International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 96-103, December-2023

https://doi.org/10.47001/IRJIET/2023.712014

An In-Depth Review of Deep Facial Expression Recognition: Obstacles, Utilizations, and Prospective Directions

[1]Raed Ibrahim Khaleel Almsari, [2]Abbas Hussein Miry, [3]Tariq M. Salman

[1]MSC Student, Electrical Engineering Department, Al-Mustansiriyah University, Baghdad, Iraq

[2,3]Assistant Professor Dr., Electrical Engineering Department, Al-Mustansiriyah University, Baghdad, Iraq

***Abstract* - Facial expression recognition (FER) is emerging as an emerging and multifaceted field of study. The use of FER in areas such as healthcare, security, and safe driving has not only enhanced the credibility of these technologies, but also their integration into human computer interaction to achieve intelligent outcomes. Computational FER seeks to replicate the skill of humans in decoding facial expressions, providing important cues that complement spoken language and aid listeners in understanding. Likewise, FER's deep learning (DL) and Artificial Intelligence (AI) methodologies are meticulously designed, incorporating advanced modules for efficiency and real time processing. In light of this background, many investigations have looked at different aspects of FER. Although current surveys focus primarily on traditional technologies and generic methodologies for on premises servers, they overlook the large field of deep learning inspired by edge vision and AI assisted FER technologies. To fill this gap, the current study conducts a comprehensive and thorough analysis of the prevailing FER literature. It carefully surveys the operational framework of FER technologies, highlighting their basic and intermediate phases, as well as the underlying pattern structures. Furthermore, the study addresses the limitations inherent in current FER surveys. The exploration extends to the FER datasets, subjecting them to thorough examination, thus revealing the attendant challenges and pitfalls. In addition, it provides a comprehensive discussion of the various metrics used to measure the effectiveness of FER methods.**

Keywords: Facial Emotion Recognition; Dataset; Deep Learning; Computer Vision.

I. INTRODUCTION

FER stands as a pivotal element in discerning an individual's emotions. Researchers in [1] have substantiated that the human countenance holds unparalleled adaptability in gauging engagement levels. Extensive exploration of FER has been undertaken within the realm of human psychology [2]. FER serves a myriad of purposes, spanning from human computer interaction intelligence and eLearning systems to marketing, stress analysis, and diverse interactive services. Such services, in order to remain pertinent, necessitate real time status updates. Thus, real time FER processing assumes a momentous role within the domain of computer vision.

Amidst this landscape, the utilization of convolutional neural network (CNN) models for FER has garnered considerable attention, primarily fueled by the leverage of extensive datasets for DL. While various survey papers have meticulously documented the evolution of FER techniques [3]. Researchers in [4] have also probed into real time FER methodologies. However, a noticeable void exists in the form of survey papers covering FER that harnesses few shot learning (FSL) technology, which stands as an alternative learning strategy. This technique exhibits the potential to address prevailing challenges in deep learning.

This study delves into recent FER systems that capitalize on FSL, a fact that bears the potential to mitigate existing challenges in the FER domain. Bringing matters to a close, the final section encapsulates the conclusions drawn from this comprehensive review. The uniqueness of this endeavor stems from the ensuing aspects:

- This denotes the debut survey paper focused on FER carried out through FSL procedures.
- We offer a thorough assessment of contemporary profound learning models applied to FER, focusing on the difficulties they wrestle with, deliberately arranging them in view of the idea of information issues and philosophies.
- We dive into the domain of FER through FSL by comparing two particular situations: the speculation to novel information and area transformation.
- We feature the potential for FER involving FSL to present time effectiveness and lessen unpredictability in various constant picture handling attempts.

