GMJ Global Mainstream Journal of Innovation, Engineering & Emerging Technology

Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

Exploring Convolutional Neural Networks for Facial Expression Recognition: A Comprehensive Survey

AKM Monzurul Islam [1]

[1] Department of Computer Science and Engineering, University of Scholars, Banani, Dhaka, Bangladesh

Abstract

Facial emotion recognition is a very sought-after field in recent times as it has many important applications such as effective calculations, video game testing and motion capture in video games, human-computer interaction through machine vision, computer research etc. Facial expression is considered a nonverbal form of communication, as it reveals an individual's internal sentiments and emotional states through changes in multiple facial landmark points. Facial identification provides a more comprehensive insight into the person's thoughts and these expressions are analyzed using deep learning methods, such as CNN. The accuracy rates achieved are compared to other methods. In this paper, a concise exploration of diverse applications within Facial Expression Recognition (FER) fields and the publicly accessible data sets employed in FER studies is outlined. FER using multiple different CNN algorithms is also presented. Finally, through comparing multiple different studies of various CNN algorithms, a table and a chart are provided for a better understanding of the rate of accuracy achieved throughout the use of different datasets.

Keywords: *Facial expression detection, Convolutional Neural Network, feature extraction, deep learning, pre-processing.*

Introduction

A facial expression is one or more motions or positions of the muscle tissues under the pores and skin of the face, consistent with one set of arguable theories, those movements bring the emotional country of an individual observer. Facial expressions are a form of nonverbal verbal exchange.

Facial expression recognition is vital for smooth human-computer interaction, enhancing communication between users and devices. Its diverse applications span across various domains such as healthcare, education, corporate marketing, and other sectors [1]. Mehrabian, a renowned psychologist from the United States, suggested that in everyday human communication, language and voice contribute to 7% and 38% of total information, respectively. In contrast, facial expressions convey a significant 55% of the overall information.[1]. Some of the key facial expressions consist of emotions like sadness, anger, joy, disgust, fear, surprise, and a neutral state. The understanding of human emotion through facial expressions holds significance in various applications such as human-computer interaction, robotics, and healthcare [7], [8], [9]. Facial Expression Recognition (FER) also finds applications in social contexts such as smart security, lie detection, and enhanced medical practices [8], [11]. However, accurately identifying facial expressions in challenging conditions, marked by factors like diverse facial positions, varying view angles, complex lighting conditions, facial wrinkles, and partial occlusions, remains a formidable task [12], [13].

CNN, or Convolutional Neural Network, represents a deep learning algorithm designed to process images directly, producing final classification results without requiring preliminary data processing. Various classifiers, including Local Binary Pattern (LBP) with SVM classification [2], Haar [3], SIFT [4], Gabor filters with Fisher linear discriminant [5], and local phase quantization (LPQ) [6], have been utilized in the past. Employing iterative methods like gradient descent, CNN autonomously learns to extract features directly from the training database, enhancing its capability for facial expression recognition.

This paper focuses on recent advancements in detecting facial emotions, employing various methods, algorithms, and architectures, particularly emphasizing the use of Convolutional Neural Networks (CNNs) for facial expression recognition. The review encompasses diverse subjects and contributions.

GMJ Global Mainstream Journal of Innovation, Engineering & Emerging Technology
Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

The paper is organized into four sections. The first provides a brief overview of Facial Expression Recognition (FER) implementation. The second delves into the fundamentals of CNN architecture. The third explores FER systems, analyzing ten studies using different CNN techniques for expression recognition, with a comparative table. The fourth section concludes by summarizing key findings from the study.

2. FACIAL EXPRESSION RECOGNITION

Within this segment, various systems and methodologies are employed for the categorization of fundamental human emotions through artificial intelligence algorithms, with a specific focus on the Facial Action Coding System (FACS) [14]. The Facial Expression Recognition (FER) architecture comprises three key components: pre-processing, feature extraction, and classification [15], [16], [17].

2.1 PRE-PROCESSING:

In the pre-processing phase, enhancements are made to refine the input data (image), minimizing duplication and improving quality [18]. The m x n-sized RGB image is converted to grayscale using the standard equation [19]. Facial contours are detected using the Haar Cascade picture library for face recognition [20], [21], [22]. Subsequently, facial expressions are cropped into rectangular regions and standardized to a consistent scale. To optimize neural network efficiency and avoid excessive density, pixel values are converted into 64x64 grayscale images [23]. Real-world data, captured under varying conditions, undergoes standardization, cropping, centralization, and additional pre-processing techniques to improve image recognition during experiments [24].

