Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

Enhanced Real-Time Facial Expression Recognition Using Deep Learning

Hafiz Burhan Ul Haq[1*], Waseem Akram[2], Muhammad Nauman Irshad[1], Amna Kosar[3], Muhammad Abid[4]

[1] Department of Electronics and Telecommunication Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi (KMUTT), 10140 Bangkok, Thailand
[2] Faculty of Information Technology, University of Central Punjab, 54000 Lahore, Pakistan
[3] Department of Computer Sciences, Faculty of Computer Sciences, Lahore Garrison University, 54000 Lahore, Pakistan
[4] Department of Mathematics, North Carolina State University, 27695 NC Raleigh, USA

[*] Correspondence: Hafiz Burhan Ul Haq (hafiz.burh@kmutt.ac.th)

Received: 10-23-2023 **Revised:** 12-29-2023 **Accepted:** 01-09-2024

Citation: H. B. U. Haq, W. Akram, M. N. Irshad, A. Kosar, and M. Abid, "Enhanced real-time facial expression recognition using deep learning," *Acadlore Trans. Mach. Learn.*, vol. 3, no. 1, pp. 24–35, 2024. https://doi.org/10.56578/ataiml030103.

Abstract: In the realm of facial expression recognition (FER), the identification and classification of seven universal emotional states, surprise, disgust, fear, happiness, neutrality, anger, and contempt, are of paramount importance. This research focuses on the application of convolutional neural networks (CNNs) for the extraction and categorization of these expressions. Over the past decade, CNNs have emerged as a significant area of research in human-computer interaction, surpassing previous methodologies with their superior feature learning capabilities. While current models demonstrate exceptional accuracy in recognizing facial expressions within controlled laboratory datasets, their performance significantly diminishes when applied to real-time, uncontrolled datasets. Challenges such as degraded image quality, occlusions, variable lighting, and alterations in head pose are commonly encountered in images sourced from unstructured environments like the internet. This study aims to enhance the recognition accuracy of FER by employing deep learning techniques to process images captured in real-time, particularly those of lower resolution. The objective is to augment the accuracy of FER in real-world datasets, which are inherently more complex and collected under less controlled conditions, compared to laboratory-collected data. The effectiveness of a deep learning-based approach to emotion detection in photographs is rigorously evaluated in this work. The proposed method is exhaustively compared with manual techniques and other existing approaches to assess its efficacy. This comparison forms the foundation for a subjective evaluation methodology, focusing on validation and end-user satisfaction. The findings conclusively demonstrate the method's proficiency in accurately recognizing emotions in both laboratory and real-world scenarios, thereby underscoring the potential of deep learning in the domain of facial emotion identification.

Keywords: Facial expression recognition; CNNs; MobileNet; Feature extraction

1 Introduction

Over the past decade, the field of artificial intelligence (AI), which aims to emulate human cognitive processes, has undergone significant advancements and encountered intriguing challenges. Among these, the analysis of subtle facial expressions represents a complex task. It is observed that the manifestation of a single emotion can vary considerably across individuals, influenced by factors such as ethnicity, age, or gender. Moreover, the interpretation of an individual's emotional state is subject to contextual variables including lighting, posture, and background. This paper delves into the intricacies of facial expression analysis in the era of AI, exploring the multitude of aspects impacting the accuracy of human emotion detection. Expression, encompassing a broad spectrum of behaviors, actions, thoughts, and feelings, is ultimately a subjective and intimate mental and physical state. The foundational work of Charles Darwin, particularly his book "The Expression of the Emotions in Man and Animals," laid the groundwork for early emotion studies. Subsequent research, notably by Ekman and Friesen in 1969, identified

cross-cultural consistencies in emotional expressions, establishing six universal emotional states: happiness, sadness, anger, contempt, surprise, and fear [1–3].

Conversely, facial expression, a non-verbal form of communication, is crucial to human perception, behavior, and interaction. Facial expressions represent morphological alterations in the face [4], and it is estimated that only about 7% of the information conveyed is through words, with vocal intonation accounting for 55% and body language for 38%. The use and interpretation of body language and facial expressions often occur subconsciously, yet they play a vital role in effective communication. The increasing relevance of emotions in human-robot interaction (HRI) has sparked interest in equipping social robots with FER capabilities. HRI amalgamates disciplines such as social sciences, robotics, AI, and natural language processing [5]. This interdisciplinary approach underlines the growing need to understand and accurately interpret facial expressions, not only in human-to-human interactions but also in the evolving domain of human-robot communication.

Emotions are fundamental in HRI, rendering social robots an increasingly studied subject due to their potential in FER. The exploration of HRI necessitates a multidisciplinary approach, incorporating fields such as AI, robotics, natural language processing, design, and social sciences. Within this scope, facial recognition technologies are crucial yet encounter several limitations including restricted processing capabilities, speed, duration, and accuracy. Challenges in 2D, 3D, and temporal facial recognition methods are prevalent, primarily owing to spatial alterations, occlusions, lighting variances, and the intensive demand for computational resources. Efforts to refine classification accuracy have been observed, with some researchers opting to simplify methodologies by minimizing feature points or adopting a more objective approach. In the realm of computer vision, traditional machine learning techniques previously demonstrated efficacy but were hindered by their inability to process direct photo inputs. Contemporary face recognition systems continue to face challenges due to varying lighting conditions, backgrounds, and postures, which can significantly alter appearances and obstruct precise expression detection. The advent of deep learning has been pivotal in addressing these challenges, enhancing the recognition performance of the six core emotional expressions—sadness, disgust, anger, happiness, fear, and surprise. However, the application of deep learning models to faces captured under divergent conditions from the training dataset remains a significant limitation. A comparative analysis of traditional and deep learning techniques in facial recognition is presented in Figure 1.

