International Journal of Basic and Applied Sciences, 14 (SI-1) (2025) 606-614

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/nx916614
Research paper

Gradient Boosting Decision Tree Classification-Based Facial Emotion Detection Using Machine Learning

Ms. K. Rajeswari [1] *, S. Tharaneedharan [2]

[1] *Assistant Professor, Department of Computer Science Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India*
[2] *PG Student, Department of Computer Science Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India*
**Corresponding author E-mail: rajeswarik@mahendra.info*

Received: July 15, 2025, Accepted: July 24, 2025, Published: November 1, 2025

Abstract

Facial Emotion Detection is an automated task of computer vision used to determine the emotion of the human face based on facial expressions in still images or video. It is done by extracting the facial features and labeling them as happy, sad, angry, or surprised. It finds universal application in human-computer interaction, mental condition examination, and surveillance mechanisms. Achieving high levels of accuracy in detecting Facial Emotion in systems using different types of lighting and occlusions is quite a challenge that current systems fail to achieve. They also have the problem of generalizing due to different facial constructions, age groups, and cultural expressions. To solve those problemsContrast Limited Adaptive Histogram Equalization (CLAHE) is used to pre-process facial expressions in images or videos. This improves the local contrast and reveals subtle emotional guidance by boosting brightness in small areas and restricting noise expansion. Further, Elephant Herding optimization (EHO) is used to optimize the learning process, find all similar features of the face, and select the most essential features in the case of feedback facial recognition. The Gradient Boosting Decision Tree (GBDT), with its effective input in the non-linear relation and amelioration in the accuracy of the prediction, was used in carrying out the classification task. The performance indicators are used to evaluate the model, including accuracy89.3%, precision87.2%, recall86.5%, specificity88.1%, AUC-ROC of 0.91, and Matthews Correlation Coefficient (MCC) of 0.84. Experimental findings indicate that the designed approach reaches the peak tenth in the ratio of the support rate of dominant emotion categories and regularly exceeds the classical schemes in most severe and complicated facialrecognition contexts, which indicates high reliability and resilience of the regimen in various challenging tasks.

Keywords: *Contrast Limited Adaptive Histogram Equalization; Elephant Herding optimization; Gradient Boosting Decision Tree; Matthews Correlation Coefficient.*

1. Introduction

Facial emotion detection has come to the forefront of computer vision and the study of affects, whereby computers can interpret the physiological state of humans through facial expression analysis. The technology has critical roles worldwide in intelligent surveillance, mental health monitoring, adaptive learning systems, and human-computer interaction. Proper identification of facial expressions is generally difficult because of the dynamic nature of facial expressions as affected by the light conditions, pose discrimination, occlusions, ethnicity, age, and cultural discrepancies. Conventional machine learning methods, including Support Vector Machines, k-Nearest Neighbors, and Decision Trees, have demonstrated exemplary performance, but in general fail to effectively analyze the high-dimensional and complex facial information.

New developments in deep learning have ushered in a significant improvement in the efficiency and stability of facial emotion identification programs. Specifically, introducing CNNs has shown better results as they can automatically learn hierarchical features directly on raw facial images. To illustrate, an efficient emotion recognition system to efficiently recognize emotions based on thermal images using CNN by Gray Wolf Optimization to adjust the parameters, so it is robust and works under changing environmental factors. Under the same line of thought, hybrid models, combining deep learning and optimization methods, have appeared as a good answer to the problem of better model generalization and avoiding overfitting. The extreme subtlety and concision of emotional cues make the XGBoost model, which performs a combination of Extreme Gradient Boosting (XGBoost) and dual-stream shallow CNN, to improve micro-expression recognition, complex [1].

Other than the facial modalities, incorporating modalities like body postures and hand movements has helped establish a more pervasive and holistic definition of human emotions. Efficient emotion recognition using hand and body half landmarks extracted using the GEMEP dataset. Their results show that a combination of several modalities will expand the feature space and improve classification performance. When it comes to recognizing emotions, on the other hand, very similar neural signal analysis techniques have already been used; the nodal efficiency method applied to EEG functional networks to detect micro-expressions could be an indicator of how neurophysiological signals can be used with traditional vision-based systems. The second efficient technique is the ensemble learning paradigm, which encompasses the integration of several models to make more reliable and precise predictions. In this case, Gradient-Boosting Decision Trees (GBDT) are the kind of predictive modeling that has been proven to work in emotion classification, particularly because of their ability to handle non-linear relations and minimize model bias [2]. focused on deep learning to classify expressions based on faces and remarked that the use of multiple models enhances classification performance [12].

1.1. Objectives

- To improve the image quality and visibility of the features, the CLAHE algorithm was implemented to accentuate local contrast and depict meaningful emotional elements within the facial areas.
- To maximize the feature selection process and enhance the classification results of facial data using Elephant Herding Optimization, the system must discover the most appropriate facial features and minimize dimensionality/redundancy in the input data.
- To apply Gradient Boosting Decision Tree because it has superior capabilities to accommodate non-linear relationships and to iteratively refine the accuracy of prediction, especially when there is a need to apply corrective measures to increasingly complex patterns of relationships.
- To measure the proposed model using the most common set of performance measures, accuracy, precision, recall, specificity, AUC-ROC, and Matthews Correlation Coefficient, and, thus, determine the effectiveness and reliability of the latter.
- Contribution of work
- The study presents a preprocessing technique called Contrast Limited Adaptive Histogram Equalization that improves the visibility and contrast of facial features in low-light or occluded images, making it easier to detect subtle cues in emotional states.
- The Elephant Herding Optimization algorithm is employed to select the most relevant facial features, ensuring maximum efficiency of the models, reducing dimensionality, and performing optimal feature selection to enhance recognition.
- The primary classification algorithm to be used is the Gradient Boosting Decision Tree because it effectively models non-linear relationships and can enhance prediction accuracy by learning weak classification models iteratively.

