



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025

DOI: <https://doi.org/10.22214/ijraset.2025.71935>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Face Recognition Using Deep Learning

Vivek Kumar¹, Ritik Raj², Dr.Suveg Moudgil³

¹Scholar, ²Professor, School of Computer Science & Engineering, Galgotias University, Greater Noida

Abstract: Facial emotion recognition, an essential aspect of artificial intelligence (AI) and computer vision (CV), focuses on detecting and interpreting human emotions through facial expressions. This paper presents an approach utilizing deep learning (DL) models, including convolutional neural networks (CNN), binary-temporal scale convolutional neural networks (DTSCNN), recurrent neural networks (RNN), and residual networks (ResNet-50), to achieve real-time and accurate emotion recognition. The main goals of the study are real-time recognition, high accuracy, minimal latency, and resilience to diverse conditions. Experiments conducted on benchmark datasets evaluate each model based on accuracy, processing speed, and facial orientations. This paper compares the performance of these models and highlights that the CNN model outperforms others, offering superior precision and robustness. This research contributes to advancing facial emotion recognition with implications for human-computer interaction, psychology, marketing, and healthcare. The CNN ensemble model represents a significant step forward in facial emotion recognition, providing a comprehensive solution with broad applications in various fields. Its effectiveness underscores the importance of continuous research and refinement of deep learning techniques for solving complex tasks in AI and CV.

Keywords: Facial emotion recognition, deep neural networks, automatic recognition, database

I. INTRODUCTION

Affective computing is a multidisciplinary field aimed at developing systems that can recognize and interpret human emotions (Banafa, 2016). It appears that you have not provided any text to paraphrase. Please provide the text you'd like me to work on. The primary motivation for research in this area is the ambition to replicate empathy, allowing machines to identify and react to users' emotional conditions. Facial expressions play a vital role in communication, expressing intricate emotional information during exchanges. In nonverbal communication, the face acts as an essential medium for conveying feelings (Darwin and Prodger, 1996).

Using machine learning methods like facial recognition, facial expressions can be examined to gather emotional information and assess a person's emotional condition. Affective computing focuses on recognizing and interpreting emotional states to enhance the quality of interactions between humans and machines. Systems possessing this ability might produce responses that are more pertinent and flexible, considering the emotional context of users (Banafa, 2016).

The applications of affective computing are wide-ranging. In marketing, understanding emotions is necessary for assessing the impact of advertisements or products on consumers. Many organizations are investing in affective computing technologies to address issues like workplace stress and to create video games that adapt based on players' emotional responses.

This paper examines the creation of software that identifies user emotion through the use of computer vision methods and artificial intelligence algorithms. It also examines multiple emotional theories and approaches for assessing emotions via various algorithms. The software emphasizes emotion recognition using computer vision, especially through the application of convolutional neural networks (CNNs). Facial recognition utilizes the Multitask Cascade Convolutional Networks (MTCNN) framework, combining emotional theories with various assessment techniques to effectively evaluate emotional states. Facial recognition technology has evolved over the past few decades from a conventional method reliant on outdated machine learning techniques to the deep learning methods that are prevalent today. Previous methods, like PCA-based Eigenfaces, established a groundwork for the existing state-of-the-art by employing statistical techniques to map facial images into a reduced dimensional space. As suggested by Turk and Pentland in 1991, Eigenfaces facilitated the extraction of facial traits but were highly vulnerable to changes in lighting and various facial expressions.

A. Facial Databases Accessible

The success of facial emotion recognition (FER) systems largely depends on the quality and variety of the datasets used for training and evaluation. Numerous publicly available datasets have been developed, each offering unique features related to emotion categories, annotation quality, demographic diversity, and image acquisition techniques.

This part explores several well-known facial emotion databases that have significantly influenced FER research.