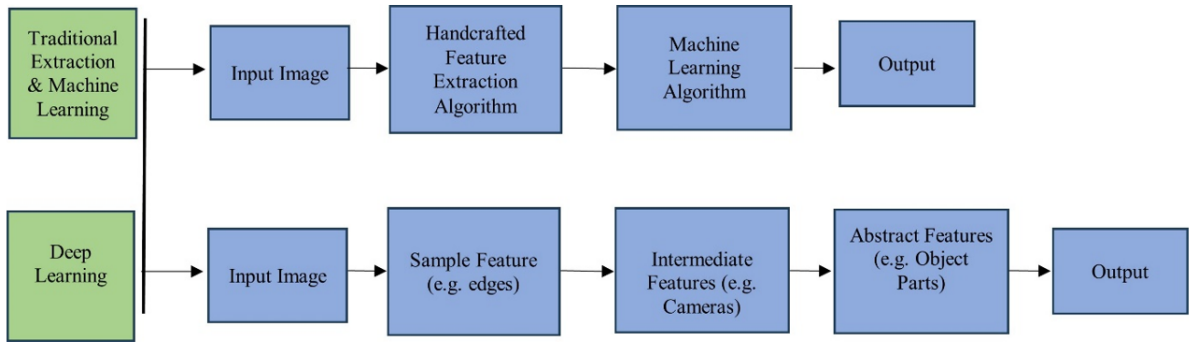

Figure 1. Comparative analysis between traditional machine learning and the deep learning model

In this study, a novel deep learning-based framework is introduced, designed to surmount the challenges inherent in real-time facial emotion recognition. The system employs deep learning algorithms for detection, coupled with CNNs for the extraction of features, thereby recognizing a spectrum of seven emotional states: happiness, sadness, anger, fear, surprise, disgust, and neutrality. The methodology incorporates current techniques while introducing several key enhancements:

• Expanded recognition capability: The model is engineered to differentiate between seven emotional categories, thereby broadening its scope to capture a more extensive range of human emotions. This expansion enables a more precise and nuanced analysis of facial expressions.

• Streamlined and resilient architecture: The system is designed with simplicity and robustness, facilitating effective real-time processing. This feature ensures the model's applicability in real-world scenarios without excessively taxing computational resources.

• Enhanced accuracy: By leveraging advanced deep learning techniques, the model achieves elevated levels of accuracy in facial emotion detection. This improvement is critical for reliable outcomes, particularly in fields such as human-computer interaction, market research, and mental health assessments.

• Rigorous evaluation and validation: The efficacy of the proposed model will be rigorously assessed using predefined datasets. This evaluation process is aimed at empirically demonstrating the system's proficiency and its capacity to yield valuable insights across various applications. The methodology outlined in this research encompasses

several critical steps, each contributing to the development of an advanced real-time facial expression detection system:

• Data collection: Initially, a comprehensive and varied dataset of facial expressions is compiled. This dataset is meticulously curated to ensure diversity and representativeness, laying a foundational basis for subsequent model training.

• Model training and evaluation: Deep learning models are then rigorously trained on the assembled datasets. The focus of this training is to enhance the models' proficiency in accurately identifying the seven predefined emotional categories. Subsequent extensive testing is conducted to refine and validate the models' performance.

• Application in real-time detection: Designed for practical, real-time scenarios, the system operates by selecting an image from the collected dataset as input. It then rapidly processes this image for emotion recognition, aiming to deliver prompt and reliable results.

In summation, the approach proposed in this study marks a significant advancement in real-time FER. It integrates innovative features, including simplicity, robustness, and high accuracy, making it a valuable asset in diverse applications where understanding human emotions is essential. These applications range from enhancing human-computer interaction to providing insights in market research and mental health assessments.

2 Literature Review

The implementation of facial recognition technology in smart devices has become increasingly prevalent, yet it imposes significant demands on storage and processing capabilities. In response to these challenges, a range of strategies and systems for expression recognition have been developed and are briefly reviewed herein:

Guo et al. [6] introduced an innovative approach utilizing DNNs with relativity learning (DNNRL). This method aims to contract the distances in the embedding space between samples representing the same expression, while concurrently expanding the gap between those of differing expressions. The training process involves the selection of an anchor, a positive sample (bearing the same expression as the anchor), and a negative sample (exhibiting a different expression). The core objective is to minimize the triplet loss, which effectively reduces the distance between the anchor and the positive instance in the embedding space, ensuring that it remains narrower than that between the anchor and the negative sample. DNNRL notably assigns greater weight to challenging instances based on the network's output, allowing for more nuanced learning. The efficacy of DNNRL has been validated using the SFEW and FER2013 datasets.

