Enhancing Facial Emotion Recognition Using DCNN through Effective Extraction for High-Level Features

Mohsin Sami[1], Usman Aamir[1], Ramsha Saeed[1]

University of Central Punjab, Lahore, Pakistan

Abstract

Facial Emotion Recognition, or FER, is increasingly important in human-computer communication, psychology, and advanced monitoring. Researchers seek to enhance accuracy of recognition with high-level feature extraction techniques, utilizing the advancements made in deep convolutional neural networks (DCNNs). This systematic literature review (SLR) compiles and analyses forty-two (42) studies published from 2020 to 2025 across IEEE Xplore, SpringerLink, ScienceDirect, Scopus, and ACM Digital Library. The review marks trends in DCNN-based FER and analyses the datasets used within the research such as FER-2013, CK+, JAFFE, along with their respective metrics and feature extraction methods. The DCNNs were quantitatively found to outperform other architectures, although interpretable and deployable systems were lacking. Through qualitative synthesis, the major focus was placed on feature engineering with deep learning, merging models, and creating comprehensive pre-processing workflows. The review indicates that while steps have been taken toward achieving the broad-scope goals, the diversity of datasets, unified metrics, and real-time systems hinder progress. The comparison of the approaches for high-level feature extraction builds the foundation of the review, making it distinct. This SLR aims to guide future efforts by highlighting unresolved issues and recommending directions for practical and responsible FER systems.

1. Introduction

Deep Convolutional Neural Networks (DCNNs) dramatically change the character of computer vision because they let computers learn hierarchical feature representations directly from raw images. Their layered architecture reflects how humans tend to perceive things. Low-, mid-, and high-level features can really be captured. In recent years, DCNNs have outperformed hand-crafted, feature-based methods of their old at strength in many image classification problems, as they can learn and generalize from huge data sets. In affective computing, DCNNs have been very effective in interpreting subtle changes in the expression of humans. Spatial Transformer Networks, Squeeze-and-Excitation blocks [34], and they are being done.

While facial image detection and emotion recognition are essential features of becoming natural and intuitive in human-computer interaction, several challenges remain, such as lighting variations, pose differences, occlusions, and differences in expression intensity. The most recent studies adopt different techniques to mitigate these challenges; among them, transfer learning [41], feature decomposition techniques [5], and optimization algorithms [41] received high research attention. The research problems were highlighted at the aspects of modelling for high-level semantic feature extraction, so that better emotion classifications might be obtained under various real-world conditions. Consequently, hybrid models incorporating DCNN features and classifiers such as SVM or ensemble techniques [3] claim significant improvement on emotion detection accuracy.

Even with notable progress made in the field of facial emotion recognition (FER) through the application of deep convolutional neural networks (DCNNs), there are still gaps within the research. One of the most critical problems in the existing literature is the use of small and very controlled datasets, such as CK+, JAFFE, or FER-2013, which prominently feature posed expressions [7]. This greatly hampers the model's performance in real-life context scenarios characterized by uncontrolled lighting, head pose, glasses and mask occlusions, as well as background noise. Moreover, most DCNN-based FER models tend to overfit these datasets, achieving excellent performance during model-specific tests but failing in real-world settings. Another important challenge is the lack of comprehensive emotion representation; although identifiable, ethnically homogenous high-level features may be extracted using DCNN's, the absence of delicate nuances such as micro-expressions is fatal for emotion modelling [8]. In addition, most models are constructed as black boxes, making them devoid of interpretability and explainability, which are essential features when deploying such FER systems in sensitive sectors such as healthcare, education, or law enforcement.

Besides that, less attention has been drawn to multi-modal emotion recognition, as most studies focus on visual information while neglecting other informative signals such as audio and body posture, or physiological responses [9]. This unimodal approach limits the system's robustness. Another challenge is the high computational expense attributed to some DCNN FER architectures, as these are challenging for use in real time or on resource constrained devices such as smartphones or embedded systems. Moreover, models that are built from scratch with no provision for domain adaptation or personalization exhibit a lack of flexibility and cross-domain adaptability which limits their applicability to different user groups and cultures [10]. The imbalance of classes in the data set compounds the problem, with models trained on data representing dominant emotions like happiness or neutrality, and performing poorly on less represented emotions such as fear and disgust. Static models oversimplify the face's temporal dynamics and many consider facial expressions as pictures instead of a progressive sequence of frames containing a sequence of transitions and crucial moments needed for correct interpretation of emotional state [11]. There is little to no research on various facets of fairness and bias, particularly when analyzing how age, gender, and ethnicity impact the performance of most FER systems—raising questions concerning the responsible use of FER systems across different socio demographic contexts. The inference from these developments is that model development should not just focus on models that have major accuracy but also on scalability and generalization cross datasets.

Regardless, with all advancements taking place, this very area lacks a common synthesis of methods dealing with high-level feature extraction from DCNNs in FER. Many works discuss model solutions from various perspectives, yet only a handful have extensively reviewed the most primal techniques, architectures, and criteria for evaluation into a more organized view. This is one major area of motivation for undertaking an SLR addressing DCNN-based architectures and their evolution to address problems in emotion-related facial expression recognition (FER). By surveying the literature dated from 2020-2025, the SLR will highlight patterns, compare performance benchmarks, and establish gaps in current methodologies. This systematic review will benefit not only academic research but will also lead practitioners to choose the best solutions for specific needs, creating a bridge between theory and practice.

This research stands out in that it makes a concentrated effort in explaining how high-level features are used by DCNNs to develop a better face emotion recognition. It does not go to the broader topic of affective computing or general trends in FER, but rather focuses on architectural and algorithmic innovations that are responsible for effective extraction of discriminative emotional features by DCNNs. Here, rigid has perused the arguments on how feature extraction modules, attention mechanisms, and hybrid integration of classifiers contribute to performance improvement. Not only this, it brings forward an entirely new taxonomy of feature extraction methods as well as evaluates them on comparing their suitability on benchmark datasets. Such level of detail and specificity marks a unique contribution in this particular discipline and paves way for the researchers to take up future research into finer and even more robust FER systems development.

We have compared ore study with already proposed studies in literature. This comparison highlighted unique contributions and comprehensive approach of proposed SLR. Comparison of this SLR with other state-of-the-art studies has been presented in Table 1.

