












Category: STEM (Science, Technology, Engineering and Mathematics)

ORIGINAL

Emotion Recognition with a Hybrid VGG-ResNet Deep Learning Model: a Novel Approach for Robust Emotion Classification

Reconocimiento de Emociones con un Modelo Híbrido de Aprendizaje Profundo VGG-ResNet: un Enfoque Novedoso para una Clasificación Sólida de las Emociones

N Karthikeyan¹  , K Madheswari¹  , Hrithik Umesh¹  , Rajkumar N²  , Viji C²  

¹Department of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India.

²Department of Computer Science and Engineering, Alliance College of Engineering and Technology, Alliance University, Bangalore, India.

Cite as: Karthikeyan N, Madheswari K, Umesh H, N R, C V. Enhancing Deep Learning for Autism Spectrum Disorder Detection with Dual-Encoder GAN-based Augmentation of Electroencephalogram Data. Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:960. <https://doi.org/10.56294/sctconf2024960>

Submitted: 15-02-2024

Revised: 05-05-2024

Accepted: 04-07-2024

Published: 05-07-2024

Editor: Dr. William Castillo-González 

ABSTRACT

The recognition and interpretation of human emotions are crucial for various applications such as education, healthcare, and human-computer interactions. Effective emotion recognition can significantly enhance user experience and response accuracy in these fields. This research aims to develop a robust emotion recognition system by integrating VGG and ResNet architectures to improve the identification of subtle variations in facial expressions. This paper proposes a hybrid deep learning approach using a combination of VGG and ResNet models. This system incorporates multiple convolutional and pooling layers along with residual blocks to capture intricate patterns in facial expressions. The FER2013 dataset was employed to train and evaluate the model's performance. Comparative analysis was conducted against other models, including VGG16, DenseNet, and MobileNet. The hybrid model demonstrated superior performance, achieving a training accuracy of 99,80 % and a validation accuracy of 66,17 %. In contrast, the VGG16, DenseNet, and MobileNet models recorded training accuracies of 54,27 %, 68,51 %, and 84,68 %, and validation accuracies of 46,58 %, 56,11 %, and 60,35 %, respectively. The proposed hybrid approach effectively enhances emotion recognition capabilities by leveraging the strengths of VGG and ResNet architectures. This method outperforms existing models, offering a significant improvement in both training and validation accuracies for emotion recognition systems.

Keywords: Emotion Detection; Image Classification; Deep Learning; CNN; Densenet; Mobilenet; VGG16; Resnet; Hybrid Model.

RESUMEN

Reconocer y comprender las emociones humanas tiene consecuencias importantes para diferentes aplicaciones, incluida la educación, la atención médica y las interacciones entre personas y computadoras. Este artículo propone un esquema híbrido basado en VGG y ResNet para identificar patrones intrincados y variaciones sutiles entre expresiones faciales utilizando muchas capas convolucionales y de agrupación, al mismo tiempo que agrega bloques residuales dentro de cada capa. Esta investigación tiene como objetivo aprovechar técnicas de aprendizaje profundo y enfoques híbridos para construir modelos robustos y confiables, que apuntan a mejorar el rendimiento de futuros sistemas de reconocimiento de emociones. Aprovechando el conjunto de datos FER2013, entrenamos y comparamos el rendimiento con diferentes modelos de aprendizaje profundo como VGG16, DenseNet y MobileNet. El enfoque propuesto superó a los modelos existentes con una precisión de entrenamiento del 99,80 %, junto con una precisión de validación del 66,17 %. Los otros modelos de comparación, como VGG16, DenseNet y MobileNet, lograron precisiones de entrenamiento del 54,27 %, 68,51 % y 84,68 %, junto con precisiones de validación del 46,58 %, 56,11 % y 60,35 % respectivamente.

Palabras clave: Detección de Emociones; Clasificación de Imágenes; Aprendizaje Profundo; CNN; Densenet; Mobilenet; VGG16; Resnet; Modelo Híbrido.

