built, where N is the number of output classes. The LibSVM library was deployed for evaluating the candidate solutions obtained in each generation of the BBO algorithm. The radial basis function (RBF) was used as a kernel in the SVM, and 10-fold cross-validation was performed to ensure the reliability of classification results.

To compare the performance of the proposed wrapper-based feature selection mechanism, two other filter-based feature selection approaches—Mutual Information and ANOVA feature selection—were also applied to all three databases. It was observed that the proposed BBO-SVM feature selection mechanism outperformed the filter-based approaches. The effect of applying a feature selection mechanism on the extracted features is evident from the results, as the recognition accuracy of the facial expression recognition system increased compared to the earlier system [2] that performed the classification task immediately after the feature extraction process.

Table 3: Recognition accuracy achieved on JAFFE dataset

Method Used	**Recognition Accuracy**
SVM classification without any feature selection method	81.39%
Mutual Information Feature Selection	83.7%
ANOVA Feature Selection	86.04%
Proposed BBO-SVM Feature Selection	95.34%

Table 3 presents the recognition accuracy values obtained for the experiments performed on the JAFFE dataset. The recognition accuracy of 81.39% [2] was achieved when no feature selection mechanism was applied to the extracted combination of deep and geometric features. The accuracy improved to 83.7% and 86.04% when Mutual Information and ANOVA feature selection methods were applied to the extracted features, as shown in Figure 10. The results demonstrate that the proposed wrapper-based BBO-SVM feature selection model outperformed the filter-based approaches. The recognition performance of the facial expression recognition system improved by 13.95% compared to the framework that did not use any feature selection mechanism.

The highest recognition accuracy of 96.92% was achieved on the MUG database using the proposed wrapper-based approach, compared to the filter-based approaches which achieved recognition accuracy values of 87.4% and 89.69%, as

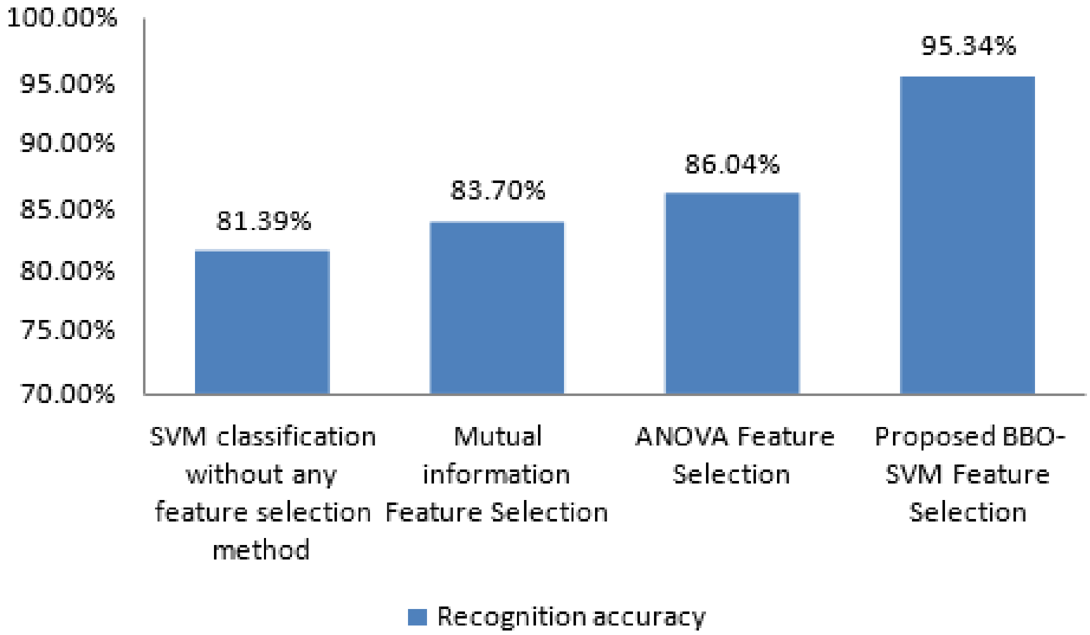

Figure 10: Comparison of recognition accuracy achieved using different methods on JAFFE dataset.

shown in Table 4. Figure 11 depicts the bar graph plot comparing the accuracies obtained after applying the feature selection mechanisms on the original feature vector.