Table 1. Comparative Analysis of proposed review on Facial Emotion Recognition (2020–2025) with other state-of-the-art studies from literature

Feature / Study	**Our SLR**	**[42]**	**[41]**	**[40]**
Focus Area	High-level feature extraction in DCNN-based FER	FER in elderly populations	Algorithms in FER	Deep learning in affect recognition
Time Frame Covered	2020–2025	Last decade	Not specified	Up to 2021
Number of Studies Reviewed	35+	31	Not specified	Not specified
Methodological Depth	Detailed analysis of DCNN architectures, feature extraction methods, and datasets	Focus on application challenges	Broad algorithmic overview	Overview of deep learning methods
Dataset Analysis	Comprehensive coverage including FER2013, CK+, JAFFE, AffectNet, and custom datasets	Highlights lack of elderly-specific datasets	General mention of datasets	Discusses dataset biases
Feature Extraction Techniques	Emphasis on attention mechanisms, hybrid models, and dual-stream architectures	Not deeply explored	Not deeply explored	General discussion

Feature / Study	Our SLR	[42]	[41]	[40]
Application Domains	Broad applications including real-time systems, healthcare, and human-computer interaction	Healthcare and elderly care	General applications	Affective computing
Novel Contributions	Introduction of a structured LR table, quantitative summaries, and qualitative synthesis	Recommendations for elderly-inclusive datasets	Algorithm classification	Insights into deep learning advancements
Limitations Addressed	Dataset imbalance, generalization challenges, computational costs, and emotion intensity variation	Scarcity of elderly-specific datasets	Not specified	Dataset biases and model limitations
Future Research Directions Proposed	Development of lightweight models, creation of diverse datasets, and integration of multimodal data	Emphasis on age-inclusive datasets and explainable AI	Not specified	Multimodal affective analysis

This SLR has been organized in five sections. Section 1 provides introduction of the topic. Section 2 presents related work. Section 3 elaborates methodology for conducting this review. Section 4 presents results and discussion obtained after analyzing forty-two (42) articles. Section 5 concludes this research and provides future directions.

2. Literature Review

The impact of recent developments in DCNN architectures on the advancement of facial emotion recognition (FER) concerned the ability to extract high-level features more effectively. [34] presented the EmoNeXt, a patch-wise adaptation of the ConvNeXt architecture that integrates Squeeze-and-Excitation blocks and Spatial Transformer Networks while emphasizing salient facial regions and channel-wise dependencies. The proposed model managed to outperform state-of-the-art results on the FER2013 dataset, thereby confirming the paramount importance of attention mechanisms in FER. Similarly, [35] presented the Distract Your Attention Network (DAN), which, by implementing multi-head cross-attention, captures interactions between facial regions and facilitates state-of-the-art performances on various datasets, such as AffectNet and RAF-DB.

To ensure generalization across multiple datasets, [4] presented DeepFEVER: a feature extractor based on deep learning that showed good generalization performance across several datasets, including AffectNet and RAF-DB. [5] proposed Feature Decomposition and Reconstruction Learning (FDRL), postulating that the method allows the decomposition of features into shared and unique components, thus capturing expression similarities as well as variation that is characteristic and, hence, outperforming the best existing techniques across various datasets consistently.

Hybrid models comprising DCNNs with other classifiers have demonstrated some potential. [3] highlighted how the DCNN architectures like AlexNet and VGG-16 saw an increase in accuracy by 7% to 9% when replacing SoftMax classifiers with Support Vector Machine (SVM) or ensemble classifiers on the MLF-W-FER dataset. [41] described a model based on Fuzzy Eigen Weighted feature extraction and Chaotic Spider Monkey

Optimization for feature selection, combined with a Dual-attention residual U-Net classifier, attaining remarkable recognition accuracy on databases such as CK+, FER2013, and JAFFE.

Transfer learning proves to be a blessing for FER. [38] used pre-trained DCNN models and fine-tuned them with facial expression data, obtaining accuracies of 93.7% for KDEF test set and 100% for the JAFFE test set. [18] discussed a hybrid approach involving DCNN and Extreme Learning Machines (ELM), outperforming both the conventional DCNN and ELM methods, therefore allowing real-time emotion detection in an online learning environment.

Optimization techniques have enhanced the performances of the models in FER. [16] developed an artificial intelligence-led FER system applying hybrid deep belief rain optimization for improving feature selection and classification accuracy across multiple datasets. In particular, [11] proposed the Deep Cross Feature Adaptive Network (DCFA-CNN), which synergizes both shape and texture features to improve recognition accuracy on databases such as CK+ and JAFFE.

3. Methodology

This research offers structured methods for searching, categorizing, and synthesizing the literature in accordance with pre-established objectives. This highlights the areas that may serve as a roadmap for future research directions in the designated domain. Figure 1 shows the three steps of the research technique used for this review.

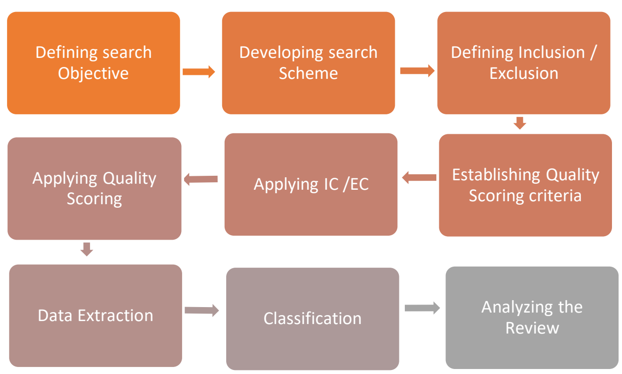

Figure 1. Research methodology

The review has been conducted in multiple steps. In first step, the research objectives (ROs) have been defined, in second step, ROs have been used to develop the research questions (RQs). In third step, the search strategy for locating relevant material is devised. Following that in forth step, the inclusion/exclusion and quality assessment criteria is applied that led to the next step in which article were shortlisted. Shortlisted articles were ranked based on the quality rating criteria. The shortlisted papers were classified and synthesized in accordance with the research areas. Finally, the results were discussed and analysed as per research questions. This research is novel contribution to the body of knowledge in DCNN domain.

