International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Systematic Review on Facial Emotion Recognition System Datasets

Ramanpreet Kaur[1], Dr. Kanwal Preet Singh[2], Dr. Harmandeep Singh[3]

[1,2,3]Assistant Professor, Computer Science and Engineering, Punjabi University Patiala, Punjab

Abstract

This review provides a comprehensive analysis of prominent facial emotion recognition (FER) datasets, including RAVDESS, FER2013, AffectNet, CK+, JAFFE, SFEW, EmotioNet, and RAF-DB. Each dataset is examined based on its size, emotional categories, demographic coverage, and relevance to real-world scenarios. While these datasets have significantly contributed to the development and benchmarking of FER models, notable limitations persist, particularly the underrepresentation of diverse demographic groups and the lack of complex or subtle emotional expressions. The review further explores emerging deep learning advancements such as Transformer-based architectures, multimodal learning techniques, and generative models, which hold promise in overcoming these challenges. By integrating these innovations with enriched and inclusive datasets, FER systems can achieve greater accuracy, adaptability, and robustness in uncontrolled environments. Such improvements are essential for enhancing the performance and fairness of emotion recognition applications across domains including healthcare, education, customer service, and human-computer interaction.

Keywords: Facial Emotion Recognition (FER), Deep Learning, Datasets, Convolutional Neural Networks (CNN), Machine Learning.

1. Introduction

Facial Emotion Recognition (FER) systems have become a pivotal component in enhancing human-computer interaction, enabling machines to interpret and respond to emotional cues in ways that can benefit various sectors, such as psychology, healthcare, marketing, and security. The advancement of these systems is deeply linked to the creation and expansion of diverse datasets, which serve as the essential foundation for training and validating machine learning models used in emotion detection [1]. By interpreting facial expressions, FER aims to replicate a form of emotional intelligence in machines, fostering more natural and responsive interactions. FER systems typically function through a multi-stage process that begins with face detection, followed by landmark localization, feature extraction, and emotion classification. The detection of facial regions, once dependent on methods like the Viola-Jones algorithm, has significantly improved with the advent of deep learning models such as Convolutional Neural Networks (CNNs), which provide greater precision. Facial landmarks including the eyes, eyebrows, nose, mouth, and jawline are identified to facilitate the extraction of features that signify emotional states. While earlier approaches relied on manually crafted features, modern deep learning frameworks automatically learn and extract subtle patterns from facial data, enhancing the recognition of complex and nuanced emotions. In the final stage, classification algorithms, increasingly based on deep neural networks, categorize expressions into basic emotional states such as Happiness, Sadness, Anger, Surprise, Fear, Disgust, and Neutral, as represented in Fig 1 [21].