2.2 EXTRACTING FEATURES FOR FACIAL EXPRESSION RECOGNITION:

The subsequent phase in Facial Expression Recognition (FER) involves the critical process of feature extraction, where the goal is to identify and represent essential features within an image for further analysis. In the realm of image processing and computer vision, feature extraction plays a pivotal role as it signifies the transition from graphical to implicit data representation, serving as input for subsequent classification. Feature extraction methods fall into five main categories: texture feature-based, edge-based, global and local feature-based, geometric feature-based, and patch-based methods. Extracting new features from the original dataset is crucial to streamline resource usage during processing without overlooking relevant information [25].

During pre-processing, the input facial image undergoes transformation to facilitate the extraction of the most representative features [26]. The outcome of the feature extraction significantly impacts the system's output [16]. Traditional algorithms for facial feature extraction are broadly divided into two categories: geometric approaches, exemplified by methods like Active Appearance Models (AAM), and appearance-based methods, such as Gabor wavelet representation and Local Binary Pattern (LBP) [27].

Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

Geometric approaches consider parameters like position, angle, and points of reference, while appearance-based methods analyze the entire input image, extracting features that best describe it [17], [28], [29].

The features extracted from images are then used to train conventional classification methods, including Support Vector Machine (SVM), Discriminant Analysis Classification (LDA), and K-Nearest Neighbors (K-NN) [20], [30]. Throughout the feature extraction process, intricate properties are formed to clarify changes in facial form or texture modification. Various characteristics, including Local Directional Ternary Pattern (LDTP), Histogram of Oriented Gradients (HOG), Spatio-Temporal Texture Map (STTM), and LBP, are employed to define alterations in facial organ texture. These extracted features are subsequently used for the classification of facial expressions [31].

2.3 CLASSIFICATION:

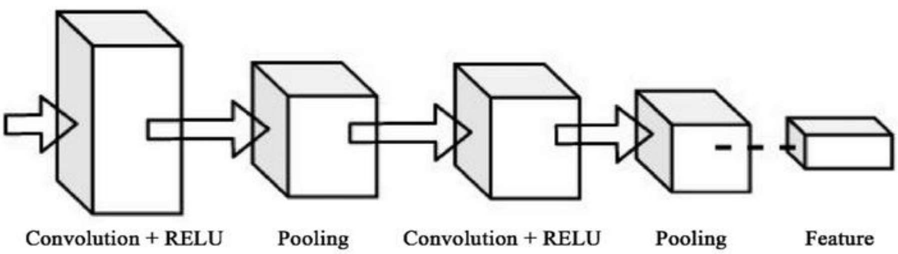
Concluding the Facial Expression Recognition (FER) process, the final step involves the precise mapping of labeled emotional action units [16]. In the domain of image classification, the recent success of deep learning (DL) algorithms has been evident [24], [32]. Diverse deep neural network architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Neural Networks (DNNs), have been explored for the identification of facial emotions [33]. Through an end-to-end model, this methodology trains a deep network structure with millions of parameters, enabling the extraction of accurate features from large datasets without the need for manually crafted features [15], [34].

To boost identification accuracy, commonly employed classification methods such as Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), and Linear Discriminant Analysis (LDA) are utilized [17], [29], [35]. Extracted features are then input into the classifier to recognize diverse facial expressions, with the efficiency of the classifier hinging on the consistency of the extracted features [27], [36]. Each classifier possesses distinct characteristics; SVM excels in target recognition and face detection applications, while the LDA method offers a means to distinguish between various groups [30].

3. CONVOLUTIONAL NEURAL NETWORK (CNN)

This section aims to elucidate the foundational principles of Convolutional Neural Networks (CNNs), a potent model widely applied in computer vision [24]. CNNs have proven highly effective in advancing the field of facial expression recognition, achieving commendable performance across various studies [10], [37], [38], [39]. An advantageous aspect of CNNs is their reduced need for pre-processing compared to alternative image classification algorithms [30]. Within the CNN framework, convolution is applied to each input filter, followed by a non-linear operation [40], [41].

Extending their utility beyond facial expression identification, CNNs find applications in diverse fields, including natural language processing [30]. The elementary structure of a CNN is characterized by three key layers: the convolution layer, the pooling layer, and the fully connected layer [42], [43]. This



Block diagram of CNN architecture.

GMJ Global Mainstream Journal of Innovation, Engineering & Emerging Technology
Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

succinct overview encapsulates the fundamental aspects of CNNs, highlighting their versatility in handling image-based tasks and pattern recognition across multiple domains.

