



Bsfoa-Dt Based Facial Emotion Recognition Using Facial Features

Zaid Hayder Akbar^{1*}, Fatima Dhamed Kazem² and Ahmed Mohammed Hussein³

⁽¹⁾ Ministry of Education, Wasit, Iraq.

⁽²⁾ Ministry of Education, Diyala, Iraq.

⁽³⁾ Ministry of Education, Babylon, Iraq.

Abstract. General Background: Automated facial emotion recognition serves as a vital component in modern human-computer interaction and affective computing systems. Specific Background: Convolutional neural networks extract high-dimensional representation features from expressive face images across various public datasets. Knowledge Gap: Existing approaches often retain redundant spatial attributes, increasing computational overhead and reducing overall classification efficiency. Aims: This paper proposes a hybrid methodology combining AlexNet deep feature extraction, Binary Starfish Optimization Algorithm feature selection, and Decision Tree classification. Results: Experimental evaluations on FER and FER+ datasets demonstrate significant dimensionality reduction from 128 to 60 features alongside a peak testing accuracy of 94.54%. Novelty: The integration of the binary starfish metaheuristic provides an optimal feature selection mechanism specifically tailored for discrete decision tree boundaries. Implications: This architecture offers a scalable and computationally efficient pipeline for real-time intelligent emotion monitoring systems.

Keywords: Facial Emotion Recognition, Deep Feature Extraction, Binary Starfish Optimization Algorithm, Decision Tree Classification, Dimensionality Reduction

OPEN ACCESS

ISSN 2503 3492 (online)

*Correspondence:

Zaid Hayder Akbar

raaed.mohammed@mohesr.edu.iq

Citation:

Zaid Hayder
Akbar and Fatima Dhamed
Kazem and Ahmed Mohammed
Hussein (2026) Bsfoa-Dt Based
Facial Emotion Recognition
Using Facial Features
Journal of Information and
Computer
Technology Education.
10.21070/jicte.v10i1.1722

INTRODUCTION

Over the past 30 years, economies worldwide have Facial expressions are among the most important indicators of emotions due to their connection to feelings. In 55% of cases [1], facial expressions convey emotions without words, providing clear evidence of whether a person is being truthful or not [2]. Therefore, emotion recognition techniques can be a valuable resource in various fields when combined with high-quality sensors, such as image processing, cybersecurity, robotics, and many other areas, including virtual reality [3].

Deep learning and machine learning algorithms have been widely used due to their efficiency and effectiveness in subject classification [4]. Furthermore, the addition of deep learning and machine learning algorithms improves the performance of most modern systems in various fields, such as medical image classification [5], [6], [7] language identification [8], [9], cloud fog networks [10], spam detection [11], emotion recognition in speech [12] [12], vehicle detection [13], and voice disorder detection [14], [15], [16].

Furthermore, deep learning and machine learning algorithms play a pivotal role in developing facial expression-based emotion recognition methods [17], [18]. The primary objective of employing these algorithms is to train models capable of classifying individuals with high accuracy and efficiency in the detection process [19],[20]. However, many current methods focus primarily on facial features while neglecting the surrounding background, which leads to the introduction of unnecessary information that can confuse the training process and negatively impact the performance of convolutional neural networks. In this context, this study focuses on five main categories of facial expressions: anger/displeasure, sadness/unhappiness, happiness/smiling, fear, and surprise/astonishment [21]. The proposed FER algorithm in this study is designed to analyse these expressions, aiming to classify images within these five basic emotional categories.

Generally, facial expression recognition techniques can be divided into two main approaches. The first approach relies on distinguishing expressions [22] by using specific classifiers to identify them, while the second approach is based on describing expressions based on features extracted from the face [23]. Within the facial movement coding system known as FACS [24], movement units are key indicators of different expressions, which can be identified by observing changes in facial muscle activity.

I. Literature Review

Moghaddam et al. [25] presented a three-stage facial expression recognition system. In this study, images of people displaying happy expressions were used. The image samples were collected from the Light Field Face Database (LFFD), which contains approximately 800 images representing a variety of facial expressions. The results showed that the highest level of accuracy achieved was 94.00%, specifically in recognizing the happy expression. However, the evaluation of this method's performance was limited to accuracy only, without considering other important evaluation metrics such as sensitivity, specificity, and mean G, which are essential for a more comprehensive assessment of classification systems.

On the other hand, Fang [26] proposed an approach to address the problem of image classification, particularly in cases where facial expressions are misclassified. This approach has focused on improving the process of distinguishing between different expressions by developing a regression algorithm that aims to identify the active pixels within the final layers of feature maps, which helps to highlight the most important areas in the image and improve the accuracy of the expression recognition process.