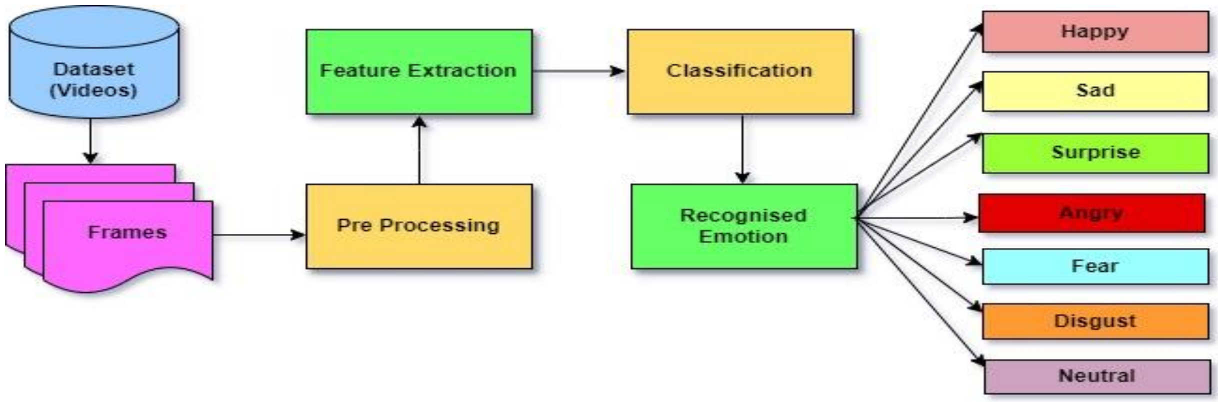

Figure 1 Steps in Facial Emotion Recognition System

By the 1990s, early computational models began integrating these psychological insights. In the 2010s, classical computer vision methods and traditional machine learning became prevalent. The landscape significantly evolved with the introduction of deep learning models, which unified feature extraction and classification processes, resulting in higher recognition accuracy. Around 2015, the adoption of Transformer architectures further improved FER capabilities by employing self-attention mechanisms that enabled the modeling of intricate facial details and contextual dependencies. Advancements from the 2020s onwards have introduced more sophisticated methods, including Generative Adversarial Networks (GANs) for synthetic data generation, multimodal learning that integrates visual and audio cues, and real-time emotion detection systems. These innovations have considerably enhanced the performance, adaptability, and real-world applicability of FER systems, as illustrated in Figure 2.