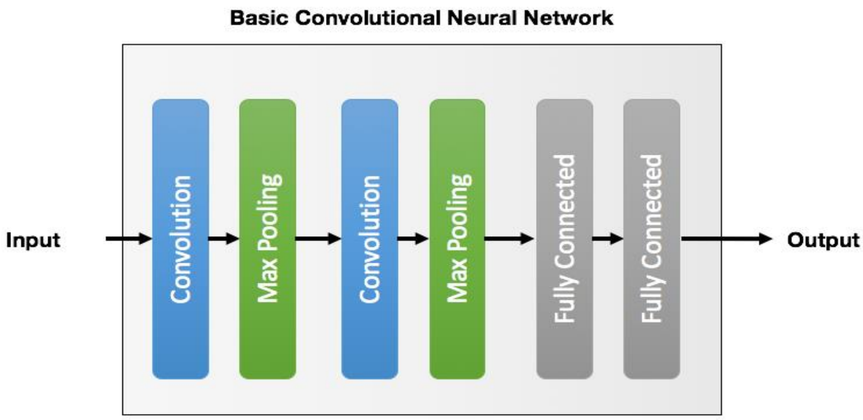


Fig: Basic CNN architecture

3.1 CONVOLUTIONAL LAYER:

This layer's primary role is to extract features from the input signal. Input images are processed by a set of learning neurons [44], and these features are then passed to the subsequent layer with neurons possessing trained weights and biases. The convolutional layer has distinct properties, with each neuron in a feature map connected to neighbors in the preceding layer. The local visual region of a neuron is termed an adjacent area from the last layer. The feature map is convolved with a learning kernel, and the output undergoes a non-linear activation function to compute a new feature map [43]. Multiple filters (kernels) convolve with the input image to extract features like edges [29], [46], [47]. The number of kernels corresponds to the feature maps obtained by filtering the preceding layer's feature map. Kernel strength directly impacts recognition accuracy and network learning. Determining the number of kernels is crucial and should be based on specific requirements [44], [48].

3.2 POOLING LAYER

It is viable to employ a computer-based screening approach to differentiate primary visual components and organize them into abstract visual features of variation. After sampling in the pooling layer, both the quantity and dimensions of output characteristic maps diminish [44]. The pooling layer functions to reduce the dimensionality of the input, utilizing either complete or average pooling to extract information [29]. The system utilizes a set of non-overlapping rectangles and employs a non-linear approach for querying within the space between convolution layers [40]. The predominant form of pooling is termed regular pooling, enhancing data representation by stacking multiple consecutive convolutional layers [43].

3.3 FULLY CONNECTED LAYER:

The preceding strata consist of a comprehensive layer of connections following multiple overlaps and pooling stages. Contrastingly, in the current strata, both neurons are mutually linked, and spatial information is not preserved across the entire connection layer [43], [49]. When amalgamating features obtained from the front, the fully connected layer integrates local knowledge along with a layer or pooling layer, contributing to category discrimination. Examples of fully connected layers encompass a fully linked input layer, the first connected layer, and a fully connected output layer. The flattening layer is often considered an interconnected input layer, transforming the output of preceding layers into

GMJ Global Mainstream Journal of Innovation, Engineering & Emerging Technology
Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

a singular vector. Following feature extraction from the convolutional layer, the network is linked to the upper connection layer, enhancing extraction capacity while constraining scalability [44], [50].

LITERATURE REVIEW OF FACIAL EXPRESSION RECOGNITION (FER)

This paper section critically examines the latest research in facial expression recognition, providing valuable insights into facial features, identification methodologies, model architectures utilized for feature extraction and classification, and the precision outcomes achieved by prominent researchers within the domain of FER.

FER using CNN with SIFT Aggregator:[51], the FER system utilized the SIFT aggregator and operated on two distinct datasets, namely FER-2013 and CK+. All images were standardized to dimensions of 48x48 pixels. Employing data augmentation enhanced the model's resilience to noise and minor alterations. The custom CNN network architecture comprised six convolutional layers and three Max-pooling layers, followed by two dense fully connected layers.

The SIFT method was employed to extract key points from facial images. The original image underwent processing through the CNN layers before being combined with either SIFT or DENSE SIFT features. Integration of SIFT and DENSE SIFT with CNN features aimed to enhance the overall performance of facial expression recognition. The SIFT descriptor was determined by partitioning the image into 4x4 squares, followed by grouping descriptors into clusters using K-means. The resulting K-vectors were then passed through a fully connected layer of size 4096, followed by dropout. Subsequently, the output was merged with the CNN model.

DENSE SIFT, in contrast, did not necessitate key point extraction. Instead, DENSE SIFT divided the image into equal pixel regions. To further enhance the model's accuracy, outputs from CNN, CNN with SIFT, and CNN with DENSE SIFT were aggregated using an average sum. The experimental results demonstrated that the proposed approach achieved an accuracy of 73.4% after aggregation on the FER-2013 dataset and an accuracy of 99.1% after aggregation on the CK+ dataset.

Facial Expression Recognition based on CNN: [52], the authors address the challenge of expression classification using a deep neural network built on Convolutional Neural Network (CNN) architecture. To achieve facial expression recognition, the paper emphasizes the necessity of face feature extraction. The Viola-Jones method is employed as the Face Detection (FD) algorithm, treating Haar characteristics as weak classifiers through a combination of AdaBoost and Haar face identification technologies.