Subsequently, the facial features that have been misclassified will be highlighted. In the context of facial emotion recognition, the activity of particular image pixels plays a crucial role in determining the classification outcomes. The facial emotion samples were sourced from the Radboud Faces Database (RaFD), which comprises 67 individuals of varying genders and ages. Furthermore, each individual contributed 120 image samples. The facial emotions incorporated in this methodology include neutral, surprise, happiness, disgust, contempt, anger, sadness, and fear. This approach has achieved a classification accuracy of 90.97% in the recognition of human facial emotions. A four-part method for human emotion recognition has been proposed by Velagapudi et al. [27]. First, the frame video is used as input. Next, the features displayed by each emotion are recovered using geometric and local binary pattern-based feature extraction techniques. Then, a feature selection method is used to avoid the features that lower the recognition technique's performance. It chooses the characteristics that yield good results during the recognition procedure. In order to recognise emotions, the deep belief network (DBN) is given the chosen features. An effective Harris Hawk optimisation approach is used to optimise the weight parameter selection of DBN. Jangid, M. and others [28], a video database was utilized for the classification of facial emotions. This approach employed a Convolutional Neural Network (CNN) for the purpose of emotion classification. The method analyzes the video content and accurately recognizes emotions based on facial expressions. The model was developed using TensorFlow and Keras. OpenCV facilitated the extraction of frames, with this process occurring for each individual frame to determine the location of the facial image. A framework known as SSD (single-shot multi-box detector) was implemented for face detection, utilizing a compact ResNet-10 model. Ultimately, the CNN model classified the emotions derived from the identified facial images. Abdul-Al, M et al. [29] They used the VGG16 model, which was already trained on the ImageNet dataset, to gather features. Feature collection is done with the trained VGG16 model on the ImageNet dataset, along with PCA for reducing dimensions. They created a new method called VGG16 PCA-NN, which aims to enhance the accuracy of facial recognition; This method is validated using the Sejong Face Database, with a 70% training, 15% testing, and 15% validation split. A three-stage technique for detection emotions from facial image was suggested by Donuk, K., et al. [30]. The Fer+ dataset is used to train the CNN-based network in the first stage. The feature vector in the fully connected layer of the CNN network trained in the second stage is subjected to feature selection using the Binary Particle Swarm Optimisation method. Support Vector Machine is used to classify Selected features. Liang et al. [31] introduced a bidirectional long-short-term memory (BiLSTM) fusion network. This network comprises three components: the Deep Spatial Network (DSN), which is utilized for spatial display features; the Deep Temporal Network (DTN), which assists in leveraging time-related information; and the BiLSTM network, which is employed to generate six fundamental expressions.

Comparative evaluations against 13 distinct methods on the CK+, Oulu-CASIA, and MMI datasets demonstrated that the proposed

approach exhibits excellent performance. Alsemawi, M. R. M., et al. [32] used Yale Face database for facial emotion detection. In addition, the histograms of oriented gradients (HOG) is used to extract features from the images. The extracted features will feed the fast learning network (FLN) algorithm for the classification part to identify the images of facial emotions with respect to their subjects. A four-phase methodology was proposed by El Dahshan, K. A. et al. [33] First, the input image is preprocessed by cropping the Region of Interest (ROI) to obtain the desired region and remove non-significant parts. Second, the ROI is split up into numerous blocks, and the best (most effective) blocks are identified using the integral image. Third, to enhance system performance, the image down sampling algorithm is modified to make the new sub image smaller. Fourth, the DBN is used to determine the class of the emotion. QPSO is used to automatically optimise the values of DBN parameters rather than manually adjusting them. Liu, X. et al. [34] have been proposed two newly established geometric features: landmark curvature and vectorized landmark for facial emotion detection. These features are derived from facial landmarks that correspond to specific components of facial muscle movements. The method introduced integrates support vector machine (SVM) classification with a genetic algorithm (GA) to address a multi-attribute optimization challenge concerning feature and parameter selection.

II. PROPOSED METHOD

In this section, we are supposed to present the methodology of the proposed system for the recognition of facial emotion. The samples of the database are images of human faces in

different emotions that include 7 subjects. The proposed method is used at first, using one of the deep neural network architectures called AlexNet for extracting the facial features of each data sample; second, As you know the metaheuristic algorithms can optimize the feature selection problem with good accuracy and in an acceptable time; Meanwhile, the more important features that have a higher impact on the accuracy of emotion recognition are selected by the starfish algorithm. The best feature vector obtained at the end of DSFOA is classified in Decision tree. The flow chart of the proposed system is shown in Fig. 1.

