Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 11s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Facial Expression Classification in The Wild by Traditional Machine Learning and Deep Learning

Deepa D. Mandave[1*], Lalit V. Patil[1]

[1]Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Vadgoan Bk.,Pune, S.P. Pune University

***Corresponding Author:**

Deepa D. Mandave

Email ID: deepa.rengade@gmail.com

Cite this paper as: Deepa D. Mandave, Lalit V. Patil, (2025) Facial Expression Classification in The Wild by Traditional Machine Learning and Deep Learning. *Journal of Neonatal Surgery*, 14 (11s), 1084-1100.

ABSTRACT

Analyzing facial expressions is a fundamental aspect of computer vision, with significant importance in applications such as human-machine communication and interaction. Recognizing expressions in uncontrolled environments (wild), with varied lighting, poses, and occlusions, is particularly challenging. This study compares traditional methods (PCA, LBP, LDA, and HOG) with deep learning technique (CNN) using CK+ (posed images) and FER2013 (candid images) datasets. Results show that CNN outperforms traditional methods, achieving 34% higher accuracy in uncontrolled conditions. While traditional approaches excel in controlled settings, deep learning proves more effective overall for FER in natural environments.

Keywords: *Facial expression classification, computer vision, deep learning, CNN, wild dataset, etc.*

1. INTRODUCTION

Facial expression classification is a growing field fuelled by developments in AI, machine learning, and imaging techniques[1]. It is widely applied in areas such as security, healthcare, smart living, and robotics. FER analyses facial expressions linked to muscle movements to identify emotions like happiness, sadness, anger, and fear. Its applications include sentiment detection, mental health treatment, stress management, gesture recognition, smart homes, and fatigue monitoring[2]. The complexity of automatic expression analysis makes it a crucial area of research across diverse domains[3]. Traditional approaches like LBP and LDA have been used for feature extraction in facial expression recognition. However, deep learning has developed as a commanding approach, with representations like CNNs and RNNs capable of simultaneous feature extraction and classification[4]. The FER2013 dataset is instrumental in transitioning FER research from controlled settings to real-world situations, allowing deep learning techniques to tackle the challenges of emotion recognition in diverse, unconstrained environments[5]. Amjad Rehman Khan [6] analysed foundational modules and trends in Facial Expression Recognition (FER), highlighting both classical ML (machine learning) and DL (deep learning) approaches. Li and Deng [7] reviewed advanced deep neural network architectures and training strategies for static and dynamic image inputs, discussing their benefits and restrictions. Ekundayo et al. [8] explored emotion intensity estimation, addressing challenges like inconsistent labelling and the integration of textual data with FER models.

Emotional intelligence combines psychology and technology to analyze sentiments, with facial expression analysis being a key method. Worldwide sentiments like happy, sad, anger, fear, disgust, and contempt are biologically ingrained and recognizable. Happiness often involves smiling, sadness reflects hurt or worry, and anger arises from pressure, potentially leading to aggression. Fear alerts to danger, while disgust is triggered by aversive stimuli [9]. Despite these associations, achieving accurate Facial Expression Recognition (FER) remains challenging due to factors like lighting, occlusion, and age-related variations in images.

Recent approaches to Facial Expression Recognition (FER) focus on two key components: feature extraction and classifier structure. Feature extraction often uses methods like PCA, LDA, and LBP for dimensionality reduction and attribute mining [10]. Hewa et al. [11] applied PCA with SVM for expression classification using the JAFFE dataset, while Verma et al.[12] combined PCA and LDA features with SVM and HMM classifiers, highlighting SVM's effectiveness. Jumani et al. [14] proposed FER-CNN and FER-HOGCNN models, with FER-CNN showing better performance due to limitations in dataset quality affecting FER-HOGCNN. Arora and Kumar [14] integrated PCA with particle swarm optimization to enhance precision. Kola and Samayamantula [15] developed a robust local binary model with variable windows, demonstrating noise resilience and recommending SVM for classification. These methods underline the diversity of FER techniques and the challenges posed by varying dataset quality and environmental factors.

Niu et al. [16] developed a facial feature recognition method combining LBP (Local Binary Pattern) and ORB (Oriented FAST and Rotated BRIEF) features, followed by Support Vector Machine (SVM) classification, showing effective and accurate results. Lakshmi and Ponnusamy [17] enhanced emotion recognition by improving HOG and LBP descriptors and using a deep group autoencoder for feature reduction, followed by multiclass SVM for classification. Ravi et al. [18] compared LBP and Convolutional Neural Networks (CNN) for FER, finding CNN achieved superior accuracy (97.32%) in recognition. Despite advancements in FER with techniques like LBP, PCA, HOG, and LDA, a gap remains in comparative analyses between traditional ML and DL methods across different datasets.

The goal of this article is to bridge the gap among traditional ML and DL methods in Facial Expression Recognition (FER), comparing their performance in both controlled and uncontrolled environments. It reviews both wild and posed datasets, highlighting their diverse characteristics and uses. The article provides a proportional analysis of outcomes from ML and DL systems to inform new researchers in the FER field. The structure includes: Section 2, which covers FER data and methods; Section 3, analyzing results and ongoing research; Section 4, discussing opportunities and challenges; and Section 5, concluding with key findings and future research directions.

2. MATERIALS AND METHODS

2.1 Dataset

This study compares feature extraction processes using two main datasets: the extended Cohn-Kanade (CK+) and FER2013. The CK+ dataset consists of 981 images collected under controlled conditions, with participants aged 18-50, 31% male and 69% female, who were instructed not to wear glasses, jewellery, or have beards. It includes seven expression categories(anger, disgust, fear, happiness, neutrality, sadness, and surprise), with images in grayscale at a resolution of 48×48 pixels [19]. In contrast, FER2013 contains 28,000 labelled images collected from the internet, intended for emotion recognition challenges [20]. These images, also in grayscale and sized at 48×48 pixels, depict seven expressions but present a more complex, uncontrolled environment compared to CK+. The lower resolution and grayscale format reduce the computational load, making training and inference faster and less resource-intensive [21].

The FER2013 dataset contains images with varying attributes such as eyeglasses, hand poses, and contrast variations, creating challenges in accurate cataloguing. FER2013 offers a larger image volume, beneficial for training, but its varied poses, lighting, backgrounds, occlusions, and accessories complicate labelling and categorization [22]. Examples illustrating emotions from each database are display in Figure 1. A more accurate and reliable dataset can lead to better facial expression recognition models [23].Researchers often use established datasets like FER2013 and CK+ to evaluate facial expression recognition (FER) algorithms [24]. Experiments conducted on these datasets contribute significantly to the development and enhancement of FER algorithms. While posed datasets like CK+ emphasize differences between expressions for easier classification, spontaneous datasets like FER2013 better reflect real-world scenarios [25]. The controlled conditions of CK+ such as the absence of glasses or beards, improves accuracy by reducing variability, lighting inconsistencies, and pose variation [26]. A predefined set of categories might be based on expressions that are more typical or recognizable in certain cultures [27].

Leveraging both datasets can help in building robust, unbiased, and generalizable models that perform well across a wide range of real-world conditions and demographic groups [28]. The models trained on such diverse data can generalize better and perform reliably in real-world scenarios [29]. By doing so, models can be trained to recognize facial expressions accurately across different ages and genders, improving their applicability in real-world scenarios [30].

