International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 4, August 2024, pp. 4046~4054
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i4.pp4046-4054

Performance analysis of deep unified model for facial expression recognition using convolution neural network

Kavita, Rajender Singh Chhillar
Department of Computer Science and Applications, Maharshi Dayanand University Rohtak, Haryana, India

Article Info

Article history:

Received Dec 31, 2023
Revised Apr 30, 2024
Accepted May 12, 2024

Keywords:

Convolution neural network
Deep learning
Facial expression
LFW dataset
Preprocessing

ABSTRACT

Facial expression recognition has gathered substantial attention in computer vision applications, with the need for robust and accurate models that can decipher human emotions from facial images. Performance analysis of a novel hybrid model combines the strengths of residual network (ResNet) and dense network (DenseNet) architectures after applying preprocessing for facial expression recognition. The proposed hybrid model capitalizes on the complementary characteristics of ResNet's and DenseNet's densely connected blocks to enhance the model's capacity to extract discriminative features from facial images. This research evaluates the hybrid model performance and conducts a comprehensive benchmark against established facial expression recognition convolution neural network (CNN) models. The analysis encompasses key aspects of model performance, including classification accuracy, and adaptability with the labeled faces in the wild (LFW) dataset for facial expressions such as anger, fear, happy, disgust, sad, surprise, along neutral. The research observes that the proposed hybrid model is more accurate and efficient computationally than existing models consistently. This performance analysis eliminates information on the hybrid model's perspective to further facial expression recognition research.

Corresponding Author:

Kavita
Department of Computer Science and Applications, Maharshi Dayanand University Rohtak 124001, Haryana, India
Email: kavitavce@gmail.com

1. INTRODUCTION

Facial expression recognition (FER) is a fascinating and pivotal area within the domain of computer vision and AI. It focuses on the ability of machines to discern and understand human emotions and mental states by analyzing facial expressions [1]. Human communication is a complex interplay of verbal and non-verbal cues, with facial expressions being prominent non-verbal channels. Recognizing and interpreting these expressions is of immense significance in various fields [2], [3]. The human face is a canvas of emotions, a window to the inner thoughts and feelings of an individual. Whether it is a heartfelt smile, a furrowed brow denoting worry, or a raised eyebrow signaling surprise, facial expressions are a rich source of information about a person's emotional state [4]. Consequently, the development of technology capable of deciphering these cues holds enormous potential in improving our interactions with machines and, by extension, with each other [5]. FER systems aim to automatically detect, analyze, and interpret facial expressions from images or video frames. This involves the identification of key facial landmarks, extraction of relevant features, and the application of machine learning (ML) [6] and deep learning (DL) algorithms to classify expressions into emotional categories, such as happiness, sadness, anger, fear, disgust, surprise, and neutrality. The ability to recognize these emotions can be applied in diverse applications, including humanrobot interaction, sentiment analysis, mental health assessment, lie detection, personalized user experiences in technology, and the development of empathetic and intuitive artificial intelligent (AI) systems [7].

Over the years, FER witnessed significant advancements due to the evolution of DL techniques, the availability of large facial expression datasets, and the increased computational power of modern hardware [8]. These advancements have enabled the development of highly accurate and robust facial expression recognition systems, some of which can even outperform humans in certain tasks. In this era of rapid technological progress, FER is not only a subject of academic research but also a fundamental tool with practical implications in a wide range of industries [9]–[11]. This introduction provides a gateway into the intriguing world of FER, highlighting its importance and the transformative impact it can have on our digital interactions. The understanding of human behavior is a field that continues to reshape the way we understand and interact with the world around us [12], [13]. To improve the accuracy different models are developed trained and tested with different datasets. Moreover, some models integrated with different approaches for better facial expressions [14], [15]. Facial expression recognition (FER) holds substantial significance in various fields and applications due to its potential to enhance human-computer interaction, psychological research, and a wide range of industries. Here are some key aspects of the significance of FER: human-computer interaction (HCI), affective computing, mental health assessment, market research, security and surveillance, education and special needs, detection and forensics, entertainment and gaming, research, and psychology. Enable machines to understand and respond to human emotions, not only enhances technological capabilities but also has the potential to improve human well-being, safety, and the overall quality of interactions between people and technology [16]–[20].