Table 4: Recognition accuracy achieved on MUG dataset

Method Used	**Recognition Accuracy**
SVM classification without any feature selection method	86.01%
Mutual Information Feature Selection	87.4%
ANOVA Feature Selection	89.69%
Proposed BBO-SVM Feature Selection	96.92%

The results obtained on the CK+ dataset also validate the proposed work, as a high recognition accuracy of 98.5% was achieved using the proposed BBO-SVM approach, as shown in Table 5. The Mutual Information and ANOVA feature selection approaches also showed promising recognition accuracies of 86.1% and 88.3%, respectively, as depicted in Figure 12.

Table 5: Recognition accuracy achieved on CK+ dataset

Method Used	**Recognition Accuracy**
SVM classification without any feature selection method	84.5%
Mutual Information Feature Selection	86.1%
ANOVA Feature Selection	88.3%
Proposed BBO-SVM Feature Selection	98.5%

The experiments performed on all three databases suggest that the recognition accuracy of the facial expression recognition system can be improved by using feature selection mechanisms, as shown in Figure 13. The proposed BBO-SVM approach has proven to achieve good recognition accuracy on all the databases used in this study. As mentioned earlier, the algorithm was run for 50 generations, as it was observed that the recognition accuracy remained constant during the last few generations and did not improve further.

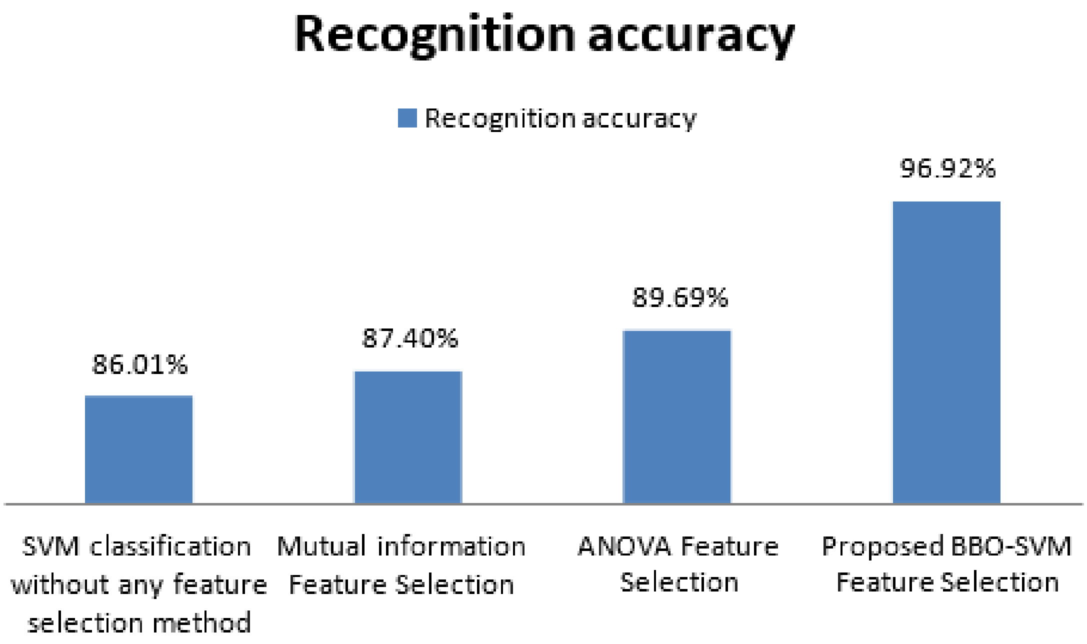

Figure 11: Comparison of recognition accuracies using different methods on the MUG dataset.