1.2. Organization of the paper

The rest of the document is organized into key sections, as outlined below: Section II reviews recent research on Gradient-Boosting Decision Tree Classification-Based Facial Emotion Detection using Machine Learning, which various researchers have utilized. Section III details the proposed methodology, including steps such as preprocessing, feature selection, and classification techniques. Section IV discusses the analysis of experimental results and model performance metrics. Finally, Section V concludes the study with a summary of findings and key takeaways.

2. Literature Survey

In the study, subjects underwent gait trials and had to recover autobiographical memories to experience five target emotions: anger, sadness, joy, fear, and neutral [8]. Leave-one-participant-out cross-validation and Synthetic Minority Over-sampling Technique to overcome the issue of imbalanced classes were also used with Machine learning algorithms, K-Nearest Neighbors, Logistic Regression and Random Forest, Multi-layer Perceptron, and extreme Gradient Boosting (XGBoost) [3]. The results achieved are catastrophic error: Machine learning models demonstrated above-chance results when classifying emotional states (59% accuracy vs. 25% chance accuracy) [7].

The author created machine learning models, such as CNN-LSTM, Random Forest, and Gradient Boosting Decision Trees, and assessed how well they estimated emotional states [4]. The findings show that decision tree-based techniques, particularly Random Forest, successfully predict emotional states from environmental data [5].

The suggested system, Exploratory Data Analysis and Principal Component Analysis visualizations, was used to extract and optimize features relevant to the time, frequency, and wavelet domains. Facial expressions were distinguished using Gabor descriptors, Histogram of Oriented Gradients, Local Binary Patterns, and Local Ternary Patterns. Using K-Nearest Neighbors, Random Forest, Decision Tree, and XGBoost, the facial emotion recognition model achieved an accuracy of 84.6%, 74.3%, 67%, and 64.5%, respectively [6]. Performance indicators such as the Receiver Operating Characteristic Curve (ROC), F1 score, and Area under Curve were calculated to assess the models [13].

The author suggested that Convolutional neural networks, a type of deep neural network, are widely employed in FER because of their built-in image feature extraction mechanism. This paper presents a novel approach based on the Transfer Learning technique, called Efficient Net- XGBoost. Efficient Net-XGBoost is a cascade of the Efficient Net and XGBoost approaches with some experimentally driven improvements that represent the work's uniqueness.

In this study, we evaluate eight popular machine learning approaches using the FER 2013 dataset to identify the most effective approach for classifying human facial expressions [9]. According to the results, some algorithms have quite satisfactory accuracy: 37% for Logistic Regression, 33% for the K-neighbors classifier, 100% for the Decision Tree Classifier, 78% for Random Forests, 57% for Ada-Boost, 100% for Gaussian NB, 33% for LDA (linear discriminant Analysis), and 99% for QDA.

In this study, we classify EEG-based features to establish the approach for feature selection for a particular form of BCI that predicts whether a person correctly detects changes in facial expressions. The influence of each feature on the identification of expression changes was examined using various feature selection techniques, and the optimal combination was chosen by utilizing several machine learning classification algorithms [11]. The highest categorization accuracy result is 71%.

The suggested model maintains stringent evaluation methods, including essential metrics like the F1 score, validation accuracy, precision, and recall rate to gauge its robustness and dependability in the actual world. A possible and highly accurate method for effective stress detection is to combine facial emotion analysis with the Conv-XGBoost Algorithm [10].

Hyperparameter optimization was used to verify and evaluate the suggested techniques and obtain similar recognition performance. When it comes to accuracy, the statistical findings indicate that Extreme Gradient Boosting is the best classification model, offering the highest accuracy (78%) for facial image prediction when compared to Adaptive Boosting (77%), Random Forest (75%), and Light Gradient Boosting Model (58%) [14][15].

Table 1: Emotion Classification Using Multimodal Machine Learning Techniques

Ref. No	Authors Name	Modality Used	Technique Model	Applications	Dataset
[16]	Joshi & Kanoongo et al. (2022)	Multimodal (text, speech, facial)	ML + Emotional AI	Depression Detection	Various Emotion Datasets
[17]	Elbawab et al. (2023)	Emotional & Non-Emotional Cues	ML (RF, SVM)	Student Attentiveness	Self-collected educational dataset
[18]	Lei & Cao et al. (2023)	Audio-Visual	Preference Learning + Multi-Label	Emotion Recognition	Custom Annotated Dataset
[19]	Martinez-Martin et al., (2025)	Body + Hand Pose	ML (SVM, RF, KNN)	Emotion Recognition	GEMEP Dataset
[20]	Nandini et al. (2023)	EEG	3D VAD + ML (SVM, RF)	Emotion Detection	DEAP Dataset
[21]	Siddiqui et al. (2023)	IR-UWB Respiration	Temporal & Spectral + ML	Emotion Classification	Collected Respiration Data
[22]	Shilaskar et al. (2023)	EEG	BCI + ML	Emotion Recognition	Public EEG Datasets
[23]	Saher et al. (2024)	Facial Micro-Expressions	Deep Learning (CNN)	Schizophrenia Detection	Custom Dataset

