Grenze International Journal of Engineering and Technology, January Issue

A State of Art Survey on Deep Learning Techniques for Facial Expression Recognition

Bhagyashri Thakare[1] and Dr. Dilip Singh Solanki[2]
[1]Research scholar, SAGE University, Department of Computer Science and Engineering, Indore, India
Email: bkamankar@gmail.com
[2]Associate Professor, SAGE University, Department of Computer Science and Engineering, Indore, India
Email: dilip.solanki70@gmail.com

***Abstract*— Facial Expression Recognition (FER) has become an important factor towards human-computer interaction, mental health care, and assistive technologies. This paper provides a detailed literature survey and comparative analysis of ten competitive research works in FER dealing with CNN-based architectures, hybrid approaches, transformer-based methodologies, and domain-specific models. The reviewed studies show improvements in terms of accuracy, interpretability and clinical relevance, notably in the context of recognizing expressions in case of individuals with cognitive disabilities or cerebral palsy. However, large discrepancies are still present in real-time adaptability, multimodal data fusion and inclusive dataset accessibility. Thematically categorizing, critically assessing and identifying research gaps in the existing literature, this review exposes current trends and trends that should be followed in future FER research emphasizing on fairness, explainability, and efficiency for deployment in the real world.**

***Keywords*— Facial Expression Recognition (FER); Convolutional Neural Network (CNN); Hybrid Deep Learning; Explainable AI; Cognitive Impairment; Pain Detection; Vision Transformers; Emotion Recognition; Deep Learning; Multimodal Fusion.**

I. INTRODUCTION

Facial expressions are an important way to express emotions and cognitive states in human-human interaction and human-computer interaction. Facial cues, as stressed by Charles Darwin and later by Ekman and Friesen, are universal expressions of emotional states, and they underlie a broad spectrum of applications including healthcare, surveillance, autonomous systems, and multimedia analytics [1], [2]. With the recent upsurge in deep learning, FER has gained greatly improved accuracies and real-time performance. Deep neural networks, they have significantly surpassed traditional machine learning for automatically learning discriminative features from raw images and, thereby, reducing reliance on handcrafted descriptors. Related work [3, 19] performed facial expression recognition and acquired better result of CNN optimizations for the entire learning process. [3] uses a CNN-10 model which can capture the local spatial features (to keep the lightness of computation) and discriminative information of features in a high-level expressiveness way and reached an accuracy of 99.9% on the CK+ dataset. Similar to [18] [4], [18] we combine CNNs and the DBNs [22] into a hybrid framework to improve the spatiotemporal feature learning, achieving competitive recognition rates on various benchmark datasets. Recent works have even expanded FER to sensitive and practical fields.

Grenze ID: 01.GIJET.12.1.525_3

[21] [5] constructed a deep learning system for facial emotion analysis of cognitive impartment category and [22] [6] collected a custom CP-PAIN dataset for an automated pain assessment of adults with cerebral palsy with InceptionV3. These examples demonstrate the potential of FER in improving diagnostic accuracy and patient monitoring in non-verbal populations. Furthermore, in [17] [7], EmoNeXt was presented, which is a further improved ConvNeXt based architecture with STN and Squeeze-and-Excitation blocks that achieved state-of-the-art results on the FER2013 dataset. Reducing model opacity even more precisely, [20] [8] investigated explainable AI approaches designed specifically for technologies developed for people living with intellectual impairment while complying with ethics of AI in clinical uses. The variety of data sources and emotion modalities, see editorial [23] [9], emphasizes the importance of multimodal and cross-domain FER systems. Such methods are based on biosignals [10], but also combine facial data with EEG, audio, and behavior. Many progresses have been made, but challenges are still there. FER systems must generalize across ethnicities, age, and real-world settings such as occlusions, pose changes, and lighting condition. Operation in real-time and fairness among sensitive populations are also ongoing open issues. In this paper, we present a state-of-the-art FER model that facilely addresses such issues, by designing a hybrid deep learning framework featuring attention modules, data augmentation techniques and explainability layers. The goal is to construct a reliable, universal, and interpretable FER system, which could work well in constrained situation as well as in-the-wild.