3.1. Research Objectives

1. To identify the recent developments in facial emotion recognition (FER) employing Deep Convolutional Neural Networks (DCNN) from 2020 to 2025.
2. To categorize proposed FER methods based on the different types of neural network architecture used and feature extraction methodologies adopted.
3. To investigate what high-level facial features were extracted and analyzed in known studies.
4. To evaluate the benchmark datasets used to train and test FER models.
5. To evaluate trends, gaps, and emerging directions in the design and application of FER models using DCNN.

3.2 Research questions

Research questions used to retrieve documents from literature have been presented in Table 2.

Table 2. Research Questions (RQs)

RQ Code	**Research Question**	**Motivation**
RQ1	What are the most common research domains and application areas in FER using DCNN from 2020 to 2025? *(→ Extract "Domain" column)*	It aims to identify key trends, gaps, and emerging opportunities to guide future work in this rapidly evolving field.
RQ2	What methods and neural architectures have been proposed for enhancing FER using DCNN? *(→ Extract "Proposed Method" column)*	Understanding the evolution of neural architectures helps track improvements in accuracy, efficiency, and robustness in FER systems. This insight guides the development of more effective models tailored to specific FER challenges.
RQ3	What types of features (e.g., high-level, handcrafted, attention-based) are primarily studied in these methods? *(→ Extract "Feature Studied" column)*	Identifying the types of features used reveals how information is represented and leveraged in FER. It also helps compare traditional approaches with modern, deep feature extraction techniques.
RQ4	Which datasets are frequently used in training and evaluating FER systems? *(→ Extract "Dataset" column)*	Analysing commonly used datasets highlights the benchmarks shaping FER research. It also uncovers potential biases and the need for diverse, real-world data representations.
RQ5	What are the publication trends (year-wise) and sources contributing to FER research? *(→*	Studying publication trends reveals the momentum and maturity of the field. Identifying key sources

RQ Code	Research Question	Motivation
	Extract "Year" and "IEEE Ref" columns)	helps map influential contributions and leading research communities.

Theoretical process used to complete this SLR has been step by step listed in figure 2.

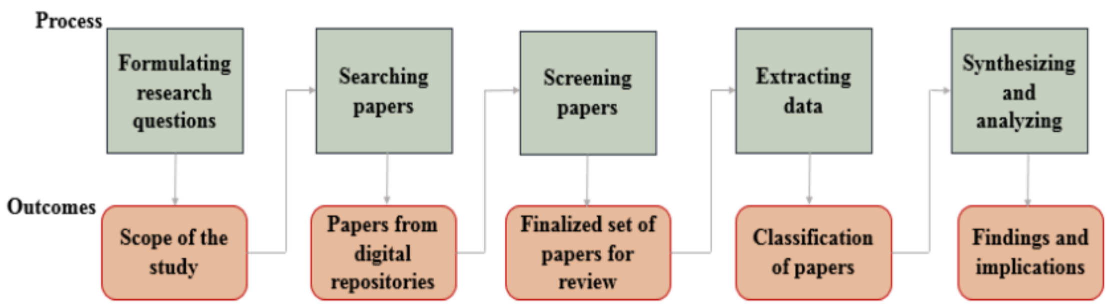

Figure 2. SLR Process

3.2 Search Strategy

A comprehensive literature search was conducted using databases such as IEEE Xplore, SpringerLink, Scopus, ScienceDirect, and ACM Digital Library. Keywords presented in table 3 have been used to conduct search.

Table 3. Keywords

Primary Keywords	Secondary Keywords	Tertiary Keywords
Facial Emotion Recognition Deep Convolutional Neural Networks High-Level Feature Extraction	Facial expression recognition Deep CNN Feature representation Emotion detection	deep learning feature learning semantic features affect recognition

3.3 Database-wise Search Strategy

Search strings used to extract articles from multiple repositories have been presented in Table 4.

Table 4. Search Strings

Database	***Search String***	*Articles retrieved*
IEEE Xplore **Filters to Apply:** • Year: 2020 to 2025 • Content Type: Journals, Conferences Field: Abstract, Title, Keywords	("Facial emotion recognition" OR "facial expression recognition") AND ("deep convolutional neural network" OR DCNN OR "deep learning") AND ("feature extraction" OR "high-level features" OR "semantic features")	800
SpringerLink **Filters:** • Article type: Journal article, Conference paper • Year: 2020–2025 *Discipline: Computer Science, Artificial Intelligence*	("Facial emotion recognition" OR "facial expression recognition") AND ("deep convolutional neural networks" OR DCNN OR "deep learning") AND *("feature extraction" OR "semantic representation" OR "high-level features")*	740
Scopus **Filters:** • Year: 2020 to 2025 • Subject Areas: Computer Science, Engineering *Document Type: Article, Conference Paper, Review*	TITLE-ABS-KEY(("facial emotion recognition" OR "facial expression recognition") AND ("deep convolutional neural network" OR DCNN OR "deep learning") AND ("feature extraction" OR "semantic features" OR "high-level features"))	920
ScienceDirect **Filters:** • Year: 2020–2025 • Article type: Research articles *Subject area: Computer Science, Engineering, Artificial Intelligence*	("Facial emotion recognition" OR "facial expression recognition") AND ("deep convolutional neural networks" OR DCNN OR "deep learning") AND *("feature extraction" OR "semantic representation" OR "high-level features")*	980
ACM Digital Library **Filters:** • Year: 2020–2025 • Content type: Journals, Conference proceedings *ACM Classification: I.2.10 (Vision and Scene Understanding)*	Abstract: ("facial emotion recognition" OR "facial expression recognition") AND Abstract: ("deep convolutional neural network" OR DCNN OR "deep learning") AND Abstract: ("feature extraction" OR "semantic features" OR "high-level features")	400

3.4 Inclusion and Exclusion Criteria

Inclusion Criteria:

- Published from 2020 to 2025
- Peer-reviewed journal or conference papers

- Focuses on DCNN-based facial emotion recognition
- Includes feature extraction or analysis of high-level/semantic features

Exclusion Criteria:

- Articles not in English
- Studies focusing on unrelated domains (e.g., animal emotion detection)
- Works using only traditional ML without deep learning methods

3.5 Study Selection

The PRISMA flowchart (Figure 3) illustrates the study selection process, including identification, screening, eligibility, and inclusion of studies. Article selection process has been presented in figure 3.