Table 1 illustrates the comparison of the recent research studies that rely on a wide range of modalities and machine learning methods for defining emotions and analyzing mental health. These methods disrupt text, facial expression, EEG, body pose, and physiological measurements. Some of the techniques used are classical models, such as SVM and RF, and more advanced models, such as deep learning and preference learning. They include the detection of depression and schizophrenia, identification of emotions in education and criminal contexts. The datasets employed are publicly available and hand-crafted, demonstrating the utility and flexibility of each model in practice. Other models have been used to identify facial emotions in previous studies, and each has a limitation that is hoped to be overcome by our suggested model. It is also worth mentioning that CNN-LSTM-based methods are prone to overfitting because deep learning models are very complex in general and require extensive volumes of labeled data. With small or imbalanced data sets, overfitting is likely to result in poor generalization. This problem is alleviated in our framework, a combination of CLAHE to perform better image preprocessing, Elephant Herding Optimization (EHO) to perform best feature selection, and Gradient Boosting Decision Trees (GBDT) to classify images. With the addition of EHO, we can cut down on the dimensionality and eliminate irrelevant information, thus making the model more general and avoiding overfitting, even when using smaller datasets.

Moreover, the application of K-Nearest Neighbors (KNN) to emotion recognition, as seen in prior research, can produce poor recall, in particular in the differentiation between united displays of emotion such as sadness and surprise. KNN is noise-sensitive and has no capacity to capture the nonlinear relationship in data. Our framework proposal can defeat this shortcoming by employing GBDT, which is an ensemble tool that can address non-linearities more clearly. GBDT makes its predictions sequentially and progressively improves its recall rate by targeting the more difficult to classify examples, and by addressing errors committed by earlier trees.

Unlike these earlier models, our CLAHE-EHO-GBDT model can give better recall and generalization, as indicated by the higher Matthews Correlation Coefficient (MCC) of 0.84 (Table 2). This improvement is especially valuable in practice, when recognition of subtle expressions of emotions is important, even in the presence of problematic conditions, such as occlusions or lighting changes.

3. Proposed Methodology

The proposed facial emotion detection system begins by capturing facial images or video frames as input. These inputs may include variations in lighting, face orientation, or obstructions. To address these issues, the images are first processed with a Contrast Limited Adaptive Histogram Equalization technique, which enhances local contrast and highlights subtle emotional details by emphasizing key facial features such as the eyes, mouth, and eyebrows, without excessively amplifying noise. After this enhancement, facial features are extracted using standard image processing methods to capture texture-based, as well as geometric features based on texture, which are crucial for emotion recognition. The next step involves applying a nature-inspired meta-heuristic algorithm called Elephant Herding Optimization (EHO) to the extracted features to identify the most relevant ones. EHO simulates the social herd behavior of elephants to select the most pertinent features and discard redundant or non-informative ones, thereby reducing computational complexity and improving classification accuracy. These optimized features are achieved for the Gradient Boosting Decision Tree (GDBT) classifier. As an ensemble learning method, GDBT constructs multiple decision trees, where errors from previous trees are used to correct subsequent ones, resulting in accurate and reliable predictions.

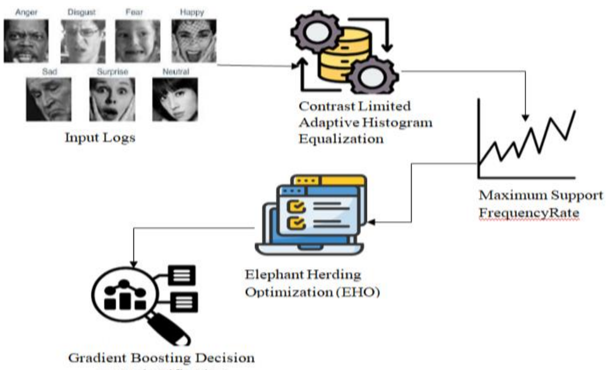

Fig. 2: Proposed Methodology Architecture.

The proposed methodology in facial emotion detection, shown in Figure 2, begins with the input facial images that have labels, such as sad, surprise, and neutral, embedded. After the pre-processing stage, the Contrast Limited Adaptive Histogram Equalization will be used to perfect local contrast further and emphasize the emotional characteristics. The system then analyzes the Maximum Support Frequency

Rate to determine the facial features that are always significant. The following technique, Elephant Herding Optimization, is used to choose the most pertinent features to eliminate dimensionality and enhance the efficiency of the learning process. These optimized characteristics are categorized by the Gradient Boosting Decision Tree, which effectively controls non-linear data and increases prediction precision.
In order to achieve some consistency between the description in Section 4.1 and Figure 2, we deal with the difference in the number of emotions. In the previous section 4.1, we have cited seven emotions: anger, disgust, fear, happiness, sadness, surprise, and not. But Figure 2 mistakenly names eight emotions, with a further category of contempt. We have replaced Figure 2 with one that is consistent with the seven emotions in the FER2013 dataset, which will fix the discrepancy between the description and the diagram.