FER systems have made strides in recent years, due to advancements in deep learning, promoting extensive research in the FER field. In this context, Khajuria *et al.* [21] focused on the convolutional neural network (CNN) and visual geometry group-16 (VGG-16) for facial emotion recognition. The accuracy achieved by the hybrid model is 91%. Khan *et al.* [22] proposed a disguise-invariant face recognition model for the challenges in an illumination variation environment. With an average runtime of 0.32 seconds, the model discovered an accuracy of 98.19%. Alonazi *et al.* [23] combined MF-based preprocessing, and a CapsNet feature extractor, model was developed for automated and accurate FER. The model achieved a better accuracy of 99.05% than previous models. Research on the suggested model's computing efficiency may be undertaken in the future. Anwarul *et al.* [24] ensemble a hybrid CNN with hyperparameter tuning was created. By using progressive training, the model's recognition accuracy was significantly improved. Hybrid ensemble CNN (HE-CNN) model obtained an accuracy of 95% on the custom-built dataset, 91.58% on the cross-pose labeled faces in the wild (LFW) dataset, and 99.35% on the LFW dataset. Methods applied such as AlexNet by SVM and residual network (ResNet-50), support vector machines (SVM) on Faculdade de Engenharia Industrial (FEI) face, LFW, YouTube, and Olivetti Research Laboratory (ORL) datasets. Almabdy and Elrefaei [25] showed that their model outperformed most state-of-the-art models in terms of accuracy. Accuracy in recognition and categorization is another area in which they want to work. Kavita and Chhillar [26] proposed a deep unified face recognition model with the integration of ResNet and DenseNet models. Compared to ResNet along with DenseNet, the proposed model achieved a better accuracy of 98.8%. The proposed innovative approach results in superior performance, classification, and feature extraction capabilities in a face recognition model. Wang and Guo [27] tested some models for face recognition using several frameworks including PyTorch, TensorFlow, and Caffe. Meena *et al.* [28] proposed a framework of sentiment identification on face expression recognition on CK+ and FER datasets with 79% and 95% accuracy respectively.

The difficulty that is now being faced with deep learning methods, more notably CNNs, is being used to accomplish the goal of accurately identifying facial expressions. The identification of facial expressions is significant in a variety of domains, including emotion analysis, and the diagnosis of mental health conditions. Existing methods often have difficulty effectively capturing minor subtleties and changes in expressions due to the complexity and variety of human emotions. Since this is the case, a more unified and robust model is required, one that is capable of handling the complexities of facial expressions across a wide range of persons and situations in an effective manner. Therefore, the goal is to construct a deep unified model that consistently recognizes and categorizes facial expressions with high accuracy, sensitivity, and generalization capabilities. This will be accomplished by harnessing the power of CNNs with the ultimate goal of contributing to breakthroughs in emotion identification technology and its applications in real-world settings.

The specific aim behind our hybrid proposed model for FER is to push the boundaries of FER technology by combining the strengths of two powerful architectures, aiming for higher accuracy, efficiency, and adaptability across diverse scenarios. This model represents a step forward in the ongoing search to create a more accurate and versatile system for understanding and responding to human emotions through facial expressions. Several key factors drive this work:

- *Enhanced feature learning:* Both ResNet and DenseNet architectures have individually demonstrated their proficiency in feature extraction for various computer vision tasks. ResNet's residual connections allow for the training of very deep networks, while DenseNet's densely connected blocks promote feature reuse and information flow. By combining these strengths, the hybrid model is motivated by the potential to achieve superior feature learning capabilities, leading to better emotion recognition.
- *Accuracy improvement:* FER demands a high level of accuracy, as recognizing subtle and nuanced facial expressions is a challenging task. The motivation behind the hybrid model is to boost recognition accuracy by leveraging the complementary features of ResNet and DenseNet. This is particularly important in applications such as mental health MH assessment, and lie detection, along with human-computer interaction, where precise emotion recognition is critical.
- *Efficiency and speed:* While accuracy is crucial, the efficiency and speed of FER models are equally significant, especially in real-time applications and resource-constrained environments. The hybrid model seeks to strike a balance between accuracy and computational efficiency, ensuring that it can be deployed in a wide range of applications without compromising real-time performance.
- *State-of-art performance:* As the field of FER evolves, there is a constant push to develop models that outperform existing methods. The motivation for the hybrid model is to be at the forefront of FER technology, offering a competitive and state-of-the-art solution that can set new benchmarks in terms of accuracy and versatility.

Section 1 has introduced the role of machine learning and deep learning in facial expression recognition along with the need for accuracy, efficiency, and speed enhancement. Section 2 is focused on a hybrid approach where ResNet and DenseNet are working in an integrated manner to perform facial expression recognition. Image acquisition, image compression, noise removal, training, and testing using a hybrid model have been discussed in this section. Section 3 presents the result and discussion part where a confusion matrix has been obtained after training and testing. Training, and testing accuracy in the case of conventional models such as CNN, DenseNet, and ResNet has been compared to the proposed hybrid approach. Section 4 presents the conclusion of the work and proposed work significance considering future scope.