To amass a diverse dataset, a web crawler, an automated program facilitating user requests for website data extraction, is utilized. Through this process, the authors collected a total of 35,887 facial expression images from the web, with 28,709 designated for the training set. Each image is formatted as a 400x500 gray-scale image with fixed-size pixels, featuring seven expressions: angry, disgusted, scared, happy, sad, surprised, and neutral.

The paper introduces a convolutional neural network model and employs an 80-20 split, randomly selecting 80% of the dataset as the training set and reserving the remaining 20% for testing purposes. The experimental results demonstrate that the proposed facial expression recognition model achieves a recognition rate exceeding 70% on the training set and surpasses 80% on the test set, indicating superior performance.

Using CNN for FER: A study of effects [53] Proposed in this study are two innovative convolutional neural network (CNN) structures, developed on the FER-2013 dataset, showcasing distinctive hyperparameter configurations throughout network layers. Both Model 1 and Model 2 architectures demonstrate human-level accuracy when applied to the FER-2013 dataset. Model 2, in particular, stands out for its streamlined design. The uniqueness of Model 1 lies in its utilization of a fixed kernel size and a specified number of filters across the network depth, with a reduction in the number of filters as the network deepens. Comparatively, Model 2 is more compact than Model 1. Both architectures employ a specific kernel size of 8, and the proposed models are compared against state-of-the-art benchmarks to determine their suitability for the FER-2013 dataset. The method presented in this paper achieves an accuracy rate of 65%.

A FER using CNN and LBP [54] Provided a thorough comparison of two prominent Facial Emotion Recognition (FER) techniques and illuminated their respective accuracies. The study focused on Local Binary Pattern (LBP) and Convolutional Neural Network (CNN) methodologies. LBP was employed solely for feature extraction, followed by classification using the Support Vector Machine (SVM) classifier. Findings revealed that CNN, utilizing its integrated softmax classifier, outperformed LBP. The achieved accuracy for the CK+ dataset was 97.32%, while for the YALE FACE dataset, it stood at 31.82%.

FER, using LGFD-based DCNN [55], A novel DCNN algorithm was devised to classify facial expressions by considering holistic factors. Prior to implementing the suggested Deep Belief Network (DCNN) model, a biological feature extractor was utilized to collect localized low-level information. The GF feature extractor (M and D) generated two local intermediate features. To ascertain the likelihood values of seven expressions, a Softmax classifier was employed at the conclusion of the proposed DCNN model. This process involves combining outputs using both measured and discrete data. According to empirical studies, success in Facial Expression Recognition (FER) missions is enhanced by the contributions of both local and holistic components. Despite achieving efficiency in a controlled laboratory setting, FER performance tends to be lower. Additionally, it is crucial to evaluate the proposed model in specific real-world scenarios. The datasets employed for testing included FER-2013, JFFE, CK+, KDEF, and RAF, with corresponding accuracies of 78%, 98%, 97.9%, 96.36%, and 82.33%, respectively.

FER with a 2-channel CNN [56] Utilized datasets such as JAFFE, CAE, Sobel-based, and McFIS to assess our neural network's performance in detecting facial expressions. Our network demonstrates the capability to achieve accurate results with minimal data compared to standard deep learning models, requiring less training effort. A notable limitation of the MCCNN involves the utilization of two predefined Sobel-based layers, particularly effective in a 3-channel topology. We conducted an evaluation of our architecture for facial expression recognition on the established JAFFE dataset, comprising 213 images depicting ten Japanese women expressing seven emotions (neutral, happiness, sadness, surprise, anger, disgust, and fear).

In our proposed architecture, the second channel incorporates two convolutional layers with max-pooling in between, sharing the same input as the primary channel. The MCCNN architecture employs multiple independent Convolutional Neural Networks in parallel, connecting them in the final layer to extract and identify patterns from images. Comparative analysis with benchmarks from existing literature underscores the state-of-the-art recognition rates achieved by our method for facial expressions. Significantly surpassing prior techniques using hand-crafted features, our approach yielded superior results.

GMJ Global Mainstream Journal of Innovation, Engineering & Emerging Technology
Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

Conclusively, averaging the performance measures across ten splits revealed that fear exhibited the lowest accuracy (90.7%) among all classes, while our method achieved a maximum accuracy of 98.6%. The combined accuracy scores for JAFFE (98.6%), CAE-based (95.8 ± 1.6), Sobel-based (93.1 ± 1.6), and McFIS [32] (87.6 ± 5.79) highlight the effectiveness of our proposed approach.