INTRODUCTION

Emotions are intrinsic to human experience, influencing perception, decision-making and interpersonal interactions. Their significance spans a multitude of disciplines including psychology, sociology, neuroscience, and artificial intelligence (AI). Recognizing and accurately interpreting emotions is pivotal for enhancing the understanding of human behaviour and well-being. The development of emotion recognition systems plays a crucial role in creating human-centric solutions capable of detecting and responding to the human emotional states effectively. In contemporary society, emotions can be expressed through various channels such as facial expressions⁽¹⁾, gestures⁽²⁾ and text.⁽³⁾ Among these, the facial expression analysis is the most extensive study offers rich and immediate insights into a person's emotional state. Research has identified a range of fundamental emotions includes happiness, sadness, fear, surprise, disgust, and anger and neutral expressions. The field of emotion recognition has seen rapid growth in recent years leading to the establishment of numerous companies and significant investment in the development of technologies aimed at identifying and interpreting the emotions. Emotion detection systems provide the foundation for intuitive and responsive interfaces that improve human-computer interactions.⁽⁴⁾ The applications of emotion detection systems are encompassing areas such as customer experience and marketing, healthcare, adaptive learning, security and surveillance, driver and attention monitoring, gaming and entertainment and in research fields like psychology and sociology.⁽⁵⁾ Each of these fields benefits from the system's capacity to effectively interpret human emotions and respond appropriately. Deep learning an instance of machine learning, makes use of multi-layered neural networks to improve prediction and decision-making processes. Deep learning is distinguished by the depth of its neural networks, which include several hidden layers that enable the model to learn complicated patterns and representations from input data. Deep learning architectures such as Convolutional Neural Networks (CNNs), automatically extract hierarchical features, enabling more abstract and sophisticated learning compared to traditional machine learning, where feature engineering is done manually. Deep learning's adaptability has led to its use in a wide range of fields, including medical picture classification and segmentation, cybersecurity, and the detection of Distributed Denial-of-Service (DDoS) attacks, among others.⁽⁶⁾ Deep learning improvements have significantly aided emotion detection, introducing highly effective models for recognizing and comprehending human emotions across a wide range of input modalities, including video, audio, and text.⁽⁷⁾

CNNs are instrumental in identifying specific features such as edges, textures, and facial attributes critical for facial expression recognition. Recurrent Neural Networks (RNNs) on the other hand it is well-suited for applications containing sequential inputs such as speech or text, because they can detect temporal patterns and grasp sentiment over time. Transfer learning is frequently employed to adapt existing models trained on large datasets to new tasks that enhances their precision and versatility in various applications. The focus of this research is to develop a precise and efficient classifier for emotions, leveraging deep learning technologies that excel in image recognition tasks. Traditional approaches frequently fall short of capturing the complexities of human emotions, necessitating new ways for categorizing and understanding these subtle signs. The aim is to create a more comfortable and intuitive interaction with AI, fostering a system that can effortlessly integrate with human emotional dynamics. Several challenges need to be addressed in this endeavour. One of the primary obstacles is identifying a dataset that is both diverse and relevant for image processing technique capable of capturing the intricate shapes and patterns required for deep feature extraction. The issue of class imbalance are certain emotions that underrepresented in the data through poses of significant challenge, as it can lead to biased model performance. Ensuring the model does not overfit to the training data while maintaining its ability to generalize to unseen data is crucial for achieving robust performance. Real-time processing handles truncated expressions and accommodating cultural variations in emotional expression are challenges specific to online communication that need to be considered. Addressing these issues requires exploring various deep learning architectures, such as CNN, ResNet, VGG16, DenseNet, and MobileNet, to determine the most effective approach for emotion classification. Enhancing the model's ability to accurately predict previously unknown facial expressions and developing techniques to manage overlapping emotions are key objectives. It is essential to implement strategies to mitigate biases, ensure fairness and address ethical concerns related to privacy, consent, and the system's impact on individuals and communities. By tackling these challenges, this research aims to contribute to the development of more reliable and ethical emotion recognition systems that can significantly improve human-AI interactions.

Related Work

Sarvakar, Ketan, et al.⁽⁸⁾ discussed the challenges in emotion classification using deep neural networks, emphasizing the need for high accuracy and reliability. They employed a Convolutional Neural Network (CNN) with six convolutional layers, two max-pooling layers, and two fully connected layers, achieving 60 % accuracy, highlighting the limitations in recognizing specific emotions due to data insufficiency. Zahara, Lulfiah, et al.⁽⁹⁾ focused on emotion recognition using deep learning models on the Raspberry Pi, achieving 65,97 % accuracy with a CNN on the FER2013 dataset. They emphasized the need for advanced CNN architectures,

more extensive training data, and better hardware for improved emotion recognition performance. Jaiswal, Akriti, et al.⁽¹⁰⁾ developed a CNN-based model for emotion recognition that excelled in feature extraction and emotion classification, achieving 70,14 % accuracy on the FER2013 dataset and 98,65 % on the JAFFE dataset. This demonstrated superior accuracy and efficiency over traditional methods. Ali, Khatun, and Turzo et al.⁽¹¹⁾ explored facial emotion recognition using neural networks and the Viola-Jones algorithm. Their model, tested on a Kaggle dataset, achieved higher accuracy than human visual systems by employing decision trees and deep learning for emotion classification, highlighting the importance of powerful databases in AI development.