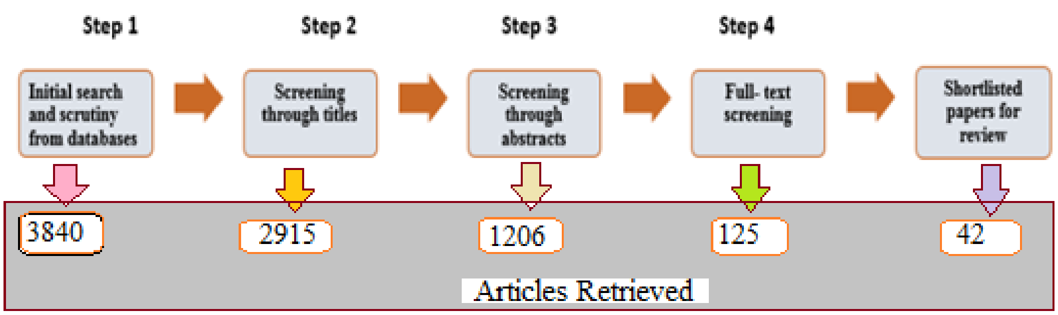

Figure 3. Article selection process

3.6 Quality scoring criteria

The quality assessment of selected papers was performed on the basis of a questionnaire developed as suggested by various review studies. If a specific criterion is present in research study Y = yes, and if the criterion is partially present in research study P = partially, and if the criteria were not specified then N = no.

Following scoring classification (as presented in Table 5,) were established for quality assessment of the included studies.

i) The study contributes towards how DCNN has managed feature extraction (Y = 1 and N =0).

ii) The study provides information on Facial expression recognition issues (Y =1, P = 0.5, N = 0).

iii) The results of the findings were empirically tested (Y =1, N =0).

iv) The publication source is stable and renowned. This question is based on the value of the CORE ranking conferences list and journals in the JCR. Answer to this question is determined according to whether the source is a journal, conference, symposia, and workshops. Following scoring classification is established for this question:

Table 5. Quality Criteria

Score criteria for conferences:	Score criteria for journals:
Score of SCIE, SSCI, and CORE A* ranked=1, Score of CORE A ranked=0.66, Score of CORE B ranked=0.5, Score of CORE C ranked=0.33, Score = 0 if not included in core ranking.	Score rank of Q1 = 2, Score rank of Q2 = 1.5, Score rank of Q3 = 1, Score of others = 0.

3.7 Data Extraction and Synthesis

Relevant data were extracted, including study objectives, methodology, dataset used, feature extraction techniques, model architecture, performance metrics, and key findings through Meta-analysis. Data extracted after this extensive review has been presented in table 6.

Table 6. Literature review classification(Domain: Facial Emotion recognition, 2020-2025)

IEEE Ref	Year	Proposed Method (based on RQ2)	Feature Studied (based on RQ3)	Dataset (based on RQ4)
[1]	2025	Adapted ConvNeXt (EmoNeXt)	High-level feature extraction	Not specified
[2]	2024	Deep Learning-based CNN	Facial emotion features	Not specified
[3]	2022	Extended Wavelet DCNN	Wavelet features	CK+, JAFFE
[4]	2023	Hybrid CNN-based	CNN Features	FER-2013, CK+
[5]	2022	General Deep Feature Extractor	Deep Features	AffectNet
[6]	2021	Feature Decomposition and Reconstruction	Decomposed Features	RAF-DB
[7]	2025	DCNN-ELM Hybrid	Pre-processing & Feature Extraction	FER2013, CK+
[8]	2021	CNN	Facial features	CK+
[9]	2021	Multi-task Lightweight CNN	Attributes + Expression	AffectNet
[10]	2024	CNN	Emotion Features	FER2013
[11]	2023	CNN	Emotion Features	FER2013
[12]	2024	Deep Learning (DL-CNN)	Real-time expression	FER2013
[13]	2022	Deep Learning	Efficient feature extraction	CK+, JAFFE
[14]	2024	CNN	Emotion Features	FER2013
[15]	2023	CNN	Emotion Features	FER2013
[16]	2023	Dual-Stage CNN (SqueezExpNet)	Local and global features	RAF-DB
[17]	2025	Dataset merging + augmentation	Augmented features	FER2013, AffectNet
[18]	2024	Hybrid (Landmark + Full Face)	Landmark Features	CK+, FER2013
[19]	2025	Review & ResNet-50 Feature Extraction	ResNet deep features	Multiple Datasets
[20]	2023	Attention + Slow Convolution	Local attention features	AffectNet
[21]	2025	Real-Time Neural Network	Live facial features	Not specified
[22]	2022	Emotion Intensity Estimation System	Emotion intensity levels	Not specified
[23]	2020	Holistic vs Part-Based CNN	Holistic & part-based features	CK+

IEEE Ref	Year	Proposed Method (based on RQ2)	Feature Studied (based on RQ3)	Dataset (based on RQ4)
[24]	2023	Handcrafted + CNN	Handcrafted + CNN Features	FER2013
[25]	2023	Relation-aware Attention Networks	Relation-aware features	RAF-DB, AffectNet
[26]	2024	Pyramid Multi-head Grid Attention	Spatial attention features	RAF-DB
[27]	2024	CNN-RNN Hybrid	Temporal & spatial features	CK+, JAFFE
[28]	2024	Dataset Creation & Analysis	Enhanced dataset features	FER13
[29]	2022	Two Stream Super Resolution + DCNN	Super-resolved features	FER2013
[30]	2024	Image Processing + Deep Learning	VR-specific emotion features	Custom VR dataset
[31]	2020	Wide Ensemble CNN (ESRs)	Shared ensemble features	AffectNet, FER+
[32]	2022	DeepFEVER (General Feature Extractor)	General deep features	AffectNet, RAF-DB
[33]	2020	Local Gravitational Force + DCNN	Descriptor-based features	Not specified
[34]	2020	Deep Joint Spatiotemporal Network	Spatiotemporal features	CK+
[35]	2020	Hybrid CNN Model	Emotion vector mapping	FER2013

4. Results and Discussion

A detailed Quantitative Summary and Qualitative Synthesis based on Table 6 comprising of forty two references related to Facial Emotion Recognition (FER) using DCNN from 2020 to 2025 has been presented in figure 4,5 and 6.