FER using CNN Ensemble [57] Utilizing the FER2013 dataset, our study successfully passed the AI plagiarism test. The approach presented is centered on a CNN-based solution for addressing Facial Expression Recognition (FER) and involves an empirical assessment of its efficacy. In contrast to a singular CNN, our approach surpasses performance by amalgamating and averaging outputs from diverse structured CNNs. The FER system is structured into three stages: firstly, face detection and localization; secondly, extraction of expression information from identified faces; and lastly, training a classifier, such as an SVM, on the extracted data to generate final expression labels. The evaluation of our method's effectiveness employs the recently released Countenance Recognition 2013 (FER-2013) dataset, which comprises 48 × 48-pixel grayscale facial images obtained through the Google Image search API with emotion-related keywords. Notably, human accuracy on this dataset is 65 ± 5%. We partitioned the FER2013 dataset into an 80% training set and a 20% validation set. Throughout this paper, we tackle the FER problem using a CNN ensemble model, featuring the design of three distinct structured CNN subnets. The comprehensive architecture, encompassing the outputs of these subnets, is introduced subsequently. Additionally, we train and assess the model's performance on the FER2013 dataset. A key advantage of our proposed model lies in the specialization of different CNNs, facilitating enhanced performance through the aggregation of results. Our achieved accuracy stands at 62.44%, with the overall model attaining 65.03% accuracy, securing the 9th and 5th positions, respectively, among all participating models.

Real-time FER using Deep learning [58], The FER system implementation utilized the FER2013 dataset sourced from the Kaggle challenge on Facial Expression Recognition (Goodfellow et al., 2013). Comprising 35,887 labeled images, the dataset was partitioned into 3,589 test and 28,709 train images. The Facial Expression Recognition process involved three key stages. Initially, the preprocessing phase involved formatting the dataset for compatibility with a generalized algorithm to yield efficient outcomes. Subsequently, the face detection stage employed OpenCV and Haar cascade classifiers, based on the method proposed by Papageorgiou et al., for real-time identification of faces in images. The final stage, emotion classification, utilized a CNN algorithm to categorize input images into one of seven expressions: Happiness, Sadness, Anger, Surprise, Disgust, Fear, and Neutral, as per the FER2013 dataset labels. The CNN training, a neural network category proven effective in image processing, yielded results indicating decreasing overtraining and test set loss across epochs. Employing a constant batch size of 256 for all experiments, the state-of-the-art model achieved a training accuracy of 79.89% and a test accuracy of 60.12%. The architecture successfully classified 22,936 out of 28,709 images from the train set and 2,158 out of 3,589 images from the test set, as illustrated in the accompanying table.

Attention Mechanism-based CNN for FER [59] It is recommended to incorporate an attention mechanism for a novel interpretation of facial expressions, where network attention layers analyze both raw images and Local Binary Pattern (LBP) characteristics. LBP features offer texture details, capturing subtle changes in skin texture crucial for recognizing gestures with nuanced differences. The researchers introduced the Nanchang University Facial Expressions (NCUFE) dataset, consisting of 490 photos depicting seven facial expressions (anger, disgust, fear, happiness, sadness, surprise, and neutral) taken by 35 individuals, captured in both RGB and depth. The study involves comprehensive research on five

Volume: 03
Issue: 02
ISSN ONLINE: 2834-2739
March, 2024
Texas, USA

datasets, including CK+, JAFFE, Oulu-CASIA, NCUFE, and FER2013. CK+ demonstrates an accuracy of 98.68%, NCUFE achieves 94.33% accuracy, and FER2013 records an accuracy of 75.82%.

Real-time Emotion Recognition [60] Proposed is a cost-effective and pragmatic strategy for real-time classification of seven distinct emotions (happy, sad, shocked, furious, hideous, terrified, and neutral) using the LeNet CNN architecture. Limited facial expression images have been utilized for CNN training, achieving a notable level of precision. The Haar Cascade library was employed to mitigate the impact of extraneous pixels beyond facial expressions. Additionally, optimizing pixel placement in image networks led to reduced training time and network complexity. The integration of three diverse datasets (KDEF, JAFFE, and a proprietary dataset) was implemented. Notably, the assessment and testing performance on the custom database surpassed that of training on existing datasets. The real-time testing model demonstrates the capability to analyze any secondary image. The overall accuracy, post-merging the three datasets (JAFFE, KDEF, and custom), stands at an impressive 96.43%.

RESULT OF ANALYSIS AND CRITICAL COMPARISON REVIEW

The presented data in the table illustrates that various studies employed different techniques and architectures of Convolutional Neural Networks (CNNs) to enhance accuracy rates. Notably, investigations conducted by researchers ([51], [52], [53], [54], [55], [56], [57], [58], [59], and [60]) were primarily centered around CNN methodologies. In [60], the author introduced a LeNet pooling layer—an older architecture—to address robust occlusion in Facial Expression Recognition (FER). This study amalgamated three datasets (JAFFE, KDEF, and custom), employing LeNet within a CNN framework, achieving an accuracy of 96.43%. In [55], the utilization of Deep CNN (DCNN) with a Softmax classifier was employed to develop dependable systems, yielding an average accuracy of 98.6% using the JAFEE dataset. [51] [61] investigated the impact of combining Scale-Invariant Feature Transform (SIFT) and Dense SIFT with CNN features on facial expression recognition. Additionally, a novel classifier for facial expression recognition was designed by amalgamating various CNN and SIFT models, showcasing state-of-the-art results on both the FER-2013 and CK+ datasets. Among the five researchers ([51], [53], [55], [57], [58]), [51] attained the highest accuracy percentage using the FER-2013 dataset. In contrast, [58] reported the lowest accuracy percentage of 60.12% based on the FER-2013 dataset in the examined research papers.

