International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimal Selection of Features for Human Emotion Identification from Face Images

N. Geetha[1], E. S. Samundeeswari[2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Facial expressions play a powerful role in human communication, serving as a highly effective non-verbal means of conveying emotions during social interactions. While humans naturally excel at interpreting these emotions, teaching machines to recognize facial expressions is a daunting task. The objective of this study is to develop a system capable of replicating human visual perception by harnessing artificial intelligence techniques to analyze input images. Facial expression recognition systems are gaining widespread application in various domains like gaming and the internet, significantly enhancing the efficiency of robots in sectors including military, healthcare, and manufacturing. Nonetheless, the abundance of features in image descriptors poses a substantial challenge for facial emotion recognition systems. Despite numerous attempts to simplify feature complexity, the intricate and diverse nature of facial expressions makes the selection of discriminative features a complex undertaking. In this paper, proposed an effective feature selection method designed to pinpoint and choose informative features from high-dimensional data, with the explicit goal of optimizing classification accuracy. Our approach leverages Particle Swarm Optimization (PSO) to identify valuable feature combinations for classification, using the accuracy determined by the K - Nearest Neighbour (KNN) and Linear Discriminant Analysis classifier (LDA) to assess fitness within the PSO algorithm.

Keywords: *Facial Emotion Recognition, Optimization Algorithms, Feature Reduction, Particle Swarm Optimization, Genetic Algorithms*.

1. Introduction

Emotion holds significant sway in our everyday lives, influencing various facets such as learning, memory, and decision-making. The ability to discern diverse emotional states carries extensive implications across domains like distance education, healthcare, assisted driving, and human-computer interaction. Consequently, this topic has garnered substantial attention from researchers in recent years, emerging as a prominent research area [1-2]. However, emotion recognition is still a challenging task because the duration of emotion varies and the ways in which different individuals express and perceive emotions vary. Effectively improving the robustness and accuracy of emotion recognition is the goal pursued by researchers [3].

Emotion is generally understood as a psychological and physiological reaction associated with a specific environmental event, with transient, intense, and situational characteristics [2]. Compared with emotion, emotion is considered to be a more stable and continuous response characteristic, and it's not restricted by the relevance of specific events, and can reflect things more comprehensively [3]. This paper focuses on the emotional recognition with situational and transient, in order to provide relevant theoretical basis for the user's emotional perception in the home environment.

[1]*Research Scholar, Department of Computer Science, Vellalar College for Women, Tamilnadu, India*
[2]*Associate Professor, Department of Computer Science, Vellalar College for Women, Tamilnadu, India*

Currently, emotion recognition methods in academia primarily fall into two categories: self-assessment through questionnaires and equipment-based evaluation using human body parameters. Emotion recognition through questionnaires heavily relies on the subjective judgments of individuals, resulting in relatively lower accuracy and reliability. On the other hand, the equipment-based evaluation method utilizing human body parameters has gained prominence in the field of emotion recognition. The outcomes of emotion computing based on this approach find widespread applications in human-computer interaction and robotics. Presently, research in the realm of emotion recognition and algorithm development primarily revolves around facial expressions, speech, text, and human body gestures. Both single-modal recognition algorithms and multi-modal comprehensive recognition algorithms are under exploration [4]. Among them, speech emotion recognition relies on speech recording data, The requirements for data clarity and noise reduction are high; facial emotion recognition relies on high-quality facial image or video data, although its emotion recognition accuracy and robustness are high, but there are problems such as light sensitivity and privacy invasion; Gesture emotion recognition generally uses wearable devices to collect emotional characteristics of different stances to automatically detect and perceive emotions. Capacitive sensors are relatively low-cost, low-power consumption, high convenience and high applicability, and wearable devices in contrast, it is less intrusive, hardly affects the user's daily behaviour, and can conveniently collect the user's natural behaviour gait. The capacity to recognize and interpret human emotions stands as a crucial element in effective human interaction and communication. Emotions are an intrinsic facet of the human experience, and accurately discerning them from facial expressions holds immense significance across numerous domains, including psychology, healthcare, human-computer interaction, and artificial intelligence. Recent years have witnessed the fusion of computer vision and machine learning, offering new avenues for automating human emotion identification from facial images. At the heart of these automated systems lies a pivotal factor: the selection of pertinent and distinguishing features from the intricate and nuanced data encoded in facial expressions [5].

This research work embarks on a comprehensive exploration of the domain of optimal feature selection for the recognition of human emotions from facial images. Particular attention is devoted to the utilization of Particle Swarm Optimization (PSO) as a potent tool for this purpose. Feature selection, a decisive phase in the design of emotion recognition systems, entails the judicious selection of a subset of informative features from the multifaceted realm of facial image data. The overarching objective is to elevate the precision, efficiency, and resilience of emotion recognition models by identifying and capitalizing on the most salient and distinctive facial cues, while simultaneously mitigating the influence of extraneous information and noise.

