

## COMPARATIVE ANALYSIS OF CNN, RESNET50, EFFICIENTNETB0, AND DENSENET121 FOR FACIAL EMOTION RECOGNITION USING THE RAF-DB DATASET

**Dr. Abdul Aziz**

MS Scholar in Data Science, Department of Robotics and AI, SZABIST University, Karachi

[msds24101101@szabist.pk](mailto:msds24101101@szabist.pk)**Dr. Shahid Khan Yusufzai**

Adjunct Faculty Member, Department of Robotics and AI, SZABIST University Karachi

[shahid.khan@szabist.edu.pk](mailto:shahid.khan@szabist.edu.pk)**Naveed Ali Rajper**

MS Scholar in Data Science, Department of Robotics and AI, SZABIST University, Karachi

[msds23101140@szabist.pk](mailto:msds23101140@szabist.pk)DOI:<https://doi.org/10.5281/zenodo.20495011>**Article History**

Received: 7 March 2026

Accepted: 24 May 2026

Published: 29 May 2026

Copyright @Author

Corresponding Author: \*

Dr. Abdul Aziz

**Abstract**

Facial emotion recognition has got significant attention in the field of computer vision. It has several applications such as healthcare, surveillance, human-computer interaction, and intelligent system. The aim of this study is to compare four deep learning models CNN, ResNet50, EfficientNetB0, and DenseNet121. All four models have been trained on RAF DB dataset to recognize seven human facial emotion classes Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. All models have been trained and evaluated on the basis of different parameters such as accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. The results suggest that the CNN model achieved the highest test accuracy of 77.36%, followed closely by ResNet50 with 77.57%, on the other hand DenseNet121 and EfficientNetB0 achieved 74.66% and 73.53% respectively. All models performed really well on the dominant classes like Happy and Surprise but poorly performed on minority classes like Disgust and Fear. The overall results suggest that pre-trained models have strong capability to extract robust feature and better generalization capabilities to recognize human facial emotions. This study confirms the usefulness of deep learning models and their capabilities to recognize human emotions.

**1. Introduction**

Computer vision field got immense importance in the recent years. It has rapidly changed the way of human life. The emergence of driverless vehicles, robots, and surveillance system are the miracles of the computer vision field. It has improved the effectiveness of robots, disease detection, and human computer interaction. The human emotion system is the natural signaling to understand the behaviors of human beings. Facial

expressions communicate the intentions, feelings, and attitude towards particular situation (Kara, Churamani, & Gunes, 2021; Essahraui, et al., 2025). (Ekman & Friesen, 1971) Suggested seven human emotions surprise, fear, disgust, happy, sad, angry, and neutral. Though there is light variations in expression due to age, gender, and cultural background but mostly these seven behaviors are identified. It is difficult to understand emotional behavior now days and it

demands tools which can better understand human emotions. In the fields of financial markets to understand investor's emotions, customer behavior, human computer interaction, and adaptive learning demand such tools to understand those emotions effectively. The main objective of this paper is to empirically evaluate FER generalization using RAF DB dataset. This study has attempted to compare four models Convolutional Neural Network (CNN) which has been trained from scratch, and three transfer learning models ResNet50, EfficientNetB0, and DenseNet121. We have evaluated

- Comparative performance of CNN and other pre trained models
- Assessment and analysis of each model

RAF DB dataset is publicly available at Kaggle website, it is human facial expression data and it is our ethical responsibility to take care the privacy of everyone. To the best knowledge of the authors, we have not found such disclaimer which state that these images cannot be included rather we have found several studies which presented image output using different models. Therefore, this study also follows the same. These images have been included for academic (non-commercial) and research purpose only, copyright and ownership remain with dataset providers.

### Literature Review

We have reviewed eleven research papers related to topic of human emotion recognition, which evaluated dataset of RAF-DB from 2025, and 2026 publications. The overall analysis of literature suggests that previous studies used CNN, fusion architectures, hybrid CNN transformers to analyze dataset of RAF-DB. The extensive review of literature further suggests that recent studies have focused on transformer-enhanced recognition, cross-domain adoption, variation inference, efficiency aware FER systems, and compound emotion molding. The overall literature review suggests that recent models performed better as compare to previously trained models.

In this research (Alzahrani, et al., 2025) have carried their research on facial expression using data RAF-DB and suggested a network which was conducive to recognize the facial expression. This deep learning model was based on dynamic cross-domain and

utilized network of dual attention to recognize expression. The network extracted both local and global features from images using global and local feature extractor; it has also adopted semantic awareness, which ameliorated the domain generalization for the model. The findings of the research suggested that model performed really well on the RAF-DB dataset.

In this research (Sofia , Ferreira, & Leite, 2026) have carried their work on the recognizing compound facial expression using modern models of computer vision as well as quantum kernel. They utilized data of RAF-DB to analyze the seven human facial expressions. They produced their results using two classical hybrid models such as VGGFace-SVM and ResNet50-SVM, 2 modern models such as ViT-B/16, EfficientNetV2-S and 3 quantum kernel models such as QSVM, QKNN, QCNN. Among all ViT-B/16 performed well and achieved 63.13% accuracy but took more time for training. The second closer model was EfficientNetV2-S with accuracy 60.9% but took lesser time for training. In quantum kernel models, QSVM performed well with 54.97% accuracy.

In this research (Lv, Gang, Zhang, & Tsoi, 2026) conducted their research on recognition of facial expression using variation inference. Their research focused on the issues related to conventional models, which failed to address psychological research, which suggested that expressions are combination of basic units, but models failed to detect intrinsic uncertainty as well as nature of emotions. To solve this problem, they proposed POSTER-Var, which used variation inference to capture uncertainty aware for FER. It is robust model, which ameliorated noisy label as well as subtle expressions. They also utilized this model on RAF-DB.

In this research (Xu, et al., 2025) carried out their work on the topic of recognizing facial expressions using combination of vision transformer and fixed weight ArcFace. They utilized ArcFace loss which is well designed for inter class separation and better performance in facial recognition. Due to its limitation angular distribution, they rather used ArcFace with weight constrained along with Vision

Transformers (ViT). It was conducive to reduce biases due to imbalance data and computationally cost effective. They carried on their work on RAF-DB dataset and FER2013 dataset and utilized different structures of networks, classical loss and ArcFace loss. This approach found computationally effective with better accuracy in recognizing facial expression.

In this research (Florestiyanto, Yanu, Surjono, & Jati, 2026) conducting their study on recognizing facial emotions using both cross corpus and multi corpus to compare results of both approaches on FER2013, RAF-DB, AffectNet, and CK+. They used four datasets and applied deep learning-based techniques to analyze facial emotion recognition. They perform preprocessing techniques to ensure quality of data, color combination and dataset bias to produce quality results. They evaluated cross corpus and multi corpus on all datasets and found multi corpus produced better accuracy as compare to cross corpus.

In this research (Lv, Gang, & Zhang, Facial Expression Recognition via Variational Inference, 2025) carried out work on this paper. The main aim of this was to ameliorate FER performance through resolving issues related to feature representation and uncertainty. They suggested a model POSTER-Var it is a variation inference-based model, which incorporates uncertainty modeling along with hierarchical feature learning. They utilized transformer-based methodology, which extracted features combining layer embedding is along with VICH in order to capture complex patterns in facial expression. The findings suggest that model obtained 92.76% accuracy on RAF-DB dataset, it was better accuracy in baseline models and it is computationally effective. This study finally recommended incorporating modules related to uncertainty awareness in FER to increase generalization of the model specially to deal noisy datasets of the real world.

In this research (Jayaswal, et al., 2025) evaluated recognition of facial expression using variation inference. They focused on the issue that certain variations on facial expressions become difficult to recognize and require better feature extraction as well as discriminative ability for the model. This study also

contradicted with studies, which focused only on intra class differences or inter class similarities but ignored inconsistencies in single image. To address this issue, they utilized variational inference-based classification head that mainly focus to capture variation in a single image. This approach helped to create robust model, which captured regional features in broader way. For the further improvement of the model, they also utilized multi stage feature fusion along with nonlinear transformation. They utilized data of RAF DB, FER2013, and AffectNet and found that model performed better as compare to previously trained models. This paper helps to test the dataset using this approach to find the variations in a single image rather relying only on traditional approaches.

In this research (Tutuianu, Ipinze, Liu, Alamäki, & Kauttonen, 2024) conducted this research in order to resolve the issues related to dataset imbalance and generalization issues among the cross domain for the improvement in robustness of FER models. They used dataset RAF-DB and compared with balanced in the wild dataset also known BTFER. They utilized methodology, which focused on validating cross-domain and distributed of balance data in order to reduce bias. Findings suggest better performance if balance datasets are used during training. These results were limited only on the dataset of RAF-DB. This paper recommended combining RAF-DB with balance dataset to generalize results across other domains in order to obtain reliability in the FER deployment.

In this research (Lin, et al., TAG: Thinking with Action Unit Grounding for Facial Expression Recognition, 2026) carried out their work on this paper to increase interpretability along with reasoning ability among FER models. They proposed TAG also called AU grounded transformer framework, which integrates predictions of facial action units. They used mix methodology supervised fine-tuning with reinforcement learning by further combing with rewards based on AU. It was found that model achieved better accuracy with RAF-DB and improved quality of visual explanation. It recommended integrating techniques of explainable AI with action unit grounding for producing reliable FER system.

In this paper (Jayaraman, Samanthisvaran, & Mahendran, 2025) introduced Deep CNN along with architecture of BiLSTM which was basically a dual mechanism to recognize facial expressions. Thorough this combined approach spatial CNN layers captured feature of face and BiLSTM, which ameliorated subtle expression classification, captured dependencies of temporal or emotional. This model focused on eyes, eyebrows and mouth, which are the most, important facial parts using RAF-DB dataset. The overall performance was very well as compared to baseline models of CNN.