A. Fer2013 and FER+ Dataset

The FER-2013 dataset, originally introduced at the International Conference on Machine Learning (ICML), was obtained through the Google image search API. There are 35,887 grayscale images of faces included in the dataset. Each of these gray images, sized at 48x48 pixels, has been classified into 7 unique emotional expressions. However, the labeling accuracy of the FER-2013 dataset is inferior when compared to that of the FER+ dataset [35]. In contrast, the FER+ dataset acts as an extension and contains emotion expressions that were labeled by 10 crowdsourced annotators instead of depending on a single-label output.

The initial FER+ dataset consists of 35,887 images categorized into 10 labels: neutral, happiness, surprise, sadness, anger, disgust, fear, contempt, unknown, and NF (non-face) [36]. Images within the dataset underwent preprocessing and cropping. In order to train our AlexNet architecture more accurately, “unknown” and “NF” labeled images were completely removed from the dataset. A selection of widely used facial expression datasets is displayed in Table 1.

The final statistical distribution of the adopted facial emotion

Table 1. Emotion expression datasets.

No.	Datasets	Reference
1	CFD	[37],[38]
2	JAFFE	[39],[40]

Table 2. Emotion expression data statistic.

	Google facial Expression	Comparison Data valid	Test
Neu	10299	1305	1302
Hap	7519	933	903
Sad	3453	451	467
Fear	678	75	66
Cont	173	21	22
Sur	359	442	423
Ang	2471	306	334
Dis	200	16	32
Total	20390	3549	3549

dataset in 8 active categories. Neu, Hap, Sad, Fear, Con, Sur, Ang, and Dis are abbreviations of Neutral, Happiness, Sadness, Fear, Contempt, Surprise, Anger, and Disgust, respectively. To rigorously evaluate the proposed model the data set is divided into three main subsets: a training set of 20,390 images, and validation and testing sets of 2,549 images each. Moreover, Table 2. shows an important class imbalance in the dataset. The distribution is highly dominated by the classes ‘Neutral’ and ‘Happiness’ with 10,299 and 7,519 images respectively while ‘Contempt’ and ‘Disgust’ are the minority classes with the least number of samples in all the phases. This detailed distribution reflects the natural data imbalance in facial expression datasets, which is an important factor that directly affects the evaluation metrics of the model and the recognition rate of minor classes.

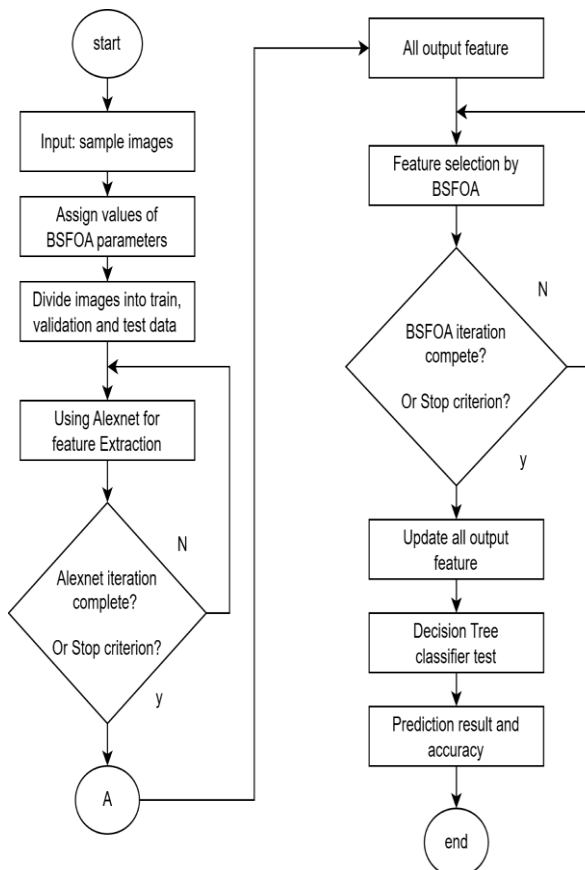


Fig. 1. Flowchart of the proposed method.

B. AlexNet architecture

AlexNet is still the base of deep convolutional neural networks for colour image recognition and classification. In most cases, the input size for this model is $224 \times 224 \times 3$ pixels. It is a huge architecture with about 62 million learning parameters distributed over eight main layers. In particular, the architecture consists of five convolutional layers, which are responsible for the main feature extraction, and three fully connected layers. Between these convolutional stages, overlapping max-pooling layers are inserted, keeping the computational cost manageable and reducing the spatial size of the feature maps.

