Journal of Innovative Technology, Vol. 6, No. 1, pp. 41-49, March 2024
http://doi.org/10.29424/JIT.202403_6(1).0004

A Review of the Deep Convolutional Neural Networks for the Analysis of Facial Expressions

Rajesh Kumar [1], Syed Ibrar Hussain [2,*]

ABSTRACT

Facial expression recognition has acquired much renouncement in recent years and is used in multiple fields of image processing, emotion detection, depression detection, criminal record findings, and medical-related fields. Facial expression recognition systems have revolutionized the field of emotion detection by allowing machines to identify and interpret human emotions, opening new avenues of research in deep learning and providing insight into the potential of artificial intelligence to understand and analyze human emotions. This study aims to thoroughly review Facial expression recognition systems, including an assessment of different methodologies and publicly available databases used in Facial expression recognition studies. Apart from the ability to diagnose emotional expressions distantly, the proposed face recognition model has potential uses in telemedicine and mental health diagnosis. All these developments are summarized, which additionally draws attention to their wider implications for healthcare services. This paper also investigates cutting-edge facial emotion identification approaches, especially deep learning and convolutional networks.

Keywords: Expression reorganization; Facial Emotion Recognition (FER); Convolution neural networks (CNN); Deep learning; Depression detection.

1. INTRODUCTION

To guess a person's facial expressions or emotions is a challenging task that involves comprehending a person's emotional state [1]. Research has indicated that written communication only conveys 7% of a person's expression, whereas speech conveys 38%, and facial expressions convey 55% [2]. This highlights the significance of facial expressions as a critical indicator of a person's emotions and thoughts. In 1971, Ekman and Friesen identified and classified a total of six (6) emotions, including fear, anger, happiness, surprise, sadness, and disgust. These emotions have a particular expression on the face that can be recognized in different cultures [3]. A person's emotions are vital to their existence and significantly impact how we perceive and understand things [4, 5, 6, 7]. This study explores the potential benefits of facial recognition for mental health assessment, highlighting its importance for enhancing telemedicine practices and extending the model's applicability. The next sections offer an in-depth investigation of these extra aspects, which enables a more in-depth understanding of the model's effects on healthcare.

Facial expression recognition (FER) has garnered significant attention and has been applied in various fields. Its leading utilization is IN computer-human interaction, which comprises corelative games, multimedia enjoyments, and artificial environment expert computer systems. FER also plays a vital role in analyzing emotions and behavior in the medical field, particularly in evaluating Autism [8], mental disorders [9], and pain assessment [10]. Moreover, it is utilized in surveillance and law enforcement to enhance security measures and aid in criminal investigations. To tackle facial recognition issues, most techniques typically involve two primary steps: feature extraction (which is an object-based technique) and classification (which categorizes the features). In the feature extraction stage, the facial images are analyzed to identify potential facial characteristics. As various expressions display distinct facial features, by conforming to these features, improved classification and dynamic applications can be achieved. [11].

The objective of this manuscript is to present a concise summary of the fundamental concepts and progress in FER as achieved by researchers in recent years. Additionally, it presents a comparative study of the latest research conducted in this field, utilizing various methods.

●This paper proposes a new vision that provides an extensive analysis of all the crucial techniques employed for detecting and identifying expressions. The uniqueness of this work lies:

●This manuscript provides a thorough evaluation and comparison of facial emotion identification approaches, models, and datasets.

●It presents an immense analysis and comparison of different methods, models, and datasets used in facial expression identification.

●The paper examines the current domains where FER is utilized and suggests potential areas for future research in FER applications.

The subsequent module of this manuscript offers an assessment of the foundational concepts related to FER, setting the stage for the subsequent sections. Module III contains an extensive study of expression recognition problems that use CNN and deep learning approaches, delving into the details of the techniques and methods employed in such approaches. Section IV features the ob-

Manuscript received December 15, 2023; revised January 16, 2024.

[1] 1 Ph.D. student, Department of Mathematics and Computer Sciences, University of Palermo, Via Archirafi 34, 90123 Palermo, Italy
[2] Ph.D. student, Department of Mathematics and Computer Sciences, University of Palermo, Via Archirafi 34, 90123 Palermo, Italy.
Research scholar (corresponding author), Department of Mathematics, University of Houston, Houston, TX, USA

servations and discussions that emerged from the study, highlighting the positivity and negativity of the different methods explored in the previous section. Finally, the paper concludes with section V, which presents the conclusion of the study and outlines potential avenues for foresight study and development in the field of FER. A key component of our face expression recognition study is the use of convolutional neural networks (CNN). Tailored for image-related tasks, CNNs autonomously extract crucial features from facial images, contributing to nuanced expression recognition. CNNs are effective for timely recognition due to their real-time processing capabilities, which emphasizes their significance in enhancing both the accuracy and efficiency of our method.

2. DATASET

Facial image and video datasets designed for emotion recognition have been available for some time now. These datasets are crucial for deep learning methods, serving as a repository of important attributes necessary for successful face emotion recognition. By aggregating datasets beforehand, researchers can ensure that they are ready for the extraction and interpretation of results. Emotion recognition algorithms also heavily rely on these datasets. Considering the crucial role of datasets in data-driven learning for expression identification [12], this module presents a comprehensive review of the major datasets utilized for recognizing emotions in images.