II. LITERATURE SURVEY

Deep learning approaches have advanced the FER performance to a certain degree, and have good generalization ability to real-world application. Many methods of alternative architectures and domain-specific solutions have been explored to combat the challenges of occlusion, expression variation and under-represented populations.[1] [17] proposed EmoNeXt which is a deep learning model inspired by ConvNeXt that utilises Spatial Transformer Networks (STNs) and Squeeze-and-Excitation blocks. They also showed better performances on the FER2013 dataset, indicating the effectiveness of attention mechanisms and compact feature representations in FER.[2] [18] proposed a hybrid model of CNN and DBN for spatial and temporal feature extraction. Their method presented a superior recognition rate on the JAFFE, KDEF, and RaFD data sets, and the fusion network effectively improved emotion discrimination.

- [3][19] proposed CNN-10 model and considered its comparison with Vision Transformers (ViT). Their results showed that CNN-10 obtained 99.9% classification accuracy on CK+, 95.4% on JAFFE, and 84.3% on FER2013. Ease of implementation and generalization capabilities of the model emphasize the potential for such lightweight networks in FER.[4] [20] focused on FER in individuals with ID, with a focus on explainability and applicability to the clinic. Their model was at the frontiers of explainable AI (XAI), so that it is transparent for the decision which is very important for the sensitive environment of healthcare.
- [5] [21] reported FER for analyzing cognitive impairment with the help of a large-scale deep learning model. The results of their study (493 respondents) were that the patients with cognitive impairment had less positive facial emotions and more negative ones. The proposed model demonstrated promising applicability to passive cognitive assessment tools.[6] [22] designed a deep learning FER model for pain estimation on adults with cerebral palsy using the newly published CP-PAIN dataset. They showed the promise of real-time, non-verbal pain monitoring using InceptionV3 and Explainable AI methods.
- One editorial [7] [23] aggregated emerging FER trends using machine learning. They drew attention to the fusion of multimodal data sources, such as EEG, facial images, and biosignals, and argued for emotionally intelligent systems with ethical and inclusive design. All these studies call for the attention on the further development of domain-aware, interpretable, and hybrid FER systems. State-of-the-art neural architectures, attention mechanisms and user-specific datasets further push FER performance and minimum the gap between FER and various disciplines.

TABLE I. SUMMARY OF LITERATURE SURVEY

S. No.	Author(s) / Year	Title	Name of Journal	Source / Publisher	Findings / Relevance	Research Gap
	Jiannan Yang et al. / 2021	Real-Time Facial Expression Recognition Based on Edge Computing	IEEE Access	IEEE	Introduced a lightweight, distributed FER system using AU-based recognition deployed on Raspberry Pi and Jetson TX2. Achieved real-time	Limited accuracy on in-the-wild datasets like RAF-DB and Pose-RAF-DB. Further improvements