3.1. Contrast-limited adaptive histogram equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization is an essential preprocessing step that greatly enhances image quality before feature extraction and classification. CLAHE improves the local contrast of facial images and aids in detecting subtle emotional cues, such as fine wrinkles, eyebrow positioning, and mouth shape. It also reduces noise amplification compared to traditional histogram equalization by applying contrast enhancement to small image regions and using a clipping threshold to prevent over-saturation. This localized brightening approach ensures consistent facial expressions, providing a uniform appearance even under different lighting conditions or low-contrast backgrounds.

$$H_{k,l}(g) = \sum_{x=k-r}^{k+r} \sum_{y=l-r}^{l+r} [I_{x,y} = g] \tag{1}$$

This histogram count of grey level g in some localized region surrounding the pixel at positionk, l is given equation, $H_{k,l}(g)$This is a square area delimited by a window with radius r, that is it expands to the length of $k + r$ horizontally and $l + r$ vertically. In this intensities of nearby pixels are investigated, and many pixels are investigated, and many pixels are precisely at the gray level g are counted.

$$C_{k,l}(g) = \sum_{k=0}^{g} H_{k,l}(k) \tag{2}$$

The equation (2) represents the cumulative frequency of gray levels from 0 to g in a localized region centered at pixel position(k, l). Here,$C_{k,l}$ is used to accumulate the histogram counts computed from the local neighborhood of the image. This accumulation helps to build a Cumulative Distribution Function for contrast enhancement. The term $H_{k,l}$ denotes the number of pixels in the local region that have a specific gray level k. By summing these values from 0 to g, the function provides the total count of pixels with intensities up to level g.

$$p_x(i) = \frac{m_i}{m}, 0 \leq i \leq L \tag{3}$$

In equation (3), represents $p_x(i)$is the relative frequency of gray level i, which is calculated by dividing m_i the number of pixels in the region with intensity. The value of i denotes the specific intensity level under consideration, ranging from 0 to L, where L is the maximum gray level value in the image, typically 255 in an 8-bit grayscale image.

$$CDF(i)_X = \sum_{j=0}^{i} p_x(j) \tag{4}$$

In this equation, $p_x(j)$ is the probability (or relative frequency) of the intensity level j, and the summation from $j = 0$ to i accumulate these probabilities. This means that $CDF(i)_X$gives the total probability of all gray levels less than or equal to i, effectively measuring how intensities are distributed up to that point. In CLAHE, this cumulative information is crucial for redistributing pixel intensities in a way that enhances local contrast.

3.2. Maximum support frequency rate

The maximum frequency of support words refers to the highest frequency at which a specific expression of emotions or keywords is used in a set of records. In the context of facial emotional identification, repetitive expressions strongly influence the overall connection to certain emotional states or processes. By studying how frequently certain features or patterns occur, researchers can determine the significance and reliability of these features when identifying a particular emotion. This frequency-based approach has the potential to identify emotionally significant episodes and enable the processing of the most important, irrelevant, or redundant emotional signals. Furthermore, it is also possible to learn more about these behavioral patterns once the maximum levels of support and confidence for these repeated emotional patterns are analyzed.

$$\text{Support }(E) = \frac{f(E)}{N} \tag{5}$$

The variable E signifies a certain emotion or facial expression, including happy, sad, angry, surprised, and the function f(E) is the can either mean the frequency, the number of repetitions of this specific emotion, E, will not be comparable in the dataset. The variable N is the total records count the sample in the dataset, $\frac{f(E)}{N}$ is expressed in terms of the size of the total data. This measure plays a critical role in determining the emotions communicated most of the time and thus is of greater statistical significance.

$$E_{max} = \arg\max \left(\frac{f(E_i)}{N}\right) \tag{6}$$

A facial emotion recognition dataset is a collection of data entries that can be used in facial emotion recognition. E_i the symbolizes every single emotion that one has, like happy, sad, surprised, these are the categories of emotions with which the system tries to recognize and distinguish. E_{max} implies the main expression of emotion, is recurrent, and perhaps points to

$$(A \rightarrow E) = \frac{f(A \cap E)}{f(A)} \tag{7}$$

In equation (7), represents a certain facial feature status, as raised eyebrows, smiling lips, or frowning forehead; these characteristics work as visual cues that can be linked with specific moods. E is the aligned associated with those facial features, $f(A \cap E)$ Is the expression of the frequency of the affectation of the facial feature?

3.3. Elephant herding optimization (EHO)

Elephant Herding Optimization, which implements the ideas of social behavior of elephant clans, is applied to optimize feature selection and hyperparameters and ensure that only the most significant facial features are employed. This simplifies the system and improves the classifier. Then GBDT, a potent ensemble learning method, classifies emotions sequentially, correcting the mistakes of the preceding trees.

$$X_i(t+1) = X_i(t) + \Delta X_i$$

Here, $X_i(t)$ represents the current position of the elephant at iteration t, which corresponds to a candidate solution in the optimization problem, such as a particular combination of features or hyperparameters. The term ΔX_idenotes the movement vector, which defines the direction and magnitude of change for that solution, based on influences from the best solution found so far and the average behavior of the elephant's clan. The result, $X_i(t+1)$is the updated position of the elephant at the next iteration, incorporating the learned adjustments.

$$\Delta X_i = \alpha. X_{best} + \beta. X_{center}$$

In this equation, ΔX_i represents the movement vector for the elephant, indicating how its position should be adjusted during the optimization process. The term $\alpha$is a control parameter that determines how strongly the elephant is influenced by the global best solution found so far, denoted asX_{best}. This encourages the elephant to move toward the most promising solution in the entire population. On the other hand, β is another weighting factor that controls the influence of the clan center, represented by X_{center}which is the average position of all elephants in the same group or clan. This term promotes social learning and local exploration within the clan.