The incorporation of PSO, an optimization algorithm inspired by the collective behaviors observed in natural swarms and groups, introduces a novel and promising dimension to the feature selection process. Initially inspired by the social dynamics of birds and fish, PSO has demonstrated notable efficacy in optimizing intricate and nonlinear challenges. By harnessing PSO's inherent parallelism and adaptability, researchers have achieved significant advancements in feature selection across a diverse spectrum of applications.

In the context of this research work, we embark on a deep dive into the intricacies of PSO-based feature selection for human emotion identification from facial images. We delve into the theoretical foundations of PSO and its tailored application to the feature selection task. Furthermore, we undertake an in-depth exploration of the impact of PSO on the precision and efficiency of emotion recognition models, shedding light on its potential for practical, real-world implementation. Through an exhaustive review of existing research and recent breakthroughs, our goal is to provide invaluable insights and guidance to researchers and practitioners striving to unlock the full potential of PSO in the realm of emotion recognition.

The remaining part of the paper is organised as follows. Section discusses the introduction of the features selection and face recognition system. Section 2 detailed the relevant literatures with work citations. Proposed feature selection method is explained in section 3 along with its flowchart and system architecture. Section 4 provides experimental results and discussion and concludes the paper in section 5 along with quoted references.

2. Literature Survey

Qian et al [6] incorporated an attention module into the ResNet model, significantly enhancing its performance. When evaluated on the CK+ dataset, their algorithm exhibits a lower recognition rate compared to the method outlined in existing literature, yet it employs fewer model parameters. In contrast to other algorithms, this model boasts a reduced parameter count while still achieving improved recognition rates, offering distinct advantages in overall performance.

According to the Ding et al [7] comparison results, the network branch added to the LBP image can provide more discriminative facial feature information, and the fusion of the two networks extracts the high-level feature information can generate a large amount of rich feature information that is helpful for expression recognition for the final expression classification and recognition, and the attention mechanism can strengthen the weight of effective information and reduce the influence of feature information that is weakly related to the current category, thereby improve the performance of lightweight network to a certain extent.

Wang et al. [8] segmented three partial images related to expression changes from the input face image, left eye, mouth and nose, and used three double-branch neural networks to extract the local expression features and features of different facial parts. The global expression features of the original image are combined with the prediction results of the three sub-networks to make the final expression prediction, and the recognition rate of the algorithm is improved by adding the discriminative features of different local face regions. High-level features are usually highly correlated with the final expression category, and most algorithms.

Li et al [9] shows that the local texture features extracted by the LBP operator can capture the small change information of the face, help to distinguish the subtle differences between different expressions, and improve the recognition rate of neural networks for expressions. The feature extraction ability of the lightweight convolutional neural network is weak, and it is difficult to extract more discriminative features compared with the general convolutional neural network, while the face image processed by LBP can reduce the interference of irrelevant information such as the background, focus more on information on key areas such as facial features. Aiming at the low performance of lightweight networks in expression recognition tasks, this paper proposes an expression recognition algorithm based on multi-feature fusion, which combines the features of LBP images and the middle-level local features and high-level global features extracted by neural networks. On the basis of the neural network, more feature information that helps to recognize expressions is captured, and the improved channel attention module is used to screen the fused feature information, while retaining the original feature information, it can more effectively strengthen the more discriminative, it can reduce the influence of weakly correlated feature information, so as to improve the recognition accuracy of the algorithm to a certain extent while maintaining a low number of network parameters.

Tonguc et al [10] said that, facial expression is the main way for human beings to express their emotions, and the research on facial expression began in the field of psychology. Psychologist MEHRABIAN research believes that in all the information people want to express, facial expressions account for 55%. In 1971, American psychologist EKMAN divided facial expressions into happiness, anger, surprise, fear, disgust and sadness. In subsequent studies, pain, neutrality, etc. were added. Nowadays, most of the explorations for facial expression recognition are based on these expressions.Gan et al. [11] proposed a new multi-attention network to simulate human's coarse and fine visual attention to improve expression recognition performance in order to extract recognition features from key facial regions; two networks were defined, from coarse to fine-grained levels binary masks are learned to localize key regions associated with salience. Although the above methods have improved the accuracy of facial expression recognition in complex environments, while solving complex situations, the complexity of the network model will inevitably increase significantly, the amount of network calculations will also increase, and the network will also require a lot of memory space and time when running. Computing resources cannot be deployed on resource-constrained devices.

Yin et al. [12] carried out dimension reduction processing on the model and embedded attention mechanism, which improved the feature extraction ability of the model and reduced the complexity of the model. In view of the above problems, this paper proposes a lightweight facial expression recognition model based on multi-region fusion. Through two branches, fine-grained feature extraction is performed on the local details and the global face image respectively; at the same time, an attention map is proposed in the global branch. To adjust the feature weights and generate masks with key points to assist in adjusting the attention map. Through the pruning strategy, the overall model is optimized to improve the recognition accuracy and speed.

Bhattacharya et al. [13] proposed the STEP classifier, which is based on the spatio-temporal graph convolution algorithm, adding a new regularization loss function, which improves the accuracy of video gait emotion recognition to 88%, and released the E-Gait Dataset. In addition to the above network, recent popular deep learning models such as Transformer and Rocket have also been used in similar time series data detection. The algorithm performs well in its corresponding tasks, but there are still some problems, if the analyzed data is usually direct walking images or skeletal points, rather than deeper indirect information like capacitive floor, the model may not be able to extract the deep connection between capacitive floor data and user emotions.