2.1 JAFEE

The JAFFE dataset stands for Japanese Female Facial Expression, which many researchers commonly use, as [62] explained it is accessible for public use and comprises images featuring ten Japanese women displaying a range of emotions, The study aimed to capture photographs of participants displaying various reaction (emotions) acting as, disgust, sad, anger, happy, surprise, fear and neutral. The photographs were taken to create a dataset for emotion identification research. The attendees were requested to portray each emotion while being photographed by professional photographers in a controlled setting to ensure high-quality images. The photographs captured various aspects of facial expressions, including the eyes, mouth, and overall facial configuration. The pictures were categorized with expressions and the dataset was made openly available for the researcher to use in expression identification investigation. During the data collection process, each participant was asked to take around 4 to 5 photographs of themselves while looking towards the camera through a semi-transparent plastic sheet, resulting in a total of 213 images. All images are of size 256x256. The JAFFE dataset contains the 7 universal emotions [13], [14], [15].

2.2 Fer 2013

The FER dataset, which is called Facial expression Recognition is a commonly used dataset for facial expression identification. Pierre-Luc Carrier and Aaron Courville developed the facial expression recognition dataset and made it available for public use in the ICML Kaggle competition. The dataset comprises a total of 35887 black and white(grayscale) images, each having a resolution of 48 by 48 (48*48) pixels. It is split into pair sets for training and testing purposes, with 28,709 and 3,589 images, respectively. The images are elucidated with seven basic facial emotions. The dataset contains a sum of 35887 images, which have been segregated into various categories based on the emotional expressions they depict. Specifically, there are seven different emotion categories, each containing a different number of images. The categories and their corresponding image counts are as follows: anger (4953), fear (5121), sad (6077), neutral (6198), happy (8989) surprise (4002), and disgust (547) images respectively.

2.3 CK AND CK+

In 2000, Cohn et al. [16] introduced a database of facial expressions that contained 486 sequences from 97 different individuals, with sixty-five percent of the subjects being women and thirty-five percent were manned. The CK dataset contains a series of sequences showing facial expressions ranging from neutrality to peak expressions, with apex expression images fully coded and labeled with an emotion. However, A point to be aware of is that the emotion label in the dataset indicates the emotions that were requested, rather than the ones expressed by the participant. The CK dataset holds 2-dimensional images acquired in a monitored environment with standardized lighting, circumstances, angle of view and all utterances were obtained through the performance of educated participants, which may differ from spontaneously acquired images. Images have 640x490 or 640x480 resolutions, accompanied by a pixel matrix containing 8-bit grayscale or color with 24-bit values. To address the limitations of the CK data set, Cohn et al. [17] proposed a prolonged version called the Cohn Kanade Extended Database (CK+). The CK+ data set includes non-posed images and uses formal expression labels based on the participant's impression of the seven primary expression categories, namely Happy, Sad, Disgust, Anger, and Surprise. The CK+ dataset has an increased number of subjects from 97 to 123 and an increased number of sequences from 486 to 593, with a 31% male and 69% female population percentage.

2.4 MMI-V

The MMI (Multimodal Meeting-Recognition) dataset is a compilation of facial expressions from 25 individuals of various ethnicities, The dataset consists of participants from a diverse age range, with individuals ranging from 19 to 62 years old. The gender distribution in the dataset is such that 44% of the participants are women, while 56% are men. In the fourth part of the dataset, annotations are available for 6 (six primary emotions, namely fear, anger, happiness, surprise, sad, neutral, and disgust. Additionally, the dataset includes annotations for facial muscle movements, allowing for a more in-depth analysis of the participants' expressions. Additionally, the dataset's Part V includes annotations for daughters who speak and those who do not. Furthermore, a distinct dataset Later on, the MMI-V dataset was incorporated into the existing MMI dataset of facial expression.

2.5 BU-3DFE

The Binghamton University database [19] is composed of 101 subjects, together with 58 women and 43 men with diverse ethnic and racial backgrounds, The database comprises separate ages between 18 to 70 years old, and it covers the six prime expressions, including neutral expression. The data was collected using a 3D face scanner, The data set includes seven facial emotions from each participant, with four levels of severity recorded for each of the six key emotions. Additionally, each emotion shape model is matched with consistent facial texture images captured from two different viewpoints. Multiple intensity levels were obtained for each emotion from each subject using several techniques.

2.6 BU-4DFE

The BU-4DFE is a comprehensive 3D active facial emotion database designed for the analysis of facial actions from static to dynamic 3D space. It features a total of 606 3D emotion arrangements comprising around 60,600 devices. The dataset was recorded from 101 subjects with diverse ethnic and racial backgrounds, consisting of 43 males and 58 females. The video sequences consist of 3D models, with each model having an approximate resolution of 35,000 vertices and a dataset that includes the corresponding texture frames, featuring a resolution of approximately 1040 x 1329 spatial resolution.

2.7 MISCELLANEOUS

In addition to the databases, there exist numerous notable datasets in the domain of facial emotion identification. These include: such as EmotiW [20], CASIA Webface [21], KDEF[22], FEEDB [23], eNTERFACE'05 [24], RaFD [25], (NIMH-ChEFS) [26], TFEID [27], Afew [28], Iemocap [29], AffectNet [30], Kdef [31], iCV-MEFED [32], ISED[33], SEMAINE [34], BAUM-1[35], EmotioNet [36], FER-Wild[37] and HAPPEI [38].

