Journal of Electrical Engineering and Computer (JEECOM)
Vol. 7, No. 1 (2025), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Facial Expression Based Emotion Recognition

Mohamed Ibrahim Abdi[1]ORCID, Burhan Ergen[1]ORCID
[1] Firat University, Engineering Faculty, Computer Engineering Department, Elazig, Türkiye.

Article Info

Article history:

Received April 12, 2025
Revised April 15, 2025
Accepted April 16, 2025

Keywords:

FacialEmotion Recognition(FER), FER2013 Dataset
VGG16
Emotion Analysis
Deep Learning

ABSTRACT

Human communication is not limited to spoken and written language; it also includes a rich layer of nonverbal cues, with facial expressions being a key element in conveying emotions. This study presents the development and evaluation of a deep learning model for Facial Emotion Recognition (FER) using the VGG-16 architecture and the FER2013 dataset, which contains over 35,000 images depicting seven distinct emotions in natural settings. The goal of this research is to enhance emotion recognition accuracy and overall performance beyond current benchmarks. To achieve this, transfer learning was applied using pre-trained VGG-16 weights, and the classification layers were restructured and fine-tuned for emotion detection. Extensive preprocessing techniques, including normalization and data augmentation, were used to enhance generalization and reduce overfitting. The proposed model achieved an accuracy of 85.77%, outperforming several previous VGG-16-based FER models. The performance was assessed using standard metrics, such as accuracy, precision, recall, and F1-score, confirming the model's robustness. The critical factors contributing to this success included hyperparameter optimization and regularization methods, particularly dropout and early stopping. The model also demonstrated its ability to extract meaningful features from low-resolution images. Future research can expand its applicability in real-world domains, such as transportation safety, surveillance, and customer behavior analysis, by incorporating more diverse datasets and exploring advanced architectures.

Corresponding Author:
Mohamed Ibrahim Abdi
Fırat University, Elazığ and 23000, Turkiye
Email: Mohamedqadar280@gmail.com

1. INTRODUCTION

In recent years, Facial Emotion Recognition (FER) technology has seen significant growth in academic research and industrial applications. FER is a computer vision-based AI technology designed to automatically analyze emotional states using facial expressions. This technology has been applied across various fields, including healthcare, education, security systems, the automotive industry, and social media analysis [1],[2]. The integration of deep and transfer learning techniques into FER systems has led to substantial improvements in their accuracy and performance [3]. FER systems intersect computer vision, artificial intelligence, and deep learning, facilitating the digitization and analysis of human emotional states. These methods enhance human-computer interactions and serve as tools in health applications, such as the early diagnosis of depression and stress [4].

The application of FER systems in diagnosing autism spectrum disorder (ASD) underscores their importance in clinical processes [5]. CNN (Convolutional Neural Networks) are at the forefront of FER applications. CNNs achieve high accuracy on complex datasets owing to their ability to learn low- and high-level features from facial expressions through a multi-layered structure [6].

For FER (Facial Emotion Recognition) systems to be effectively implemented in real-time applications, it is imperative that they are optimized to accommodate resource constraints such as processing power, memory, and energy. This necessitates efforts to design lightweight and efficient model architectures [8, 9]. Furthermore, there is a pressing need to develop innovative solutions for data preprocessing techniques, model configurations, and learning strategies to enhance the overall performance of FER systems [10].

In evaluating MRI technologies, it is crucial to consider technical precision and socio-cultural and ethical dimensions for sustainable and ethical application. Facial expressions can vary according to cultural norms, demographics, and social context. Studies using diverse datasets are recommended to enhance the generalizability of the models [11]. FER systems have diverse applications in various fields. In education, they can assess students' emotional states to customize teaching strategies [12]. In the automotive field, systems that detect driver fatigue or distraction can enhance road safety [13].

This study aims to surpass the success rates reported in the existing literature by developing a deep learning model based on the VGG-16 architecture and utilizing the FER2013 dataset. The model's hyperparameters were meticulously optimized, resulting in a significant enhancement compared with previous findings. In the domain of FER (facial emotion recognition), this model is anticipated to enhance passenger comfort and mitigate safety risks by assessing passengers’ emotional states during long-distance travel. Additionally, the model's capability to conduct emotion analysis over video streams indicates its potential applicability across various industries to enhance customer experience.