It can be seen from the table given by Guo et al [14] that the classification accuracy of the proposed method is higher than that of literature using the same data set. The classification accuracy of the method is better than that of the proposed method, but it uses facial video (or image) data. Its advantage is that it can obtain sufficient facial behavior characteristics for depression recognition, but its disadvantages may lead to problems such as privacy disclosure of subjects and high computational cost. The proposed method can obtain better MDD recognition results under the premise of protecting the privacy of the subjects.

Table 1. Methodologies Proposed and Limitations of Various

SNo	Author Name	Pros	Limitations
1	Qian et al [6]	- Added an attention module to the ResNet model, improving model performance. - Lower number of parameters compared to some algorithms. - Improved recognition rate compared to other algorithms. - Advantages in overall performance.	-Recognition rate on CK+ dataset is lower than some other algorithms. - Lack of detailed quantitative performance metrics.

2	Ding et al [7]	- The network branch added to the LBP image provides more discriminative facial feature information. - Fusion of networks extracts high-level feature information. - Attention mechanism strengthens the weight of effective information. - Improves performance of lightweight network.	- Lack of clarity on the extent of performance improvement.
3	Wang et al. [8]	- Segmentation of facial parts and use of double-branch neural networks. - Improved recognition rate through discriminative features of different local face regions. - Combines local and global features for better prediction.	- Limited discussion on potential trade-offs.
4	Li et al [9]	- LBP operator captures small change information and improves recognition rate. - LBP reduces interference of irrelevant information. - Multi-feature fusion enhances recognition. - Improved channel attention module for better feature selection.	- Weak feature extraction ability of lightweight CNN not quantified. - Potential challenges in feature fusion not addressed.
5	Tonguc et al [10]	- Historical context of facial expression research provided. - Mention of the importance of facial expressions for emotions.	- Lack of discussion on recent advancements in the field.
6	Gan et al. [11]	- Proposes a multi-attention network for expression recognition. - Attempts to simulate human coarse and fine visual attention.	- Mention of increased complexity and resource requirements but no details.
7	Yin et al [12]	- Dimension reduction and embedded attention mechanism for improved feature extraction. - Proposes a lightweight model based on multi-region fusion. - Fine-grained feature extraction on local details and global face image. - Attention map for feature weight adjustment.	- Lack of discussion on potential model optimization challenges.
8	Bhattacharya et al. [13]	- Introduces the STEP classifier based on spatio-temporal graph convolution for gait emotion recognition. - Claims improved accuracy and releases a dataset. - Mentions the use of popular deep learning models.	- Lack of information on dataset limitations.
9	Guo et al [14]	- Higher classification accuracy compared to some methods on the same dataset. - Protects subjects' privacy.	- Lack of information on scalability to other datasets.

2.1 Problem Statement

The primary goal of this literature review is to explore and assess various strategies, methods, and models proposed within the realm of facial expression recognition. The recognition of facial expressions holds significant importance in computer vision and emotion analysis, with practical applications spanning human-computer interaction, psychology, and affective computing. The central challenge lies in devising efficient and precise algorithms and models capable of autonomously detecting and categorizing facial expressions from images or videos, thus facilitating machines in comprehending human emotions.

The objectives of this survey encompass the following:

1. Identify and succinctly summarize the noteworthy contributions and advancements in facial expression recognition.

2. Assess the strengths and weaknesses inherent in the diverse array of approaches and models introduced by researchers.

3. Shine a spotlight on the shared difficulties and constraints encountered within the field, including issues pertaining to accuracy, computational complexity, and privacy considerations.

4. Offer insights into the techniques deployed to enhance recognition rates, streamline model intricacy, and elevate the overall efficacy of facial expression recognition systems.

5. Delve into the integration of attention mechanisms, feature extraction methodologies, and the fusion of multi-modal data within the context of facial expression recognition.

6. Engage in a discourse concerning the potential practical applications of facial expression recognition technology and its ramifications for real-world scenarios.

Through this survey, the objective is to cultivate a comprehensive grasp of the cutting-edge developments in facial expression recognition, with the ultimate aim of steering forthcoming research endeavors and contributing to the evolution of more resilient and proficient facial expression recognition systems.

3. Facial Emotion Recognition System

The system is structured around three core phases: feature extraction, feature selection, and feature classification. These fundamental stages, as illustrated in Figure 1 below, are preceded by an initial process involving face detection and preprocessing. In the initial stage, referred to as feature extraction, pertinent information is collected from the input data. This process involves the identification and extraction of essential attributes or characteristics from the raw data.

Equations

Following the completion of the feature extraction process, the subsequent phase entails feature selection. During this stage, a careful selection of the most relevant features is carried out from the previously extracted set. The primary goal in this phase is to reduce dimensionality and enhance computational efficiency. Finally, the ultimate phase, which is feature classification, is entered into. In this phase, the selected features are utilized in sorting or categorizing the data into distinct groups or classes, based on predetermined criteria. This systematic approach, which encompasses the aforementioned phases as well as the crucial initial steps of face detection and preprocessing, serves as the foundational framework for the entire process. It plays a pivotal role in facilitating effective and precise data analysis and decision-making.