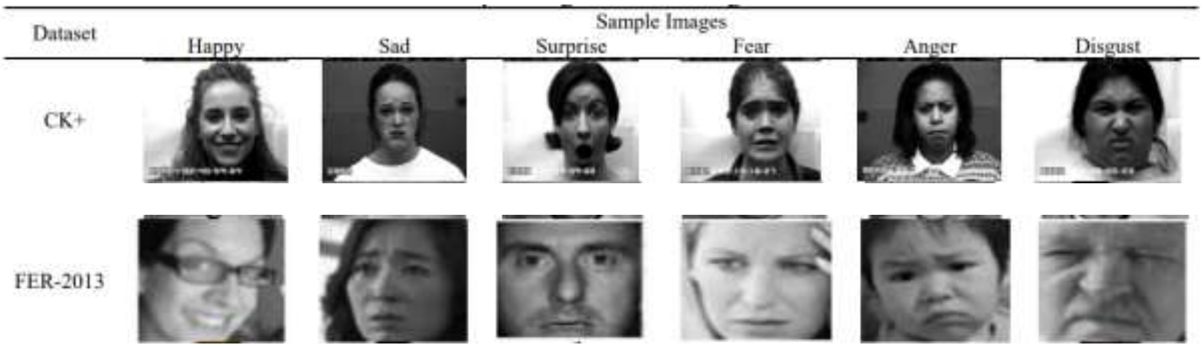


Fig.1 Samples of the Ck + and FER2013 datasets

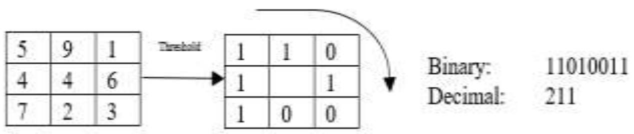
2.2 Feature Extraction

2.2.1. Traditional Feature Extraction Techniques

Feature extraction is an essential technique in the realm of computer vision, mainly in the area of FER. Features refer to patterns or structures within an image that aid in its classification. These features encompass attributes such as corners, edges, regions of interest points, ridges, etc. [31]. Given their widespread application in facial recognition tasks, algorithms such as PCA, LBP, LDA, and HOG are recognized as prominent feature extraction techniques. It serves as a foundational step by extracting discriminative attributes from facial images, which are essential for precise classification in FER systems.

a. Local Binary Patterns:

The LBP operator, primarily proposed by Ojala et al. [32], stands out as a widely utilized method for extracting local texture features from images. LBP is instrumental in delineating the texture properties of an image, typically by generating binary codes through comparisons of intensity values between a center pixel and its neighbouring pixels [33]. During this process, a binary digit (1) is assigned if the intensity of a neighbour surpasses or equals that of the center pixel; otherwise, a binary digit (0) is assigned. This binary sequence serves as the LBP code for the center pixel, effectively encoding the texture pattern within its local vicinity. This computation is iteratively performed for every pixel in the image, resulting in a comprehensive LBP representation of the entire image. Subsequently, this representation is harnessed to analyze and delineate the texture attributes of the image, thereby facilitating the identification of facial expressions. Nevertheless, as highlighted by Liao et al. [34], uniform patterns in Local Binary Patterns (LBP) may suffer from information loss within intricate images, despite their effectiveness in capturing fundamental patterns. To mitigate this limitation and bolster its robustness and discriminatory capabilities, several variants of LBP have been presented. For instance, Mistry et al. [35] presented an improved LBP variant that compares the horizontal and vertical neighbourhood pixels, thereby producing a more discriminative representation for facial emotion recognition. Shan et al. [36] observed that LBP features generated by a boosted LBP approach perform consistently and robustly, particularly over lower-resolution facial images captured in real-world environments, yielding promising results. The Local Binary Patterns (LBP) descriptor allocates a code to each pixel of an image by analyzing a 3×3 cell, comparing the values of the neighbouring pixels to that of the center pixel, and representing the results as a binary code, as depicted in Fig. 6. The histogram of these LBP labels computed over an area serves as a texture descriptor. This resulting binary information captures the characteristics of local regions containing curved edges, flat areas, spots, and more. To extract shape information from images, the images are separated into smaller sections to compute LBP histograms, as illustrated in Fig. 7. The benefits of the LBP operator contain its robust performance even under varying illumination conditions and computational easiness. Previous studies have demonstrated the efficacy and usefulness of the LBP algorithm in FER. However, one limitation of this approach is its inability to process dynamic or temporal information inherent in video sequences [37]. To overcome this limitation, Zhao et al. [38] introduced LBPTOP, a method that calculates LBP histograms from various orthogonal planes to improve expression recognition accuracy. LBPTOP then combines these LBP histograms to generate spatial-temporal motion LBP features for expression recognition. Additionally, Zhu et al. [39] proposed a FER algorithm that integrates CNN features with enhanced LBP characteristics, known as Fusion features. This method aims to leverage the strengths of both CNN and LBP methods for more accurate expression recognition.



Figures: - 2. The LBP operator [29]

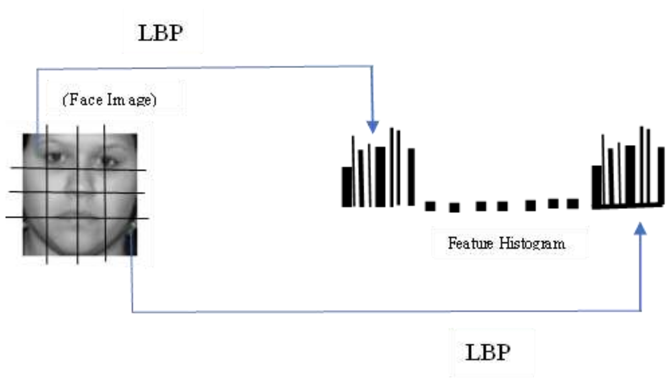


Figure: -3 A facial image is partitioned into distinct sections from which Local Binary Pattern (LBP) histograms are derived and then combined into a unified spatially enriched feature histogram

b. PCA (Principal Component Analysis):

PCA is a standard technique utilized for attribute mining and data visualization, commonly applied in pattern recognition and computer graphics. Its core aim is to decrease large-sized data into a small-sized feature space. PCA operates based on eigenfaces, enabling it to identify and classify faces by associating facial features with known individuals. This methodology views face recognition as a two-dimensional challenge, where features like mouth contours, eyebrows, and eyes are derived from eigenfaces and employed for expression recognition [40]. In the context of image representation, data is organized into matrices containing highly correlated information. PCA effectively operates on these correlated data matrices to extract relevant features. It achieves this by transforming the image matrix into a lower-dimensional Eigen subspace through various computational steps. Initially, the covariance matrix is computed using equation (1), representing the relative variance among the pixels in the image. Next, the eigen vectors of this covariance matrix, calculated using equation (2), yield the basis vectors, known as eigenfaces. These eigenfaces are selected based on their corresponding eigenvalues, with higher eigenvalues indicating greater significance. Consequently, the image is transformed into a feature space [41].

$$\mathrm{C} = E\{(X-\mu_x)(X-\mu_x)^T\} \quad (1)$$

$$\mathrm{C}_x\, e_i = \lambda_i e_i \qquad , i = 1,2,3, \dots, n \quad (2)$$

where, X is the 2D range, μ^x is row-wise average of X, C is the matrix of Covariance of X; e_i is Eigen vector of C, $\lambda_i$is Eigen code of C [41]. One of the significant advantages of PCA is its eigenface method, which enables the reduction of the dataset size required for detecting test images, thereby enhancing efficiency and computational performance.

c. LDA (Linear Discriminant Analysis):

LDA is a feature extraction scheme based on linear projection principles. Its goal is to identify projection vectors that maximize the distance between different categories while minimizing the variation within each category, thereby reducing the dimensionality of the feature space [42]. Unlike PCA, LDA emphasizes linear discriminatory criteria, making it highly effective in face expression recognition tasks. The main objective of LDA is to optimize the ratio between the determinant of the interclass scatter matrix and the determinant of the intraclass scatter matrix of the projected examples [43]. By doing so, LDA identifies a subspace of features that effectively discriminate between different classes of facial expressions. It achieves this by gathering similar class images closely together while creating distinct separations between images of different expression categories. In real-world applications, researchers such as Shih et al. [44] have integrated LDA with SVM for facial expression recognition. In this approach, each input test sample is compared with every training sample and classified based on its proximity to the nearest training sample. This fusion of LDA with SVM improves the precision and effectiveness of FER systems.

***d.* HOG (*Histogram Oriented Gradient*):**

The HOG algorithm extracts local image features from the Region of Interest (ROI) within an image, typically represented by small cells of 8 × 8 pixels [45]. This process involves calculating the distribution of local edge directions within these cells, as illustrated in Figure 4. The resulting signal encapsulates the gradient orientations present in the image region and is commonly employed for tasks such as edge detection and modelling of facial muscles.