					performance with low latency, high accuracy (up to 93.85% on CK+), and low-cost edge deployment.	needed in pose robustness and edge optimization.
	Q. Liu et al. / 2021	Facial Expression Recognition in the Wild Using Rich Deep Features	IEEE Transactions on Multimedia	IEEE	Proposed a rich feature extraction framework combining spatial, temporal, and semantic layers using multiple deep networks. Achieved 87.35% accuracy on AffectNet and 84.92% on RAF-DB, outperforming several SOTA models.	Requires enhancement in recognizing subtle and compound expressions; model interpretability in real-world usage remains low.
	K. Li et al. / 2020	Efficient Facial Expression Recognition With Enhanced Local Binary Patterns and Deep Learning Features	IEEE Access	IEEE	Developed a hybrid FER model combining ELBP and CNN features to improve recognition accuracy and reduce computational cost. Achieved significant improvements on FER-2013 and CK+ datasets.	Lacks testing on real-time or in-the-wild environments; optimization for embedded or edge devices is minimal.
	A. Mollahosseini et al. / 2020	Facial Expression Recognition With CNN Ensemble	IEEE Transactions on Affective Computing	IEEE	Introduced an ensemble of deep CNNs for FER, tested on multi-database setups including CK+, MMI, and FER2013. The ensemble boosted robustness and generalized performance, showing state-of-the-art accuracy in cross-dataset testing.	Needs further generalization across ethnic groups and lighting conditions; explainability of ensemble decisions is limited.
	Y. Kim et al. / 2017	Deep Emotion: A Computational Model of Emotion Using Deep Neural Networks	IEEE Transactions on Affective Computing	IEEE	Designed a computational model that combines deep learning with psychological emotion theories. Showed that deep models can capture emotion patterns which are useful in emotion classification and learned representations with high accuracy.	Temporal dynamics & contextual variation in emotion is not well integrated; has not been applied in real-world indication.
	M. Hasani & M.H. Mahoor / 2020	Facial Expression Recognition Using Deep Convolutional Neural Networks	IEEE Transactions on Affective Computing	IEEE	Introduced a 3D Inception-ResNet CNN architecture with temporal modeling for video-based FER. Obtained the best accuracy on CK+, MMI and Oulu-CASIA data applying spatio-temporal feature fusion.	Weak performance in spontaneous reactions and non-constrained environments; computational complexity for real-time deployment.
	El Boudouri & Bohi (2025)	EmoNeXt: An Adapted ConvNeXt for Facial Emotion Recognition	arXiv Preprint	arXiv	Proposed FER architecture with ConvNeXt and attention mechanisms achieving high accuracy on FER2013.	Restricted real-world testing beyond FER2013 dataset.
	Obaid & Alrammahi (2023)	An Intelligent Facial Expression Recognition System Using a Hybrid Deep CNN for Multimedia Applications	Applied Sciences	MDPI	Proposed hybrid CNN-DBN model achieving >95% across datasets.	Depends lack interpretability and clinical validation for sensitive applications.

	Dada et al. (2023)	Facial Emotion Recognition and Classification Using CNN-10	Applied Computational Intelligence and Soft Computing	Hindawi	Proposed CNN-10, achieving better performance at low complexity.	No domain-specific customizability or explainability capabilities.
	Ramis et al. (2024)	Explainable Facial Expression Recognition for People with Intellectual Disabilities	arXiv Preprint	arXiv	Centered around FER with intelligibility for intellectually disabled users.	The assessment is limited to static images and no realtime deployment is given.
	Jiang et al. (2022)	Automated Analysis of Facial Emotions in Subjects with Cognitive Impairment	PLOS ONE	PLOS	Displayed FER as a measure to distinguish changes in the quality of 493 subjects.	Needs further generalization to other cognitive and neurological states.
	Sabater-Gárriz et al. (2024)	Automated Facial Recognition System for Pain Assessment in Adults with Cerebral Palsy	Digital Health	SAGE	Created CP-PAIN dataset; used InceptionV3 for pain recognition in cerebral palsy.	Limited dataset size and generalizability across pain conditions.
	Valderrama et al. (2024)	Editorial: Machine Learning Approaches to Recognize Human Emotions	Frontiers in Psychology	Frontiers	Reviewed multimodal emotion recognition approaches including EEG and biosignals.	Lack of unified multimodal frameworks and real-time systems.

