International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

An Efficient Deep Learning Framework for Adaptive Emotion Detection in Diverse Environments

Sara Mishra
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

Muskan
Department of Computer Science and Engineering
Galgotias University Greater Noida, India

Ashish Shrivastava
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

Pradeep Kumar
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

Neeraj Kumar
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

Pawan Kumar
Department of Computer Science and Engineering
Galgotias University
Greater Noida, India

***Abstract*—Facial expression recognition has become increasingly relevant in the last few years particularly due to the abundance of communication is now done online, rather than in person. Even though many challenges still exist, there are already a lot of approaches available that perform well in case the environment is uncontrolled. Things like poor lighting, shadows and occluded facial region make the results unreliable.**

There are five models, which include the traditional model in this study to determine machine-learning and deep-learning approaches which are more effective in the real world. The best model with the best accuracy of 87.9% was ResNet50. MobileNetV2 was also quite successful and appeared more efficient. According to the results, deep learning, in particular, transfer learning, would be better in practical emotion-recognition systems.

***Index Terms*—Facial Expression Recognition, Deep Learning, Real-Time FER, CNN, Emotion Classification, Image Processing**

I. Introduction

The facial expressions are one of the most direct forms of expressing emotion, which can be more accurate than even words or actions. As virtual environments have quickly developed, gadgets with the ability to consistently recognize emotions have increasingly become important- most notably in the fields of online classrooms and assistive technology. Earlier methods were based on the use of handcrafted features, which only worked in controlled environments but had no generalization. Whereas deep-learning models learned the characteristics themselves, resulting in increased accuracy, but the facial-emotion data are still problematic in terms of cultural, light, pose, and individual differences. Therefore, in this paper the strength of various models is considered to determine which one is able to cope with the variability in the real life.

II. Literature Survey

Facial Expression Recognition (FER) research has experienced an apparent change in the past few years. Initially, the bulk of the work was done based on handcrafted features. Most older systems were based on the Facial Action Coding System (FACS) of Ekman and Friesen, which is, in essence, a system that links certain movements of the face muscles to an emotional state [11]. Such models as SVM, Random Forest, and k-Nearest Neighbours were all common at that stage. They worked well, but chiefly in the cases where the pictures were made when it was light and it was possible to see the face. As soon as faces did not match perfectly or the light changed, the level of performance dropped at a rather rapid rate. It can be observed that most of the earlier approaches were particularly problematic due to the lack of diversity of the data they worked on.

The change began with the popularity of deep learning. Convolutional Neural Networks (CNNs) allowed the training of models on the actual images instead of the features that were manually designed. Popular architectures, such as AlexNet [19], VGGNet [5], and ResNet [6], had contributed towards strong baselines. Approximately, at the same period, datasets such as FER-2013 [9], AffectNet [3], and RAF-DB [10] emerged and provided researchers with a lot more to work with. The application of deep learning models led to a rapid development in the field.

The further development of the research involved the examination of more complicated architectures and methods in order to achieve better generalization. As an example, Mollahosseini et al. trained deeper CNN models and demonstrated that their

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

training on multiple datasets led to higher robustness [16]. Subsequently, Li and Deng identified a number of obstacles that remain topical these days, including the demographic disparity, lighting problems, and the noise in labels [2]. To address these issues, scholars came up with attention mechanisms, multi-branch CNNs and even Capsule Networks which attempt to preserve the spatial relationship among features intact [7]. There were papers that did not agree on the best model and this was so when it comes to subtle displays such as fear or disgust.

Generative Adversarial Networks (GANs), as well, started to be used in FER research. The GANs were used to produce new samples in order to resolve class imbalance and recreate conditions such as occlusion or bad lighting [21]. The second concept that seems interesting to me was soft-label learning, in which there are no one-to-one labels on emotions, rather probabilities, and this is reasonable in case some expressions do not belong to a single emotion [4].

Transformer-based models and CNN-Transformer hybrids are receiving attention in the past years. Global self-attention is employed in these models, and appears to perform effectively on small scale emotional cues [25]. However, simultaneously, transformers are very computationally heavy and thus, numerous practicable systems continue to depend on transfer learning as it provides a superior tradeoff amid training and performance [2]. It is true that the issues of deployment are as significant as accuracy is.

The issue of efficiency became of interest later when FER was used in the real-time applications. Lightweight models such as MobileNetV2 [20] and GhostNet demonstrated that one can attain a high accuracy with a small model. Zhang et al. and Xia et al. studies [17] and [23] have demonstrated these smaller models to be effective on devices such as smartphones. When FER goes mainstream and can run on constrained hardware, then it is likely that performance by running smoothly will be more important than teasing the final few accuracy bits out of it.