2. MATERIALS AND METHODS

To develop a hybrid ResNet and DenseNet model for facial expression recognition there is a need to involve a combination of materials and methods from deep learning and computer vision. High-performance GPUs are essential for training deep neural networks efficiently. Common choices include NVIDIA GPUs such as the GeForce RTX series or Tesla GPUs. Deep learning frameworks like TensorFlow, PyTorch, and Keras are used to implement and train the model. Additionally, Python libraries for data preprocessing, such as OpenCV, are essential. The face recognition system makes use of the labeled faces in the wild (LFW) [29] dataset for training of deep learning models. The Labeled faces in the wild LFW dataset is a popular benchmark dataset for face recognition algorithms. It is comprised of over 13,000 tagged face photos gathered from various online sources.

The present research has undergone a preliminary pre-processing phase by compressing image size and then using a Gaussian noise reduction mechanism. The impact of noise has been measured to assess the picture's overall quality. Due to reduced size, the time consumption of training got reduced. The pictures are sent to a convolution neural network model, trained using a hybrid model illustrated in Figure 1. The goal of this simulation experiment is to evaluate the relative accuracy of several CNN models on simulated data sets that have been subjected to various filtering operations. The process described involves a multi-step approach for facial expression image analysis, compression, and noise filtering using a hybrid proposed model. Here's a summary of the process:

- *Input images:* The initial step involves acquiring facial expression images that are often used for applications like emotion recognition. These images may be captured in various settings and conditions.
- *Image compression:* The first part of the model involves image compression. The input images are passed through the DenseNet model, which is known for its feature extraction capabilities. DenseNet extracts relevant features from the images to reduce their dimensionality and encode important information.
- *Noise filtering:* After compression, the images are passed through the ResNet model. ResNet is renowned for its ability to handle deep neural networks and is often used for image recognition tasks. It further processes the compressed images, potentially reducing noise and enhancing the quality of the images.
- *Hybrid model:* The hybrid proposed model combines the strengths of DenseNet and ResNet architectures to capitalize on their complementary characteristics. Specifically, the noise-filtered images after compression serve as input to the hybrid model. The model architecture may involve parallel or

sequential integration of DenseNet and ResNet components, where features learned by each architecture fused or concatenated at certain stages of the network.

- *Feature extraction and fusion:* In the hybrid model, ResNet and DenseNet components extract hierarchical features from the noise-filtered images. These features capture different levels of abstraction, ranging from low to high-level details of semantic information related to facial expressions.
- *Training:* The hybrid model is trained using a suitable optimization algorithm and loss function, to minimize classification errors in facial expression recognition. Transfer learning technique utilized to initialize the model weights and fine-tuning procedures adapts the hybrid model to the specific characteristics of facial expression images, thereby enhancing its discriminative power.
- *Testing:* After training, the performance of the trained model is evaluated on a test set to assess its accuracy and generalization capability.