2. METHOD

Facial emotion recognition (FER) has evolved significantly, shifting from traditional methodologies to deep learning-based methods. Conventional methods rely on manual feature extraction from facial images, followed by classification, whereas deep learning techniques integrate these processes into a unified framework, resulting in improved accuracy and robustness. In recent years, CNN(Convolutional Neural Networks) have become increasingly prominent in FER literature because of their ability to effectively capture complex visual patterns.

This study employs the VGG-16 architecture, a widely recognized The Visual Geometry Group (VGG) at the University of Oxford developed a CNN [20]. as shown in Figure 1 VGG-16 is a deep and homogeneous structure comprising 16 layers—13 convolutional layers and three fully connected layers—with all filters uniformly set to 3 × 3 size. Its straightforward yet efficient design has facilitated notable success in various computer-vision tasks, including image classification. Owing to its depth, VGG-16 excels at learning intricate hierarchical structures in visual data, making it particularly well-suited for emotion recognition from facial expressions. Figure 1 illustrates the architecture of the VGG16. Each convolutional layer employed a kernel size of 3 × 3. The initial layers incorporated 64 filters, which increased to 128 filters in the subsequent layers. In the later stages, 256 filters were utilized, whereas the final layers implemented 512 filters. The diagram depicts 2 × 2 max pooling layers applied following certain convolutional layers. The output from the max pooling layers is connected to the fully connected layer, which plays a role in determining the class or label for the model's classification task.

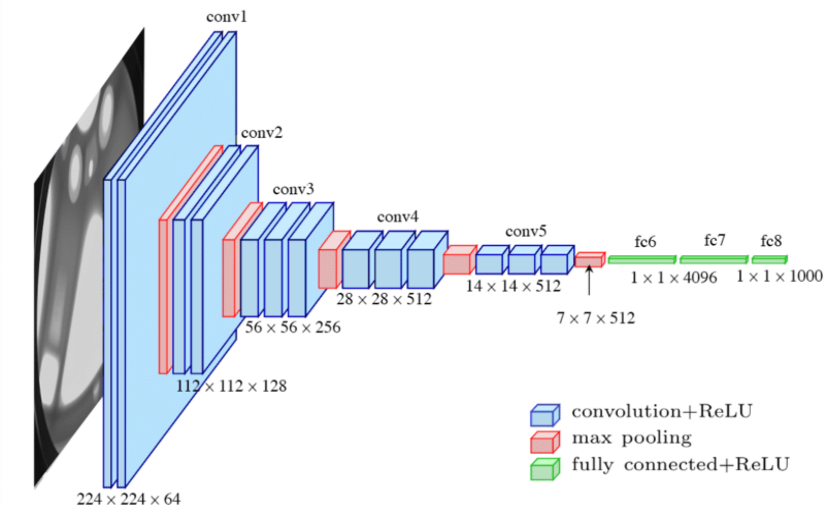


Figure 1 The architecture of VGG16 [20].

2.1. Dataset

Facial Emotion Recognition (FER) is a well-established area of research, supported by a multitude of available datasets. This study employs the FER2013 dataset as the primary source of data. FER2013 is a highly esteemed dataset, having been utilized in ICML competitions and numerous scholarly articles. It is recognized as one of the more difficult datasets, with human-level precision estimated to be around 75±5%, while leading published studies have achieved a test accuracy of 75.8% [14]. This dataset, accessible via Kaggle, comprises 35,887 grayscale images, each standardized to a resolution of 48x48 pixels. However, FER2013 exhibits an uneven distribution across its seven facial expression categories, with the following counts: According to the data, emotional responses are as follows: Happy is most frequent with 8,989 occurrences, followed by Neutral with 6,198, and Sad with 6,077. Fear has 5,121 instances, anger 4,953, and surprise 4,002. Disgust is least common, with 547 instances recorded. [7].

2.2. data preprocessing

The FER2013 dataset was categorized into three distinct usage groups: Training, Validation, and Testing. The pixel values underwent image resizing to 48 × 48 dimensions, conversion from grayscale to color with three channels, and normalization. Furthermore, the labels were transformed using one-hot encoding to facilitate their use in categorical format. Data augmentation is a crucial step in the training process, as it enables the model to learn from a broader range of variations. The images were subjected to operations such as rotation, panning, horizontal and vertical translation, brightness adjustment and truncation. These procedures mitigated the overfitting issue and enhanced the generalization capability of the model.