One of the main reasons this network performs so well is that it moves away from standard functions like Tanh. Instead it employs the Rectified Linear Unit (ReLU) activation function, which has a huge advantage in the training speed. For example, a ReLU network can achieve 25% error on datasets like CIFAR-10 in almost 6 times less time than a Tanh network. Of course, 62 million parameters is a lot, which carries a high risk of overfitting. To this end, the fully connected layers of the architecture employ the Dropout [44] technique that disables randomly some of the neurones during training to force the network to learn more generalised patterns.

AlexNet has proved its ability to achieve very high accuracy on difficult datasets due to its strong structural setup. We do not use it for the final classification, rather, we use AlexNet as a feature extractor only in our work. The pre-processed images are given to the network for the deep layers to extract the distinctive facial features for each input sample which sets the stage for the actual emotion recognition process [44].

C. SFOA and BSFOA

In the past few decades, Metaheuristic optimization algorithms have attracted the attention of many researchers for providing optimal or near optimal solutions in various engineering, medical and ... issues. SFOA is one of these algorithms, presented in 2025 by Zhang et al. [45]. The inspiration for this algorithm is innovations and mathematical power. In particular, this algorithm is suitable for solving complex, scalable and multidimensional problems. This algorithm has exploration, hunting, and re generation processes for optimization problems. In the exploration phase of the algorithm when the search space has more than 5 dimensions, the algorithm must use all five arms to explore all around environment. In such situations, starfish's position is updated according to Eq. (1).

$$Y_{i,p}^T = \begin{cases} X_{i,p}^T + \alpha_1 (X_{best,p}^T - X_{i,p}^T) \cos \theta & , r \leq 0.5 \\ X_{i,p}^T - \alpha_1 (X_{best,p}^T - X_{i,p}^T) \sin \theta & , r > 0.5 \end{cases} \quad (1)$$

Where $Y_{i,p}^T, X_{i,p}^T$ denote a starfish's obtained and current positions, respectively. $X_{best,p}^T$ represent best position, where p is the set of five dimensions randomly selected from the D dimensional space. r is a random number between 0 and 1. α_1 and θ are calculated with Eq. (2) and Eq. (3).

$$\alpha_1 = (2r - 1)\pi \quad (2)$$

$$\theta = \frac{\pi}{2} \cdot \frac{T}{T_{max}} \quad (3)$$

T , T_{max} denote current iteration and maximum iteration, respectively. When search space is less than 5 dimensions, the starfish uses one-dimensional search strategy to update position in exploration phase. The updated starfish location is expressed by Eq. (4).

$$Y_{i,q}^T = E_t X_{i,p}^T + A_1 (X_{k1,p}^T - X_{i,p}^T) + A_2 (X_{k2,p}^T - X_{i,p}^T) \quad (4)$$

$$E_t = \frac{T_{max} - T}{T_{max}} \cdot \cos \theta \quad (5)$$

$X_{k1,p}^T$, $X_{k2,p}^T$ are the p-dimensional positions of two randomly selected starfish from the population. A_1 and A_2 are two random numbers within the range $(-1,1)$. p is a randomly selected dimension from the D dimensions, and E_t represents the energy of the starfish, formulated in Eq. (5). In SFOA in exploitation phase has hunting and regeneration strategy to find global solutions. First, five randomly selected starfish from population, then calculate distance between the best location with each five randomly selected starfish as:

$$d_m = (X_{best}^T - X_{mp}^T) \quad m = 1, \dots, 5 \quad (6)$$

$$Y_i^T = X_i^T + r_1 d_{m1} + r_2 d_{m2} \quad (7)$$

In Eq. (6) is the obtained distance between the global best and the other five starfish, mp is the five randomly selected starfish. The update rule of each starfish in its hunting behavior is modelled in Eq. (7). r_1 and r_2 are random numbers between $(0,1)$. d_{m1} and d_{m2} are randomly selected individuals among d_m . In the SFOA, the regeneration phase focuses on the last individual and integrated with the exploitation step in each iteration. This phase is to avoid local solutions and improve the global convergence of the SFOA. Eq. (8) updates the rule of the regeneration stage and the position.