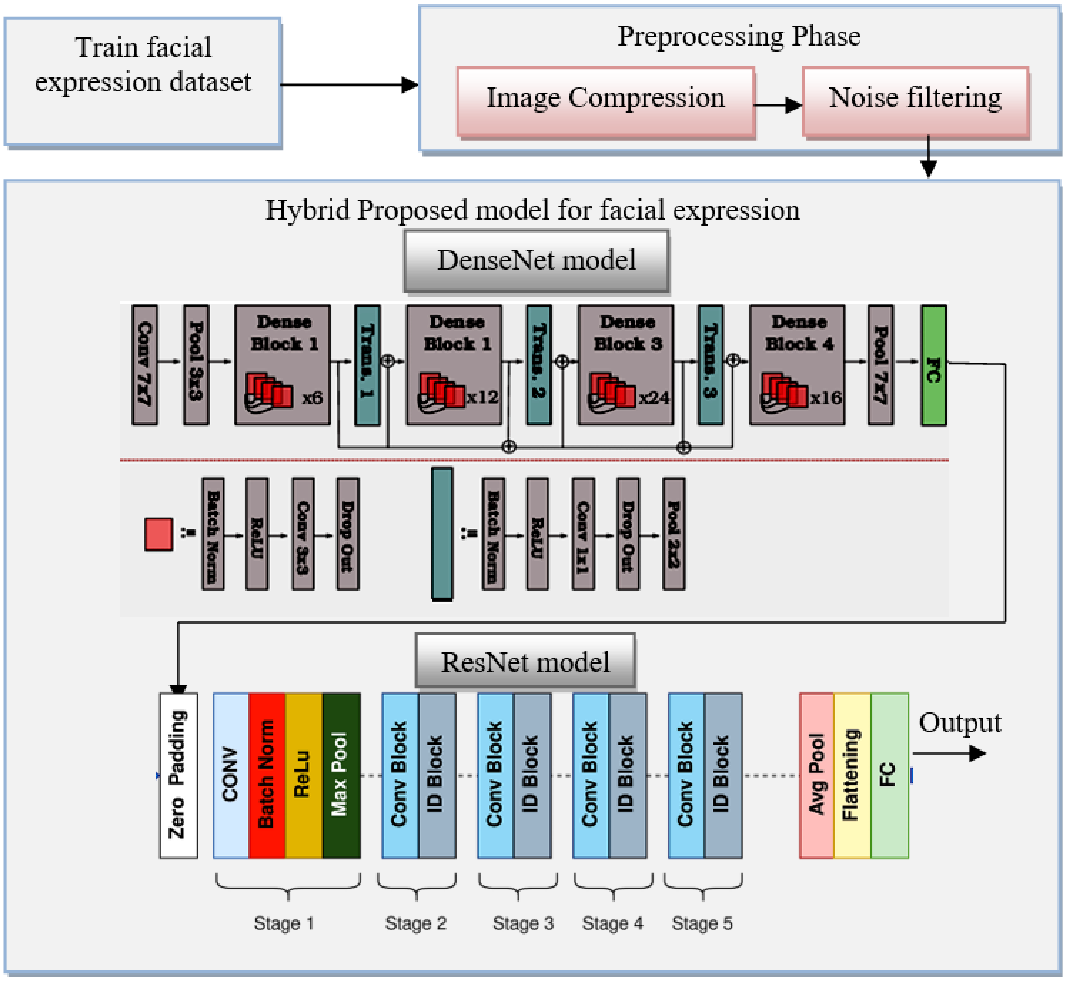


Figure 1. Proposed work of hybrid model of ResNet and DenseNet

The adopted methodology for testing the hybrid ResNet and DenseNet model has utilized both train-test split and cross-validation techniques to evaluate its performance on the given dataset. We have improved hybrid model performance via transfer learning. We used k-fold cross-validation to verify the model resilience across data subsets, assuring accurate performance measures. Before model training, the dataset is separated into training and test sets to avoid overfitting. 80% of data is allocated in the case of the training model, while the remaining 20% is reserved in the case of testing. The hybrid ResNet+DenseNet model is trained along with evaluated k times, with each fold serving as a test set once along with the remaining folds as training sets for 0,001 iterations. By averaging the results across multiple iterations, cross-validation helps to mitigate variability in model performance due to random partitioning of the dataset. We carefully modify hyperparameters such as epoch size 50, and batch size 32 respectively. Simulation is performed on Adam optimizer settings during training for optimum convergence. To assess model effectiveness, post-training assessment analyzes accuracy. Misclassifications are examined to enhance model design, hyperparameters, and preprocessing.

This comprehensive approach allows the construction of a complex face expression detection system that smoothly blends cutting-edge deep learning architectures, robust preprocessing approaches, and rigorous validation processes for higher performance and dependability. The hybrid proposed model for facial expression recognition leverages the capabilities of DenseNet and ResNet architectures while addressing the challenges posed by noise-filtered images after compression. By integrating complementary features learned from these architectures, the hybrid model aims to achieve robust and accurate recognition of facial expressions, thus facilitating applications in emotion detection, human-computer interaction, and affective computing.

3. RESULTS AND DISCUSSION

Firstly, seven different emotional states of anger, fear, happiness, disgust, sadness, surprise, along neutral were considered. It is worthy of praise that the hybrid model showed impressive performance in predicting emotional characteristics during the simulation. Throughout the simulation process in-depth analysis of the model's predictions in comparison to the ground truth labels is shown in the confusion matrix. The confusion matrix presents that the rate of accuracy has been increased whereas the rate of error has been reduced. Table 1 illustrates the confusion matrix for the hybrid proposed model, which has 1,095 correct predictions for anger faces, 1,262 correct values for disgust faces, 1,206 true values for fear face expressions, 1,210 correct predictions for happy faces, 1308 faces are detected as Sad faces, 1,412 true values with Surprise faces, and 1,122 correct values obtained for neutral facial expression. After simulation, the model considered 8,615 true positives out of 8,825.

Table 1. Prediction for hybrid proposed model

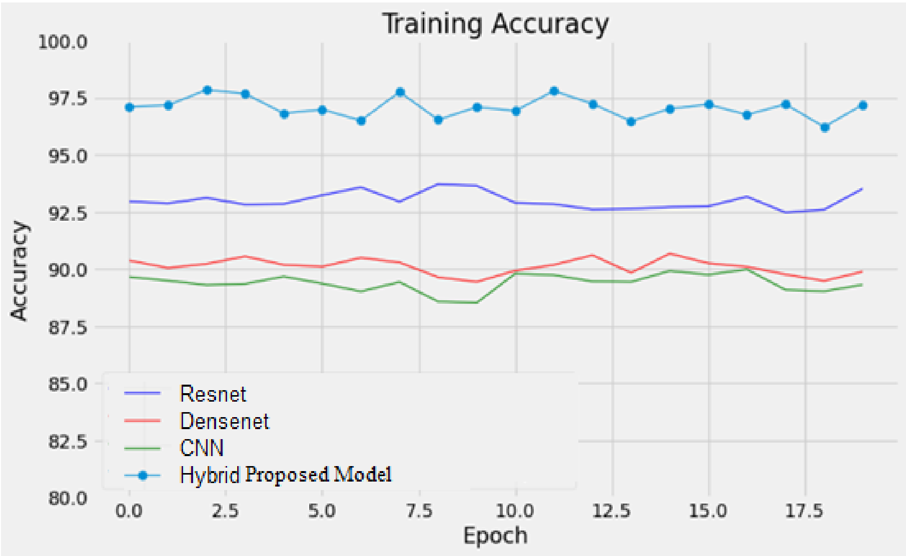
Confusion matrix								
		Predicted Label						
		Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
True Label	Anger	**1,095**	0	0	0	0	0	0
	Disgust	11	**1,262**	16	0	0	0	1
	Fear	0	2	**1,206**	0	0	0	0
	Happy	0	2	0	**1,210**	2	2	1
	Sad	7	1	78	0	**1,308**	0	0
	Surprise	1	11	44	3	27	**1,412**	0
	Neutral	0	0	0	0	1	0	**1,122**

