Advances in Technology Innovation, vol. 10, no. 4, 2025, pp. 395-406

Comparative Analysis of Facial Expression Recognition Using Image-Based and Landmark-Based Methods

Thanawat Srikaewsiew, Sarunya Kanjanawattana*

School of Computer Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand

Received 05 December 2024; received in revised form 22 March 2025; accepted 31 March 2025

DOI: https://doi.org/10.46604/aiti.2025.14589

Abstract

This study compares the effectiveness of image-based and landmark-based methods for facial expression recognition (FER) in classifying hurt and normal facial expressions, utilizing datasets from the Delaware Pain Database and UTKFace. Five machine learning models are assessed, including convolutional neural networks (CNN), support vector machines (SVM), random forest classifier (RFC), logistic regression classifier (LRC), and gradient boosting classifier (GBC). The findings indicate that CNN achieves the highest accuracy at 95% when using landmark-based features, while SVM and GBC also perform admirably with these features. Conversely, LRC exhibits inconsistent results, especially when relying on image-based features. These findings offer valuable insights into the strengths and weaknesses of each approach, guiding the selection of effective FER techniques.

Keywords: face expression recognition, machine model comparison, image-based classification, landmark-based classification, expression dataset

1. Introduction

Facial expression recognition (FER) is a field that combines computer vision, artificial intelligence, and psychology to interpret human emotions from facial cues. Although humans naturally recognize expressions, training machines to do the same continues to be a challenge. With the rise of human-computer interaction, emotionally intelligent technologies are becoming increasingly important.

FER [1] is an interdisciplinary field that bridges computer vision [2], artificial intelligence [3], and psychology [4], aiming to decode the intricate language of human emotions conveyed through facial expressions. Facial expressions, a fundamental mode of nonverbal communication [5], help humans express emotions, intentions, and perceptions. While humans can naturally interpret these expressions, enabling machines to interpret them remains a significant challenge [6]. With the growing importance of human-computer interaction, the need for emotionally intelligent technologies capable of recognizing and responding to human emotions is increasingly apparent.

Image-based methods leverage the transformative power of deep learning [7], with convolutional neural networks (CNN) at the forefront. CNN models, trained on large-scale datasets like the Delaware Pain Database [8] and UTKFace, excel at capturing global and local features from facial images, enabling robust emotion recognition. In contrast, landmark-based learning [9] focuses on extracting key facial points, such as the eyes, nose, and mouth, as representative features of facial expressions. These methods utilize geometric relationships between landmarks, offering robust models for emotion classification under varying poses and lighting conditions.

* Corresponding author. E-mail address: Sarunya.k@sut.ac.th

While both image-based [10] and landmark-based methods [11] for FER have been extensively explored, there has been limited research directly comparing their effectiveness across diverse facial expression datasets. Most studies tend to focus on one approach, without offering a clear comparison that highlights the relative advantages and drawbacks of each. This study addresses that gap by conducting a direct comparison, providing valuable insights into which method performs better under various conditions.

The scarcity of such comparisons can be attributed to the complexity and variation in how facial expressions are captured and analyzed. Image-based methods typically process raw data and rely on deep learning for feature extraction. In contrast, landmark-based methods extract key facial points, offering a more computationally efficient approach. By combining both methodologies in one study, this research aims to uncover their complementary strengths and limitations.

Despite progress in both areas, significant gaps remain in the literature. Most research focuses exclusively on either image-based or landmark-based techniques, without comparing their effectiveness across multiple machine learning (ML) models. Furthermore, few studies explore the practical application of these methods in real-world contexts, such as rehabilitation and diagnostic systems. These gaps highlight the need for a thorough evaluation of these approaches to better understand their strengths and limitations in practical applications.

FER is a crucial task in affective computing and human-computer interaction, with applications in healthcare and emotional AI. While existing FER techniques leverage both image-based and landmark-based approaches, there remains a gap in systematically comparing these methods under standardized conditions using diverse ML models. This study aims to bridge this gap by conducting a comprehensive comparison of image-based and landmark-based FER techniques using five distinct ML models: CNN [12], logistic regression classifier (LRC) [13], support vector machines (SVM) [14], random forest classifier (RFC) [15], and gradient boosting classifier (GBC) [16]. A key aspect of this research is the evaluation of image-based features versus landmark-based feature extraction, assessing their effectiveness in recognizing both normal and hurt emotions.

The specific objectives of this study include:

(1) Comparing the performance of image-based and landmark-based FER approaches using standardized datasets.

(2) Assessing the strengths and weaknesses of CNN, LRC, SVM, RFC, and GBC in processing different facial features.

(3) Measuring accuracy, precision, recall, and F1-score to determine the robustness and efficiency of each approach.

The research contributes to the field in several ways. Firstly, it provides a direct comparison of the performance of image-based methods, which use a histogram of oriented gradients (HOG) features, and landmark-based methods, which rely on facial landmarks for feature extraction. This comparison is based on a large and diverse dataset, including normal (neutral) and hurt emotions. Secondly, the study assesses the performance of a variety of widely used ML algorithms, identifying the most effective models for each technique. This evaluation serves as a comprehensive guide for selecting the optimal method for different FER tasks. Lastly, the study employs a range of performance metrics, including accuracy, precision, recall, and F1-score, ensuring a thorough and balanced assessment of the model's effectiveness in FER.