2.3. Model Architecture and Training Process

The model introduced in this study represents an adaptation of the VGG-16 architecture, originally pretrained on ImageNet for the classification of 1000 categories, and has been specifically modified for facial emotion recognition (FER) utilizing the FER2013 dataset. It processes input images with dimensions of 48 × 48 × 3 in color, employing five sequential blocks of 3 × 3 convolutional filters with ReLU activation, followed by 2 × 2 max-pooling to facilitate deep hierarchical feature extraction. The original dense layers of VGG-16 were replaced with newly designed fully connected layers tailored to categorize images into seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutrality. This categorization was achieved using a softmax activation function in the output layer, which consisted of seven neurons corresponding to these categories. During data preparation, preprocessing steps such as resizing, cropping, and normalization were applied to ensure a clean data set. The training process involved fine-tuning by freezing the lower layers of VGG-16 and reconfiguring the new fully connected layers for emotion recognition. The Adam optimizer was employed with categorical cross-entropy as the loss function to enhance the model performance. The model was evaluated using metrics such as accuracy, precision, recall, AUC, and macro F1-score. In the testing phase, a new face image is input, the facial region is cropped, and the trained model outputs probability scores, selecting the emotion class with the highest predicted probability as the final output class. The training and testing workflow depicted in Figure 2 illustrates a systematic method for adapting a general-purpose deep-learning model for FER. The combination of fine-tuning and robust pre-training enables the model to effectively extract emotional cues from low-resolution images. This architecture ensures reusability and supports real-time applications, demonstrating its strong potential for use in human-computer interaction, mental health monitoring, and enhancing user experience.

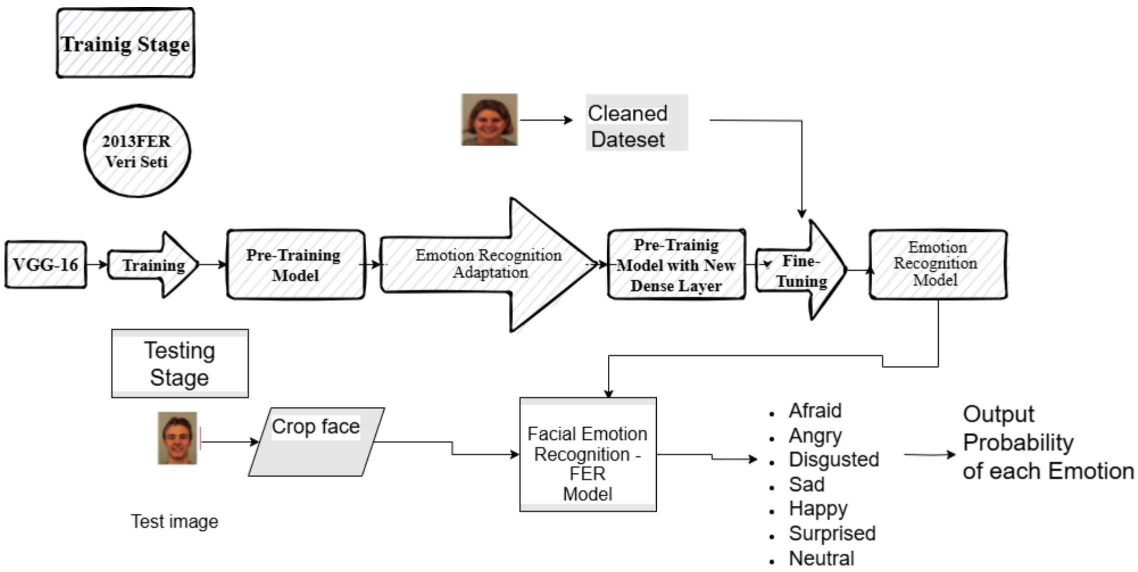


Figure 2. **Model's Architecture for Facial Emotion Recognition**

3. RESULTS AND DISCUSSION

A VGG-16-based deep learning model was developed to perform facial emotion recognition (FER) tasks using the FER2013 dataset. Transfer learning was applied during model training, whichein the classification layers were restructured using pre-trained VGG-16 weights. To enhance the generalization capacity of the model and substantially mitigate overfitting, data augmentation techniques, hyperparameter optimization, dropout layers, and early stopping strategies were systematically integrated.

