Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 2, November 2025, pp. 1001~1010
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i2.pp1001-1010

Panic detection through facial recognition paradigm using deep learning tools

Sameerah Faris Khlebus[1], Mohammed Salih Mahdi[2], Monji Kherallah[3], Ali Douik[4]
[1]Department of Computer Science, National School of Electronics and Communications, University of Sfax, Sfax, Tunisia
[2]Business Information College, University of Information Technology and Communication, Baghdad, Iraq
[3]Faculty of Science, University of Sfax, Sfax, Tunisia
[4]Department of Industrial Computing, National Engineering School, University of Sousse, Sousse, Tunisia

Article Info

Article history:

Received Jan 22, 2025
Revised Aug 1, 2025
Accepted Oct 15, 2025

Keywords:

Annotation
Convolutional neural network
FER2013
MobileNet
Panic
ResNet

ABSTRACT

Recently, panic detection has become essential in security, healthcare, and human-computer interaction. Automatic panic detection (APD) systems are designed to monitor physiological signals and behavioral patterns in real-time to detect stress responses. APD is increasingly adopted across many sectors, including disaster preparedness, COVID-19, and terror attacks. Their integration with various applications reduces human efforts and saves costs. However, most studies rely on existing models with fewer new ones or techniques. This study proposes a vision-based panic detection model using MobileNet, ResNet, and convolutional neural network (CNN). The FER2013 dataset is used for the model training and testing. The results indicate that MobileNet is the most effective model for image-based panic detection across ten folds with an accuracy of 90%, recall of 96.9%, and mean accuracy of 0.032. MobileNet also showed a mean absolute error (MAE) between 0.02 and 0.04. This study has been to confirm MobileNet's suitability for image-based panic detection. The findings contribute to developing more reliable and accurate image-based panic detection systems in real-world applications. It offers valuable insights and lays the groundwork for future deep-leaning-based panic detection studies.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Sameerah Faris Khlebus
Department of Computer Science, National School of Electronics and Communications
University of Sfax
Sfax, Tunisia
Email: sameerah.alradhi@uoitc.edu.iq

1. INTRODUCTION

In an increasingly complex world, emotional state detection has gained significant attention in security, healthcare, and human-computer interaction. Automatic panic detection (APD) is increasingly adopted across many sectors [1]. ADP systems are designed to monitor physiological signals and behavioral patterns in real-time to detect stress responses [2]. They leverage artificial intelligence and machine learning to identify panic or anxieties in individuals [3], [4]. ADP has become critical to disaster preparedness. Additionally, the models have become essential in recent years due to crises such as COVID or terror attacks. They are integrated into social media, surveillance systems, and various applications to reduce human effort and save the technology adaptation cost [5].

Image-based sentiment detection has become crucial considering the shift towards more visual expressions in online communication. In emotional recognition systems, panic detection enhances timely intervention in critical situations like responding to human emotions [6]. It is integral to many applications,

such as surveillance and security. Facial expression recognition 2013 [7] dataset is widely used in computer vision and machine learning for tasks related to facial expression recognition [8]. Researchers at the University of Toronto developed it, comprising 35,887 grayscale images measuring 48x48 pixels [9]. The images are categorized into seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral [8]. The dataset features posed and spontaneous expressions from diverse public platforms like internet images, movies, and TV series. The 28,709 images are split into public and private test sets, each comprising 3,589 images [9]. Moreover, FER2013 [7] distributes images across emotion categories, enabling robust training and evaluation of facial expression recognition models.

Despite its applicability, FER2013 has some limitations. The images have low resolution and grayscale, which limits their ability to capture the full complexity of real-world facial expressions [7]. Hence, researchers integrated FER2013 with machine learning algorithms and deep learning architectures to recognize facial expressions in images accurately. Hence, this study applied MobileNet, ResNet, and convolutional neural network (CNN) to a vision-based panic detection model for facial recognition. It explores how pictures and visuals shared on social media platforms contribute to the overall sentiment. These findings aim to uncover new perspectives, address challenges, and enhance understanding of sentiments within panic detection systems.

2. EXISTING LITERATURE

Emotion recognition and crowd behavior analysis are essential research areas due to their wide applications, including safety systems, public surveillance, smart homes, and healthcare. Spatio-temporal feature extraction for video emotion recognition applies five datasets to improve the process [10]. It integrates body gestures and facial expressions for emotion recognition. Although this method is adaptable, its limited datasets affect its duplicability. Driver facial expression recognition (DFER) techniques are a hybrid method that monitors drivers' emotions in real-time [11]. DFER utilizes VGGNET and optical flow reconstruction to address occlusions and lighting changes.

Furthermore, the technology uses extended Cohn Kanade (CK+) and KMU facial expression database (KMU-FED) datasets to improve driver safety [12]. It can detect fatigue, distraction, and aggressive or emotional driving. However, the system is limited by complexity and computational resources. Additionally, it has data privacy and security concerns and does not account for emotions represented in different cultures.