While both approaches in this study offer distinct advantages, they also come with limitations that could affect the results. The image-based approach, particularly when using deep learning models like CNN, demands significant computational resources and may be sensitive to variations in image quality and lighting conditions. Additionally, the feature extraction process, such as with HOG, might fail to capture subtle facial expressions better represented by the geometry of facial landmarks. On the other hand, the accuracy of the landmark-based approach is heavily reliant on the quality and precision of the landmark detection process. Factors like occlusions, misalignments, or low-resolution images can compromise the reliability of feature extraction. Furthermore, landmark-based methods may not effectively capture the dynamic changes in facial expressions, as image-based methods can. To address these limitations, this study employs a diverse dataset and carefully selected models, ensuring that the comparison remains as balanced and robust as possible.

The research concept is shown in Fig. 1, which visually represents the comparative framework for FER using image-based and landmark-based approaches. It highlights key features such as "Normal" and "Hurt" as critical markers for evaluating the effectiveness of different methods.

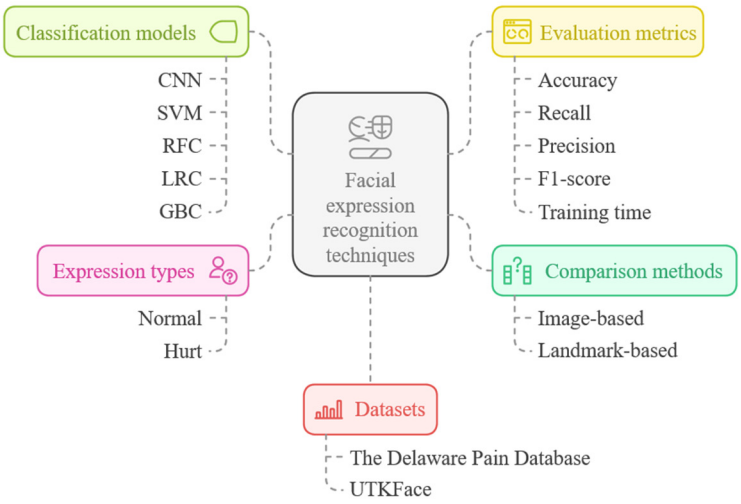


Fig. 1 FER comparative framework

2. Literature Review

FER has been an important area of research in computer vision, artificial intelligence, and human-computer interaction. This section reviews significant contributions in FER areas, aligning with the comparative works in Table 1 and positioning the current study within this research landscape. Early work, Buck et al. [5], focused on manual coding of facial expressions, which provided foundational insights but lacked automation. The introduction of ML in FER began with Vapnik [17], who developed SVM, a key advancement in classification methods. However, early ML applications in FER were limited in scope and required further refinement.

As artificial intelligence advanced, various approaches to FER were explored. Lisetti and Schiano [6], along with Picard et al. [18], expanded the field by integrating affective computing, enabling machines to automatically interpret emotional states. Bartlett et al. [19] later demonstrated the effectiveness of CNN for real-time FER, though their study was constrained by the availability of labeled data. Other foundational works, such as those of Lu and Weng [10], along with Hilbe [13], examined classification techniques and statistical modeling, but did not provide direct comparisons between traditional ML and deep learning models in FER.

Beyond general FER applications, researchers have explored its use in healthcare and specialized domains. Lee and Park [20] developed a fast region-based convolutional neural networks (R-CNN)-based FER model for diagnosing depressive disorders, demonstrating its potential in mental health assessments. However, their study relied only on image-based features, without considering how landmark-based features might enhance diagnostic accuracy. Similarly, Bekele et al. [21] introduced a virtual reality-based FER system, while Assari and Rahmati [22] focused on drowsiness detection in driver monitoring systems. While these studies highlighted important applications, they did not compare landmark-based and image-based methods, limiting their scope.

Many studies have concentrated on image-based FER techniques, refining classification models to improve accuracy. Hamester et al. [23] proposed a two-channel CNN architecture, where one channel processed raw image data and the other extracted features using an autoencoder. While this approach improved classification accuracy, it lacked comparisons with traditional ML classifiers. Alsubari et al. [24] incorporated wavelet transforms and local binary patterns (LBP) to enhance feature extraction and fusion, improving recognition rates. However, the study did not explicitly analyze the individual contribution of landmark-based features to FER performance.

Alongside image-based approaches, scholars have also explored landmark-based FER. Söylemez et al. [25] applied three-dimensional facial landmark distances with SVM, showing strong performance in recognizing expressions from 3D face models. However, their study did not compare landmark-based methods to traditional image-based techniques. Munasinghe [26] used RFC for landmark-based FER, demonstrating efficient training times but relying on a relatively small dataset. More recently, Sharma et al. [9] optimized landmark-based SVM models, improving robustness but highlighting their sensitivity to errors in facial landmark detection. Di Luzio et al. [27] proposed deep neural networks (DNNs) for processing hierarchical landmark features, though their study did not include a direct comparison between landmark-based and image-based models. Hangaragi et al. [28] addressed this limitation by implementing a Face Mesh-based deep learning model, which achieved real-time efficiency but lacked comparative analysis with landmark-based classifiers.