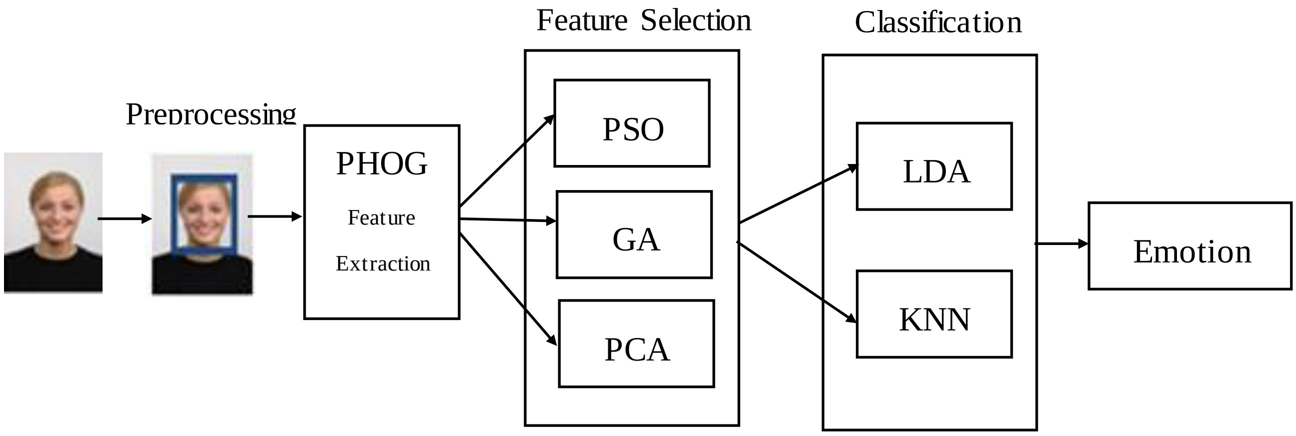


Figure 1. Overall Block Diagram for Face Emotion Recognition

3.1 Pre-processing

The system is organized into three fundamental phases: feature extraction, feature selection, and feature classification. These crucial stages, as illustrated in Figure 1, are preceded by an initial step involving face detection and data preprocessing. In the initial phase, known as feature extraction, the system collects relevant information from the input data by identifying and extracting essential attributes or characteristics from the raw dataset. Subsequently, in the following phase, which is referred to as feature selection, the system meticulously identifies and chooses the most significant features from the extracted set. The primary objective here is to reduce dimensionality and optimize computational efficiency. Finally, the system progresses to the feature classification phase, where the selected features are employed to categorize the data into distinct classes or categories based on predefined criteria. This systematic approach, encompassing these critical phases in conjunction with the essential initial steps of face detection and preprocessing, establishes the fundamental framework of the entire process, enabling effective and precise data analysis and decision-making.

3.2 Feature Extraction

The Pyramid Histogram of Oriented Gradient (PHOG) is a feature extraction technique that is characterized by similarities to the Histogram of Oriented Gradient (HOG) approach. However, the introduction of a unique structural element known as a pyramid in PHOG allows for the organization of feature extraction in a hierarchical manner. After normalizing the input image, it is divided into smaller cells, and for each pixel within these cells, a gradient histogram is computed. These histograms are subsequently concatenated to create the feature descriptor. It is worth emphasizing that, for this study, we have adopted the PHOG features as detailed by Bosch et al [13]. The specific parameters chosen for PHOG encompass three hierarchical levels, a 360-degree angle range, and 16 bins for the histogram. This configuration enables us to efficiently capture and represent essential features from the input images, thus facilitating robust data analysis and decision-making processes.

3.3 Feature Selection using Particle Swarm Optimization

In lieu of conventional feature selection techniques, we advocate for the adoption of a bio-inspired approach—Particle Swarm Optimization (PSO). PSO offers a unique and nature-inspired method for selecting features, diverging from the more traditional approaches typically employed in this context. This choice aligns with our aim to harness the collective intelligence of a particle swarm to efficiently optimize and select the most relevant features, enhancing the overall effectiveness and performance of our feature selection process.

At present, the main feature selection strategies are wrapper methods. Among them, the Filter strategy does not rely on the recognition model, but analyzes the correlation coefficient, mutual information and information entropy between features. The disadvantage is that it is separated from the subsequent classification algorithm, resulting in the accuracy of the results cannot be guaranteed when the combination of features screened out is used for classification recognition. In contrast, the Wrapper strategy used in this work directly takes the correct classification rate of the classifier as the evaluation index, and then combines the particle swarm algorithm for heuristic search, which can not only ensure the accuracy of the final result, but also take into account the search efficiency. The algorithm design idea is as follows:

1) Feature extraction is performed using the PHOG technique. The data is then normalized, and an appropriate number of samples is randomly selected.