$$x_{new}, c_i, j = \beta \times X_{center}, c_i$$

This update is based on the clan's central tendency rather than the individual's previous position. The term β is a scaling or influence factor that controls how strongly the elephant is attracted toward the clan's central position. X_{center}, c_iRefers to the center position of the clan c_i, which is calculated as the average of all elephants' positions within that clan

3.4. Gradient boosting decision tree classification

Gradient Boosting Decision Tree (GBDT) classification is an individually simple but highly effective ensemble learning method used to construct an accurate predictive model by assembling many sequentially learned weak learners. After learning from observed facial data, GBDT is applied to classify happiness, sadness, anger, or surprise in facial emotion detection. Every single decision tree in the sequence is trained to fix the mistakes of the previous ones, until the overall accuracy of predictions is increased steadily. Instead of one significant sample per iteration, the model looks at the harder ones, making it quite capable of catching the subtle differences in the facial expression. This approach is suitable especially for high-dimensional and high-level facial data since it provides better results in accuracy, robustness, and generalization.

$$\{(x_1, y_1) \ldots . (x_n, y_n)\}$$

The variable x_1 refers to the feature vector, the sample represented by the facial details, key landmark points, pixel intensity values, or texture-based descriptors after the extraction holds the visual information of a facial image, the variable y_1 refers to the target label, in the emotional category, and the symbol n is the total number of training datasets.

$$f_k(x) := \sum_{k-1}^{k} T(x; \theta_k)$$

In this regard, $f_k(x)$ the aggregate output of the model is the addition of the prediction, and this is denoted by (x). k is the sequence of decision trees, k+1, and every single decision tree is referred to as $T(x; \theta_k)$ is the collection of parameters, the decision node split thresholds, and the corresponding prediction to be made in the area of a tree.

$$\theta_k = \text{argmin} \sum_{i=1}^{n} L\,(y_i, f_{k-1}(x) + T(x; \theta_k))$$

The loss function $L\,(y_i, f_{k-1})$ has an important role to play in the process of learning. It is the difference between the actual emotion labels, y_i is the actual class and the output that should come out, y_i is produced by the model. Asked about their choice of route, k is the prior prediction is used as the basis of the cumulative prediction on the previous iteration, whereupon the present prediction is constructed. θ_k The aim is to optimize 0 of this new tree with the aim of minimizing the total loss.

4. Result and Discussion

The Facial Emotion Detection model results have shown considerable advances in all the significant performance scores in the experimental results. The model was found to be accurate in proving its overall accuracy in classifying the emotions. This level of precision and recall scores reveals that it can correctly identify the emotional expressions while having low false positives and false negatives, respectively. The specificity score shows how well it categorizes non-target emotions and keeps a low rate of identifying neutral or irrelevant expressions. An AUC-ROC value of a higher value denotes a good discriminating ability of various emotional classes, even in cases of overlapping or ambiguity. In addition, the Matthews Correlation Coefficient also confirms the soundness of the model since it considers both the true and false positives and negatives proportionately.

4.1. Dataset description

The data set contains images of people showing seven emotions (anger, disgust, fear, happiness, sadness, surprise). Each image represents a specific emotion, so researchers and machine learning practitioners can research and develop emotion recognition and analysis models.

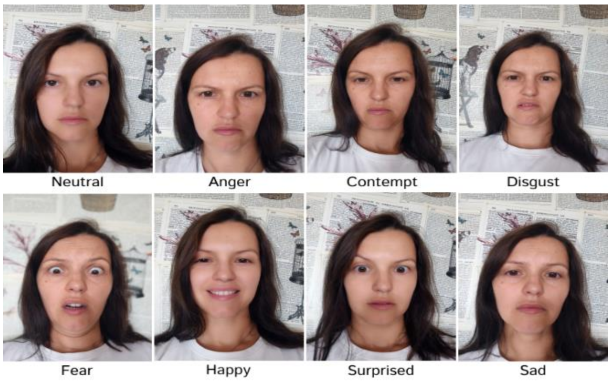

Fig. 2: Emotion Recognition Dataset.

Figure 2 presents an excerpt of the Emotion Recognition Dataset that presents a person's facial expression sequence under restricted circumstances. The data covers eight core emotions: neutral, anger, contempt, disgust, fear, happiness, surprise, and sad. In each picture, different facial muscle reactions to each emotion are emphasized, like the opening of eyes and elevated eyebrows in case of fear, smiling with happiness, and turned down lips in case of sadness. These differences in facial expression create the basic characteristics that the emotion recognition algorithms should recognize and classify.
This study uses the FER2013 dataset, which is a publicly available dataset that is typically used in facial emotion recognition challenges. There are a total of 35,887 facial images in this dataset, and each of them has been labeled with one of the seven following emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. The photos are gathered in a wide range of sources, such as the internet, and represent a wide range of people of different ages, genders, and races. These photos are in various lighting settings and portray diverse facial expressions, and the dataset can be utilized in research on emotion recognition.

4.2. Comparison results

Table 2: Performance Comparison Table

Models	Accuracy	Precision	Recall	Specificity	AUC-ROC	Matthews Correlation Coefficient (MCC)
CNN-LSTM	85.6%	83.4%	82.1%	86.0%	0.88	0.79
XG-Boost	86.9%	84.7%	83.6%	87.4%	0.89	0.81
KNN	81.3%	79.0%	77.5%	82.5%	0.84	0.74
Proposed GBDT	89.3%	87.2%	86.5%	88.1%	0.91	0.84

In Table 2, a comparison of four machine learning models applied to facial emotion recognition is provided. The other models perform worse than the proposed GBDT model across the seven metrics. The proposed GBDT model demonstrates an accuracy of 89.3% and an AUC-ROC of 0.91, indicating a high classification capacity and robustness. The CNN-LSTM and XG-Boost models are not far behind; however, they slightly underperform in recall and MCC, suggesting balanced prediction and generalization. The KNN model exhibits relatively poor results, particularly in terms of recall (77.5%), reflecting low sensitivity to the characteristics of emotions.