To generate the HOG descriptor, the gradient information of pixels is aggregated within these small cells, forming one-dimensional histograms that capture the distribution of gradient orientations [46]. These histograms are subsequently concatenated to create feature vectors, which act as inputs for subsequent classification algorithms. This method enables the extraction of discriminative features from images, thereby facilitating tasks like facial expression recognition. Let L denote the intensity function (grayscale) defining the image under analysis. The image is partitioned into large N × N pixels (as illustrated in Figure 4), and the gradient direction θx,y at each pixel is computed using the following code:

$$\theta_{x,y} = \tan^{-1}\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \quad (1)$$

In the HOG algorithm, the orientation of each cell is computed and recorded in a histogram known as the M-box histogram. These histograms from all cells are then analysed and combined to generate a unified HOG histogram, marking the completion of the algorithm. This consolidated histogram represents the distribution of gradient orientations across the entire image and presents the final outcome of the HOG attribute extraction process.

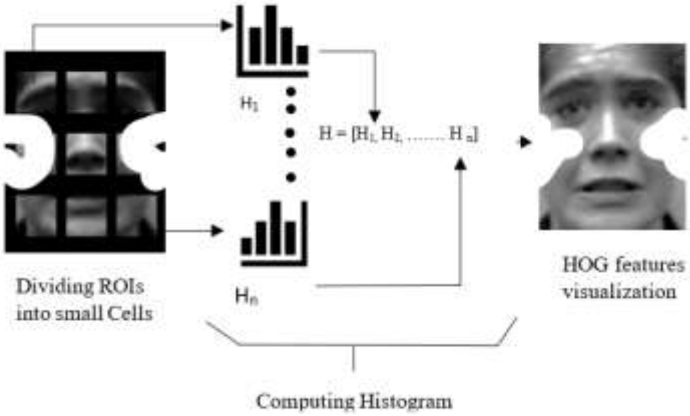


Fig. 4 Extracting HOG feature from ROI

2.2.2 CNN-based features

A typical neural network architecture consists of an input layer, multiple hidden layers, and an output layer. The hidden layers are generally categorized into three primary types: convolutional, pooling, and fully connected layers [47]. Convolutional layers accomplish mathematical operations on pixel values and kernel matrices to generate feature maps that capture essential patterns in the input. Pooling layers, meanwhile, decrease the size of the feature maps by selecting maximum or average values within specific regions, without introducing additional weights. Finally, the fully connected layers aggregate the features extracted from the previous layers to make classification decisions. These layers map low-level features, like edges, to higher-level attributes, such as textures, and continue to explore increasingly complex aspects of the input data. Over the years, various architectures and techniques have been developed, leading to a multitude of convolutional neural networks (CNNs). In this study, CNN was chosen, and different methodologies were employed to enhance classification accuracy. Essentially, different neural networks are capable of learning specific features of emotions, contributing to the overall effectiveness of the model [48]. Çayir et al. [49] introduced hybrid models that amalgamate CNN with traditional machine learning algorithms for classification, yielding improved results compared to traditional algorithms alone.

In this study, the CNN model has been comprised four learning layers. In the first convolutional layer, a 24x24 space is convoluted with 32 different kernels of size 3x3, with a stride of 1 pixel, resulted in 64 feature maps. The outcome of this layer feed as the input to the second convolutional layer, which convolutes the maximum pooled feature map from the preceding layer with 128 kernels of size 3x3, maintaining a stride of 1 pixel, resulting in 128 feature maps. Similarly, the third convolutional layer employs 256 kernels of size 3x3, connected to the maximum pooled feature map from the second convolutional layer. Subsequently, the fourth convolutional layer utilizes 512 kernels of size 3x3, connected to the maximum pooled feature map from the third convolutional layer.

Following the convolutional layers, the first fully connected layer comprises 4608 neurons, linked to the output map without any connections between the four convolutional layers. Finally, the last layer is a 7-softmax layer, responsible for estimating the probability of the 7 labels in the class. The implemented construction of the CNN is illustrated in Fig. 2. Throughout the training, the model learned iteratively, as proved by the decreasing loss and growing accuracy, as depicted in Fig. 3. This trend demonstrated that the model has effectively improving its performance over time.