Mukhopadhyay et al.⁽¹²⁾ analyzed emotions in online classes by using a CNN model on the FER2013 dataset. They identified common emotions like dissatisfaction and happiness, achieving 65 % accuracy but struggled with dataset imbalance and variations in lighting and camera angles. Gaddam et al.⁽¹³⁾ utilized the ResNet50 architecture for facial emotion detection on the FER2013 dataset, achieving 75,45 % training accuracy and 54,56 % test accuracy. They emphasized the need for larger and balanced datasets to improve the precision of emotion detection models. Prasad and Chandana presented a DenseNet-based approach for thermal face emotion recognition, achieving 95,97 % accuracy on the RGB-D-T dataset. Their model outperformed existing models like SSD and YOLOv3, suggesting further refinement for improved performance. Zhang et al.⁽¹⁴⁾ proposed the ERLDK model for real-time emotion recognition in conversational contexts using reinforcement learning and domain knowledge. Their model excelled in recognizing emotions during dialogues, outperforming existing methods on challenging datasets. Singh et al.⁽¹⁵⁾ developed a CNN-based facial emotion recognition system, achieving 61,7 % accuracy on the FER2013 dataset without pre-processing steps. They highlighted the potential of dropout layers for improving model performance and suggested future research on data preprocessing and network optimization.

Rakshith et al.⁽¹⁶⁾ designed the ConvNet-3 model for emotion recognition, achieving 88 % training accuracy and 61 % validation accuracy. Their study emphasized the need for improved generalization and highlighted the challenges of overfitting when using smaller datasets like CK+48.

Irmak et al.⁽¹⁷⁾ explored emotion analysis from facial expressions using CNNs, achieving 70,62 % training accuracy on the FER2013 dataset. They emphasized the need for diverse datasets and called for further research to enhance bias awareness and model efficiency. Huang et al.⁽¹⁸⁾ study on facial emotion recognition using CNNs focused on three FER datasets, achieving 63,10 % accuracy on the FER2013 dataset. They discussed challenges like label errors and data variability, highlighting the need for better data quality and algorithm accuracy. The research on emotion recognition using various methods like CNN, DenseNet, VGG16, and MobileNet highlights the progress in achieving accurate and reliable emotion analysis tools. These advancements have broad applications in safety, human-computer interaction, healthcare, and psychology, indicating the potential for further development in this field.

METHOD

Study Design

This research leverages advanced deep learning techniques to develop an effective Facial Expression Recognition (FER) system. The research involves implementing and comparing three established architectures—DenseNet, MobileNet, and VGG16—with our proposed hybrid model, and model diagram is shown in figure 1. Our goal is to improve FER technology for applications such as human-computer interaction, affective computing, and mental health monitoring.

Research Location and Timeframe

The study was conducted at the Department of Computer Science, VIT University, Chennai from January 2023 to April 2024.

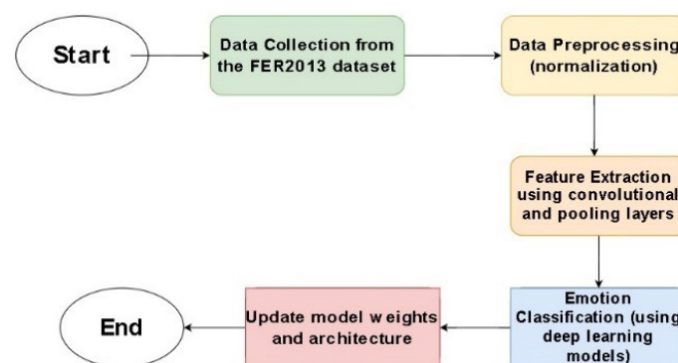


Figure 1. Flow diagram of workflow research

Data Collection and Preprocessing

We utilized the FER2013 dataset from Kaggle, which contains 28,729 training images and 7,178 test images across seven emotion categories: happy, sad, neutral, surprise, anger, fear, and disgust. Each greyscale image was standardized to 48,48 pixels. The dataset was normalized to reduce variations and enhance feature extraction. This preprocessing stage minimized noise and optimized training performance. The figure 2 shows sample images from different random classes in the dataset.



Figure 2. Sample images of the FER2013 dataset

Feature Extraction Methods

The next phase after data preprocessing is feature extraction. It involves using convolutional and pooling layers, along with the use of activation functions on the pre-processed images, to extract features like edges, textures, and facial landmarks. In this research, we make use of some specific methods to optimize the process of feature extraction as follows:

- Transfer Learning: we utilized pre-trained models (DenseNet, MobileNet, and VGG16) on ImageNet to leverage existing feature extraction capabilities for our FER task.
- Activation Function (ReLU): the ReLU function introduced non-linearity, expediting training and addressing the vanishing gradient problem.
- Global Average Pooling (GAP): this technique reduced feature map dimensions while retaining critical information, mitigating overfitting.
- Max Pooling: max pooling reduced spatial dimensions, retaining essential characteristics and improving computational efficiency.
- SoftMax: the SoftMax function converted raw output into probabilities for each category, facilitating classification.