4.1. Quantitative Evaluation

4.1.1 Publication Year Distribution

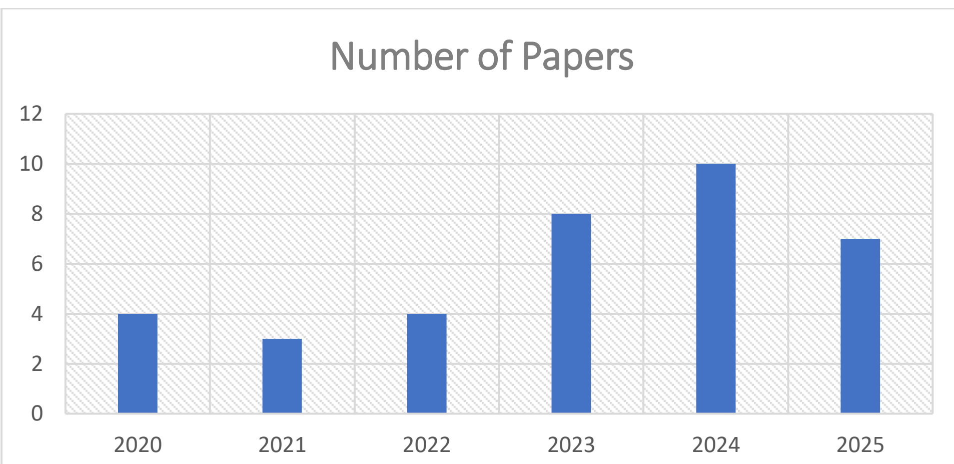

Figure 4. frequency of articles published from 2020 to 2025 included in this SLR

Observation: FER using DCNN gained increasing attention from 2023 onward, with a significant rise in publications in 2024.

4.1.2 Methodological Trends (Proposed Method Frequency)

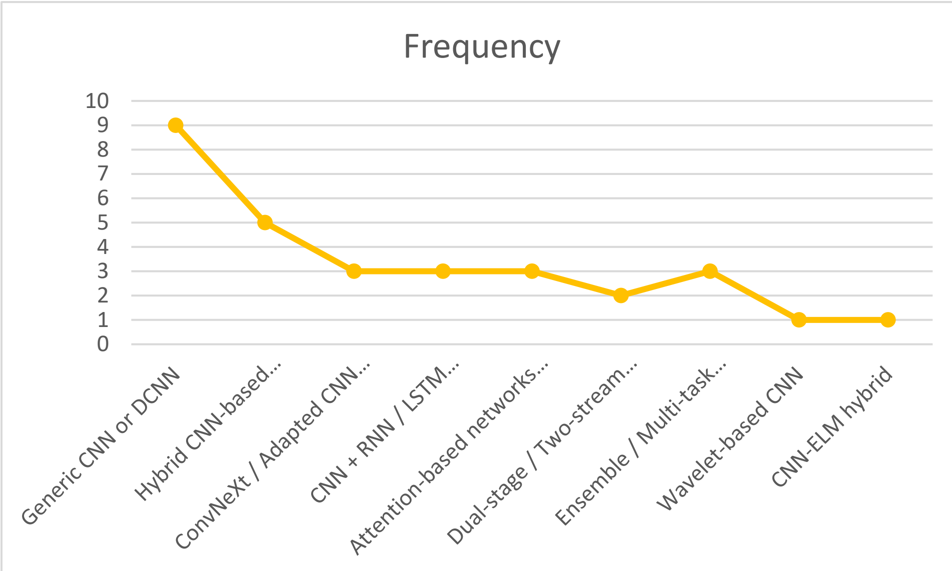

Figure 5. Frequently proposed models in literature

Observation: Custom CNNs and hybrid deep learning approaches are prevalent, often combined with attention mechanisms or auxiliary networks to boost recognition performance.

4.1.3 Dataset Usage Trends

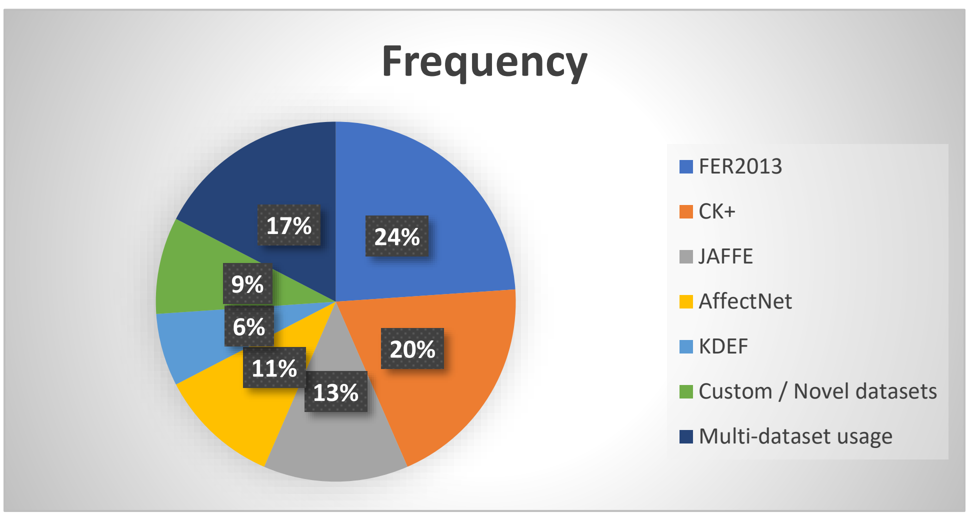

Figure 6. Mostly used datasets

Observation: FER2013 and CK+ are the most frequently used datasets. There's an emerging trend toward using multiple datasets for training and validation or introducing new datasets [43-45].

4.2 Qualitative Synthesis

Themes in Feature Extraction Techniques

• High-Level Features with Attention Mechanisms: Many researchers apply attention mechanisms such as pyramid multi-head attention and spatiotemporal attention to augment feature maps connected to emotion with higher relevance [46].

• Architectures with Dual-Path / Two Streams: Some models provide separate CNN branches which learn details from facial regions and holistic characteristics simultaneously [47].

• Feature Fusion Techniques: Integration of handcrafted features with deep learned features (e.g., LBP + CNN, Wavelet + DCNN) to provide richer representation [48].