Despite extensive research on both image-based and landmark-based FER, a notable gap remains: few studies provide direct performance comparisons between the two approaches. Most prior work has focused on a single classifier, making it difficult to assess how different models generalize across datasets. Additionally, while studies such as Lee and Park [20] demonstrated the potential of image-based FER for medical diagnosis, they did not explore landmark-based techniques, which could enhance interpretability and robustness in clinical settings.

To address these gaps, this study conducts a detailed evaluation of both image-based and landmark-based FER methods across multiple classifiers, including CNN, SVM, LRC, RFC, and GBC. Furthermore, by using diverse datasets such as the Delaware Pain Database and UTKFace, the study aims to improve generalizability across different demographics and application domains. This research extends the findings of Lee and Park [20], Hangaragi et al. [28], and Di Luzio et al. [27], providing a comprehensive performance analysis of image-based and landmark-based FER techniques across multiple ML classifiers.

Table 1 Chronological comparison of studies

Study	Methodology	Models used	Key findings	Limitation	Novelty in study
Buck et al. [5], 1972	Image-based	Human-coded expressions	Early facial expression analysis	No automated recognition	Introduces ML-based automated recognition
Vapnik [17], 1999	Image-based	SVM	Foundational ML theory	No FER application	Test SVM on FER tasks
Lisetti and Schiano [6], 2000	Image-based	AI-based interpretation	Human-computer interaction focus	No ML comparison	Broadens scope to modern ML models
Picard et al. [18], 2001	Image-based	Physiological state analysis	Early ML in affective computing	No landmark feature use	Evaluates landmark features
Bartlett et al. [19], 2003	Image-based	CNN	Real-time FER	Limited dataset	Uses multiple datasets for comparison
Lu and Weng [10], 2007	Image-based	Image classification survey	Overview of classification methods	No specific FER application	Direct application to FER
Hilbe [13], 2009	Image-based	LRC	Foundational statistical modeling	No comparison with deep learning	Compares classical ML vs. DL models
Assari and Rahmati [22], 2011	Image-based	FER-based drowsiness detection	Application-specific	No direct comparison with landmark-based methods	Generalizes findings beyond drowsiness detection
Hamester et al. [23], 2015	Image-based	CNN	2-channel CNN for expression recognition	No comparison with classical ML models	Broadens evaluation across ML models
Bekele et al. [21], 2017	Image-based	VR-SAAFE for FER	Application-specific	No classical ML comparison	Expands analysis to classical ML models
Söylemez et al. [25], 2017	Landmark-based	SVM	Distance-based features for 3D FER	No image-based comparison	Extends analysis to both 2D and image-based methods

Table 1 Chronological comparison of studies (continued)

Study	Methodology	Models used	Key findings	Limitation	Novelty in study
Alsubari et al. [24], 2017	Image-based	wavelet transform + LBP	Feature fusion improves recognition	No landmark feature analysis	Incorporates landmark features for comparison
Munasinghe [26], 2018	Landmark-based	RFC	Fast training	Limited dataset	Extends analysis with a larger dataset
Michael Revina and Sam Emmanuel [1], 2021	Image-based	CNN	High accuracy	Computationally expensive	Direct comparison with landmark-based methods
Lee and Park [20], 2022	Image-based	Fast R-CNN	FER-based depressive disorder diagnosis	No landmark-based evaluation	Applied FER to clinical diagnostics
Sharma et al. [9], 2023	Landmark-based	SVM	Improved robustness	Sensitive to detection errors	Evaluates multiple classifiers
Di Luzio et al. [27], 2023	Landmark-based	DNN	Hierarchical features	No comparison with image-based methods	Direct performance comparison
Hangaragi et al. [28], 2023	Image-based	Face Mesh + DNN	Efficient for real-time applications	Lacks comparison with landmark-based models	Extends analysis with landmark-based methods
This study	Image-based and landmark-based	CNN, SVM, LRC, RFC, and GBC	Landmark-based features enhance performance	Computational trade-offs	First detailed evaluation of both methods across multiple classifiers

3. Machine Learning Techniques

This section provides an overview of the ML algorithms used in this study, the preprocessing steps, and training procedures for each approach. The comparative setup of these methods is summarized in Table 2. Each algorithm, including CNN, RFC, LRC, SVM, and GBC, was selected for its ability to handle both image and landmark data, with distinct preprocessing techniques applied to optimize model performance.

(1) Convolutional neural networks (CNN)

CNN are widely used in image processing due to their ability to learn spatial hierarchies from pixel data. In this study, CNNs were applied in both image-based and landmark-based approaches. For the image-based CNN, raw images were resized to 128×128 pixels and normalized. The architecture consisted of three convolutional layers with Rectified Linear Unit (ReLU) activation, followed by max-pooling operations to reduce spatial dimensions. The output was flattened and passed through fully connected layers, ending with a sigmoid-activated neuron for binary classification. The Adam optimizer and binary cross-entropy loss were used for training. For the landmark-based CNN, images were processed using MediaPipe to extract 468 facial landmarks, which were reshaped into a 1D input format. The CNN architecture included a 1D convolutional layer, followed by max-pooling, flattening, and dense layers, similar to the image-based approach. Feature standardization was applied using the StandardScaler.

(2) Random forest classifier (RFC)

RFC is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to enhance accuracy and reduce overfitting. For the image-based approach, HOG was used to extract texture features before classification. In the landmark-based approach, the extracted facial landmarks were used directly. Standard scaling was applied to the extracted features before training. The model was trained using default RFC parameters, with the number of trees set to 100.