Driver panic detection using EEG signals simulates a driving environment to detect the various emotional states of a driver, which may affect their driving. Despite achieving a binary classification accuracy of 91.5%, it is yet to be tested in real-world situations. DL-based object detection in crowd analysis is a CNN-based method for crowd analysis. It is useful in crowded environments. Meanwhile, multimodal emotion recognition (MER) systems use deep learning architectures and information fusion techniques to analyze multiple input data or modalities [13]. It examines facial expressions, eye movements, physiological signals, posture, and gestures to understand a person’s emotional state [14]. The system uses multiple data sources to increase its accuracy (over 91%); however, it has yet to be tested with real-world datasets.

Similarly, crowd behavior analysis focuses on anomalies and analyses crowd behavior using deep learning-based anomaly detection [15]. On the other hand, real-time facial emotion-based security for smart homes is used for detection in smart home security systems [16]. Facial expression recognition with big data technology uses integral graph methods, weak classifiers, and dynamic sequence models [17], [18]. Nevertheless, the system is limited to smart homes.

The literature shows advancements in emotion recognition using deep learning techniques. Table 1 (in Appendix) demonstrates different techniques applied in emotion and crown recognition. Many studies review existing methods rather than introduce new models or techniques [19]-[23]. Therefore, multimodal data integration is needed to improve emotion recognition in real-world environments and determine the most appropriate model.

3. METHOD

The training, testing, and validation of datasets are done to ensure the proper format of the images is used in the dataset for model prediction. Three proposed models, MobileNet, ResNet, and CNN, were used in this study to determine the results using 10-fold cross-validation. The FER2013 dataset is used for the model training and testing.

The three models' accuracy, mean accuracy, recall, and average mean absolute error (MAE) are assessed. Accuracy measures the overall percentage of correct predictions. A higher accuracy indicates that the model effectively identifies instances of panic and non-panic states. Recall measures the ability of the

model to identify true positive cases (actual instances of panic) out of all actual positive instances. High recall is significant in panic detection, as missing a panic instance could have critical consequences. MAE measures the average magnitude of the errors between the model's predictions and the actual values without considering their direction (whether the error is positive or negative). In this study, MAE provides insight into how far off the model's predictions are from the true values, on average. A lower MAE indicates better model performance, indicating that its predictions are closer to the actual values.

3.1. MobileNet

MobileNet is CNN designed for mobile and embedded vision applications [24]. The algorithm can effectively be adapted for emotional face recognition tasks based on several processes [25]. The dataset for facial images has various emotional expressions for diversity and is pre-processed [26]. The images are resized to fit the MobileNet model dimensions. The MobileNet model is pre-trained on a large-scale dataset like ImageNet. The parameters of the early layers of the model are locked to prevent changes during training. Furthermore, the output is adjusted to house the emotions like happiness, sadness, anger, surprise, fear, and disgust. When training the MobileNet model, mini-batch gradient descent and learning rate schedules optimize performance and prevent overfitting. The model's performance and effectiveness are evaluated using accuracy, precision, recall, and F1-score. Moreover, the model's hyperparameters and architecture are fine-tuned based on evaluation results and experiments with data augmentation techniques to increase robustness. The steps involved in MobileNet are highlighted in Figure 1.