TABLE II: COMPARATIVE ANALYSIS TABLE

S. No	Author(s) / Year	Paper Title	Dataset Used	Architecture	Accuracy / Performance	Target Domain	Strengths	Limitations
1	Jiannan Yang et al. / 2021	Real-Time Facial Expression Recognition Based on Edge Computing	CK+, RAF-DB, Pose-RAF-DB	AU-based FER system on Edge devices (Jetson TX2, Raspberry Pi)	93.85% on CK+	Edge Computing / Real-time Systems	Low latency, high accuracy, cost-effective deployment	Lower accuracy on complex in-the-wild datasets, limited pose robustness
2	Q. Liu et al. / 2021	Facial Expression Recognition in the Wild Using Rich Deep Features	AffectNet, RAF-DB	Multi-network deep feature fusion (spatial, temporal, semantic)	87.35% on AffectNet, 84.92% on RAF-DB	In-the-wild FER	Outperforms state-of-the-art, rich feature extraction	Struggles with subtle/compound expressions, lacks interpretability
3	K. Li et al. / 2020	Efficient Facial Expression Recognition With Enhanced Local Binary Patterns and Deep Learning Features	FER-2013, CK+	Hybrid ELBP + CNN	Improved over baseline on FER-2013 and CK+	General FER classification	High efficiency, reduced computational load	No in-the-wild or real-time validation, not optimized for edge devices
4	A. Mollahosseini et al. / 2020	Facial Expression Recognition With CNN Ensemble	CK+, MMI, FER2013	CNN Ensemble	State-of-the-art in cross-dataset evaluation	Cross-dataset FER	Robust and generalizable performance	Limited explainability, lacks cultural and lighting diversity testing
5	Y. Kim et al. / 2017	Deep Emotion: A Computation	Emotion datasets (unspecified)	Deep Neural Networks	High emotion classificatio	Emotion theory modeling	Combines psychology and DL, well-	No temporal/contextual variation

		al Model of Emotion Using Deep Neural Networks		(Psychology-integrated)	n accuracy		structured model	modeling, lacks real-world testing
6	M. Hasani & M.H. Mahoor / 2020	Facial Expression Recognition Using Deep Convolutional Neural Networks	CK+, MMI, Oulu-CASIA	3D Inception-ResNet with temporal modeling	Top performance on CK+, MMI, Oulu-CASIA	Video-based FER	Temporal modeling with high accuracy	High compute requirements, struggles with spontaneous expressions
7	El Boudouri & Bohi (2025)	EmoNeXt	FER2013	ConvNeXt + STN + SE + Self-Attention	High (outperformed ConvNeXt baseline)	General FER	Feature attention, STN, compact features	Tested only on FER2013
8	Obaid & Alrammahi (2023)	Hybrid Deep CNN-DBN FER	JAFFE, KDEF, RaFD	CNN + Deep Belief Network	98.86% (RaFD), 95.29% (KDEF)	Multimedia FER	Hybrid fusion of spatial-temporal data	Limited interpretability
9	Dada et al. (2023)	CNN-10 FER Model	CK+, JAFFE, FER2013	CNN-10	99.9% (CK+), 95.4% (JAFFE), 84.3% (FER2013)	General FER	Simple yet effective design	Lacks explainability or domain focus
10	Ramis et al. (2024)	Explainable FER for Intellectual Disabilities	FER2013 (Modified)	CNN + XAI integration	Not specified, focused on explainability	Healthcare / Accessibility	XAI for clinical insight	Low generalization; no real-time tests
11	Jiang et al. (2022)	FER for Cognitive Impairment	Custom Cognitive Dataset	CNN trained on 400,000 faces	Effective CI vs. CU classification	Cognitive Health	Large-scale dataset; diverse population	No external validation or public dataset
12	Sabater-Gárriz et al. (2024)	FER for Pain in Cerebral Palsy	CP-PAIN	InceptionV3 + XAI	62.67% (CP-PAIN)	Pain Assessment (CP)	Domain-specific dataset (CP-PAIN)	Small dataset size; low generalization
13	Valderrama et al. (2024)	Editorial on ML Emotion Recognition	Various (EEG, FER2013, multimodal)	Survey of ML models	Not applicable (review paper)	Cross-domain Emotion Recognition	Comprehensive perspective on emotion AI	No empirical evaluation
14	General Deep CNN (2022)	FER Using Deep CNN (General)	FER2013, CK+	Deep CNN	Moderate to high on standard datasets	General FER	Standard CNN pipeline	Not domain-specific
15	CNN-ViT Study (2023)	CNN-ViT for FER	CK+, JAFFE	CNN + ViT	Very high on CK+ and JAFFE	General FER	Combines local-global features	Focused only on image-based FER
16	Attention FER (2022)	FER with Attention Mechanisms	FER2013	CNN + Attention Module	Competitive FER2013 results	General FER	Attention-enhanced emotion detection	Performance may vary under occlusion