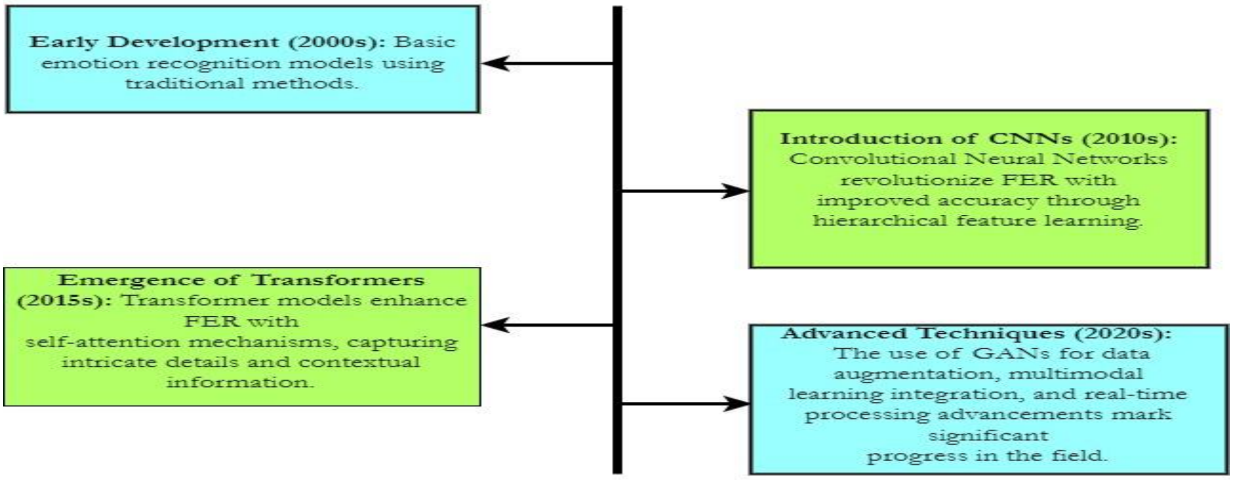

Figure 2 Timeline diagram of FER System

1.1 Applications of FER Systems

The applicability of FER spans across numerous sectors, highlighting its versatility and growing relevance. In the healthcare domain, FER is increasingly being utilized for monitoring mental health conditions, including the detection of depression, anxiety, and stress, by observing a patient's facial expressions over time. It also serves an important function in assistive technologies, supporting individuals with speech or emotional communication disorders in expressing or understanding emotions more effectively. In the context of security and surveillance, FER systems contribute to behavioural analysis by evaluating emotional cues in high-risk scenarios, potentially aiding in the identification of suspicious or abnormal activities. The technology also holds significant value in marketing and consumer research, where it is used to gauge customer responses and emotional reactions to products, services, or advertisements, enabling organizations to refine their strategies and enhance user engagement [5]. Moreover, FER plays a central role in human-computer interaction, where it supports the development of emotionally responsive systems, such as intelligent virtual assistants and social robots, which adjust their interactions based on the emotional state of the user. This enhances overall user experience by fostering more empathetic and adaptive communication. The integration of FER into everyday applications and smart devices further underscores its role in creating emotionally aware technology. In the fields of entertainment and gaming, FER enhances user engagement by dynamically adjusting content or gameplay in response to the player's emotional reactions. Within educational environments, the technology assists in evaluating students' emotional involvement and attentiveness, offering educators valuable feedback for improving teaching strategies. Collectively, the definition and scope of FER underscore its transformative potential across diverse sectors, positioning it as a critical component in the evolution of emotionally intelligent and context-aware technologies.

1.2 Challenges in FER Systems

Facial Emotion Recognition (FER) systems, despite their growing capabilities, continue to face a variety of technical and practical challenges that affect their precision and dependability. A prominent issue is the sensitivity of these systems to lighting conditions, which can significantly impact the visibility and clarity of facial features. Since FER relies heavily on detecting subtle facial expressions, variations in illumination, such as shadows, glare, or insufficient light, can obscure crucial facial landmarks like the eyes or mouth. These inconsistencies are especially problematic in real-world settings where lighting is often unpredictable and not easily controlled, thus increasing the likelihood of misclassification. Occlusion represents another major obstacle, particularly in situations where portions of the face are blocked by objects such as eyeglasses, facial hair, masks, or even hairstyles. Such obstructions reduce the effectiveness of FER by hiding key facial regions required for accurate emotional interpretation. For example, widespread mask usage during the COVID-19 pandemic has made it difficult to detect emotions that rely on mouth movements, such as surprise or disgust. Similarly, sunglasses can conceal the eyes, which are central to identifying emotions like fear or anger. As a result, systems must often infer emotional states from incomplete data, increasing the risk of errors. The inherent variability in how emotions are displayed across individuals adds yet another layer of complexity [6]. Emotional expressions are not uniform; one person's expression of happiness might be a wide smile, while another's may be a restrained grin. These differences are influenced by both individual behaviour and cultural norms. In some cultures, emotions are deliberately masked in public spaces, making it more challenging for FER technologies to interpret true emotional states accurately. Even with training on diverse datasets, FER models struggle to maintain consistent performance across all demographic and cultural groups. Moreover, temporal dynamics in facial expressions, such as micro-expressions that occur within milliseconds, are often difficult for FER systems to capture and interpret. These fleeting expressions are crucial in conveying authentic emotional cues but may be missed by systems not designed for high-speed real-time processing. This limitation becomes even more pronounced in live applications where video streams must be analyzed continuously, requiring both high accuracy and rapid computation [10]. Beyond the technical hurdles, FER systems must also address significant ethical and privacy concerns. The monitoring of facial expressions, often without the individual's informed consent, raises issues surrounding data protection, consent, and potential misuse. Developing FER technologies that are not only technically robust but also ethically responsible remains a pressing need. Progress in this field will depend on advancements in both machine learning techniques and ethical design frameworks, ensuring FER systems are adaptable, inclusive, and respectful of privacy in real-world environments.