Discussion

Author	**Method**	**Performance (Accuracy)**	**Data Type /Dataset**
Mundher Al et al.[51]	CNN with SIFT & CNN with DENSE SIFT(aggregated)	99.1%	FER-2013, CK+
Mingjie Wang et al, Pengcheng, Xin.[52]	CNN, Viola-Jones method	70%	web crawler
Agrawal et al.[53]	CNN	65%	FER-2013
Ravi et al.[54]	LBP and CNN SVM	CK+ 32% YALE FACE 31.82%	CK+ and YALE FACE

Mohan et al. [55]	DCNN Softmax classifier	FER213 78%, JFFE 94%, Ck+ 979%, KDEF 96.36%, RAF 82.33%	FER2013, JFFE, CK+, KDEF, RAF
Dennis Hamster, Pablo Barros, Stefan et al.[56]	MCCNN	98.6%	JAFFE dataset
Kuang Liu1, Minming, Zhigeng et al.[57]	SVM CNN	65.03%	FER-2013
Isha Talegaonkar, Kalyani, Shreya, Rucha , Anagha et al.[58]	Haar cascade CNN	60.12%	FER2013
Li et al.[59]	LBP	NCUFFE 94.33%, CK+ 98.68%, FER-2013 75.82%	CK+, JAFFE, Oulu-CASIA, NCUFE
Ozdemir et al.[60]	LeNet CNN	96.43%.	KDEF, JAFFE, and their custom dataset(merged)

Fig: Table of Comparison

Conclusions

Facial expression recognition (FER) is recognized as a potent method for gaining insights into emotions. However, its efficacy is often confined to the recognition of six fundamental emotions and neutral expressions. This article delves into contemporary FER analysis techniques, examining multiple CNN architecture models. The observed recognition rates fluctuate across different datasets due to the utilization of diverse architectural models. Notably, each algorithm studied exhibits varying degrees of efficiency when applied to similar datasets, underscoring the nuanced performance of these FER methods.

Exploring Sustainable Initiatives and Prospects for the Future

Researchers with shared interests in this specific field can conduct additional studies on the advancement of human-robot interaction development and applications based on facial expressions, among other related areas.

References

[1] - Mingjie Wang , Pengcheng Tan , Xin Zhang , Yu Kang , Canguo Jin5, Jianying Cao1.: Facial expression recognition based on CNN (J. Phys). Conf. Ser. 1601 052027 (2020)

[2] - Shan, C. et al.: Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study. Image Vis. Comput. 27, 6, 803–816 (2009).

[3] -Whitehill, J., Omlin, C.W.: Haar Features for FACS AU Recognition. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06). p. 5 pp.-101 (2006).

[4] - Berretti, S. et al.: A Set of Selected SIFT Features for 3D Facial Expression Recognition. In: 2010 20th International Conference on Pattern Recognition (ICPR). pp. 4125–4128 (2010).

[5] - Liu, C., Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE Trans. Image Process. 11, 4, 467–476 (2002).

[6] - Wang, Z., Ying, Z.: Facial Expression Recognition Based on Local Phase Quantization and Sparse Representation. In: 2012 Eighth International Conference on Natural Computation (ICNC). pp. 222–225 (2012).

[7] - Bailly, K. and Dubuisson, S. (2017). Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. IEEE Transactions on Affective Computing, 10, 167-181

[8] - Liu, D., Ouyang, X., Xu, S., Zhou, P., He, K. and Wen, S. (2020). SAANet: Siamese action-units attention network for improving dynamic facial expression recognition. Neurocomputing, 413, 145-157

[9] - Zhi, R. and Wan, M. (2019). Dynamic Facial Expression Feature Learning Based on Sparse RNN. 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 1373-1377

[10] - Wen, G., Chang, T., Li, H. and Jiang, L. (2020). Dynamic Objectives Learning for Facial Expression Recognition. IEEE Transactions on Multimedia, 22, 2914-2925

[11] - Lasri, I., Solh, A. R. and El Belkacemi, M. (2019). Facial Emotion Recognition of Students Using Convolutional Neural Network. 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 1-6

[12] - Sun, X., Xia, P., Zhang, L. and Shao, L. (2020). An ROI-guided deep architecture for robust facial expression recognition. Information Sciences, 522, 35-48