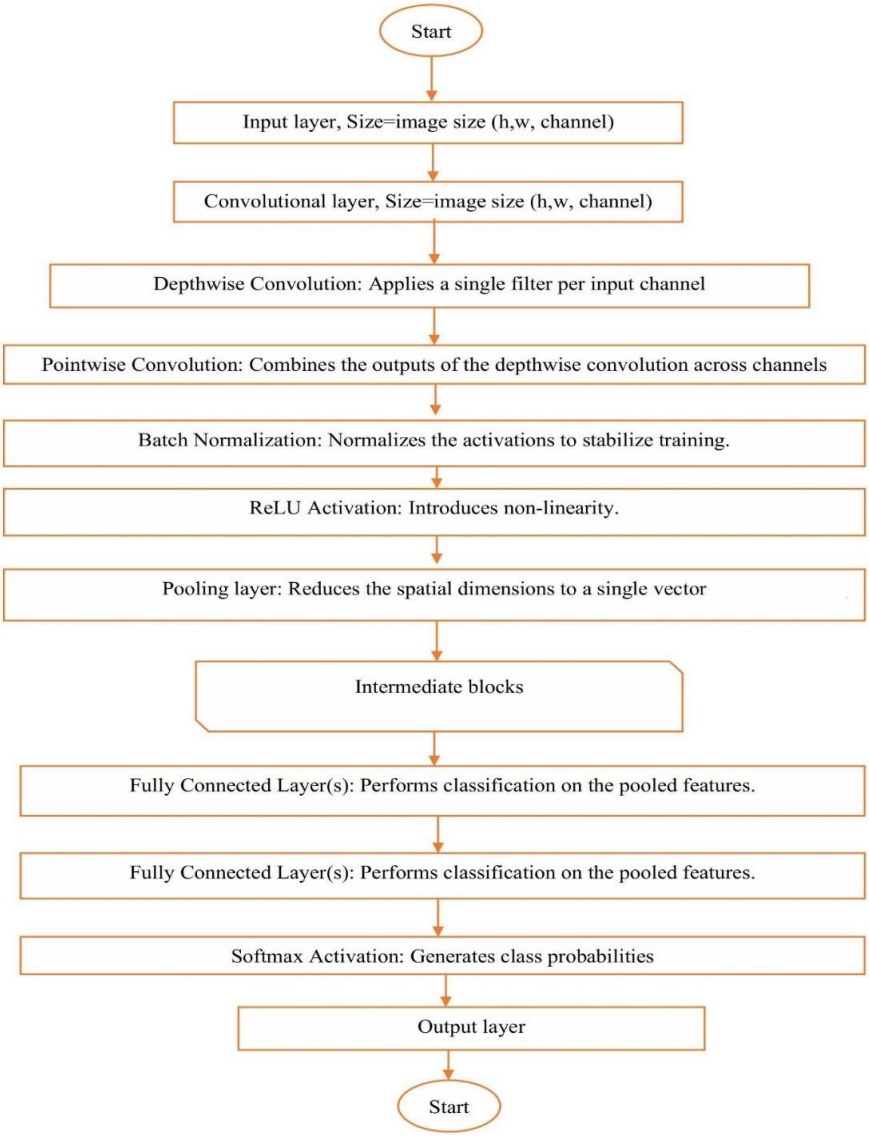

Figure 1. The process of mobilenet's working mechanism

3.2. Residual network

Residual network (ResNet) is based on CNN architecture to address the problem of vanishing gradients in very deep networks [27]. Diminishing gradient diminishes can result in performance stagnation or decline [28]. It is a groundbreaking computer vision model that accommodates several convolutional layers. The algorithm leverages previous layer activations and consolidates the network into fewer layers. The ResNet models skip two or three layers in succession by integrating nonlinearity and batch normalization [29]. While retraining, the network expands, and the residual parts and skipped layers further explore the feature space of input images. The number of layers to skip is determined by skip weights. Additionally, the residual block is a critical component of ResNet. They are designed for efficient training and improved performance. In addition, the algorithm adds an intermediate input to the output of convolution blocks (output = F(x) + x), where x represents the input to the residual block and the output from the previous layer. F(x) is the multiple convolutional blocks. Furthermore, ResNet has scalability to 50, 100, or 150 layers without computational burden by smoothing out gradient flow during backpropagation [30]. The ResNet process is depicted in Figure 2.

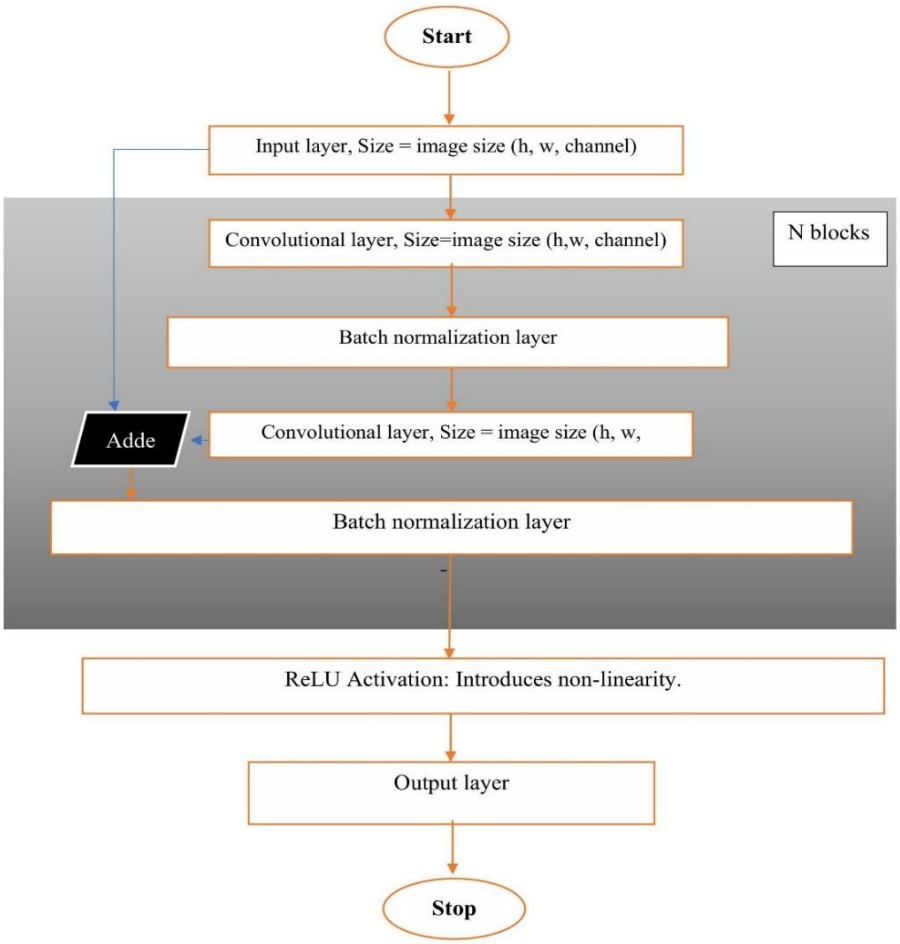

Figure 2. ResNet's working mechanism process

3.3. Convolutional neural networks

CNNs are deep learning algorithms tailored to process input images by convolving them with filters or kernels [31]. This algorithm extracts pertinent features [31]. CNNs are an evolved iteration of artificial neural networks (ANNs) employed for feature extraction from matrix-based datasets [32]. They are essential in visual datasets such as images or videos where discerning data patterns is essential [33]. Within CNNs, convolution layers, including learnable filters or kernels, are characterized by diminutive dimensions like the input volume's depth, typically 3 for image input. When an image (N*N) encounters a filter (f*f) through convolution, the operation identifies consistent features across the entire image. Subsequently, the window