Feature extraction, a critical step following face detection in FER, is heavily dependent on the quality of the features extracted. Subtle or pronounced deformations in facial features such as eyebrows, lips, eyes, and nose can induce changes in facial expressions. Feature extraction methods are categorized into two types: non-geometric and geometric-based features [7]. Geometric feature extraction focuses on quantifying the size and position of facial features, including the nose, lips, forehead, chin, and eyes. These attributes are encapsulated within a facial geometry feature vector. Geometric feature extraction employs various geometric interactions, such as points, stretches, and angles, between these components to encode the features. In contrast, appearance-based feature extraction employs either a single image filter or a combination of filters applied to the entire image or specific regions to discern changes in texture and shape [8]. Furthermore, a range of computational models and methods for processing visual data are employed in feature extraction. These include tools like fuzzy logic and neural networks. Feature extraction strategies are broadly classified into four types: feature-based, appearance-based, template-based, and part-based approaches [9].

Li et al. [10] have explored the application of the k-nearest neighbor (KNN) strategy, augmented by center loss and locality-preserving loss (LP-loss), for clustering deep features and ensuring intra-class compactness. The employed deep locality-preserving CNN (DLP-CNN) maintains the local representation of each sample in the embedding space. During training, Euclidean distance is utilized to ascertain the KNN for each data point, aiming to minimize the sample's distance from the mean of its KNNs. The effectiveness of LP-loss has been evaluated using datasets such as CK+, SFEW, MMI, and RAF-DB. Center loss, while promoting intra-class compactness and consequently aiding in inter-class separation, may still permit overlap among feature regions in the embedding space. Building on this concept, Cai et al. [11] enhanced center loss by integrating an additional objective function. This modified center loss, termed as island loss, merges the original center loss with the pairwise cosine distance between class centers in the embedding space. The approach aims to increase cosine distance, thereby angularly separating the class centers. Island loss has been assessed using datasets including CK+, MMI, and Oulu-CASI. Recent advancements in facial emotion recognition have been significantly influenced by deep learning algorithms. Jain et al. [12] introduced single deep neural network (DNN) incorporating convolution layers and deep residual blocks. Lopes et al. [13], in a similar vein, presented a multiple CNN framework, complemented by a specialized image pre-processing stage for emotion recognition.

The application of FER in dynamic environments was addressed by Jain et al. [14] through the deployment of a hybrid convolution-recurrent neural network technique. A comparative analysis was conducted by Sajjanhar et

al. [15] on the performance of pre-trained facial recognition algorithms, Visual Geometry Group (VGG)-facial and Inception, both initially developed for object detection. Wen et al. [16] employed a convolutional rectified linear layer as the initial layer in their CNN aggregate for facial emotion recognition, incorporating multiple hidden maxout layers to modify the architecture of each CNN. Despite notable advancements in the field of FER, research predominantly focuses on devising strategies to enhance outcomes presented in one or more datasets independently. The investigation into the impact of cross-dataset fine-tuning on performance was conducted by Zavarez et al. [17]. For this purpose, the VGG face deep CNN model was adapted for facial emotion recognition. Cross-dataset experiments were meticulously designed, utilizing one dataset as the test set while employing others for training, to ensure the reliability of the results. Wang et al. [18] proposed an innovative approach integrating FER technology with online course platforms. In this method, student facial expressions were captured using device cameras during an online course and processed through a FER algorithm (CNN model), categorizing them into eight emotional states: anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral. This approach was tested in an online course using Tencent Meeting with 27 students, demonstrating consistent performance across diverse scenarios. The applicability of this concept extends beyond online educational settings, suggesting potential in various interactive environments.

Pise et al. [19] have applied contemporary deep learning models to the evolving field of automated emotion recognition within computational intelligence. This research demonstrates the integration of deep learning-based FER with architectural methods and databases, yielding highly accurate results. A diverse array of machine learning and deep learning methodologies are employed in this investigation. Saeed et al. [20] discussed a technique to enhance accuracy in facial recognition. Their proposed CNN method (fall detection-CNN), incorporating two hidden layers and four convolutional layers, serves as an automated framework. Utilizing the expanded Cohn-Kanade (CK+) dataset, which includes images portraying a range of emotions from various individuals, the process encompasses pre-processing, feature extraction, and categorization. The model's effectiveness is evaluated through metrics such as F1-score, recall, and precision, with respective values of 84.07%, 78.22%, and 94.09%. Additionally, numerous studies employing machine learning methods have contributed to this field [21–27]. Despite advancements, certain facial recognition approaches encounter challenges, including poor lighting, shadows, partial facial visibility, camera orientation issues, and lower recognition rates. This project aims to develop a CNN-based FER system enhanced with data augmentation. The proposed system is designed to classify the seven principal emotions, anger, contempt, fear, happiness, neutrality, sadness, and surprise, from visual data.

3 Proposed Methodology

Figure 2 delineates the architecture of the proposed emotion recognition model. The methodology comprises the following principal components:

(a) Data collection: This phase involves the accumulation of a diverse dataset, encompassing images that represent a range of emotions.

(b) Data preprocessing: The dataset undergoes classification, categorizing images into seven emotional states: anger, happiness, fear, disgust, neutrality, sadness, and surprise.

(c) Emotion prediction: Utilizing a deep learning model, emotion predictions are executed on the images.

(d) Performance evaluation: The final stage involves assessing the model's performance in accurately predicting emotions.