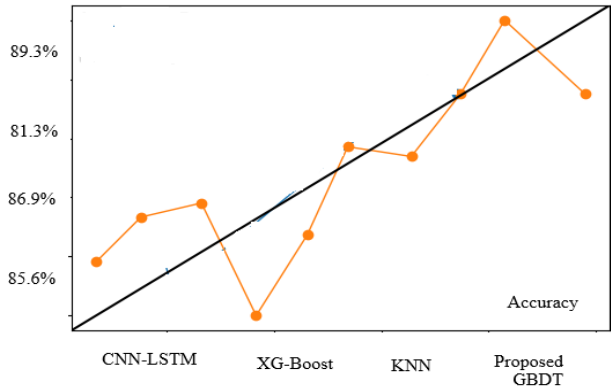

Fig. 3: Illustrates Accuracy Performance.

Figure 3 represents the accuracy of performance of different machine learning models, such as CNN-LSTM, XG-Boost, KNN, and the proposed GBDT model. The proposed GBDT model showed the most accurate result, reaching 89.3%, giving greater performance than other techniques. The XG-Boost model 86.9% came next with a precision almost like that of CNN-LSTM, 85.6 %. Conversely, KNN has the lowest accuracy of 81.3%, which shows that it was not capable of classifying very well.

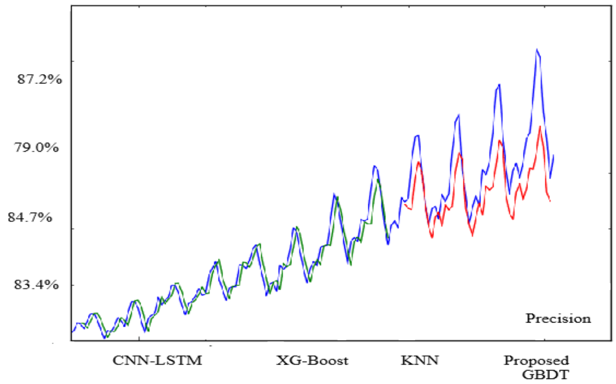

Fig. 4: Illustrates Precision Performance.

Figure 4 illustrates the accuracy of various machine learning models, including CNN-LSTM, XG-Boost, KNN, and the proposed GBDT model. The GBDT model is ranked highest with an 87.2% precision level, demonstrating its ability to predict and capture relevant positive cases with high accuracy. XG-Boost follows with a precision of 84.7%, while CNN-LSTM is close behind at 83.4%. Both show relatively strong performance. In contrast, KNN demonstrates the lowest precision at only 79.0%, leading to more false positives.

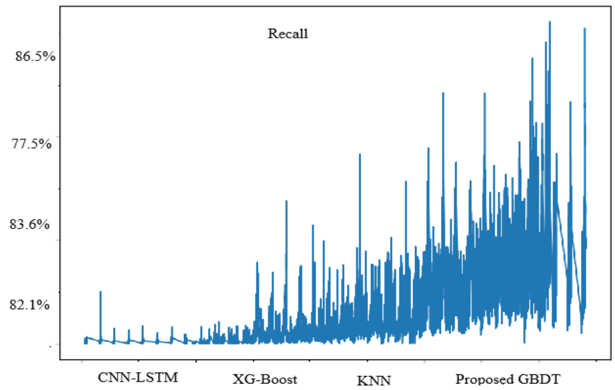

Fig. 5: Illustrates Recall Performance.

Figure 5 indicates the recall performance of four models: CNN-LSTM, XG-Boost, KNN, and the proposed GBDT. In the case of the Proposed GBDT model, the tested recall of the model is the highest of all at 86.5%, which indicates a great ability of the model to find most of the actual positive cases. XG-Boost and CNN -LSTM have decent effectiveness in terms of 83.6% and 82.1 % recall, respectively. Nevertheless, KNN ranks lowest in recall of 77.5%, which means it fails to identify most positives.

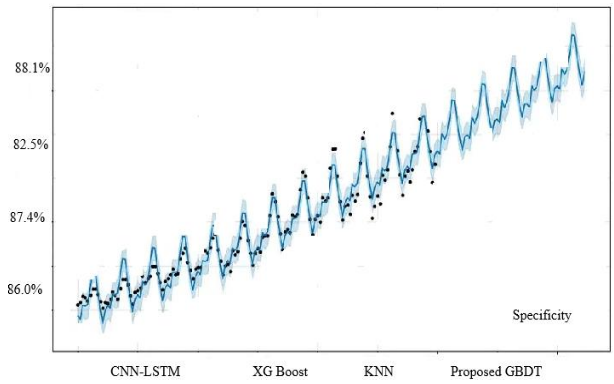

Fig. 6: Illustrates Specificity Performance.

Figure 6 presents the specificity performance of the CNN-LSTM, XG-Boost, KNN, and proposed GBDT models. The Proposed GBDT model is the best in terms of specificity, with 88.1%, which means it could perform very well in detecting negative cases and reducing false positive results. XG-Boost kicked it off with a specificity of 87.4 %, slightly above CNN-LSTM at 86.0%, both signifying a good performance. Conversely, KNN observes the lowest specificity value of 82.5 %, implying a high chance of misclassifying negative cases.