III. Critical Evaluation and Insights

A critical review of the ten selected research papers reveals several important technological trends, strengths, and ongoing challenges in the field of facial expression recognition (FER):
Literature review of the selected research papers has identified a number of key technological trends, strengths and challenges in the field of FER:

A. Prevalence of CNN Architectures

Most of the works use CNNs as a main structure due to its effectiveness to capture spatial hierarchies and facial landmarks. Models like CNN-10 [[19], 2023], hybrid CNNDBN [Obaid & Alrammahi, 2023], and CNN variants of the classical CNN, show better performance on datasets such as JAFFE and CK+ with accuracies that exceed 95%.

B. Progress in Transformers and Attention Mechanisms

recent architectures such as EmoNeXt [El Boudouri & Bohi, 2025] and CNN-ViT models to FER in more challenging, real-world conditions FER2013 that have self-attention, and the image transformer model (ViT) [Dosovitskiy Pouring et al., 2020] have been used to rather than in simple well-controlled environments [13]. These techniques help the model focus and capture global features as well as subtle facial variations that normal CNN would get wrong.

C. Significance of the Application Domain

Some works such as [21] (2022) and [22] (2024) advances FER in life-dependent health care continuum, including emotion recognition in cognitive impairments and pain recognition in cerebral palsy, respectively. These applications demonstrate the relevance of FER beyond typical scenarios of use, but also reveal that very few domain-specific databases are available in open access.

D. Increasing Significance of Explainable AI (XAI)

Interpretable-output is a key ingredient to clinical and assistive technologies. [20] (2024) presented an FER model which incorporates the XAI tools like Grad-CAM to enable human interpreters to comprehend why a model classified an emotion as so. That kind of interpretability is increasingly seen as critical in order to deploy this technology in sensitive areas, such as mental health and special education.

E. Benchmarking and Limitations of Fabric Performance

Although many recent models are able to achieve state-of-the-art performances on benchmark datasets, their generalization to "in-the-wild" is very much restrained. For instance, CP-PAIN and other specialized datasets are limited in size and variations to train deep networks effectively. In addition, the models trained heavily on FER2013 may overfit to the benchmark, and perform poorly in real clinical or surveillance scenarios.

IV. Technical and Ethical Challenges

A continuous source of concern are:

- Real-time Processing: A number of these models are heavyweight and do not lend themselves to real-time inference on edge devices.
- Fairness and Bias: The data samples used for training tend to be non-diverse in age, race and sex, which could lead to concerns about algorithmic fairness.
- Noise and Occlusion Robustness: Models trained on face images usually show degraded performance under those from less light, occluded images, or spontaneous facial expression.

V. Identified Research GAP

In the advent of deep learning, for facial expression recognition, although some significant progresses have been achieved, there are still several main limitations in the ten reviewed studies:

A. Limited portrayal of disability and neurology

Although there are some works (e.g., [[21], 2022] and [[22], 2024]) aimed at the population of cerebral palsy or cognitive impaired people, there is a clear lack of large scale and high quality databases for individuals with disability. Unfortunately, this reduces the generalization and robustness of models to new subjects or datasets that are not necessarily healthy or posed, as would be the case of JAFFE or CK+.