One notable dataset in facial expression recognition is the dataset [21], which comprises almost 500,000 images from over 10,000 subjects gathered from the internet. Another dataset, [23], features video sequences captured using a Microsoft Kinect sensor, designed for facial recognition and emotion studies. The [20] dataset is comprised of two distinct sub-databases, The Acted Facial Expression in the Wild (AFEW) and Static Facial Expression in the Wild (SFEW) are two sub-databases included in the larger Facial Expression Recognition and Analysis Challenge (FERA) dataset. Both sub-databases contain samples of the six primary emotions (sad, fear, anger, happy, disgust, and surprise), as well as neutral expressions. The images in SFEW have a resolution of 128 x 128 pixels. Another dataset commonly used in facial expression recognition research is the database described in [22]. The images feature the six basic emotions (happiness, sadness, anger, disgust, fear, surprise) as well as neutral expressions. The images were captured from diverse perspectives across two separate sessions. The database [24] comprises video sequences (with audio) of 42 subjects from 14 different nationalities, with 81% of men and 19% of women, wearing glasses and beards in 31% and 17% of cases, respectively. The sequences have a resolution of 720 × 576 and cover the six key expressions. Alternatively [25] is a dataset of pictures featuring 67 models simultaneously captured from five different camera angles, The dataset records eight emotional expressions, namely disgust, contempt, neutral, sad, surprise, fear, and happiness, with three gaze directions, and an image size of 1024 × 681. Lastly, the [26] dataset comprises 482 high-fidelity colored images of children exhibiting five emotions and different gaze directions. The Brain Mapping Laboratory and Integrated Brain Research Unit developed the [27] dataset, which features eight emotions and 40 models, but it is only available for scientific research purposes. In contrast to other lab-controlled datasets, proposes a face image extraction method from movies [28]. Additionally, the database is a vast repository of multimodal emotion detection resources, which provides an extensive collection of 12 hours of recorded audio-visual data [29]. This dataset includes data captured from 10 human participants engaged in conversation. The dataset comprises recordings of multiple modalities, including video and audio, providing an excellent resource for researchers and practitioners seeking to develop and evaluate novel techniques for multimodal emotion detection. This database is one of the utmost far-reaching and freely accessible resources, enabling researchers to explore and experiment with various techniques to advance the field of multimodal emotion detection. The IEMOCAP dataset stands out from other facial emotions datasets because it not only includes the key emotions of anger, fear, sadness, and neutral, but also incorporates notations for valence, dominance, and the degree of emotional activation about the context. This dataset is widely used in the field of facial emotion recognition due to its large size and comprehensive labeling [30]. The KDEF dataset comprises 4,900 images classified with seven fundamental facial emotions (i.e., fear, neutral, anger, disgust, and happy) and acquired from 35 male and 35 female participants from five distinctive viewpoints [31]. The iCV-MEFED dataset features a combination of emotions, as each of the 31,250 facial images is described by a dominant and complementary emotion [32]. The 125 participants were instructed to elicit 50 distinct emotional states, thereby resulting in the production of 50 types of complex emotional expressions. The dataset contains 428 videos of 50 Indian subjects who were induced to experience emotions by watching emotional videos, and their self-evaluation for each emotion (happy, disgust, sad, and neutral) was collected simultaneously [33].

The SEMAINE dataset, as described in [34], contains 130,695 frames from typical sessions of Solid SAL and semi-automatic SAL. The study involving 24 undergraduate and postgraduate students encompassed individuals whose ages ranged from 22 to 60 years old. In contrast, the dataset presented in the dataset described in [35] is composed of 1,184 multimodal facial video clips captured from 31 Turkish participants, showing a variety of emotions and mental states, such as joy, anger, grief, disgust, surprise, fear, boredom, disdain, confused, neutral, thinking, concentration, and concerned. With many modalities present in the movies in the dataset—visual, auditory, and physiological signals—the collection offers a wealth of data for creating facial emotion detection systems that are more reliable and precise. Furthermore, the dataset [36] contains a set of one million pictures depicting facial expression images procured from the internet, categorized into 23 types of basic and compound emotions, annotated with either the corresponding emotion category or with associated Aus. The FER-Wild [37] dataset comprises 24,000 web images obtained from popular search engines such as Bing, Google, Yahoo, Yandex, and Baidu. These images are divided into nine categories: no-face, sad, fear, neutral, anger, surprise, disgust, and uncertain. The 'no-face' category includes instances where no face is present in the image, the face is partially covered by the bounding box, a watermark is present on the face, the face is drawn or animated, or the 'none' category is assigned when the face is significantly beyond the distortion of a normal or natural shape, making it difficult to infer any expression. When images do not display any of the six key expressions that are neutral, Nonetheless, 'uncertain' category is utilized when annotation experts are unclear of facial movements. Additionally, the dataset described in [38] contains 4,886 images of 8,500 faces sourced from Flickr and annotated by four human annotators.

Table 1 A summary of some FER datasets with some statistical details.

Sno.	Dataset	Size	Facial expressions	type
1	JAFFE	213	7	Images
2	Fer 2013	35887	7	Images
3	CK AND CK+	Sequence: 486 &593	8	Images
4	MMI-V	2900	6	Videos

5	BU-3DFE	2500	7	Images
6	BU 4DFE	606	7	videos
7	EmotiW	1268,700	7	Videos & Images
8	CASIA Web-face	494,414	NA	Images
9	KDEF	4900	7	Images
10	FEEDB	NA	NA	Videos
11	eNTER-FACE'05	1166	6	Videos
12	RaFD	29672	8	Images
13	NIMH-ChEFS	482	5	Images
14	TFEID	112234	8	Images
15	Afew	1809	7	Images
16	Iemocap	12h	4	Videos
17	AffectNet	1000000	8	Images
18	iCV-MEFED	31250	50	Images
19	ISED	428	4	Videos
20	SEMAINE	31250	4	Videos
21	BAUM-1	184	13	Videos
22	EmotioNet	1000000	23	Images
23	FER-Wild	24000	6	Images
24	HAPPEI	4886	6	Images