$$Y_i^T = \exp\left(-\frac{T * N}{T_{max}}\right) X_i^T \quad (8)$$

where T , T_{max} , and N represent the current iteration, the maximum iteration number, and the population size, respectively. Traditional SFOA use for continuous values, but in this paper, we want to use it for feature section, so we need binary starfish optimization algorithm (BSFOA). In BSFOA, unlike SFOA, candidate starfish for solution are denoted by "0" and "1". Traditional SFOA use for continuous values, but in this paper, we want to use it for feature section, so we need binary starfish optimization algorithm (BSFOA). In BSFOA, unlike SFOA, candidate starfish for solution are denoted by "0" and "1". When position of starfishes' are updated, they converted into probability values between 0-1 by passing a Sigmoid function. After that, their values obtained are compared with a random number and the new position of starfish

is updated as “0”, “1”. the position is calculated in BSFOA are given in Eq. (9) and Eq. (10) [46],[47].

$$Sig(Y[t+1]) = \frac{1}{1 + e^{-[Y(t+1)]}} \quad (9)$$

$$Y([t+1]) = \begin{cases} 0, & \text{if } rand() \geq Sig(Y[t+1]) \\ 1, & \text{if } rand() < Sig(Y[t+1]) \end{cases} \quad (10)$$

D. Decision Tree

Decision tree learning is an approach used to approximate target functions with discrete values. Where the resulting function is depicted as a decision tree. These trees can alternatively be expressed as a series of if-then rules, enhancing their interpretability for human users. Decision tree learning is a method for getting close to target functions that have discrete values. Where the resulting function is shown as a decision tree, it can also understand these trees better if they are written as a series of "if-then" rules. Because they are simple, easy to understand, and work well in many real-world situations, common algorithms are used in machine learning. When looking at loan applications, this algorithm has been used in many fields, including medical diagnostics and credit risk analysis. In decision trees, the classification process uses a tree structure that starts with the root node and moves through a series of internal nodes until it reaches the final node, which shows the classification result. Each node stands for a test that is linked to a data attribute, and the branches that come out of it show the possible values for that attribute. The ID3 algorithm is one of the most well-known algorithms for making decision trees. It is often used for classification tasks. The algorithm starts with the whole dataset and then splits it up piece by piece based on the attribute that gives the most information gain. Entropy is a measure of how uncertain a dataset is, and information gain is a number that comes from it. ID3 effectively reduces uncertainty subsets by choosing attributes that maximise information gain. The process continues until one of the stopping criteria is met, such as when all instances in a node belong to the same class or when there are no more attributes to evaluate. We use ID3 to sort the test samples [48].

RESULTS

Facial emotion detection is considered the main objective of this proposed system. In this system, we have used 2 datasets with names FER and FER+ as input data. The features of these image samples have been extracted using the AlexNet architecture. Subsequently, the BSFOA algorithm has been used to select the best features from the extracted features. Finally, the decision tree has been performed to classify the facial emotion images. All tests run by using MATLAB. To reduce the risk of overfitting, the training was concluded after 65 iterations by applying early stopping. The model achieved a training accuracy of 97.82% and a validation accuracy of 96.19% for FER dataset and training accuracy of 96.37% and a validation accuracy of 94.23% for FER+ dataset, indicating strong generalization capabilities. Evaluation on the test dataset further confirmed the model's robustness, yielding an accuracy of 96.03% and 94.54% for FER and FER+ dataset respectively. We used Confusion matrix to evaluate the

classification performance of our propose method, the confusion matrix is a widely accepted and reliable evaluation metric. The confusion matrix derived from the test set for FER and FER+ dataset is presented in Fig. 2 and Fig. 3, respectively. A detailed analysis of this matrix provides insight into how well the model distinguishes between different emotion categories.

As you can see in Fig. 2 for instance the classification result for “fearful” is 99.12%, “sad” is 97.62% .Moreover, in Fig.3 the classification result for “happiness” is 98.08% , “fear” is 96.67% .

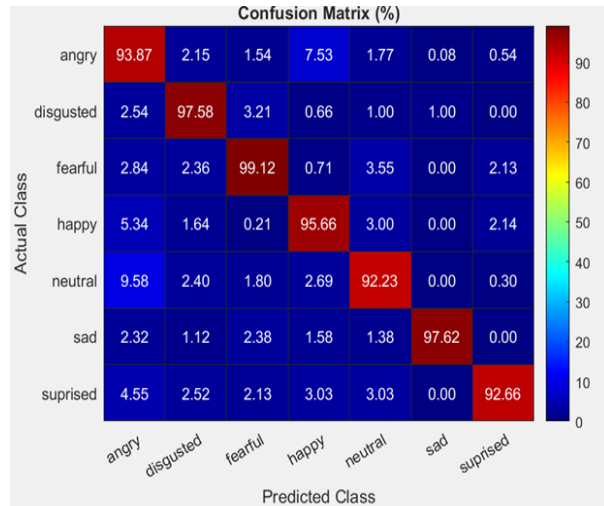


Fig. 2. Test dataset confusion matrix for FER dataset.