Fig. 5. Proposed convolutional neural network structure

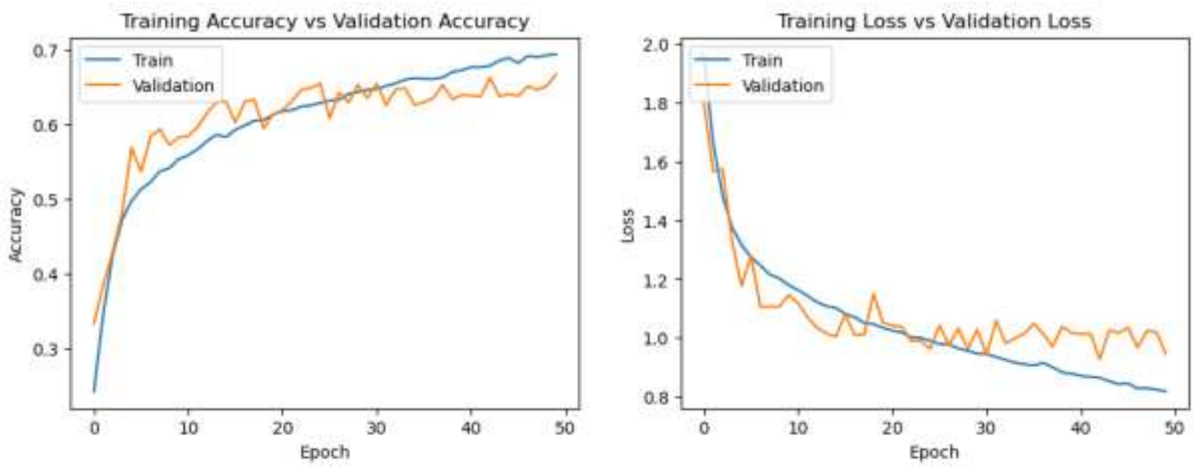


Fig. 6 Accuracy and loss of the CNN model with FER2013

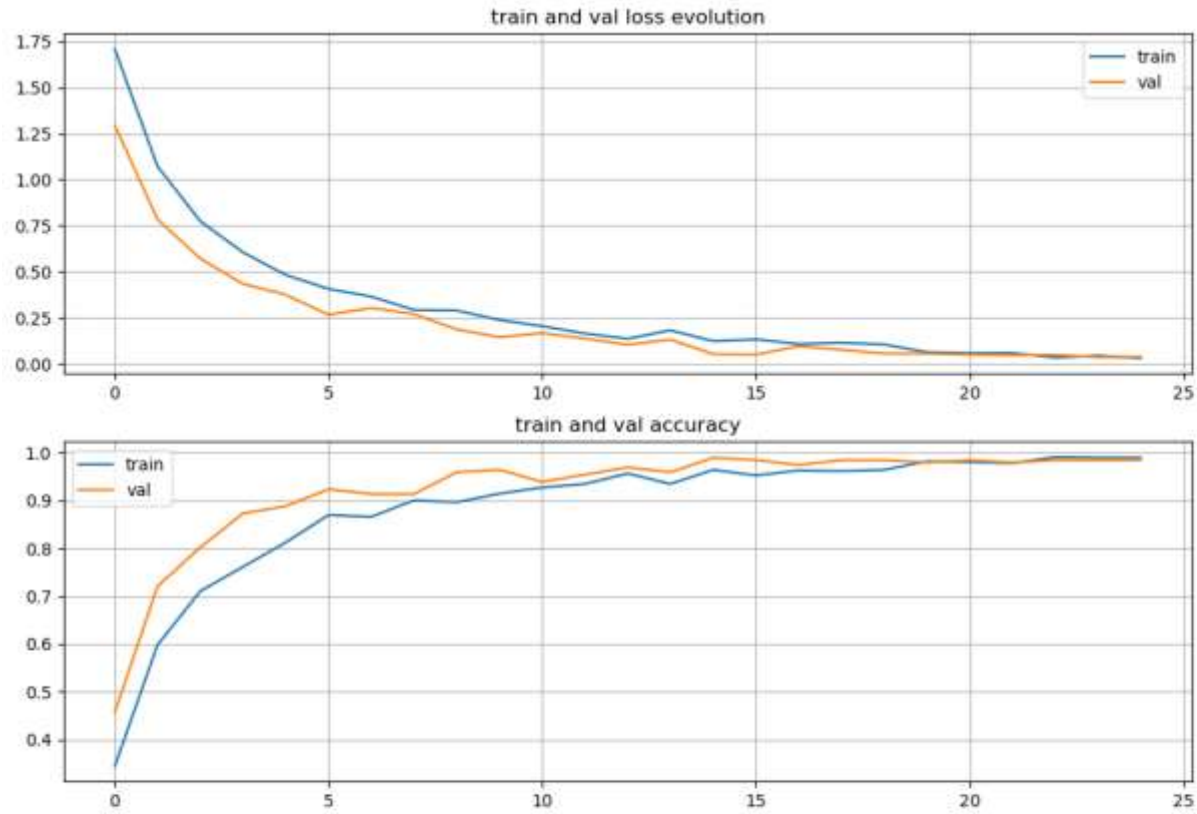


Fig. 7 Accuracy and loss of the CNN model with CK+

2.3 Data augmentation

Deep learning relies heavily on having a large amount of training dataset to accomplish optimum performance in FER tasks. However, many existing datasets for FER lack sufficient images for effective training. As a solution, DA (data augmentation) becomes indispensable in enhancing the performance of deep FER systems. DA techniques can be generally separated into two classes: dynamic and offline [50]. Dynamic DA encompasses the application of transformations to input images during the training phase. This often entails randomly cropping the four corners and the center of the input image and horizontally flipping them. Through dynamic DA, researchers can significantly augment the size of the training dataset, occasionally expanding it up to ten times its original size. In addition to dynamic DA, various offline DA approaches are utilized to further diversify and increase the train data. These techniques encompass a range of transformations such as rotation, translation, scaling, noise addition, contrast adjustment, and dithering. However, it's essential to acknowledge that generating and storing multiple augmented versions of the original images can significantly escalate the storage requirements and computational burden, potentially impacting the performance of the neural network. By leveraging both dynamic and offline DA techniques, researchers can substantially augment the variety and size of the train dataset. Consequently, this enhances the generalization and robustness of deep facial expression recognition models [51].

2.4 Model classifier

Supervised ML comprises two primary classes: traditional (non-CNN) classification methods and neural networks (NNs). Traditional classifiers like SVM, LDA, k-NN, and naive Bayes have a long-standing history of usage and their efficacy has been extensively scrutinized across various classification tasks. In the context of this study, four distinct feature extraction algorithms were employed to evaluate classification accuracy, with SVM being selected as the classifier. SVM, employing a "one-to-one" strategy, was chosen to represent traditional methods due to its well-documented performance and efficiency [52]. SVM operates as a supervised learning process that purposes to discover a hyperplane that maximizes the margin between two support vectors, effectively delineating the data points belonging to different classes.

2.5 Traditional Vs deep learning

Deep learning, a specialized area within machine learning, employs neural networks to emulate the brain's learning processes. It consists of algorithms organized in layers of increasing complexity and abstraction. Each layer transforms its input non-linearly and uses the learned information to build a statistical model. This process is repeated iteratively until achieving a desired level of accuracy. A notable feature of deep learning is its scalability. In the realm of computer vision, CNNs are widely used. Originating from ANN (Artificial Neural Networks) and announced in 1998, CNNs excel in supervised classification tasks, including real-time facial emotion recognition (FER). These networks are particularly effective in end-to-end learning, enabling the direct transformation of raw data into actionable insights [53]. Convolutional Neural Networks (CNNs) have distinct features like local connectivity and weight sharing, which reduce the number of parameters, speed up training, and provide a regularizing effect. Challenges such as occlusion and pose variation, which can obscure or distort facial expressions, pose significant obstacles for facial emotion recognition (FER), particularly in uncontrolled environments. However, deep learning-based methods, including CNNs, substantially minimize the need for extensive image preprocessing and feature extraction. They are also more resilient to varying conditions like changes in lighting and obstructions, enabling them to significantly beat traditional methods. Moreover, these approaches are well-suited to manage large volumes of data effectively [54].

Although it requires more data for training, DL systems can adapt to novel situations and self-correct. In contrast, machine learning can work with lesser datasets but needs more human involvement for learning and error correction [55]. Deep learning algorithms improve their performance with more data exposure. The architecture of deep learning is designed for comprehensive feature extraction and modification. Initial layers process incoming data to learn simple features, while subsequent layers handle the learning of more complex features. This makes deep learning well-suited for managing larger datasets and intricate problems. Companies like Tesla, Apple, and Nissan are leveraging deep learning for developing autonomous technologies [56].

The high accuracy observed with traditional methods can often be attributed to smaller data sizes and static environments, where they excel. However, their performance tends to degrade in more challenging environments with larger datasets. On the other hand, while deep learning methods also achieve high performance, they particularly excel with larger datasets, which allow the models to learn and represent features more effectively. The performance of deep learning improves as more data becomes available. Traditional machine learning models, typically binary classifiers, face significant challenges when adapting to tasks involving non-linear and high-dimensional features, such as facial emotion recognition (FER) [57]. In contrast, neural networks (NNs), particularly Convolutional Neural Networks (CNNs), use non-linear activation functions like Rectified Linear Unit (ReLU). This non-linearity allows CNNs to catch complex relations and patterns in the data that linear models might miss [58]. In summary, CNNs outperform traditional techniques due to their capability to repeatedly and hierarchically extract features, model complex non-linear relationships, handle unstructured raw data, perform end-to-end learning, utilize large-scale data effectively, and maintain robustness to variability. These advantages make CNNs the go-to choice for a extensive range of CV (computer vision) and image-related jobs. Compared to traditional methods, CNNs offer greater flexibility, adaptability, generalization capability, and scalability, making them more suitable for real-world applications in diverse and uncontrolled settings [59].

In contemporary years, there has been a discernible trend towards employing DNN systems for analysis of expression within the realm of FER research [60].In the last decade, Deep Learning (DL) algorithms, notably CNNs, have brought about a paradigm shift in computer vision [61]. A pivotal advantage of deep learning approaches lies in their capacity to diminish dependence on handcrafted features, thus streamlining the preprocessing and feature extraction phases. However, despite their effectiveness, challenges persist in applying deep learning to FER.

Deep NN need extensive training datasets to mitigate the risk of overfitting. This constraint highlights the necessity for expansive and varied datasets to bolster the training of DL models intended for FER tasks [62]. Furthermore, DL systems demand significant computation sources and storage capacity for both train and test models, distinguishing from traditional machine learning methods. Consequently, there is a pressing need to optimize computation time during the inference process of DL models [63]. However, this method requires gathering data from multiple tasks, leading to progressively complex training procedures due to the increased number of tasks involved [64]. FER research continues to advance to address real-world applications such as driver drowsiness recognition, distance learning assistance, clinical patient monitoring, and teaching robots, as well as healthcare systems for children with autism. These applications highlight the necessity of developing FER systems that can effectively handle challenges related to illumination, lighting variations, pose variations, and aging effects for real-world expression classification systems. The key intention of this article is to inspect and contrast FER using both traditional ML and DL methodologies, with the intention of identifying areas of opportunity and challenges within the domain for emerging researchers. This article helps the researchers to handle the real-time expression dataset for real-time applications.

Many algorithms designed for facial emotion recognition (FER) produce impressive results on frontal or nearly frontal images. Yet, facial images are frequently captured from multiple angles and in spontaneous settings, making non-frontal or in-the-wild FER significantly more challenging and less explored. Non-frontal FER is particularly difficult due to issues like precise alignment of non-frontal faces, accurate localization of facial points [65]. In a controlled scenario, algorithms designed to detect posed expressions can be particularly advantageous due to their higher accuracy. Research by Li & Deng [66] demonstrated that their projected DLP-CNN beats state-of-the-art handcrafted features for expression classification in the wild. Mahony et. al. [67] examined the benefits and limitations of classical computer vision and deep learning techniques, showing that hybrid methodologies can enhance performance and address issues unsuitable for deep learning alone. KARYPIDIS et al. [68] also compared the most common ML and DL techniques for 2D object classification tasks in computer vision, highlighting the strengths and weaknesses of each method. Experiments conducted on the FER2013 dataset, which is known for its complexity and large size, demonstrate that the approach outperformed others. In this study, SIFT features were integrated with a CNN model, resulting in a 75.2% accuracy rate. This recognition rate surpasses that of traditional methods and purely deep learning-based predictions on the FER2013 dataset [69]. Quan et al. [70] introduced a model called the K-order emotional intensity model (K-EIM), which utilizes K-Means clustering to measure emotional intensity in an unsupervised manner with accuracy of 88.32%.