Proposed model architecture: VGG-ResNet

The proposed model architecture combines VGG and ResNet to construct a deep convolutional neural network suitable for image classification applications. The process begins with convolutional layers, like VGG networks⁽¹⁹⁾ and then includes batch normalization and ReLU⁽²⁰⁾ activation to generate features from input images. Furthermore, following every convolutional layer, a residual block comprising two convolutional layers with batch normalization and ReLU activations, is inserted. Additionally, there is an immediate link that connects input directly to the output without passing through the layers. This design trains deeper networks by eliminating the vanishing gradient problem and enabling the learning of arbitrary and high-degree characteristics. Convolutional layers are accompanied by max-pooling layers, which minimize the spatial dimensions of feature maps while retaining only the necessary details. Convolutions and pooling compress feature maps, which are then transmitted through fully linked layers with ReLU activation for classification. The output layer includes a SoftMax function, which generates probabilistic distributions for the classes. This architecture is designed to leverage the strong points of the VGG and ResNet architectures, thus resulting in a strong model, which can

recognize complicated patterns and obtain high accuracy on image recognition problems. Figure 3 shows the architecture of our suggested hybrid model, providing an overview of how it might function in this instance.

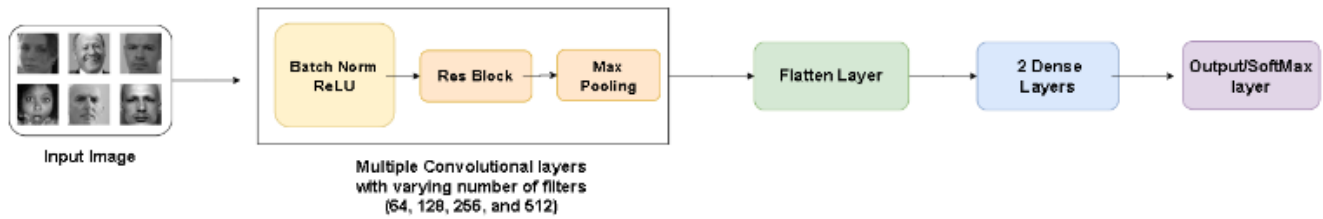


Figure 3. Model architecture of our proposed VGG-ResNet hybrid

To showcase the efficiency of our proposed hybrid model, we compare its performance with other existing deep learning architectures in the experimental part.

Performance Metrics

The standard evaluation tools are essential for evaluating how well the models have performed, as follows.

Accuracy

In classification models, accuracy is the percentage of properly identified samples compared to the overall dataset. It's calculated with the following formula:

$$Accuracy = \frac{\text{number of correctly classified samples}}{\text{total number of samples}} \quad (1)$$

Loss

Loss measures the difference between the actual probabilities of classes and predictions made. In most multi-class classification problems, the loss function utilized is categorical cross-entropy and to alter weights during model training as shown in equation 2.

$$Loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (2)$$

In this equation, N represents the total number of samples, M represents the number of classes, y_{ij} indicates if class_j is the proper classification for samples (1 if true, 0 otherwise), and p_{ij} represents the anticipated probability of samples belonging to class_j based on the model. Model training involves monitoring and recording accuracy and loss after each epoch to assess performance and convergence.

Confusion Matrices

These matrices visually represent the model's accuracy and errors across different classes, highlighting areas for improvement.

RESULTS

This research used the Kaggle platform and the GPU T4 x2 accelerator to train deep learning models for multi-class categorization. To optimize the process, we used the Adam optimizer and category cross-entropy loss function. To ensure convergence, each model is trained with a batch size of 64 and a learning rate of 0,0001. The major evaluation parameter was accuracy, and confusion matrices were used to analyze the models' performance across different emotion categories. VGG16, MobileNet, and DenseNet models are used to test and find the emotion recognition and categorization. The VGG16 network, were noticed due to its simplicity and depth, producing a certain amount of uncertainty. During training, it obtained accuracy of 54,27 %, while the loss is 1,2267, which is quite high the function was to detect image patterns that spread light on it. Yet, it lagged in the validation phase, achieved 46,58 % of accuracy and a loss of 1,4278. The training phase went smoothly by modeling MobileNet.^(21, 22) With an accuracy of 84,68 % and a loss of 0,4442. The validation procedure had further verified MobileNet performance because of a 60,35 % validation accuracy and a 1,2353 validation loss. DenseNet⁽²³⁾ exhibited competitive performance. Although DenseNet had training accuracy of 68,42 % and loss of 0,8706, it succeeded in having low training loss and high training accuracy. On top of that, the model also had problems during validation so that the validation accuracy was low (56,11 %), as well as the

validation loss was high, which proved the model’s low generalizability for unseen samples. Figures 4-6 display the accuracy and loss graphs of each model, and figure 7 depicts the confusion matrices of all the models used.