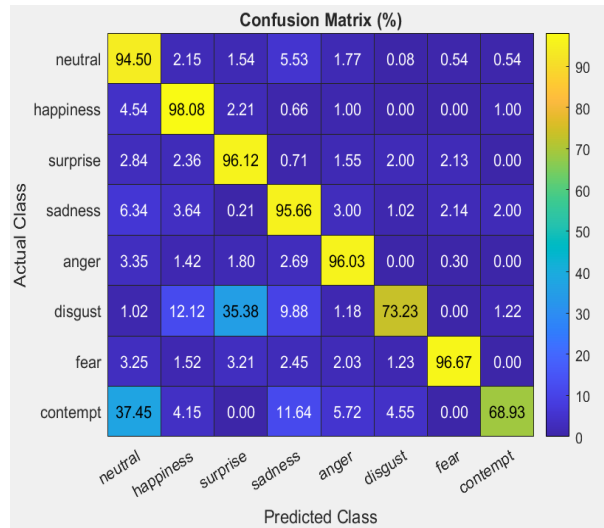


Fig. 3. Test dataset confusion matrix for FER+ dataset.

III. Discussions

The confusion matrices were analysed and show the model's capacity to distinguish between the different emotion categories. The values obtained in the confusion matrix are decreased in the FER+ dataset since the contempt and disgust classes have the least amount of data as compared to other classes. This means that the data imbalance directly affects the recognition rate for minor classes.

Table 3 presents a comparison of the accuracy of our proposed model with other approaches used to solve this problem previously. Our proposed BSFOA-DT method obtained the accuracy of 94.54% on the FER+ dataset. This result shows that our method has better performance than existing methods such as BPSO+SVM (85.74%), SENet Teacher (88.80%), and others in the comparison. The BSFOA algorithm incorporated successfully

reduced the dimensionality of the extracted features, which optimised the decision tree classification and also greatly improved the overall recognition accuracy.

CONCLUSIONS:

Table 3. Comparison of the proposed method accuracy with existing methods

Methods	Accuracy
BPSO+SVM [30]	85.74%
VGG16-PLD [36]	84.99 ± 0.37%
Pre-trained VGG-face [49]	81.73 %
SHCNN [50]	86.54%
ESR-9 [51]	87.15 ± 0.1%
SENet Teacher [52]	88.80 ± 0.3%
ResNet18 + FC [53] [53]	83.40%
proposed method on FER+	94.54 %

In this paper, a system is proposed to increase the accuracy and computational efficiency emotion expression classification. The proposed system consists of AlexNet, BSFOA and Decision tree. Using the developed AlexNet model, 128 features representing the images from the FER+ dataset were extracted. Classification was then performed using Decision Tree with these extracted features. The Decision Tree classification achieved superior accuracy compared to the AlexNet. To further enhance accuracy, the BSFOA optimization method was performed. A hybrid approach combining the Double Starfish Optimization Algorithm (BSFOA) and the Decision Tree algorithm was proposed to improve the performance of emotion recognition systems using the FER and FER+ datasets. In this method, BSFOA was used to select an optimized subset of features, reducing an initial set of 128 features extracted from images to just 60. These selected features were then used to train the Decision Tree classifier. The experimental results showed that the BSFOA+DT hybrid model outperformed the Decision Tree alone in both classification accuracy and computational efficiency. These results indicate that the proposed system can improve the performance of emotion recognition models, particularly when applied to the FER and FER+ datasets.

Acknowledgment

No acknowledgment.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets analyzed during the current study are publicly available in the FER and FER+ repositories.

Author Contribution

Zaid Hayder Akbar conceptualised and designed the study,

conducted the data collection and analysis, and wrote the manuscript.

Declarations

Ethical Approval

This study used publicly available datasets; therefore, formal ethical approval was not required.

Consent for Publication Participants.

Not applicable

Competing Interests

The author declare no competing interests.

