World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

An Interdisciplinary Mathematical Optimization and Neural Networks for facial emotion recognition in pharmaceutical and clinical research

Neha Mathur [1, 2, *], Paresh Jain [3] and Pankaj Dadheech [4]

[1] *Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India.*
[2] *Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur, Rajasthan, India.*
[3] *Department of Electrical and Electronics Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India.*
[4] *Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur, Rajasthan, India.*

World Journal of Advanced Engineering Technology and Sciences, 2026, 18(01), 204-211

Publication history: Received on 06 December 2025; revised on 12 January 2026; accepted on 14 January 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.18.1.0011

Abstract

This paper presents a unified framework involving integration of different principles of calculus and transform functions on the deep neural networks and using them for efficient emotion recognition. Facial emotion recognition is the application of significant use in computer vision, the evaluation of pharmaceutical research and mental health related drug development. It's one of the very significant applications in computer science area and also to human-computer interaction, security, and psychological analysis. This research presents an interdisciplinary framework that integrates mathematical optimization techniques with the deep neural networks to optimize learning rate for the ResNet architectures to be better for FER on the benchmark dataset FER2013+ for emotion classification. Trade-offs in the model size with computational efficiency in recognition performance shall be addressed together with feasibility and potential applications to deploy such a ResNet model on resource-limited devices. To measure the models' ability to recognize complex emotional features under resource constraints, experiments were conducted. The findings highlight the potential of optimized deep neural networks as supportive tools in pharmaceutical and healthcare research, particularly for patient-centered studies, real-time emotion monitoring, and data-driven assessment of treatment responses. Higher models, such as ResNet50 and ResNet101, recorded a higher accuracy rate in complicated emotions but relied on more computing resources. ResNet18 and ResNet34 were more efficient and thereby useful in embedded applications. The fit-one-cycle method gave enhanced training efficiency for all the architectures.

Keywords: Emotion Recognition; Behavioral Sciences; System Theory; Computer Science; Neural Emotion Regression; Residual Networks; Differential Networks; Transfer Learning; Transformation Function; Function Equation; Pharmaceutical Technology; Telepharmacy Applications; Recurrent Learning Rate Function

1. Introduction

Emotion recognition involves analyzing the statistical pattern of behavioral sciences. It involves Facial Emotion Recognition (FER) [1] is one of the most important applications of AI and computer vision in that it helps in the automatic detection and classification of human emotions through facial expressions. FER systems enrich human-computer interaction as interfaces are designed to match emotional states. Thus, the interface is intuitive and responsive [2,3]. In security and surveillance, FER systems identify individuals showing tension, angst, or other emotional states that may give potential dangers. Healthcare applications leverage FER for diagnostic and monitoring purposes, especially in mental health, where understanding a patient's emotional state can provide valuable insights

* Corresponding author: Neha Mathur

[4,5]. This paper focuses on benchmarking the ResNet family of deep learning models, popular for their resistance to vanishing gradient issues and ability to train deep networks [6]. ResNet18, ResNet34, ResNet50 [7], and ResNet101 [8] are used in this study in order to investigate the impact of the depth of models on FER performance over the FER2013+ dataset. The problems of variability, efficiency, and model scalability are all addressed in the hope of deriving actionable insights regarding the selection of optimal FER models for application in human-computer interaction, security, and healthcare [9].

2. Literature Review

Ashraf, A. et al., 2021, [10] conducted the classification of Alzheimer's disease by transferring learning using a CNN; this application supplied an accuracy of 99.05% with ADNI features using the DenseNet. In a related fashion, Gerczuk, M. et al., 2021, [11] applied deep transfer learning to the multi-corpus area of speech emotion recognition. Recently, Loey, M., et al., 2021, [12] proposed a novel hybrid model, DL-CNNXGB, which enhanced deep learning with classic machine learning for face mask detection. Due to the combined straightforward CNN-ResNet50 architecture with regression-to-classification conversion and ensemble methods, high accuracy is achieved. Recently, Sahoo, K. K. et al. 2021, [13] TLEFuzzyNet has been reported as a transfer-learning-based pipeline of CNN and presented the state-of-the-art performance on a large number of datasets within speech-based emotion recognition. Xie, B. et al., 2021, [14] explored multimodal approaches to transformer-based emotion recognition, reaching up to 65% accuracy on the MELD dataset. Finally, Dresvyanskiy, D. et al., 2022, [15] demonstrated in-the-wild emotion recognition using audiovisual deep learning and the AffWild2 database. Bashath, S. et al. (2022) [16] outlined some of the difficulties in deep text learning while they presented a new nomenclature in the model of transfer learning, with the models' taxonomy graphically in images. Amin, M. et al. (2022), [17] developed deep transfer learning models in ECG signals used for the detection of stress from drivers. It got a 98.11% precision Xception model. Helaly, R. et al. (2023), [18] developed a deep CNN-based system for facial emotion recognition and reported an improvement to 98% on the CK+ dataset and to 83% on FER2013. Sultana, A. et al. (2023), [19] used deep transfer learning for facial expression recognition and reported 94.8% accuracy on CK+ and 93.7% on JAFFE. Finally, Meena, G. et al. (2023), [20] applied fine-tuned InceptionV3 for image-based [21] sentiment analysis, showing 99.5% accuracy on CK+, 86% on JAFFE, and 73% on FER2013 [22].