In this research (Mishra, et al., 2026) introduced a lightweight model hybrid CNN attention named EmoVisioNet which was useful to recognize facial emotions in real time. This model extracted features using MobileNet, spatial focus through self-attention and the enhancement of local features through patch-based techniques. The model produced more than 96% accuracy on the RAF-DB dataset and its performance was optimal on the edge devices, which are used in real time.

Recent 2025–2026 review of literature suggests major shift from traditional CNN models to hybrid system and transformer-driven models. These models focus on uncertainty, fairness, explain ability, cross-domain generalization and compound emotion recognition. The major emphasis is on the lightweight deployment of the model to achieve operational efficiency. These studies utilized different types of techniques but mainly emphasized on the improvement in computational cost along with better accuracy. Some studies focused on comparative study and some emphasized on improving single image rather focusing only on intra class differences or inter class similarities. The major consideration is on robustness and efficiency of the model with accuracy-focused benchmarking.

Despite the fact, recent studies have achieved better accuracy, explainability, deployment efficiency, and cross culture, bias but there is stillroom for improvement. The future research can be conducted to further improve the above parameters as well as it can also evaluate trustworthiness of AI models,

recognition of adaptive emotions in real time and multimodal FER.

### Research Methodology

This study proposes a Convolutional Neural Network (CNN) based framework, ResNet50, EfficientNetB0, and DenseNet121 for human emotion recognition using facial images from the RAF-DB dataset. The CNN based models are capable of extracting hierarchical facial features automatically. Along with CNN, pre trained models have also been trained because data size was approximately 15339 images that were very low for deep learning models. To train CNN from scratch may inclined towards over-fitting, the pre-trained model will be conducive to validate the CNN results. These experiments have been carried using Anaconda Jupyter notebook with TensorFlow, Keras, Seaborn, OpenCV, NumPy, Matplotlib, and Scikit-learn libraries.

The Real-world Affective Faces Database (RAF-DB) was utilized as the benchmark dataset with 15339 images. The dataset contains real-world facial expressions categorized into seven emotional classes: Surprise, Fear, Disgust, Happy, Sad, Angry, and Neutral. Images were organized into training and testing folders according to their labels. Initially both training and testing folders have been concatenated and then split into three folders 70% of the training data, 15% testing, and 15% validation using stratified splitting to preserve class distribution.

The CNN model from scratch was run on original image size 100x100x3 because several studies suggested to train model from scratch did not require resizing but for pre-trained models, we resized images to 224x224x3 because all three models have been pre-trained using same resolution. Pixel values were normalized between 0 and 1 to improve convergence during training. Corrupted or unreadable images were excluded automatically during the loading process. Inclusion criteria consisted of properly labeled facial images belonging to one of the seven RAF-DB classes, while blurred, damaged, or unreadable images were excluded. Batch processing and prefetching were applied to optimize computational performance.

The CNN model was trained using the Adam optimizer and categorical cross-entropy loss function

for 50 epochs with batch normalization, dropout, Relu activation function in hidden layers, and Softmax in output layer. The pre-trained models were fine-tune to make them compatible with current dataset.

Performance evaluation was conducted using validation accuracy, training loss, validation loss, and test accuracy metrics.

3. Data Analysis and Discussion

3.1 CNN architecture

Table 4.1- Accuracy Results of Train -Validation-Test

	Train	Validation	Test
Accuracy	0.8249	0.7962	0.7736

Table 4.2-Classification Report

	Precision	Recall	F1-score	Support
Angry	0.68	0.62	0.65	130
Disgust	0.46	0.27	0.34	132
Fear	0.83	0.47	0.60	53
Happy	0.90	0.89	0.90	894
Neutral	0.67	0.80	0.73	480
Sad	0.67	0.76	0.71	369
Surprise	0.88	0.72	0.79	243
Accuracy			0.77	2301
Macro avg	0.73	0.65	0.67	2301
Weighted avg	0.77	0.77	0.77	2301

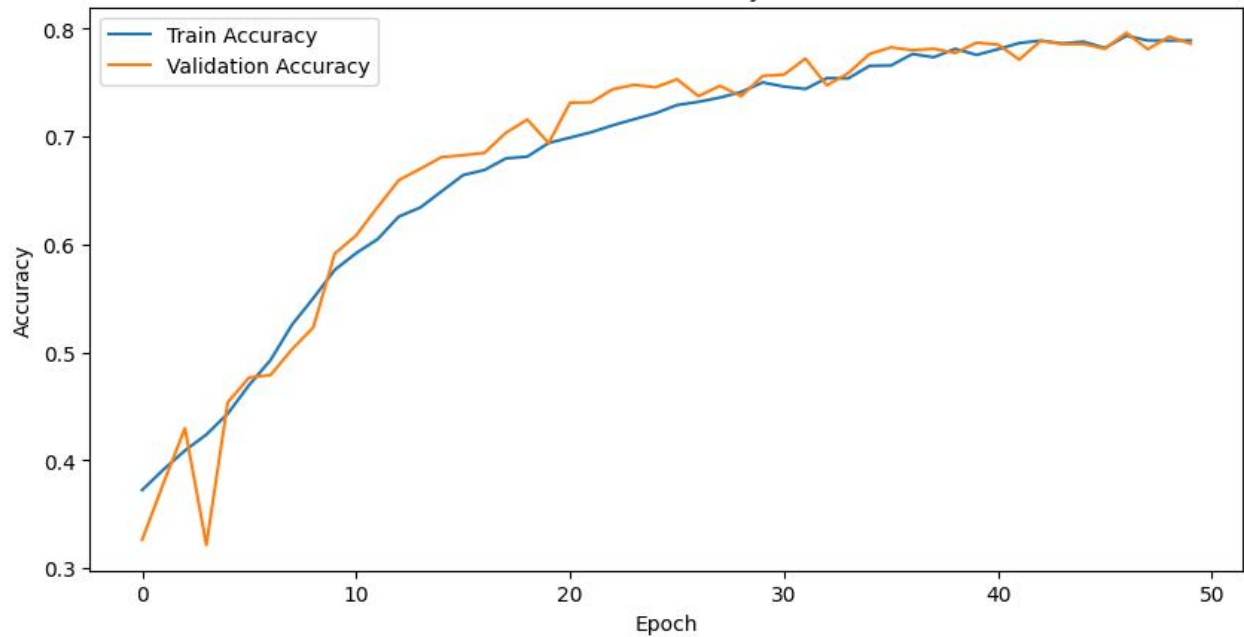
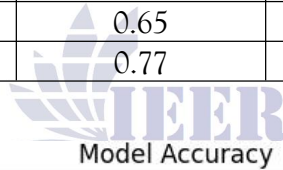