REFERENCE

- [1] Mehrabian, Nonverbal communication. London, U.K.: Routledge, 2017.
- [2] M. Bartlett, G. Littlewort, E. Vural, K. Lee, M. Cetin, A. Ercil, and J. Movellan, "Data mining spontaneous facial behavior with automatic expression coding," in Verbal and nonverbal features of human-human and human-machine interaction, Berlin, Germany: Springer, 2008, pp. 1–20.
- [3] R. R. Asaad, "Review on deep learning and neural network implementation for emotions recognition," Qubahan Academic Journal, vol. 1, no. 1, pp. 1-4, 2021.
- [4] M. H. H. Khairi et al., "Detection and classification of conflict flows in SDN using machine learning algorithms," IEEE Access, vol. 9, pp. 76024-76037, 2021.
- [5] M. A. A. Albadr, M. Ayob, S. Tiun, F. T. Al-Dhief, and M. K. Hasan, "Gray wolf optimization-extreme learning machine approach for diabetic retinopathy detection," Frontiers in Public Health, vol. 10, p. 925901, 2022.
- [6] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. Al-Dhief, K. Omar, and F. A. Hamzah, "Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection," PLoS ONE, vol. 15, no. 12, p. e0242899, 2020.
- [7] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," Int. J. Eng. Technol., vol. 7, no. 4.36, pp. 160-166, 2018.
- [8] M. A. A. Albadr, S. Tiun, M. Ayob, and F. T. AL-Dhief, "Spoken language identification based on optimised genetic algorithm-extreme learning machine approach," Int. J. Speech Technol., vol. 22, pp. 711-727, 2019.
- [9] M. A. A. Albadr, S. Tiun, M. Ayob, M. Mohammed, and F. T. AL-Dhief, "Mel-frequency cepstral coefficient features based on standard deviation and principal component analysis for language identification systems," Cogn. Comput., vol. 13, pp. 1136-1153, 2021.
- [10] Lakhan, M. A. Mohammed, S. Kadry, K. H. Abdulkareem, F. T. Al-Dhief, and C. H. Hsu, "Federated learning enables intelligent reflecting surface in fog-cloud

- enabled cellular network," *PeerJ Comput. Sci.*, vol. 7, p. e758, 2021.
- [11] M. A. Mohammed et al., "An anti-spam detection model for emails of multi-natural," 2019.
- [12] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. AL-Dhief, K. Omar, and M. K. Maen, "Speech emotion recognition using optimized genetic algorithm-extreme learning machine," *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 23963-23989, 2022.
- [13] F. Abbas, U. U. Sheikh, F. T. Al-Dhief, and M. N. H. Mohd, "A comprehensive review of vehicle detection using computer vision," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 3, pp. 838-850, 2021.
- [14] T. AL-Dhief et al., "Voice pathology detection using machine learning technique," in *Proc. IEEE 5th Int. Symp. Telecommun. Technol. (ISTT)*, Nov. 2020, pp. 99-104.
- [15] T. Al-Dhief et al., "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms," *IEEE Access*, vol. 8, pp. 64514-64533, 2020.
- [16] M. A. Mohammed et al., "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.
- [17] N. Navaz, S. M. Adel, and S. S. Mathew, "Facial image pre-processing and emotion classification: A deep learning approach," in *Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2019, pp. 1-8.
- [18] K. M. Rajesh and M. Naveenkumar, "A robust method for face recognition and face emotion detection system using support vector machines," in *Proc. Int. Conf. Electr., Electron., Commun., Comput. Optim. Techniques (ICEECCOT)*, Dec. 2016, pp. 1-5.
- [19] T. Al-Dhief et al., "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77293-77306, 2021.
- [20] F. T. AL-Dhief et al., "Voice pathology detection using support vector machine based on different number of voice signals," in *Proc. 26th IEEE Asia-Pacific Conf. Commun. (APCC)*, Oct. 2021, pp. 1-6.
- [21] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies," *Psychol. Bull.*, vol. 115, no. 1, p. 102, 1994.
- [22] Y. Gizatdinova and V. Surakka, "Automatic detection of facial landmarks from AU-coded expressive facial images," in *Proc. 14th Int. Conf. Image Anal. Process. (ICIAP)*, 2007, pp. 419-424.
- [23] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors*, vol. 17, no. 4, p. 712, 2017.
- [24] R. Ekman, *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. New York, NY, USA: Oxford Univ. Press, 1997.
- [25] Sepas-Moghaddam, A. Etemad, F. Pereira, and P. L. Correia, "Facial emotion recognition using light field images with deep attention-based bidirectional LSTM," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3367-3371.
- [26] X. Fang, "Understanding deep learning via backtracking and deconvolution," *J. Big Data*, vol. 4, no. 1, Dec. 2017.
- [27] Sreenivas, V. Namdeo, and E. V. Kumar, "Modified deep belief network based human emotion recognition with multiscale features from video sequences," *Softw. Pract. Exper.*, vol. 51, no. 6, pp. 1259-1279, 2021.
- [28] M. Jangid, P. Paharia, and S. Srivastava, "Video-based facial expression recognition using a deep learning approach," in *Adv. Comput. Commun. Comput. Sci.: Proc. IC4S 2018, 2019*, pp. 653-660.
- [29] M. Abdul-Al, G. K. Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "A Novel Approach to Enhancing Multi-Modal Facial Recognition: Integrating Convolutional Neural Networks, Principal Component Analysis, and Sequential Neural Networks," 2024.
- [30] K. Donuk, A. Ari, M. F. Özdemir, and D. Hanbay, "Deep feature selection for facial emotion recognition based on BPSO and SVM," *Politeknik Dergisi*, vol. 26, no. 1, pp. 131-142, 2023.
- [31] D. Liang, H. Liang, Z. Yu, and Y. Zhang, "Deep convolutional BiLSTM fusion network for facial expression recognition," *Visual Computer*, vol. 36, no. 3, pp. 499-508, 2020.
- [32] M. R. M. Alsemawi, M. H. Mutar, E. H. Ahmed, H. O. Hanoosh, and A. H. Abbas, "Emotions recognition from human facial images based on fast learning network," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, pp. 1478-1487, 2023.
- [33] K. A. El Dahshan, E. K. Elsayed, A. Aboshoha, and E. A. Ebeid, "Recognition of Facial Emotions Relying on Deep Belief Networks and Quantum Particle Swarm Optimization," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, 2020.