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 96-103, December-2023
https://doi.org/10.47001/IRJIET/2023.712014

II. FER DATASETS

Numerous datasets catering to facial expression recognition (FER) are at our disposal within the FER domain. Nevertheless, our focus narrows down to a select subset of datasets, particularly those tailored for FER through FSL methodologies. In the realm of deep learning driven FER, the acquisition of an optimal volume of data stands as a pivotal challenge. This stems from the considerable variations that facial images exhibit due to factors like age, gender, and ethnicity. Datasets can be neatly bifurcated into 'in the lab' and 'in the wild' categories. An 'in the lab' dataset typically furnishes pristine, high quality data [2] alongside meticulously structured and annotated emotional cues, often mapped through the Facial Action Code [5], a system tailored for controlled environments. Conversely, 'in the wild' datasets encapsulate spontaneous facial expressions captured across diverse environmental contexts [6].

FER datasets can be further categorized into image based and video based datasets. The latter, characterized by sequences of emotions, enables a more comprehensive information accrual compared to single emotion images.

The onset of initiatives like the Emotion Recognition in the Wild Challenge (EmotiW) [7] has facilitated the access to a plethora of unconstrained facial images sourced from the digital realm. However, it's noteworthy that FER models formulated within controlled environments often demonstrate performance degradation when applied to the real world, where unscripted and sequential images prevail. Consequently, a thrust has been placed on the exploration of deep learning-based models to harness real world datasets.

Within this segment, we delve into diverse datasets pertinent to FSL centric FER tasks. The FER 2013 dataset [8], FER+ [9], AffectNet [5], Moreover, the Reality Full of Emotion Faces Dataset (RAFDB) [13] comprises images depicting six primary emotions (anger, disgust, fear, joy, sadness, and surprise), accompanied by a neutral sentiment. In contrast, the extensive Cohn-Kanade (CK+) dataset [10] comprises seven fundamental emotions (anger, contempt, disgust, fear, joy, sadness, and surprise). Additionally, considering the challenge of domain shift assumption frequently faced in Few-Shot Learning (FSL), several experiments involve training on datasets containing basic emotions and subsequently evaluating on datasets containing complex emotions. As a result, we also delve into the RAFDB and the Compound Looks of Emotion (CFEE) dataset [11], both encompassing a spectrum of compound emotions such as 'sadly surprised' and 'fearfully angry'. Table 1 summarizing the characteristics of each dataset.

Table 1: Emotion datasets summary

Dataset	Contents	Emotion Categories	Data Collection	Annotations	Size	Usage
CK+ [10]	593 sequences 123 subjects	7 basic emotions + contempt	Lab controlled	Neutral to peak expression	Medium	Widely used for FER evaluation
CFEE [11]	21 emotion categories	Compound emotions + basic categories	Various production methods	Facial Action Coding System	Small	Study of compound emotions
FER2013 [8]	35,887 images 3 groups	6 basic expressions + neutral	Google API & OpenCV	Human labelers	Large	Training, validation, testing
FER+ [9]	Improved FER dataset	Standard emotions + probabilities	Crowdsourced taggers	10 labelers per image	Medium	Multi-label algorithms
EmotioNet [12]	1 million facial images	Various expressions	AU detection model	Automatic & manual annotation	Large	AU detection model evaluation
AffectNet [2]	1,000,000+ facial images	8 basic expressions	Internet search	Manual annotation	Large	Categorical & dimensional models
RAFDB [13]	29,672 diverse images	Common expressions + compound	Internet download	Not specified	Medium	Real world expressions
AFEW [14]	957 videos	6 basic expressions + neutral	Collected from movies	Manual annotation	Small	Challenging environment analysis

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 96-103, December-2023

https://doi.org/10.47001/IRJIET/2023.712014

III. FACE FEELING ACKNOWLEDGMENT TECHNIQUES AND ADVANCES

The process of identifying facial emotions involves a series of distinct stages, starting from face detection and culminating in emotion classification. Initially, face detection involves isolating the face from its surroundings through the identification of key facial features such as the eyes, nose, and mouth. Once this step is completed, the system progresses to the task of categorizing the emotions [15].

3.1 Face Detection

Face detection is a foundational concept in computer vision that involves identifying and locating human faces within images or video frames [16]. This process relies on algorithms that analyze the visual patterns and unique attributes associated with human faces, such as the arrangement of eyes, nose, mouth, and other facial features. Effective face detection is essential for various applications, including facial recognition, emotion analysis, and even in digital cameras for focusing and exposure adjustment. Modern face detection methods often utilize machine learning techniques, particularly convolutional neural networks (CNNs), to achieve high accuracy and robustness in detecting faces across different poses, lighting conditions, and backgrounds. [17][18].