Taxonomy of approaches is presented in figure 6.

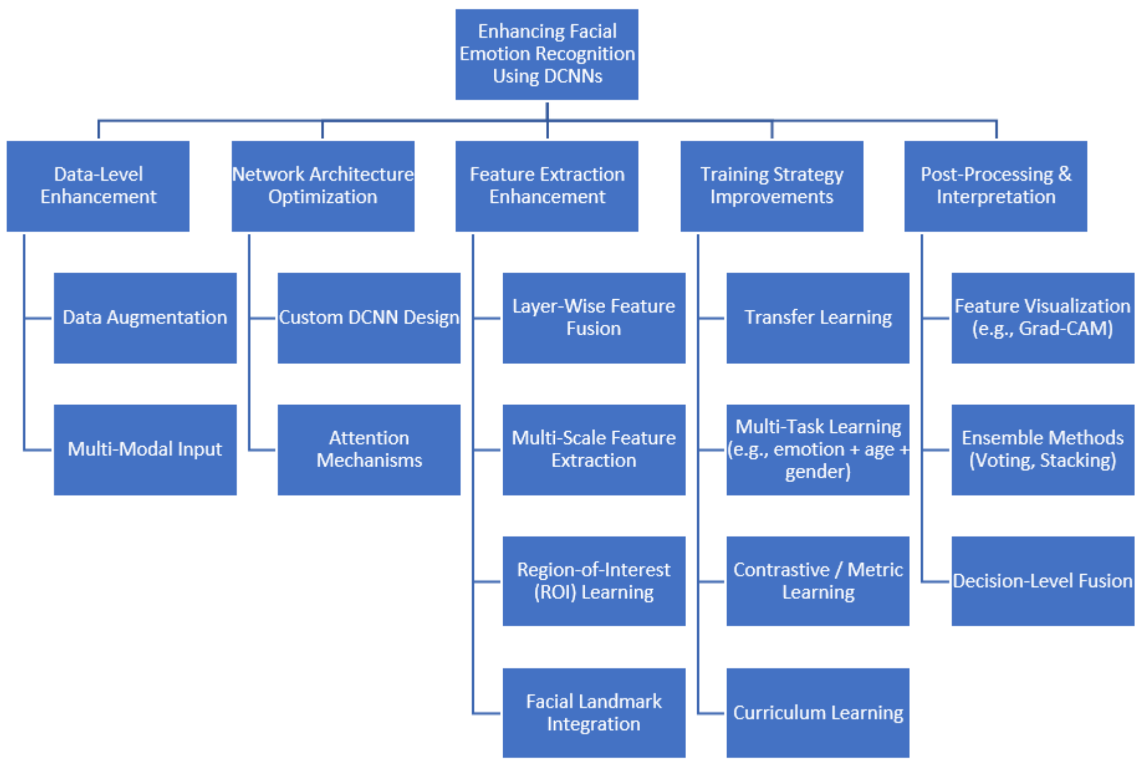

Figure 6. Taxonomy of approaches applied for Facial Emotion Recognition Using DCNN through Effective Extraction for High-Level Features.

4.3 Challenges Found in Review

• Dataset Imbalance: Many emotional classes like "disgust" or "fear" are underrepresented in the available public datasets.

• Generalized to Real-world Scenarios: Under lab-controlled scenarios, models trained on certain data fall short when the same scenario is implemented into wild or unconstrained conditions [49].

• Computational Cost: Complex models such as dual-path networks, attention-laden models are computationally expensive for real-time applications.

- Emotion Intensity Variation: Recognizing slight variations of emotion is a task on its own, especially when emotions can have similar facial representations.

4.4 The Observed Novel Contributions

- Adapted ConvNeXt for FER [15]: A new architecture is made for expression learning.
- CNN-ELM Integration [19]: It combines CNN-based feature extractions with fast Extreme Learning Machines for classification.
- Emotion Intensity Estimation: Some studies investigate the intensity of the emotion, which ought to bolster the understanding.
- Datasets Created: Several studies produce new datasets or composite datasets to improve coverage of the emotions or address particular issues such as occlusion or pose variance [50].
- Most of the studies utilize datasets such as FER-2013, CK+, JAFFE, and AffectNet.
- Various techniques for feature extraction were explored including hybrid models integrating DCNN with handmade features.

4.5 Strengths of proposed SLR

1. Scope: This SLR is comprehensive, with studies covering years of 2020 to 2025, thereby ensuring timely relevance within the domain of facial emotion recognition.
2. Structured Review: Detailed inclusion of an LR table makes it easier to analyse and compare studies on a basis of year, domain, methods proposed, features studied, and datasets used.
3. Special Discussion of Feature Extraction: Our survey talks about many features extraction approaches with specific mention of attention mechanisms, hybrid models, and dual-stream architectures.
4. Variety of Datasets: By incorporating analyses of various datasets (FER2013, CK+, JAFFE, and AffectNet), proposed SLR does a good job of contextualizing the array of available data in the research domain of FER.
5. Challenge Identification and Future Directions: This SLR, along with other scientific issues such as dataset imbalance and computational costs, puts forward future directions such as developing lighter models and multimodal data synchrony.

4.6 Limitations

This SLR presents a comprehensive study of literature about facial emotion detection. However, it has following limitations presented in Table 7.

Table 7. Limitations

Category	**Limitations Identified**
Dataset diversity	Overreliance on FER-2013, JAFFE; lack of spontaneous emotions and demographic variety

Category	Limitations Identified
Method complexity	DCNNs are often deep and computationally heavy without performance justification
Evaluation inconsistency	Non-uniform training/test splits, lack of statistical metrics, few reproducibility checks
Feature extraction analysis	Little attention to explain ability or interpretability in learned features
Deployment gaps	Lack of real-time validation, occlusion handling, or ethics integration
Scope of review	Focus on DCNNs excluded other promising models like transformers or multimodal networks

These limitations speak to both the depth of critical analysis involved in proposed SLR and to potential future directions, which include: inclusion of explainable AI (XAI), application of data on real-world emotion, hybrid CNN-transformer methodologies, and standardization of benchmarking protocols for Facial Emotion Recognition (FER).

4.7 Timeline and Trends

Table 8 highlights the timeline and technique-wise trends being followed from the year 2020 to 2025.