2.6 FER in healthcare

Facial expressions are a primary means of recognizing human emotions, essentially involving pattern recognition to identify regularities in the analyzed data, enabling the recognition of both faces and emotions [71]. Facial Emotion Recognition (FER) systems are garnering significant interest, particularly in healthcare, where they hold transformative potential. FER applications in healthcare include improving patient-practitioner communication and aiding in the diagnosis and treatment of mental health issues. This technology offers a new way to understand patient emotions, providing insights that were previously difficult to access. Such capabilities are crucial in mental health care, where monitoring emotional fluctuations can lead to more personalized and effective interventions. Additionally, FER systems are valuable in contexts where patients may have difficulty expressing their emotions, such as pediatric care or among individuals with neurodegenerative diseases, by helping to understand and address their needs. Building a robust FER system that can recognize basic emotions requires several preprocessing steps. These tasks are complicated by various factors such as pose, lighting surroundings, gender, oldness, and facial hair. Integrating FER into healthcare also presents challenges, including ethical considerations like privacy, data security, and informed consent, due to the sensitive nature of the data involved. The accuracy and reliability of FER systems are critical, as incorrect data can lead to adverse clinical outcomes. Additionally, it is crucial to address and mitigate potential biases in FER systems that may arise from the training datasets used. FER technology shows great promise in healthcare, but its implementation must be carefully managed to address technical, ethical, and practical challenges [72].

3. RESULTS AND DISCUSSION

This research presented cutting-edge techniques aimed at addressing both controlled and uncontrolled environments. Feature extraction methods like LBP, HOG, LDA, and PCA were utilized, followed by SVM for classifying facial expressions in static images. Two datasets, FER2013 and CK+, are employed to compare traditional and deep learning methods. FER2013 and CK+ can provide valuable insights for future data collection and preprocessing strategies in facial expression recognition (FER) research [73]. In the domain of DL, the data is initially separated into train and test sets using an 8:2 ratio, which can be adjusted as required. The training of a CNN model was performed using the train data and validated using the test data for expression classification into seven classes, with accuracy computed accordingly. Table 1 shows that the model has a total of 2,821,959 parameters. Out of these, 2,819,015 are trainable, meaning they are adjusted during the training process to lessen the loss function and advance the performance of the model. The remaining 2,944 parameters are non-trainable. Table 2 furnishes an overview of different existing FER systems employing varied feature extraction and classification techniques, alongside their respective accuracies. Table 3 juxtaposes the accuracy of implemented traditional methods with implemented CNN based on the FER2013 and CK+ datasets for facial expression classification. Table 2 demonstrates that most traditional machine learning (ML) methods have predominantly utilized images from controlled laboratory environments for Facial Expression Recognition (FER), resulting in higher accuracy. LBP is best suited for high-resolution images, making it ideal for controlled environments, medical imaging, and high-resolution security applications. On the other hand, HOG is more effective for low-resolution images, offering robustness to variability and computational efficiency, making it appropriate for real-time processing, surveillance, and applications in less controlled environments [74]. However, when applied to challenging datasets, the performance of traditional ML methods tends to decrease, as shown in Table 3. In contrast, CNNs consistently surpass traditional methods in both controlled and wild situations. Table 3 summarizes the experiments and results of FER using both traditional and DL methods. The experiment utilizing Local LBP features with a SVM classifier accomplished an average accuracy rate of 97% on the CK+ dataset. In contrast, the same method resulted in a lower recognition rate of 31.9% on the FER2013 dataset. Traditional methods show high accuracy in controlled environments and perform competitively compared to deep learning approaches. However, DL models tend to exceed traditional approaches in terms of recognition rates. The proposed CNN accomplished an average accuracy of 98% on the CK+ dataset. On the more challenging and larger FER2013 dataset, the proposed CNN accomplished an accuracy of 66.67%.

Traditional methods regularly need physical feature engineering, which is not only time-consuming but also computationally intensive as it may involve multiple iterations of trial and error [75]. The disparity between training and testing accuracy and loss in a CNN model, as depicted in Figure 6, offers critical perceptions into the model's potential for overfitting. Overfitting happens when a model learns the train data too well, catching noise and details that do not generalize to new data, leading to high train accuracy but low testing accuracy. The disparity where training loss is much lower than validation loss shows that the model is fitting the train data too well and not performing as well on the testing data.

Table 1 CNN architecture details

Layer (type)	Output Shape	Parameters
sequential (Sequential)	(None, 48, 48, 3)	0
conv2d (Conv2D)	(None, 48, 48, 32)	896
conv2d_1 (Conv2D)	(None, 48, 48, 64)	18496
batch normalization (BatchNormalization)	(None, 48, 48, 64)	256
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d_2 (Conv2D)	(None, 24, 24, 128)	73856
batch_normalization_1 (BatchNormalization)	(None, 24, 24, 128)	512
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 128)	0
dropout_1 (Dropout)	(None, 12, 12, 128)	0
conv2d_3 (Conv2D)	(None, 12, 12, 256)	295168
batch_normalization_2 (BatchNormalization)	(None, 12, 12, 256)	1024
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 256)	0
dropout_2 (Dropout)	(None, 6, 6, 256)	0
conv2d_4 (Conv2D)	(None, 6, 6, 512)	1180160
batch_normalization_3 (BatchNormalization)	(None, 6, 6, 512)	2048
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 256)	1179904
batch_normalization_4 (BatchNormalization)	(None, 256)	1024
dropout_4 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 256)	65792
batch_normalization_5 (BatchNormalization)	(None, 256)	1024
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 7)	1799

Total params: 2,821,959

Trainable params: 2,819,015

Non-trainable params: 2,944

____

Table 2 Various existing FER systems

Literature	Dataset	Feature Extraction Method	Classification	Accuracy (%)
2005[38]	CK	LBP	SVM	87.6
2016[21]	Video data	PCA	Kalman Filter	Improved accuracy
2017[12]	CK+	LBP+ Gabour	SVM	95.8
2017[23]	CK+	Gradient Descent	CNN	96.76
2018[24]	JAFFE	Gabor filters.	SVM	More accuracy
2019[25]	CK+	FACS	MNN+SVM	90
2020[18]	CK+	HOG	LDA	94.66
2020[27]	Image dataset	EEG signal	SVM	96
2017[26]	JAFFE	CNN	CNN	86.38
2019[45]	JAFFE	Deep CNN	Deep CNN	97.01

Table 3 Comparison of Accuracy using traditional ML Techniques with CNN in a posed and wild environment

CK+		**FER2013**	
Feature Extraction+ classification Technique	**Accuracy (%)**	**Feature Extraction+ classification Technique**	**Accuracy (%)**
LBP+SVM	97	LBP+SVM	31.9
HOG+SVM	98.66	HOG+SVM	28.9
PCA+SVM	84	PCA+SVM	18.7
LDA+SVM	96.66	LDA+SVM	27.93
CNN+FC	**98**	**CNN+FC**	**66.67**