3.2 Feature Extraction

Face image feature extraction is a crucial process in computer vision that involves distilling meaningful and distinctive information from facial images. It aims to capture key characteristics such as facial expressions, shapes, textures, and structures. By converting complex facial data into a more compact and representative form, feature extraction facilitates efficient analysis and recognition tasks. Various techniques, including deep learning approaches like convolutional neural networks (CNNs) and traditional methods like Principal Component Analysis (PCA), are employed to extract these features. The extracted features play a vital role in applications ranging from facial emotion recognition and identity verification to surveillance and human-computer interaction. [15].

3.3 Data Augmentation

Data augmentation aims to improve emotion classification outcomes by enhancing the dataset. Common practices include altering contrast and brightness, as well as resizing images to a uniform format. Additional techniques involve filtering and edge detection, often using methods like the Sobel edge detection [17].

3.4 Emotion Classification

Emotion classification is a significant aspect of both psychology and artificial intelligence, aiming to decipher and categorize human emotional states based on various cues such as facial expressions, voice tone, and physiological signals. In the realm of AI, emotion classification involves training models to recognize and differentiate between different emotional states, such as happiness, sadness, anger, fear, and surprise, among others [19]. This is achieved through supervised learning techniques, where machine learning algorithms learn from labeled datasets containing examples of different emotions. These models are then capable of analyzing input data, such as images or audio recordings, and assigning them to appropriate emotion categories. Emotion classification finds applications in diverse fields including human-computer interaction, sentiment analysis, mental health monitoring, and the development of empathetic AI systems [20].

3.5 Real Time Emotion Recognition

Real-time emotion recognition is a dynamic technology that involves the instantaneous detection and interpretation of human emotional states as they occur. Utilizing various data sources such as facial expressions, speech patterns, and physiological signals, real-time emotion recognition systems employ sophisticated algorithms to rapidly analyze and categorize emotions like happiness, sadness, anger, and surprise [21]. This technology holds immense potential in applications like video conferencing, virtual reality, and human-computer interaction, where understanding and responding to users' emotions in real time can enhance user experience and engagement. Real-time emotion recognition is challenging due to the need for speed and accuracy, often requiring advanced machine learning models and efficient processing pipelines to ensure seamless and meaningful emotion interpretation in dynamic contexts [22].

IV. STATE OF THE ART TECHNOLOGIES, MEASURES, AND MODELS

Several cutting-edge technologies, measures, and models contribute to advanced emotion recognition systems. Table 2 describing each of the mentioned technologies, methods, and models.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 96-103, December-2023

https://doi.org/10.47001/IRJIET/2023.712014

Table 2: Advanced technologies, methods, and models for facial emotion recognition

Name	Description	Key Features
CAERNet	A model for context aware emotion recognition that considers both facial and contextual cues [23].	Focuses on face and attentive context regions. Enhances emotion recognition with contextual information.
SASEFE	Dataset featuring congruent and incongruent facial expressions, revealing insights into various emotional types [23].	Provides congruent and incongruent expressions. Offers understanding of complex emotional types.
Microsoft Hololens	Hardware device with a depth camera for 3D face detection, utilizing the Microsoft Azure Face API application [23].	Employs depth camera for accurate 3D detection. Integrates with Azure Face API for enhanced functionality.
Multimodal Systems	Integrates image, video, and voice inputs for improved emotion determination accuracy [24]	Combines various input modalities for comprehensive analysis. Increases accuracy by leveraging multiple sources of data.
Strrn (Spatial Temporal Recurrent Neural Network)	Deep learning framework combining spatial and temporal features for signal sources [24].	Incorporates spatial temporal feature learning. Enables analysis of complex signal data.
Stationary Wavelet Entropy	Extracts facial expression features using stationary wavelet entropy [25].	Applies wavelet based method for feature extraction. Captures distinctive features from expressions.
Island Loss	Technique enhancing discriminative features by reducing intra class variations and increasing interclass differences [11].	Focuses on enhancing feature separability. Improves emotion representation within features.
Deeply Supervised CNN	Merges models like Feed Forward Neural Network and Naïve Bayes for heightened accuracy [15].	Combines different models for improved classification. Enhances accuracy through collaborative learning.
Facial Expression Sentence Generating Model	Generates descriptive sentences to interpret recognized facial emotions [26].	Converts emotion recognition into human in temper table language. Provides comprehensive descriptions of emotional states.
Compound Emotion Recognition	Identifies complex emotions formed by combining basic emotions [27].	Detects and labels multifaceted emotional states. Expands emotional recognition beyond basic categories.