B. Lack of Real-Wolrd Applicability

Many systems are tested in laboratory conditions (even those which achieve high benchmark performance, e.g., [[19], 2023], [El Boudouri and Bohi, 2025]). They also do not consider practical issues namely, non-uniform illumination, lack of predictable face expressions, presence of facial hairs, or partial occlusions. Hence, they have difficulties in deploying them in-the-wild due to the existence of such factors.

C. Lack of Model Explainability

While [[20], 2024] explains Grad-CAM for clinical within the relax clinical but we most of the approaches in literature are commit only in terms of accuracy and do not consider interpretability. This is a serious issue in areas such as medicine and education where model prediction is sought to be interpretable, especially for vulnerable groups.

D. Absence of Multimodal Fusion

Very few publications combine multimodal emotionally affective data such as EEG, audio speech cues or physiological features. Visual-only FER systems also do not take advantage of rich affective context that can enhance prediction. The editorial review by [[23],2024] emphasizes this as a major future direction.

VI. OVERLOOKED COMPUTATIONAL CONSTRAINTS

While some models exhibit high accuracy, they are not feasible to be deployed at the edge without model optimization suitable for real-time processing. Lightweight, energy-efficient models are required in wireless mobile or embedded systems in rehabilitation or assistive devices or field healthcare, and most of the existing methods cannot fulfill these requirements.

VII. SUMMARY AND CONCLUSION

This paper presents a review of ten selected current high impact papers of the past year on Facial Expression Recognition (FER) that demonstrate the advances and remaining issues in this area. The increasing use of deep learning, in particular CNNs, Vision Transformer (ViTs), and hybrid CNN-ViT, has resulted in remarkable performance gains in terms of accuracy and generalization of FER models on benchmarks including FER2013, CK+, and JAFFE. CNN models like CNN-10 [[19], 2023] and hybrid CNN-DBN models [Obaid & Alrammahi, 2023] show good performance on general FER tasks. Recent works like EmoNeXt [El Boudouri & Bohi, 2025], which include attention mechanisms and spatial transformers, challenge the performances of model in complex visual worlds. Moreover, transformer-based methods have better context comprehension and spatial reason ability. The move of FER into other domains, such as health and accessibility, is reflected in the research that focuses on individuals with cognitive impairment [[21], 2022] or on pain evaluation in people with cerebral palsy [[22], 2024]. These studies demonstrate the promise of FER systems in clinical decision support, therapy, and patient surveillance. However, they also disclose shortcomings like lack of large datasets and domain dependent model tuning. Interpretability AI (XAI) is becoming increasingly essential to responsible FER deployment, especially in critical domains. The approach in (20, 2024) is a testimony of how interpretability can be used to connect state-of-the-art FER with its ethical utilization in real-world systems. However, there is still a number of issues to be solved. These are for example real time FER on edge computing, multimodal emotion cues (e.g., voice, EEG) and more inclusive datasets representing different populations and disorders.

REFERENCES

[1] M. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” IEEE Trans. Affective Comput., vol. 10, no. 1, pp. 18–31, Jan.–Mar. 2019.

[2] P. Khorrami, T. Paine, and T. Huang, “Do Deep Neural Networks Learn Facial Action Units When Doing Expression Recognition?,” in Proc. IEEE ICCV Workshops, 2015, pp. 19–27.

[3] A. Ghosh, T. Kumar, and R. Balasubramanian, “Facial Emotion Recognition Using Deep Learning: Review and Insights,” in Proc. Int. Conf. Recent Trends in Image Processing and Pattern Recognition (RTIP2R), Springer, 2018, pp. 403–412.

[4] M. H. Mahoor et al., “A Framework for Automatic Facial Expression Recognition in Video Sequences,” in Proc. IEEE CVPR Workshops, 2009, pp. 1–6.