2) The particle swarm is initialized, and the dot product of individual particles with the feature vectors in the sample within the particle swarm is calculated to create a feature subset. This subset is randomly split into a training set and a test set.

3) The classifier is trained using the training set, and the test set data is fed into the classifier to calculate the correct classification rate, which serves as the fitness value.

4) Individual particles and group particles are calculated.

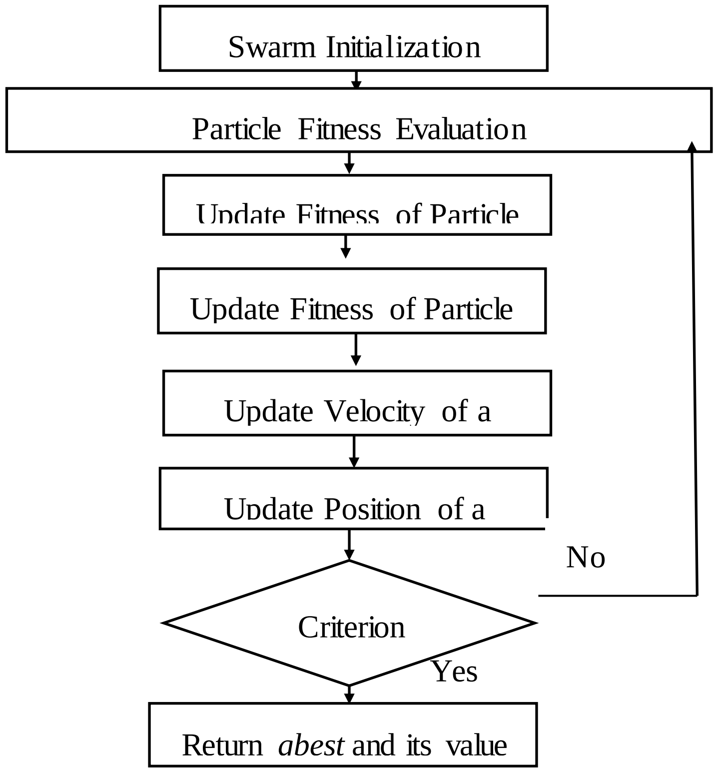
5) Particle velocity and position are updated, along with the fitness value.

6) A comparison is made between the current and previous generation of individual particles. If specific discriminant conditions are met, the individual and group particles are updated, and the particle velocity and position that led to the maximum performance are saved.

7) If the maximum number of iterations has not been reached, the process returns to step 3. Otherwise, the search terminates.

8) The optimal solution is then output.

The particle swarm optimization techniques flow chart is shown in Figure 2.



In our approach, the evaluation of fitness relies on the accuracy of the classifier. Specifically, we have chosen Linear Discriminant Analysis (LDA) as our classifier. To assess the fitness of each chromosome, we employ the K-fold cross-validation technique, which comprises several essential steps. Initially, the training dataset is divided into K subsets. Each of these subsets is then tested using the remaining (K-1) subsets for training. The accuracy is subsequently determined by calculating the average of the results obtained across all subsets.

For this research, we have adopted a 10-fold cross-validation strategy, which entails splitting the data into ten distinct subsets to conduct a comprehensive evaluation. This approach ensures a robust and thorough assessment of our model's performance, effectively measuring its ability to generalize and make accurate predictions.

3.4 Classification

3.4.1 Linear Discriminant Analysis

In the realm of image processing research, Linear Discriminant Analysis (LDA) has demonstrated its effectiveness as a classifier in a wide range of applications, with face recognition being a notable example, as demonstrated by Belhumeur et al [14]. The core objective of LDA is to enhance class separation by maximizing the distance between classes while minimizing the distance within each class. The success of LDA can be attributed to its utilization of second-order moments, specifically covariance and mean, from the class distribution. This simplified approach focuses on essential statistical characteristics, making it a powerful tool in various domains, particularly in tasks like face recognition.

3.4.2The K Nearest Neighbors

Types of Graphics

K-Nearest Neighbors (KNN) is widely recognized as one of the most frequently utilized classification algorithms. Its fundamental principle involves assigning unlabeled samples to categories by evaluating their similarity to the samples within the training dataset [15]. Thus, when dealing with unclassified samples, KNN aims to identify the K nearest neighbors among the categorized samples and allocate them to the nearest group. It's important to note that KNN relies on a distance metric to quantify similarity, and in our specific study, we have opted to utilize the cosine distance measure. This particular metric assists in quantifying the proximity between data points, enabling accurate and effective classification based on the similarity of feature spaces.