V. TRANSFER LEARNING IN FER

Face emotion recognition based on transfer learning is a powerful and widely adopted approach in the field of computer vision and affective computing. This methodology leverages pre-trained deep neural networks, often originating from large-scale image classification tasks, to extract meaningful features from facial images and subsequently infer emotional states. Transfer learning capitalizes on the knowledge acquired by these pre-trained models and adapts it to the specific task of emotion recognition [28].

The process typically involves fine-tuning the pre-trained model on a dataset that contains labeled facial expressions, where each image is associated with a particular emotion such as happiness, sadness, anger, fear, disgust, or surprise. During fine-tuning, the weights of the network are adjusted to optimize its performance on the emotion recognition task. This allows the model to learn relevant facial features, such as the arrangement of facial landmarks, the shape of the mouth, or the intensity of eye expressions, that are indicative of different emotional states [29].

One key advantage of transfer learning in face emotion recognition is the ability to achieve high accuracy even when dealing with limited training data, as the model has already learned general features from a vast amount of data during pre-training. This reduces the risk of overfitting and can lead to more robust emotion recognition systems. Additionally, transfer learning enables researchers and developers to avoid the computational cost and time associated with training deep neural networks from scratch, making it a practical and efficient approach for real-world applications such as human-computer interaction, sentiment analysis, and healthcare [30].

However, the choice of the pre-trained model, the fine-tuning strategy, and the selection of an appropriate emotion recognition dataset all play crucial roles in the success of this approach. The field continues to evolve, with ongoing research aimed at improving the performance, robustness, and generalization capabilities of transfer learning-based face emotion recognition systems, making it an exciting area of study with significant potential for various applications in both academia and industry [31].

VI. RELATED WORKS FER

The field of facial expression recognition has witnessed significant advancements through a multitude of studies focusing on various datasets, feature extraction techniques,

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 96-103, December-2023
https://doi.org/10.47001/IRJIET/2023.712014

and classification methods. Notably, in the work presented by authors [32], they employed the Local Fisher Discriminant Analysis (LFDA) for feature extraction from the JAFFE and MUG datasets, achieving remarkable recognition rates of 94.37% and 95.24%, respectively, using a 1-nearest-neighbor classification approach. In a similar vein, researchers [33] concentrated solely on the JAFFE dataset, utilizing Gabor filters for feature extraction and Bayesian classification techniques, yielding an impressive recognition rate of 96.73%. Extending the scope to multiple datasets, [34] harnessed Gabor techniques for both JAFFE and Yale datasets, coupling them with a neural network back-propagation algorithm. This hybrid approach led to recognition rates of 96.83% for JAFFE and 92.22% for Yale.

Furthermore, the application of diverse feature extraction methods remained a key avenue of exploration. For instance, [35] delved into the combination of Gabor wavelet transform, Principal Component Analysis (PCA), and Local Binary Pattern (LBP) for feature extraction from JAFFE, demonstrating a recognition rate of 90% using the k-Nearest-Neighbor (k-NN) classification technique. A departure from image-based techniques, [36] focused on the CK+ dataset and introduced kernel PCA (KPCA) for feature extraction, attaining recognition rates of 76.5% and 72.3% for KPCA and PCA, respectively.