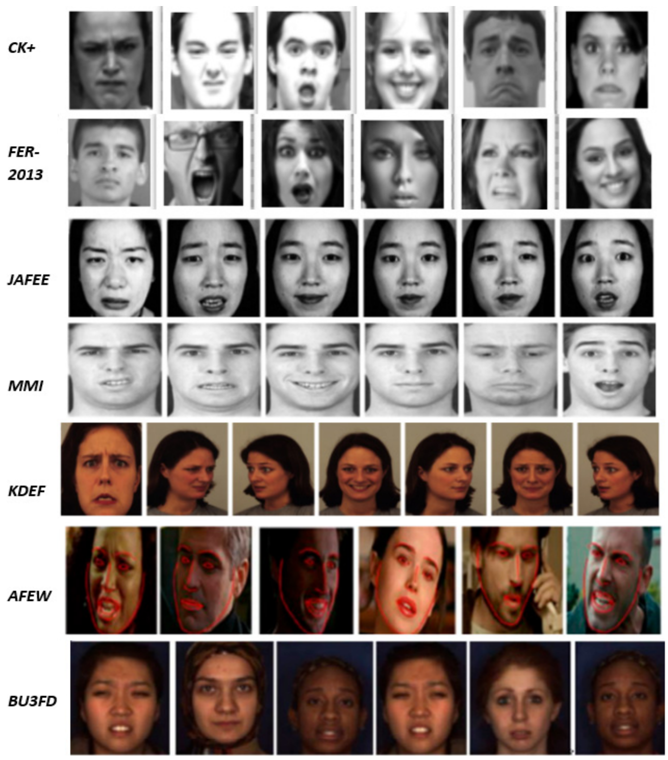

Fig. 1 Facial samples from different datasets.

3. RECOGNITION OF FACIAL EMOTIONS USING MACHINE LEARNING

Understanding facial expressions is essential to understanding human behavior [39,40], as illustrated in Figure 1. Figure 1 presents the visual representation of facial expressions from several known datasets. Psychological research has established that facial emotion recognition involves measuring the location and movement of the eyes, nose, and mouth. Figure 2 shows that emotions are identified from facial images, which go through a series of stages including preprocessing, to reduce irrelevant information, facial emotion recognition typically involves three main steps: face detection, feature extraction, and classification. In the past, distance-based methods were commonly used to estimate facial emotion intensity. Spatial volumetric distinction patterns are employed along with a multidimensional rate transformation in these techniques, which aid in the categorization and quantification of facial expressions. In the case of video-based systems, Principal Component Analysis (PCA) is frequently employed to represent features of facial expressions, allowing for the recognition of action units that express different facial emotions [41]. Additionally, PCA is used to structure and recognize other facial expressions by providing a facial action unit [42].

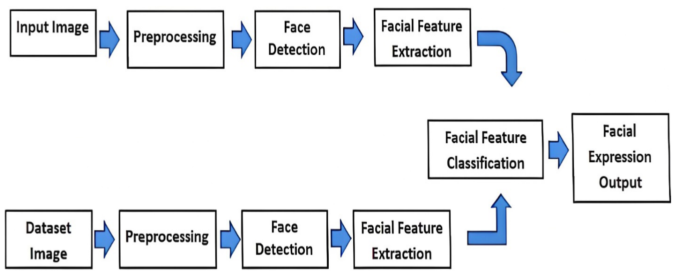

Fig. 2 Conventional approach for facial expression

Siddiqi et al. [43] used the active contour model to identify and draw out the facial segment in their study. They used the energy functions of Bhattacharyya and Chan-Vese to reduce span between the face and its adjuncts and minimize the variants inside the face. The team also used wavelet decomposition to reduce noise and extract features of the geometric appearance of facial emotions, as well as features of facial motion using optical flow. Unlike convolutional neural network (CNN) methods, conventional machine learning (ML) approaches do not require high computing power and memory. Therefore, further research is needed to adapt these methods for use in immersive devices capable of performing real-time classification with low computing power while providing satisfactory results.

4. CONSTRAINT OF ORDINARY METHODOLOGY

The traditional strategies for facial emotion identification have their obstacles when it comes to feature classification, extraction, and processing of abundant data. On the other hand, deep learning-based techniques rely on automatic feature extraction and pre-processing of images, making them more effective in domains with varying parameters like radiance and obstruction. Furthermore, Deep learning can handle large amounts of data, which sets it apart from traditional techniques that can only manage limited forms and ranges of data. To overcome the limitations of conventional approaches, deep learning techniques have been developed. Conventional techniques rely on models that use physical facial activity, and they require pre-processing techniques for facial images, and manual feature extraction. By utilizing an "end-to-end" learning approach, deep learning significantly reduces the need for manual effort [44]. However, deep learning approaches require more power and memory compared to conventional approaches, their multimodal nature makes them unsuitable for embedded systems but ideal for emotion recognition based on multimodal input. Further, conventional approaches cannot be combined to improve performance at the same time [45] and building testing and training models require more computational load as the components are designed individually and physically. Table 2 and Table 3 express the summary average accuracy rate of conventional facial expression recognition techniques with different datasets by various researchers.