3.1 Data Collection

The data collection process was facilitated by a data acquisition layer, responsible for aggregating data from various online sources. This research utilized information gathered from links, data repositories, and additional internet resources. Figure 3 presents a selection of the data samples amassed for this study. The methodology incorporated two primary datasets: the FER-2013 [28] and a Random dataset [29]. The FER-2013 dataset comprises grayscale images, each measuring 48×48 pixels. It encompasses a training set of 28,000 labeled images, a development set consisting of 3,500 labeled images, and a test set with another 3,500 labeled images. This dataset encapsulates seven emotional states: happiness, sadness, anger, fear, surprise, disgust, and neutral. In contrast, the Random dataset includes a compilation of 350 images, both in color and grayscale, further categorized into six emotional categories: happiness, sadness, anger, fear, surprise, disgust, and neutral. Figure 3 showcases representative images from both datasets, illustrating the diversity and range of emotions covered.

3.2 Proposed Model

In this phase of the research, the focus is on the utilization of deep learning models, specifically MobileNet, for real-time prediction of seven emotional categories: happiness, sadness, anger, fear, surprise, disgust, and neutrality. The MobileNet architecture is leveraged due to its efficiency in processing and reduced parameter count compared to conventional convolutional networks. Bounding boxes are employed to highlight the facial regions where emotions are detected. MobileNet, a variant of CNNs developed by Google, employs depth-separable convolutions, significantly

reducing the number of parameters required. This reduction enables the deployment of DNNs on portable devices, making MobileNet an ideal foundation for compact and rapid classifiers. The architecture of MobileNet comprises several depth-separable convolutional layers, each consisting of a depth-wise convolution followed by a point-wise convolution. In total, a MobileNet architecture contains 28 layers when depth-wise and point-wise convolutions are considered separately. Furthermore, the adaptability of MobileNet is enhanced by the width multiplier hyperparameter, which allows for the adjustment of the network's complexity. Typically, a standard MobileNet comprises approximately 4.2 million parameters, with input dimensions of 224×224×3. Figure 4 presents the architectural diagram of MobileNet, highlighting its structural components [30, 31].

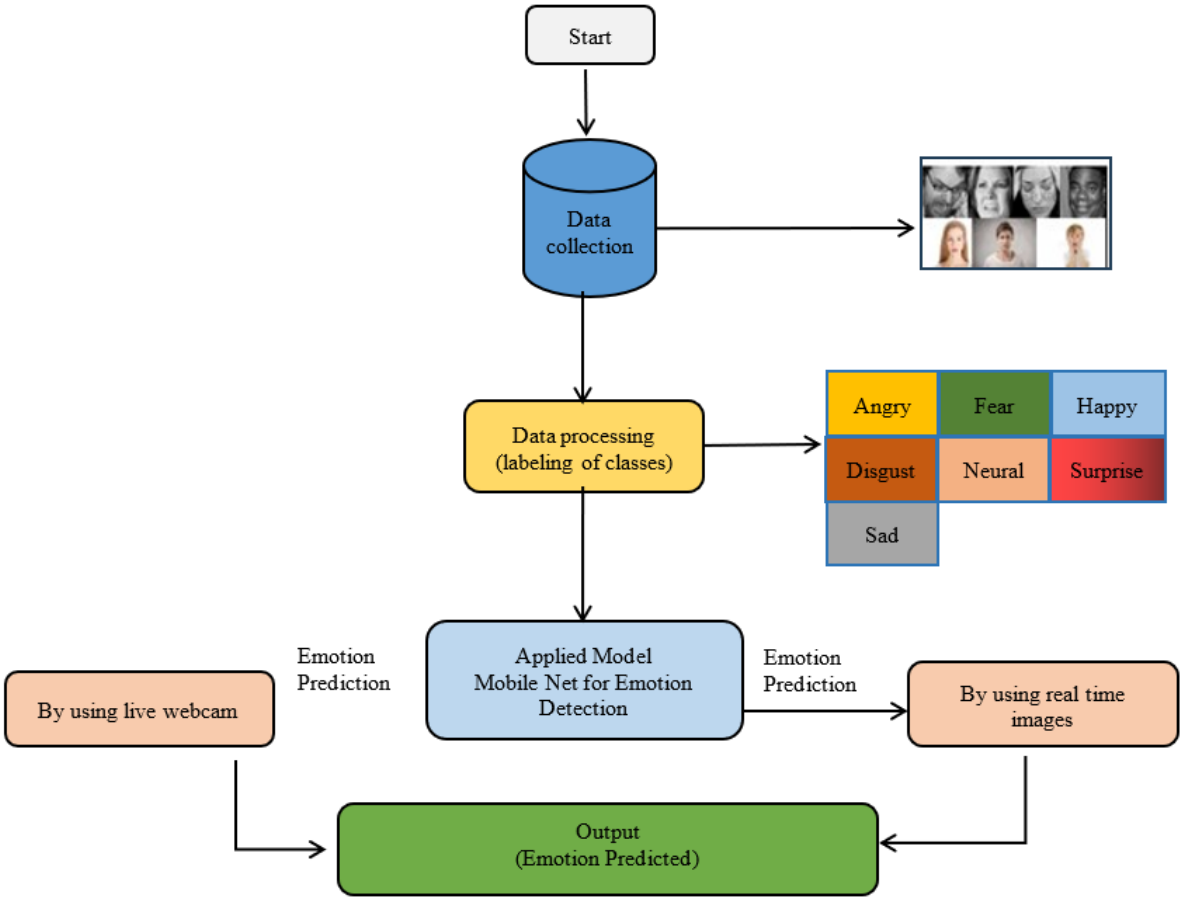

Figure 2. Architectural diagram of the proposed model

Figure 3. Sample images from the collected datasets

3.3 Speed Comparison Between MobileNet and Other Models

In the evaluation of object detection models, MobileNet is distinguished by its exceptional speed performance. Contrasting with its counterparts, which typically operate at a frame rate of 5 frames per second, MobileNet excels by achieving a remarkable 22 frames per second. This rapid processing capability significantly elevates MobileNet above