slides iteratively learn features to produce the feature maps. These maps contain the local receptive field of the image and operate under the principle of shared weights. For instance, when applying convolution to an image with dimensions (34*34*3), filters can assume dimensions of *a*a*3*, where (a) represents values like 3, 5, or 7. The values are usually smaller than the image's dimensions.

In the forward pass, each filter navigates the entire input volume incrementally with values like 2, 3, or 4 for a high-dimensional, stride image [34]. It computes the dot product between kernel weights and input volume patches. As filters navigate, they produce 2-dimensional (2D) outputs, which are stacked to form an output volume mirroring the number of filters employed [35]. As in Figure 3, the network acquires knowledge of all filters to optimize its features within the input data through the iterative process.

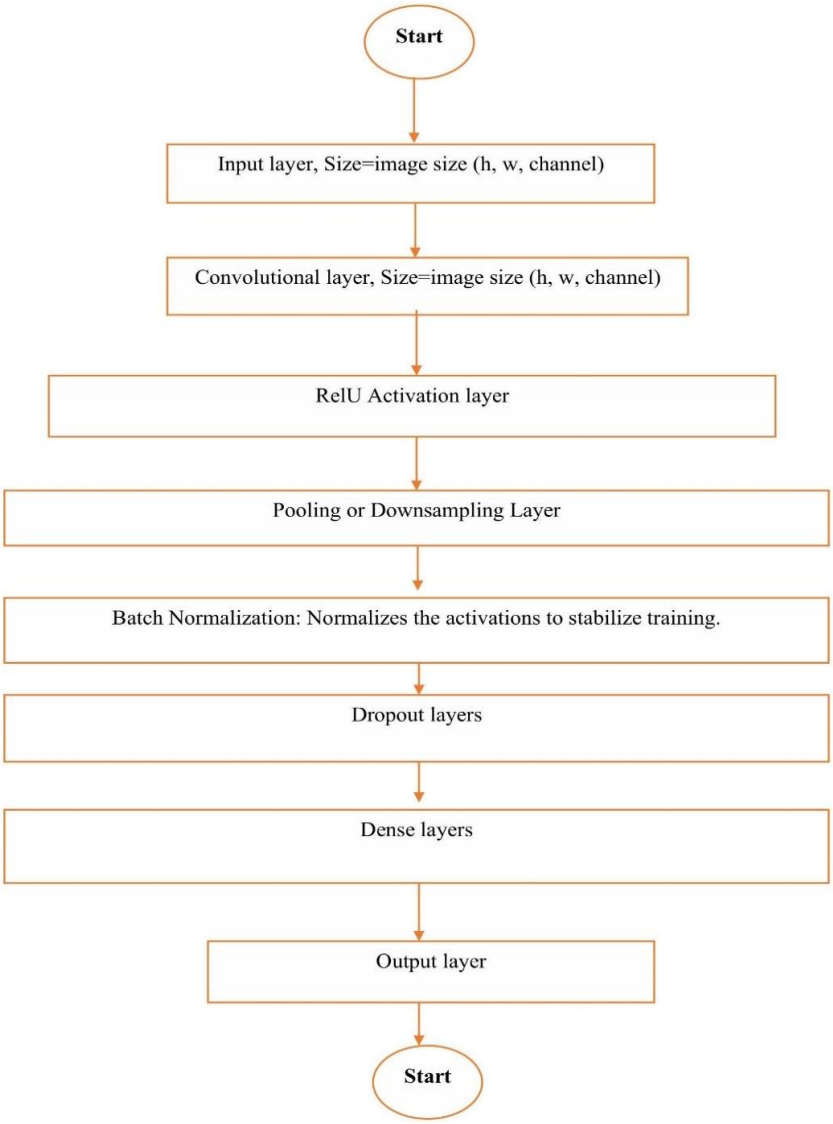

Figure 3. CNN working mechanism process

4. FINDINGS

The accuracy of the three proposed models, MobileNet, ResNet, and CNN, is determined using 10-fold cross-validation. The three models' accuracy, mean accuracy, recall, and average MAE are recorded. The results in Table 2 are derived from the three models. Interestingly, Table 2 shows that MobileNet consistently achieved the highest accuracy of about 99% across all validation folds. This finding aligns with previous studies [25], [36], [37]. The finding highlights MobileNet's reliability and robustness, especially in tasks where high accuracy is crucial. ResNet’s performance is slightly lower but still robust, with a 94% accuracy.

Meanwhile, CNN has the lowest accuracy, suggesting better models for this task may exist. However, CNN models are typically more straightforward and faster to train. Hence, they could be more useful in situations that require greater computational efficiency or deployment speed. Hence, the outcome confirms that MobileNet is more consistent across all validation folds.