2. Datasets for Facial Emotion Recognition Systems

These datasets must reflect a broad spectrum of variables including ethnicity, age, gender, illumination, head poses, and cultural nuances that influence the way emotions are expressed and interpreted. A well-rounded dataset enables FER models to generalize more effectively across different demographic groups and environments. However, achieving such comprehensiveness brings its own set of challenges, such as standardization, annotation consistency, and the potential trade-off between diversity and data quality. In response to the expanding interest in FER, numerous datasets have been developed, each designed to address specific technical and contextual challenges. For example, some focus on addressing occlusions like glasses or face masks, while others emphasize subtle emotional variations or real-time usability in dynamic settings. Although these resources have advanced FER research, notable limitations persist. Certain datasets provide high-quality images with clearly labelled emotional states, but may lack demographic variety, limiting the model's applicability across diverse populations [2]. Conversely, datasets emphasizing inclusivity often sacrifice resolution or the granularity of emotion annotations, creating a tension between inclusiveness and labelling precision. The intended use-case of a FER system further influences the selection of appropriate datasets. Applications in healthcare, such as early diagnosis of psychological disorders, necessitate datasets capable of capturing subtle and nuanced expressions that could correlate with mental health symptoms. In contrast, datasets for retail or customer service applications may prioritize naturalistic expressions and real-time emotional shifts observed in uncontrolled settings like shopping malls or online meetings. This contextual dependency underscores the need for datasets that align with the specific demands of each domain. Moreover, ethical considerations play a pivotal role in dataset development and usage. Training models on imbalanced datasets may result in biased systems that perform inequitably across populations. For example, if a dataset over represents certain ethnic or age groups, the corresponding FER system may show reduced accuracy or fairness when applied to individuals from underrepresented demographics. Such disparities emphasize the importance of both inclusive data collection and the implementation of algorithmic strategies to reduce bias. Addressing these concerns is essential for developing FER technologies that are both technically robust and socially responsible. Conducting a systematic analysis of existing FER datasets provides vital insights into their relative strengths and shortcomings. Such reviews guide researchers in selecting the most suitable datasets for their specific applications, whether in academic investigations, commercial systems, or assistive technologies, and also highlight the necessity for future datasets to be ethically sound, demographically inclusive, and capable of representing complex emotional states in realistic contexts [3]. This review thus evaluates a wide array of FER datasets, taking into account parameters such as dataset size, diversity, annotation precision, and relevance to various practical applications. Through this evaluation, it aims to support practitioners and researchers in making informed decisions regarding dataset selection and to encourage the continued development of datasets that can bridge existing gaps in the field. As FER systems advance and find broader application, the role of well-structured, high-quality datasets will remain central to ensuring their effectiveness, fairness, and adaptability.

2.1 Dataset Description

Several benchmark datasets have played a pivotal role in the advancement of Facial Emotion Recognition (FER) by providing the empirical foundation for training and evaluating machine learning and deep learning models. These datasets vary significantly in terms of size, diversity, modalities, and application scenarios, each contributing uniquely to the field. RAVDESS offers a multimodal collection of 7,356 recordings comprising emotional speech and song performed by 24 professional actors (12 male and 12 female), using two lexically-matched statements in a standard North American accent. Emotions such as calm, happy, sad, angry, fearful, surprise, and disgust are represented in both normal and strong intensities. Available in audio-only, video-only, and audio-visual formats, RAVDESS is especially valuable for multimodal emotion recognition research. The FER2013 dataset, created for a