Figure 4.1 Train-Validation Accuracy

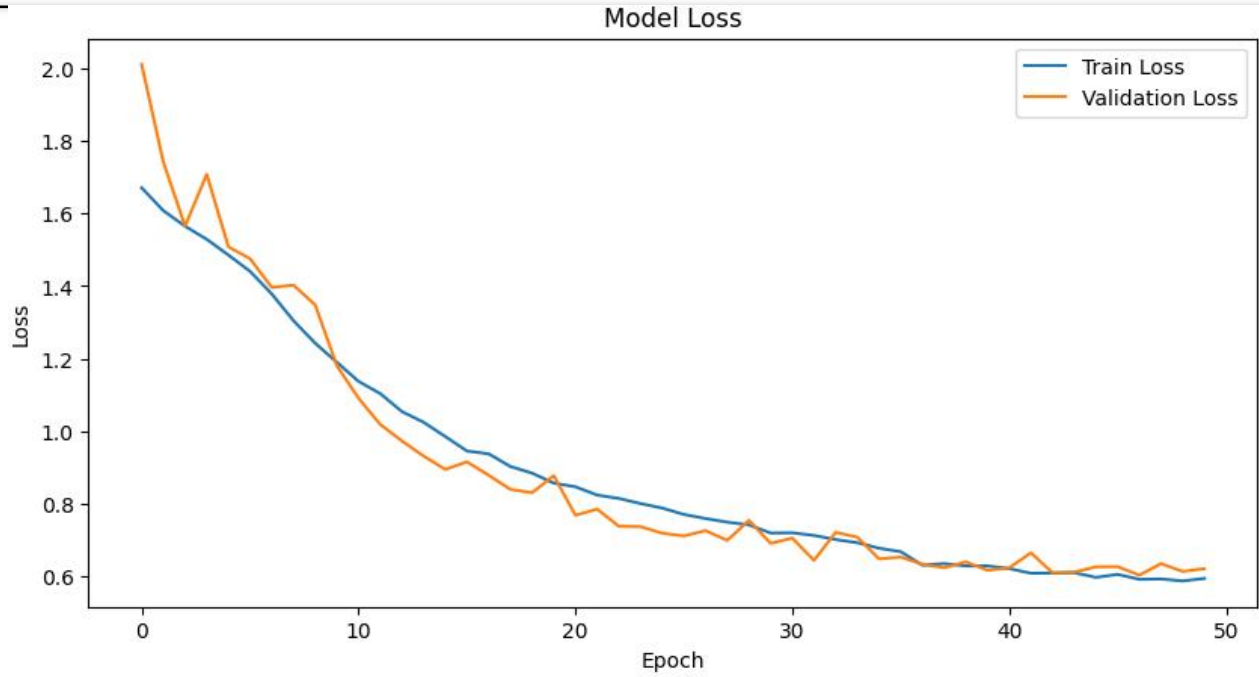


Figure 4.2 Train-Validation Loss



Figure 4.3 CNN Confusion Matrix

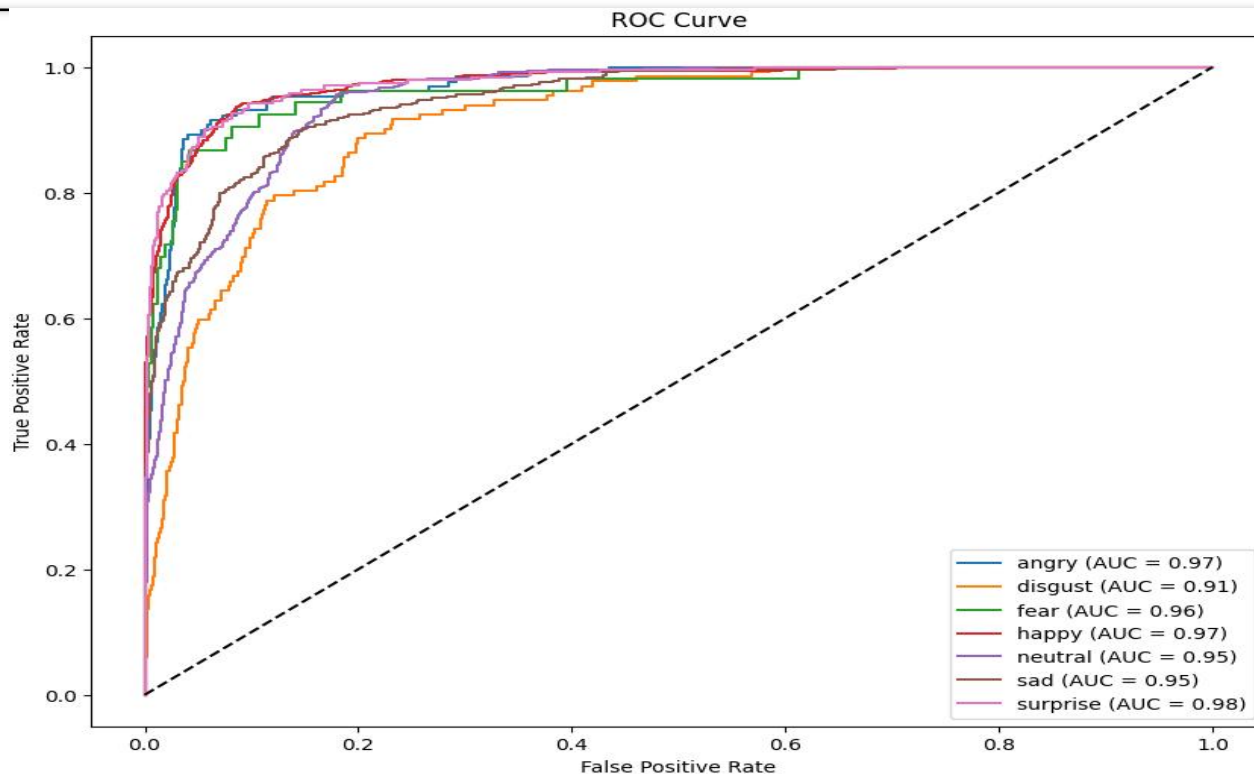


Figure 4.4 CNN ROC Curve

The CNN model performed well on RAF DB dataset to recognize seven emotions. There were several convolutional layers along with Batch Normalization, Dropout, Global Average Pooling, and Max Pooling. This model has approximately 652,583 trainable parameters. Table 4.1 and figure 4.1 suggest that model has gradually been converged towards global minima, its validation accuracy was 32.64% at first epoch but increased to 79.62% at epoch 47. The validation AUC touched to the level of 0.969 that confirmed that it has good generalization capability to recognize different emotion classes. The F1-scores

presented in table 4.2 for Happy and Surprise is 0.90 and 0.79 respectively suggest model performance for both classes is exceptional. On the other hand, model poorly performed for Disgust and Fear classes. There may be reasons of imbalance or complexity of class. It has also been noticed that loss value gradually reduced and converged smoothly it indicates regularization techniques, such as Dropout and adopted learning rate, performed well and overcome over fitting issue. In this case it can be concluded that CNN model performed well to recognize human emotion recognition applied on RAF-DB dataset.

3.2 ResNet50

Table 4.3-Accuracy Results of Train -Validation-Test

	Train	Validation	Test
Accuracy	0.8844	0.7658	0.7757

Table 4.4-Classification Report

	Precision	Recall	F1-score	Support
Angry	0.77	0.63	0.69	130
Disgust	0.57	0.43	0.49	132
Fear	0.70	0.53	0.60	53
Happy	0.92	0.88	0.90	894
Neutral	0.72	0.67	0.69	480

Sad	0.67	0.83	0.74	369
Surprise	0.68	0.84	0.75	243
Accuracy			0.78	2301
Macro avg	0.72	0.69	0.71	2301
Weighted avg	0.78	0.78	0.77	2301