Table 2. Accuracy measure of the image-based panic detection

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
MobileNet	0.97	0.98	0.97	0.98	0.98	0.99	0.98	0.98	0.97	0.98
ResNet	0.92	0.93	0.92	0.94	0.93	0.93	0.94	0.93	0.92	0.93
CNN	0.89	0.90	0.88	0.91	0.90	0.91	0.92	0.90	0.89	0.90

The average accuracy of the proposed algorithms is shown in Table 3. Table 3 revealed that the MobileNet model outperformed the other models with an accuracy of 98%. This also accords with our earlier observations, which showed that the MobileNet model has a higher accuracy [36]-[38]. This indicates that MobileNet is highly effective at detecting panic through image-based inputs. Thus, MobileNet can reliably distinguish panic-related patterns in images with minimal error. Its performance demonstrates suitability for critical applications like real-time panic detection systems in public spaces or healthcare settings requiring high precision. For instance, detecting panic early through video surveillance is critical in airport security or crowd control at festivals or national events. MobileNet's high accuracy ensures fewer false alarms and more reliable results [39]. In such environments, decision-making needs to be quick and precise.

Table 3. The mean accuracy measure of the image-based panic detection

Model	Average accuracy (%)
MobileNet	0.980
ResNet	0.930
CNN	0.900

ResNet had an accuracy of 93%, while CNN had 90 %. Although ResNet might be a viable option, the system can help detect panic signs or distress in patients or students in local clinics or smaller school health centers. Meanwhile, the CNN model was observed to have less accuracy than other models. However, it could be suitable for smaller-scale, cost-effective, low-stakes projects prioritizing effortlessness and speed over accuracy.

The Recall of the proposed three models, MobileNet, ResNet, and CNN, are determined using 10-fold cross-validation. The results are displayed in Table 4. Concerning recall, MobileNet is the most reliable model, with an average recall of 96.8%. This indicates that most true panic cases were detected. With the lowest recall, CNN is suitable for applications where detecting panic instances is crucial [40]. Thus, MobileNet is appropriate for tasks requiring a balance between precision and recall, as it minimizes false negatives. In Table 5, MobileNet outperformed the other models by scoring a recall of 96.9 %, followed by ResNet and CNN with 91.9% and 88%, respectively. Similarly, Al Reshan *et al.* [41] showed that MobileNet had a high recall rate in detecting pneumonia from chest X-ray images.

Table 4. Recall the measure of image-based panic detection

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
MobileNet	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.03	0.04	0.03
ResNet	0.08	0.07	0.08	0.06	0.07	0.07	0.06	0.07	0.08	0.07
CNN	0.11	0.10	0.12	0.09	0.10	0.09	0.08	0.10	0.11	0.10

Table 5. Mean Recall measure of the image-based panic detection

Model	Average recall
MobileNet	0.969
ResNet	0.919
CNN	0.880

The MAE of MobileNet, ResNet, and CNN are determined using 10-fold cross-validation. The results are extracted and presented in Table 6. The finding shows that MobileNet MAE values range between 0.02 and 0.04. This indicates that MobileNet has higher accuracy with lower errors across all folds. Besides, CNN had less accuracy and higher errors than the other models.

The average MAE of the proposed algorithms is shown in Table 7. MobileNet outperformed the other models with an MAE of 0.032. The other models had MAEs of 0.071 (ResNet) and 0.100 (CNN).

Table 6. MAE of the image-based panic detection

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
MobileNet	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.03	0.04	0.03
ResNet	0.08	0.07	0.08	0.06	0.07	0.07	0.06	0.07	0.08	0.07
CNN	0.11	0.10	0.12	0.09	0.10	0.09	0.08	0.10	0.11	0.10

Table 7. MAE of the image-based panic detection

Model	Average MAE
MobileNet	0.032
ResNet	0.071
CNN	0.100

5. CONCLUSION

Four models, namely MobileNet, ResNet, and CNN, were analyzed to identify panic detection through deep learning. The results indicate that MobileNet is the most effective model for image-based panic detection across ten folds. The accuracy using MobileNet is 90%. Also, the model had a recall of 96.9% and a mean accuracy of 0.032. Additionally, the MAE of MobileNet was between 0.02 and 0.04. ResNet had an accuracy of 93% and a recall of 91.9%. Concerning CNN, while not as accurate (90%), it may still be helpful in environments where computational efficiency or simplicity is required. The contribution of this study has been to confirm MobileNet's suitability for image-based panic detection. The findings contribute to developing more reliable and accurate image-based panic detection systems in real-world applications. It offers valuable insights and lays the groundwork for future studies in deep-leaning-based panic detection.