Other researchers have also experimented with hybrid methods such as, deep learning will extract features, but classification will be performed through conventional models. Tang discovered that a combination of deep features with a linear SVM could even outperform pure traditional models and the standalone CNNs [8].

It has also changed to multi-modal strategies, that is the use of facial expression combined with speech or pose or even physiological expression [12], [18]. And since FER is under consideration in such domains as healthcare and education, explainability has taken a significance. Such tools as Grad-CAM can be used to reveal the areas of the face that contributed to a prediction [14], [22].This solution will help increase model understandability by reducing black-box properties of emotion recognition systems.

Despite all these developments, there are still a number of challenges that have not been addressed. The cultural variations, noisy data, imbalance in the dataset and the real-life variability continue to be big challenges. It is actually not clear what approach is certainly the best one - and the sphere is still dynamic and developing. These shortcomings and advancements observed in prior research were used to influence the rationale behind this research: the construction of a system not only accurate under ideal test conditions but also functional under realistic conditions.

III. Methodology

A. Dataset Selection

This study was conducted using three sets of data:

- FER-2013: A noisy dataset having a large amount of real-world variations.
- CK+: A significantly cleaner dataset which is well-articulated.
- RAF-DB: A variety of data containing realistic facial expressions.

The fact that more than one dataset was used assisted the model in learning both regulated and natural facial differences.

B. Preprocessing

Prior to training, every image underwent a number of processes. The face was identified and cut off. Then, the image was scaled, normalized, grayed out where necessary. Information addition, including rotations and horizontal flips, was included to assist the model in handling variations. This step was noteworthy since certain emotions (such as fear or disgust) did not appear as often as others.

C. Model Design

Both classical models and deep-learning were considered in the study ones:

- SVM and Random Forest are considered classical
- Deep learning: A benchmark CNN, MobileNetV2, and ResNet50

Transfer-learning models would have done better since they begin with pretrained knowledge.

IV. Experiments and Results

The four standard measures that were used to assess the performance of the proposed models were accuracy, precision, recall, and F1-score. Classical machine-learning solutions, as well as deep-learning models, were experimented, to discuss their performance in the realistic environment.

TABLE I
Performance Comparison of Different Models

Model	Accuracy	Precision	Recall	F1-score
SVM	58.2%	56.9%	54.3%	55.5%
Random Forest	62.4%	60.2%	59.7%	60.1%
Custom CNN	74.8%	73.3%	72.4%	73.0%
MobileNetV2	85.4%	84.8%	83.9%	84.3%
ResNet50	**87.9%**	**87.2%**	**86.8%**	**87.0%**

Deep-learning models beat traditional ML, and transfer-learning architectures outshine the custom CNN. Visual graphs made metric comparisons straightforward.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

A. Accuracy

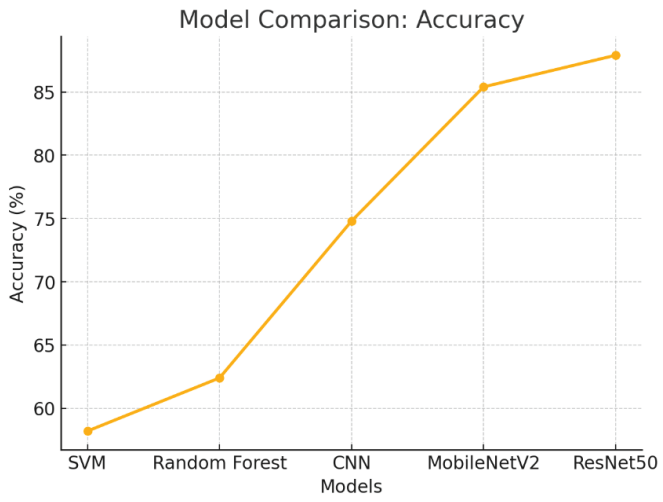

Fig. 1. Comparison of all of the assessed models in terms of accuracy.

The graph indicates an increase in the accuracy with the increase in model complexity. MobileNetV2 and ResNet50 were far more successful compared to traditional methods with SVM and Random Forest lagging well behind.

Among them is the high improvement in the performance of the transfer-learning models. The transition between the traditional CNN and MobileNetV2 and ResNet50 is rather evident, and it can indicate that pretrained features become extremely useful when it comes to performing the task of emotion recognition.