Kaggle competition, contains 35,887 grayscale images of faces annotated with seven emotions: Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral. Despite being composed of relatively low-resolution images (48×48 pixels), it remains a standard benchmark for developing and comparing emotion recognition models due to its large scale and class balance. AffectNet is among the largest FER datasets, comprising approximately one million images collected from the web. Each image is annotated with one of eight emotional states: Angry, Disgust, Fear, Happy, Sad, Surprise, Contempt, and Neutral. The dataset's sheer volume and diverse sources make it suitable for developing deep learning models that require large-scale, real-world data for robust generalization [16]. The Cohn-Kanade Plus (CK+) dataset provides 593 video sequences from 123 individuals, capturing the temporal progression of facial expressions from a neutral baseline to peak emotion. The dataset includes both posed and spontaneous expressions covering emotions such as Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Contempt, making it highly suitable for dynamic expression analysis. The JAFFE dataset contains 213 grayscale images from 10 Japanese female models. Each image is labeled with one of seven emotions, Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral. JAFFE is particularly noted for its focus on cultural and gender-specific aspects of emotion expression, offering a niche yet important perspective in FER development [15]. Social Face Expression in the Wild (SFEW) comprises approximately 1,500 images extracted from movies and social contexts, presenting facial expressions under uncontrolled lighting, occlusion, and head pose conditions. It includes seven primary emotions and serves as a benchmark for evaluating FER systems in more natural, unconstrained environments. The EmotioNet dataset consists of around one million images with labels for 11 distinct emotional categories, including Shame, Guilt, and Embarrassment, emotions often omitted in other datasets. By including a broader set of affective states, EmotioNet contributes to developing more nuanced FER models capable of recognizing complex emotions in diverse, real-world contexts [20]. RAF-DB includes 30,000 facial images annotated for seven fundamental emotions, Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral. Collected from uncontrolled environments and a wide demographic range, RAF-DB provides high intra-class variation, making it ideal for training FER models for in-the-wild applications [17]. Together, these datasets serve as essential tools for advancing FER research. Each dataset brings unique strengths and limitations, ranging from the controlled conditions of JAFFE and CK+ to the diverse and unconstrained nature of AffectNet, SFEW, and EmotioNet, thereby enriching the methodological and empirical foundation upon which robust and context-aware emotion recognition systems are built.

3. Comparison and Analysis of Datasets

A comparative analysis of prominent Facial Emotion Recognition (FER) datasets highlights the varying strengths and limitations that influence their suitability for different research and application contexts. The FER2013 dataset, comprising 35,887 grayscale images annotated with seven basic emotions, is frequently used for baseline model training due to its balanced classes and standardized format. However, its relatively low resolution and limited diversity constrain its generalizability in real-world applications. In contrast, AffectNet**,** with approximately one million annotated images across eight emotions, offers a significantly larger and more diverse dataset. Sourced from the internet, it captures a wide array of facial variations under natural conditions, making it ideal for training robust deep learning models that require extensive data coverage for generalization. The Cohn-Kanade Plus (CK+) dataset provides 593 video sequences from 123 subjects and is particularly valuable for its temporal depth, capturing the progression of facial expressions from neutral to peak intensity. This makes CK+ especially suitable for studying dynamic and spontaneous emotional expressions in FER. JAFFE, although small in size with just 213 images, offers a focused view on cultural and gender-specific expression, featuring Japanese female subjects expressing seven emotions. Its controlled conditions and well-annotated expressions make it useful for niche studies on demographic sensitivity in emotion recognition. The Social Face Expression in the Wild (SFEW) dataset contributes 1,500 images extracted from movies and other social contexts, providing valuable insights into FER performance in naturalistic, uncontrolled environments. Despite its limited size, the diversity of expressions and contextual backgrounds enhances model robustness in real-world applications. EmotioNet distinguishes itself with around one million images and annotations for 11 emotions, including complex affective states like guilt, shame, and embarrassment. Its extensive coverage makes it a critical resource for developing advanced FER systems capable of recognizing nuanced emotional expressions across diverse settings. Finally, RAF-DB (Real-world Affective Face Database) includes 30,000 real-world images annotated with seven fundamental emotions. Its strength lies in its demographic diversity and varied environmental conditions, making it suitable for building models with improved real-world applicability. Overall, each dataset presents unique advantages, whether in terms of volume, diversity, emotion granularity, or temporal resolution, that must be strategically aligned with the intended application of FER systems. The appropriate selection and use of these datasets directly impact the performance, fairness, and reliability of emotion recognition technologies in practice.