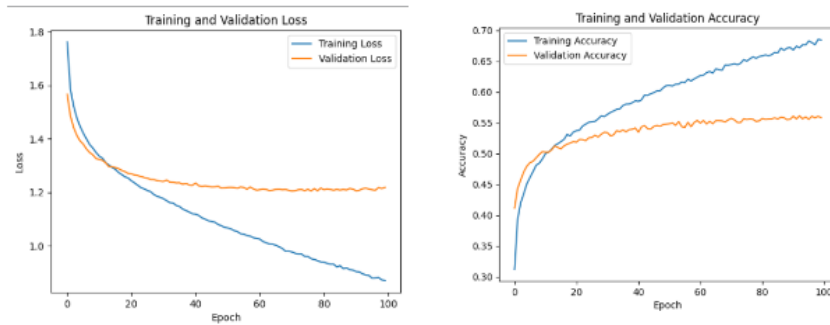


Figure 4. Performance graph of DenseNet

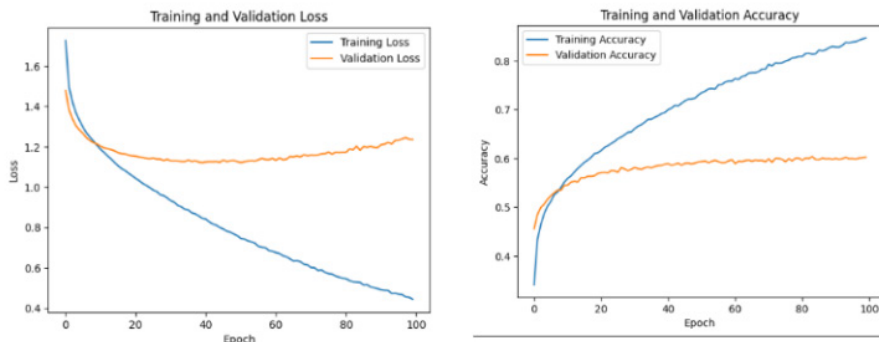


Figure 5. Performance graph of MobileNet

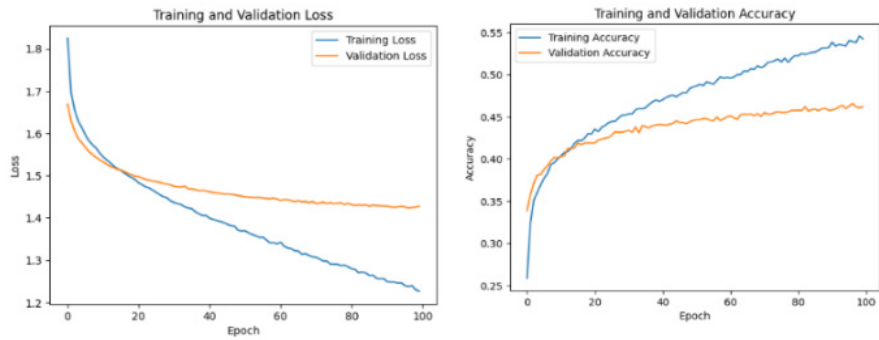


Figure 6. Performance graph of VGG16

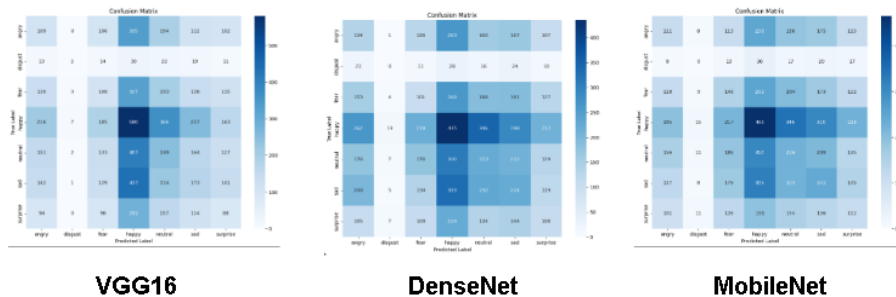


Figure 7. Confusion matrices of the three models

The proposed hybrid VGG-ResNet architecture delivered extraordinary training accuracy figure of 99,80 %. A training loss of 0,0046, which indicates that it has a higher effectiveness in minimizing the errors within the training process is revealed. Upon validation, the hybrid model achieved an accuracy of 66,17 %, along with a loss of 1,3828. With this observation, our proposed model can distinguish between different facial expressions and correctly classify emotions. Figures 8-10 depict the accuracy and loss graphs of the proposed model, as well as the confusion matrix of the model.