Classifiers also played a pivotal role in achieving robust recognition performance. [37] adopted an Eigen face approach with Euclidean distance as the classification metric, resulting in an average recognition rate of 85.38%. In contrast, [38] explored Active Shape Models (ASM) in conjunction with RBF kernel Support Vector Machines (SVM) and Hidden Markov Models (HMM), achieving recognition rates of 70.6% and 65.2%, respectively. Noteworthy is the innovative use of Biorthogonal Wavelet Entropy (BWE) by [39], coupled with Fuzzy Multiclass SVM (FMSVM), attaining an impressive recognition rate of 96.77% ± 0.10%.

Incorporating audio signals into facial expression recognition, [40] combined Gabor filtering for images and Mel-Frequency Cepstral Coefficients (MFCC) for audio, achieving recognition rates of 84.68% for the CK dataset, 80.68% for the Berlin dataset, and 81.58% in real-time scenarios using SVM. In the era of deep learning, [41] employed Convolutional Neural Networks (CNN) to process the JAFFE and CK+ datasets, achieving recognition rates of 76.7442% and 80.303%, respectively.

Additionally, [42] introduced a hybrid feature set consisting of Gabor and Local Binary Pattern (LBP), and examined its effectiveness with various kernel SVMs, achieving impressive recognition rates ranging from 94.45% to 97.42% across different SVM kernels and class distributions. [43] further expanded the study across multiple datasets—CK+, JAFFE, and BU-3DFE—attaining recognition rates of 96.76%, 82.10%, and 82% respectively, through CNN-based approaches.

Even novel techniques like Discrete Wavelet Transform (DWT) found their place in facial expression recognition, as demonstrated by [44], who employed a Single-hidden-layer Neural Network (NN) for classification and achieved a recognition rate of 89.49% ± 0.76%.

Table 3: Related works summary

Paper Reference	Dataset	Feature Extraction	Classification	Recognition Rate
[32]	JAFFE, MUG	LFDA	1-NN	JAFFE: 94.37%, MUG: 95.24%
[33]	JAFFE	Gabor filter	Bayesian	96.73%
[34]	JAFFE, Yale	Gabor techniques	Neural network	JAFFE: 96.83%, Yale: 92.22%
[35]	JAFFE	Gabor wavelet, PCA, LBP	k-NN	90%
[36]	CK+	KPCA	KPCA, PCA	KPCA: 76.5%, PCA: 72.3%
[37]	Private	Eigen face	Euclidean distance	Avg: 85.38%
[38]	CK+	Active Shape Models	SVM, HMM	SVM: 70.6%, HMM: 65.2%
[39]	Private	Biorthogonal Wavelet Entropy	Fuzzy Multiclass SVM	96.77% ± 0.10%
[40]	CK, Berlin	Gabor, MFCC	SVM	CK: 84.68%, Berlin: 80.68%, Real-Time: 81.58%
[41]	JAFFE, CK+	CNN	CNN	JAFFE: 76.7442%, CK+: 80.303%

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 96-103, December-2023
https://doi.org/10.47001/IRJIET/2023.712014

[42]	CK+	Gabor, LBP	SVM (Linear, RBF, Poly)	See description
[43]	CK+, JAFFE, BU-3DFE	CNN	CNN	CK+: 96.76%, JAFFE: 82.10%, BU-3DFE: 82%
[44]	Private	DWT	Single-hidden-layer NN	89.49% ± 0.76%

VII. CONCLUSIONS

The comprehensive survey on deep facial expression recognition provides valuable insights into the rapidly evolving field and highlights various facets of the subject. The study begins by acknowledging the increasing significance of facial expression recognition (FER) in diverse domains, such as healthcare, security, and human-computer interaction. The integration of FER with deep learning and artificial intelligence techniques has propelled its capabilities, enabling real-time processing and intelligent outcomes. The survey underlines the importance of computational FER as a means of replicating human abilities in decoding facial expressions, offering essential cues that complement verbal communication and enhance understanding.