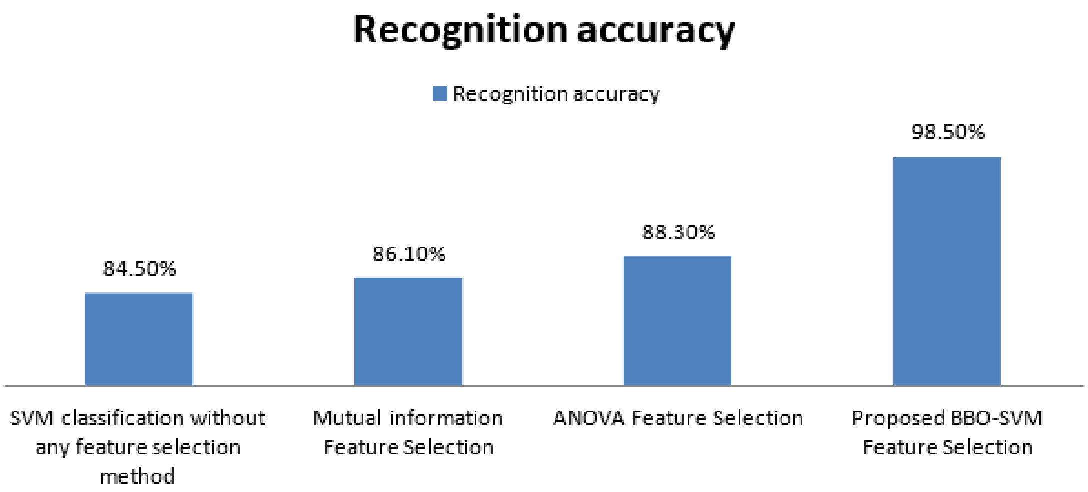

Figure 12: Comparison of recognition accuracies achieved using different methods on CK+ dataset.

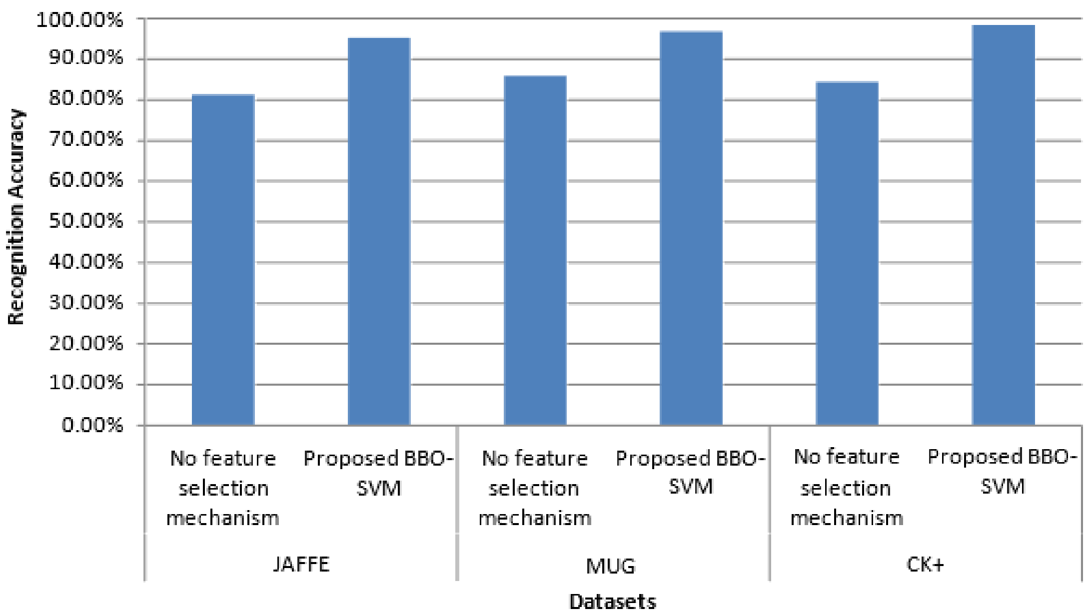

Figure 13: Effect of using feature selection mechanism on recognition accuracy of the facial expression recognition system.

The proposed system is also compared with other similar works available in the literature, as shown in Table 6. The results indicate that the proposed work has achieved comparable and even better results than some of the similar systems that perform the expression classification task.