other models in terms of efficiency. To illustrate this, consider the comparison with models such as regions with CNN features (R-CNN) and its enhanced version, Fast R-CNN. While these models exhibit higher accuracy rates, capturing more detailed information than MobileNet, they lag in processing speed. The defining advantage of MobileNet lies in its speed, making it a preferred option for applications where prompt and efficient object detection is crucial. This aspect is particularly vital in real-world scenarios where time-sensitive detection is paramount [32]. Figure 5 provides a comparative analysis of MobileNet against various object detection techniques, emphasizing the speed differential.

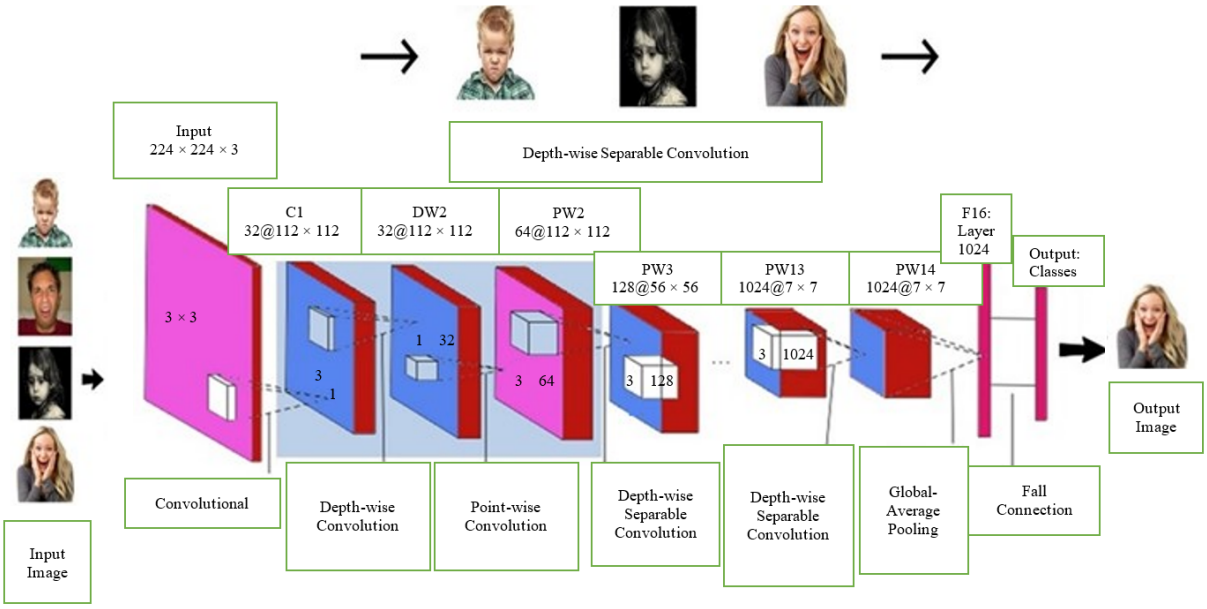

Figure 4. MobileNet architecture [31]

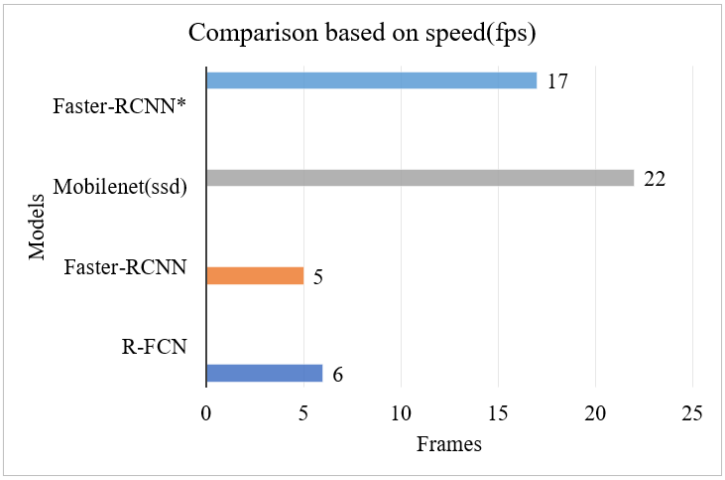

Figure 5. Speed comparison of deep learning models

3.4 Experimental Analysis and Results

The experimental evaluation was conducted on a computer equipped with an Intel Core i5-6200U CPU, operating at 2.4 GHz and supported by 8 gigabytes of RAM. Python, chosen for its versatility and efficiency, served as the programming language for the implementation of the models. To assess the accuracy of the developed models, a comprehensive evaluation was performed using a test dataset with well-established target features. The model outputs were systematically compared against these known ground truths, facilitating a detailed analysis of their performance. A key instrument in this evaluation was the utilization of a confusion matrix. This matrix provided both a visual and numerical representation of the model's performance, indicating not only the predicted instances for each class but also the accuracy of these predictions. Moreover, various assessment parameters were calculated using specific

mathematical formulae, further elucidating the models' effectiveness. These calculations and their corresponding formulae are detailed in Eq. (1), which offers a comprehensive view of the analytical methods employed in this study.