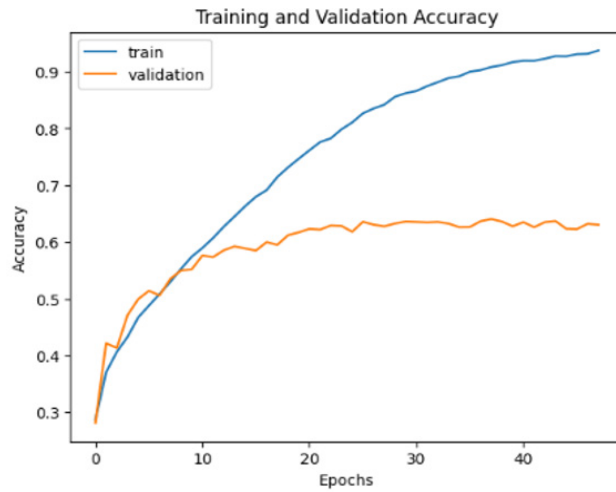


Figure 8. Accuracy graph of VGG+ResNet

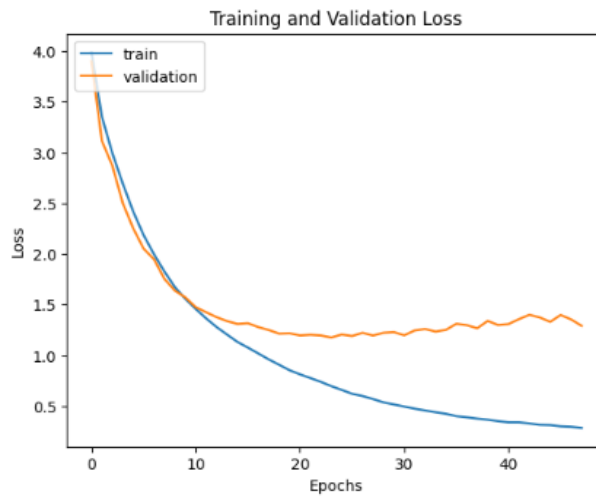


Figure 9. Loss graph of VGG+ResNet

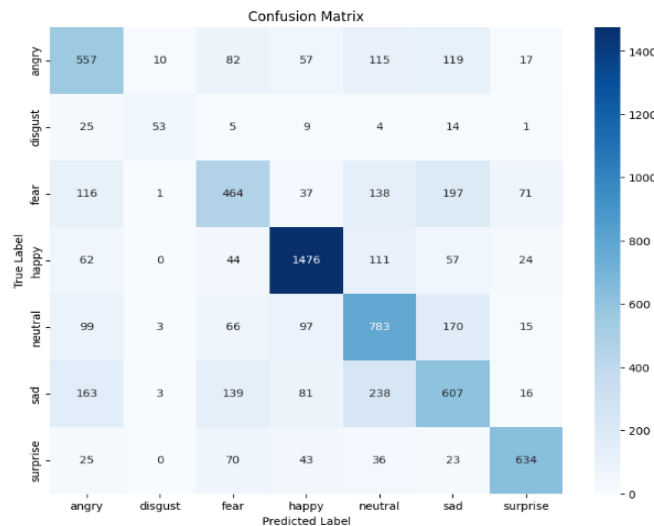


Figure 10. Confusion matrix of VGG+ResNet

The Various architectures have pros and cons concerning accuracy, computation power, and versatility. Table 1 shows us the evaluated performance of each model used in the paper. Figures 11 and 12 display the comparison graphs of the performance of each model implemented in this research.

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
Proposed model	99,80 %	0,0046	66,17 %	1,3828
VGG16	54,27 %	1,2267	46,58 %	1,4278
MobileNet	84,68 %	0,4442	60,35 %	1,2353
DenseNet	68,42 %	0,8706	56,11 %	1,2181