The survey brings to light the limitations of existing FER surveys, which often focus on traditional methods and technologies, disregarding the vast potential of deep learning and AI-assisted FER. To address this gap, the study extensively reviews the current literature on FER, taking into account the operational framework, phases, pattern structures, and challenges associated with FER technologies. This encompasses both the basic and intermediate stages of FER processes, enabling a holistic understanding of the field's advancements.

Furthermore, the survey dives into the realm of FER datasets, providing a thorough examination of the datasets commonly used in FER tasks. The challenges posed by variations in facial images due to factors like age, gender, and ethnicity are acknowledged. The study categorizes datasets into "inthelab" and "inthewild" categories, highlighting the distinctions between controlled and real-world environments. Moreover, the significance of multimodal datasets that incorporate image, video, and voice inputs is emphasized for improving accuracy.

The survey meticulously explores the technologies, methods, and models employed in FER, offering insights into face detection, feature extraction, data augmentation, emotion classification, and real-time FER. Notably, deep learning frameworks like Convolutional Neural Networks (CNNs) are identified as pivotal for their ability to analyze patterns within data and excel in image recognition tasks. The discussion on advanced technologies further underlines the importance of context-aware emotion recognition, multimodal systems, and novel approaches to handling complex and compound emotions.

A significant contribution of the survey lies in its focus on FER through few-shot learning (FSL), a relatively less-explored area in the existing literature. The study elucidates the potential of FSL to address challenges in deep learning-based FER and presents a comprehensive review of recent FER systems that leverage FSL techniques. This examination sheds light on the versatility of FSL in generalizing to novel data and adapting to different domains, thus improving the efficiency and applicability of FER methods.

REFERENCES

[1] J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, “The faces of engagement: Automatic recognition of student engagement from facial expressions,” IEEE Trans. Affect. Comput., vol. 5, no. 1, pp. 86–98, 2014, doi: 10.1109/TAFFC.2014.2316163.

[2] C. L. Kim and B. G. Kim, “Few-shot learning for facial expression recognition: a comprehensive survey,” J. Real-Time Image Process., vol. 20, no. 3, pp. 1–18, 2023, doi: 10.1007/s11554-023-01310-x.

[3] O. S. Ekundayo and S. Viriri, “Facial Expression Recognition: A Review of Trends and Techniques,” IEEE Access, vol. 9, pp. 136944–136973, 2021, doi: 10.1109/ACCESS.2021.3113464.

[4] S. Deshmukh, M. Patwardhan, and A. Mahajan, “Survey on real-time facial expression recognition techniques,” IET Biometrics, vol. 5, no. 3, pp. 155–163, 2016, doi: 10.1049/iet-bmt.2014.0104.

[5] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” IEEE Trans. Affect. Comput., vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.

[6] A. Dhall, O. V. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon, “Video and image based Emotion recognition challenges in the wild: EmotiW 2015,” ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact., no. December, pp. 423–426, 2015, doi: 10.1145/2818346.2829994.

[7] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, “From individual to group-level emotion recognition: Emoti W 5.0,” ICMI 2017 - Proc. 19th

ISSN (online): 2581-3048
Volume 7, Issue 12, pp 96-103, December-2023
https://doi.org/10.47001/IRJIET/2023.712014

ACM Int. Conf. Multimodal Interact., vol. 2017-Janua, pp. 524–528, 2017, doi: 10.1145/3136755.3143004.

[8] I. J. Goodfellow et al., “Challenges in representation learning: A report on three machine learning contests,” Neural Networks, vol. 64, pp. 59–63, 2015, doi: 10.1016/j.neunet.2014.09.005.

[9] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” ICMI 2016 - Proc. 18th ACM Int. Conf. Multimodal Interact., pp. 279–283, 2016, doi: 10.1145/2993148.2993165.

[10] 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,” 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010, no. May 2014, pp. 94–101, 2010, doi: 10.1109/CVPRW.2010.5543262.

[11] S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” Proc. Natl. Acad. Sci. U. S. A., vol. 111, no. 15, pp. 1–9, 2014, doi: 10.1073/pnas.1322355111.

[12] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, “EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, no. June 2016, pp. 5562–5570, 2016, doi: 10.1109/CVPR.2016.600.