Table 1. Emotion and crown recognition techniques

No	Technique	Aim	Dataset type	Outcome	Advantages	Limitation
1	SISTCM and two-stream LSTM for facial expressions, ACCM for body gestures.	Improve video emotion recognition by extracting spatio-temporal features.	Five common datasets.	A significant improvement over the state-of-the-art method.	Utilizes facial expressions and body gestures for comprehensive emotion recognition.	Limited details on specific datasets used.
2	Comparison of machine learning, deep learning, and hybrid methods.	Driver facial expression recognition (DFER).	Various datasets from 2018 to 2023.	Highlights the importance of improving road safety.	Provides a comprehensive analysis of recent techniques.	It does not introduce new techniques, only reviews existing ones.
3	Two-stage and one-stage CNN-based methods.	Review DL-based object detection, focusing on crowd analysis.	Various datasets for object detection and crowd analysis.	Significant advances in object detection for various applications.	Efficient and effective DL techniques for object detection.	It focuses more on reviewing existing methods rather than innovations.
4	Review of state-of-the-art models, DL architectures, and information fusion techniques.	A systematic review of DL-based MER systems.	Various multimodal datasets.	Identifies key challenges and suggests future research directions.	Comprehensive overview of recent advancements.	Lack of specific new methodologies proposed.
5	Integral graph method, weak classifier, dynamic sequence model, optical flow.	Improve facial expression recognition with big data technology.	Not specified, used simulations.	Achieved 91.78% accuracy in simulations.	Enhances robustness in expression categorization.	Limited real-world dataset validation.

Table 1. Emotion and crown recognition techniques (*Continue…*)

No	Technique	Aim	Dataset Type	Outcome	Advantages	Limitation
6	DFEER system with VGGNet, optical flow reconstruction with PMVO.	Address limitations in driver facial emotion recognition in IoT context.	CK+ and KMU-FED.	High accuracy, recall, precision, and f-measure.	Tackles issues like occlusions and lighting changes.	Complex systems may require high computational resources.
7	Review of techniques using physical and physiological signals.	A systematic review of emotion recognition using various signals.	142 journal articles reviewed.	Detailed analysis of existing studies and datasets.	Broad coverage of emotion recognition techniques.	No new methods focus on reviewing the past decade's literature.
8	iSecureHome with EmoFusioNet, stacked, and late fusion methodologies.	Develop real-time facial emotion-based security for smart homes.	Experimental datasets for SH security.	Achieved 98.48% training and 98.43% test accuracy.	High accuracy in real-time emotion detection.	Specific to smart home security, not generalizable.
9	Review of deep learning for crowd anomaly detection.	Propose taxonomy for crowd behavior analysis.	Various datasets for crowd analysis.	Emphasizes the need for real-world challenging datasets.	Brings emotional aspects into crowd behavior studies.	Lacks new methodological contributions.
10	Advanced Fake Image-Feature Network (AFIFN) with DCT and Y Cr Cb.	Detect forged images to address digital media security.	Not specified.	Outperforms existing models in image forgery detection.	Effective in distinguishing real and fake images.	Limited information on the dataset and real-world applicability.
11	Facial emotion recognition and text emotion recognition.	Identify and predict anxiety in university students.	AMAS-C for validation.	Facial: 84.21% precision, Text: 86.84% precision.	Effective early detection of anxiety.	Limited to an academic setting, may need to generalize better.
12	Review of techniques in various modalities and information fusion.	Review of MER techniques.	Various unimodal and multimodal datasets.	Comprehensive understanding of emotion recognition progress.	Covers diverse domains and applications.	It focuses on reviewing rather than proposing new methods.
13	Multimodal feature representation with GCN and ensemble learning.	Improve sentiment recognition of online public opinion.	Sina Weibo data (COVID-19 context).	F1-score: 84.13% (sentiment polarity), 82.06% (fine-grained).	Enhances sentiment recognition accuracy.	Specific to online public opinion, limited generalizability.
14	GNN with functional connectivity and attention mechanisms.	Explore the connection between panic emotion and driving ability using EEG.	Simulated driving environment data.	Binary classification: 91.5% accuracy.	Effectively monitors emotional state in a driving context.	Limited to a simulated environment, real-world validation is needed.
15	Review of visual, auditory, linguistic modalities and joint representations.	Review of unimodal and MER techniques.	Various video and multimodal datasets.	Identifies gaps and suggests future research directions.	Broad review across multiple modalities.	Lacks focus on proposing new methods, more of a review.

ACKNOWLEDGEMENTS

The authors received no specific funding for this study. The authors declare no conflicts of interest to report regarding the present study.

FUNDING INFORMATION

Authors state no funding involved.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] D. Guo *et al.*, "Research on cam–kalm automatic tracking technology of low, slow, and small target based on Gm-APD LiDAR," *Remote Sensing*, vol. 17, no. 1, p. 165, Jan. 2025, doi: 10.3390/rs17010165.

[2] A. Albaloushi, "The effective use of social media in crime dectection and prevention: the promotion of public trust in the UAE police-the case of the Abu Dhabi Police," *Doctoral dissertation, Cardiff Metropolitan University*, 2019.

[3] B. Aldissi and H. Ammar, "Real-time frequency-based detection of a panic behavior in human crowds," *Multimedia Tools and Applications*, vol. 79, no. 33–34, pp. 24851–24871, Jun. 2020, doi: 10.1007/s11042-020-09024-z.