Table 2 Summary of conventional Facial expression recognition methods

Initiated by	Extraction of features	Datasets	Method
Zeng et al. [46]	LBP, HOG, and SIFT	CK+	DSAE
(Kim et al. [47]	EHMM	CK, JAFE	ASM
Sajjad et al. [48]	ORB, SIFT, SURF	MMI JAFFE CK+	SVM
Al-Agha et al. [49]	Geometric descriptor	BOSPHORUS	Euclidean distance
Chang et al. [50]	AAMS	CK+	SVR
(Ghamire et al 51)	Kanade-Lucas-Tomaci Tracker, Elastic bunch graph matching algorithm,	CK+ MMI MUG	SVM and ELM
Yang and Wang [36]	LBP	BVTKFER B Curtin Faces	Classifier: Random Forest

Table 3 The Feature extraction accuracy using conventional methods.

Initiated by	Datasets	Accuracy	Technique
Zeng et al. [46]	CK +	95.80%	DSAE, HOG
Kim et al. [47]	CK, JAFE	77%, 64%	ASM
Agha et al. [48]	BOSPHOIRUS	85.0%	Euclidean distance
Sajad et al. [49]	JAFFE MMI CK+	98.10% with MMI accuracy and 92.15% with JAFFE, Accuracy: 90.21% with CK +	SVM
Chang et al. [50]	CK +	AAM 0.306. Gabor 0.339. AAM 0.334, MAEs, Gabor: 0.651	SVR, AAMS
Ghamire et al. 51	MMI, CK +, MUG	77.2 %, 97.8%, 95.5%	ELM and SVM, KLT

5. EMOTION RECOGNITION WITH DEEP LEARNING

The area of FER using deep learning has garnered substantial interest from researchers because of its potential for accurate and automatic recognition. A multitude of techniques have been developed for recognizing emotions from facial images and videos. As illustrated in Figure 3 [52], these techniques generally entail preprocessing the input data, extracting features, and distinguishing expressions. The same approaches can be applied for both tasks and can identify the six dominant emotions, namely anger, fear, surprise, disgust, sadness, and happiness by analyzing facial micro-expressions. Some of the most used techniques include:

I)CNN
II)LSTM
III)GAN
IV)RNN
V)RBFN

CNN: The CNN is a commonly utilized framework for classifying feature learning and emotion identification. This framework integrates feature extraction and classification to develop an artificial expression system for the facial emotion recognition [53].

LSTM: LSTM is a technique that extracts the feature that solves long term reliance problems using short term memory. It is specifically designed for temporal data and can process input sequences of facial points, indicators, and portraits of facial components. The LSTM can grasp from protracted arrangements and each block holds a recollection cell that can update the details about the state of the cell [52]. By encompassing the secular features of continuous images, LSTM boosts the precision of expression scrutiny and can evaluate the spatial representation of feeling in multimodal circumstances [54].

Generative Adversarial Network (GAN): GAN is a productive modeling approach that utilizes deep learning CNNs to extract the features classification. GAN structure includes two networks, namely a source network and an intolerance network, which work together to produce realistic data. The generator network creates fake samples, and the discriminator network evaluates the samples to determine their authenticity. Both networks are trained simultaneously, with the generator network trying to create samples that can fool the discriminator network, while the discriminator network tries to discriminate between true and false samples [55].

Recurrent Neural Network (RNN): As with CNN, RNN analyzes data into a sequential pattern of facial features based on temporal and spatial information [56]. Therefore, RNN is suitable for tasks such as speech and text recognition, and feature extraction from the facial images. This technique has an intrinsic memory system, which makes it more versatile and powerful as a networked memory. RNN contains neural network processes that help with memory, which is its great advantage.

RBFN: The radial basis function network (RBFN) is deep learning (Convolution neural network) used for pattern recognition and function approximation tasks. It includes ternary layers an input layer, a hidden layer that uses radial basis functions to map the input data into a higher-dimensional feature space, and an output layer that is a linear combination of the radial basis functions with the associated weighting. factors, which are learned during the training phase. RBFNs are well known for their accuracy, and ability to generalize new data, and have been employed in multiple areas such as speech and image processing, and financial forecasting.

Li et al. [57] suggested a 3D CNN architecture to identify expressions from recorded clips or videos. The team extricated detailed aspects and evaluated network performance on three benchmark data sets, specifically, CASME II, SMIC, and CASME. In another study, Li et al. [54] used cropping and face rotation methods to extract features using a CNN and tested their approach on the CK+ and JAFFE databases. [58] also presented an in-depth CNN approach for automatic facial expression recognition, which was evaluated on two data sets, JAFFE and CK+, and compared with the K-nearest neighbor (KNN) algorithm. resulting in significantly higher accuracy. The accuracy rates on JAFFE and CK+ were recorded as 76.7442% and 80.303%, respectively. One major advantage of the approach is exact accuracy even in low susceptivity situations, without requiring human control. improve the discrimination ability by focusing on non-occluded regions and interpreting the face region that is obscured. CNN merges different representations of facial regions of interest (ROIs) and applies a gate unit to weigh each representation based on its importance. Two types of ACNN were developed: patch based (pACNN) and global-regional-based (gACNN). Table 4 and Table 5 show the summary and average accuracy rate of facial expression identification based on deep learning for various datasets by different researchers in this field.