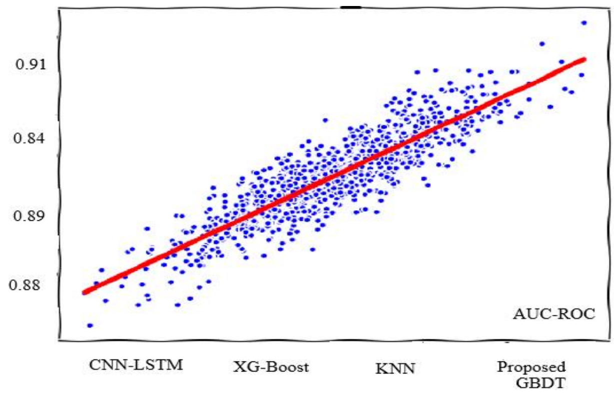

Fig. 7: Illustrates AUC-ROC Performance.

Figure 7 illustrates the AUC-ROC (Area under the Receiver Operating Characteristic Curve) performance of CNN-LSTM, XG-Boost, KNN, and the proposed GBDT models. The proposed GBDT diagnostic tool achieves the highest AUC-ROC value of 0.91, indicating its superior ability to classify positive and negative classes. XG-Boost follows with a score of 0.89, while CNN-LSTM ranks third with a score of 0.88, both reflecting good classification performance. In contrast, KNN has the lowest AUC-ROC value at 0.84, demonstrating relatively weak discrimination power.

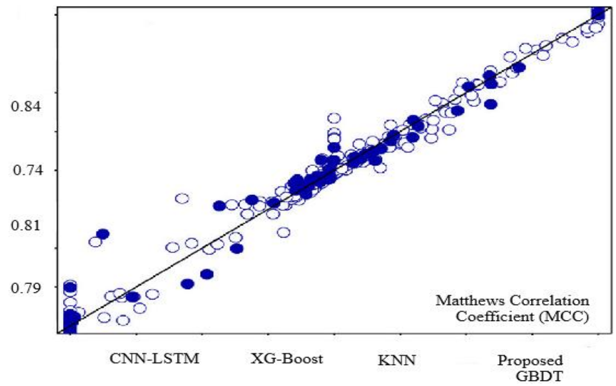

Fig. 8: Illustrates MCC Performance.

The performance of the Matthews Correlation Coefficient (MCC) on four models, including CNN-LSTM, XG-Boost, KNN, and the GBDT under the proposal, is shown in Figure 8. The Proposed GBDT model attains the best MCC of 0.8,4, corresponding to the high overall predictive quality regarding both true and false positives and negatives. XG-Boost follows it with an MCC of 0.81 and CNN-LSTM with a score of 0.79, in terms of performance. KNN, on the other hand, has the lowest score of 0.74, which shows that KNN is a relatively less effective algorithm in a balanced classification task.

4.3. Limitations

Although the proposed CLAHE-EHO-GBDT framework has good points, some limitations should be considered in the research in the future. A possible weakness is the computational cost of Elephant Herding Optimization (EHO). Although EHO is particularly effective at assisting feature selection by decreasing dimensionality and increasing model accuracy, its complexity can become a concern when implementing the framework in real-time systems or resource-constrained systems, e.g., mobile devices or embedded systems. The future work may be dedicated to the optimization of the EHO algorithm to minimize its computational cost, e.g., by using parallel processing methods or by approximating the process of its optimization to increase its scalability.

The other difficulty is when there are extreme occlusions. Although CLAHE is useful in improving faces in images with different lighting conditions, the system might fail to read emotions when faces are significantly blocked in images (e.g., using glasses, masks, or hands over the face). Even though EHO aids in the selection of the most relevant features, occlusions continue to remain a big challenge in practice. To address this drawback, future research might examine the incorporation of deep learning models specifically trained to be occlusion resistant, or might examine multimodal information (i.e., audio or body posture data in addition to facial expressions).

5. Conclusion

In conclusion, the suggested facial emotion detection framework is an effective replica of a strong solution, which combines different advanced technologies to address the typical issues pertinent to emotion perception. Contrast Limited Adaptive Histogram Equalization Technique considerably hikes the quality of input photos, locally enhancing the contrast, and thus helps to capture barely comprehensible emotional messages in murky conditions, minimal visibility, or interfered sight. The Elephant Herding Optimization algorithm is an efficient approach to perform feature selection where only the most relevant features of the face are used in the classification, and thus, learn the optimization process effectively, only by using the appropriate features. Gradient Boosting Decision Tree is a good classification model that works well with a non-linear relationship between emotions and facial features that are often more complicated. The system provides high scores, such as an accuracy of 89.3%, precision of 87.2%, recall of 86.5%, and AUC-ROC of 0.91, which supports that the system can easily classify emotions in unfavorable conditions. These results depict the applicability and usefulness of the recommended model, and in this way, it becomes a strong and dependable substitute for real-life facial emotion finding.