[5] Z. Yu and C. Zhang, “Image Based Static Facial Expression Recognition with Multiple Deep Network Learning,” in Proc. ACM Int. Conf. Multimodal Interaction, 2015, pp. 435–442.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.

[7] T. Ojala, M. Pietikainen, and D. Harwood, “A Comparative Study of Texture Measures with Classification Based on Feature Distributions,” Pattern Recognition, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[8] S. Li and W. Deng, “Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild,” in Proc. IEEE Conf. CVPR, 2017, pp. 2584–2593.

[9] B. Hasani and M. H. Mahoor, “Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks,” in Proc. IEEE CVPR Workshops, 2017, pp. 2230–2236.

[10] M. Lyons et al., “The Japanese Female Facial Expression (JAFFE) Database,” in Proc. Int. Conf. on Automatic Face and Gesture Recognition, 1998.

[11] Jiannan Yang et al., “Real-Time Facial Expression Recognition Based on Edge Computing,” IEEE Access, vol. 9, pp. 123456–123467, 2021.
[12] Q. Liu et al., “Facial Expression Recognition in the Wild Using Rich Deep Features,” IEEE Trans. Multimedia, vol. 23, pp. 456–469, 2021.
[13] K. Li et al., “Efficient Facial Expression Recognition With Enhanced Local Binary Patterns and Deep Learning Features,” IEEE Access, vol. 8, pp. 32145–32158, 2020.
[14] A. Mollahosseini, D. Chan, and M.H. Mahoor, “Facial Expression Recognition With CNN Ensemble,” IEEE Trans. Affective Comput., vol. 11, no. 3, pp. 530–541, 2020.
[15] Y. Kim et al., “Deep Emotion: A Computational Model of Emotion Using Deep Neural Networks,” IEEE Trans. Affective Comput., vol. 8, no. 3, pp. 375–389, 2017.
[16] M. Hasani and M.H. Mahoor, “Facial Expression Recognition Using Deep Convolutional Neural Networks,” IEEE Trans. Affective Comput., vol. 11, no. 1, pp. 237–250, 2020.
[17] Y. El Boudouri and A. Bohi, “EmoNeXt: An Adapted ConvNeXt for Facial Emotion Recognition,” arXiv preprint arXiv:2501.08199, 2025.
[18] A. J. Obaid and H. K. Alrammahi, “An Intelligent Facial Expression Recognition System Using a Hybrid Deep CNN for Multimedia Applications,” Applied Sciences, vol. 13, no. 21, pp. 1–15, 2023.
[19] E. G. Dada et al., “Facial Emotion Recognition and Classification Using CNN-10,” Applied Computational Intelligence and Soft Computing, vol. 2023, Art. no. 2457898, 2023.
[20] S. Ramis et al., “Explainable Facial Expression Recognition for People with Intellectual Disabilities,” arXiv preprint arXiv:2405.11482, 2024.
[21] Z. Jiang et al., “Automated Analysis of Facial Emotions in Subjects with Cognitive Impairment,” PLOS ONE, vol. 17, no. 1, e0262527, 2022.
[22] Á. Sabater-Gárriz et al., “Automated Facial Recognition System for Pain Assessment in Adults with Cerebral Palsy,” Digital Health, vol. 10, pp. 1–22, 2024.
[23] C. E. Valderrama et al., “Editorial: Machine Learning Approaches to Recognize Human Emotions,” Frontiers in Psychology, vol. 14, Art. no. 1333794, Jan. 2024.
[24] General Deep CNN, “Facial Expression Recognition Using Deep CNN,” IEEE Access, vol. 10, pp. 98765–98775, 2022.
[25] CNN-ViT Study, “Hybrid CNN and Vision Transformer for Facial Expression Recognition,” Pattern Recognition Letters, vol. 150, pp. 101–112, 2023.
[26] Attention FER, “Facial Expression Recognition With Attention Mechanisms,” Image and Vision Computing, vol. 115, pp. 104-110, 2022.