B. FER2013 (Facial Expression Recognition 2013)

The FER2013 dataset was introduced at the ICML2013 Challenge and consists of 35,887 grayscale images showing facial expressions, with each image having a resolution of 48×48 pixels. The dataset is divided into training, public testing, and private testing subsets, with the images organized into seven emotional categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. It is frequently used due to its large size and accessibility. However, the method of collecting data—using the Google image search API—introduces noise and potential mislabeling due to the automated nature of retrieving images.

C. Enlarged Cohn-Kanade (CK+) Dataset

The CK+ dataset acts as a standard in facial expression recognition (FER), consisting of 593 sequences from 123 subjects. Each sequence begins with a neutral expression and concludes with a peak face. Emotions are divided into seven categories: rage, scorn, aversion, dread, joy, sorrow, and astonishment. CK+ features detailed Action Unit (AU) annotations based on the Facial Action Coding System (FACS), enabling accurate analysis of expressions. While CK+ is known for its structured and tidy format, it is limited by a relatively small number of participants and a lack of ethnic diversity.

D. JAFFE (Japanese Women's Facial Expression Dataset)

The JAFFE dataset contains 213 images posed by 10 Japanese female models, each exhibiting seven facial expressions: happiness, sadness, surprise, anger, disgust, fear, and neutrality. The images are evaluated by Japanese participants, offering personal emotion labels. Owing to its straightforwardness and transparency, JAFFE is often employed for preliminary evaluations of FER models. Nonetheless, its constraints involve a limited dataset size and absence of demographic diversity.

E. AffectNet

AffectNet is among the largest datasets of facial expressions, comprising more than 1 million images sourced from the web. It comprises both categorical labels (eight distinct emotions) and dimensional annotations (valence and arousal scores). The dataset presents considerable variety regarding age, ethnicity, gender, and surrounding circumstances. While some annotations were produced by automated systems, a significant portion has been manually checked, rendering it appropriate for both classification and regression tasks in FER.

F. EmotioNet

The EmotioNet dataset includes more than 1 million facial images tagged with action units and emotional categories. It is intended to examine the identification of facial expressions in natural settings. A portion of the dataset has been annotated by hand, whereas most depends on automated AU detectors. This extensive database aids in creating FER systems that can identify both simple and complex facial expressions.

G. RAF-DB (Real-world Affective Facial Expressions Database)

RAF-DB comprises more than 29,000 authentic facial images annotated with fundamental and complex emotions. Labels are sourced from the crowd and cross-verified to guarantee equality. The database is especially remarkable for its incorporation of nuanced and intricate emotions, rendering it perfect for detailed emotion recognition tasks.

H. Oulu-CASIA NIR&VIS Facial Expression Dataset

The Oulu-CASIA database includes videos and static images of individuals exhibiting six fundamental facial expressions under different lighting scenarios and across both visible and near-infrared wavelengths. It comprises information from 80 subjects, obtained in carefully regulated settings. This dataset is especially valuable for research that focuses on multimodal emotion recognition and illumination-invariant facial analysis.

I. Comparative Examination

Compact datasets like CK+ and JAFFE are advantageous for preliminary benchmarking because of their precise annotations and regulated environments. In contrast, extensive datasets such as AffectNet and FER2013 are vital for training deep learning models that generalize more effectively across various real-world situations.

Additionally, the presence of Action Unit annotations in datasets such as EmotioNet and CK+ allows for the investigation of hybrid models that merge AURECognition with emotion classification.

II. COMPARISON AND RESEARCH GAP

Traditional machine learning methods were mostly used in the early years of facial recognition research. Principal Component Analysis (PCA) was used in techniques such as Eigenfaces, which were introduced by Turk and Pentland in the early 1990s, to lower the dimensionality of face images. This technique made it simpler to identify and categorize facial photos by projecting them into a feature space. Eigenfaces, however, frequently performed worse in uncontrolled settings due to their extreme sensitivity to changes in lighting and facial expressions. Based on Linear Discriminant Analysis (LDA), Fisherfaces then enhanced Eigenfaces by optimizing the ratio of within-class variation to between-class variance. As a result, Fisherfaces were more resilient to illumination variations than Eigenfaces, but they were still not completely resistant to all types of variability in real-world situations.