Performance comparison of ResNet, DenseNet, traditional CNNs, and the proposed ResNet, and DenseNet architectures involves considering their respective strengths and trade-offs across various tasks and datasets. Traditional CNNs form the basis for subsequent advancements in convolutional neural network architectures. While their performance can be task-dependent, they may struggle with training very deep networks due to issues like vanishing gradients. In practice, the choice between ResNet, DenseNet, traditional CNNs, or hybrid architectures depends on the specifics of the application, dataset characteristics, and the available computational resources. Performance evaluations are often conducted through empirical experiments on the target dataset, and the selection of the most suitable architecture may involve careful consideration of these factors in Table 2.

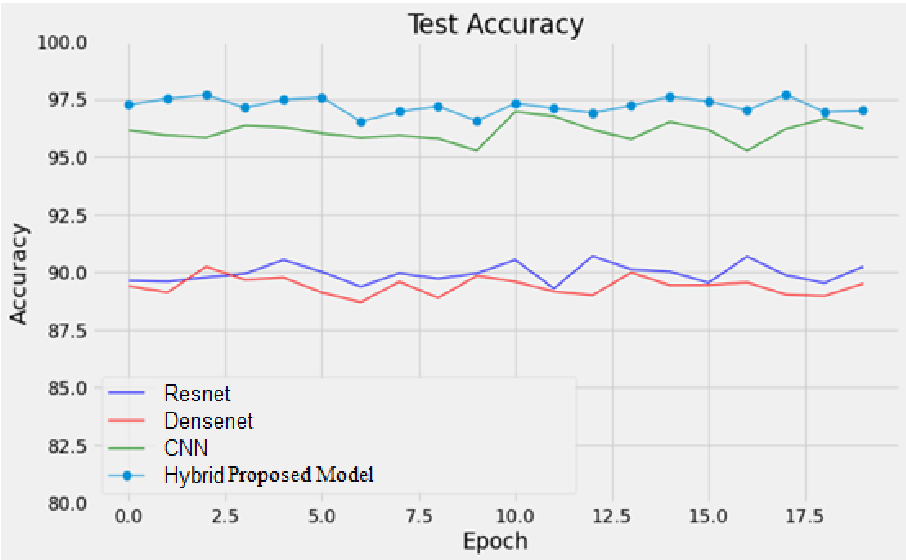
In the performance comparison, four models have been trained and tested considering the dataset of labeled faces in the wild illustrated in Figure 2 during training accuracy in sub-Figure 2(a) and testing accuracy in sub-Figure 2(b). These models are ResNet, DenseNet, CNN, and the Hybrid proposed model. Several key metrics were considered, providing insights into their effectiveness for classifying emotional traits. ResNet exhibited a notable advantage in overall accuracy, showcasing its capability to correctly predict emotional states. However, this accuracy came at the cost of a slightly higher computational burden, as indicated by a longer training time. The error rate for ResNet is comparatively low, signifying a high precision in its predictions. The true positive counts for emotional traits are commendable, reinforcing its proficiency in capturing instances of specific emotions. It is observed that testing and training accuracy of the Hybrid proposed model is maximum whereas the conventional model yields yielding least accuracy during training and testing.

Table 2. Comparative analysis of models

ResNet [2], [3], [14], [15], [22]-[24], [27]	DenseNet [22], [23], [25]-[27]	Hybrid proposed work
- ResNet, with its introduction of residual connections, has addressed the challenges associated with training deep networks. - The use of skip connections facilitates the training of extremely deep architectures, allowing ResNet to achieve state-of-the-art performance in various image classification tasks. - The simplicity and effectiveness of ResNet's architecture have made it a widely adopted and influential model in the field.	- DenseNet, on the other hand, takes a different approach by introducing dense connections between layers. - This architecture promotes feature reuse and helps mitigate the vanishing gradient problem. DenseNet often requires fewer parameters compared to traditional architectures, leading to more parameter-efficient models. - The ability to connect each layer to all preceding layers contributes to its competitive performance on image classification tasks.	- Hybrid architectures that combine elements of ResNet and DenseNet have also been explored to leverage the strengths of both. - Such hybrids may use residual connections in some parts of the network and dense connections in others, offering a flexible approach depending on specific requirements of the task. - The design of hybrid architecture depends on factors such as computational resources, dataset size, along desired trade-offs between model complexity and performance.