B. Precision

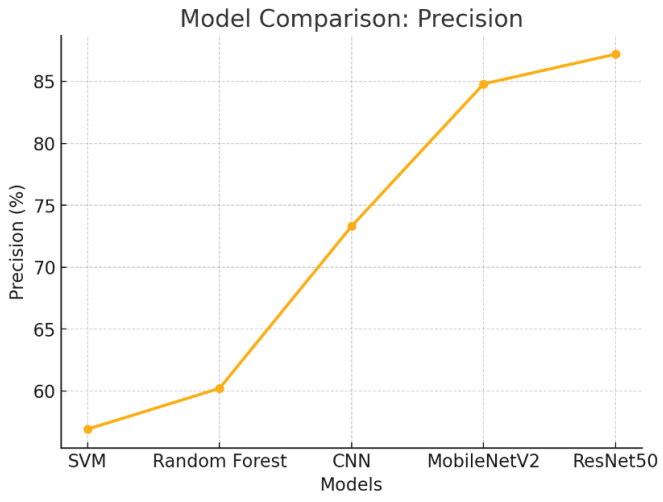

Fig. 2. Precision of various machine-learning and deep-learning models.

Precision is also similar to the accuracy results. The models that are most difficult to reach the traditional ones, and the highest value of precision is obtained by ResNet50.

Precision informs us on the rate with which the predicted emotion was true. Again, the leader is ResNet50, and MobileNetV2 follows it by a narrow margin. This implies that the number of false positives is lower than in the previous models.

C. Recall

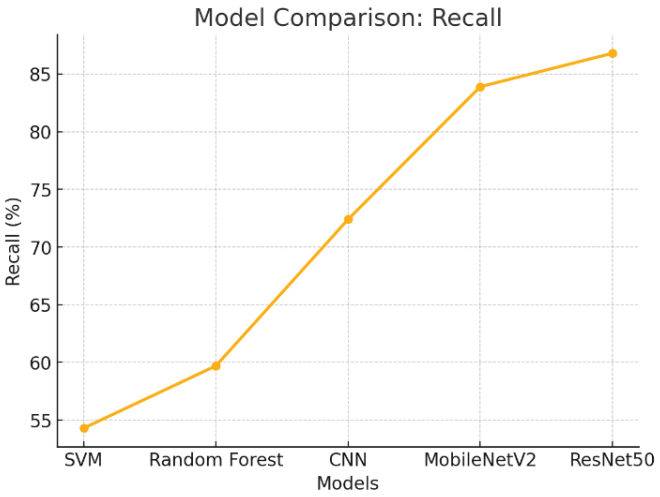

Fig. 3. Compare between all models.

The more in-depth models have an advantage here too, which is that they were more efficient at recognizing emotions without false cases.

Fear and disgust were more difficult to discern. Even stronger models did not pass without difficulties, which is also characteristic of other FER research.

D. F1-Score

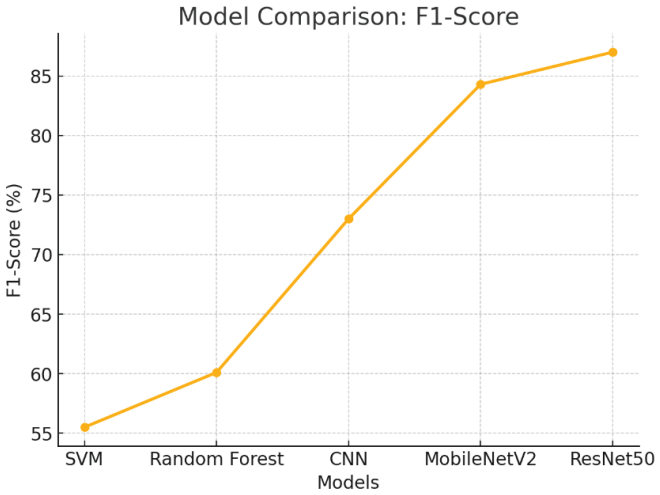

Fig. 4. Values of F1-scores of all tested models.

F1-score is a combination of recall and precision and thus it provides a well-rounded view of performance. In this case,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

once again we see transfer-learning models lead the pack in terms of results.

The pattern of the F1-score agrees with the earlier graphs. It is possible to note that ResNet50 is the highest-performing, and MobileNetV2 comes close to it. It is evident that the classical models do not perform well in face recognition task in the real world.

V. Discussion

Upon a closer examination of the results, one can notice the difference between classical models and deep-learning models becomes very apparent. The classic models were not working well, particularly when faces were faint or shadows were unearthed. The custom CNN did get the better of things, but the greatest leap occurred with transfer-learning models.