The introduction of deep learning brought about a dramatic change in the face recognition field.

Facebook's 2014 launch of DeepFace, which used a deep neural network trained on a sizable dataset to recognize faces with previously unheard-of precision, was one of the first projects in this field. DeepFace significantly outperformed conventional techniques, achieving near-human performance on the Labeled Faces in the Wild (LFW) benchmark.

Later, Google's FaceNet introduced the idea of face embeddings, pushing the limit even further. FaceNet mapped faces into a compact Euclidean space using a deep convolutional neural network (CNN), where facial similarities were directly correlated with distances. Face recognition was transformed by this method, which made identification and verification procedures scalable and effective.

Finally, the problem of cross-domain generalization is still open. Models often perform well on the specific datasets they are trained on but do not generalize well to new, unseen datasets or real-world environments. This limits the scalability and effectiveness of face recognition systems and points to the need for further research into more universal and adaptable algorithms.

These research gaps underscore the challenges of developing face recognition technology that is efficient, fair, secure, and ethically responsible, thereby paving the way for more advanced and reliable systems in the future.

III. IMPLEMENTATION

Face Emotion Recognition (FER) is a rapidly evolving area within AI and CV (computer vision) dedicated to determining human emotion through facial expressions. Recognizing emotions from facial cues is essential for various applications, including human-computer interaction, security, marketing, and healthcare. While early emotion recognition systems relied on manual feature extraction techniques, advancements in deep learning have been crucial in significantly improving the accuracy and efficiency of these systems.

The emergence of deep learning, especially Convolutional Neural Networks (CNNs), has transformed the field of FER. CNNs are capable of learning spatial hierarchies of features from images automatically, making them exceptionally effective for tasks like image classification. Unlike traditional methods, which necessitate handcrafted features such as the shape of the eyebrows or mouth, deep learning models can directly extract meaningful features from raw pixel data. This ability allows for the recognition of subtle facial expressions, which are crucial for detecting emotions like happiness, sadness, surprise, or anger.

Finally, emotion classification assigns an emotional label to the face based on the extracted features. The result is typically one of several predefined emotions, such as happiness, sadness, surprise, anger, disgust, or fear.

Training deep learning models for FER requires large, diverse, and accurately labeled datasets. Widely used datasets in this field include FER-2013, AffectNet, and EmotioNet, which contain millions of labeled images depicting various facial expressions. These datasets are essential for training models that can recognize emotions across different lighting conditions, angles, and ethnic backgrounds, thus enhancing the model's robustness and generalization. However, challenges remain in ensuring that models are not biased toward specific ethnicities or facial features.

A. CNN—convolutional neural networks

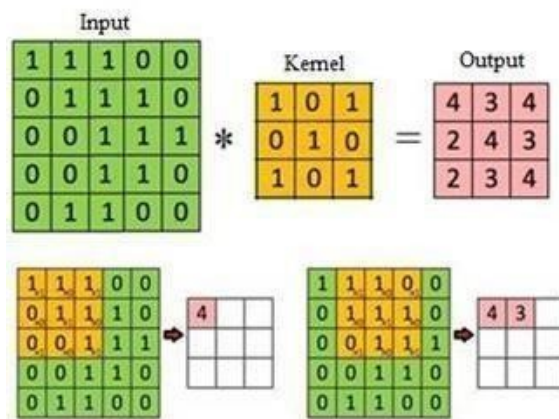
Convolutional Neural Networks (CNNs) consist of several layers of neurons, with the convolutional layer being the key component. This layer handles an input vector composed of pixel values, employing a range of filters that traverse the image to generate the layer's output (Simonyan and Zisserman, 2014.) achieved groundbreaking success by winning the 2012 ImageNet competition.