Table 6: Comparison of the proposed facial expression recognition system with other works available in the literature

Study	JAFFE	CK+	MUG
[27]	89.6%	88.9%	–
[28]	–	94.93% (Highest accuracy)	–
[29]	92.31%	–	–
[30]	92.39%	–	92.60%
[31]	94.37%	–	95.24%
[32]	–	98.0%	97.2%
[33]	–	97.8%	95.5%
Proposed BBO-SVM	95.34%	98.5%	96.92%

5 Conclusion

In this paper, an optimization-based feature selection approach is presented for building a robust expression classification system that classifies input images into one of the seven universal emotion classes: angry, disgust, fear, happy, sad, surprise, and neutral. The meta-heuristic Biogeography-Based Optimization (BBO) algorithm is utilized to obtain an optimized subset of the feature vector, and the Support Vector Machine (SVM) is deployed as the classifier, integrated within the BBO algorithm. The cross-validation accuracy obtained for each combination of features in the selected subset is used as the cost or fitness value in the algorithm, which is maximized by performing migration and mutation operations within the BBO framework. The proposed BBO-SVM method works by improving the quality of habitats by selecting the most appropriate features and discarding the weaker ones. Additionally, the number of features selected in each generation is not fixed, allowing for dynamic generation of feature subset dimensions after each generation to ensure all relevant features are selected. The method was evaluated using three publicly available datasets: JAFFE, MUG, and CK+. The results demonstrate that the proposed method not only outperformed filter-based approaches such as Mutual Information and ANOVA feature selection but also surpassed various similar systems available in the literature. The highest recognition rate was achieved on the CK+ database with an accuracy of 98.5%, followed by 96.97% on the MUG dataset and 95.34% on the JAFFE dataset. In the future, other meta-heuristic algorithms such as Ant Colony Optimization, Particle Swarm Optimization, and Grey Wolf Optimization could be explored for feature selection. Additionally, hybrid methods combining two or more meta-heuristic techniques could be proposed. Furthermore, different classification models such as Random Forest and Naive Bayesian Classifier could be used in the wrapper approach to potentially enhance the performance of the system.

6 Acknowledgment

The authors express their gratitude to the Management of The NorthCap University and Sushant University for providing excellent computational facilities. The authors also extend their thanks to the anonymous referees for their valuable comments and suggestions, which significantly improved the quality of this paper.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding Declaration

This research did not receive any specific grants from governmental, private, or nonprofit funding bodies.

Author Contribution

Garima Sharma: Conceptualization, Methodology, Investigation, Visualization, Writing - original draft, review, and editing. **Latika Singh**: Resources, Validation, Writing - original draft, review, and editing.

References

[1] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers and Electrical Engineering*, vol. 40, pp. 16–28, Jan. 2014.

[2] G. Sharma, R. Vig, and L. Singh, “Facial expression recognition with fused deep and geometric features,” *International Journal of Disaster Recovery and Business Continuity*, vol. 12, no. 1, pp. 930–944, 2021.

[3] S. Gharsalli, B. Emile, H. Laurent, and X. Desquesnes, “Feature selection for emotion recognition based on random forest,” in *Scitepress*, pp. 610–617, Apr. 2016.

[4] M. Ghosh, R. Guha, R. Sarkar, and A. Abraham, “A wrapper-filter feature selection technique based on ant colony optimization,” *Neural Computing and Applications*, vol. 32, pp. 7839–7857, Jun. 2020.

[5] S. M. Lajevardi and Z. M. Hussain, “Feature selection for facial expression recognition based on optimization algorithm,” 2009.

[6] S. M. Lajevardi and Z. M. Hussain, “Automatic facial expression recognition: Feature extraction and selection,” *Signal Image and Video Processing*, vol. 6, pp. 159–169, Mar. 2012.

[7] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, and F. Battisti, *European Workshop on Visual Information Processing*. IEEE New Jersey, 2013.

[8] N. P. Nirmala Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, and R. Boothalingam R, “Grey wolf optimisation-based feature selection and classification for facial emotion recognition,” *IET Biometrics*, vol. 7, no. 5, pp. 490–499, 2018.

[9] J. C. Zwick and L. Wolkenstein, “Facial emotion recognition, theory of mind and the role of facial mimicry in depression,” *Journal of Affective Disorders*, vol. 210, pp. 90–99, Mar. 2017.

[10] L. Zhang, K. Mistry, S. C. Neoh, and C. P. Lim, “Intelligent facial emotion recognition using moth-firefly optimization,” *Knowledge-Based Systems*, vol. 111, pp. 248–267, Nov. 2016.