4. Experimental Results and Discussion

The experiments were conducted using the Extended Cohn-Kanade (CK+) [16].The dataset encompasses 593 video sequences featuring a total of 123 distinct individuals, spanning an age range from 18 to 50 years, encompassing diverse genders and ethnic backgrounds. Each video chronicles the transformation of a neutral facial expression into a specific targeted peak expression. These recordings are conducted at a frame rate of 30 Frames Per

Second (FPS) and exhibit a resolution of either 640x490 or 640x480 pixels. Among these video sequences, 327 are annotated with one of seven predefined expression categories: anger, contempt, disgust, fear, happiness, sadness, and surprise. The CK+ database stands as the most widely employed laboratory-controlled repository for facial expression classification, serving as a fundamental resource in the development of numerous facial expression classification methodologies. The dataset effectively captures a wide spectrum of emotions, with each sample depicting eight distinct facial expressions and three unique gaze directions, as visually represented in Figure 5. These emotional states include neutrality, anger, fear, disgust, happiness, contempt, surprise, and sadness. The database's adherence to the Facial Action Coding System (FACS) standard codes for expressing emotions renders it highly suitable for research in the field of facial emotion recognition.

The dataset consists of 981 images, which are categorized into various emotional expressions. Anger is depicted in 135 images, Contempt in 54, Disgust in 177, Fear in 75, and Happiness is represented by 207 images. Furthermore, Sadness is expressed in 84 images, and Surprise is illustrated through 249 images. These images collectively constitute a rich dataset for the study of a wide range of emotional states. It's noteworthy that these test images were entirely novel and had not been utilized during the training phase. As depicted in Figure 4, our study focused on seven core emotions: fear, anger, happiness, surprise, disgust, contempt and sadness. This selection allowed us to conduct a comprehensive investigation and assessment of the performance of facial emotion recognition.

Figure 3. Few Sample from CK+Database Shows Seven Basic Emotional Expressions

Figure 4 provides a visual representation of the impact of employing feature selection techniques such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA). Notably, these methods have led to a significant reduction in the number of features. Initially, there were 1360 features, but through the application of PSO and GA, this number has decreased to 750 and 800 features, respectively. This translates to a reduction rate of 44.85% and 41.18%. Such a substantial decrease in feature dimensions underscores the efficacy of these techniques in streamlining and optimizing the feature space, ultimately enhancing the efficiency and performance of the system.

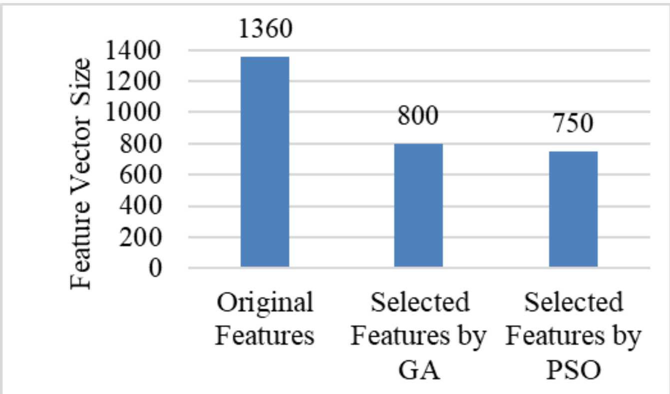


Figure 4. Comparison in Feature Reduction by PSO and GA

When evaluating feature sets generated by Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), a noteworthy disparity becomes evident. Precisely, the feature set obtained via PSO demonstrates a significant advantage, producing results that are six times more favorable in comparison to the feature set derived through GA. This discrepancy underscores the remarkable effectiveness and efficiency of PSO in the domain of feature selection, resulting in substantially enhanced performance and outcomes in this comparative analysis.

4.1 The Overall Accuracy

In Figure 5, we can observe the depiction of the average recognition rates achieved in our study. What's noteworthy is that feature selection driven by Particle Swarm Optimization (PSO) has demonstrated superior performance compared to Genetic Algorithms (GA) and other methods that rely on Principal Component Analysis (PCA) for feature selection. This outcome highlights the effectiveness of PSO in optimizing feature selection, leading to improved recognition rates when compared to GA and PCA-based approaches.

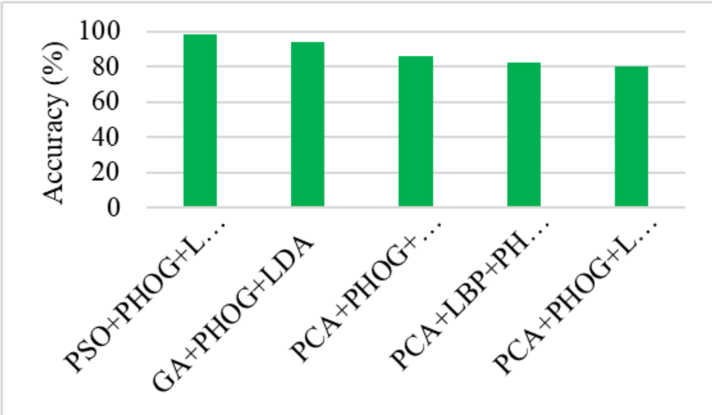


Figure 5. The Overall Accuracy improvement of Feature Selection Methods

From the Figure 5, accuracy of PSO is 4.25% better than GA+PHOG+LDA, and it has 13.95% higher accuracy compare with PCA+PHOG+KNN, PSO has 19.51% higher accuracy compare with PCA+LBP+PHOG+LDA and PSO has 22.5% better than PCA+PHOG+LDA .