Fig. 1. Different deep learning methods proposed by Yichu et al. [28] and (b) Liang et al. [33]

(Convolution algorithm functions through the application of a 3x3 kernel.)

Along with the filter size utilized in the convolutional layers, two important parameters influence their function: stride and padding, which entails surrounding the output generated by the filters with zeros (Mathworks, 2023). Scattered among the convolutional layers are intermediate layers that assist in managing non-linearity. By maintaining dimensions, these layers improve the neural network's resilience and aid in minimizing the likelihood of overfitting. They frequently employ the ReLU activation function, which is more computationally efficient than conventional functions such as Sigmoid or Tanh (Thomas et al., 2005). As the network explores further into additional convolutional layers, the activation maps increasingly identify more intricate features over broader areas of the image. The initial layers excel at recognizing basic structures in specific regions, while deeper layers derive abstract representations by expanding on the previously identified features (Simonyan and Zisserman, 2014). A significant milestone in computer vision, the convolutional neural network known as AlexNet

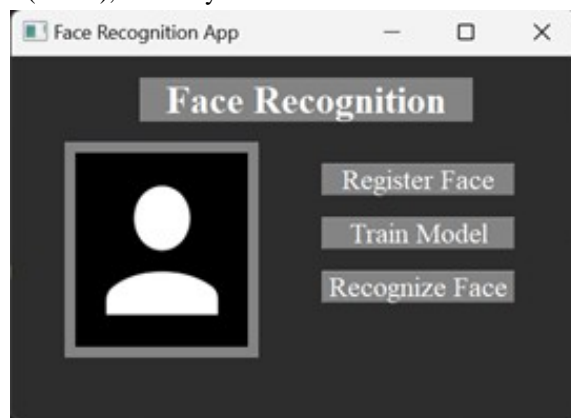


(Krizhevsky et al., 2017)

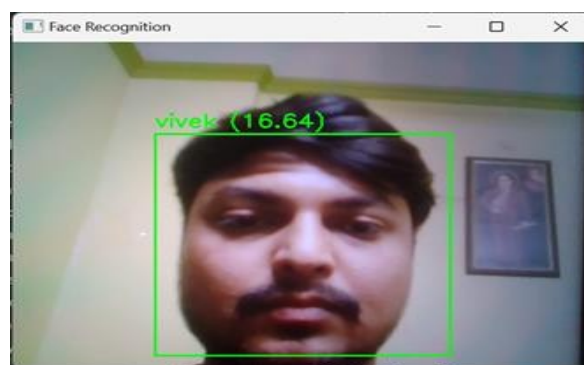
IV. RESULT

CNNs represent a type of neural network known for their remarkable effectiveness in tasks like face recognition and classification. CNNs are a form of multilayered, feed-forward neural networks.

CNNs are made up of neurons that have learning weights. Every filter accepts specific inputs, processes them, and then applies a non-linearity [41]. As illustrated in Fig. 1, a standard CNN architecture is demonstrated. The architecture of CNN comprises layers such as convolution, pooling, rectified linear unit (ReLU), and fully connected.



The convolution layer serves as the fundamental component of a convolutional network, handling the majority of intense computational tasks. The primary function of the convolution layer is to extract characteristics from input data that is image-based. By analyzing image features through small segments of the input image, convolution maintains the spatial relationship among pixels.

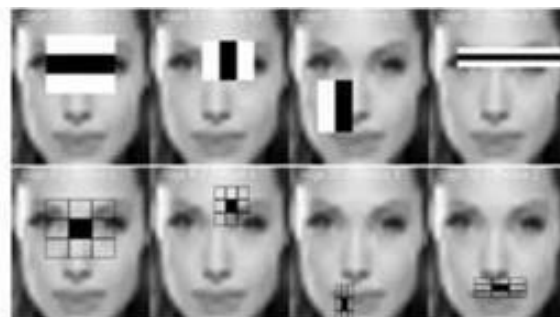


A collection of trainable neurons is employed to compress the input image.