Table 1 Comparison of various facial Emotion Recognition Datasets

Dataset	Size (Samples)	Types of Emotion
RAVDESS	7356 files	calm, happy, sad, angry, fearful, surprise, disgust
FER2013	35,887 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral
AffectNet	1,000,000 images	Angry, Disgust, Fear, Happy, Sad, Surprise, Contempt, and Neutral
CK+ (Cohn-Kanade Plus)	593 video sequences, with 123 subjects	Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Contempt
JAFFE (Japanese Female Facial Expression)	213 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral
SFEW (Social Face Expression in the Wild)	1,500 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral
EmotioNet	1,000,000 images	Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness, Surprise, Shame, Guilt, and Embarrassment

RAF-DB (Real-world Affective Face Database)	30,000 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral

4. Future Directions

4.1 Dataset Enhancement: Improving Existing FER Datasets

Enhancing the utility and effectiveness of existing facial emotion recognition (FER) datasets requires a multi-faceted approach. A primary strategy involves increasing demographic diversity by incorporating a broader spectrum of age groups, ethnicities, and genders. This is essential to mitigate algorithmic bias and ensure model generalizability across global populations. Another important improvement lies in expanding the datasets to include a wider range of real-world conditions, such as varied lighting environments, facial occlusions (e.g., masks, glasses), and spontaneous expressions, to better simulate everyday scenarios. Enhancing the visual quality of the data through high-resolution images and dynamic video sequences can further enable models to learn from fine-grained facial details and temporal expression patterns. Additionally, broadening the emotional taxonomy to include complex and nuanced emotions, such as blended or context-specific affective states, enriches the dataset and facilitates more accurate emotion modelling. Annotating these datasets with detailed metadata, such as emotion intensity levels and situational context, also improves training precision. Regular updates to these datasets with culturally and temporally relevant content ensure ongoing relevance in evolving social environments. Collectively, these enhancements contribute to the development of FER systems that are not only technically robust but also ethically and contextually sound.

4.2 New Dataset Needs: Addressing Gaps in FER Research

Despite the availability of several established FER datasets, there remain significant gaps that impede the development of truly inclusive and adaptable emotion recognition systems. A major shortcoming is the underrepresentation of diverse demographic groups, which can lead to biased model performance and limited applicability across cultural contexts. Furthermore, existing datasets often focus predominantly on basic emotions, neglecting the subtle, mixed, or culturally specific affective states that are crucial for real-world emotional intelligence. Another limitation is the prevalence of static images in many datasets, which fail to capture the dynamic evolution of facial expressions. This restricts their effectiveness in applications involving video analysis or real-time interaction. The lack of contextual and multimodal information, such as situational background, speech cues, or physiological signals, also constrains the depth of emotional understanding achievable through these datasets. To address these issues, there is a pressing need for new datasets that are demographically diverse, emotionally nuanced, temporally rich, and environmentally varied. Such datasets would significantly enhance the ecological validity and cross-cultural generalizability of FER systems, paving the way for more accurate and inclusive emotion detection technologies.

4.3 Technological Advancements: Deep Learning Innovations in FER

Emerging advancements in deep learning offer substantial potential to overcome current limitations in FER. Transformer-based architectures, known for their powerful attention mechanisms, hold promise for capturing subtle spatial and temporal dependencies within facial expressions, enabling the detection of fine-grained emotional cues. Multimodal learning, which integrates visual data with complementary signals such as voice, text, or physiological metrics, can provide a holistic understanding of emotional states by embedding facial expressions within broader behavioural contexts. The application of Generative Adversarial Networks (GANs) is another promising direction, particularly for data augmentation. GANs can generate synthetic yet realistic facial expressions, thereby addressing dataset imbalance and scarcity. Moreover, self-supervised and semi-supervised learning approaches can leverage vast amounts of unlabelled data to improve model performance without relying entirely on manual annotation, which is often labour-intensive and prone to subjectivity. These technological innovations not only promise to enhance the accuracy and efficiency of FER systems but also expand their practical deployment across domains such as healthcare, education, security, and human-computer interaction.