(a)



(b)

Figure 2. Comparative analysis of accuracy for different models (a) training accuracy (b) testing accuracy

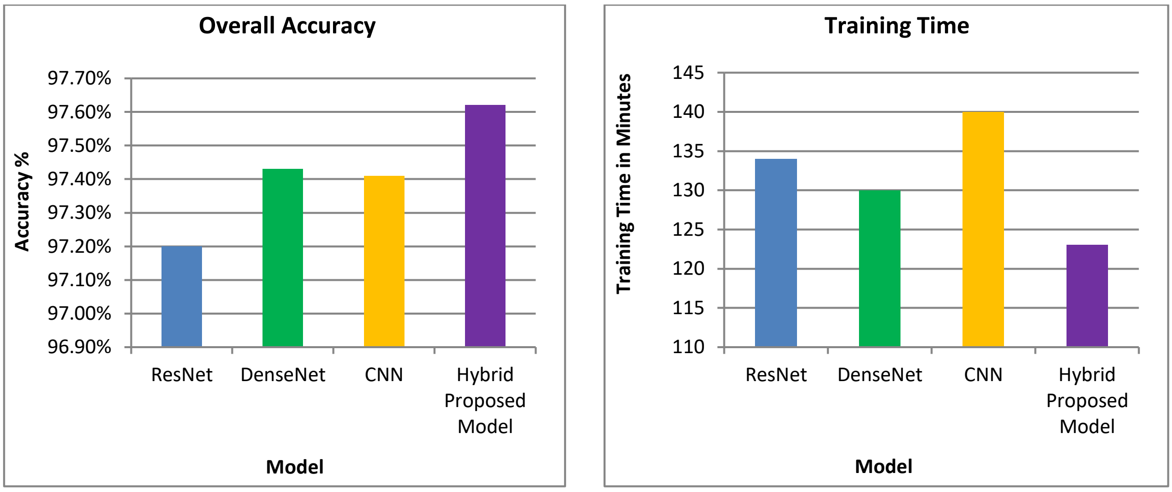
Table 3 presents the comparative analysis of model performance. Each model exhibited strengths and weaknesses across different performance metrics. DenseNet stood out for its high overall accuracy, while ResNet offered a good balance between accuracy and training efficiency. The traditional CNN model provided a reliable baseline, and the hybrid proposed model capitalized on the strengths of multiple architectures. The choice of the best model may depend on specific requirements such as computational resources, training time constraints, and the desired trade-off between accuracy and efficiency. The overall accuracy, error rate, and training time obtained for the ResNet model is 97.20%, and 2.80%, in 134 minutes as these are 97.43%, and 2.57% in 130 minutes for DenseNet respectively. On the other hand, the overall accuracy, error rate, and training time for the CNN model are 97.11%, 2.89%, and 140 minutes as these are 97.62%, 2.38%, and 123 minutes for the hybrid proposed model respectively. DenseNet excelled in overall accuracy, DenseNet provided a balance between accuracy and training efficiency, the traditional CNN model offered a reliable baseline, and the hybrid proposed model leveraged a combination of strengths to deliver competitive performance across key metrics.

Table 3. Comparative analysis of models performance

Models	Overall Accuracy	Error Rate	Training Time	Correct Prediction
ResNet [27]	97.20%	2.80%	134	8626
DenseNet [24]	97.43%	2.57%	130	8599
CNN [22]	97.11%	2.89%	140	8596
Hybrid Proposed Model	97.62%	2.38%	123	8615

Figure 3, is a graphical representation illustrating a comparison of various parameters, including overall accuracy and training time in sub-Figures 3(a), and 3(b) across different models such as ResNet, DenseNet, CNN, and the hybrid proposed model. The superiority of the hybrid proposed model suggests that its combination of features from ResNet, DenseNet, and potentially other techniques results in enhanced performance compared to standalone architectures like ResNet, DenseNet, and CNN. The overall accuracy figure shows accuracy in the case of reset DenseNet, CNN, and the proposed model where it is observed that the accuracy of the hybrid proposed work is higher than conventional models. On the other hand, training time chart presents the performance and it is found that the training time of the hybrid proposed model is higher than that of conventional models.