Following the training and testing phases, the model achieved an accuracy of **85.77%** on the FER2013 dataset, which represents a significant performance improvement compared with similar VGG-16-based studies in literature. The model's success can be attributed to the selection of a robust pretrained architecture, application of effective data preprocessing and augmentation techniques, integration of transfer learning, and successful implementation of overfitting prevention strategies. Notably, the VGG-16 architecture demonstrated a strong capability for extracting discriminative features, even from low-resolution images, which played a crucial role in achieving high classification performance.

A comprehensive evaluation was conducted using multiple metrics, including **accuracy, precision, recall, and F1-score**, providing an in-depth analysis of the model performance. Furthermore, the adoption of transfer learning facilitated efficient training, enabling the model to produce reliable results in a shorter time frame. These findings underscore the effectiveness of the proposed approach in advancing FER systems, particularly in scenarios that require robust feature extraction and generalization across diverse facial expressions. Figure 4 depicts the accuracy and loss metrics of the model during training. The accuracy graph shows the values for the training and validation datasets at each epoch. The blue line denotes the training accuracy, which initially fluctuated but stabilized at approximately 0.86. The orange line represents the validation accuracy, which is lower than the training accuracy. This discrepancy suggests that the model fits the training data well but struggles with the test data. The loss graph presents the training and validation loss values per epoch. The training loss starts high and decreases rapidly, whereas the validation loss remains stable at lower levels. These observations indicate effective learning during the training.

.

Figure 3 The emotional prediction images obtained from the FER2013 dataset

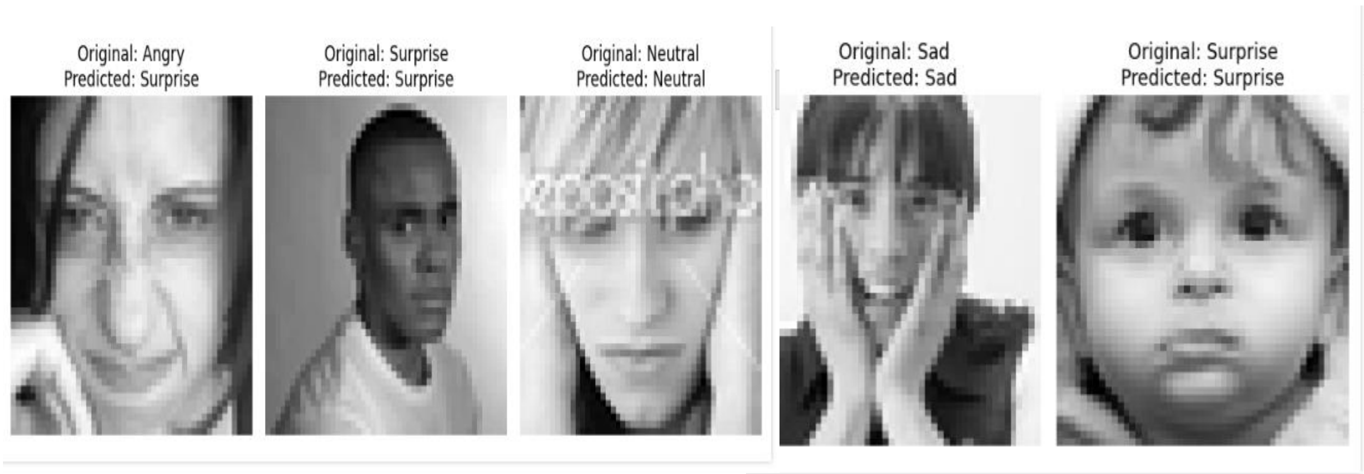
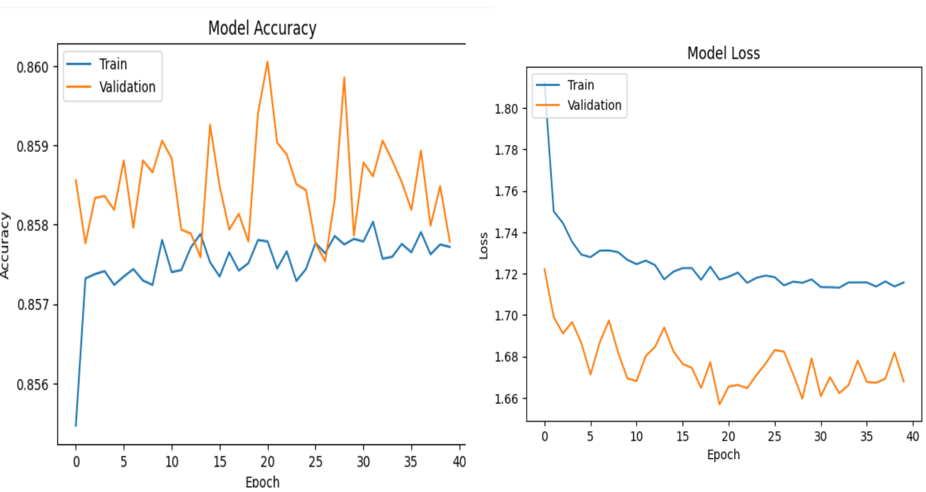