$$\begin{aligned} \text{Accuracy} &= (TP + TN)/(TP + TN + FP + FN) \\ \text{Precision} &= \text{TP}/(\text{TP} + \text{FP}) \\ \text{Recall} &= \text{TP}/(\text{TP} + \text{FN}) \end{aligned} \tag{1}$$

In this research, the model's performance was evaluated using a subjective assessment approach, wherein manually created images and frames depicting emotions were compared individually. The datasets utilized for testing and training encompassed the following emotional classes: happiness, sadness, anger, fear, surprise, disgust, and neutrality. Table 1 provides an extensive description of the dataset.

Table 1. Dataset composition for testing/training

Classes	FER-2013 Dataset	Random Dataset
Happiness	879	50
Sadness	594	50
Anger	491	50
Fear	528	50
Surprise	416	50
Disgust	55	50
Neutrality	626	50

For the experimental analysis, images were sourced from online platforms, system directories, and those specifically collected for this study. While some images from the training data were used preliminarily to check for duplicates, the primary focus was on images not included in the training set. The proposed model's efficacy was tested across a range of image resolutions. A unique aspect of the assessment involved contrasting the proposed model with a manual approach, where an individual subjectively classified images into respective emotional categories. This method entailed manual predictions which, despite initial accuracy, exhibited uncertainty in certain cases due to behavioral similarities, such as mistaking an image of a newborn for fear when it might also be interpreted as surprise. Subsequently, these images were processed through the proposed model, and the outcomes from both methods were compared. This process constituted an image-level comparison. Figure 6 illustrates this comparison, showcasing how each method categorized images across the seven emotional classes: happiness, sadness, anger, fear, surprise, disgust, and neutrality.

As depicted in Figure 7, it was observed that the proposed model did not accurately identify certain emotional states. This limitation was primarily due to the visual similarities between different emotions. For instance, an image that predominantly exhibited characteristics of fear was erroneously classified under the category of surprise by the model. Similarly, another image, which ideally belonged to the disgust category, was incorrectly identified as sadness. This misclassification stemmed from the visual resemblance of the image to those typically associated with sadness, as perceived by the unaided eye. Figure 7 presents a selection of instances where the model's predictions were impeded by such behavioral similarities. These examples highlight the challenges faced in accurately distinguishing between emotions that share common visual traits.

3.4.1 Results

The results of the proposed model, as depicted in Figure 8, demonstrated a remarkable accuracy of 100% during validation and 97.9% during training. These statistics indicate the model's successful generalization from the training dataset to the validation dataset. However, challenges arose when the model was applied to real-world images sourced from various platforms such as online resources, system directories, and captured photographs. These images encompassed a spectrum of emotional states: happiness, sadness, anger, fear, surprise, disgust, and neutrality.

The manual assessment method, which relied on human judgment to classify images, also faced difficulties in accurately identifying emotions, especially in instances where images exhibited similar emotional traits. For example, an image of a baby, which might appear fearful, could also be interpreted as showing surprise due to the ambiguity in facial expressions. This issue underscores the subjective nature of human visual perception in emotion recognition tasks. Discrepancies were noted between the model's predictions and manual assessments. The model, due to its focus on behavioral similarities, misclassified certain images into incorrect emotional categories. An image that visually suggested fear was sometimes predicted as surprise, while an image that initially appeared sad was classified as disgust. These misclassifications were attributed to the model's challenge in discerning subtle differences in emotional expressions. Figure 8 illustrates instances where the model struggled with reliable predictions due to the proximity

of the emotional expressions depicted in the images. Despite its high accuracy in training and validation settings, the model's performance in real-world scenarios highlighted the intricacies of emotion recognition in photographs, particularly when dealing with minute variations and nuances.

(a)

(b)

Figure 6. Comparative outputs (a) Manual method; (b) Proposed model

In summary, while the model exhibited commendable performance during training and validation phases, it encountered difficulties in accurately discerning emotions in real-world images. The findings emphasize the significance of acknowledging the inherent challenges and limitations in emotion detection tasks, particularly with

images exhibiting a range of similar emotional expressions. Further research and refinement may be necessary to enhance the model's capability in such scenarios.

Figure 7. Speed comparison of deep learning models

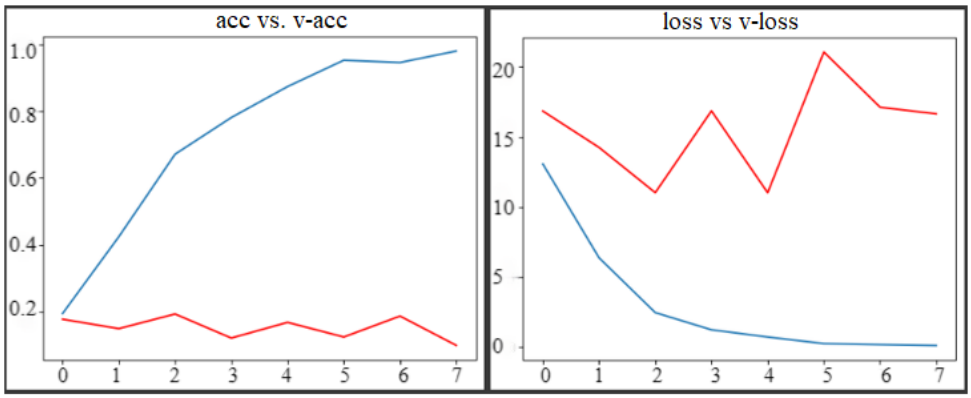

Figure 8. Accuracy curve and validation loss

4 Results

Table 2 presents a comparative analysis between the proposed MobileNet-V1 model and existing techniques in emotion recognition. The comparison, based on accuracy, is drawn from a range of studies and models:

Table 2. Comparative analysis with other techniques

Authors	Model	Accuracy
Barsoum et al. [33]	VGG13(MV)	83.86%
Li et al. [34]	TFE-JL	84.29%
Georgescu et al. [35]	CNNs and BOVM + global SVM	87.76%
Huang [36]	ResNet + VGG	87.4%
Wang et al. [18]	SCN + ResNet18	88.01%
Nan et al. [37]	A-MobileNet	88.11%
Proposed model	MobileNet-V1	**97.9%**

The proposed MobileNet-V1 model demonstrates a significantly higher accuracy rate of 97.9%, surpassing the accuracy levels of other models cited in the comparative study. This superior accuracy rate is indicative of the model's effectiveness in emotion recognition tasks. The comparative analysis underscores the compactness and reliability of the proposed model, especially in contrast to the existing methodologies.

5 Conclusions

The method presented in this article for real-time emotion recognition incorporates user behavior analysis and is capable of differentiating between seven behavioral categories: happiness, sadness, neutrality, disgust, fear, surprise, and anger. Existing methods in the literature for content extraction and behavior recognition, while useful, are often hindered by high hardware demands and slow processing speeds. A detailed comparison of emotion recognition techniques is provided, demonstrating the simplicity, optimization, precision, and reliability of the proposed model relative to current methods. A key innovation in this study is the accurate and efficient extraction of seven distinct behaviors. The primary objective of developing an emotion detection system based on seven classifications using images or shots has been successfully achieved with the proposed approach. For assessment purposes, the experimental study utilized two datasets, FER2013 and Random datasets, both comprising images categorized into seven emotional states. In comparison to the subjective evaluation method, where an observer manually identifies the behavior from an image, the proposed model demonstrated superior performance. Extensive experiments have shown that the proposed method achieved an accuracy of 97.7% while maintaining high processing speed and efficiency. Future research directions include exploring additional classes, enhancing accuracy, and implementing real-time facial emotion recognition using cameras. A user-friendly interface is also planned for integration into an application utilizing the developed model.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] M. Grimm, D. Dastidar, and K. Kroschel, "Recognizing emotions in spontaneous facial expressions," in *Proceedings of the International Conference on Intelligent Systems and Computing, Cyprus*, 2006.

[2] P. Ekman, "Facial expression and emotion," *Am. Psychol.*, vol. 48, pp. 384–392, 1993. https://doi.org/10.1037/0003-066X.48.4.384

[3] C. Busso, Z. G. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, USA*, 2004, pp. 205–211. https://doi.org/10.1145/1027933.1027968

[4] S. Goodfellow and S. Nowicki Jr, "Social adjustment, academic adjustment, and the ability to identify emotion in facial expressions of 7-year-old children," *J. Genet. Psychol.*, vol. 170, no. 3, pp. 234–243, 2009. https://doi.org/10.1080/00221320903218281

[5] F. Furrer, M. Burri, M. Achtelik, and R. Siegwart, "RotorS—A modular gazebo MAV simulator framework," in *ROS: Studies in Computational Intelligence*. Cham: Springer, 2016. https://doi.org/10.1007/978-3-319-26054-9_23

[6] Y. Guo, D. P. Tao, J. Yu, H. Xiong, Y. T. Li, and D. C. Tao, "Deep neural networks with relativity learning for facial expression recognition," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA*, 2016, pp. 1–6. https://doi.org/10.1109/ICMEW.2016.7574736

[7] M. Sharma, J. Anuradha, H. K. Manne, and G. S. C. Kashyap, "Facial detection using deep learning," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 263, p. 042092, 2017. https://doi.org/10.1088/1757-899X/263/4/042092

[8] A. J. Shepley, "Deep learning for face recognition: A critical analysis," *arXiv preprint arXiv:1907.12739*, 2019. https://doi.org/10.48550/arXiv.1907.12739

[9] A. K. Sharma, U. Kumar, S. K. Gupta, U. Sharma, and S. L. Agrwal, "A survey on feature extraction technique for facial expression recognition system," in *2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India*, 2018, pp. 1–6. https://doi.org/10.1109/CCAA.2018.8777550

[10] S. C. Li, W. H. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, 2017, pp. 2584–2593. https://doi.org/10.1109/CVPR.2017.277

[11] J. Cai, Z. Meng, A. S. Khan, Z. Y. Li, J. O'Reilly, and Y. Tong, "Island loss for learning discriminative features in facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China*, 2018, pp. 302–309. https://doi.org/10.1109/FG.2018.00051