This generates a feature map in the output image and subsequently supplies the feature maps to the next convolution layer as input data.

V. CONCLUSION AND FUTURE WORK

In this study, we examined the application of deep learning methods, particularly Convolutional Neural Networks (CNNs), along with the Haar Cascade algorithm for recognizing facial emotions (FER). Through the integration of these technologies, we demonstrated that emotion recognition systems can achieve high accuracy and efficiency, especially when combined with the robust face detection capabilities provided by Haar Cascade. The research highlights the growing potential of deep learning in enhancing FER systems and opens up avenues for future advancements in the field.



(Hair Characteristics)

The CNN-based approach is particularly effective for FER due to its ability to learn hierarchical features from raw image data. Unlike traditional methods that relied on manually crafted features, CNNs automatically capture intricate patterns in facial expressions, leading to better performance in recognizing emotions. Moreover, CNNs are highly adaptable and can be trained on large, diverse datasets, improving their generalization capabilities. The Haar Cascade algorithm, while relatively older, remains a widely used technique for face detection in FER systems due to its speed and efficiency. This combination of CNNs for emotion classification and Haar Cascade for face detection proves to be a powerful tool in real-world applications.

Despite these advancements, numerous challenges continue to exist in the field of FER. One of the major challenges is the handling of variations in facial expressions across different demographics, including age, gender, and ethnicity. Additionally, real-time emotion recognition remains a difficult task due to the computational demands of deep learning models, especially when applied to video streams or large datasets. Facial occlusions, such as glasses, masks, or hair, also pose difficulties for accurate emotion recognition. These issues can affect the overall performance of FER systems in real-world environments.

As we look ahead, there are many crucial aspects where enhancements can be implemented to improve the precision and resilience of FER systems. Initially, employing more sophisticated face detection algorithms like Multi-task Cascaded Convolutional Networks (MTCNN) could tackle some constraints of Haar Cascade, especially in identifying faces in difficult angles or with partial obstructions. Furthermore, incorporating attention mechanisms into CNN architectures might enhance the focus on key facial features, thereby boosting emotion recognition precision, particularly in instances of nuanced or blended expressions.

Regarding real-time processing, additional studies are required to enhance deep learning models for quicker and more effective emotion recognition. Methods like model pruning, quantization, and hardware acceleration (for instance, utilizing GPU or dedicated AI hardware) may facilitate quicker processing rates without compromising accuracy. Moreover, investigating the possibilities of transfer learning, which involves adjusting pre-trained models for particular tasks, may lessen the time and data needed to develop effective FER systems.

Finally, the application of FER systems in real-world scenarios, such as healthcare, education, and human-robot interaction, presents exciting opportunities. As FER technology continues to improve, it could be integrated into systems that provide emotional feedback, enhance user experiences, or assist in diagnosing emotional and psychological conditions.

However, ensuring the ethical use of FER technology, especially in sensitive contexts, will be crucial in shaping the future of this field.

In conclusion, while deep learning techniques, particularly CNNs combined with the Haar Cascade algorithm, have made significant strides in FER, there is still considerable room for improvement. By addressing current challenges and exploring new methodologies, future FER systems can become more accurate, inclusive, and capable of real-time, real-world applications.