Figure 3 The emotional prediction images obtained from the FER2013 dataset



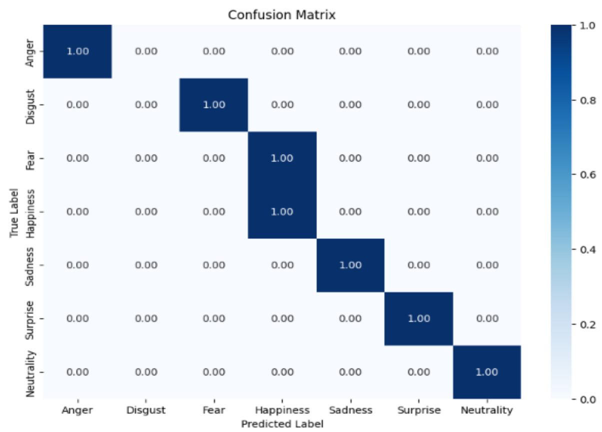


Figure 4 Model Accurancy and Model loss **figure 5 Confusion Matrix**

Therefore, we developed a VGG16-based deep learning model for facial emotion recognition (FER), and its performance was evaluated on the FER2013 dataset. As a result of the experiments, the model achieved an accuracy rate of 85.77%. This result can be considered as a very successful performance when compared to similar studies in literature.

Comparison with VGG-16-Based Studies in Literature

Table I presents a comparative overview of previous studies utilizing the VGG-16 architecture in conjunction with the FER2013 dataset. This table enables an objective evaluation of the proposed model's performance relative to existing methods. Study [17] reported an accuracy of 75.8%, while [18] and [19] achieved 69.40% and 72.30% accuracy, respectively. Study [20] documented an accuracy of 73.28%, representing one of the highest results in the current literature. In contrast, the model developed in this study achieved an accuracy of 85.77% using the same architecture and dataset, indicating a significant performance enhancement.

This improvement can be attributed to the effective integration of transfer learning, comprehensive preprocessing procedures, and a structured training pipeline incorporating regularization techniques such as dropout and early stopping. These components collectively enhanced the generalization capability and learning efficiency of the model. Furthermore, the use of transfer learning reduced the training time and improved the performance on previously unseen data.

The findings confirm the reliability of the VGG-16-based approaches and highlight the advancements introduced by the proposed method. The results demonstrate strong potential for real-time applications in fields requiring accurate emotional recognition, including driver monitoring systems, healthcare diagnostics, and adaptive human-computer interaction technologie

Table 1 **Facial Emotion Expression Recognition Results with VGG16 Using FER2013 Data Set**

Studies	Model	Accuracy Rate	Data Set
[17]	VGG16	75.8%	FER2013
[18]	VGG16	69.40%	FER2013
[19]	VGG16	74.11%	FER2013
[20]	VGG16	73.28%	FER2013
Our study	**VGG16**	**85.77%**	**FER2013**

This analysis demonstrates that the proposed model achieves a superior accuracy rate compared to analogous studies in the existing literature. Notably, the implementation of data augmentation and hyperparameter optimization substantially enhanced the model performance.

Future research indicates the potential for effective application of Facial Emotion Recognition (FER) technology across various domains, including security systems, public transportation, customer service, and psychological assessment. To further augment the efficacy of the model, it is anticipated that the integration of larger and more diverse datasets, alongside advanced hardware infrastructures and innovative deep learning techniques, will be pursued.