5. Conclusion

The field of facial emotion recognition (FER) has witnessed substantial progress, driven by the availability of diverse datasets such as RAVDESS, FER2013, AffectNet, CK+, JAFFE, SFEW, EmotioNet, and RAF-DB. Each of these datasets contributes uniquely to the advancement of FER systems, offering various combinations of emotional labels, data modalities, and contextual diversity. However, critical gaps remain, including limited demographic representation, inadequate coverage of subtle or culturally specific emotions, and reduced performance in uncontrolled, real-world environments. Bridging these gaps requires the development of enriched datasets that reflect greater diversity and realism, as well as the integration of cutting-edge deep learning techniques, such as Transformer architectures, multimodal fusion strategies, and real-time processing capabilities. These innovations collectively promise to enhance the precision, fairness, and adaptability of FER systems. As the domain continues to mature, such enhancements will not only improve the technical performance of emotion recognition technologies but also broaden their applicability across sectors like healthcare, education, human-computer interaction, and social robotics, ultimately contributing to a deeper, more empathetic understanding of human emotions.

6. Data Availability

The data presented in this study are derived from publicly available datasets, including RAVDESS, FER2013, AffectNet, CK+, JAFFE, SFEW, EmotioNet, and RAF-DB. These datasets can be accessed through their respective official sources or repositories. Any additional data used or analyzed during this study are available from the corresponding author upon reasonable request

7. Conflict of Interest

The authors declare that they do not have any conflict of interest.

8. Funding Source

None.

9. Acknowledgements

The authors would like to thank the Department of Computer Science and Engineering, Punjabi University Patiala, for providing the support and resources required to carry out this study.

10. Authors' Biography

[1]Ramanpreet Kaur

Ramanpreet Kaur , working as an Assistant Professor at Govt Bikram College of Commerce Patiala is a researcher with over 16 years of teaching experience. She holds an M.Tech degree in Computer Science and Engineering and pursuing Ph.D in the field of Image Processing in Computer Science and Engineering.

[2]Dr. Kanwal Preet Singh

Dr. Kanwal Preet Singh Attwal is a seasoned academic with over 22 years of teaching and 12 years of research experience in Computer Science and Engineering. He holds an M.Tech. and Ph.D. with a focus on Data Mining. His research spans data analytics, machine learning, and agricultural informatics, emphasizing crop yield prediction, privacy-preserving classification, and sustainable computing. Dr. Attwal has published over 30 papers in reputed international journals and conferences, including IEEE proceedings and book chapters.

[3]Dr. Harmandeep Singh

Dr. Harmandeep Singh is an Assistant Professor at Punjabi University, Patiala, specializing in Computer Networking and Network Security. His research focuses on secure and efficient communication systems, addressing issues like cyber threats, intrusion detection, and resilient data transmission. He explores protocols, firewalls, and cryptographic methods to safeguard wired and wireless networks. Dr. Singh has published extensively in reputed journals and conferences and actively mentors postgraduate and doctoral scholars.