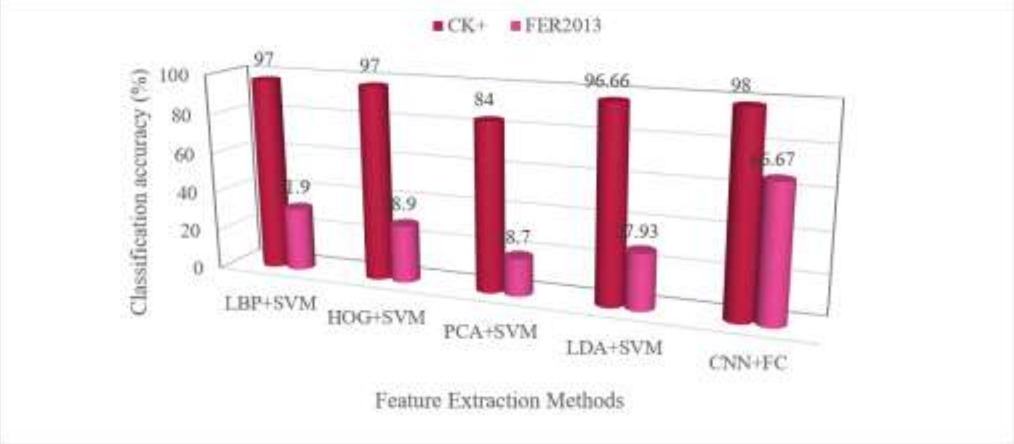


Fig. 8 Comparison of accuracy of traditional techniques with CNN

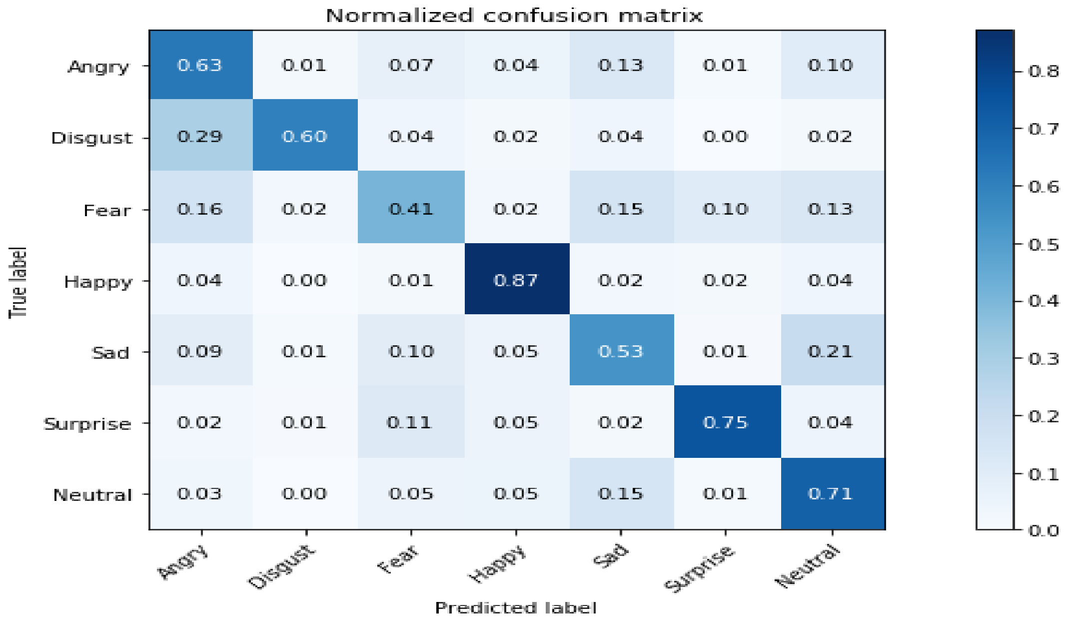


Fig. 9 Confusion matrix of 7 classes FER results obtained by ConvNet on the FER2013 dataset.

Implementing strategies like DA, regularization techniques, early stopping, simplifying the model, using pre-trained models, increasing training data, and batch normalization can support to mitigate overfitting and expand the model's generalization to unobserved data [76]. Figure 6 and 7 demonstrates that CNN achieves superior accuracy compared to traditional techniques in both controlled and uncontrolled environments. Facial emotion recognition holds pivotal significance across various domains, spanning from human-computer interface to healthcare.

The confusion matrix offers a detailed breakdown of the predictions made by a classification model. In this instance, normalized values between 0 and 1 are used, with each cell color-coded based on its value. The higher the number of predictions, the greater the color intensity. A color scale accompanies the matrix, indicating the range of colors and values. Ideally, the model's performance is reflected by high color intensity along the main diagonal of the matrix, indicating accurate predictions, while the other cells should have lower color intensities, as illustrated in the fig.9. It also provides insights into the accuracy of individual classes. In fig. 9, the highest recognition rates are typically observed in the happy and surprise categories, while the sad and fear categories often have the lowest recognition rates [77].

4. CHALLENGES AND OPPORTUNITIES

As the focus of facial expression recognition (FER) data shifts towards addressing the complexities of real-world environments, many researchers are turning to deep learning techniques to tackle challenges such as transitions, occlusions, off-frontal poses, personal biases, and detecting less-intense expressions. Nonetheless, a significant hurdle in deep facial expression recognition (FER) remains the limited availability of high-quality training data, which is deficient both in terms of quantity and diversity [78]. FER is inherently data-centric, and training deep networks capable of detecting subtle changes associated with expressions requires a vast amount of annotated dataset. The main challenge lies in acquiring a comprehensive expression database containing abundant samples with accurate facial characteristic labels, considering the diverse ways in which individuals across diverse age groups, cultures, and genders express and understand facial expressions. Capturing large volumes of varied image data in natural conditions is a significant challenge for training [79]. Moreover, class imbalance presents a prevalent challenge in facial expression datasets, reflecting the practicalities of data collection. While collecting data for common expressions like smiles may be straightforward, acquiring sufficient data for less frequent expressions such as disgust and anger can be challenging, leading to imbalanced class distributions in the dataset. It is imperative to address these issues to enhance the effectiveness and generalization capability of DL models for FER tasks. Exploring the classification accuracy of FER in aging adults warrants thorough investigation. Furthermore, the challenge in classification becomes more apparent when analyzing expressions of young and middle-aged subjects [80]. This complexity may arise from the heightened intensity with which these individuals' express emotions compared to aging adults [81]. Future research in facial expression classification should focus on enhancing data diversity, improving preprocessing techniques, exploring advanced model architectures, leveraging temporal and contextual information, addressing bias and fairness, increasing model explainability, and optimizing for real-life deployment. By handling these fields, scholars can build more robust, accurate, and fair facial expression recognition systems that perform well across a wide range of scenarios and populations.

5. CONCLUSIONS

This study explores the efficacy of facial expression classification through two distinct approaches: a) Traditional feature extraction techniques (including LBP, PCA, LDA, HOG, and SVM classifiers). b) CNN-based feature extraction. The evaluation is carried out using the FER2013 and CK+ databases. The summarized results of these techniques on both databases are presented in Table 2. In the realm of traditional techniques, LBP demonstrates superior performance on high-resolution images sourced from extensive datasets, while HOG achieves higher success rates on low-resolution samples with straightforward poses, particularly from the CK+ dataset. The direct extraction of features from the layers of CNN architecture helped to enhance classification accuracy further. Moreover, computational time is significantly reduced compared to traditional methods, particularly for large datasets. CNN showed improved accuracy for large datasets across both controlled and uncontrolled environments, thereby aiding in the lessening of computation time. Additionally, the study compared the accuracy of FER using traditional ML methods and CNN in posed and uncontrolled environments. Incorporating additional modalities like infrared images, depth data from 3D face models, and physical information may use to yield substantial enhancements to the investigation. These modalities offer substantial complementarity in capturing facial expressions, thereby enhancing the overall analysis and understanding of human emotions. DL is considered state-of-the-art due to its exceptional performance across various types of data, including static, sequential, in-the-wild, and dynamic data. Many recent FER models are hybrid systems, combining different approaches. Hybrid networks are increasingly developed. Future advancements in deep learning for FER are likely to involve various enhancement methods. These include integrating handcrafted features with deep learning features, using classifiers such as DT(decision trees), RF (random forests), and SVMs at the output layer of deep learning models, employing network cascading, utilizing generative networks, pre-training of datasets in an unsupervised manner with fine-tuned in a supervised manner, and applying optimization techniques. Enhancing deep learning models for FER, particularly in uncontrolled settings, remains an active area of research.

Financial and ethical disclosures: This research does not receive funding or support from any organization or agency.

Conflict of interest: The authors declare no conflicts of interest related to the publication of this paper.

REFERENCES

[1] Nisha V, Shilpa T., Facial Emotion Recognition using Machine Learning Techniques, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 23, Issue 5, Ser. I (Sep. –Oct. 2021), PP 41-49.