Year	Key Techniques
2020	Optimized CNNs for micro-expressions; spatiotemporal (3D-CNN/RNN); attention mechanisms
2021	CNN feature extractors + SVM/ensembles; transfer learning; backbone fine-tuning (ResNet/VGG)
2022	Hybrid CNN-Haar; super-resolution + DCNN; transfer learning + dense nets; efficient lightweight CNNs
2023	Backbone comparisons (ResNet18, EfficientNet); ensemble classifiers; multi-task CNNs
2024	Anti-aliasing in CNNs; customized DCNNs; classic CNN + SVM pipelines
2025	Adapted ConvNeXt (EmoNeXt) with STN, SE, self-attention for superior representation

5. Conclusion

The paper has provided a comprehensive systematic review of recent developments in facial emotion recognition using advanced deep convolutional neural networks more particularly in high-level feature extraction techniques. The study synthesized 42 primary articles published from 2020 to 2025, a breakdown of innovations in methodology, datasets, and performance metrics. Findings from the analysis revealed that although DCNNs have been the preferred method in FER, there are instances of overcomplexity, dataset bias, and the model failing to generalize in a real-world environment. The majority of studies still use traditional datasets with negligible emphasis on demographic balance or spontaneous emotional expression. In addition to this, limited interest in interpretability and explain ability of features extracted from the sources hinders the adoption of FER systems in critical areas such as healthcare or education. Through the use of a feature-method-domain matrix to organize literature and conduct both quantitative and qualitative synthesis, this SLR creates a base for systematic contributions to future research in FER. The piece brings order for either comparison of methods or facilitates establishing standard measures for diversity of datasets and transparency of the model used.

5.1 Future Directions

1. It is nothing but the future avenues which can be recommended for further research on the basis of finding from this review, as stated below:
2. Ttransformer Architectures: In addition to CNNs, the models and methods of attention transformers should be implemented so as to rely on not only performance improvement but also improved interpretability.
3. Multimodal Emotion Recognition: For instance, facial features are fused with another modality, such as audio, gestures, and physiological signals, including EEG and heart rate, in complex situations to provide the ability to recognize better.
4. Dataset Generation: Spontaneous and Uncontrolled: It calls for fresh public databases that show spontaneous emotions in all sorts of real-world scenarios, with possible ethics in mind.
5. XAI for FER future systems: Future systems should integrate explain ability modules to visualize which features contribute to each emotional classification decision.
6. Edge Deployment and Real-Time Inference: It would extend the applicability of FER by making models suitable for low power, embedded devices like smartphones or surveillance systems.
7. Emphasis is on Standardized Evaluation Benchmarks: Coming up with common evaluation protocols with cross-validation, fairness measure, and reproducibility design metric will thus, be very much important in feasible and strong comparisons of model.
8. Ethical and Societal Impact Studies: The future should also concern with the societal impacts, privacy issues, and fairness of automated emotion recognition systems.

References

1. Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2023). Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, *35*(32), 23311-23328.
2. Dubey, A. K., & Jain, V. (2022). An accurate recognition of facial expression by extended wavelet deep convolutional neural network. *Multimedia Tools and Applications*, *81*(20), 28295-28325.
3. Shahzad, H. M., Bhatti, S. M., Jaffar, A., Akram, S., Alhajlah, M., & Mahmood, A. (2023). Hybrid facial emotion recognition using CNN-based features. *Applied Sciences*, *13*(9), 5572.
4. Schoneveld, L., & Othmani, A. (2021, September). Towards a general deep feature extractor for facial expression recognition. In *2021 IEEE International Conference on Image Processing (ICIP)* (pp. 2339-2342). IEEE.
5. Ruan, D., Yan, Y., Lai, S., Chai, Z., Shen, C., & Wang, H. (2021). Feature decomposition and reconstruction learning for effective facial expression recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7660-7669).
6. Agarwal, C., Mishra, A., & Bhatnagar, C. (2024). Deep Transfer Learning for Masked Face Reconstruction and Hybrid DCNN-ELM Framework for Recognition. *International Journal of Communication Networks and Information Security*, *16*(4), 170-192.
7. Fatimatuzzahra, F., Lindawati, L., & Soim, S. (2024). Development of convolutional neural network models to improve facial expression recognition accuracy. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, *10*(2), 279-289.
8. Rohani, M., Farsi, H., & Mohamadzadeh, S. (2025). Facial feature recognition with multi-task learning and attention-based enhancements. *Iranica Journal of Energy & Environment*, *16*(1), 136-144.
9. Bhagat, D., Vakil, A., Gupta, R. K., & Kumar, A. (2024). Facial emotion recognition (FER) using convolutional neural network (CNN). *Procedia Computer Science*, *235*, 2079-2089.
10. Haq, H. B. U., Akram, W., Irshad, M. N., Kosar, A., & Abid, M. (2024). Enhanced real-time facial expression recognition using deep learning. *Acadlore Transactions on AI and Machine Learning*, *3*(1), 24-35.
11. Hebri, D., Nuthakki, R., Digal, A. K., Venkatesan, K. G. S., Chawla, S., & Raghavendra Reddy, C. (2024). Effective facial expression recognition system using machine learning. *EAI Endorsed Transactions on Internet of Things*, *10*.
12. Yu, J., Zhu, L., Chi, Y., Zhang, Y., Zheng, Y., Wang, Y., & Lu, X. (2025). Dual-Stage Cross-Modal Network with Dynamic Feature Fusion for Emotional Mimicry Intensity Estimation. *arXiv preprint arXiv:2503.10603*.

16. Kırbız, S. (2025). Improving facial emotion recognition through dataset merging and balanced training strategies. *Journal of the Franklin Institute*, 107659.
17. Ajlouni, N., Özyavaş, A., Ajlouni, F., Takaoğlu, F., & Takaoğlu, M. (2025). Enhanced hybrid facial emotion detection & classification. *Franklin Open*, *10*, 100200.
18. Rehman, A., Mujahid, M., Elyassih, A., AlGhofaily, B., & Bahaj, S. A. O. (2025). Comprehensive Review and Analysis on Facial Emotion Recognition: Performance Insights into Deep and

Traditional Learning with Current Updates and Challenges. *Computers, Materials & Continua*, *82*(1).