Future Directions

To develop the power of the CLAHE-EHO-GBDT framework, there are several major research questions and applications that should be explored. To begin with, the framework might be used to increase the practical applicability of video analysis in real-time. Dynamic environments, e.g., live surveillance, interactive human-computer interfaces, or emotional analysis in healthcare environments, require real-time emotion recognition. Future studies might investigate methods to optimize the framework to deal with such a computational burden as continuous video streams, including temporal correlations in facial expressions. Cross-cultural datasets are also another potential direction. Although the existing framework has been performing very well on the available facial emotion recognition databases, there is a possibility that culture affects the model in question by altering the facial expression across different cultures. Working with cross-cultural data sets would assist in testing the framework to determine its strength in relation to the various people and to improve its accuracy in multicultural settings. The research questions might be specific to determine cultural differences in emotional expressions and investigate how such differences might be modeled during the training process to enhance cross-cultural generalization.

References

[1] Neelakandan, S., & Paulraj, D. (2020). A gradient boosted decision tree-based sentiment classification of twitter data. *International Journal of Wavelets, Multiresolution and Information Processing*, *18*(04), 2050027.

[2] Sun, R., Wang, G., Zhang, W., Hsu, L. T., & Ochieng, W. Y. (2020). A gradient boosting decision tree based GPS signal reception classification algorithm. *Applied Soft Computing*, *86*, 105942.

[3] Priyadarshini, R. K., Banu, A. B., & Nagamani, T. (2019, April). Gradient boosted decision tree based classification for recognizing human behavior. In *2019 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 1-4). IEEE.

[4] Rodrigues Filho, L. F., & Martins, A. S. (2023). EMOTION DETECTION USING MACHINE LEARNING: ASSESSMENT OF EFFECTIVENESS AND APPLICABILITY IN DIFFERENT CONTEXTS. *Humanidades & Inovação*, *10*(21), 273-284.

[5] Arsirii, O. O., & Petrosiuk, D. V. (2023). Pseudo-labeling of transfer learning convolutional neural network data for human facial emotion recognition. *Вісник сучасних інформаційних технологій*, *6*(3), 203-214.

[6] Jaiswal, S., & Nandi, G. C. (2020). Hyperparameters optimization for Deep Learning based emotion prediction for Human Robot Interaction. *arXiv preprint arXiv:2001.03855*.

[7] Goyal, K. Multimodal Attention CNN for Human Emotion Recognition.

[8] Chakraborty, A., & Nagar, A. K. Emotion Recognition from Facial Action Points by Principal Component Analysis.

[9] Karali, A., Bassiouny, A., & El-Saban, M. (2015, September). Facial expression recognition in the wild using rich deep features. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 3442-3446). IEEE.

[10] Uyulan, Ç., Gümüş, A. E., & Güleken, Z. (2022). EEG-induced Fear-type Emotion Classification Through Wavelet Packet Decomposition, Wavelet Entropy, and SVM. *Hittite Journal of Science and Engineering*, *9*(4), 241-251.

[11] Marques, L. S., Gresse von Wangenheim, C., & Hauck, J. C. (2020). Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, *19*(2), 283-321.

[12] Ahn, H., Kim, S., & Kim, J. K. (2014). GA-optimized support vector regression for an improved emotional state estimation model. *KSII Transactions on Internet & Information Systems*, *8*(6).

[13] YIN, H., LIU, P., & ZHANG, L. (2025). Preliminary development and evaluation of the Chinese self-conscious emotions nonverbal behavior expression stimulus set, and its application in research. *Acta Psychologica Sinica*, *57*(10), 1745.

[14] Zhang, Z., & Jung, C. (2020). GBDT-MO: Gradient-boosted decision trees for multiple outputs. *IEEE transactions on neural networks and learning systems*, *32*(7), 3156-3167.

[15] Si, S., Zhang, H., Keerthi, S. S., Mahajan, D., Dhillon, I. S., & Hsieh, C. J. (2017, July). Gradient boosted decision trees for high dimensional sparse output. In *International conference on machine learning* (pp. 3182-3190). PMLR.

[16] Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings*, *58*, 217-226.

[17] Elbawab, M., & Henriques, R. (2023). Machine Learning applied to student attentiveness detection: Using emotional and non-emotional measures. *Education and information technologies*, *28*(12), 15717-15737.

[18] Lei, Y., & Cao, H. (2023). Audio-visual emotion recognition with preference learning based on intended and multi-modal perceived labels. *IEEE Transactions on Affective Computing*, *14*(4), 2954-2969.

[19] Martinez-Martin, E., & Fernández-Caballero, A. (2025). Improved human emotion recognition from body and hand pose landmarks on the GEMEP dataset using machine learning. *Expert Systems with Applications*, *269*, 126427.

[20] Nandini, D., Yadav, J., Rani, A., & Singh, V. (2024). Enhancing emotion detection with non-invasive multi-channel EEG and hybrid deep learning architecture. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, *48*(3), 1229-1248.

[21] Siddiqui, H. U. R., Zafar, K., Saleem, A. A., Raza, M. A., Dudley, S., Rustam, F., & Ashraf, I. (2023). Emotion classification using temporal and spectral features from IR-UWB-based respiration data. *Multimedia Tools and Applications*, *82*(12), 18565-18583.

[22] Shilaskar, S., Bobby, D., Dusane, A., & Bhatlawande, S. (2023, June). Fusion of EEG, EMG, and ECG signals for accurate recognition of pain, happiness, and disgust. In *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)* (pp. 142-149). IEEE.

[23] Saher, A., Gilanie, G., Cheema, S., Latif, A., Batool, S. N., & Ullah, H. (2024). A Deep Learning-Based Automated Approach of Schizophrenia Detection from Facial Micro-Expressions. *Intelligent Automation & Soft Computing*, *39*(6).