[13] S. Li and W. Deng, “Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition,” IEEE Trans. Image Process., vol. 28, no. 1, pp. 356–370, 2019, doi: 10.1109/TIP.2018.2868382.

[14] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, “Collecting large, richly annotated facial-expression databases from movies,” IEEE Multimed., vol. 19, no. 3, pp. 34–41, 2012, doi: 10.1109/MMUL.2012.26.

[15] P. Naga, S. Das Marri, and R. Borreo, “Facial emotion recognition methods, datasets and technologies: A literature survey,” Mater. Today Proc., vol. 80, no. xxxx, pp. 2824–2828, 2023, doi: 10.1016/j.matpr.2021.07.046.

[16] A. S. Aljoloud, H. Ullah, and A. A. Alanazi, “Facial emotion recognition using neighborhood features,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 1, pp. 299–306, 2020, doi: 10.14569/ijacsa.2020.0110137.

[17] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment,” Procedia Comput. Sci., vol. 125, no. November 2022, pp. 2–10, 2018, doi: 10.1016/j.procs.2017.12.003.

[18] E. Dandıl and R. Özdemir, “Real-time Facial Emotion Classification Using Deep Learning,” Data Sci. Appl., vol. 2, no. 1, pp. 13–17, 2019.

[19] G. Tonguç and B. Ozaydın Ozkara, “Automatic recognition of student emotions from facial expressions during a lecture,” Comput. Educ., vol. 148, no. January, p. 103797, 2020, doi: 10.1016/j.compedu.2019.103797.

[20] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, and A. Akan, “Real time emotion recognition from facial expressions using CNN architecture,” TIPTEKNO 2019 - Tip Teknol. Kongresi, no. October, 2019, doi: 10.1109/TIPTEKNO.2019.8895215.

[21] A. Mahmood, S. Hussain, K. Iqbal, and W. S. Elkilani, “Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion,” Math. Probl. Eng., vol. 2019, p. 9185481, 2019, doi: 10.1155/2019/9185481.

[22] M. R. M. Shahane*, R. Sharma .K, and M. S. Siddeeq, “Emotion Recognition using Feed Forward Neural Network & Naïve Bayes,” Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 2, pp. 2487–2491, 2019, doi: 10.35940/ijitee.b7070.129219.

[23] S. Jyoti, G. Sharma, and A. Dhall, “Expression empowered residen network for facial action unit detection,” Proc. - 14th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2019, no. 1, pp. 1–8, 2019, doi: 10.1109/FG.2019.8756580.

[24] M. Malygina, M. Artemyev, A. Belyaev, and O. Perepelkina, Overview of the Advancements in Automatic Emotion Recognition: Comparative Performance of Commercial Algorithms. 2019. doi: 10.31234/osf.io/x7bvd.

[25] S. H. Wang, P. Phillips, Z. C. Dong, and Y. D. Zhang, “Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm,” Neurocomputing, vol. 272, no. September 2020, pp. 668–676, 2018, doi: 10.1016/j.neucom.2017.08.015.

[26] S. Palaniswamy and S. Tripathi, “Emotion recognition from facial expressions using images with pose, illumination and age variation for human-computer/robot interaction,” J. ICT Res. Appl., vol. 12, no. 1, pp. 14–34, 2018, doi: 10.5614/itbj.ict.res.appl.2018.12.1.2.

[27] S. Tripathi, S. Tripathi, and H. Beigi, “Multi-Modal Emotion recognition on IEMOCAP Dataset using Deep Learning,” arXiv1804.05788v3 [cs.AI] 6 Nov 2019 MULTI-MODAL, 2018.