[11] X. Fan and T. Tjahjadi, “A dynamic framework based on local zernike moment and motion history image for facial expression recognition,” *Pattern Recognition*, vol. 64, pp. 399–406, Apr. 2017.

[12] Y. Yang, G. Wang, H. Kong, and P. Liatsis, “Self-learning facial emotional feature selection based on rough set theory,” *Mathematical Problems in Engineering*, p. 29, 2009.

[13] J. Wei, R. Zhang, Z. Yu, R. Hu, J. Tang, C. Gui, and Y. Yuan, “A bpso-svm algorithm based on memory renewal and enhanced mutation mechanisms for feature selection,” *Applied Soft Computing*, vol. 58, pp. 176–192, 2017.

[14] S. Zhang, X. Zhao, and B. Lei, “Facial expression recognition based on local binary patterns and local fisher discriminant analysis,” 2012.

[15] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, “The japanese female facial expression (jaffe) database,” in *Proceedings of third international conference on automatic face and gesture recognition*, pp. 14–16, 1998.

[16] N. Aifanti and A. Delopoulos, “Linear subspaces for facial expression recognition,” *Signal Processing: Image Communication*, vol. 29, no. 1, pp. 177–188, 2014.

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *2010 ieee computer society conference on computer vision and pattern recognition-workshops*, pp. 94–101, IEEE, 2010.

[18] Y.-Q. Wang, “An analysis of the viola-jones face detection algorithm,” *Image Processing on Line*, vol. 4, pp. 128–148, Jun. 2014.

[19] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Incremental face alignment in the wild,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1859–1866, 2014.

[20] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, “Normalized mutual information feature selection,” *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.

[21] R. M. Mehmood, R. Du, and H. J. Lee, “Optimal feature selection and deep learning ensembles method for emotion recognition from human brain eeg sensors,” *IEEE Access*, vol. 5, pp. 14797–14806, 2017.

[22] D. Simon, “Biogeography-based optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.

[23] L. Goel, D. Gupta, and V. K. Panchal, “Two-phase anticipatory system design based on extended species abundance model of biogeography for intelligent battlefield preparation,” *Knowledge-Based Systems*, vol. 89, 2015.

[24] W. L. Lim, A. Wibowo, M. I. Desa, and H. Haron, “A biogeography-based optimization algorithm hybridized with tabu search for the quadratic assignment problem,” *Computational Intelligence and Neuroscience*, 2016.

[25] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

[26] O. Chapelle, P. Haffner, and V. N. Vapnik, “Support vector machines for histogram-based image classification,” 1999.

[27] M. Song, D. Tao, Z. Liu, X. Li, and M. Zhou, “Image ratio features for facial expression recognition application,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, pp. 779–788, Jun. 2010.

[28] M. B. Abdulrazaq, M. R. Mahmood, S. R. M. Zeebaree, M. H. Abdulwahab, R. R. Zebari, and A. B. Sallow, “An analytical appraisal for supervised classifiers’ performance on facial expression recognition based on relief-f feature selection,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021.

[29] J. Kumari, R. Rajesh, and K. M. Pooja, “Facial expression recognition: A survey,” in *Procedia Computer Science*, pp. 486–491, Elsevier, 2015.

[30] M. A. Jaffar, “Facial expression recognition using hybrid texture features based ensemble classifier,” *International journal of advanced computer science and applications*, vol. 8, no. 6, 2017.

[31] Y. Rahulamathavan, R. C.-W. Phan, J. A. Chambers, and D. J. Parish, “Facial expression recognition in the encrypted domain based on local fisher discriminant analysis,” *IEEE transactions on affective computing*, vol. 4, no. 1, pp. 83–92, 2012.

[32] H. Ghazouani, “A genetic programming-based feature selection and fusion for facial expression recognition,” *Applied Soft Computing*, vol. 103, May 2021.

[33] D. Ghimire, J. Lee, Z. N. Li, S. Jeong, S. H. Park, and H. S. Choi, “Recognition of facial expressions based on tracking and selection of discriminative geometric features,” *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 3, pp. 35–44, 2015.