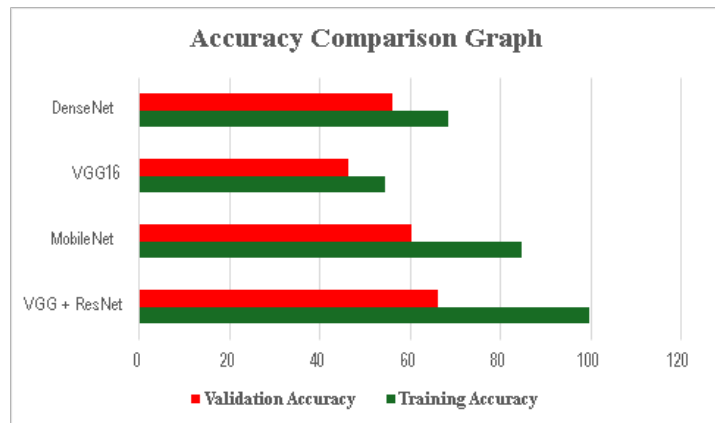


Figure 11. Accuracy comparison graph of existing models with the proposed scheme

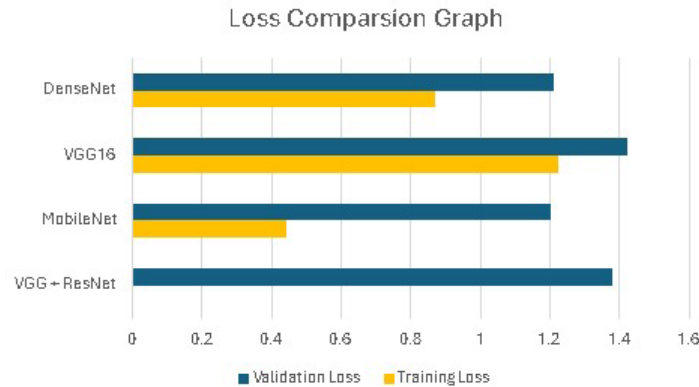


Figure 12. Loss comparison graph of existing models with the proposed scheme

By means of sophisticated deep learning systems, this paper goes in depth into exploring facial emotion recognition techniques. Our hybrid model outperforms previous models with training and validation accuracy of 99,80 % and 66,17 %, respectively. The discussion is illustrated to enhance higher levels of accuracy with the resulting models. This causes the model to misclassify the least dominant emotions like ‘fear’, ‘disgust’ and ‘surprise’. Table 2 shows us the model performance as compared with the existing literature, and figure 13 shows us the comparison of the proposed system with the existing literature on the FER2013 dataset.

Model	Validation Accuracy
Proposed model (VGG + ResNet)	66,17 %
VGG16	46,58 %
MobileNet	60,35 %

DenseNet	56,11 %
Sarvakar et al. ⁽⁸⁾	60 %
Zahara et al. ⁽⁹⁾	65,97 %
Jaiswal et al. ⁽¹⁰⁾	70,14 %
Mukhopadhyay et al. ⁽¹²⁾	65 %
Gaddam et al. ⁽¹³⁾	54,56 %
Singh et al. ⁽¹⁵⁾	61,7 %
Rakshith et al. ⁽¹⁶⁾	61 %
Irmak et al. ⁽¹⁷⁾	70 %
Huang et al. ⁽¹⁸⁾	63,10 %

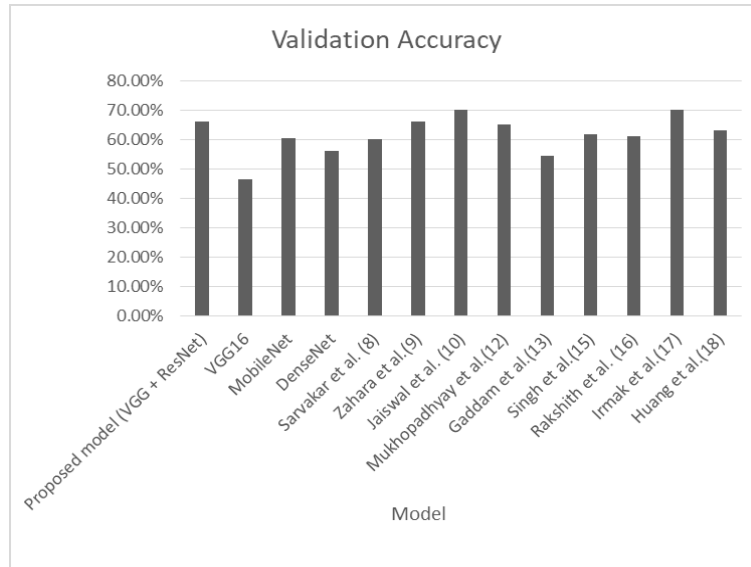


Figure 13. Graph comparing the proposed model with existing work on FER2013 dataset

CONCLUSIONS

A various deep learning models for recognizing facial emotions are experimented and analyzed in this paper. The proposed hybrid model excelled during training but fell short in validation accuracy compared to the highest reported benchmarks. This highlights the iterative nature of research and the need for ongoing optimization. While our model did not reach the highest validation accuracy, it underscores the importance of continuous innovation and refinement in deep learning for facial emotion recognition. This study lays a foundation for future advancements in image classification, integrating informatics and artificial intelligence to achieve more precise and effective models.

BIBLIOGRAPHIC REFERENCES

1. Akhand MAH, Roy S, Siddique N, Kamal MAS, Shimamura T. Facial emotion recognition using transfer learning in the deep CNN. *Electronics*. 2021;10:1-19. <https://doi.org/10.3390/electronics10091036>
2. Wu J, Zhang Y, Zhao X, Gao W. A generalized zero-shot framework for emotion recognition from body gestures. *arXiv preprint arXiv:2010.06362*. 2020. <https://arxiv.org/abs/2010.06362>
3. Majeed A, Mujtaba H, Beg MO. Emotion detection in Roman Urdu text using machine learning. In *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*. 2020. p. 125-130. <https://doi.org/10.1145/3417113.3423375>
4. Ekman P, Keltner D. Universal facial expressions of emotion. *Calif Mental Health Res Dig*. 1970;8(4):151-158. <https://www.paulekman.com/wp-content/uploads/2013/07/Universal-Facial-Expressions-of-Emotions1.pdf>
5. Saxena A, Khanna A, Gupta D. Emotion recognition and detection methods: A comprehensive survey. *J Artif Intell Syst*. 2020;2(1):53-79. <https://doi.org/10.33969/AIS.2020.21005>
6. Gadze JD, Bamfo Asante AA, Agyemang JO, Nunoo-Mensah H, Opare KAB. An investigation into the <https://doi.org/10.56294/sctconf2024960>

application of deep learning in the detection and mitigation of DDOS attack on SDN controllers. *Technologies*. 2021;9(1):1-22. <https://doi.org/10.3390/technologies9010014>

7. Corbella S, Srinivas S, Cabitza F. Applications of deep learning in dentistry. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2021;132(2):225-238. <https://doi.org/10.1016/j.oooo.2020.11.003>.