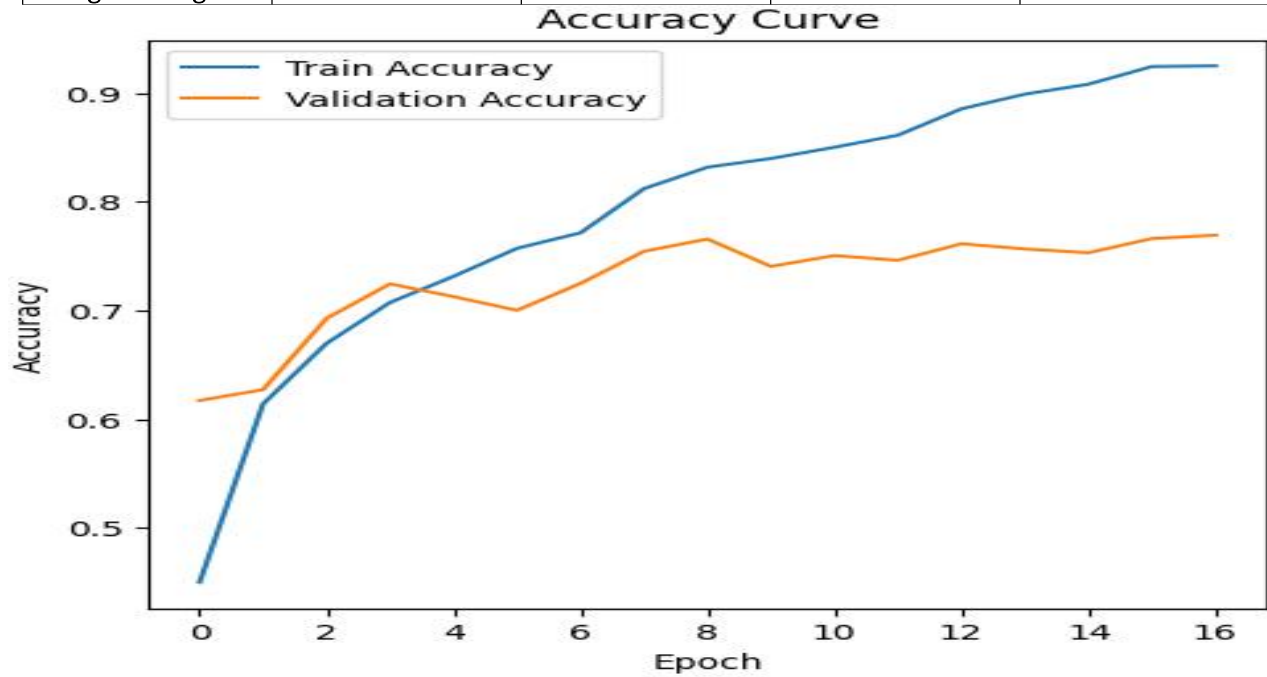


Figure 4.5 Train-Validation Accuracy

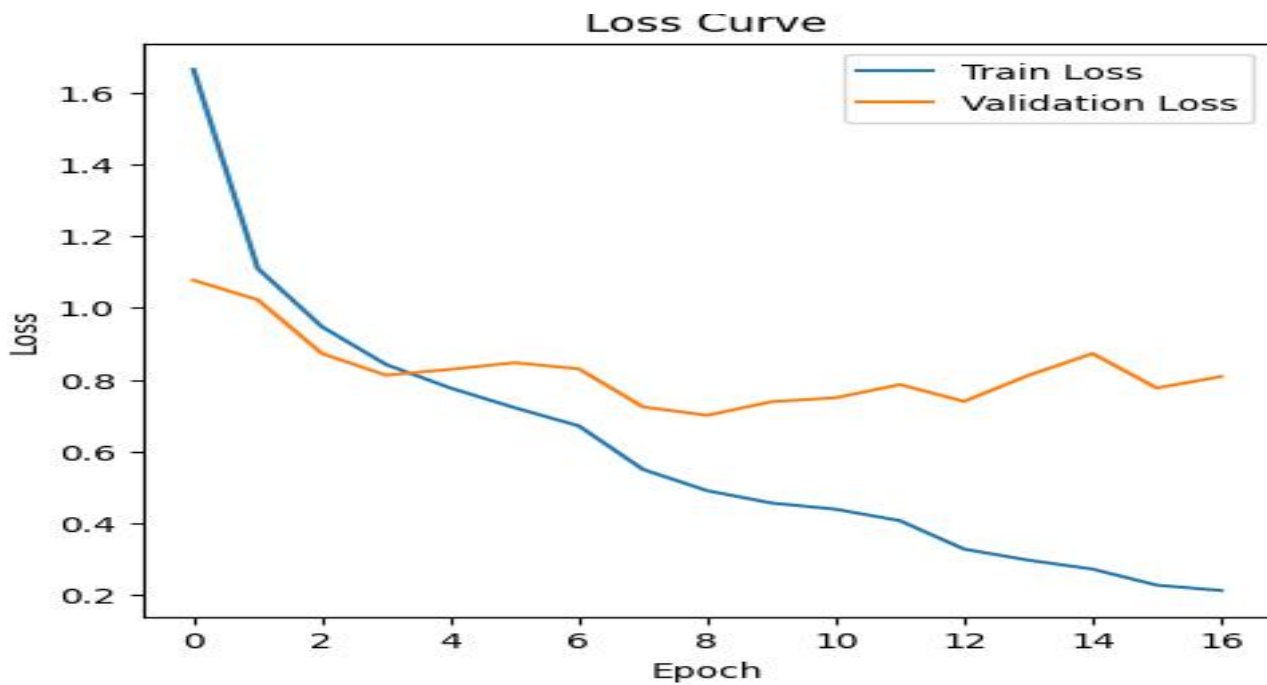


Figure 4.6 Train-Validation Loss

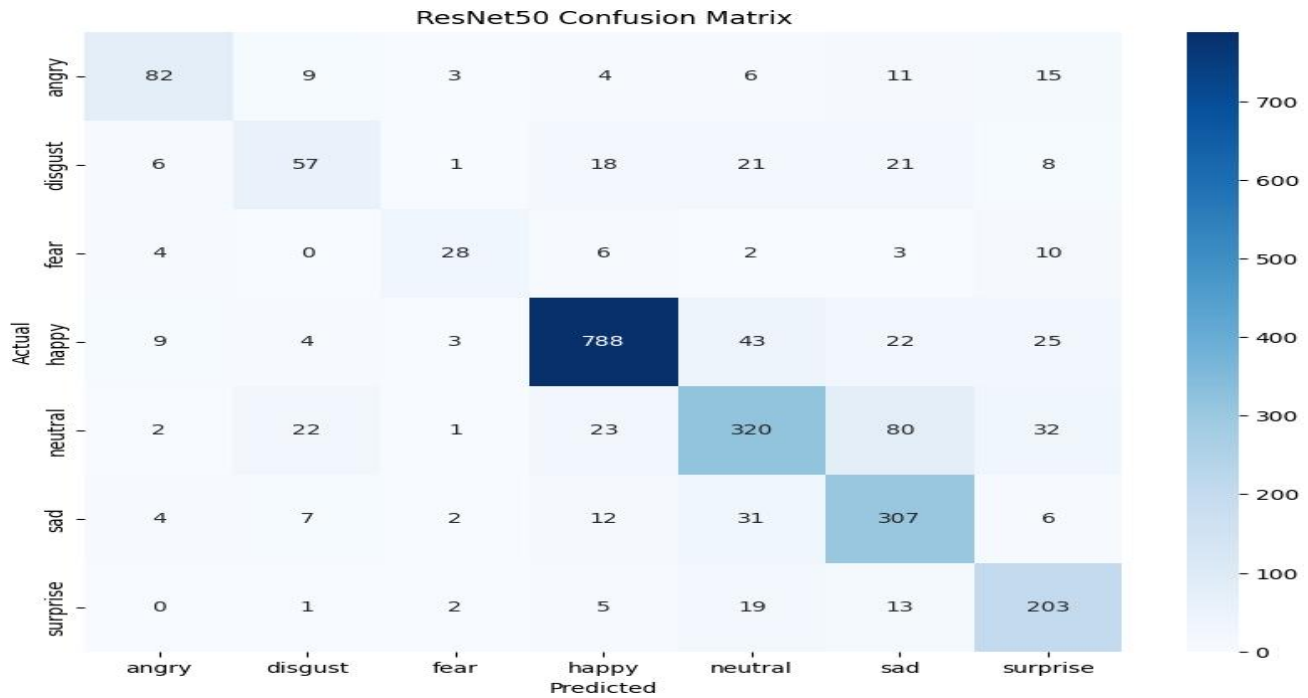


Figure 4.7 ResNet50 Confusion Matrix

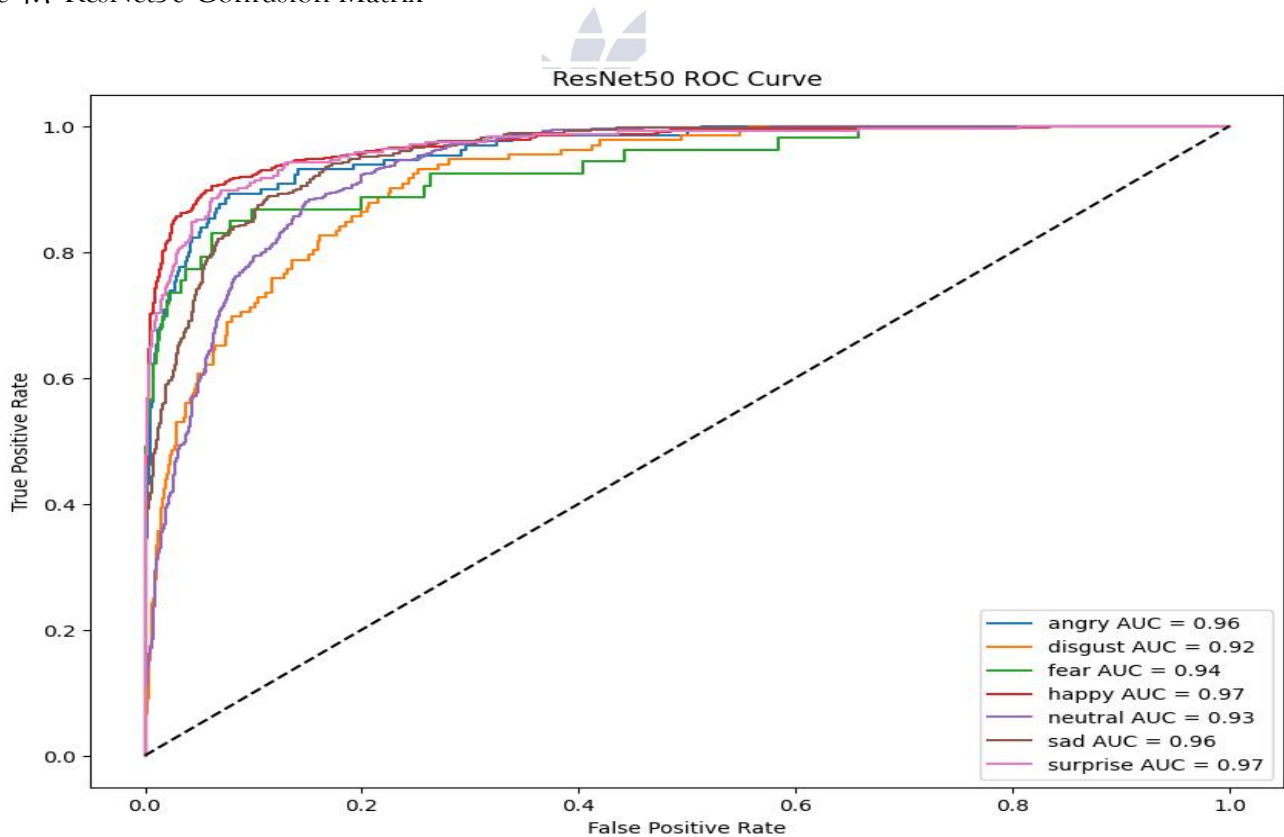
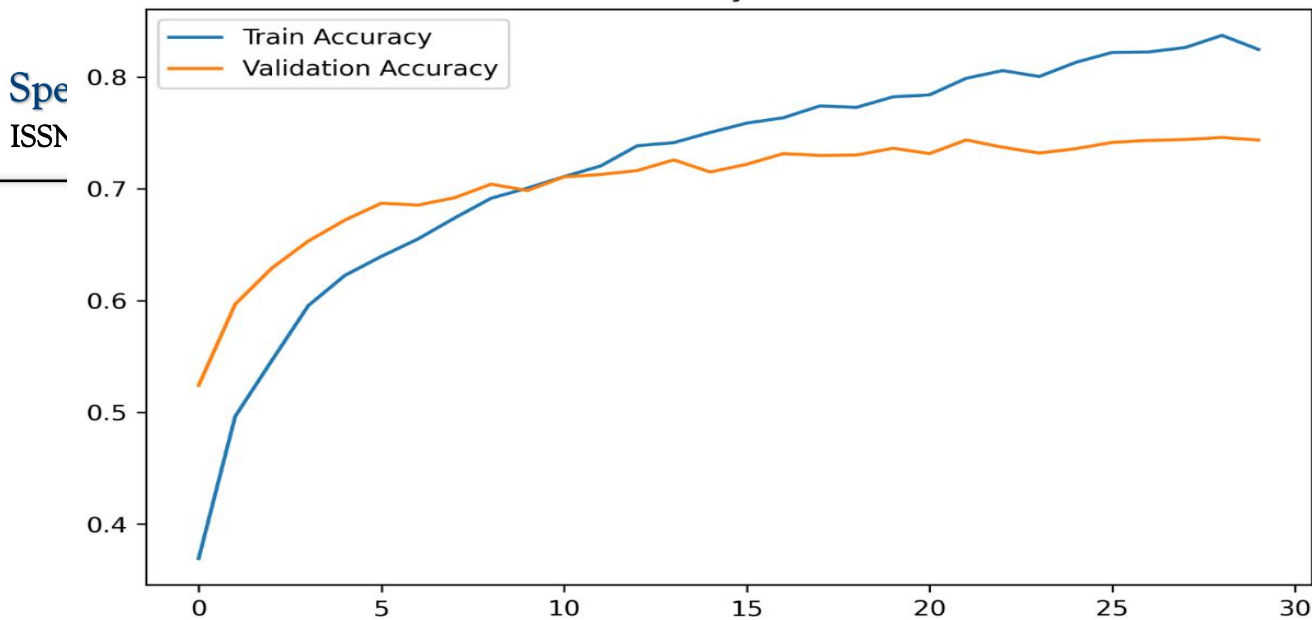


Figure 4.8 ResNet50 ROC Curve

The second model we applied was the ResNet50, it is transfer-learning model and applied on RAF DB

dataset. It is a pre-trained model trained on ImageNet data. Out dataset contains 10737 training images, and

Accuracy Curve



2301 for both testing and validation. To produce better result, it was fine-tuned. The custom classification head was added along with global average pooling, fully connected dense layer, and batch normalization. The model convergence behavior during training was very well. Its validation accuracy at first epoch was 61.71% which touched to 76.92% and test accuracy is 77.57% with validation AUC is 0.9576. The ReduceLRonPlateau, scheduling strategy for learning rate performed really well and generalization

ameliorated at later epoch level. The F1-scores presented in table 4.4 for Happy and Surprise is 0.90 and 0.75 respectively suggest model performance for both classes is exceptional. On the other hand model poorly performed for Disgust and Fear class. There may be reasons of imbalance or complexity of class. In this case it can be concluded that ResNet50 model performed well compared to CNN but there is problem of over-fitting which shows limitation of this model.