[13] - Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C.-H., Xiang, Y., et al. (2019). A review of automatic facial expression recognition systems assisted by multimodal sensor data. Sensors, 19, 1863

[14] - Babajee, P., Suddul, G., Armoogum, S. and Foogooa, R. (2020). Identifying Human Emotions from Facial Expressions with Deep Learning. 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), 36-39

[15] - Chen, J., Lv, Y., Xu, R. and Xu, C. (2019). Automatic social signal analysis: Facial expression recognition using difference convolution neural network. Journal of Parallel and Distributed Computing, 131, 97-102

[16] - Ekundayo, O. and Viriri, S. (2019). Facial expression recognition: a review of methods, performances and limitations. 2019 Conference on Information Communications Technology and Society (ICTAS), 1-6

[17] - Ganapathy, N., Veeranki, Y. R. and Swaminathan, R. (2020). Convolutional Neural Network-based Emotion Classification using Electrodermal Activity Signals and Time-Frequency Features. Expert Systems with Applications, 113571

[18] - Ekundayo, O. and Viriri, S. (2019). Facial expression recognition: a review of methods, performances and limitations. 2019 Conference on Information Communications Technology and Society (ICTAS), 1-6

[19] - Kaviya, P. and Arumugaprakash, T. (2020). Group Facial Emotion Analysis System Using Convolutional Neural Network. 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), 643-647

[20] - Mohan, K., Seal, A., Krejcar, O. and Yazidi, A. (2020). Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks. IEEE Transactions on Instrumentation and Measurement, 70, 1-12

[21] - Sulaiman, D. M., Abdulazeez, A. M., Haron, H. and Sadiq, S. S. (2019). Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization. 2019 International Conference on Advanced Science and Engineering (ICOASE), 82-87

[22] - Othman, G. and Zeebaree, D. Q. (2020). The Applications of Discrete Wavelet Transform in Image Processing: A Review. Journal of Soft Computing and Data Mining, 1, 31-43

[23] - Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R. and Akan, A. (2019). Real-Time Emotion Recognition from Facial Expressions Using CNN Architecture. Medical Technologies Congress (TIPTEKNO), 1-4

[24] - Bargshady, G., Zhou, X., Deo, R. C., Soar, J., Whittaker, F. and Wang, H. (2020). Enhanced deep learning algorithm development to detect pain intensity from facial expression images. Expert Systems with Applications, 149, 113305

[25] - Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends, l. 1, 56-70

[26] - Le, T. T. Q., Tran, T.-K. and Rege, M. (2020). Dynamic image for micro-expression recognition on region-based framework. 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 75-81

[27] - Li, J., Jin, Zhou, K., Kubota, N. and Ju, Z. (2020). Attention Mechanism-based CNN for Facial Expression Recognition. Neurocomputing, 411, 340-350

[28] - Chinchanikar, N. A. (2019). Facial Expression Recognition Using Deep Learning: A Review. International Research Journal of Engineering and Technology (IRJET), 6, 3274-3281

[29] - Hussein, E. S., Qidwai, U. and Al-Meer, M. Emotional Stability Detection Using Convolutional Neural Networks, 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 136-140

[30] - Fei, Z., Yang, E., Li, D. D.-U., Butler, S., Ijomah, W., Li, X., et al. (2020). Deep convolution network-based emotion analysis towards mental health care. Neurocomputing, 388, 212-227, 2020

[31] - Baddar, W. J., Lee, S. and Ro, Y. M. (2019). On-the-Fly Facial Expression Prediction using LSTM Encoded Appearance-Suppressed Dynamics. IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2957465

[32] - Najat, N. and Abdulazeez, A. M. (2017). Gene clustering with partition around mediods algorithm based on weighted and normalized Mahalanobis distance. 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 140-145

[33] - Lee, J. R. H. and Wong, A. (2020). TimeConvNets: A Deep Time Windowed Convolution Neural Network Design for Real-time Video Facial Expression Recognition. 17th Conference on Computer and Robot Vision (CRV), 9-16

[34] - Zeebaree, D. Q., Abdulazeez, A. M., Hassan, O. M. S., Zebari, D. A. and Saeed, J. N. (2020). Hiding Image by Using Contourlet Transform, The Mattingley Publishing Co., Inc 83, 16979 - 16990

[35] - Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H. and Hamed, H. N. A. (2020). Improved Threshold and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. IEEE Access, 8, 203097-203116

[36] - Deng, L., Wang, Q. and Yuan, D. (2019). Dynamic Facial Expression Recognition Based on Deep Learning. 2019 14th International Conference on Computer Science & Education (ICCSE), 32-37

[37] - Chen, L., Ouyang, Y., Zeng, Y. and Li, Y. (2020). Dynamic facial expression recognition model based on BiLSTM-Attention. 15th International Conference on Computer Science & Education (ICCSE), 828-832