(3) Logistic regression classification (LRC)

LRC is a simple yet effective classifier for binary classification problems. It models the probability of an instance belonging to a specific class using the logistic sigmoid function. For the image-based approach, HOG was used for feature extraction, while for the landmark-based approach, facial landmarks were directly utilized. Standardization was applied using the StandardScaler. The model was trained with default parameters, including L2 regularization to prevent overfitting.

(4) Support vector machine (SVM)

SVM aim to find the optimal hyperplane that maximizes the margin between different classes. For the image-based approach, HOG was used for feature extraction, while for the landmark-based approach, landmark coordinates were used. Feature normalization was applied using the StandardScaler. The model was trained using default parameters with a radial basis function (RBF) kernel to handle non-linear decision boundaries effectively.

(5) Gradient boosting classifier (GBC)

GBC is an ensemble method that sequentially builds decision trees to correct the errors of previous iterations. For the image-based approach, HOG was used for feature extraction, whereas the landmark-based approach used facial landmark coordinates. Standardization was applied to the extracted features. The model was trained using default parameters, including a learning rate of 0.1 and 100 estimators.

Table 2 Summary of machine learning methods and preprocessing steps

Model	Feature extraction	Preprocessing	Hyperparameters
CNN (image-based)	Raw image data	Resizing, normalization	Custom architecture (Conv2D layers, Adam optimizer)
CNN (landmark-based)	468 facial landmarks	Standardization	Custom architecture (Conv1D layers, Adam optimizer)
RFC	HOG (image); Landmarks (landmark-based)	Standardization	Default (100 trees)
LRC	HOG (image); Landmarks (landmark-based)	Standardization	Default (L2 regularization)
SVM	HOG (image); Landmarks (landmark-based)	Standardization	Default (RBF kernel)
GBC	HOG (image); Landmarks (landmark-based)	Standardization	Default (learning rate = 0.1)

4. Experiment

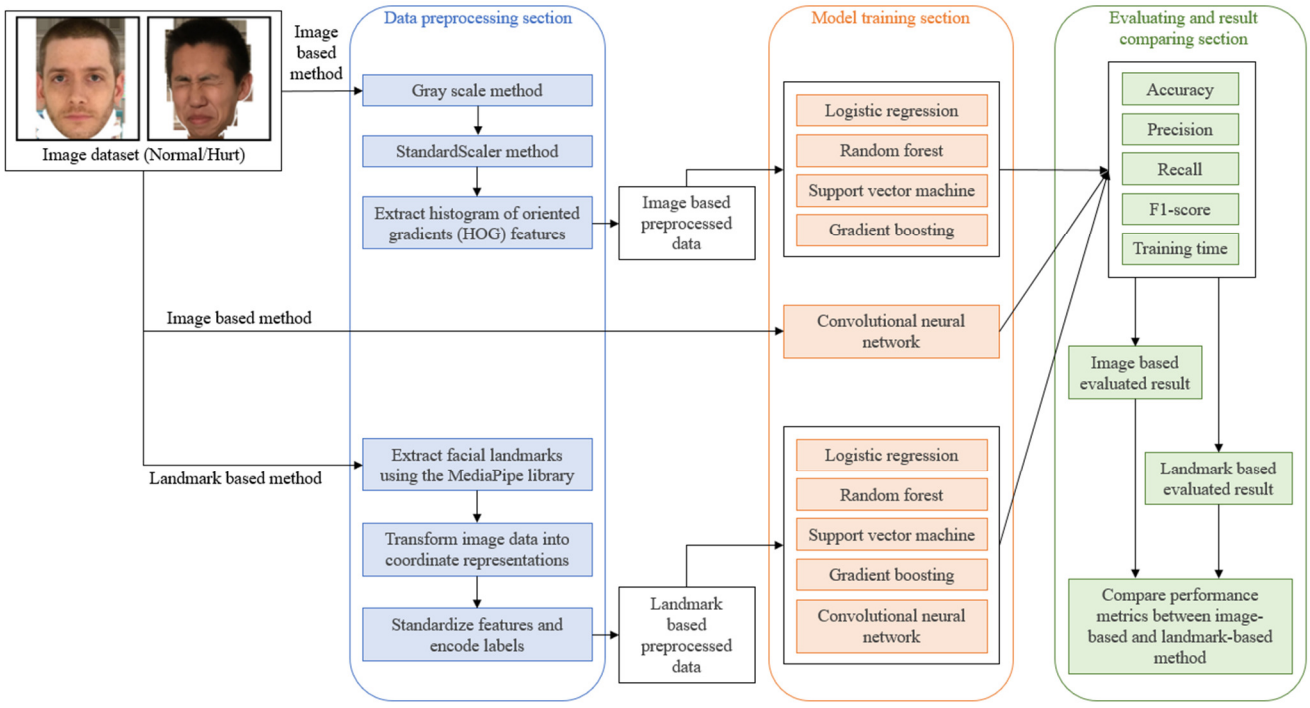


Fig. 2 Experiment pipeline

This section outlines the experimental framework used to evaluate the image-based and landmark-based FER methods. The experimental pipeline, depicted in Fig. 2, demonstrates the step-by-step process, from preprocessing and feature extraction to classification and performance evaluation. This experiment aims to compare the effectiveness of five classification models—

CNN, LRC, SVM, RFC, and GBC—on two distinct input types: image-based features using HOG and landmark-based features extracted from facial landmarks. Both approaches are assessed through a comprehensive set of performance metrics, including precision, recall, F1-score, and training time, to determine the most suitable method for reliable FER.

4.1. Dataset

This study utilized publicly available human image datasets, namely the Delaware Pain Database and UTKFace, which contain images depicting various facial expressions. The Delaware Pain Database consists of high-resolution images (1200×1200 pixels), while UTKFace provides lower-resolution images (200×200 pixels). To ensure consistency in model training, all images were resized to 128×128 pixels. For this study, only images displaying normal and hurt expressions were selected, resulting in 1,200 images. The dataset was divided into 800 images for training, 200 for validation, and 200 for testing. Example images from each category are shown in Fig. 3 and Fig. 4.

Fig. 3 Normal face example (the Delaware Pain Database and UTKFace)

Fig. 4 Hurt face example (the Delaware Pain Database and UTKFace)

4.2. Image-based method

The image-based method involved five classification models, like CNN, LRC, SVM, RFC, and GBC. Since LRC, SVM, RFC, and GBC do not inherently process raw image data, a feature extraction technique was required to convert images into numerical representations.

For preprocessing, all images were converted to grayscale and normalized by scaling pixel values between 0 and 1. Feature extraction was performed using the HOG technique, which captures edge and texture information by analyzing the distribution of gradient orientations. In this study, HOG parameters were set to a cell size of (8×8), a block size of (2×2 cells), and nine orientation bins. The extracted feature vectors were then standardized to improve model performance and ensure compatibility with classification algorithms. Once features were extracted, the LRC, SVM, RFC, and GBC models were implemented using scikit-learn. These models were trained with their default hyperparameters, as the focus of this study was not on fine-tuning but rather on evaluating their baseline performance for this classification task.

For CNN-based classification, a deep learning model was developed using the Keras framework. The CNN architecture consisted of three convolutional layers, designed to progressively capture hierarchical features within the image. The first layer employed 16 filters of size (3×3) with a stride of 1, followed by a ReLU activation function. The subsequent layers increased

the number of filters to 32 and 16, respectively. Max-pooling layers were incorporated to reduce spatial dimensions while preserving essential information. The extracted features were flattened and passed through a fully connected layer with 256 neurons, utilizing ReLU activation to capture high-level patterns. Finally, the output layer consisted of a single neuron with a sigmoid activation function, making it suitable for binary classification. The CNN model was trained using the Adam optimizer with a learning rate of 0.001, and binary cross-entropy loss was used to measure classification error. The training process spanned 50 epochs with a batch size of 32, incorporating early stopping to prevent overfitting.

The CNN architecture used in this study is illustrated in Fig. 5. The performance of all models was evaluated using precision, recall, and F1-score, providing a comprehensive assessment of classification effectiveness. Additionally, confusion matrices were generated to analyze misclassifications and model behavior. To assess computational efficiency, the training time for each model was recorded.

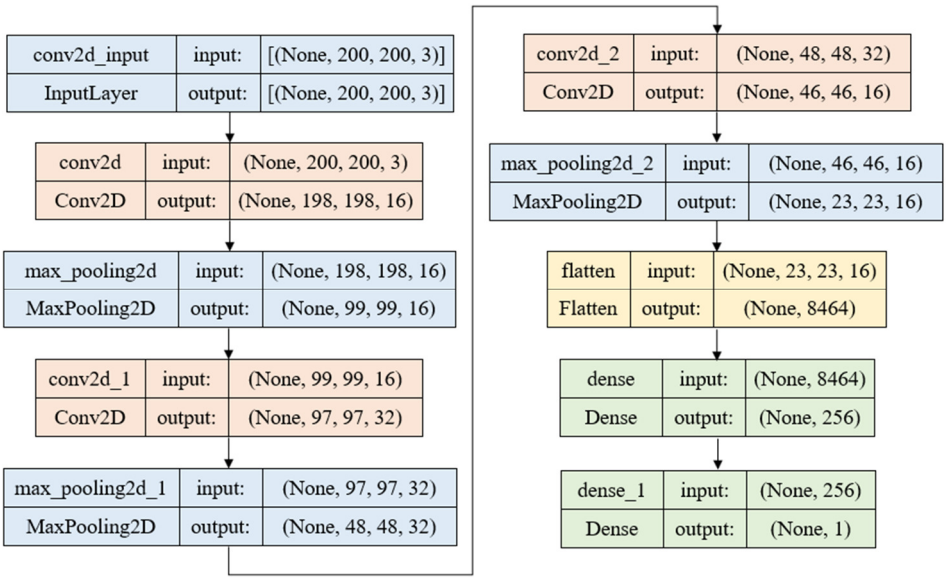


Fig. 5 Image-based CNN model architecture

4.3. Landmark-based method

In the landmark-based method, facial images were transformed into structured numerical representations by extracting key facial points. The MediaPipe library was employed to extract 468 facial landmarks, capturing the relative positions of key facial features along the x and y axes. This approach enabled the conversion of unstructured image data into a standardized format for ML classification.

Following preprocessing, the extracted landmarks were stored as structured data and normalized using feature standardization, ensuring values were scaled to a comparable range. The labels were encoded as binary values, with normal expressions labeled as 0 and hurt expressions labeled as 1. For CNN-based classification, the landmark data was reshaped into a format compatible with 1D convolutional layers, allowing the model to process sequential patterns in the landmark coordinates. As in the image-based method, the LRC, SVM, RFC, and GBC models were trained using default parameters, ensuring that the evaluation remained focused on baseline performance rather than fine-tuning.

For the CNN model, the architecture was adapted to handle 1D landmark data, replacing 2D convolutional layers with 1D convolutional layers. The model consisted of an initial 1D convolutional layer with 32 filters and a kernel size of 3, followed by ReLU activation and max-pooling to downsample the feature representation. A flattening layer was then used to convert the extracted spatial features into a one-dimensional vector. The fully connected layer comprised 128 neurons with ReLU activation, followed by an output layer with sigmoid activation for binary classification. The model was compiled using the Adam optimizer with a learning rate of 0.001, and training was performed with binary cross-entropy loss.

The CNN architecture for the landmark-based method is illustrated in Fig. 6. As with the image-based approach, model performance was assessed using precision, recall, and F1-score. Confusion matrices were utilized to further evaluate classification accuracy, while training time was recorded to compare computational efficiency.

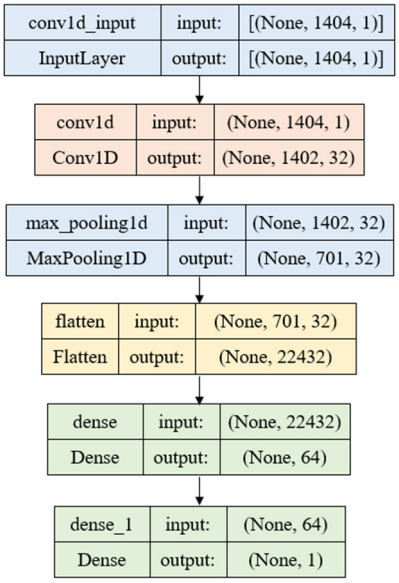
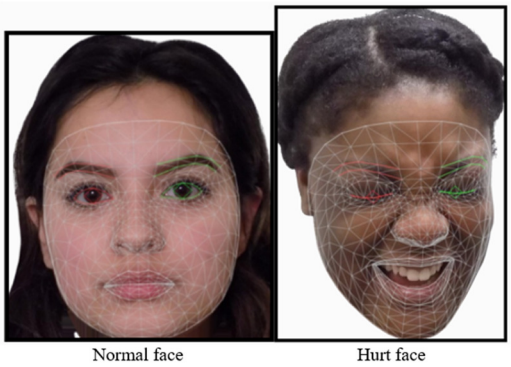


Fig. 6 Landmark-based CNN model architecture

5. Experimental Results

This study compares image-based and landmark-based FER techniques using various ML models, highlighting the distinct advantages and challenges of each approach, with the input data from both methods illustrated in Fig. 7, and the results, including performance metrics and model comparisons, summarized in Table 3.



B	C	D	E	
Expression	x1	x2	x3	x4
hurt	0.2081438	0.208931863	0.209579527	
hurt	0.166195542	0.166501015	0.169461668	
hurt	0.171509892	0.17361173	0.17438671	
hurt	0.202058852	0.202364177	0.205554873	
hurt	0.223162234	0.225129992	0.225894004	
hurt	0.186670214	0.188379377	0.189788938	
hurt	0.210388958	0.210486054	0.211479843	
hurt	0.208351076	0.209630042	0.212174743	
hurt	0.195476502	0.196630239	0.197068989	
hurt	0.21428436	0.214924812	0.225100398	
hurt	0.178058118	0.178706259	0.181971818	
hurt	0.212763101	0.213768989	0.214690089	
hurt	0.175352424	0.176146328	0.176562876	

Landmark-based

Fig. 7 Research input data illusion

Table 3 Machine learning model performance

Model	Techniques	Training time (seconds)	Accuracy	Precision	Recall	F1-score
SVM	Landmark	0.24	0.8550	0.8876	0.8550	0.8519
	Image-HOG-scaling	0.19	0.8200	0.8363	0.8200	0.8178
GBC	Landmark	31.72	0.8850	0.8965	0.8850	0.8842
	Image-HOG-scaling	20.93	0.8700	0.8724	0.8700	0.8698
RFC	Landmark	1.34	0.8950	0.9018	0.8950	0.8946
	Image-HOG-scaling	1.36	0.8600	0.8601	0.8600	0.8600
LRC	Landmark	0.11	0.8850	0.8939	0.8850	0.8843
	Image-HOG-scaling	0.09	0.7900	0.8047	0.7900	0.7874
CNN	Landmark	2.01	0.9500	0.9541	0.9541	0.9541
	Image-HOG-scaling	88.89s	0.8700	0.8558	0.8900	0.8725

The classification results of the five models on landmark-based and image-HOG-scaling features in Table 3 are summarized below:

(1) The SVM model performed strongly with the landmark-based technique (accuracy 0.8550, precision 0.8876, recall 0.8550, and F1-score 0.8519) but slightly decreased with the image-HOG-scaling method (accuracy 0.8200, precision 0.8363, recall 0.8200, and F1-score 0.8178).

(2) The GBC showed high performance with the landmark-based technique (accuracy 0.8850, precision 0.8965, recall 0.8850, and F1-score 0.8842) and also performed well with the image-HOG-scaling method (accuracy 0.8700, precision 0.8724, recall 0.8700, and F1-score 0.8698).

(3) The RFC achieved strong results with the landmark-based technique (accuracy 0.8950 and precision 0.9018) but lower performance with the image-HOG-scaling method (accuracy 0.8600 and precision 0.8601).

(4) The LRC was efficient in training time with the landmark-based technique (accuracy 0.8850, precision 0.8939, recall 0.8850, and F1-score 0.8843) but underperformed with the image-HOG-scaling method (accuracy 0.7900, precision 0.8047, recall 0.7900, and F1-score 0.7874).

(5) The CNN outperformed other models with the landmark-based technique (accuracy 0.9500, precision 0.9541, recall 0.9541, and F1-score 0.9541), while the image-HOG-scaling method also yielded strong results (accuracy 0.8700, precision 0.8558, recall 0.8900, and F1-score 0.8725).

These results meet the objective of comparing the performance of image-based and landmark-based FER techniques, providing insight into the effectiveness of different ML models for emotion recognition. The findings suggest that while landmark-based methods, such as those using facial feature points, offer higher accuracy and precision, image-based methods like HOG-scaling remain viable alternatives, although slightly lower in performance. This comparison informs future work on optimizing emotion recognition systems for real-world applications across various datasets and conditions.

6. Discussion

This study emphasizes the critical role of selecting appropriate ML models and feature extraction methods to achieve optimal classification outcomes. The performance of each model is summarized as follows:

(1) The SVM model exhibited sensitivity to the feature extraction techniques employed, with the landmark-based approach significantly outperforming the image-based method. This finding underscores the importance of landmark features in enhancing the SVM's ability to identify complex patterns in facial expression data, corroborating the conclusions of both Sharma et al. [9] and Michael Revina and Sam Emmanuel [1].

(2) The GBC demonstrated robust performance across both feature extraction techniques, indicating its adaptability to diverse data representations. Although the training duration was longer, particularly when using the landmark-based method, the GBC's superior accuracy, precision, recall, and F1-scores justified the extended training period. This finding aligns with Di Luzio et al. [27], who highlighted GBC's efficacy in managing complex datasets.

(3) The RFC displayed consistent performance across both feature extraction approaches, achieving high precision, recall, and F1-scores while requiring relatively short training times. This combination of efficiency and effectiveness positions RFC as a reliable model for classification tasks, echoing the results observed by Munasinghe [26] and Canedo and Neves [2], who demonstrated RFC's reliability in image classification.

(4) The LRC exhibited remarkable efficiency in training time, especially with the landmark-based technique, where training required only 0.11 seconds. However, the notable decline in performance with the image-based method suggests that LRC is more sensitive to the quality of feature selection. This aligns with findings by Kanjanawattana et al. [3], who identified limitations in the use of LRC for more complex classification tasks.

(5) The CNN emerged as the top-performing model, particularly with the landmark technique, achieving superior accuracy, precision, recall, and F1-scores. Despite the longer training time, particularly with the image-based technique (88.89 seconds), the substantial improvements in key performance metrics justify the increased computational cost. The CNN's ability to automatically learn hierarchical features from raw image data positions it as an effective choice for FER, as demonstrated by Hilbe [13] and Bodini [11].

In summary, these findings emphasize the importance of careful model and feature selection. The researcher highlights the trade-offs between computational time and performance. For higher accuracy, computationally intensive models like CNN, particularly when paired with landmark-based features, offer substantial benefits.

7. Conclusions

This study compared five machine learning models (CNN, SVM, RFC, LRC, and GBC) for recognizing facial expressions. Both image-based HOG features and landmark-based methods were evaluated to classify normal and hurt expressions using data from the Delaware Pain Database and UTKFace. The main findings from this research are:

(1) Landmark-based methods consistently performed better than image-based HOG methods across all models, demonstrating greater effectiveness in distinguishing facial expressions.

(2) CNN achieved the best results with landmark-based features, reaching 95% accuracy, 95.41% precision, recall, and F1-score, making it the most effective method for applications requiring high accuracy.

(3) Traditional machine learning models (SVM, RFC, GBC) showed strong performance with landmark features, achieving accuracies between 85.5% and 89.5%, making them good alternatives when computing resources are limited.

(4) Large performance differences existed between feature extraction methods, with LRC showing the biggest gap (88.5% vs. 79% accuracy for landmark vs. image-based features).

(5) Analysis of computing efficiency showed trade-offs between accuracy and training time, with LRC having the fastest training (0.11 seconds) while CNN needed longer training time but provided better classification results.

(6) Facial landmark points offer more reliable geometric information for expression recognition compared to texture-based HOG features.

Future research should examine hybrid approaches that integrate geometric and appearance-based features, explore deep learning architectures tailored to landmark-based classification, and evaluate performance on larger, more diverse datasets, including cultural variations in pain expression. Additionally, real-time implementation studies and cross-dataset generalization capabilities warrant further investigation to enhance practical applicability.

Statement of Ethical Approval

For this type of study, statement of human rights is not required.

Statement of informed consent

For this type of study, informed consent is not required.

References

[1] I. Michael Revina and W. R. Sam Emmanuel, "A Survey on Human Face Expression Recognition Techniques," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 6, pp. 619-628, 2021.

[2] D. Canedo and A. J. R. Neves, "Facial Expression Recognition Using Computer Vision: A Systematic Review," Applied Sciences, vol. 9, no. 21, article no. 4678, 2019.
[3] S. Kanjanawattana, P. Kittichaiwatthana, K. Srivisut, and P. Praneetpholkrang, "Deep Learning-Based Emotion Recognition through Facial Expressions," Journal of Image and Graphics, vol. 11, no. 2, pp. 140-145, 2023.
[4] S. Schindler, C. Tirloni, M. Bruchmann, and T. Straube, "Face and Emotional Expression Processing under Continuous Perceptual Load Tasks: An ERP Study," Biological Psychology, vol. 161, article no. 108056, 2021.
[5] R. W. Buck, V. J. Savin, R. E. Miller, and W. F. Caul, "Communication of Affect through Facial Expressions in Humans," Journal of Personality and Social Psychology, vol. 23, no. 3, pp. 362-371, 1972.
[6] C. L. Lisetti and D. J. Schiano, "Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect," Pragmatics & Cognition, vol. 8, no. 1, pp. 185-235, 2000.
[7] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, and A. C. P. de Leon Ferreira de Carvalho, "Deep Learning for Biological Image Classification," Expert Systems with Applications, vol. 85, pp. 114-122, 2017.
[8] P. Mende-Siedlecki, J. Qu-Lee, J. Lin, A. Drain, and A. Goharzad, "The Delaware Pain Database: A Set of Painful Expressions and Corresponding Norming Data," Pain Reports, vol. 5, no. 6, article no. e853, 2020.
[9] U. Sharma, K. N. Faisal, R. R. Sharma, and K. V. Arya, "Facial Landmark-Based Human Emotion Recognition Technique for Oriented Viewpoints in the Presence of Facial Attributes," SN Computer Science, vol. 4, no. 3, article no. 273, 2023.
[10] D. Lu and Q. Weng, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," International Journal of Remote Sensing, vol. 28, no. 5, pp. 823-870, 2007.
[11] M. Bodini, "A Review of Facial Landmark Extraction in 2D Images and Videos Using Deep Learning," Big Data and Cognitive Computing, vol. 3, no. 1, article no. 14, 2019.
[12] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," First International Conference on Secure Cyber Computing and Communication, pp. 278-282, 2018.
[13] J. Hilbe, Logistic Regression Models, Boca Raton: Chapman & Hall/CRC, 2009.
[14] M. A. Chandra and S. S. Bedi, "Survey on SVM and Their Application in Image Classification," International Journal of Information Technology, vol. 13, no. 5, pp. 1-11, 2021.
[15] N. M. Abdulkareem and A. M. Abdulazeez, "Machine Learning Classification Based on Radom Forest Algorithm: A Review," International Journal of Science and Business, vol. 5, no. 2, pp. 128-142, 2021.
[16] Z. He, D. Lin, T. Lau, and M. Wu, "Gradient Boosting Machine: A Survey," https://doi.org/10.48550/arXiv.1908.06951, 2019.
[17] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., New York: Springer, 1999.
[18] R. W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175-1191, 2001.
[19] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," Conference on Computer Vision and Pattern Recognition Workshop, pp. 53-53, 2003.
[20] Y. S. Lee and W. H. Park, "Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN," Diagnostics, vol. 12, no. 2, article no. 317, 2022.
[21] E. Bekele, D. Bian, J. Peterman, S. Park, and N. Sarkar, "Design of a Virtual Reality System for Affect Analysis in Facial Expressions (VR-SAAFE); Application to Schizophrenia," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 6, pp. 739-749, 2017.
[22] M. A. Assari and M. Rahmati, "Driver Drowsiness Detection Using Face Expression Recognition," IEEE International Conference on Signal and Image Processing Applications, pp. 337-341, 2011.
[23] D. Hamester, P. Barros, and S. Wermter, "Face Expression Recognition with a 2-Channel Convolutional Neural Network," International Joint Conference on Neural Networks, pp. 1-8, 2015.
[24] A. Alsubari, D. N. Satange, and R. J. Ramteke, "Facial Expression Recognition Using Wavelet Transform and Local Binary Pattern," 2nd International Conference for Convergence in Technology, pp. 338-342, 2017.
[25] Ö. F. Söylemez, B. Ergen, and N. H. Söylemez, "A 3D Facial Expression Recognition System Based on SVM Classifier Using Distance Based Features," 25th Signal Processing and Communications Applications Conference, pp. 1-3, 2017.
[26] M. I. N. P. Munasinghe, "Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier," IEEE/ACIS 17th International Conference on Computer and Information Science, pp. 423-427, 2018.
[27] F. Di Luzio, A. Rosato, and M. Panella, "A Randomized Deep Neural Network for Emotion Recognition with Landmarks Detection," Biomedical Signal Processing and Control, vol. 81, article no. 104418, 2023.
[28] S. Hangaragi, T. Singh, and N. N, "Face Detection and Recognition Using Face Mesh and Deep Neural Network," Procedia Computer Science, vol. 218, pp. 741-749, 2023.