[4] A.-F. A. Mentis, D. Lee, and P. Roussos, "Applications of artificial intelligence−machine learning for detection of stress: a critical overview," *Molecular Psychiatry*, vol. 29, no. 6, pp. 1882–1894, Apr. 2023, doi: 10.1038/s41380-023-02047-6.

[5] A. Arora, P. Chakraborty, and M. P. S. Bhatia, "Problematic use of digital technologies and its impact on mental health during COVID-19 pandemic: assessment using machine learning," in *Emerging Technologies During the Era of COVID-19 Pandemic*, Springer International Publishing, 2021, pp. 197–221.

[6] M. Aboualola, K. Abualsaud, T. Khattab, N. Zorba, and H. S. Hassanein, "Edge technologies for disaster management: a survey of social media and artificial intelligence integration," *IEEE Access*, vol. 11, pp. 73782–73802, 2023, doi: 10.1109/access.2023.3293035.

[7] O. Aydın, K. Balıkçı, F. P. Çökmüş, and P. Ünal Aydın, "The evaluation of metacognitive beliefs and emotion recognition in panic disorder and generalized anxiety disorder: effects on symptoms and comparison with healthy control," *Nordic Journal of Psychiatry*, vol. 73, no. 4–5, pp. 293–301, Jun. 2019, doi: 10.1080/08039488.2019.1623317.

[8] L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim, and S. Bahri Musa, "The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–9, doi: 10.1109/icic50835.2020.9288560.

[9] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with CNN ensemble," in *2016 International Conference on Cyberworlds (CW)*, Sep. 2016, pp. 163–166, doi: 10.1109/cw.2016.34.

[10] S. Sahoo, S. Mishra, B. Panda, A. K. Bhoi, and P. Barsocchi, "An augmented modulated deep learning based intelligent predictive model for brain tumor detection using GAN ensemble," *Sensors*, vol. 23, no. 15, p. 6930, Aug. 2023, doi: 10.3390/s23156930.

[11] D. Nguyen, K. Nguyen, S. Sridharan, A. Ghasemi, D. Dean, and C. Fookes, "Deep spatio-temporal features for multimodal emotion recognition," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar. 2017, pp. 1215–1223, doi: 10.1109/wacv.2017.140.

[12] I. Saadi, D. W. cunningham, A. Taleb-Ahmed, A. Hadid, and Y. El Hillali, "Driver's facial expression recognition: A comprehensive survey," *Expert Systems with Applications*, vol. 242, p. 122784, May 2024, doi: 10.1016/j.eswa.2023.122784.

[13] M. Jeong and B. C. Ko, "Driver's facial expression recognition in real-time for safe driving," *Sensors*, vol. 18, no. 12, p. 4270, Dec. 2018, doi: 10.3390/s18124270.

[14] K. Tang, Y. Tie, T. Yang, and L. Guan, "Multimodal emotion recognition (MER) system," in *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2014, pp. 1–6, doi: 10.1109/ccece.2014.6900993.

[15] S. Kalateh, L. A. Estrada-Jimenez, S. Nikghadam-Hojjati, and J. Barata, "A systematic review on multimodal emotion recognition: building blocks, current state, applications, and challenges," *IEEE Access*, vol. 12, pp. 103976–104019, 2024, doi: 10.1109/access.2024.3430850.

[16] P. Bour, E. Cribelier, and V. Argyriou, "Crowd behavior analysis from fixed and moving cameras," in *Multimodal Behavior Analysis in the Wild*, Elsevier, 2019, pp. 289–322.

[17] A. Bhattacharya, P. Dash, M. Jain, and A. Jothimani, "Smart home security system using emotion detection," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 5, 2020.

[18] O. S. Ekundayo and S. Viriri, "Facial expression recognition: a review of trends and techniques," *IEEE Access*, vol. 9, pp. 136944–136973, 2021, doi: 10.1109/access.2021.3113464.

[19] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/taffc.2020.2981446.

[20] G. A.V., M. T., P. D., and U. E., "Multimodal emotion recognition with deep learning: advancements, challenges, and future directions," *Information Fusion*, vol. 105, p. 102218, May 2024, doi: 10.1016/j.inffus.2023.102218.

[21] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "Multimodal emotion recognition using deep learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 73–79, May 2021, doi: 10.38094/jastt20291.

[22] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689–694, 2020, doi: 10.1016/j.procs.2020.07.101.

[23] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: a review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019, doi: 10.1109/access.2019.2936124.

[24] H. Ranganathan, S. Chakraborty, and S. Panchanathan, "Multimodal emotion recognition using deep learning architectures," Mar. 2016, doi: 10.1109/wacv.2016.7477679.

[25] H.-Y. Chen and C.-Y. Su, "An enhanced hybrid MobileNet," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 308–312, doi: 10.1109/icawst.2018.8517177.

[26] D. Sinha and M. El-Sharkawy, "Thin MobileNet: an enhanced MobileNet architecture," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, Oct. 2019, pp. 0280–0285, doi: 10.1109/uemcon47517.2019.8993089.

[27] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, "A novel image classification approach via dense-MobileNet models," *Mobile Information Systems*, vol. 2020, pp. 1–8, Jan. 2020, doi: 10.1155/2020/7602384.

[28] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021, doi: 10.1016/j.procs.2021.01.025.