8. Sarvakar K, Senkamalavalli R, Raghavendra S, Kumar JS, Manjunath R, Jaiswal S. Facial emotion recognition using convolutional neural networks. *Mater Today Proc*. 2023;80:3560-3564. <https://doi.org/10.3390/electronics12224608>

9. Zahara L, Musa P, Wibowo EP, Karim I, Musa SB. The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. *Int Conf Informatics Comput (ICIC)*. 2020. p. 1-9. <https://doi.org/10.1109/ICIC50835.2020.9288560>

10. Jaiswal A, Raju AK, Deb S. Facial emotion detection using deep learning. *Int Conf Emerg Technol (INCET)*. 2020. p. 1-5. <https://doi.org/10.1109/INCET49848.2020.9154121>

11. Ali MF, Khatun M, Turzo NA. Facial emotion detection using neural network. *Int J Sci Eng Res*. 2020;11(8):1318-1325. https://www.researchgate.net/publication/344331972_Facial_Emotion_Detection_Using_Neural_Network

12. Mukhopadhyay M, Pal S, Nayyar A, Pramanik PKD, Dasgupta N, Choudhury P. Facial emotion detection to assess learner's state of mind in an online learning system. In *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*. 2020. p. 107-115. <https://doi.org/10.1145/3385209.3385231>

13. Gaddam DKR, Ansari MD, Vuppala S, Gunjan VK, Sati MM. Human facial emotion detection using deep learning. In *ICDSMLA 2020: Proceedings of the 2nd International Conference on Data Science, Machine Learning and Applications*. Springer. 2022. p. 1417-1427. https://doi.org/10.1007/978-981-16-3690-5_136

14. Zhang K, Li Y, Wang J, Cambria E, Li X. Real-time video emotion recognition based on reinforcement learning and domain knowledge. *IEEE Trans Circuits Syst Video Technol*. 2021;32(3):1034-1047. <https://doi.org/10.1109/TCSVT.2021.3072412>

15. Singh S, Nasoz F. Facial expression recognition with convolutional neural networks. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. 2020. p. 0324-0328. <https://doi.org/10.1109/CCWC47524.2020.9031283>

16. Rakshith MD, Kenchannavar HH, Kulkarni UP. Facial emotion recognition using three-layer ConvNet with diversity in data and minimum epochs. *Int J Intell Syst Appl Eng*. 2022;10(4):264-268. <https://ijisae.org/index.php/>

17. Irmak MC, Tas MBH, Turan S, Hasloğlu A. Emotion analysis from facial expressions using convolutional neural networks. *Int Conf Comput Sci Eng (UBMK)*. 2021. p. 570-574. <https://doi.org/10.1109/UBMK52708.2021.9558917>

18. Huang H. A facial expression recognition method based on convolutional neural network. *Front Comput Intell Syst*. 2022;2(1):116-119. <https://doi.org/10.54097/fcis.v2i1.3178>

19. Atabansi CC, Chen T, Cao R, Xu X. Transfer learning technique with VGG-16 for near-infrared facial expression recognition. *J Phys Conf Ser*. 2021;1873:1-11. <https://doi.org/10.1088/1742-6596/1873/1/012033>

20. Eberle S, Jentzen A, Riekert A, Weiss GS. Existence, uniqueness, and convergence rates for gradient flows in the training of artificial neural networks with ReLU activation. *arXiv preprint arXiv:2108.08106*. 2021. <https://doi.org/10.3934/era.2023128>

21. Hu L, Ge Q. Automatic facial expression recognition based on MobileNetV2 in real-time. *J Phys Conf Ser*. 2020;1549(2):1-7. <https://doi.org/10.1088/1742-6596/1549/2/022136>

22. Khasoggi B, Ermatita E, Sahmin S. Efficient MobileNet architecture as image recognition on mobile and embedded devices 2019;16(1):389-394 <http://doi.org/10.11591/ijeecs.v16.i1.pp389-394>

23. Prasad BR, Chandana BS. Human face emotions recognition from thermal images using DenseNet. *Int J Electr Comput Eng Syst*. 2023;14(2):155-167. <https://doi.org/10.32985/ijeces.14.2.5>

FINANCING

No financing.

CONFLICT OF INTEREST

None.

AUTHOR CONTRIBUTIONS

Conceptualization: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Data curation: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Formal analysis: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Research: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Methodology: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Drafting - original draft: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.

Writing - proofreading and editing: N Karthikeyan, K Madheswari, Hrithik Umesh, Rajkumar N, Viji C.