### 3.3 EfficientNetB0

Table 4.5- Accuracy Results of Train -Validation-Test

	Train	Validation	Test
Accuracy	0.8750	0.7505	0.7353

Table 4.6-Classification Report

	Precision	Recall	F1-score	Support
Angry	0.75	0.61	0.67	130
Disgust	0.41	0.34	0.37	132
Fear	0.61	0.43	0.51	53
Happy	0.88	0.86	0.87	894
Neutral	0.66	0.71	0.68	480
Sad	0.69	0.69	0.69	369
Surprise	0.61	0.75	0.67	243
Accuracy			0.73	2301
Macro avg	0.66	0.63	0.64	2301
Weighted avg	0.73	0.73	0.73	2301

Figure 4.9 Train-Validation Accuracy

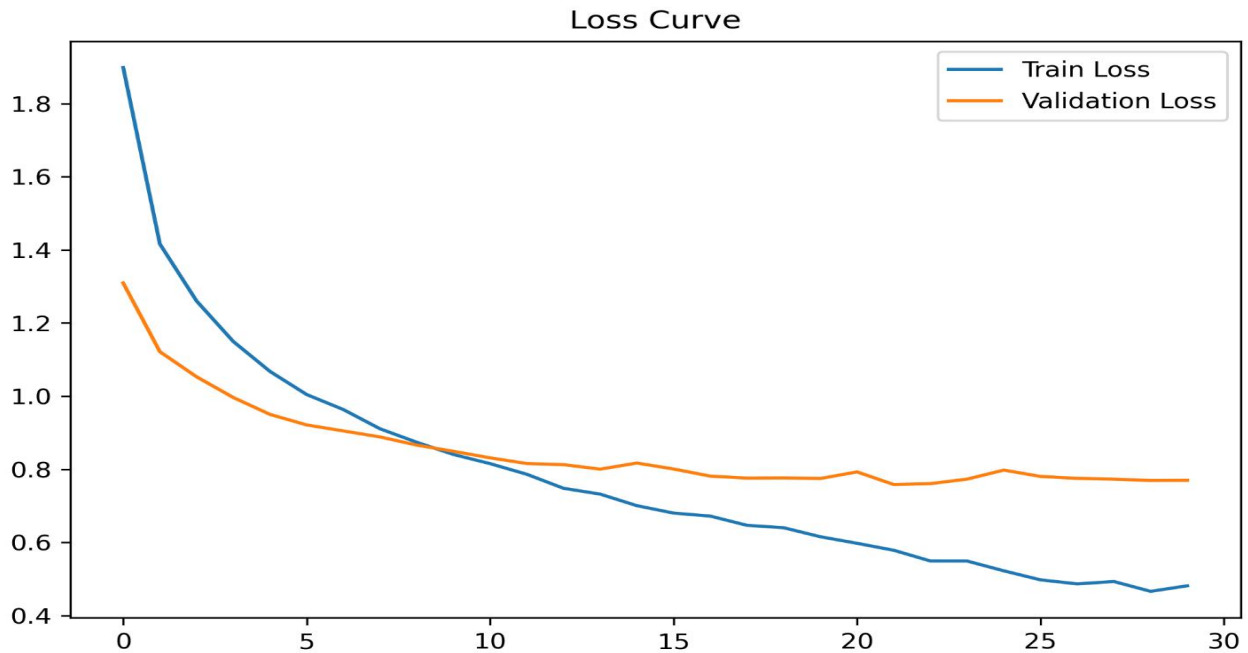


Figure 4.10 Train-Validation Loss

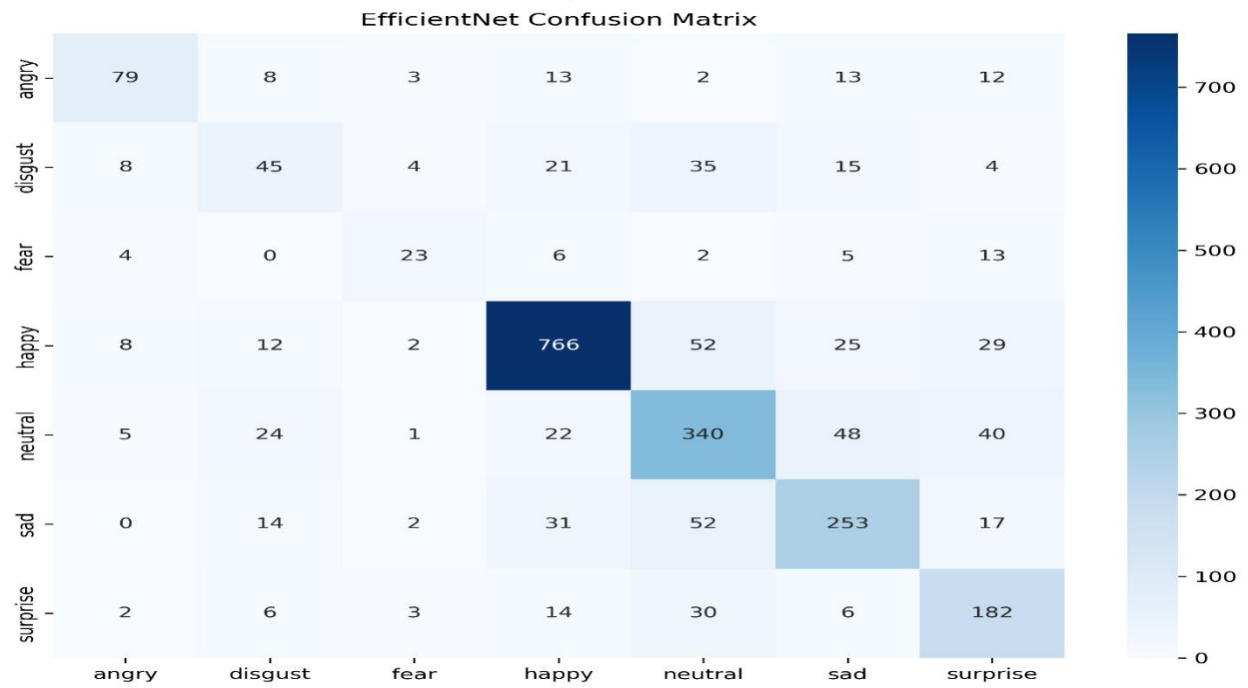


Figure 4.11 EfficientNetB0 Confusion Matrix

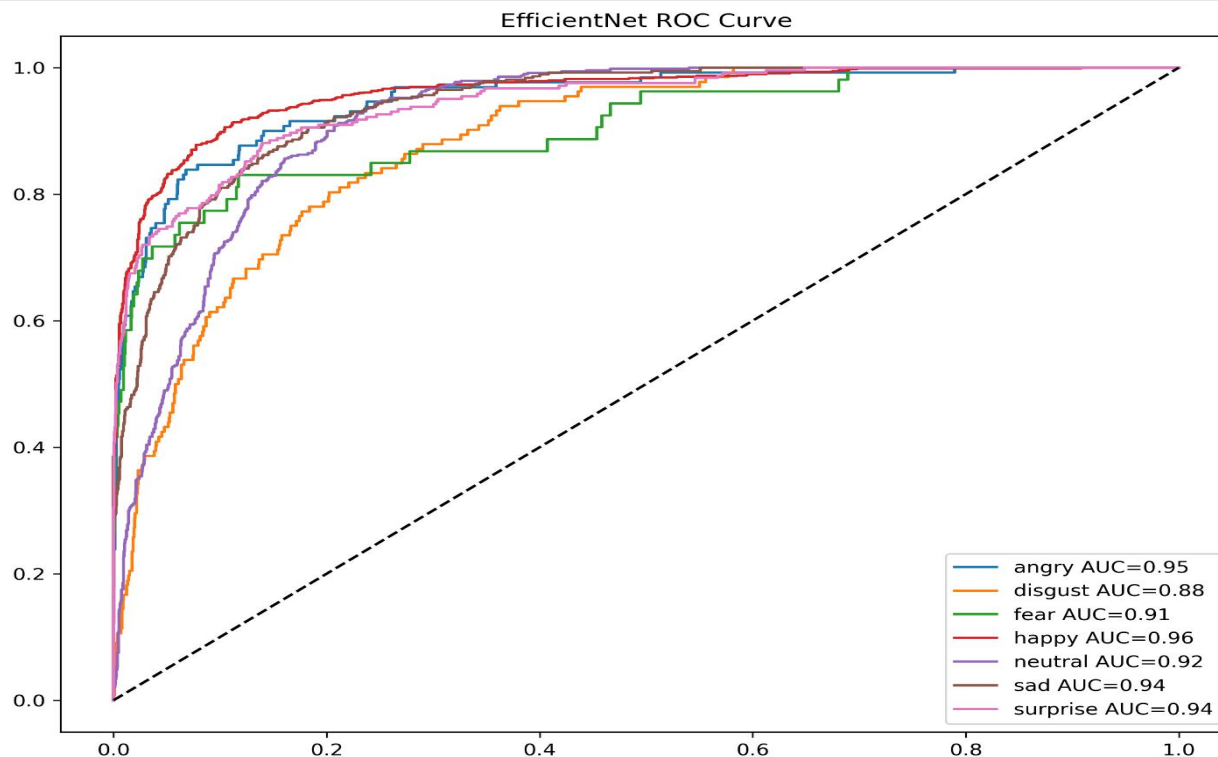


Figure 4.12 EfficientNetB0 ROC Curve

The third model we applied was the EfficientNetB0, it is transfer-learning model and applied on RAF DB dataset. Its performance is very good with training, validation, and testing accuracies of 87.5%, 75.05%, and 73.53% respectively. There is 14% gap between training and testing accuracies highlights over-fitting issue. The F1-scores presented in table 4.6 for Happy is 0.87 suggests model performance for Happy class is exceptional. There is moderate performance of Angry, Neutral and Sad classes. On the other hand, model

poorly performed for disgust and fear class. There may be reasons of imbalance or complexity of class. In this case it can be concluded that EfficientNetB0 model training accuracy is better than ResNet50 and CNN but the gap between training and test accuracies indicates slight over-fitting issue. This model is computationally effective than ResNet50 and CNN. It took lesser time to get trained with training accuracy is around 87.5%.