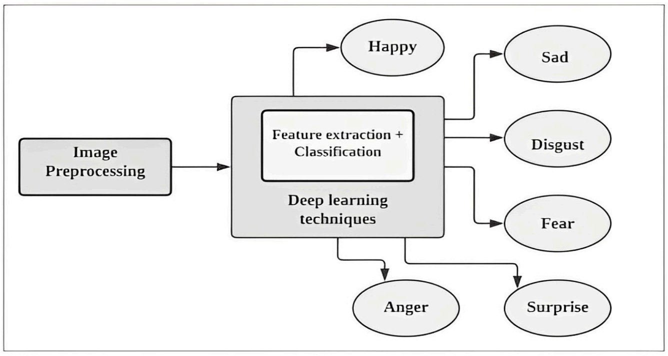

Fig. 3 Facial expression identification based on CNN

Table 4 Summary of Facial expression identification based on deep learning

Proposed by	Dataset	Method
(Ucar et al.[66]	CK, JAFE	OSELM,RBFN
(Minaee et al.[65]	FER 2013, CK+, JAFE	ACN
Li et al. [64]	CK+ JAFFE	CNN
Jain et al. [60]	CK+,MMI	LSTM,CNN
Yu and Zhang [63]	FER 2013,SFEW	DCNN
Al-Shabi [62]	FER 2013,CK+,	SIFT,CNN
Chu et al. [61]	BPD4	CNN.LSTMS

Table 5 Average rate of correctness on deep learning techniques.

Proposed by	Dataset	Technique	Accuracy
(Ucar et al.[66]	CK, JAFE	OS-ELM,RBFN	95.71%,94.65%
(Minaee et al.[65]	FER 2013,CK+,-JAFE	CAN	70.02%,98.0%
Li et al. [64]	CK+ JAFFE	CNN	97.38%,97.18%
Jain et al. [60]	CK+,MMI	LSTM,CNN	96.17%,98.72%
Yu and Zhang [63]	FER 2013, SFEW	DCNN	55%,61.29%
Al-Shabi [62]	CK+ and FER 2013	CNN, sift	99.1%,73.4%
Chu et al. [61]	BPD4	CNN. LSTMS	82.5%

The methods mentioned earlier, particularly those utilizing deep learning, demand considerable computational resources. Additionally, these methods are designed to categorize specific emotions, which makes them unsuitable for classifying other emotional states. As a result, creating a new framework that can be used to identify the entire range of features would be highly pertinent and can be extended to identify intricate FE.

6. CONCLUSION

This research paper provides an extensive analysis of differences among traditional neural networks and convolutional neural network techniques with models proposed for facial and speech-based emotion recognition, highlighting those that have achieved the highest accuracy. This research highlights the need for more advanced algorithms, models, and datasets to get beyond current limitations in emotion identification, while also acknowledging the vital part that feature extraction techniques and classifier selection play in reaching high accuracy rates. Following that, research ought to delve into the creation of unidimensional or multiple-dimensional models that incorporate diverse input data modalities, such as facial, textual, and audio-visual inputs. Emotion recognition is used in industries for promotional research, gathering consumer reactions to items, since emotions are recognized as important markers of moods, thoughts, behavior, and gestures. Emotion recognition in security has significance because it can be used to identify criminals, perpetrators, or thieves and prevent crime.

To sum up, this research study carried out a detailed analysis of the differences between convolutional neural network techniques and classic neural networks in relation to the identification of emotions using speech and facial recognition. We've highlighted the benefits and drawbacks of both strategies by highlighting models that have proven to be more accurate. Although conventional techniques frequently rely on facial recognition, feature and classification, extraction, CNNs provide an "end-to-end" learning method; nonetheless, there may be limitations on time during the training and testing stages.

Future directions for emotion recognition research include extensive testing of deep learning and multi-layer methods in artificial intelligence. By leveraging face, audio, and textual data, these methods seek to get high identification accuracy rates in several emotion recognition domains. This predicted testing has the possibility of helping enhance the accuracy and effectiveness of emotion recognition systems.

CONFLICT OF INTEREST

Neither of the authors have any conflicts of interest to disclose.

REFERENCES

[1] Hu, L., Li, W., Yang, J., Fortino, G. and Chen, M. (2019). "A Sustainable Multimodal Multi-layer Emotion-aware Service at the Edge." *IEEE Transactions on Sustainable Computing,* doi: 10.1109/TSUSC.2019.2928316.

[2] Mehrabian, A. (1968). "Communication without words." *Psychol. Today,* Vol. **2**, pp. 535.

[3] Ekman, P., & Friesen, W.V. (1971). "Constants across cultures

in the face and emotion." *J. Pers. Soc. Psychol.,* **17**(2), pp. 124–129.

[4] Vandervoort, D. J. (2006). "The importance of emotional intelligence in higher education." *Current Psychology: Development,* Learning, Personality, Social, **25**(1), 4–7.

[5] Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). "The Role of Emotions in Marketing." *Journal of the Academy of Marketing Science*, **27**(2), 184–206. doi: 10.1177/0092070399272005.

[6] Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor." *Journal of Educational Media,* **29**(3), 241-250. doi: 10.1080/1358165042000283101.

[7] Nussbaum, M. (2004). "Emotions as Judgments of Value and Importance. In R. C. Solomon (Ed.), Series in affective science. " *Thinking about feeling: Contemporary philosophers on emotions* (pp. 183-199). New York, NY, US: Oxford University Press.

[8] Harms, M.B., Martin, A., & Wallace, G.L. (2010). "Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies." *Neuropsychology review,* **20**(3), 290-322.

[9] Sprengelmeyer, R., Young, A., Mahn, K., Schroeder, U., Woitalla, D., B¨uttner, T., Kuhn, W., & Przuntek, H. (2003). "Facial expression recognition in people with medicated and unmedicated Parkinson's disease." *Neuropsychologia,* **41**(8), 1047-1057.