References

1. B. G. K. Reddy, P. Yashwanthsaai, A. R. Raja, A. Jagarlamudi, N. Leeladhar, and T. T. Kumar, “Emotion recognition based on convolutional neural network (CNN),” in *Proc. Int. Conf. Adv. Electr., Electron., Commun., Comput. Autom.*, IEEE, pp. 1–5, 2021. https://doi.org/10.1109/ICAECA52838.2021.9675688
2. A. Sharma, K. Sharma, and A. Kumar, “Real-time emotional health detection using fine-tuned transfer networks with multimodal fusion,” *Neural Comput. Appl.*, vol. 2, 2022.
3. M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, “Deep learning-based facial emotion recognition for human–computer interaction applications,” *Neural Comput. Appl.*, vol. 8, 2021. https://doi.org/10.1007/s00521-021-06012-8
4. K. Wang, Y. Ho, Y. Huang, and W. Fang, “Design of intelligent EEG system for human emotion recognition with convolutional neural network,” in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, pp. 142–145, 2019.
5. W. Liu, J. Qiu, W. Zheng, and B. Lu, “Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 14, no. 2, pp. 715–729, 2022.
6. L. Sandra, Y. Heryadi, W. Lukas, W. Suparta, and A. Wibowo, “Deep learning based facial emotion recognition using multiple layers model,” in *Proc. Int. Conf. Adv. Mechatron., Intell. Manuf. Ind. Autom.*, IEEE, pp. 137–142, 2021. https://doi.org/10.1109/ICAMIMIA54022.2021.9809908
7. A. Landowska *et al.*, “Automatic emotion recognition in children with autism: A systematic literature review,” *Sensors*, vol. 22, no. 4, pp. 1–30, 2022. https://doi.org/10.3390/s22041649
8. S. Al-asbaily and K. Bozed, “Facial emotion recognition based on deep learning,” in *Proc. IEEE 2nd Int. Maghreb Meet. Conf. Sci. Tech. Autom. Control Comput. Eng. (MI-STA)*, vol. 22, no. 16, pp. 534–538, 2022. https://doi.org/10.3390/s22166105
9. S. Kaur and N. Kulkarni, “A deep learning technique for emotion recognition using face and voice features,” in *Proc. IEEE Pune Section Int. Conf. PuneCon*, pp. 1–6, 2021. https://doi.org/10.1109/PuneCon52575.2021.9686510
10. M. U. Khan *et al.*, “Deep learning based intelligent emotion recognition and classification system,” in *Proc. Int. Conf. Front. Inf. Technol. (FIT)*, IEEE, pp. 25–30, 2021. https://doi.org/10.1109/FIT53504.2021.00015
11. A. I. Siam, N. F. Soliman, A. D. Algarni, F. E. Abd El-Samie, and A. Sedik, “Deploying machine learning techniques for human emotion detection,” *Comput. Intell. Neurosci.*, 2022. https://doi.org/10.1155/2022/8032673

12. S. Palaniswamy and Suchitra, "A robust pose illumination invariant emotion recognition from facial images using deep learning for human-machine interface," in *Proc. IEEE Int. Conf. Syst., Inf., Technol., Softw. (CSITSS)*, 2019. https://doi.org/10.1109/CSITSS47250.2019.9031055
13. S. Yuvaraj, J. V. Franklin, V. S. Prakash, and A. Anandaraj, "An adaptive deep belief feature learning model for cognitive emotion recognition," in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, pp. 1844–1848, 2022. https://doi.org/10.1109/ICACCS54159.2022.9785267
14. G. Chartrand *et al.*, "Deep learning: A primer for radiologists," *Radiographics*, vol. 37, no. 7, pp. 2113–2131, 2017. https://doi.org/10.1148/rg.2017170077
15. L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, 2021. https://doi.org/10.1186/s40537-021-00444-8
16. L. Arnold, S. Rebecchi, S. Chevallier, and H. Paugam-Moisy, "An introduction to deep learning," in *Proc. Eur. Symp. Artif. Neural Netw.*, pp. 477–488, 2021. https://doi.org/10.1201/9780429096280-14
17. D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Appl. Sci.*, vol. 9, no. 21, pp. 1–31, 2019. https://doi.org/10.3390/app9214678
18. A. Saravanan, G. Perichetla, and D. K. S. Gayathri, "Facial emotion recognition using convolutional neural networks," *arXiv*, 2019.
19. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, pp. 1–6, 2017. https://doi.org/10.1109/ICEngTechnol.2017.8308186
20. X. Sun, P. Xia, L. Zhang, and L. Shao, "A ROI-guided deep architecture for robust facial expressions recognition," *Inf. Sci.*, vol. 522, pp. 35–48, 2020. https://doi.org/10.1016/j.ins.2020.02.047
21. C. Dalvi, M. Rathod, S. Patil, S. Gite, and K. Kotecha, "A Survey of AI-based Facial Emotion Recognition: Features, ML & DL Techniques, Age-wise Datasets and Future Directions," *IEEE Access*, pp. 1–1, 2021. https://doi.org/10.1109/ACCESS.2021.3131733