3.4 DenseNet121

Table 4.7- Accuracy Results of Train -Validation-Test

	Train	Validation	Test
Accuracy	0.8291	0.7362	0.7466

Table 4.8-Classification Report

	Precision	Recall	F1-score	Support
Angry	0.67	0.58	0.63	130
Disgust	0.33	0.33	0.33	132
Fear	0.71	0.57	0.63	53
Happy	0.93	0.83	0.88	894
Neutral	0.65	0.76	0.70	480
Sad	0.67	0.76	0.71	369

Surprise	0.77	0.72	0.75	243
Accuracy			0.75	2301
Macro avg	0.68	0.65	0.66	2301
Weighted avg	0.76	0.75	0.75	2301

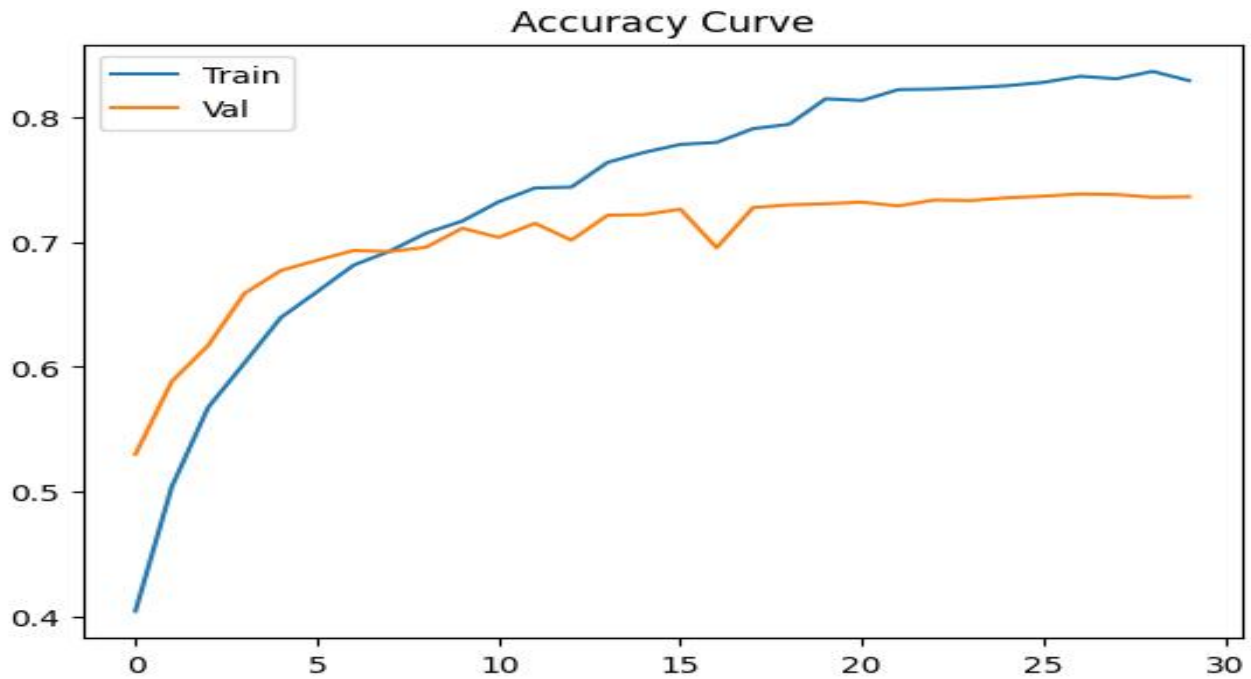


Figure 4.13 Train-Validation Accuracy

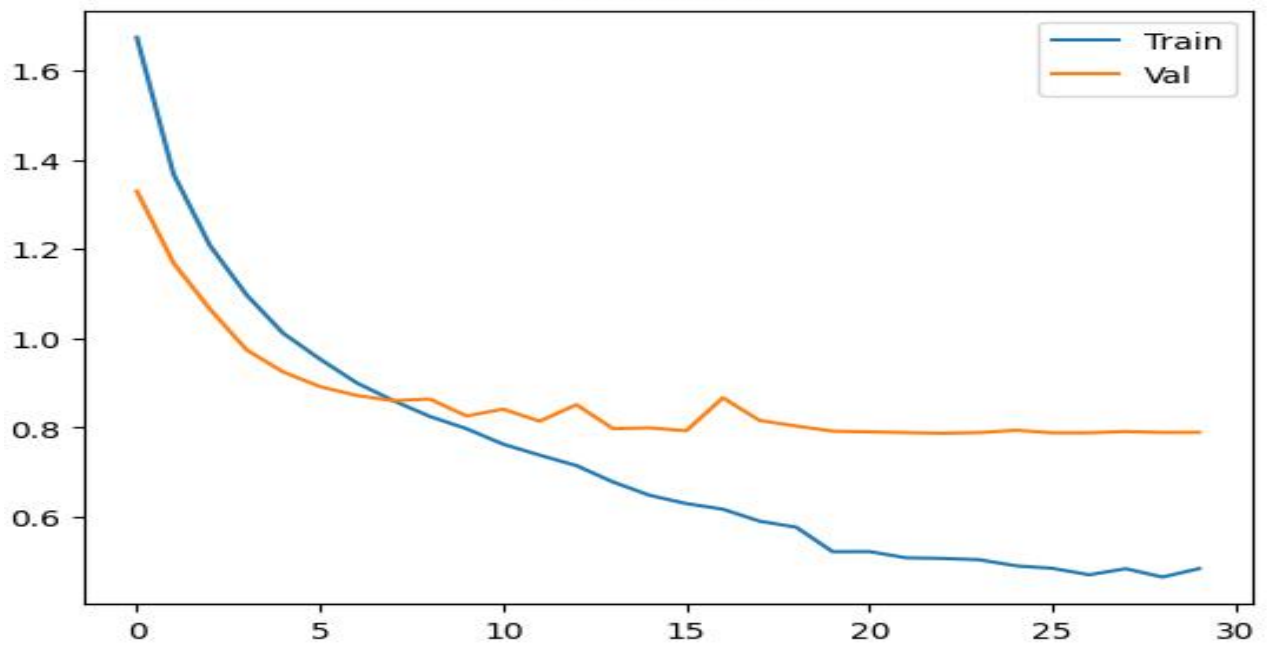


Figure 4.14 Train-Validation Accuracy

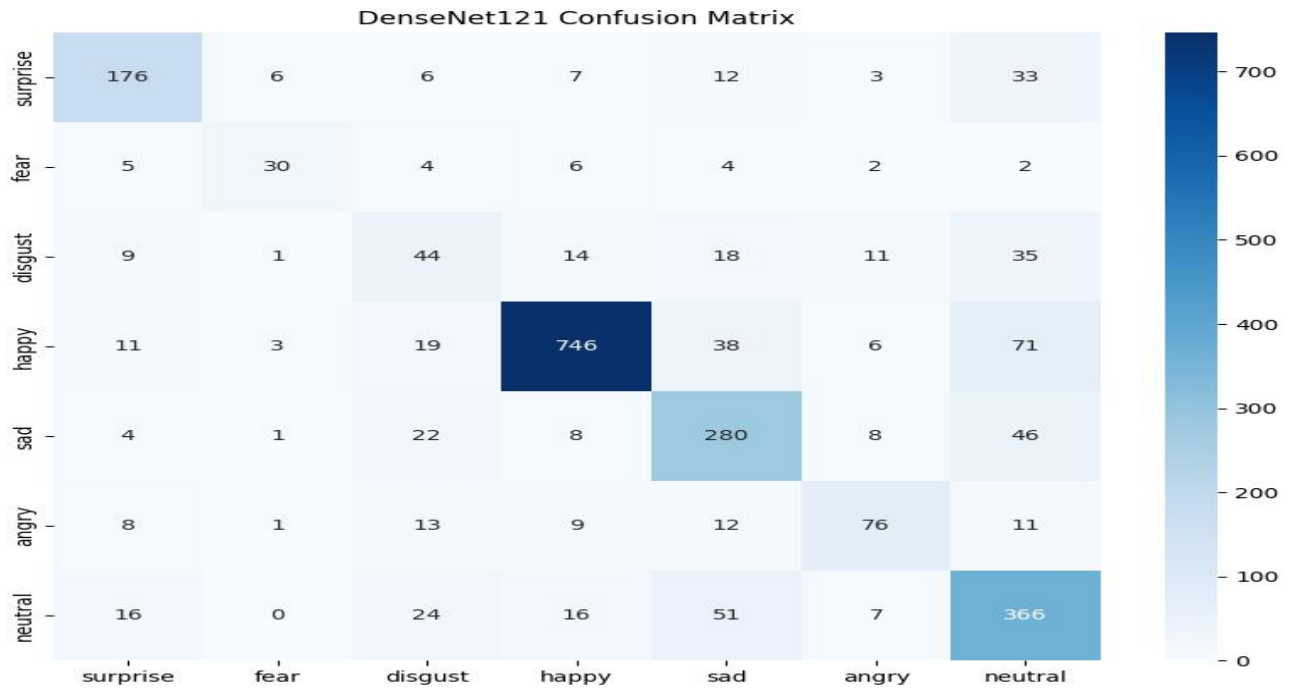


Figure 4.15 DenseNet121 Confusion Matrix

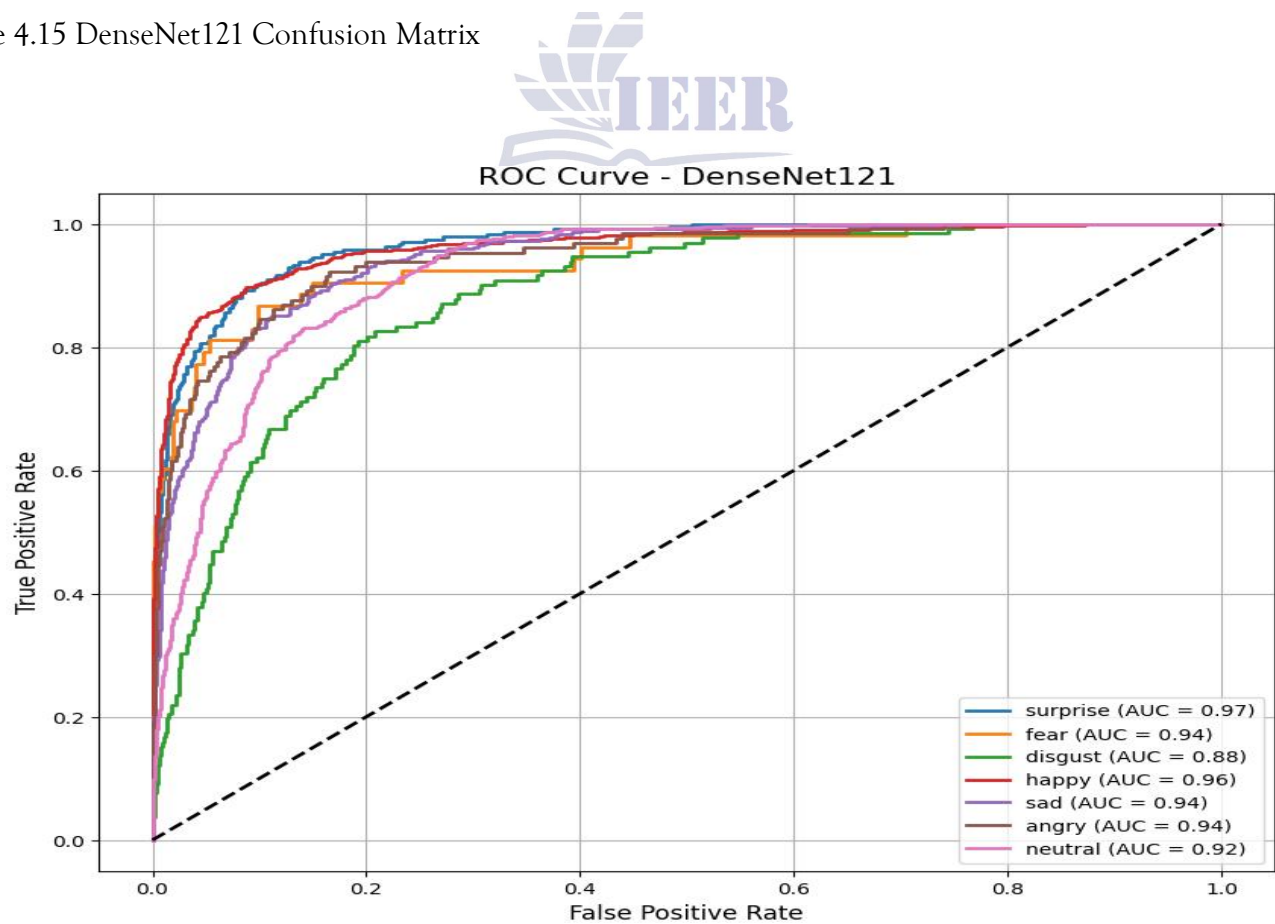


Figure 4.16 DenseNet121 ROC Curve

The fourth model we applied was the DenseNet121, it is transfer-learning model and applied on RAF DB dataset. Its performance is very good with training, validation, and testing accuracies of 82.91%, 73.62%, and 74.66% respectively. There is 8% gap between training and testing accuracies highlights limited over-fitting issue. The F1-scores presented in table 4.8 for Happy is 0.88 suggests model performance for Happy class is exceptional. Similarly, there is moderate performance of Angry, Fear, Neutral and Sad classes. On the other hand, model poorly performed for

Disgust class. There may be reasons of imbalance or complexity of class. In this case it can be concluded that EfficientNetB0 model training accuracy is lower than all other three models. The gap between training and test accuracies indicates limited over-fitting issue. As compare to other pre-trained model this model produced lower gap between training and test accuracies indicate this model is comparatively better than ResNet50 and EfficientNetB0.