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

[39] - Kong, F. (2019). Facial expression recognition method based on deep convolutional neural network combined with improved LBP features. Personal and Ubiquitous Computing, 23, 531-539

[40] - Meryl, C. J., Dharshini, K., Juliet, D. S., Rosy, J. A. and Jacob, S. S. (2020). Deep Learning based Facial Expression Recognition for Psychological Health Analysis. 2020 International Conference on Communication and Signal Processing (ICCSP), 1155-1158

[41] - Abdulazeez, A., Salim, B., Zeebaree, D. and Doghramachi, D. (2020). Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol, 2020

[42] - Ozcan, T. and Basturk, A. (2020). Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization. Multimedia Tools and Applications, 79, 26587-26604

[43] - Nie, Z. (2020). Research on facial expression recognition of robots based on CNN convolutional neural network. 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), 1067-1070

[44] - Cheng, S. and Zhou, G. (2020). Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network. International Journal of Pattern Recognition and Artificial Intelligence, 34, 2056003

[45] - Verma, M., Vipparthi, S. K., Singh, G. and Murala, S. (2019). LEARNet: Dynamic imaging network for microexpression recognition. IEEE Transactions on Image Processing, 29, 1618-1627

[46] - Wu, M., Su, W., Chen, L., Pedrycz, W. and Hirota, K. (2020). Two-stage Fuzzy Fusion Based-Convolution Neural Network for Dynamic Emotion Recognition. IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2020.2966440

[47] - Zebari, N. A., Zebari, D. A.D., Zeebaree, Q. and Saeed, J. N. (2021). Significant features for steganography techniques using deoxyribonucleic acid: a review. Indonesian Journal of Electrical Engineering and Computer Science, 21, 338-347

[48] - Zeebaree, D. Q., Haron, H., Abdulazeez, A. M. and Zebari, D. A. (2019). Trainable Model Based on New Uniform LBP Feature to Identify the Risk of Breast Cancer. 2019 International Conference on Advanced Science and Engineering (ICOASE), 106-111

[49] - Bargarai, F., Abdulazeez, A., Tiryaki, V. and Zeebaree, D. (2020).. Management of Wireless Communication Systems Using Artificial Intelligence-Based Software Defined Radio. International Journal of Interactive Mobile Technologies (iJIM), 14, 107-133

[50] - Zeebaree, D. Q., Haron, H., Abdulazeez, A. M. and Zebari, D. A. (2019). Machine learning and Region Growing for Breast Cancer Segmentation. 2019 International Conference on Advanced Science and Engineering (ICOASE), 88-93

[51] - Mundher Al-Shabi, Wooi Ping Cheah, Tee Connie. Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator.International Workshop on Multi-disciplinary Trends in Artificial Intelligence. DOI:10.1007/978-3-319-69456-6_12 (2017)

[52] - - Mingjie Wang , Pengcheng Tan , Xin Zhang , Yu Kang , Canguo Jin5, Jianying Cao1.: Facial expression recognition based on CNN (J. Phys). Conf. Ser. 1601 052027 (2020)

[53] - Agrawal, A. and Mittal, N. (2020). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. The Visual Computer, 36, 405-412

[54] - Ravi, R. and Yadhukrishna, S. (2020). A face expression recognition using CNN & LBP. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 684-689

[55] - Mohan, K., Seal, A., Krejcar, O. and Yazidi, A. (2020). Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks. IEEE Transactions on Instrumentation and Measurement, 70, 1-12

[56] -Dennis Hamester, Pablo Barros, Stefan et al. Face Expression Recognition with a 2-Channel Convolutional Neural Network, IEEE, INSPEC Accession Number: 15503791, DOI: 10.1109/IJCNN.2015.7280539, Electronics ISBN: 978-1-4799-1960-4, PoD (2015)

[57] - Kuang Liu1, Minming, Zhigeng et al. Facial Expression Recognition with CNN Ensemble. IEEE, INSPEC Accession Number: 16581018, DOI: 10.1109/CW.2016.34, Electronics ISBN: 978-1-5090-2303-5, PoD ISBN: 978-1-5090-2304-2 (2016)

[58] - Isha Talegaonkar, Kalyani, Shreya, Rucha, Anagha et al. Real-Time Facial Expression Recognition using Deep Learning. SSRN Electronic Journal, DOI:10.2139/ssrn.3421486 (2019)

[59] - Li, J., Jin, Zhou, K., Kubota, N. and Ju, Z. (2020). Attention Mechanism-based CNN for Facial Expression Recognition. Neurocomputing, 411, 340-350

[60] - Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R. and Akan, A. (2019). Real-Time Emotion Recognition from Facial Expressions Using CNN Architecture. Medical Technologies Congress (TIPTEKNO), 1-4

[61] - Shamim, M. M. I., & Khan, M. H. (2022). Cloud Computing and AI in Analysis of Worksite. *Nexus*, *1*(03).