[2] Tsalera, E.; Papadakis, A.; Samarakou, M.; Voyiatzis, I. Feature Extraction with Handcrafted Methods and Convolutional Neural Networks for Facial Emotion Recognition. *Appl. Sci.* **2022**, *12*, 8455.

[3] Htay, Moe., Feature extraction and classification methods of facial expression: a survey. Computer Science and Information Technologies. 2. 2021,26-32.

[4] Wafa Mellouk, Wahida Handouzi, Facial emotion recognition using deep learning: review and insights, Procedia Computer Science, Volume 175, 2020, Pages 689-694.

[5] Debnath, T., Reza, M.M., Rahman, A. et al. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Sci Rep 12, 6991 (2022).

[6] Khan, Amjad Rehman., Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. Information. 13. 2022, 268.

[7] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022.

[8] O. S. Ekundayo and S. Viriri, "Facial Expression Recognition: A Review of Trends and Techniques," in IEEE Access, vol. 9, pp. 136944-136973, 2021.

[9] Ekaterina Ivanova, Georgii Borzunov, Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions, Procedia Computer Science, Volume 169,2020, Pages 244-248.

[10] Liu, Zhenhai & Wang, Hanzi & Yan, Yan & Guo, Guanjun., Effective Facial Expression Recognition via the Boosted Convolutional Neural Network. 2015, 179-188.

[11] A. Hewa,0. Nomir, A. Saleh, Facial Expression Recognition Based on Principal Components Analysis, International Journal of Intelligent Computing and Information Science, Vo1.16 No. 4 October 2016.

[12] Varma, Satishkumar, Megha Shinde, and Satishkumar S. Chavan. "Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers." In Techno-Societal 2018: Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications-Volume 1, pp. 109-119. Springer International Publishing, 2020.

[13] Zafar, Sahar & Ali, Fayyaz & Guriro, Subhash & Ali, Irfan & Khan, Asif & Zaidi, Adnan. (2019). Facial Expression Recognition with Histogram of Oriented Gradients using CNN. Indian Journal of Science and Technology. 12. 2019.

[14] Arora Malika; Kumar, Munish, AutoFER: PCA and PSO based automatic facial emotion recognition Multimedia Tools and Applications; Dordrecht Vol. 80, Iss. 2, (Jan 2021): 3039-3049.

[15] Kola, Durga & Samayamantula, Srinivas., A novel approach for facial expression recognition using local binary pattern with adaptive window. Multimedia Tools and Applications. 80. 2021,1-20.

[16] Ben Niu, Zhenxing Gao, Bingbing Guo, "Facial Expression Recognition with LBP and ORB Features", Computational Intelligence and Neuroscience, vol. 2021, Article ID 8828245, 10 pages, 2021.

[17] D. Lakshmi, R. Ponnusamy, Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders, Microprocessors and Microsystems, Volume 82,2021,103834, ISSN 0141-9331.

[18] R. Ravi, S. V. Yadhukrishna and R. Prithvi raj, "A Face Expression Recognition Using CNN & LBP," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 684-689.

[19] www.kaggle.com, accessed on. 23.02.2024.

[20] Chaudhari, Aayushi & Bhatt, Chintan & Adiraju, Achyut & Mazzeo, Pier Luigi. (2022). ViTFER: Facial Emotion Recognition with Vision Transformers. Applied System Innovation. 5. 80. 10.3390/asi5040080.

[21] G. K. Sahoo, J. Ponduru, S. K. Das and P. Singh, "Deep Leaning-Based Facial Expression Recognition in FER2013 Database: An in-Vehicle Application," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10040121.

[22] Neha A. Chinchanikar, Facial Expression Recognition Using Deep Learning: A Review, International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 06 | June 2019.

[23] C. Mejia-Escobar, M. Cazorla and E. Martinez-Martin, "Improving Facial Expression Recognition Through Data Preparation and Merging," in IEEE Access, vol. 11, pp. 71339-71360, 2023, doi: 10.1109/ACCESS.2023.3293728.

[24] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." arXiv preprint arXiv:2105.03588 (2021).

[25] Huang Y, Chen F, Lv S, Wang X. Facial Expression Recognition: A Survey. Symmetry. 2019; 11(10):1189. https://doi.org/10.3390/sym11101189.

[26] Kim JH, Poulose A, Han DS. The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance. Sensors. 2021; 21(6):2026. https://doi.org/10.3390/s21062026

[27] Mejia-Escobar C, Cazorla M, Martinez-Martin E. Towards a Better Performance in Facial Expression Recognition: A Data-Centric Approach. Comput Intell Neurosci. 2023 Nov 3;2023:1394882. doi: 10.1155/2023/1394882. PMID: 37954097; PMCID: PMC10637848.

[28] Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychol Sci Public Interest. 2019 Jul;20(1):1-68. doi: 10.1177/1529100619832930. Erratum in: Psychol Sci Public Interest. 2019 Dec;20(3):165-166. doi: 10.1177/1529100619889954. PMID: 31313636; PMCID: PMC6640856.

[29] N. Mishra and A. Bhatt, "Feature Extraction Techniques in Facial Expression Recognition," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1247-1251, doi: 10.1109/ICICCS51141.2021.9432192.

[30] S. Li and W. Deng, "A Deeper Look at Facial Expression Dataset Bias," in IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 881-893, 1 April-June 2022, doi: 10.1109/TAFFC.2020.2973158.

[31] Wafi, Muhammad & Bachtiar, Fitra & Utaminingrum, Fitri. (2023). Feature extraction comparison for facial expression recognition using adaptive extreme learning machine. International Journal of Electrical and Computer Engineering (IJECE). 13. 2023, pp1113-1122.

[32] T. Ojala, M. Pietikainen, and D. Harwood, “A comparative ̈ study of texture measures with classification based on featured distributions,” Pattern Recognition, vol. 29, no. 1, pp. 51–59, 1996.

[33] I.Michael Revina, W.R. Sam Emmanuel, A Survey on Human Face Expression Recognition Techniques, JOURNAL OF KING SAUD UNIVERSITY: COMPUTER AND INFORMATION SCIENCES (JUL 2021), Vol. 33, no. 6, pp. 619 – 628.

[34] S. Liao, M. W. K. Law, and A. C. S. Chung, “Dominant local binary patterns for texture classification,” IEEE Trans. Image Process., vol. 18, no. 5, pp. 1107–1118, May 2009. [crossref]

[35] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim and B. Fielding, "A Micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition," in IEEE Transactions on Cybernetics, vol. 47, no. 6, pp.

1496-1509, June 2017.

[36] Shan, Caifeng & Gong, Shaogang & Mcowan, Peter. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing. 27.2009. 803-816.

[37] Zixiang Fei, Erfu Yang, David Day-Uei Li, Stephen Butler, Winifred Ijomah & Huiyu Zhou. A survey on computer vision techniques for detecting facial features towards the early diagnosis of mild cognitive impairment in the elderly, Systems Science & Control Engineering, 7:1, 2019. 252-263.

[38] Lei Zhao, Zengcai Wang, Guoxin Zhang, "Facial Expression Recognition from Video Sequences Based on Spatial-Temporal Motion Local Binary Pattern and Gabor Multiorientation Fusion Histogram", Mathematical Problems in Engineering, vol. 2017, Article ID 7206041, 12 pages, 2017.

[39] Zhu, Dimin & Fu, Yuxi & Zhao, Xinjie & Wang, Xin & Yi, Hanxi. (2022). Facial Emotion Recognition Using a Novel Fusion of Convolutional Neural Network and Local Binary Pattern in Crime Investigation. Computational Intelligence and Neuroscience. 2022. 1-14.

[40] S. S. Meher and P. Maben, "Face recognition and facial expression identification using PCA," 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, India, 2014, pp. 1093-1098.

[41] Saket Karve, Vasisht Shende, Rizwan Ahmed, A comparative analysis of feature extraction techniques for face recognition, 2018 International Conference on Communication, Information & Computing Technology (ICCICT), Feb. 2-3, Mumbai, India.

[42] Sahoo, Ramesh & Pradhan, Sateesh. (2019). A Comparative Study on Hopfield Network with LBP, PCA and LDA for Face Recognition in Distorted Face Images. 102-110.