REFERENCES

- [1] E.Sariyanidi, H.Gunes, et A.Cavallaro, « Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, Oct. 2014, doi: 10.1109/TPAMI.2014.2366127.
- [2] C.-N. Anagnostopoulos, T. Iliou, et D. Giannoukos, « Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011 », *Artif. Intell. Rev.*, vol. 43, no. 2, pp. 155-177, févr. 2015, doi: 10.1007/s10462-012-9368-5.
- [3] D. K. Jain, P. Shamsolmoali, et P. Sehdev, « Extended deep neural network for facial emotion recognition », *Pattern Recognit. Lett.*, vol. 120, p. 69-74, avr. 2019, doi: 10.1016/j.patrec.2019.01.008.
- [4] D.H. Kim, W.J. Baddar, J. Jang, et Y. M. Ro, « Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition », *IEEE Trans. Affect. Comput.*, vol. 10, no. 2, p. 223-236, Apr. 2019, doi: 10.1109/TAFFC.2017.2695999.
- [5] W. Zhao et al., « Face Recognition: Literature Survey », *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399-458, 2003.
- [6] K. Delac, *Recent Advances in Face Recognition*. 2008.
- [7] A. S. Tolba, A. H. El-baz, and A. A. El-Harby, « Face Recognition: A Literature Review », *Int. J. Signal Process.*, vol. 2, no. 2, pp. 88-103, 2006.
- [8] C. Geng and X. Jiang, « Face recognition using SIFT features », in *Proceedings - International Conference on Image Processing, ICIP*, pp. 3313-3316, 2009.
- [9] S.J. Wang, J. Yang, N. Zhang, and C. G. Zhou, « Tensor Discriminant Color Space for Face Recognition », *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2490-501, 2011.
- [10] S.N. Borade, R.R. Deshmukh, and S. Ramu, « Face recognition using fusion of PCA and LDA: Borda count approach », in *24th Mediterranean Conference on Control and Automation, MED 2016*, pp. 1164-1167, 2016.
- [11] M.A. Turk and A.P. Pentland, « Face Recognition Using Eigenfaces », *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 72-86, 1991.
- [12] M.O. Simón, « Improved RGB-D-T- based face recognition », *IET Biometrics*, vol. 5, no. 4, pp. 297-303, Dec. 2016.
- [13] O. Dniz, G. Bueno, J. Salido, and F. DeLaTorre, « Face recognition using Histograms of Oriented Gradients », *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598-1603, 2011.
- [14] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, « Robust face recognition via sparse representation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210-227, 2009.

- [16] C.Zhou,L.Wang,Q.Zhang,andX. Wei, “Face recognition based on PCA image reconstruction and LDA,” Opt. - Int. J. LightElectronOpt.,vol.124,no. 22, pp. 5599–5603, 2013.
- [17] Z.Lei,D.Yi,andS.Z.Li,“Learning Stacked Image Descriptor for Face Recognition,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 9,pp.1685–1696, Sep. 2016.
- [18] P. Sukhija, S. Behal, and P. Singh, “Face Recognition System Using Genetic Algorithm,” in Procedia Computer Science, vol. 85, 2016.
- [19] S. Liao, A. K. Jain, and S. Z. Li, “Partial face recognition:t-free approach,” IEEE Trans. Pattern Anal.Mach.Intell.,vol.35,no.5,pp. 1193–1205, 2013.
- [20] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Learning Deep Representation for Face Alignment with Auxiliary Attributes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 5, pp. 918–930, 2016.
- [21] G. B. Huang, H. Lee, and E. Learned-Miller, “Learning hierarchicalrepresentationsforface verification with convolutional deep belief networks,” in Proceedings of the IEEE Computer Society ConferenceonComputerVisionand PatternRecognition,pp.2518–2525, 2012.
- [22] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and D. Back, “Face recognition: a convolutional neural- network approach,” IEEE Trans. NeuralNetworks,vol.8,no.1,pp.98–113, 1997.
- [23] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,”inProceedingsofthe British Machine Vision Conference 2015, pp. 41.1-41.12, 2015.
- [24] Z.P.Fu,Y.N.Zhang,andH.Y. Hou, “Survey of deep learning in face recognition,” in IEEE International Conference on Orange Technologies,ICOT2014,pp.5–8,2014.
- [25] X.Chen,B.Xiao,C.Wang,X.Cai,Z. Lv, and Y. Shi, “Modular hierarchical feature learning with deep neural networks for face verification,” Image Processing (ICIP),201320thIEEEInternational Conference on. pp. 3690–3694, 2013.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)