CNN-ResNet50-EfficientNetB0-DenseNet121 Actual and Predicted Results



Figure 4.17 CNN-ResNet50-EfficientNetB0-DenseNet121 Actual and Predicted Results

3.6 Comparative Analysis

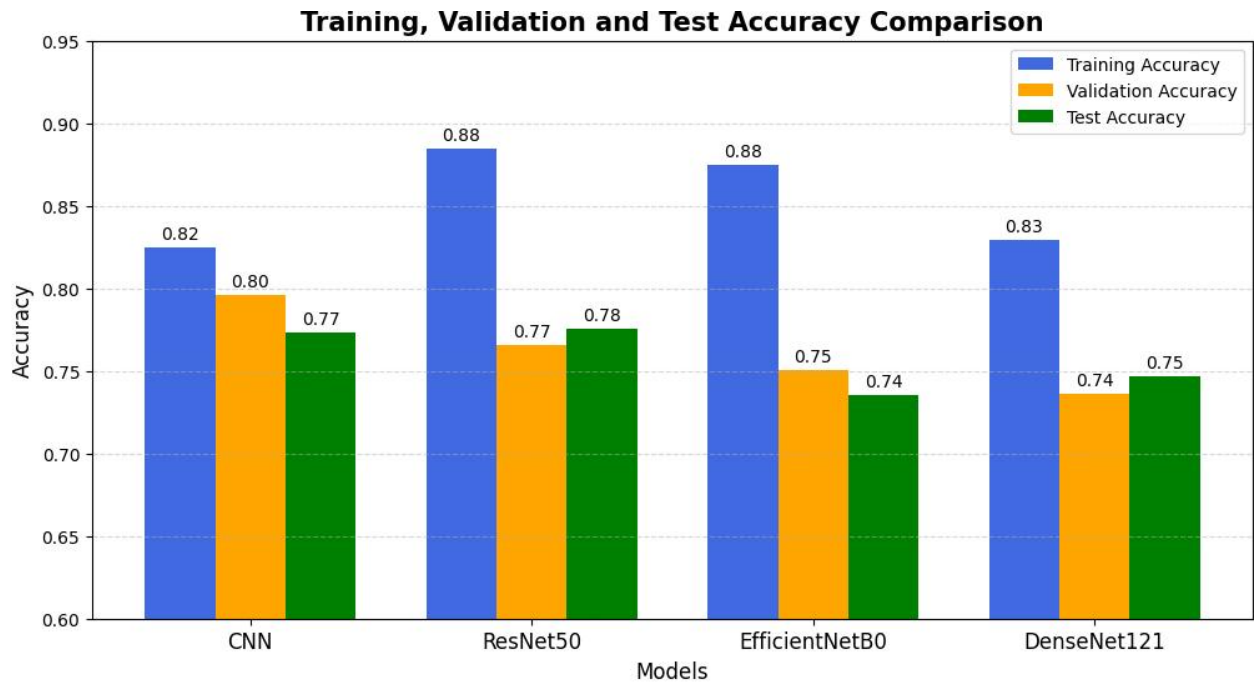


Figure 4.18 Training-Validation and Test Accuracy Comparison

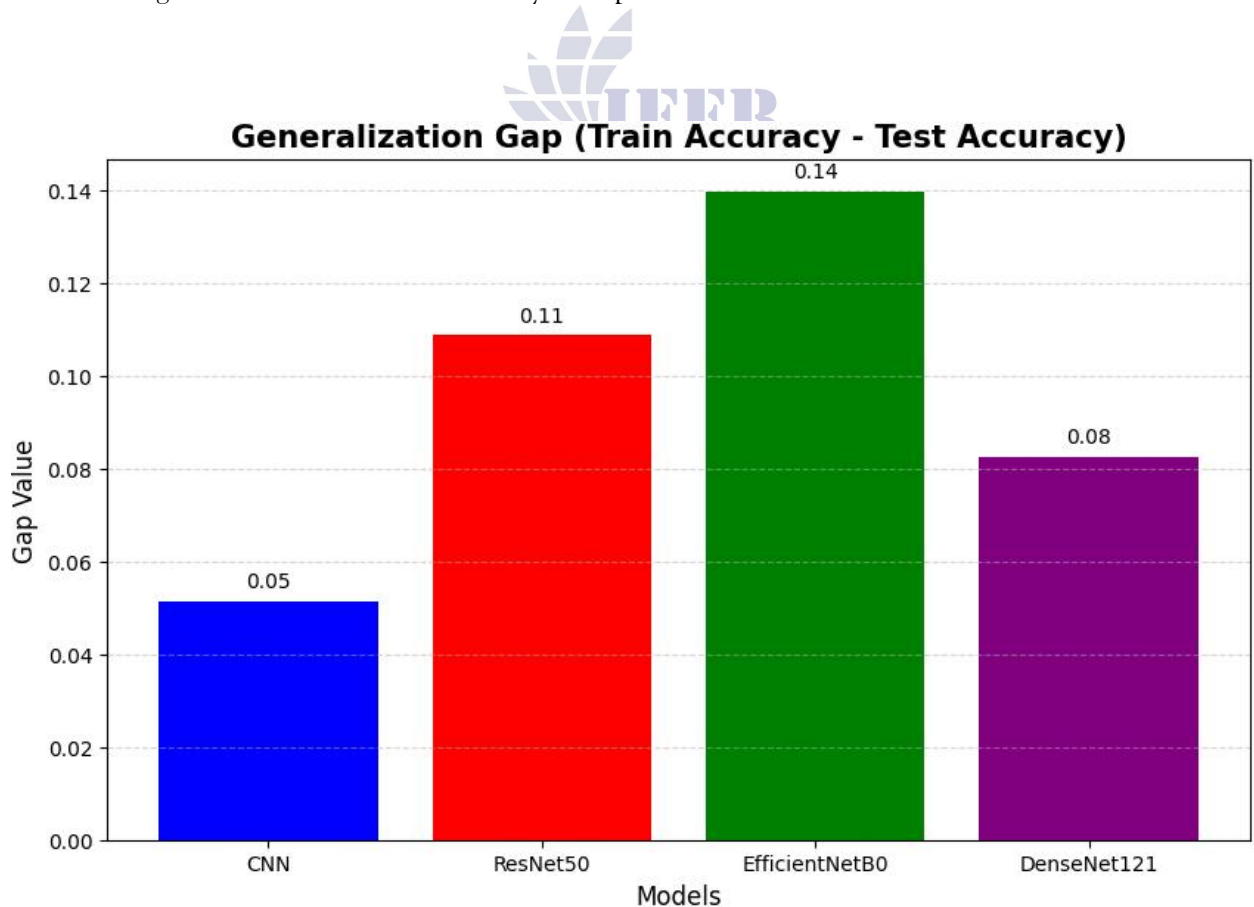


Figure 4.19 Training-Test Accuracy Gap

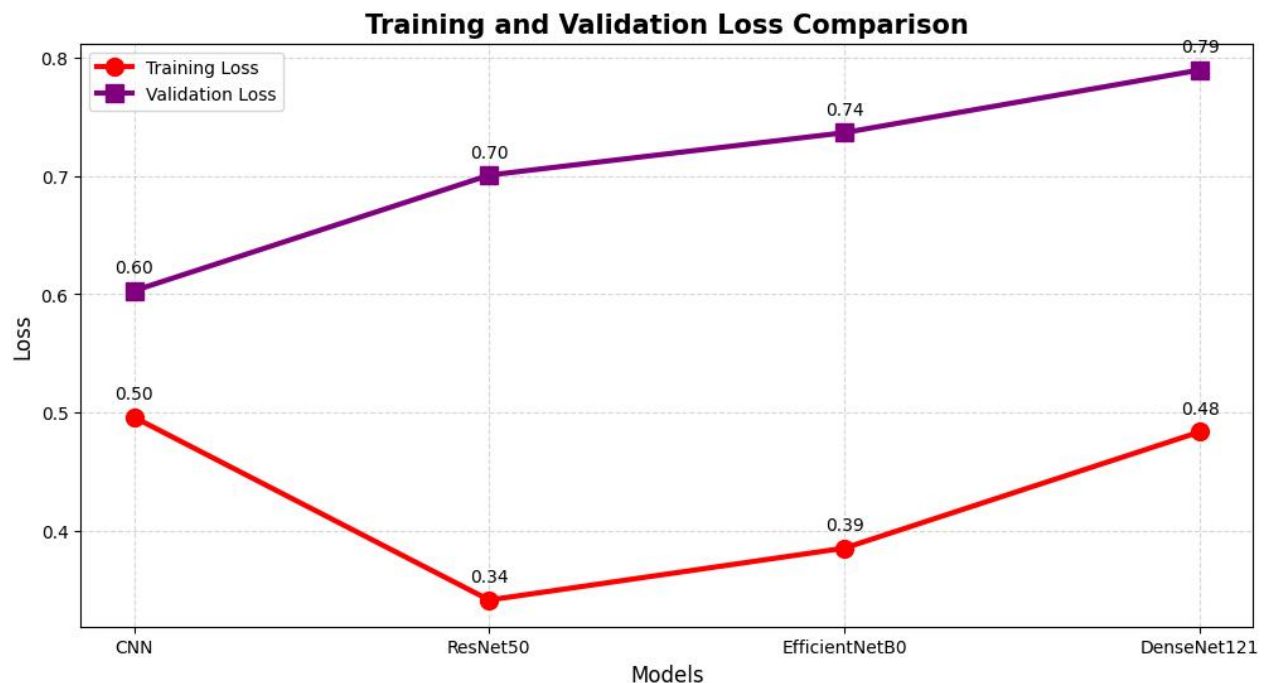


Figure 4.20 Training and Validation Loss Comparison

The overall performance of CNN is better because there is low gap between training and test accuracies indicate no issues of over-fitting but this model is computationally expensive. On the other hand, DenseNet121 indicate limited over-fitting issue as compared to other two pre-trained models. The EfficientNetB0 is computationally more effective than ResNet50 and ResNet50 is more effective than DenseNet121.

#### 4. Conclusion

This aim of this study was to compare four deep learning models CNN, ResNet50, EfficientNetB0, and DenseNet121. All four models were trained on RAF DB dataset to recognize human facial emotions. It was found that all models performed better to identify human emotions with slight variations among different emotion classes. Among all model, ResNet50 produced highest test accuracy and CNN was second highest test accuracy indicated that it had strong capability to generalize. On the other hand, EfficientNetB0 and DenseNet121 also provided better results especially for those classes which had a greater number of images It was noticed models struggled to

recognizes minority classes or visually close emotion like Disgust and Fear classes and light variations in the emotions. Overall, this study confirmed importance of CNN and pre-trained models and their use of real-life images. The further research can be carried out to ameliorate minority classes using augmentation techniques or some larger datasets, this may increase performance of these models.

#### References

- irani, A. O., Ahmed , M. A., Ashraf, M. U., Ilyas , I., Sarwar, N., Alzahrani, A., & Alarood, A. A. (2025). A Novel Facial Expression Recognition Framework Using Deep Learning Based Dynamic Cross-Domain Dual Attention Network. *Peer Journal Computer Science*, 11, 28-66. doi:doi:10.7717/peerj-cs.2866.
- in, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Person Social Psychology*, 17(2), 124-129.
- irau, S., Lamaakal , I., El Hamly , I., Maleh, Ouahbi, I., Pławiak, . . . El-Latif, A. A. (2025). Real-Time Driver Drowsiness Detection Using Facial Analysis and Machine Learning Techniques. *Sensors*, 25(3), 812-831.

- Florestiyanto, Yanu, M., Surjono, H. D., & Jati, H. (2026). Benchmarking Quantum Kernels and Modern Vision Models for Compound Facial Expression Recognition. *Scientific Reports*, *16*(1), 1126-1141. doi:doi:10.1038/s41598-026-41514-2
- Jayaraman, Samanthisvaran, & Mahendran, A. (2025). An Interactive Information Based DCNN-BiLSTM Model with Dual Attention Mechanism for Facial Expression Recognition. *Scientific Reports*, *15*(1), 262-287. doi:doi:10.1038/s41598-025-09709-1
- Jayaswal, Ruchi, Aquib, M. A., Dixit, M., Singh, K. D., & Ahmad, S. (2025). Advances in Facial Expression Recognition Technologies for Emotion Analysis. *Discover Computing*, *28*(1), 203-218. doi:doi:10.1007/s10791-025-09699-8.
- Kara, O., Churamani, N., & Gunes, H. (2021). Towards Fair Affective Robotics: Continual Learning for Mitigating Bias in Facial Expression and Action Unit Recognition. *arXiv*, 92-133.
- Lin, Haobo, Bai, T., Zhang, J. C., Lu, S., Gu, F., . . . Zhang, W. (2026). TAG: Thinking with Action Unit Grounding for Facial Expression Recognition. doi:doi:10.48550/arXiv.2602.18763
- Lin, Haobo, Bai, T., Zhang, J., & Chan, X. (n.d.). . . 2026. "TAG: Thinking with Action Unit Grounding for Facial Expression Recognition. doi:doi:10.48550/arXiv.2602.18763.
- Lv, Gang, & Zhang, J. (2025). Facial Expression Recognition via Variational Inference. doi:doi:10.21203/rs.3.rs-7577763/v1.
- iang, Zhang, J., & Tsoi, C. (2026). Facial Expression Recognition via Variational Inference. *Scientific Reports*, *16*(1), 7323-7345. doi:doi:10.1038/s41598-026-38734-x.
- ra, Gargi, Bajpai, S. S., Jain, R., Jain, D. K., & Štruc., V. (2026). EmoVisioNet: A Hybrid Network Unifying Lightweight CNN and Attention-Based Vision Model for Facial Emotion Detection. *Neurocomputing*, *655*(1), 1322-1334. doi:doi:10.1016/j.neucom.2025.132224.
- , C., Ferreira, A., & Leite, N. (2026). Facial Emotion Recognition: A Comparative Study with Cross-Corpus and Multi-Corpus Training". In *Proceedings of the 15th International Conference on Pattern Recognition Applications and Methods* (pp. 651-658). Marbella Spain: SCITEPRESS - Science and Technology Publications. doi:doi:10.5220/0014639300004067
- ianu, Ipinze, G., Liu, Y., Alamäki, A., & Kauttonen, J. (2024). Benchmarking Deep Facial Expression Recognition: An Extensive Protocol with Balanced Dataset in the Wild. *Engineering Applications of Artificial Intelligence*, 136-163. doi:doi:10.1016/j.engappai.2024.108983.
- unhao, Duan, X., Fan, P., Zhao, Z., & Guo, X. (2025). Combining Fixed-Weight ArcFace Loss and Vision Transformer for Facial Expression Recognition. *Sensors*, *25*(23), 7166-7181. doi:doi:10.3390/s25237166.