- [34] X. Liu, X. Cheng, and K. Lee, "GA-SVM-based facial emotion recognition using facial geometric features," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11532-11542, 2020.
- [35] J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59-63, 2015.
- [36] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution," in *Proc. 18th ACM Int. Conf. Multimodal Interact. (ICMI)*, 2016, pp. 279-283.
- [37] D. S. Ma, J. Correll, and B. Wittenbrink, "The Chicago face database: A free stimulus set of faces and norming data," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1122-1135, 2015.
- [38] "Chicago Face Database," Accessed: May 19, 2021. [Online]. Available: <https://chicagofaces.org/default>
- [39] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese Female Facial Expression (JAFPE) Dataset," 1998.
- [40] "(JAFPE) Dataset | Zenodo," Accessed: May 19, 2021. [Online]. Available: <https://zenodo.org/record/3451524#.YKTd3bczZp8>
- [41] "Google Facial Expression Comparison Dataset – Google Research," Accessed: May 19, 2021. [Online]. Available: <https://research.google/tools/datasets/google-facial-expression/>
- [42] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2005, pp. 317–321.
- [43] Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 2106-2112.
- [44] M. Swapna, Y. K. Sharma, and B. M. G. Prasad, "CNN Architectures: Alex Net, Le Net, VGG, Google Net, Res Net," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 953-960, 2020.
- [45] Zhong, G. Li, Z. Meng, H. Li, A. R. Yildiz, and S. Mirjalili, "Starfish optimization algorithm (SFOA): a bio-inspired metaheuristic algorithm for global optimization compared with 100 optimizers," *Neural Comput. Applicat.*, vol. 37, no. 5, pp. 3641-3683, 2025.
- [46] K. Kumari, J. P. Singh, Y. K. Dwivedi, and N. P. Rana, "Multi-modal aggression identification using Convolutional Neural Network and Binary Particle Swarm Optimization," *Future Gener. Comput. Syst.*, vol. 118, pp. 187-197, 2021.
- [47] N. S. Dagar and P. K. Dahiya, "Edge Detection Technique using Binary Particle Swarm Optimization," *Procedia Comput. Sci.*, vol. 167, pp. 1421-1436, 2020.
- [48] B. De Ville, "Decision trees," *Wiley Interdiscip. Rev.: Comput. Stat.*, vol. 5, no. 6, pp. 448-455, 2013.
- [49] M. I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64827-64836, 2019.
- [50] S. Miao, H. Xu, Z. Han, and Y. Zhu, "Recognizing facial expressions using a shallow convolutional neural network," *IEEE Access*, vol. 7, pp. 78000-78011, 2019.
- [51] H. Siqueira, S. Magg, and S. Wermter, "Efficient facial feature learning with wide ensemble-based convolutional neural networks," in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, pp. 5800-5809.
- [52] S. Albanie, A. Nagrani, A. Vedaldi, and A. Zisserman, "Emotion recognition in speech using cross-modal transfer in the wild," in *Proc. ACM Multimedia Conf. (MM)*, 2018, pp. 292-301.
- [53] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," *IEEE Trans. Affect. Comput.*, 2018.

Conflict of Interest Statement: The authors declare that there search was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2025 Rambu Ayu Eryani, Rahmania Sri Untari. This is an open-access article distributed under the terms of the Creative

Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms