International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Advancing emotion recognition in facial expressions through PCA, RFE, and MLP Integration

Irfan Qutab [1, *], Zheng Jiangbin [1], Muhammad Aqeel [1], Unaiza Fatima [1] and Imtiaz Ahmed Butt [2]

[1] *School of Software, Northwestern Polytechnical University, Xi'an, China.*
[2] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.*

International Journal of Science and Research Archive, 2024, 12(02), 2531–2542

Publication history: Received on 12 July 2024; revised on 19 August 2024; accepted on 22 August 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1549

Abstract

Emotion recognition through facial expressions is vital for enhancing human-computer interaction, making systems more intuitive and responsive to user needs. This study introduces an innovative approach to emotion detection, leveraging Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) for feature extraction and optimization. The methodology focuses on refining facial feature representations to improve classification accuracy, which is critical for accurately detecting emotions like happiness, sadness, fear, and surprise. The approach was applied to two widely recognized facial emotion datasets: the Cohn-Kanade (CK) and the Japanese Female Facial Expression (JAFFE) datasets. By integrating PCA and RFE, the model efficiently selects the most relevant features, enhancing the overall performance of emotion recognition. A comparative analysis with existing deep learning models highlights the advantages of the proposed method. The effectiveness of this approach is further supported by the results, where the model achieves an accuracy of 98.49% on the CK dataset and 96.03% on the JAFFE dataset. These results demonstrate a significant improvement over recent methods, indicating the model's potential for real-world applications.

Keywords: Facial Emotion Recognition; Principal Component Analysis (PCA); Recursive Feature Elimination (RFE); Multilayer Perceptron (MLP); Emotion Detection

1. Introduction

Emotion, a state of mind, is intricately tied to the nervous system, encompassing feelings, perceptions, behavioral responses, and levels of pleasure or discomfort [1]. A prominent application of AI, particularly through neural networks, involves recognizing faces in images and videos for diverse uses. These techniques primarily analyze visual data to identify common facial patterns. Face detection has practical applications in surveillance by law enforcement and crowd control. This paper introduces a method to identify seven specific emotions - neutral, disgust, anger, fear, happiness, surprise, and sad - through facial imagery. Prior studies have leveraged deep learning to develop models that correlate facial expressions with emotional states [2]. Traditional human-computer interaction (HCI) often overlooks the emotional state of the user, leading to significant information loss during interaction. In contrast, HCI systems sensitive to emotions are more effective and sought after. Recently, the field of emotional computing has garnered increased interest, driven by the growing needs in entertainment, commerce, health, and educational sectors. Consequently, numerous emotionally responsive HCI systems have been developed in recent years, yet a definitive solution in this research area remains elusive. Emotions, serving as a key component in non-verbal communication, transcend linguistic barriers.

They are a global phenomenon. Often, our facial expressions inadvertently reveal our inner feelings. Facial expressions contribute to 55% of the communication regarding emotions and attitudes. Ekman [3] was a pioneer in studying how

* Corresponding author: Irfan Qutab

facial characteristics map to six fundamental emotions: happiness, disgust, sadness, fear, surprise, and anger. This study laid the groundwork for the possibility of identifying emotions through facial expressions, a field known as Facial Emotion Recognition (FER). FER holds substantial promise both in academic research and commercial applications. Emotions, complex psychological aspects, significantly influence individual interactions and impact human behavior [4]. The capability to identify and comprehend emotions during communication is essential for effective interactions. Single-modality emotion recognition, utilizing only facial expressions or speech, often leads to inaccuracies [5] [6]. To overcome this limitation, multimodal emotion recognition methods have emerged [7] [8]. These techniques aim to enhance the precision of emotion recognition systems by using diverse data sources, such as facial expressions [9] [10], and textual content [11] [12]. Achieving a more accurate depiction of an individual's emotional state may be accomplished by combining data from several sensory inputs. This has led to a growing focus on developing advanced multimodal systems for the identification of human emotions in recent times.

This study aims to introduce an innovative approach for detecting emotions present in facial expressions by using PCA (Principal Component Analysis) and RFE (Recursive Feature Elimination) methods that are described in section 3. Training stage of proposed model involves Laplacian filtering, extracting features, and optimizing these features. In Section 4, we present the findings and outcomes of the conducted experiments. Subsequently, our observations are summarized and concluded in section 5.

2. Literature Review

Studies in the field of facial emotion detection have concentrated on the identification of human emotions through images or video datasets. Some latest research endeavors have aimed at face recognition in images or videos, these methods did not employ PCA with RFE for feature extraction and feature optimization from the images. Agrawal et al. [13] provided a foundational review of methodologies for recognizing facial emotions, offering a comparative study of different techniques for natural facial expression recognition. Similarly, the authors [14] employed transformation technique on a narrowed subset comprising 117 landmarks for the purpose of emotion analysis.

In the context of form-based emotion recognition, a collection of features is derived from pixel images to train the classification model. Nandani et al. [15] proposed a system for emotion detection. They employed two different models CNN and combination of PCA and SVM. Their proposed system achieved the accuracy of 95.35 percent on JAFFEE and 81.81 percent on CK+ datasets respectively. Additionally, Chiranjeevi et al. [16], demonstrated an effective method for detecting facial expressions, focusing on the varied movements of facial attributes such as eyes, nose, and mouth. The features they employed are notable for their relevance to adjacent spatial positions. A comprehensive overview of contemporary research in facial expression recognition is also documented [17]. A framework for recognizing emotions has been developed, as detailed in [18], utilizing the critical movements of facial muscles.

The authors implemented the K-nearest neighbour algorithm for categorization, successfully identifying four emotions—happiness, disgust, anger, and surprise with a precision rate of 80%. In a different approach, Black and Yacoob [19] focused on tracking the movements around key facial areas like the mouth, eyes, and eyebrows for extracting features. Their method effectively classified six fundamental emotions, achieving an accuracy of 88%. Further, Shang et al. [20] aimed at identifying Action Units (AU). Chernykh et al. [21] established a model employing point tracking for emotion classification in sequences. Techniques involving hybrid CNNs for emotion identification in images were introduced by Krestinskaya and James [22], Lopes et al. [23], and Jain et al. [24].

Kumar et al. [25] enhanced the fdlibmex technique's performance through preprocessing, including the use of median filtering to refine feature extraction. Murtaza et al. [26] presented a method for segmenting facial features into three sections, focusing the processing on areas with significant variation. They utilized facial motion coding and animation parameters to boost performance. More recent works include application-focused methods. For instance, Virrey et al. [27] designed a comprehensive approach using various facial traits for pain detection, achieving 85.66% accuracy on the FER dataset they compiled. Zhang et al. [28] introduced an innovative spatial-temporal Recurrent Neural Network that integrates features derived from both spatial and temporal data of signal sources. This model combines EEG-based emotion recognition with facial emotion recognition to identify emotions. The integration of RNN components enables the model to learn and capture dependencies existing in both spatial and temporal dimensions.

Face mining refers to the automated identification of human faces and their distinct features, such as eyes, eyebrows, and lips. In their research, the cutting-edge FPD algorithm and the innovative GLCM algorithm have been utilized to extract attributes from facial images. FPD operates based on the bounding frame principle, while GLCM employs an affinity invariant. The performance metrics used in this context are precision and the duration of feature extraction. Khorrami et al. [29] introduced emotion recognition on video dataset through the utilization of CNNs and RNNs. The study thoroughly examined the impact of various hyperparameters on the overall performance of the model. The outcomes demonstrated superior performance compared to alternative methods in the field. In our study, we have evaluated the effectiveness of PCA and RFE in categorizing seven distinct facial expressions (sadness, neutrality, irritation, happiness, anger, fear, and surprise) in two individuals using the JAFFE (Japanese Female Face Expression) database. Our aim is to identify superior methods for facial emotion detection. This research underscores the practicality of facial expression analysis in real-world scenarios, including surveillance and human-computer interactions.

3. Proposed Methodology

The methodology depicted in the figure 1 illustrates a systematic process for detecting emotions from facial expressions. The process begins with the input of a facial image, which undergoes preprocessing to enhance key features. This is achieved using a Laplacian filter, which focuses on the Region of Interest (ROI) to highlight significant facial features such as the eyes, nose, and mouth. Following this, the system identifies and isolates the area of interest, essential for accurate feature extraction. Next, feature extraction is performed using Principal Component Analysis (PCA), a method that reduces data dimensionality while preserving the most important features. These extracted features are then refined through Recursive Feature Elimination (RFE), which selects the most relevant features for the task. Finally, the optimized features are used in the emotion classification stage, where the system identifies the emotions depicted in the facial expressions. The output consists of the detected emotions, effectively concluding the sequential process from input to emotion recognition. This methodology highlights the iterative process of enhancing and refining the emotion detection system to achieve higher precision in classification.

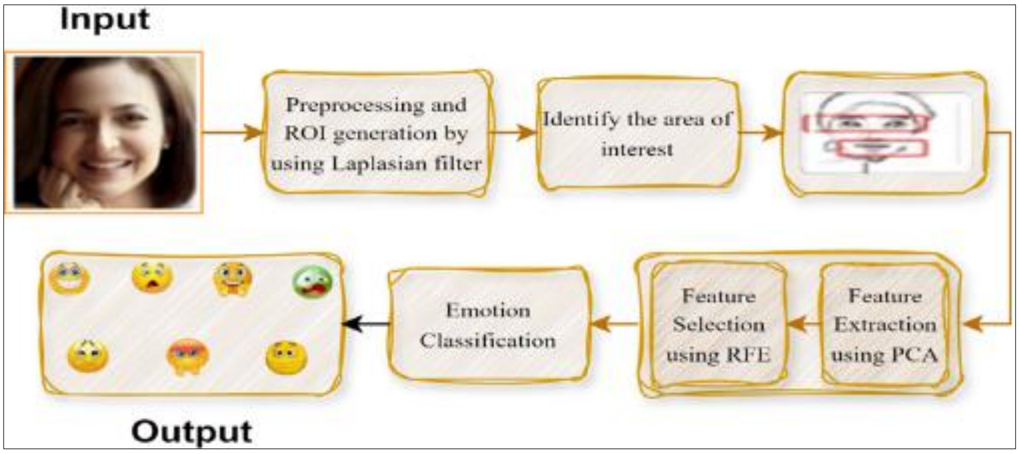

Figure 1 Proposed Methodology of the System

3.1. Dataset

We employed two significant datasets: CK and JAFFE datasets, each featuring seven distinct emotional expressions. These datasets have been selected for their extensive and varied portrayal of human emotions, ranging from happiness and sadness to anger, fear, surprise, disgust, and a neutral state. The CK [30] dataset is widely recognized for its varied expression dynamics, predominantly from adult subjects, and is valued for its high-quality images that adeptly depict the evolution of facial expressions. Contrastingly, the JAFFE dataset consists of images of Japanese female models, offering a varied collection of static facial expressions.

Figure 2 CK dataset sample

Figure 3 JAFFEE dataset sample

In order to improve the ability to apply learned knowledge to new situations, we carried out training using datasets that were not identical to each other. The CK dataset consists of 486 video clips showcasing 97 distinct individuals posing. The JAFFE dataset has 213 photos, representing seven face expressions, which include six primary emotions and one neutral expression. These emotions were presented by 10 Japanese female models, and the pictures had dimensions of 256 × 256-pixel values. Every expression is shown by a range of 2 to 4 samples. The experimental dataset consists of 6485 photos, of which 5992 are used for training, while the remaining images are reserved for testing and validation. Refer Table 1 for a comprehensive summary of the dataset.

Table 1 Dataset description.

Base	Dataset	No. of images
Training	CK	5849
	JAFFEE	143
Testing	CK	332
	JAFFEE	34

3.2. Preprocessing

Using image normalization as a crucial pre-processing approach is necessary to reduce the discrepancy in features across different classes, which might appear as consistent differences in intensity. The offsets stay constant throughout the immediate area. The Laplacian filter was used for preprocessing. The model shown exhibits the capacity to automatically adapt to different picture sizes without the need for human interaction. The cropping area is determined using a vertical factor of 4.7, with 1.4 allocated for the upper region (above the eyes) and 3.3 for the bottom region. This successfully accounts for the distance among the midpoint of the eyes and the center of the right eye. Similarly, the horizontal cropping portion is determined by a factor of 2.5, which corresponds to the same distance. The values of these particular factors were found by empirical investigation. Figure 4 shows an instance of this process.

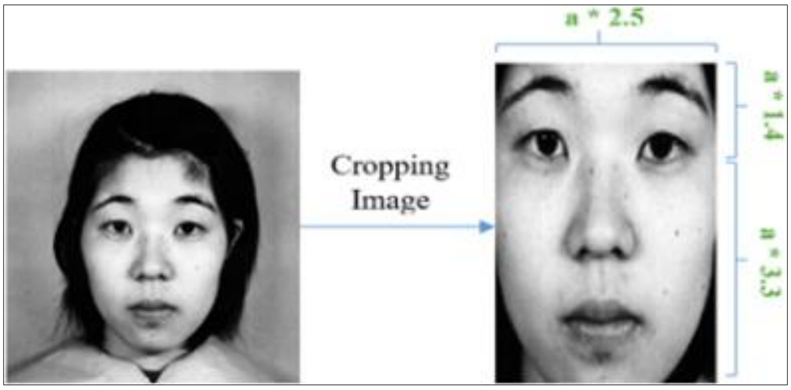

Figure 4 Example of image cropping with the goal of eliminating non-expression features, including background and hair, from the image

3.3. Laplacian Filter

In our proposed methodology, we integrated the Laplacian filter for normalization, edge detection and image enhancement. The Laplacian filter, based on the concept of second-order derivatives, is instrumental in highlighting areas of rapid intensity change within images, thereby effectively identifying edges. This is mathematically represented by the Laplacian operator, $\nabla^2 I$, defined as the sum of the second partial derivatives with respect to the x and y coordinates of the image function I(x, y). The operator is given by the equation:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad \text{.......................... (1)}$$

Variations in image intensity and contrast are evident even within images of the same person expressing the same emotion. This diversity in the feature vector introduces complexity to the classification task for each emotion. To address these challenges, intensity normalization is employed. This method, derived from [31] and referred to as contrastive equalization, operates as a two-step process. Initially, variations in intensity and contrast are mitigated, and subsequently, the normalization procedure is executed.

The (5*5) kernel is applied to each pixel of the image, enhancing edges by highlighting the differences in intensity among a pixel and its surrounding neighbors. Given the Laplacian filter's sensitivity to noise, we precede its application with a Gaussian smoothing step, we adopted a Laplacian of Gaussian (LoG) approach. Firstly, the weighted average of neighboring pixels is subtracted from the value of each pixel. Subsequently, the pixel value is separated by the typical deviation of its adjacent pixels. Figure 5 illustrates an instance of this process.

Figure 5 Intensity Normalization

3.4. Feature Extraction

During the feature extraction step, the pixel data is converted into representations that capture the motion, shape, texture, color, and spatial arrangement of the face or its individual components. The purpose of extracting facial characteristics is to discern the most relevant data from face photographs for the purpose of recognizing expressions. Multiple scholars have proposed different kinds of features and their combinations, including geometrical properties, optical flow, active shape models, and Gabor wavelets. Geometric, statistical, and spectral-transform-based characteristics are often used as alternative representations of face expressions before they are classified. The choice of feature extraction algorithms is mostly determined by the desired level of recognition accuracy and the processing resources needed. For this work, we have opted to use the Laplacian filter as an initial stage and Principal Component Analysis (PCA) for extracting features in the identification of facial expressions.

3.5. Principal Component Analysis

PCA is a method employed for decreasing the number of variables or minimizing the feature dimensionality. Particularly prevalent in the realm of facial recognition, PCA utilizes an orthogonal transformation. This transformation statistically converts observations of corelated variables into a series of uncorrelated variables, linearly known as principal components. In the context of facial emotion recognition, this technique aims to distill facial features into a unique set of attributes representations termed Eigen faces. These representations are instrumental in identifying both existing and new faces. The training dataset comprises N images, all standardized to the same scale. This standardization is achieved by transforming each image matrix into an equivalent image vector i, thereby normalizing the dataset. The matrix comprises a collection of vectors as denoted in equation (2).

$$T = T_1, T_2, T_3, \dots, T_n \text{ .... (2)}$$

The average face X represents the arithmetic vector, as indicated by Equations written below.

$$X = {}^{1}/_{M} \sum_{k=1}^{m} T_i \text{ .... (3)}$$

$$D = T_i - X \text{ .... (4)}$$

The distinction helps in identifying the unique characteristics of emotional samples while eliminating shared features. The Eigenvector associated with facial emotions is determined through the covariance matrix, as illustrated in Equation (5).

$$C = X \times X^T \text{ .... (5)}$$

Thus, the eigenvectors are represented in equation (6).

$$U_i = X \times C_i \text{ .... (6)}$$

Therefore, the Eigen faces are defined as the resulting principal components that produce unique scores, as shown in the following equation:

$$R = U_1, U_2, U_3, \dots, U_N \text{ .... (7)}$$

3.6. Feature Optimization

Recursive Feature Elimination (RFE) is a feature selection method particularly effective in refining models by identifying and eliminating non-contributory features. It is especially useful in complex datasets where determining the most impactful features can significantly enhance model performance and interpretability. RFE starts with a base model, which can be any supervised learning algorithm capable of estimating feature importance. In the case of MLP, coefficients assigned to each feature can indicate their importance. The model is trained on the initial set of attributes, and each feature is assigned a ranking based on its importance. In MLP, this could be based on the absolute value of the coefficients, where larger values indicate higher importance. Recursive Elimination Process consists of several steps:

- The least important feature (or features) is identified based on the ranking.
- This feature is then removed from the feature set.
- The model is retrained with the remaining features, and the importance of each feature is reassessed.

The process of elimination and retraining is iterated. After each iteration, the feature set becomes progressively smaller. The iterations continue until a specified number of features is reached, or another stopping criterion (like a threshold in model performance) is met. The output of RFE is an optimized subset of features that contribute most significantly to the prediction task. In the context of MLP, this results in a more streamlined model that only includes features with the strongest relationships to the target variable.

- The mathematical representation of RFE, particularly in an MLP context is as follows:
- Let the full feature set be $\mathrm{X} = \{\mathrm{x}_1, x_2, \dots, x_n\}$.
- Train the MLP model on (X), resulting in a set of coefficients $\mathrm{C} = \{\mathrm{c}_1, c_2, \dots, c_n\}$.
- At each step, eliminate the feature (x_i) with the smallest $|c_i|$.
- Retrain the model on the reduced feature set and update the coefficients.
- Continue until the desired number of features is retained.

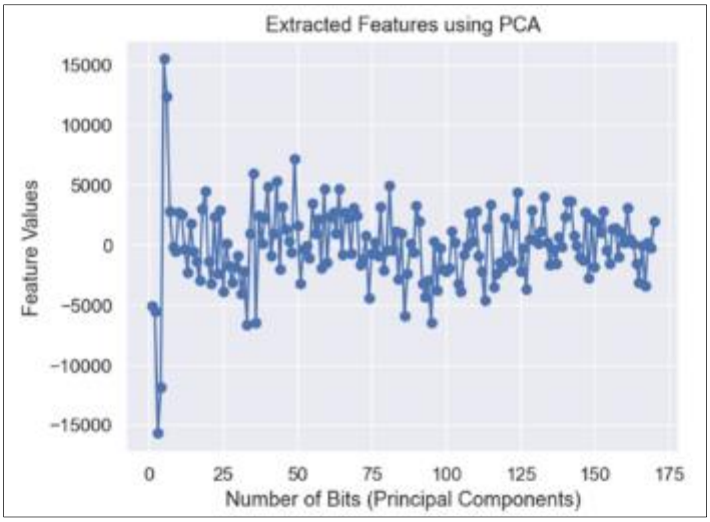

Figure 6 Features Extracted by using PCA

3.7. Proposed Algorithm

- Step 1: Initialize datasets with facial images and emotion labels: Training dataset $S = \{S_1, S_2, S_3, \dots, S_N\}$ and testing dataset $Y = \{Y_1, Y_2, Y_3, \dots, Y_N\}$.
- Step 2: Apply a Laplacian filter to each image: $L(I) = \nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$.
- Step 3: Flatten each image I of size $(m \times n)$ into a one-dimensional array $(I_{flattened})$ of size $(1 \times mn)$.
- Step 4: Perform PCA on the flattened images:
  - Compute the covariance matrix: $\sum = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)(X_i - \mu)^T$.
  - Eigen decomposition: $\sum V = V\lambda$.
  - Select top 'k' eigenvalues and eigenvectors: λ_k, V_k.
  - Project data onto selected eigenvectors: $T = XV_k$.
- Step 5: Encode emotion labels into integers: Map label l_i to an integer k using a function $f(l_i) = k$.
- Step 6: Scale features to zero mean and unit variance: $X_{scaled} = \frac{X - \mu x}{\sigma x}$.
- Step 7: Apply RFE using Random Forest Classifier to select features $F \subset T$ based on importance.
- Step 8: Train an MLP Classifier:
  - Define function $f(\cdot)$: $Y = f(X; W, b)$ with weights 'W' and biases 'b'.
  - Optimize loss function $L(Y, \hat{Y})$.
- Step 9: Predict and evaluate on the testing set:
  - Predictions: $\hat{Y} = f(Y_{test}; W, b)$.
  - Compute accuracy: $Accuracy = \frac{\sum_{i=1}^{M} (\hat{y}_i = y_i)}{M}$.

Figure 7 Processed image of Neutral Category

4. Experimental Results and Discussions

For the experimental phase, the authors utilized samples of (JAFFE) [32] and CK. The datasets are accessible to the public for non-commercial purposes. JAFFE encompasses images of 256 × 256 dimensions and CK consists of images of 48×48 dimensions. After conducting feature extraction using PCA, each facial image's feature vector is obtained, measuring 1 × 150. This process ensures a comprehensive representation of facial expressions, facilitating further analysis and interpretation in the context of emotion recognition.

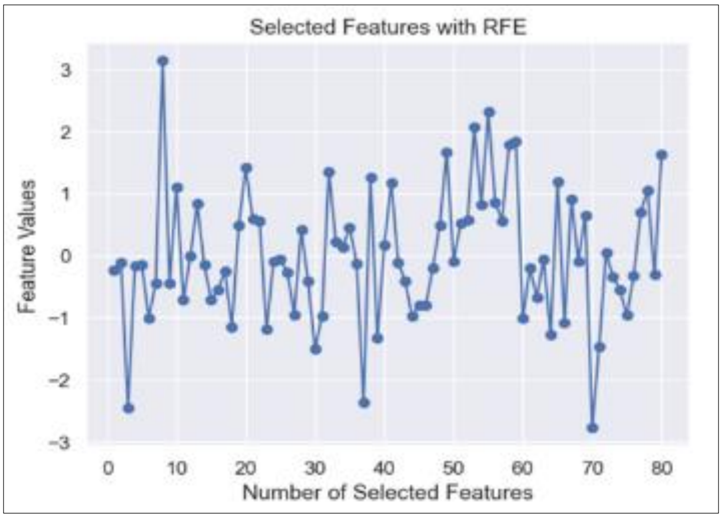

Figure 8 Optimized Features by Using RFE

Figure 9 Testing of emotion related to happy category

In Figure 6, an illustration is presented, showcasing the features extracted from the images by applying Principal Component Analysis technique. In Figure 7, the feature vector corresponding to each uploaded sample within the neutral category is displayed. This feature vector plays an important role in the subsequent testing phase for facial emotion classification. The optimized feature vectors, derived through Recursive Feature Elimination (RFE), are visually represented in Figure 8. Figure 9 showcases testing facial images representing emotions in the happiness category, contributing to the comprehensive exploration of various emotional expressions in the experimental context.

Two classifiers, Multinomial Logistic Regression (MLR) and the Multilayer Perceptron model (MLP), have been employed for facial emotion classification. Specifically, MLR demonstrates average classification rates of 95.38% on CK dataset and 88.46% on JAFFE dataset. Meanwhile, MLP achieves higher accuracies, with rates of 98.49% on CK and 96.03% on JAFFE. To assess the effectiveness of the MLP model, a comparison with other state-of-the-art approaches is detailed in Table 2. Despite notable achievements, facial expression recognition remains a challenging task. Factors such as age rejuvenation or cosmetic alterations can impact the performance of a facial expression detection algorithm [2, 3]. Additionally, challenges arise in scenarios where the face is either painted or covered, adding complexity to the recognition process.

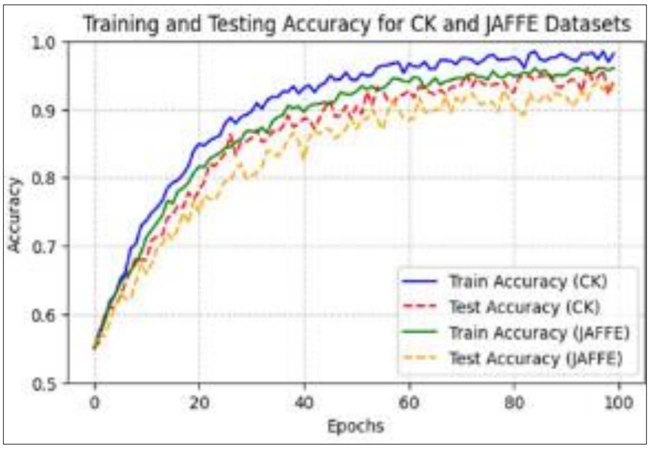

Figure 10 Training and Testing Accuracy

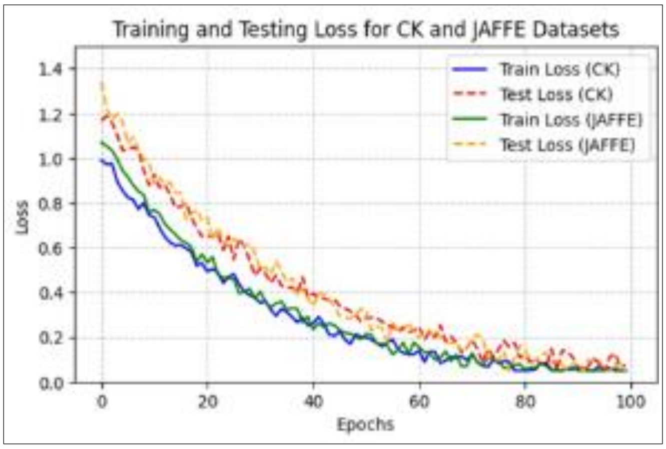

Figure 11 Training and Testing Loss

The above figures show training and testing accuracy and loss curves for the CK and JAFFE datasets over 100 epochs. The accuracy curves indicate steady learning, while the loss curves show a consistent decrease, reflecting effective model training and convergence for both datasets. While figure 12 and 13 shows confusion matrices of the model. The proposed performs well on both the CK and JAFFE datasets, accurately classifying emotions like "Happy" and "Neutral". The CK dataset has better accuracy than JAFFE, as reflected in the concentration of correct classifications along the diagonal.

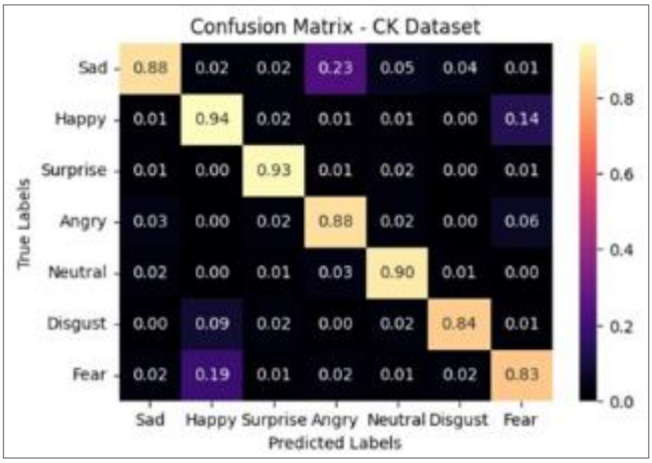

Figure 12 Training and Testing Accuracy

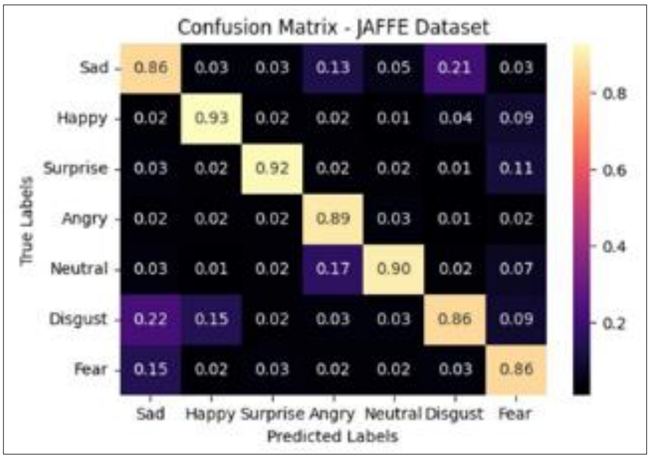

Figure 13 Training and Testing Loss

Table 2 presents a comparison of the performance of the proposed MLP model with other current deep learning models using the CK and JAFFE datasets. The proposed approach outperforms existing models. The suggested model achieves an accuracy of 98.49% on the CK dataset and 96.03% on the JAFFE dataset, outperforming all the other techniques mentioned. This clearly illustrates the efficacy and supremacy of the suggested method in identifying and understanding emotions.

Table 2 Comparison with Existing Approaches

Method	Accuracy JAFFEE	Accuracy CK
Chernykh et al. [21]	73%	70.12%
Krestinskaya and James [22]	94.89%	94.89%
Lopes et al. [23]	94.86%	94.86%
Jain et al. [24]	94.91%	94.91%

Zhang et al. [28]	94.89%	94.89%
Khorrami et al. [29]	82.43%	82.43%
Multinomial Logistic Regression	95.38%	88.46%
Proposed Model	96.03%	98.49%

5. Conclusion

This work introduces an effective approach for recognizing facial emotions. It does this by integrating Principal Component Analysis (PCA) with Recursive Feature Elimination (RFE) to enhance the process of extracting relevant features and classifying them accurately. The suggested technique underwent thorough testing on the Cohn-Kanade (CK) and Japanese Female Facial Expression (JAFFE) datasets, with a specific emphasis on refining the selection of facial characteristics to improve the accuracy of emotion categorization. The confusion matrices revealed strong classification capabilities across various emotions, particularly on the CK dataset. The comparative analysis demonstrated that the proposed model outperforms several existing approaches, providing superior accuracy. Specifically, the model achieved an accuracy of 98.49% on the CK dataset and 96.03% on the JAFFE dataset, showcasing its effectiveness in real-world applications. These findings underscore the potential of this approach for practical uses in surveillance, healthcare, and HCI, where accurate emotion recognition is crucial. Future research could aim to further refine the model to address challenges such as variability in facial expressions due to factors like age and occlusions, thereby enhancing its robustness in diverse scenarios.

Compliance with ethical standards

Acknowledgments

I wish to convey my sincere appreciation to my research fellows for their invaluable guidance and support during the progression of this research endeavor. Their expertise and encouragement have played a pivotal role in shaping the outcome of this study.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] S. Li and W. Deng, Deep facial expression recognition: A survey, IEEE transactions on affective computing, vol. 13, no. 3, pp. 1195-1215, 2020.

[2] E. Correa, A. Jonker, M. Ozo and R. Stolk, Emotion recognition using deep convolutional neural networks, Tech. Report IN4015, 2016.

[3] P. Ekman and W. V. Friesen, Constants across cultures in the face and emotion, Journal of personality and social psychology, vol. 17, no. 2, p. 124.

[4] D. Mamieva, A. B. Abdusalomov, A. Kutlimuratov, B. Muminov and T. K. Whangbo, Multimodal Emotion Detection via Attention-Based Fusion of Extracted Facial and Speech Features, Sensors, vol. 23, no. 12, p. 5475, 2013.

[5] N. Ahmed, Z. Al Aghbari and S. Girija, A systematic survey on multimodal emotion recognition using learning algorithms, Intelligent Systems with Applications, vol. 17, p. 200171, 2023.

[6] X. Gu, Y. Shen and J. Xu, Multimodal Emotion Recognition in Deep Learning: A Survey, International Conference on Culture-oriented Science & Technology (ICCST). IEEE, pp. 77-82, 2021.

[7] G. Tang, Y. Xie, K. Li, R. Liang and L. Zhao, Multimodal emotion recognition from facial expression and speech based on feature fusion, Multimedia Tools and Applications, vol. 82, no. 11, pp. 16359-16373, 2023.

[8] C. Luna-Jiménez, D. Griol, Z. Callejas, R. Kleinlein, J. M. Montero and F. Fernández-Martínez, Multimodal emotion recognition on RAVDESS dataset using transfer learning, Sensors, vol. 21, no. 22, p. 7665, 2021.

[9] M. Sajjad, F. U. M. Ullah, M. Ullah, G. Christodoulou, F. A. Cheikh, M. Hijji and J. J. Rodrigues, A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines, Alexandria Engineering Journal, vol. 68, pp. 817-840, 2023.

[10] Z. Song, Facial expression emotion recognition model integrating philosophy and machine learning theory, Frontiers in Psychology, vol. 12, p. 759485, 2021.

[11] F. Makhmudov, M. Mukhiddinov, A. Abdusalomov, K. Avazov, U. Khamdamov and Y. I. Cho, Improvement of the end-to-end scene text recognition method for text-to-speech conversion, International Journal of Wavelets. Multiresolution and Information Processing, vol. 18, no. 6, p. 2050052, 2020.

[12] A. Vijayvergia and K. Kumar, Selective shallow model's strength integration for emotion detection using GloVe and LSTM, Multimedia Tools and Applications, vol. 80, no. 18, pp. 28349-28363, 2021.

[13] D. D. Agrawal, S. R. Dubey and A. S. Jalal, Emotion recognition from facial expressions based on multi-level classification, International Journal of Computational Vision and Robotics, vol. 4, no. 4, pp. 365-389, 2014.

[14] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, Spatial–temporal recurrent neural network for emotion recognition, IEEE transactions on cybernetics, vol. 49, no. 3, pp. 839-847, 2018.

[15] S. Nandani, R. Nanavati and M. Khare, Emotion Detection Using Facial Expressions., In Futuristic Trends in Networks and Computing Technologies: Select Proceedings of Fourth International Conference on FTNCT 2021. Singapore: pp. 627-640, 2022.

[16] P. Chiranjeevi, V. Gopalakrishnan and P. Moogi, Neutral face classification using personalized appearance models for fast and robust emotion detection., IEEE Transactions on Image Processing, vol. 24, no. 9, pp. 2701-2711, 2015.

[17] A. Sánchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernández and J. J. Pantrigo, Differential optical flow applied to automatic facial expression recognition, Neurocomputing, vol. 74, no. 8, pp. 1272-1282, 2017.

[18] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, IEEE Transactions on pattern analysis and machine intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[19] M. J. Black and Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, International Journal of Computer Vision, vol. 25, pp. 23-48, 1997.

[20] S. Chen, Y. Sun and B. Yin, A novel hybrid approach based on sub-pattern technique and extended 2dpca for color face recognition, 11th IEEE International Symposium on Multimedia.IEEE, pp. 630-634, 2019.

[21] V. Chernykh and P. Prikhodko, Emotion recognition from speech with recurrent neural networks, arXiv preprint arXiv:1701.08071, 2017.

[22] O. Krestinskaya and A. P. James, Facial emotion recognition using min-max similarity classifier, International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE., pp. 752-758, 2017.

[23] A. T. Lopes, E. De Aguiar, A. F. De Souza and T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern recognition, vol. 61, pp. 610-628, 2017.

[24] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali and M. Zareapoor, Hybrid deep neural networks for face emotion recognition, Pattern Recognition Letters, vol. 115, pp. 101-106, 2018.

[25] S. Kumar, S. Singh and J. Kumar, Automatic live facial expression detection using genetic algorithm with haar wavelet features and SVM, Wireless Personal Communications, vol. 103, pp. 2435-2453, 2018.

[26] M. Murtaza, M. Sharif, M. AbdullahYasmin and T. Ahmad, Facial expression detection using six facial expressions hexagon (SFEH) model, 9th annual computing and communication workshop and conference (CCWC).IEEE, pp. 0190-0195, 2019.

[27] R. A. Virrey, C. D. S. Liyanage, M. I. B. P. H. Petra and P. E. Abas, Visual data of facial expressions for automatic pain detection, Journal of Visual Communication and Image Representation, vol. 61, pp. 209-217, 2019.

[28] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, Spatial–temporal recurrent neural network for emotion recognition, IEEE transactions on cybernetics, vol. 49, no. 3, pp. 839-847, 2018.

[29] P. Khorrami, T. Le Paine, K. Brady, C. Dagli and T. S. Huang, how deep neural networks can improve emotion recognition on video data, In 2016 IEEE international conference on image processing (ICIP). IEEE., pp. 619-623, 2016.

[30] M. Liu, R. Wang, S. Li, S. Shan, Z. Huang and X. Chen, Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild, In Proceedings of the 16th International Conference on multimodal interaction, pp. 494-501, 2014.

[31] B.A. Wandell, Foundations of Vision, First ed., Sinauer Associates Inc, Sunderland, Mass, 1995.

[32] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba and J. Budynek, The Japanese female facial expression (JAFFE) database, In Proceedings of third international conference on automatic face and gesture recognition, pp. 14-16, 1998.