[10] Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., & Prkachin, K.M. (2010). "Automatically detecting pain in video through facial action units." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* **41**(3), 664-674.

[11] Kumar, R. (2023). "A Deep Learning Approach To Recognizing Emotions Through Facial Expressions." *In 2023 Global Conference on Wireless and Optical Technologies (GCWOT)* (pp. 1-5). Malaga, Spain. doi: 10.1109/GCWOT57803.2023.10064654.

[12] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning. " *IEEE Transactions on Medical Imaging,* **35**, 1285-1298. doi: 10.1109/TMI.2016.2528162.

[13] X. Zhongzhao, Y. Li, X. Wang, Z. Liu, "Convolutional neural networks for facial expression recognition with few training samples," in 37th Chinese Control Conference (CCC), China, 2018, pp. 1-6.

[14] H. Ali, M. Hariharan, and S. Yaacob, "Facial emotion recognition using empirical mode decomposition," *Expert Systems with Applications,* vol. **42**, pp. 1261-1277, 2015.

[15] S. Minaee and A. Abdolrashidi, "Deep-emotion: facial expression recognition using attentional convolutional network," arXiv abs/1902.01019, 2019.

[16] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, France, 2000.

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, San Francisco, CA, USA, pp. 94–10.

[18] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands, 2005, pp. 5.

[19] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, vol. 2006, pp. 211–216, 2006.

[20] A. Dhall, O. R. Murthy, R. Goecke, J. Joshi, and T. Gedeon, "Video and image based emotion recognition challenges in the wild: Emotiw 2015," in Proceedings of the 2015 ACM on International Conference On Multimodal Interaction, Seattle, WA, USA, 2015, pp. 423–426.

[21] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:*1411.7923, 2014.

[22] M. G. Calvo and D. Lundqvist, "Facial expressions of emotion (KDEF): Identification under different display-duration conditions," *Behavior Research Methods,* vol. **40**, pp. 109–115, 2008.

[23] M. Szwoch, "FEEDB: A multimodal database of facial expressions and emotions," in 2013 6th International Conference on Human System Interactions (HSI), Sopot, Poland, 2013, pp. 524-531.

[24] Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006). "The eNTERFACE'05 audio-visual emotion database. In Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06)," *Atlanta, GA, USA,* 3–7 April 2006; p. 8.

[25] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). "Presentation and validation of the Radboud Faces Database." *Cognition and Emotion,* **24**(8), 1377-1388.

[26] Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). "The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): A new set of children's facial emotion stimuli. International Journal of Methods in Psychiatric Research," **20**(3), 145-156.

[27] Chen, L. F., & Yen, Y. S. (2007). "Taiwanese facial expression image database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan.

[28] Dhall, A., Goecke, R., Lucey, S., Gedeon, T., et al. (2015). "Collecting large, richly annotated facial-expression databases from movies." *IEEE Multimedia,* **22**(4), 72-77.

[29] Ahmad, I., Ilyas, H., Hussain, S.I. and Raja, M.A.Z., 2023. "Evolutionary Techniques for the Solution of Bio-Heat Equation Arising in Human Dermal Region Model." *Arabian Journal for Science and Engineering,* pp.1-26.

[30] Tautkute, I., Trzcinski, T., & Bielski, A. (2018). "I know how you feel: Emotion recognition with facial landmarks." *In Pro-*

ceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1878-1880.

[31] Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R., & Akan, A. (2019). "Real time emotion recognition from facial expressions using CNN architecture. In Medical Technologies Congress (TIPTEKNO) Izmir," *Turkey,* 1-4.

[32] Ahmad, I., Hussain, S.I., Ilyas, H., Zoubir, L., Javed, M. and Zahoor Raja, M.A., 2023. "Integrated Stochastic Investigation of Singularly Perturbed Delay Differential Equations for the Neuronal Variability Model." *International Journal of Intelligent Systems,* 2023.

[33] Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2016). "The Indian Spontaneous Expression Database for Emotion Recognition." *IEEE Transactions on Affective Computing,* **8**(2), 131–142.

[34] McKeown, G., Valstar, M., Cowie, R., Pantic, M., & Schroder, M. (2012). "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent." *IEEE Transactions on Affective Computing,* **3**(1), 5–17.

[35] Ahmad, I., Hussain, S.I., Raja, M.A.Z., Shoaib, M. and Qurratulain, 2023."Transportation of hybrid MoS 2–SiO 2/EG nanofluidic system toward radially stretched surface." *Arabian Journal for Science and Engineering,* **48**(1), pp.953-966.

[36] C.F. Benitez-Quiroz, R. Srinivasan, A.M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR16), Las Vegas, NV, USA, 27-30 June 2016, pp. 5562-5570.

[37] A. Mollahosseini, B. Hasani, M.J. Salvador, H. Abdollahi, D. Chan, M.H. Mahoor, "Facial expression recognition from world wild web," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Las Vegas, NV, USA, 26 June-1 July 2016, pp. 58-65.

[38] A. Dhall, R. Goecke, T. Gedeon, "Automatic group happiness intensity analysis," *IEEE Trans. Affect. Comput.,* vol. **6**, pp. 13-26, 2015.

[39] M. Sharif, F. Naz, M. Yasmin, M.A. Shahid, A. Rehman, "Face recognition: A survey," *J. Eng. Sci. Technol. Rev.,* vol. **10**, pp. 166-177, 2017.

[40] Butt, Z.I., Ahmad, I., Raja, M.A.Z., Hussain, S.I., Shoaib, M. and Ilyas, H., 2023. Neuro-Heuristic Computational Intelligence Approach for Optimization of Electro-Magneto-Hydrodynamic Influence on a Nano Viscous Fluid Flow. International Journal of Intelligent Systems, 2023.