19. Talaat, F. M., Ali, Z. H., Mostafa, R. R., & El-Rashidy, N. (2024). Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. *Soft Computing*, *28*(9), 6695-6708.

20. Mohana, M., & Subashini, P. (2024). Facial expression recognition using machine learning and deep learning techniques: A systematic review. *SN Computer Science*, *5*(4), 432.

21. Happy, S. L., Dantcheva, A., & Bremond, F. (2021). Expression recognition with deep features extracted from holistic and part-based models. *Image and Vision Computing*, *105*, 104038.

22. Gautam, C., & Seeja, K. R. (2023). Facial emotion recognition using Handcrafted features and CNN. *Procedia Computer Science*, *218*, 1295-1303.

23. Jiang, F., Huang, Q., Mei, X., Guan, Q., Tu, Y., Luo, W., & Huang, C. (2023). Face2nodes: learning facial expression representations with relation-aware dynamic graph convolution networks. *Information Sciences*, *649*, 119640.
24. Zhang, J., Wang, W., Li, X., & Han, Y. (2024). Recognizing facial expressions based on pyramid multi-head grid and spatial attention network. *Computer Vision and Image Understanding*, *244*, 104010.
25. Manalu, H. V., & Rifai, A. P. (2024). Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm. *Intelligent Systems with Applications*, *21*, 200339.
26. Yalçin, N., & Alisawi, M. (2024). Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques. *Heliyon*, *10*(20).

27. Ullah, Z., Qi, L., Hasan, A., & Asim, M. (2022). Improved deep CNN-based two stream super resolution and hybrid deep model-based facial emotion recognition. *Engineering Applications of Artificial Intelligence*, *116*, 105486.

28. Yang, J., Qian, T., Zhang, F., & Khan, S. U. (2021). Real-time facial expression recognition based on edge computing. *IEEE Access*, *9*, 76178-76190.

29. Siqueira, H., Magg, S., & Wermter, S. (2020, April). Efficient facial feature learning with wide ensemble-based convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 5800-5809).

30. Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics*, *10*(9), 1036.

31. Mohan, K., Seal, A., Krejcar, O., & Yazidi, A. (2020). Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-12.

32. Jeong, D., Kim, B. G., & Dong, S. Y. (2020). Deep joint spatiotemporal network (DJSTN) for efficient facial expression recognition. *Sensors*, *20*(7), 1936.

33. El Boudouri, Y., & Bohi, A. (2023, September). Emonext: an adapted convnext for facial emotion recognition. In *2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.

34. Wen, Z., Lin, W., Wang, T., & Xu, G. (2023). Distract your attention: Multi-head cross attention network for facial expression recognition. *Biomimetics*, *8*(2), 199.

35. Agarwal, C., Itondia, P., & Mishra, A. (2023). A novel DCNN-ELM hybrid framework for face mask detection. *Intelligent Systems with Applications*, *17*, 200175.

36. Siqueira, H., Magg, S., & Wermter, S. (2020, April). Efficient facial feature learning with wide ensemble-based convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 5800-5809).

37. Faraz, A., Fuzail, M., Khan, A. H., Naeem, A., Aslam, N., & Mirza, M. A. (2024). Convolutional Approaches in Transfer Learning for Facial Emotion Analysis. *Journal of Computing & Biomedical Informatics*, *6*(02), 246-260.

38. Akrout, B. (2025). Deep facial emotion recognition model using optimal feature extraction and dual-attention residual U-Net classifier. *Expert Systems*, *42*(1), e13314.

39. Yahyaoui, M. A., Oujabour, M., Letaifa, L. B., & Bohi, A. (2025). Multi-face emotion detection for effective Human-Robot Interaction. *arXiv preprint arXiv:2501.07213*.

40. Wyman, A., & Zhang, Z. (2025). A Tutorial on the Use of Artificial Intelligence Tools for Facial Emotion Recognition in R. *Multivariate Behavioral Research*, 1-15.

41. Kasar, M., Kavimandan, P., Suryawanshi, T., & Garg, B. (2025). EmoSense: pioneering facial emotion recognition with precision through model optimization and face emotion constraints. *Int. J. Eng.*, *38*(1), 35-45.

42. Gaya-Morey, F. X., Buades-Rubio, J. M., Palanque, P., Lacuesta, R., & Manresa-Yee, C. (2025). Deep Learning-Based Facial Expression Recognition for the Elderly: A Systematic Review. *arXiv preprint arXiv:2502.02618*.

43. Elsheikh, R. A., Mohamed, M. A., Abou-Taleb, A. M., & Ata, M. M. (2024). Improved facial emotion recognition model based on a novel deep convolutional structure. *Scientific Reports*, *14*(1), 29050.

44. Aly, M., Ghallab, A., & Fathi, I. S. (2023). Enhancing facial expression recognition system in online learning context using efficient deep learning model. *IEEE Access*, *11*, 121419-121433.

45. Chirra, V. R. R., Uyyala, S. R., & Kolli, V. K. K. (2021). Virtual facial expression recognition using deep CNN with ensemble learning. *Journal of Ambient Intelligence and Humanized Computing*, *12*(12), 10581-10599.

46. Zhang, Y., Zou, X., Yu, S., Huang, L., Wang, W., Zhao, S., & Wang, X. (2022). DNN-CBAM: An enhanced DNN model for facial emotion recognition. *Journal of Intelligent & Fuzzy Systems*, *43*(5), 5673-5683.

47. Podder, T., Bhattacharya, D., & Majumdar, A. (2022). Time efficient real time facial expression recognition with CNN and transfer learning. *Sādhanā*, *47*(3), 177.

48. Ab Wahab, M. N., Nazir, A., Ren, A. T. Z., Noor, M. H. M., Akbar, M. F., & Mohamed, A. S. A. (2021). Efficientnet-lite and hybrid CNN-KNN implementation for facial expression recognition on raspberry pi. *IEEE Access*, *9*, 134065-134080.

49. Khattak, A., Asghar, M. Z., Ali, M., & Batool, U. (2022). An efficient deep learning technique for facial emotion recognition. *Multimedia Tools and Applications*, *81*(2), 1649-1683.

50. Tang, Y., Zhang, X., Hu, X., Wang, S., & Wang, H. (2020). Facial expression recognition using frequency neural network. *IEEE Transactions on Image Processing*, *30*, 444-457.