The improved accuracy implies that the hybrid proposed model effectively captures both local and global features, leveraging the strengths of ResNet's skip connections and DenseNet's dense connectivity. Moreover, the reduced training time highlights the model's efficiency, indicating faster convergence during the training phase. Overall, Figure 3 provides compelling evidence that the hybrid proposed model offers a superior solution for facial expression recognition compared to existing architectures, thereby showcasing its potential for practical deployment in real-world applications.



(a) (b)

Figure 3. Comparison of performance parameters for different models where (a) overall accuracy (b) training time

4. CONCLUSION AND FUTURE WORK

Present research mainly focuses on preprocessing and a hybrid proposed model to reduce the size of images with fine quality so that the time consumption of training is reduced. The hybrid model's efficacy in recognizing a wide range of facial expressions, from basic emotions to nuanced expressions, is explored. The overall accuracy for the hybrid proposed model is 97.62%. Moreover, the accuracy of the proposed work is higher than conventional. Simulation results conclude that hybrid proposed model is performing fast as compared to conventional CNN, ResNet, along DenseNet models. Sometimes, the system is confused with fear and sad facial expressions. In the future, research directions may include further exploration of hybrid architectures and optimization techniques to push the boundaries of model performance and efficiency. The evaluation can be performed with various datasets to analyze the robustness of the hybrid model. Further, an improved approach can be added to classify the emotions that will have helpful potential for practical deployment in real-world applications.

REFERENCES

[1] F. Ozdamli, A. Aljarrah, D. Karagozlu, and M. Ababneh, "Facial recognition system to detect student emotions and cheating in distance learning," *Sustainability*, vol. 14, no. 20, p. 13230, Oct. 2022, doi: 10.3390/su142013230.

[2] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications," *Neural Computing and Applications*, vol. 35, no. 32, pp. 23311–23328, Apr. 2021, doi: 10.1007/s00521-021-06012-8.

[3] Y. Pratama, L. M. Ginting, E. H. Laurencia Nainggolan, and A. E. Rismanda, "Face recognition for presence system by using residual networks-50 architecture," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5488-5496, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5488-5496.

[4] Y. Said and M. Barr, "Human emotion recognition based on facial expressions via deep learning on high-resolution images," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 25241–25253, Apr. 2021, doi: 10.1007/s11042-021-10918-9.

[5] N. Christou and N. Kanojiya, "Human facial expression recognition with convolution neural networks," in *Third International Congress on Information and Communication Technology*, Springer Singapore, 2018, pp. 539–545.

[6] D. Ammous, A. Chabbouh, A. Edhib, A. Chaari, F. Kammoun, and N. Masmoudi, "Designing an efficient system for emotion recognition using CNN," *Journal of Electrical and Computer Engineering*, vol. 2023, pp. 1–11, Sep. 2023, doi: 10.1155/2023/9351345.

[7] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[8] M. Z. Khan, S. Harous, S. U. Hassan, M. U. Ghani Khan, R. Iqbal, and S. Mumtaz, "Deep unified model for face recognition based on convolution neural network and edge computing," *IEEE Access*, vol. 7, pp. 72622–72633, 2019, doi: 10.1109/ACCESS.2019.2918275.

[9] Y. El Madmoune, I. El Ouariachi, K. Zenkouar, and A. Zahi, "Robust face recognition using convolutional neural networks combined with Krawtchouk moments," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4052-4067, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4052-4067.

[10] A.-Q. Bi, X.-Y. Tian, S.-H. Wang, and Y.-D. Zhang, "Dynamic Transfer exemplar based facial emotion recognition model toward online video," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 2s, pp. 1–17, Jun. 2022, doi: 10.1145/3538385.

[11] A. Litvin, K. Nasrollahi, S. Escalera, C. Ozcinar, T. B. Moeslund, and G. Anbarjafari, "A novel deep network architecture for reconstructing RGB facial images from thermal for face recognition," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 25259–25271, May 2019, doi: 10.1007/s11042-019-7667-4.

[12] J. Tang, Q. Su, B. Su, S. Fong, W. Cao, and X. Gong, "Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition," *Computer Methods and Programs in Biomedicine*, vol. 197, Art. no. 105622, Dec. 2020, doi: 10.1016/j.cmpb.2020.105622.

[13] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "A unified deep model for joint facial expression recognition, face synthesis, and face alignment," *IEEE Transactions on Image Processing*, vol. 29, pp. 6574–6589, 2020, doi: 10.1109/tip.2020.2991549.

[14] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More trainable inception-ResNet for face recognition," *Neurocomputing*, vol. 411, pp. 9–19, Oct. 2020, doi: 10.1016/j.neucom.2020.05.022.

[15] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 57–64, Jun. 2021, doi: 10.1016/j.ijcce.2021.02.002.

[16] C. Zhu, "Real-time monitoring and assessment system with facial landmark estimation for emotional recognition in work," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 8, pp. 1–13, Sep. 2023, doi: 10.17762/ijritcc.v11i8.7737.