[12] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended DNN for facial emotion recognition," *Pattern Recognit. Lett.*, vol. 120, pp. 69–74, 2019. https://doi.org/10.1016/j.patrec.2019.01.008

[13] A. T. Lopes, E. De Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, 2017. https://doi.org/10.1016/j.patcog.2016.07.026

[14] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid DNNs for face emotion recognition," *Pattern Recognit. Lett.*, vol. 115, pp. 101–106, 2018. https://doi.org/10.1016/j.patrec.2018.04.010

[15] A. Sajjanhar, Z. Y. Wu, and Q. Wen, "Deep learning models for facial expression recognition," in *2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, ACT, Australia*, 2018, pp. 1–6. https://doi.org/10.1109/DICTA.2018.8615843

[16] G. H. Wen, Z. Hou, H. H. Li, D. Y. Li, L. J. Jiang, and E. Xun, "Ensemble of DNNs with probability-based fusion for facial expression recognition," *Cognit. Comput.*, vol. 9, pp. 597–610, 2017. https://doi.org/10.1007/s12559-017-9472-6

[17] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos, "Cross-database facial expression recognition based on fine-tuned deep convolutional network," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, Brazil*, 2017, pp. 405–412. https://doi.org/10.1109/SIBGRAPI.2017.60

[18] K. P. Wang, X. H. Peng, J. Yang, S. J. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA*, 2020, pp. 6897–6906. https://doi.org/10.1109/CVPR42600.2020.00693

[19] A. A. Pise, M. A. Alqahtani, P. Verma, D. A. Karras, and A. Halifa, "Methods for facial expression recognition with applications in challenging situations," *Comput. Intell. Neurosci.*, vol. 2022, 2022. https://doi.org/10.1155/2022/9261438

[20] S. Saeed, A. A. Shah, M. K. Ehsan, M. R. Amirzada, A. Mahmood, and T. Mezgebo, "Automated facial expression recognition framework using deep learning," *J. Healthc. Eng.*, vol. 2022, 2022. https://doi.org/10.1155/2022/5707930

[21] H. B. U. Haq and M. Saqlain, "Iris detection for attendance monitoring in educational institutes amidst a pandemic: A machine learning approach," *J. Ind. Intell.*, vol. 1, no. 3, pp. 136–147, 2023. https://doi.org/10.56578/jii010301

[22] D. Baig, W. Akram, H. B. U. Haq, and M. Asif, "Cloud gaming approach to learn programming concepts," *Artif. Intell. Appl.*, 2022. https://doi.org/10.47852/bonviewAIA32021378

[23] S. Nawaz, A. N. Akhtar, and H. B. U. Haq, "Cloud computing services and security challenges: A review," *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 2, pp. 17–28, 2023. https://doi.org/10.54692/lgurjcsit.2023.0702459

[24] H. B. U. Haq and M. Saqlain, "An implementation of effective machine learning approaches to perform sybil attack detection (SAD) in IoT network," *Theor. Appl. Comput. Intell.*, vol. 1, no. 1, pp. 1–14, 2023. https://doi.org/10.31181/taci1120232

[25] M. Saqlain, "Sustainable hydrogen production: A decision-making approach using VIKOR and intuitionistic hypersoft sets," *J. Intell. Manag. Deci.*, vol. 2, no. 3, pp. 130–138, 2023. https://doi.org/10.56578/jimd020303

[26] M. Abid and M. Saqlain, "Decision-making for the bakery product transportation using linear programming," *Spectr. Eng. Manag. Sci.*, vol. 1, no. 1, pp. 1–12, 2023. https://doi.org/10.31181/sems1120235a

[27] M. N. Jafar, K. Muniba, and M. Saqlain, "Enhancing diabetes diagnosis through an intuitionistic fuzzy soft matrices-based algorithm," *Spectr. Eng. Manag. Sci.*, vol. 1, no. 1, pp. 73–82, 2023. http://doi.org/10.31181/sems1120238u

[28] "Fer-2013," Kaggle, 2013. https://www.kaggle.com/datasets/msambare/fer2013

[29] "Dataset," DropBox. https://www.dropbox.com/s/w3zlhing4dkgeyb/train.zip?dl=0

[30] "MobileNet V1 Architecture," OpenGenus IQ. https://iq.opengenus.org/mobilenet-v1-architecture

[31] A. Pujara, "Image classification with MobileNet," 2020. https://medium.com/analytics-vidhya/image-classification-using-mobilenet-in-the-browser-b69f2f57abf

[32] J. Hui, "Object detection: Speed and accuracy comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3)," 2017. https://medium.com/@jonathan_hui/object-detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359

[33] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan*, 2016, pp. 279–283. https://doi.org/10.1145/2993148.2993165

[34] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial expression recognition with identity and emotion joint learning," *IEEE Trans. Affect. Comput.*, vol. 12, pp. 544–550, 2018. https://doi.org/10.1109/TAFFC.2018.2880201

[35] M. I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64 827–64 836, 2019. https://doi.org/10.48550/arXiv.1804.10892

[36] C. Huang, "Combining CNNs for emotion recognition," in *2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA*, 2017, pp. 1–4. https://doi.org/10.1109/URTC.2017.8284175

[37] Y. H. Nan, J. G. Y. Ju, Q. Hua, H. M. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," *Alex. Eng. J.*, vol. 61, no. 6, pp. 4435–4444, 2022. https://doi.org/10.1016/j.aej.2021.09.066