[43] Bhattacharyya, Suman & Rahul, Kumar. Face recognition by linear discriminant analysis. International Journal of Communication Network Security. 2. 2013. 31-35.

[44] Shih, Frank Y., Chao-Fa Chuang, and Patrick SP Wang. "Performance comparisons of facial expression recognition in JAFFE database." International Journal of Pattern Recognition and Artificial Intelligence 22.03 (2008): 445-459.

[45] Greche, L., Akil, M., Kachouri, R. et al. A new pipeline for the recognition of universal expressions of multiple faces in a video sequence. J Real-Time Image Proc 17, 1389–1402 (2020).

[46] Carcagnì, P., Del Coco, M., Leo, M.: Distante, Cosimo: Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus 4(1), 645 (2015).

[47] Jeniffer Xin-Ying Lek, Jason Teo, "Academic Emotion Classification Using FER: A Systematic Review", Human Behavior and Emerging Technologies, vol. 2023, Article ID 9790005, 27 pages, 2023.

[48] AKSOY, Orhan & GÜNEY, Selda. Sentiment Analysis from Face Expressions Based on Image Processing Using Deep Learning Methods. Journal of Advanced Research in Natural and Applied Sciences. 8. 2022.

[49] Çayir, Aykut, Işil Yenidoğan, and Hasan Dağ. "Feature extraction based on deep learning for some traditional machine learning methods." 2018 3rd International conference on computer science and engineering (UBMK). IEEE, 2018, pp. 494-497.

[50] Anil Audumbar Pise, Mejdal A. Alqahtani, Priti Verma, Purushothama K, Dimitrios A. Karras, Prathibha S, Awal Halifa, "Methods for Facial Expression Recognition with Applications in Challenging Situations", Computational Intelligence and Neuroscience, vol. 2022, Article ID 9261438, 17 pages, 2022.

[51] Podder T, Bhattacharya D, Majumder P, Balas VE. 2023. A feature boosted deep learning method for automatic facial expression recognition. PeerJ Comput. Sci. 9:e1216.

[52] Nur Alia Syahirah Badrulhisham and Nur Nabilah Abu Mangshor 2021 J. Phys.: Conf. Ser. 1962 012040.

[53] O. S. Ekundayo and S. Viriri, "Facial Expression Recognition: A Review of Trends and Techniques," in IEEE Access, vol. 9, pp. 136944-136973, 2021, doi: 10.1109/ACCESS.2021.3113464.

[54] Jyoti Kumari, R. Rajesh, K.M. Pooja, Facial Expression Recognition: A Survey, Procedia Computer Science,Volume 58,2015,Pages 486-491,ISSN 1877-0509

[55] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53. doi: 10.1186/s40537-021-00444-8. Epub 2021 Mar 31. PMID: 33816053; PMCID: PMC8010506.

[56] Taye MM. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers. 2023; 12(5):91. https://doi.org/10.3390/computers12050091.

[57] Walecki, R., Rudovic, O., Schuller, B., Pavlovic, V., & Pantic, M. (2017). Deep Structured Learning for Facial

Expression Intensity Estimation.

[58] Browne, Matthew & Ghidary, Saeed & Mayer, Norbert. (2007). Convolutional Neural Networks for Image Processing with Applications in Mobile Robotics. 10.1007/978-3-540-75398-8_15.

[59] Canedo D, Neves AJR. Facial Expression Recognition Using Computer Vision: A Systematic Review. Applied Sciences. 2019; 9(21):4678. https://doi.org/10.3390/app9214678.

[60] André Teixeira Lopes a, Edilson de Aguiar b, Alberto F. De Souza a, Thiago Oliveira-Santos, “Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order”, Elsevier, Pattern Recognition 61 (2017) 610–628.

[61] D. Venkataraman and N. S. Parameswaran, “Extraction of Facial Features for Depression Detection among Students”, International journal of pure and applied mathematics,2018, vol.118.

[62] Caifeng Shan, Shaogang Gong and P. W. McOwan, "Robust facial expression recognition using local binary patterns," IEEE International Conference on Image Processing 2005, 2005, pp. II-370.

[63] S. Shakya, S. Sharma and A. Basnet, "Human behavior prediction using facial expression analysis," 2016 International Conference on Computing, Communication and Automation (ICCCA), 2016, pp. 399-404.

[64] Zixiang Fei, Erfu Yang, David Day-Uei Li, Stephen Butler, Winifred Ijomah & Huiyu Zhou (2019) A survey on computer vision techniques for detecting facial features towards the early diagnosis of mild cognitive impairment in the elderly, Systems Science & Control Engineering, 7:1, 252-263.

[65] F. Zhang, T. Zhang, Q. Mao and C. Xu, "Geometry Guided Pose-Invariant Facial Expression Recognition," in IEEE Transactions on Image Processing, vol. 29, pp. 4445-4460, 2020, doi: 10.1109/TIP.2020.2972114.

[66] S. Li and W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," in IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 356-370, Jan. 2019, doi: 10.1109/TIP.2018.2868382.

[67] O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2020). Deep learning vs. traditional computer vision. In Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1 (pp. 128-144). Springer International Publishing.

[68] Karypidis, E., Mouslech, S. G., Skoulariki, K., & Gazis, A. (2022). Comparison Analysis of Traditional Machine Learning and Deep Learning Techniques for Data and Image Classification. arXiv preprint arXiv:2204.05983.

[69] Georgescu, Mariana & Ionescu, Radu Tudor & Popescu, Marius. (2019). Local Learning With Deep and Handcrafted Features for Facial Expression Recognition. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2917266.

[70] C. Quan, Y. Qian, and F. Ren, Dynamic facial expression recognition based on K-order emotional intensity model, in Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO), Dec. 2014, pp. 11641168.

[71] Mehta D, Siddiqui MFH, Javaid AY. Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. Sensors (Basel). 2018 Feb 1;18(2):416. doi: 10.3390/s18020416. PMID: 29389845; PMCID: PMC5856132.

[72] Bansal, Sonal & Rustagi, Aditya & Kumar, Anupam. (2021). Alzheimer's disease diagnosis based on feature extraction using optimised crow search algorithm and deep learning. International Journal of Computer Applications in Technology. 65. 325. 10.1504/IJCAT.2021.117272.

[73] Meena, Lakshminarayanan & Thambusamy, Velmurugan. (2023). Optimizing Facial Expression Recognition through Effective Preprocessing Techniques. Journal of Computer and Communications. 11. 86-101. 10.4236/jcc.2023.1112006.

[74] Y. Lin et al., "Large-scale image classification: Fast feature extraction and SVM training," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 1689-1696, doi: 10.1109/CVPR.2011.5995477.

[75] Krichen M. Convolutional Neural Networks: A Survey. Computers. 2023; 12(8):151. https://doi.org/10.3390/computers12080151.

[76] Jubair F, Al-Karadsheh O, Malamos D, Al Mahdi S, Saad Y, Hassona Y. A novel lightweight deep convolutional neural network for early detection of oral cancer. Oral Dis. 2022 May;28(4):1123-1130. doi: 10.1111/odi.13825. Epub 2021 Mar 5. PMID: 33636041.

[77] Yunfei Lai, A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition 2019 *J. Phys.: Conf. Ser.* 1314 012148

[78] M M, A M. Facial geometric feature extraction based emotional expression classification using machine

learning algorithms. PLoS One. 2021 Feb 18;16(2): e0247131. doi: 10.1371/journal.pone.0247131. PMID: 33600467; PMCID: PMC7891769.

[79] Chaudhari, Aayushi & Bhatt, Chintan & Adiraju, Achyut & Mazzeo, Pier Luigi. (2022). ViTFER: Facial Emotion Recognition with Vision Transformers. Applied System Innovation. 5. 80. 10.3390/asi5040080.

[80] Garcia-Garcia, Jose & Penichet, Victor & Lozano, María. Emotion detection: a technology review. 2017. 1-8.

[81] Caroppo, A., Leone, A. & Siciliano, P. Comparison Between Deep Learning Models and Traditional Machine Learning Approaches for Facial Expression Recognition in Ageing Adults. J. Comput. Sci. Technol. 35, 1127–1146 (2020). https://doi.org/10.1007/s11390-020-9665-4.