4.2 The Accuracy Over Seven Basic Emotions

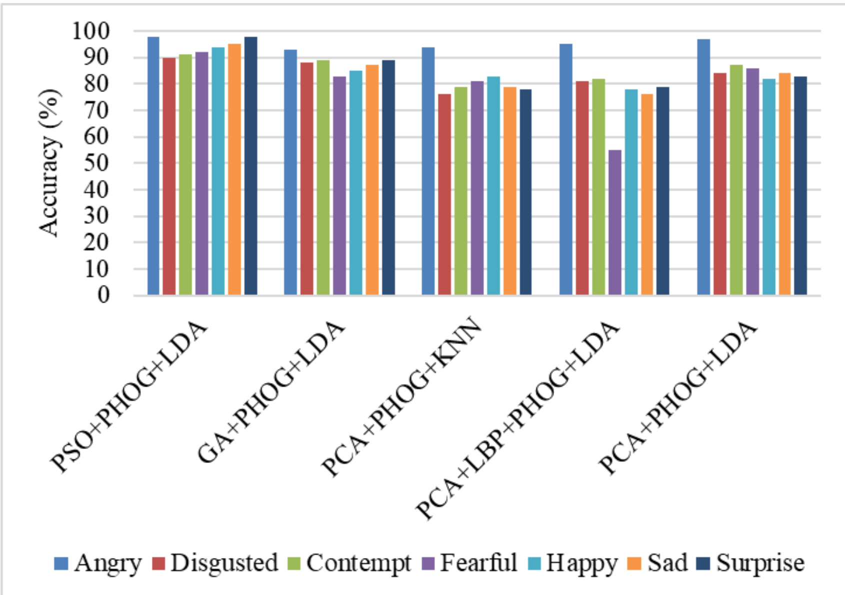


Figure 6.Comparison of Recognition Rate between Various Feature Selection

From the Figure 6 the angry recognition of PSO is 5.37% better than GA+PHOG+LDA and 4.25% better than PCA+PHOG+KNN, in Disgusted recognition of PSO is 11.11% better than PCA+LBP+PHOG+LDA and 7.14% better than PCA+PHOG+LDA at same time fearful recognition of the PSO is 6.97% higher than PCA+PHOG+LDA and 67.27% higher than PCA+LBP+PHOG+LDA and the happy recognition of

PSO 13.25% higher than PCA+PHOG+KNN,10.58% higher than GA+PHOG+LDA

In sad recognition of the PSO is 13.09% and 25% higher than PCA+PHOG+LDA and PCA+LBP+PHOG+LDA respectively. Surprise recognition of PSO is 10.11% and 25.64% better than GA+PHOG+LDA and PCA+PHOG+KNN respectively.

Assessed the classification performance by computing several performance metrics, including the True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR). The corresponding results are presented in Table 2.

Table 2. TPR, TNR, FPR and FNR Analysis

Methods		Angry	Disgusted	Contempt	Fearful	Happy	Sad	Surprised
TPR (%)	**PSO+PHOG+LDA**	**100**	**100**	**100**	**100**	**99**	**100**	**100**
	GA+PHOG+LDA	98	100	95	96	100	98	99
	PCA+PHOG+KNN	88	92	90	90	88	90	92
	PCA+LBP+PHOG+LDA	94	96	95	90	92	90	86
	PCA+PHOG+LDA	92	92	93	94	90	92	90
TNR (%)	**PSO+PHOG+LDA**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	GA+PHOG+LDA	99	99	99	98	98	96	98
	PCA+PHOG+KNN	86	88	87	86	78	80	75
	PCA+LBP+PHOG+LDA	96	94	92	92	90	90	92
	PCA+PHOG+LDA	90	92	93	86	88	92	80
FPR (%)	**PSO+PHOG+LDA**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	GA+PHOG+LDA	1	1	1	2	2	4	2
	PCA+PHOG+KNN	14	12	13	14	22	20	25
	PCA+LBP+PHOG+LDA	4	6	5	8	10	10	8
	PCA+PHOG+LDA	10	8	8	14	12	8	20
FNR (%)	**PSO+PHOG+LDA**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
	GA+PHOG+LDA	2	0	0	4	0	2	1
	PCA+PHOG+KNN	12	8	7	10	12	10	8
	PCA+LBP+PHOG+LDA	6	4	4	10	8	10	14
	PCA+PHOG+LDA	8	8	8	6	10	8	10

The tabulated results presented in Table 2 offer valuable insights into the performance of various feature selection and classification methods. The True Positive Rate (TPR), which reflects the accuracy of identifying positive cases, consistently demonstrates exceptional performance, nearly reaching 100% for most emotions in the PSO+PHOG+LDA method, with a 99% TPR for "Happy." The GA+PHOG+LDA method also performs admirably, achieving TPR values exceeding 96% for most emotions. In contrast, the PCA-based approaches exhibit relatively lower TPR values, particularly in recognizing "Happy" and "Sad."

When examining the True Negative Rate (TNR), which measures the ability to correctly identify negative cases, both the PSO+PHOG+LDA and GA+PHOG+LDA methods excel, with TNR values surpassing 98% for most emotions. Conversely, the PCA-based techniques tend to display lower TNR values, with PCA+PHOG+LDA showing limitations in distinguishing "Happy" and "Surprised."