4. CONCLUSION

A deep learning model based on the VGG-16 architecture was developed for facial emotion recognition (FER) using the FER2013 dataset, achieving an accuracy of 85.77%, surpassing comparable methodologies in the literature. This improvement was due to transfer learning, data preprocessing, hyperparameter optimization, and regularization techniques, such as dropout and early stopping. The model effectively extracts features from low-resolution facial images, thereby enhancing its robustness and classification performance. The findings validate the efficacy of the VGG-16 architecture for FER tasks and its potential for real-time applications in driver monitoring, healthcare diagnostics, and adaptive human-computer interaction. This technology also shows promise in public transportation and security. Specifically, facial emotion analysis can monitor bus and taxi drivers in real time to detect stress, fatigue, or distraction, which is crucial for road safety. In security-sensitive environments, such as terminal checkpoints, integrating facial emotion recognition with surveillance systems can facilitate early threat detection, enabling quicker security responses. Additionally, monitoring drivers' emotions in commercial transport may enhance customer experience, as a driver's emotional state influences service quality. These applications highlight the broad utility of FER technologies in enhancing operational safety and user satisfaction across sectors.

ACKNOWLEDGEMENTS

I would like to thank my superviso, Prof. Dr. Burhan Ergen, for his guidance, valuable advice, and motivation throughout this research.

REFERENCES

[1] A. R. Khan, "Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges," Information, vol. 13, no. 6, p. 268, 2022.

[2] Nidhi and B. Verma, "From methods to datasets: a detailed study on facial emotion recognition," Applied Intelligence, vol. 53, no. 24, pp. 30219–30249, 2023.

[3] N. Kumari and R. Bhatia, "Saliency map and deep learning based efficient facial emotion recognition technique for facial images," Multimedia Tools Appl., vol. 83, no. 12, pp. 36841–36864, 2024.

[4] J. Zhu, Y. Ding, H. Liu, K. Chen, Z. Lin, and W. Hong, "Emotion knowledge-based fine-grained facial expression recognition," Neurocomputing, vol. 610, p. 128536, 2024.

[5] S. Saurav, R. Saini, and S. Singh, "EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild," Applied Intelligence, vol. 51, pp. 5543–5570, 2021.

[6] F. M. Talaat, Z. H. Ali, R. R. Mostafa, and N. El-Rashidy, "Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children," Soft Computing, vol. 28, pp. 6695–6708, 2024.

[7] I. J. Goodfellow et al., “Challenges in representation learning: A report on three machine learning contests,” in Proc. Int. Conf. Neural Information Processing (ICONIP), Daegu, Korea, 2013, pp. 117–124.

[8] M. C. Gürsesli et al., “Facial emotion recognition (FER) through custom lightweight CNN model: Performance evaluation in public datasets,” IEEE Access, vol. 12, pp. 45543–45553, 2024.

[9] J. Rodríguez-Antigüedad et al., “Facial emotion recognition deficits are associated with hypomimia and related brain correlates in Parkinson’s disease,” J. Neural Transm., pp. 1–7, 2024.

[10] A. Bhattacharyya et al., “A deep learning model for classifying human facial expressions from infrared thermal images,” Scientific Reports, vol. 11, p. 20696, 2021.

11] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, “Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey,” IEEE Trans. Instrum. Meas., vol. 72, p. 5006631, 2023.

[12] B. Bakariya et al., “Facial emotion recognition and music recommendation system using CNN-based deep learning techniques,” Evolving Systems, vol. 15, pp. 641–658, 2024.

[13] X. Fan, M. Jiang, A. R. Shahid, and H. Yan, “Hierarchical scale convolutional neural network for facial expression recognition,” Cognitive Neurodynamics, vol. 16, pp. 847–858, 2022.

[14] A. Khanzada, C. Bai, and F. T. Celepcikay, “Facial expression recognition with deep learning,” arXiv preprint arXiv:2004.11823, 2020.

[15] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition using transfer learning in the deep CNN,” Electronics, vol. 10, no. 9, p. 1036, 2021.

[16] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 5188–5196.

[17] A. Khanzada, C. Bai, and F. T. Celepcikay, “Facial expression recognition with deep learning,” arXiv preprint arXiv:2004.11823, 2020.

[18] G. P. Kusuma, J. Jonathan, and A. P. Lim, “Emotion recognition on FER-2013 face images using fine-tuned VGG-16,” Adv. Sci., Technol. Eng. Syst. J., vol. 5, no. 6, pp. 315–322, 2020.

[19] S. Saurav, R. Saini, and S. Singh, “EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild,” Applied Intelligence, vol. 51, no. 8, pp. 5543–5570, 2021.

[20] Y. Khaireddin and Z. Chen, “Facial emotion recognition: State of the art performance on FER2013,” arXiv preprint arXiv:2105.03588, 2021.