[41] J.Y.R. Cornejo, H. Pedrini, F. Flórez-Revuelta, "Facial expression recognition with occlusions based on geometric representation," in Iberoamerican Congress on Pattern Recognition, Springer, Cham, Switzerland, 2015, pp. 263-270.

[42] J. Mahata, A. Phadikar, "Recent advances in human behaviour understanding: A survey," in Proceedings of the Devices for Integrated Circuit (DevIC), Kalyani, India, 23-24 March 2017, IEEE, Piscataway, NJ, USA, 2017, pp. 751-755.

[43] M.H. Siddiqi, R. Ali, A.M. Khan, E.S. Kim, G.J. Kim, S. Lee, "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Multimed. Syst.,* vol. **21**, pp. 541-555, 2015.

[44] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, M. Pantic, "Deep structured learning for facial action unit intensity estimation," *in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition,* 2017.

[45] Hussain, S.I., Ahmad, I., Raja, M.A.Z. and Umer, C.M.Z., 2023. "A computational convection analysis of SiO2/water and MoS2-SiO2/water based fluidic system in inverted cone. " *Engineering Reports,* p.e12660.

[46] Zeng, N.; Zhang, H.; Song, B.; Liu, W.; Li, Y.; Dobaie, A.M. (2018). "Facial expression recognition via learning deep sparse autoencoders." *Neurocomputing,* **273**, 643–649.

[47] Kim, D. J. (2016). "Facial expression recognition using ASM-based post-processing technique." *Pattern Recognition and Image Analysis,* **26**(3), 576–581.

[48] Sajjad, M.; Nasir, M.; Ullah, F.U.M.; Muhammad, K.; Sangaiah, A.K.; Baik, S.W. (2019). "Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services." *Information Sciences,* **479**, 416–431.

[49] Al-Agha, S.A.; Saleh, H.H.; Ghani, R.F. (2017). "Geometric-based feature extraction and classification for emotion expressions of 3D video film. Journal of Advanced Information Technology," **8**, 74–79.

[50] Chang, K.Y.; Chen, C.S.; Hung, Y.P. (2013). "Intensity rank estimation of facial expressions based on a single image. In Proceedings of the IEEE International Conference on Systems, "*Man, and Cybernetics (SMC), Manchester, UK,* 13–16 October 2013; pp. 3157–3162.

[51] Ghimire, D.; Lee, J.; Li, Z.N.; Jeong, S. (2017). "Recognition of facial expressions based on salient geometric features and support vector machines." *Multimedia Tools and Applications,* **76**, 7921–7946.

[52] Ko, B. C. (2018). "A brief review of facial emotion recognition based on visual information." *Sensors (Switzerland)*, **18**(2), 20.

[53] Rasamoelina, A. D., Adjailia, F., & Sincak, P. (2019). "Deep Convolutional Neural Network for Robust Facial Emotion Recognition." *In IEEE International Symposium on Innovations in Intelligent Systems and Applications.*

[54] Tang, H., Liu, W., Zheng, W. L., & Lu, B. L. (2017). "Multimodal Emotion Recognition Using Deep Neural Networks. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)."

[55] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2020). "Generative adversarial networks. Advances in Neural Information Processing Systems," **63**(11), 139–144.

[56] Rangulov, D., & Fahim, M. (2020). "Emotion Recognition on large video dataset based on Convolutional Feature Extractor and Recurrent Neural Network. In International Conference on Image Processing," *Applications and Systems.*

[57] Li, J.; Wang, Y.; See, J.; Liu, W. Micro-expression recognition

based on 3D flow convolutional neural network. Pattern Anal. Appl. 2019, **22**, 1331–1339.

[58] Shan, K., Guo, J., You, W., Lu, D., & Bie, R. (2017). "Automatic facial expression recognition based on a deep convolutional-neural-network structure. In 15Th IEEE/ ACIS International Conference on Software Engineering Research," *Management and Applications.*

[59] Li, Y.; Zeng, J.; Shan, S.; Chen, X. Occlusion aware facial expression recognition using CNN with attention mechanism. IEEE Trans. Image Processing 2018, **28**, 2439–2450.

[60] Jain, D.K.; Zhang ZHuang, K. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. Pattern Recognit. Lett. 2020, **139**, 157–165.

[61] Chu, W.S.; De La Torre, F.; Cohn, J.F. Learning spatial and temporal cues for multi-label facial action unit detection. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 25–32.

[62] Al-Shabi, M.; Cheah, W.P.; Connie, T. Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator. arXiv 2016, arXiv:1608.02833.

[63] Yu, Z.; Zhang, C. Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015.

[64] Li, K.; Jin, Y.; Akram, M.W.; Han, R.; Chen, J. Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. Vis. Comput. 2020, **36**, 391–404.

[65] Minaee, S., Minaei, M., & Abdolrashidi, A. (2021). "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors," **21**(9), 3046. https://doi.org/10.3390/s21093046.

[66] Ucar, A., Demir, Y., & Guzelis, C. (2016). "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering." *Neural Computing and Applications,* **27**(1), 131–142.

[67] Butt, Z. I., Ahmad, I., Hussain, S. I., Raja, M. A. Z., Shoaib, M., & Ilyas, H. (2024). "Intelligent computing paradigm for unsteady magneto nano-polymeric Casson nanofluid with Ohmic dissipation and thermal radiation." *Chinese Journal of Physics.*