The False Positive Rate (FPR), indicating the rate of erroneously classifying negative cases as positive, is notably low in both the PSO+PHOG+LDA and GA+PHOG+LDA methods, indicating strong specificity. Conversely, PCA+PHOG+KNN tends to yield higher FPR values, suggesting a relatively higher rate of false positive classifications.

The False Negative Rate (FNR), which measures the rate of erroneously classifying positive cases as negative, highlights the strength of the PSO+PHOG+LDA method, with consistently low FNR values, typically below 1%, except for "Happy" (1%). The PCA-based methods tend to yield higher FNR values, implying a higher likelihood of missing positive cases, particularly in the case of PCA+LBP+PHOG+LDA.

In summary, the PSO+PHOG+LDA method consistently exhibits remarkable classification performance, characterized by high TPR and TNR values, along with low FPR and FNR values across most emotions. The GA+PHOG+LDA method also delivers strong performance, while the PCA-based methods demonstrate varying performance levels across different emotions. These results underscore the effectiveness of PSO in feature selection and its positive impact on the recognition of emotions from facial images.

5. Conclusion

In this research paper, a fresh approach to feature selection is presented, which incorporates Particle Swarm Optimization (PSO) within a Wrapper framework. This distinct method contrasts with the Filter strategy that evaluates features independently from the recognition model. Instead, it directly integrates with the classifier by employing the classifier's correct classification rate as the assessment metric. This approach combines the merits of PSO for heuristic search, ensuring both precision and computational efficiency. The algorithm encompasses feature extraction, the initialization of PSO, fitness evaluation based on classifier performance, and a series of iterative updates. Experiments, carried out using the Extended Cohn-Kanade Database, encompassed 981 images depicting seven fundamental emotions with three different gaze directions. The utilization of PSO-driven feature selection yielded a substantial reduction in feature dimensionality, downsizing from 1360 to 750 features, marking a noteworthy 44.85% reduction. This underscores its potency in optimizing feature subsets for facial emotion recognition within the confines of a Wrapper-based methodology.

References

[1] Peng X L, Zhang H X, Hu Q Y., "Facial/paralysis expression recognition based on multitask learning of deep convolution neural network", Journal of Northwest University (Natural Science Edition), 2019, 49 (2): 187-192.

[2] Mumenthaler C, Sander D, Manstead A S R., "Emotion recognition in simulated social interactions", IEEE Transactions on Affective Computing, 2018, 11 (2): 308-312.

[3] Zhang T, Zheng W M, Cui Z, et al., "Spatial – temporal recurrent neural network for emotion recognition", IEEE Transactions on Cybernetics, 2019, 49 (3): 839-847.

[4] D'Mello S K, Kory J., "A review and meta analysis of multimodal affect detection systems", ACM Computing Surveys, 2015, 47 (3) 43.

[5] Poria S, Cambria E, Bajpai R, et al., "A review of affective computing : from unimodal analysis to multimodal fusion", Information Fusion, 2017, 37: 98-125.

[6] Qian Z, Mu J, Zhang J., "Facial expression recognition based on sasnet attention mechanism", International Conference on Computer Network, Electronic and Automation, Xian: IEEE, 2021: 159-163.

[7] Ding H, Zhou P, Chellappa R., "Occlusion-adaptive deep network for robust facial expression recognition", IEEE International Joint Conference on Biometrics. Houston:IEEE, 2020: 1-9.

[8] Wang K, Peng X, Yang J, et al., "Region attention networks for pose and occlusion robust facial expression recognition", IEEE Transactions on Image Processing, 2020, 29: 4057-4069.

[9] Li J, Jin K, Zhou D, et al., "Attention mechanism-based CNN for facial expression recognition", Neurocomputing, 2020, 411: 340-350.

[10] Tonguc G, Ozayd N, Ozkara B., "Automatic recognition of student emotions from facial expressions during a lecture", Computers & Education, 2020, 148 : 103797.

[11] Gan Y L, Chen J Y, Yang Z K, et al., "Multiple attention network for facial expression recognition", IEEE Access, 2020, 8: 7383-7393.

[12] Yin P B, Pan W M, Zhang H J., " Lightweight facial expression recognition method based on convolutional attention", Laser & Optoelectronics Progress, 2021, 58 (12): 1210023.

[13] Bhattacharya U, Mittal T, Chandra R, et al., "STEP: spatial temporal graph convolutional networks for emotion perception from gaits", Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 1342-1350.

[14] Guo W T, Yang H W, Liu Z Y, et al., "Deep neural networks for depression recognition based on 2D and 3D facial expressions under emotional stimulus tasks", Frontiers in Neuroscience, 2021, 15:[2022-12-02]. https:doi. org/10.3389/fnins.2021.609760.

[15] Bosch, A. Zisserman, and X. Munoz, "Representing shape witha spatial pyramid kernel," ACM International Conference on Image and Video Retrieval, 2007.

[16] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Risherfaces: Recognition Using Class Specific Linear Projection," In ECCV (1) Conference Proceedings, pages 45-58, 1996.