ResNet50 did the most generally, which is likely due to the reason that it achieves greater patterns by its structure. MobileNetV2 was also remarkable, particularly in view of its lightness and speed it is. In case a person wanted to process real-time emotion recognition MobileNetV2 would be the most used on a phone or an embedded system likely be the better choice.

One of the things that left an impression was that certain feelings were constantly more difficult to categorize, particularly fear and disgust. Even during labeling, the following categories appeared too often similar.

VI. Conclusion

According to the findings, it is evident that deep learning, especially when it is used in conjunction with transfer learning, is much better than any classical machine-learning approach based on facial emotion recognition. ResNet50 was the most precise, and MobileNetV2 is simpler to implement in real-time. Overall, the approach investigated in this paper may be used as a starting point in building efficient and deployment-ready FER systems.

VII. Future Work

There are certain ways in which it might be improved. One of them is to express facial expression in line with the voice or text to be more accurate. The other would be testing models that use transformers, because these have worked well in vision tasks in recent times. It also possesses room to optimize model size, which allows it to be used on low-power machines. Lastly, the investigation must be carried out in more depth in order to make sure that the system is fair to different age brackets and cultures.

References

[1] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[2] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.

[3] A. Mollahosseini, D. Chan, and M. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.

[4] A. Barsoum, C. Zhang, C. Canton-Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowdsourced label distribution," in *Proc. ACM Int. Conf. Multimodal Interaction (ICMI)*, 2016, pp. 279–283.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 [cs.CV], 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[7] S. Sabour, N. Frosst, and G. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 3856–3866.

[8] Y. Tang, "Deep learning using linear support vector machines," arXiv:1306.0239 [cs.LG], 2013.

[9] Kaggle, "FER-2013 facial expression recognition dataset," 2013. [Online]. Available: https://www.kaggle.com/datasets/msambare/fer2013. Accessed: Nov. 27, 2025.

[10] S. Li *et al.*, "RAF-DB: Real-world affective faces database," *IEEE Transactions on Affective Computing*, 2017.

[11] P. Ekman and W. Friesen, *Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[12] S. Albanie, A. Nagrani, A. Vedaldi, and A. Zisserman, "Emotion recognition in speech using cross-modal transfer in the wild," in *Proc. ACM Multimedia Conf.*, 2018, pp. 292–301.

[13] T. Baltrusaitis, A. Zadeh, Y. Lim, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *Proc. IEEE Conf. Autom. Face Gesture Recognit. (FG)*, 2018.

[14] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.

[15] D. Kollias *et al.*, "Deep affect prediction in-the-wild: Aff-Wild database and challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017.

[16] A. Mollahosseini, B. Hasani, and M. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2016, pp. 1–10.

[17] H. Zhang *et al.*, "Lightweight CNN architectures for real-time facial expression recognition," *Sensors*, vol. 20, no. 13, p. 3781, 2020.

[18] N. Majumder *et al.*, "Multimodal emotion recognition using deep learning," *IEEE Transactions on Affective Computing*, 2021.

[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2012, pp. 1097–1105.

[20] A. Howard *et al.*, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520.

[21] Z. Huang, S. Wang, and Q. Wu, "Occlusion-robust facial expression recognition using deep generative models," *Pattern Recognition*, vol. 102, p. 107249, 2020.

[22] R. Panda, C. Qi, and A. Das, "Explainable AI for facial emotion recognition," *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2023.

[23] W. Xia *et al.*, "FER-Net: A robust facial expression recognition network under unconstrained conditions," *IEEE Access*, vol. 9, pp. 145–158, 2021.

[24] L. Ji, H. Li, and J. Sun, "A comprehensive review on real-time emotion recognition systems," *Information Fusion*, vol. 87, pp. 35–52, 2022.

[25] S. Tripathi, S. Singh, and A. Dhall, "Deep learning for emotion recognition: A survey," arXiv:2302.06529 [cs.CV], 2023.

[26] A. Rahman, S. S. Sohail, M. S. Alam, A. Sharma, and W. Mansoor, "Detecting brain cancer using explainable AI," in *Proc. 7th Int. Conf. Signal Processing and Information Security (ICSPIS)*, Nov. 2024, pp. 1–6, doi: 10.1109/ICSPIS63676.2024.10812596.

[27] Izna, A. Sharma, and R. P. Pandey, "Findings and discussions of the application of IoT and machine learning towards smart agriculture in India," in *Proc. 11th Int. Conf. System Modeling & Advancement in Research Trends (SMART)*, Dec. 2022, pp. 302–305, doi: 10.1109/SMART55829.2022.10046787.