[29] F. He, T. Liu, and D. Tao, "Why ResNet works? residuals generalize," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5349–5362, Dec. 2020, doi: 10.1109/tnnls.2020.2966319.

[30] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: a survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, Sep. 2022, doi: 10.3390/app12188972.

[31] Y. Ma, M. Kim, Y. Cao, S. Vrudhula, and J. Seo, "End-to-end scalable FPGA accelerator for deep residual networks," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4, doi: 10.1109/iscas.2017.8050344.

[32] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[33] J. S. Kumar, S. Anuar, and N. H. Hassan, "Transfer learning based performance comparison of the pre-trained deep neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022, doi: 10.14569/ijacsa.2022.0130193.

[34] P. G. Brodrick, A. B. Davies, and G. P. Asner, "Uncovering ecological patterns with convolutional neural networks," *Trends in Ecology & Evolution*, vol. 34, no. 8, pp. 734–745, Aug. 2019, doi: 10.1016/j.tree.2019.03.006.

[35] G. Seetharaman, "High-performance computing in computer vision," in *Computer Vision*, Springer International Publishing, 2021, pp. 564–567.

[36] K. V. Savant, G. Meghana, G. Potnuru, and V. Bhavana, "Lane detection for autonomous cars using neural networks," in *Machine Learning and Autonomous Systems*, Springer Nature Singapore, 2022, pp. 193–207.

[37] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, "Convolutional neural networks using MobileNet for skin lesion classification," in *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jul. 2019, pp. 242–247, doi: 10.1109/jcsse.2019.8864155.

[38] J. Bethge, C. Bartz, H. Yang, Y. Chen, and C. Meinel, "MeliusNet: an improved network architecture for binary neural networks," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021, pp. 1438–1447, doi: 10.1109/wacv48630.2021.00148.

[39] B. Khasoggi, E. Ermatita, and S. Samsuryadi, "Efficient mobilenet architecture as image recognition on mobile and embedded devices," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 1, p. 389, Oct. 2019, doi: 10.11591/ijeecs.v16.i1.pp389-394.

[40] Y. Kaya and E. Gürsoy, "RETRACTED ARTICLE: A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection," *Soft Computing*, vol. 27, no. 9, pp. 5521–5535, Jan. 2023, doi: 10.1007/s00500-022-07798-y.

[41] M. S. Al Reshan *et al.*, "Detection of Pneumonia from chest X-ray images utilizing MobileNet model," *Healthcare*, vol. 11, no. 11, p. 1561, May 2023, doi: 10.3390/healthcare11111561.

BIOGRAPHIES OF AUTHORS

Sameerah Faris Khlebus is a PhD student at University of Sfax, Sfax, Tunisia. He holds a BSc in Computer Science and an MSc in Data Security from the Department of Computer Science, University of Technology, Baghdad, Iraq. His research areas are computer security, image processing, artificial intelligence, machine learning, deep learning, natural language processing, internet of things, cloud computing, and digital signal processing. Sameerah has published in server journals. She can be contacted at email: sameerah.alradhi@uoitc.edu.iq.

Dr. Mohammed Salih Mahdi received the B.Sc. degree in Computer Science, M.Sc. degree in Cloud Security, and the Ph.D. degree in Computer Science from University of Technology, Iraq. He is an Assistant Professor at the Business Information College, University of Information Technology and Communications, Iraq. He has authored or coauthored more than 40 publications, with 12 H-index and over 400 citations. Her research interests include include data security, steganography, image processing, data compression, artificial intelligence, data mining, machine learning, deep learning, internet of things, cloud computing, quantum computing and blockchain technology. Hee can be contacted at email: mohammed.salih@uoitc.edu.iq.

Monji Kherallah received his received the Diploma Ing, PhD, and HU degrees in electrical engineering from ENIS, Universityof Sfax, Sfax, Tunisia from 1989-2012. He is a Professor with the Faculty of Science, University of Sfax, Tunisia. He served as an Engineer at the Biotechnology Center, University of Sfax for fourteen years. Dr. Monji is the founder of a professional master's degree: Metrology and Industrial Instrumentation at the Faculty of Sciences, University of Sfax, Tunisia. He has authored or coauthored more than 190 publications, with 28 H-index and over 2900 citations. His research interest includes machine learning, deep learing, signal and image processing. He can be contacted at email: monji.kherallah@fss.usf.tn.

Prof. Ali Douik received his B.S., M.S., and Ph.D. degrees in electrical engineering from ENSET, Tunis from 1988-1996. He also has a HDR degree in electrical engineering from the University of Monastir, Monastir, Tunisia, in 2010. He is attached to the Faculty of Biotechnology and Biomolecular Sciences. He is currently a full professor with the Department of Industrial Computing, National Engineering School of Sousse. Previously, he worked at the National Engineering School of Monastir, from September 1991 to September 2014. He has authored or coauthored more than 100 publications, with 15 H-index and over 900 citations. Prof. Ali research area include digital image processing, artificial intelligence, machine learning, deep learning, automatic control, optimization, and evolutionary algorithms. He can be contacted at email: ali.douik@eniso.u-sousse.tn.