[17] G. Sunitha, K. Geetha, S. Neelakandan, A. K. S. Pundir, S. Hemalatha, and V. Kumar, "Intelligent deep learning based ethnicity recognition and classification using facial images," *Image and Vision Computing*, vol. 121, p. 104404, May 2022, doi: 10.1016/j.imavis.2022.104404.

[18] X. Li, S. Lai, and X. Qian, "DBCFace: towards pure convolutional neural network face detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1792–1804, Apr. 2022, doi: 10.1109/tcsvt.2021.3082635.

[19] Smitha, P. S. Hegde, and Afshin, "Face recognition based attendance management system," *International Journal of Engineering Research and*, vol. V9, no. 05, Jun. 2020, doi: 10.17577/ijertv9is050861.

[20] H. Benradi, A. Chater, and A. Lasfar, "A hybrid approach for face recognition using a convolutional neural network combined with feature extraction techniques," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 2, pp. 627-640, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp627-640.

[21] O. Khajuria, R. Kumar, and M. Gupta, "Facial emotion recognition using CNN and VGG-16," *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023, pp. 472-477, doi: 10.1109/icict57646.2023.10133972.

[22] M. J. Khan, M. J. Khan, A. M. Siddiqui, and K. Khurshid, "An automated and efficient convolutional architecture for disguise-invariant face recognition using noise-based data augmentation and deep transfer learning," *The Visual Computer*, vol. 38, no. 2, pp. 509–523, Jan. 2021, doi: 10.1007/s00371-020-02031-z.

[23] M. Alonazi, H. J. Alshahrani, F. A. Alotaibi, M. Maray, M. Alghamdi, and A. Sayed, "Automated facial emotion recognition using the pelican optimization algorithm with a deep convolutional neural network," *Electronics*, vol. 12, no. 22, Art. no. 4608, Nov. 2023, doi: 10.3390/electronics12224608.

[24] S. Anwarul, T. Choudhury, and S. Dahiya, "A novel hybrid ensemble convolutional neural network for face recognition by optimizing hyperparameters," *Nonlinear Engineering*, vol. 12, no. 1, Jan. 2023, doi: 10.1515/nleng-2022-0290.

[25] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Applied Sciences*, vol. 9, no. 20, Oct. 2019, doi: 10.3390/app9204397.

[26] Kavita and R. S. Chhillar, "Innovative integration of convolutional neural networks for enhanced face recognition," *Journal of Electrical Systems*, vol. 20, no. 3s, pp. 1941–1950, Mar. 2024, doi: 10.52783/jes.1740.

[27] Q. Wang and G. Guo, "Benchmarking deep learning techniques for face recognition," *Journal of Visual Communication and Image Representation*, vol. 65, p. 102663, Dec. 2019, doi: 10.1016/j.jvcir.2019.102663.

[28] G. Meena, K. K. Mohbey, A. Indian, M. Z. Khan, and S. Kumar, "Identifying emotions from facial expressions using a deep convolutional neural network-based approach," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 15711–15732, Jul. 2023, doi: 10.1007/s11042-023-16174-3.

[29] A. Anand, "LFW - People (Face Recognition)," *Kaggle.Com*, 2020. https://www.kaggle.com/datasets/atulanandjha/lfwpeople (accessed Nov. 15, 2020).

BIOGRAPHIES OF AUTHORS

Kavita is a research scholar, pursuing a Ph.D. in computer science from the Department of Computer Science and Applications at Maharshi Dayanand University, Rohtak, India. She is working as an assistant professor in the CSE Department of Vaish College of Engineering, M.D. University, Rohtak, India. She received an M.Tech. degree in computer science and engineering from Guru Jambheshwar University, Hisar, India. Her research interests include image processing, face recognition, machine learning, and deep learning. She has work experience of more than 15 years. She is a member of the International Association of Engineers (IAENG) She can be contacted at email: kavitavce@gmail.com.

Rajender Singh Chhillar is a professor and former head of the Department of Computer Science and Applications at Maharshi Dayanand University, Rohtak, India. He obtained his Ph.D. in computer science from M.D. University, Rohtak, India, and master's degree from Kurukshetra University, Kurukshetra, India. Additionally, he holds his MBA degree from Sikkim Manipal University, Sikkim, India. His research interests include software engineering, software testing, software metrics, web metrics, biometrics, data mining, computer networking, software design, image processing, artificial intelligence, machine learning, deep learning, and IoT. He has more than 35 years of experience. He has published more than 100 journal and 65 conference papers in recent years and has also written two books in the fields of software engineering and information technology. He is a director of the CMAI Asia Association in New Delhi and a senior member of the IACSIT in Singapore. Furthermore, he is a member of the Computer Society of India. He can be contacted at email: chhillar02@gmail.com.