[28] A. Pellacani, M. Graziano, and M. Suatoni, “Design ,

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 12, pp 96-103, December-2023

https://doi.org/10.47001/IRJIET/2023.712014

Development , Validation and Verification of GNC technologies," Eucass2019, no. July, p. 28760, 2019, doi: 10.13009/EUCASS2019-38.
[29] B. Li, "Facial expression recognition via transfer learning," EAI Endorsed Trans. e-Learning, no. December, p. 169180, 2018, doi: 10.4108/eai.8-4-2021.169180.
[30] S. Khandelwal, "Research Paper on Facial Emotion Recognition Using Transfer Learning," pp. 1–6.
[31] S. Sidharth, A. A. Samuel, H. Ranjana, J. T. Panachakel, S. P. K, and S. P. Jun, "Emotion Detection from EEG using Transfer Learning," arXiv:2306.05680v1, 2023.
[32] Y. Piparsaniyan, V. Sharma, and K. Mahapatra, Robust facial expression recognition using Gabor feature and Bayesian discriminating classifier. 2014. doi: 10.1109/ICCSP.2014.6949900.
[33] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," Expert Syst. Appl., vol. 41, no. 7, pp. 3383–3390, 2014, doi: 10.1016/j.eswa.2013.11.041.
[34] M. Abdulrahman, T. R. Gwadabe, F. J. Abdu, and A. Eleyan, "Gabor wavelet transform based facial expression recognition using PCA and LBP," 2014 22nd Signal Process. Commun. Appl. Conf. SIU 2014 - Proc., no. August 2015, pp. 2265–2268, 2014, doi: 10.1109/SIU.2014.6830717.
[35] D. K. Hu, A. S. Ye, L. Li, and L. Zhang, "Recognition of facial expression via kernel PCA network," Appl. Mech. Mater., vol. 631–632, pp. 498–501, 2014, doi: 10.4028/www.scientific.net/AMM.631-632.498.
[36] A. De, A. Saha, and M. C. Pal, "A human facial expression recognition model based on eigen face approach," Procedia Comput. Sci., vol. 45, no. C, pp. 282–289, 2015, doi: 10.1016/j.procs.2015.03.142.
[37] M. Suk and B. Prabhakaran, "Real-time facial expression recognition on smartphones," Proc. - 2015 IEEE Winter Conf. Appl. Comput. Vision, WACV 2015, no. November, pp. 1054–1059, 2015, doi: 10.1109/WACV.2015.145.
[38] Y. D. Zhang et al., "Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation," IEEE Access, vol. 4, no. c, pp. 8375–8385, 2016, doi: 10.1109/ACCESS.2016.2628407.
[39] K. Slimani, M. Kas, Y. El Merabet, R. Messoussi, and Y. Ruichek, "Facial emotion recognition," Int. J. Adv. Res. Comput. Eng. Technol., vol. 7, no. 11, pp. 88–94, 2018, doi: 10.1145/3177148.3180092.
[40] Y. Sun and J. Yu, "Facial expression recognition by fusing gabor and local binary pattern features," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10133 LNCS, no. January 2017, pp. 209–220, 2017, doi: 10.1007/978-3-319-51814-5_18.
[41] A. K. Katsaggelos, S. Bahaadini, and R. Molina, "Audiovisual Fusion: Challenges and New Approaches," Proc. IEEE, vol. 103, no. 9, pp. 1635–1653, 2015, doi: 10.1109/JPROC.2015.2459017.
[42] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," Pattern Recognit., vol. 61, no. October 2017, pp. 610–628, 2017, doi: 10.1016/j.patcog.2016.07.026.
[43] S.-H. Wang, W. Yang, Z. Dong, P. Phillips, and Y. Zhang, "Facial Emotion Recognition via Discrete Wavelet Transform, Principal Component Analysis, and Cat Swarm Optimization," Lect. Notes Comput. Sci., vol. 10559, pp. 203–214, Sep. 2017, doi: 10.1007/978-3-319-67777-4_18.
[44] E. BATTINI SÖNMEZ, "An Automatic Multilevel Facial Expression Recognition System," Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg., vol. 0, no. 0, p. 10, 2018, doi: 10.19113/sdufbed.50007.

Citation of this Article:

Raed Ibrahim Khaleel Almsari, Abbas Hussein Miry, Tariq M. Salman, "An In-Depth Review of Deep Facial Expression Recognition: Obstacles, Utilizations, and Prospective Directions" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 12, pp 96-103, December 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.712014

****