3. Research Methodology

It involves four basic models: ResNet18, ResNet34, ResNet50, and ResNet101. The deepest ResNet architecture is only growing in this order with an increment in network capacity and hence complexity. It is based on the concept of optimization of the computational resources and system theory. ResNet18 is the shallowest with only 18 layers and can thus be used for comparison as a baseline model, considering its rather simple structure. ResNet34 is giving a somewhat more complex model with 34 layers but still within computational feasibility. ResNet50 has 50 layers and uses bottleneck blocks for enhancing performance and feature extraction. Finally, ResNet101 consists of 101 layers; it is quite an intricate model to capture the most intricate features from data. Each of the models used pre-trained weights from the ImageNet dataset to take advantage of transfer learning that allows faster training and better initial performance.

3.1. Proposed Approach

Input: FER2013+ dataset, ResNet architectures (ResNet18, ResNet34, ResNet50, ResNet101) .

Output: Model performance metrics (accuracy, precision, recall, F1-score).

Step 1: Data Preprocessing Load the FER2013+ dataset. **Divide dataset into training and validation sets:** Training set: 80% Validation set: 20% Apply data augmentation techniques: Random horizontal flips, rotations, and zooms Normalize images using ImageNet statistics:

$$\text{Normalized} = \frac{\text{Image} - \mu}{\sigma} \qquad (1)$$

whereμ and σ are the mean and standard deviation of the ImageNet dataset, respectively.

Step 2: Model Initialization Initialize ResNet architectures with pre-trained weights from ImageNet:

$$\text{Model}_i (\text{where} i = 18,34,50,101) \qquad (2)$$

Define the ResNet model for each architecture:

$$\text{ResNet}_i = \text{ResNet}(L_i) \qquad (3)$$

whereL_i represents the number of layers for ResNet18, ResNet34, ResNet50, and ResNet101.

Step 3: Training Procedure Set learning rate range using learning rate finder:

$$\text{LR}_{\text{min}}\text{to}\text{LR}_{\text{max}} \quad (4)$$

Train the model using the one-cycle learning rate policy:

$$\text{LR}_{\text{cycle}}(t) = \begin{cases} \frac{t}{T/2} \cdot (\text{LR}_{\text{max}} - \text{LR}_{\text{min}}) + \text{LR}_{\text{min}} & \text{for} 0 \leq t < T/2 \\ \text{LR}_{\text{max}} - \frac{t-T/2}{T/2} \cdot (\text{LR}_{\text{max}} - \text{LR}_{\text{min}}) & \text{for} T/2 \leq t < T \end{cases} \quad (5)$$

where$T$ is the total number of iterations, and t is the current iteration.

Use Adam optimizer for model training:

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} \quad (6)$$

whereα is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the estimates of the first and second moments, and ϵ is a small constant to avoid division by zero.

Step 4: Evaluation Metrics Calculate accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (7)$$

Compute precision, recall, and F1-score for each emotion class:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \quad (10)$$

whereTP is true positives, FP is false positives, and FN is false negatives.

Generate confusion matrix to visualize model performance:

$$\text{Confusion Matrix} = \begin{bmatrix} TP_1 & FP_1 & \cdots \\ FN_1 & TP_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (11)$$

where each element represents the count of true or false classifications for each class.

Step 5: Visualization and Analysis Generate Class Activation Maps (CAMs) to visualize the model's attention:

$$\text{CAM} = \text{ReLU}\big(\text{Global Average Pooling}(\text{Gradients} \times \text{Activations})\big) \quad (12)$$

where gradients are the gradients of the target class with respect to feature maps, and activations are the feature maps from the last convolutional layer.

Step 6: Fit-One-Cycle Policy:

Use the Fit-One-Cycle policy to train the model effectively, which involves:

- Initial Phase: Start with a low learning rate and gradually increase it to a maximum value.
- Second Phase: Reduce the learning rate quickly from the maximum value to a low value.
- Purpose: This approach helps the model converge faster and achieve better performance by adapting the learning rate dynamically.

Step 7: Transfer Learning

- Utilize Transfer Learning: In these pre-trained models are used after that on which various functional equations are applied to make them optimized.
- Leverage pre-trained ResNet models to adapt them to the facial emotion recognition task.
- Fine-tune the models by training on the FER2013+ dataset with a lower learning rate to adjust the pre-trained weights specifically for the new task.
- Evaluate Transfer Learning Effectiveness:
- Compare the performance of models trained from scratch versus those using transfer learning to assess improvements in model accuracy and training efficiency.

Step 8: Comparative Analysis Compare performance metrics (accuracy, precision, recall, F1-score) across ResNet architectures. Analyze computational efficiency and model complexity for each ResNet variant.

4. Result Analysis

4.1. Dataset Used

FER2013+ is a more advanced version of the FER2013, which is a specific dataset for facial emotion recognition. The sign "+" denotes advancements like extra data or better fine-tuned annotation for improved models compared to those of the primary FER2013 dataset showed in figure 1.

Figure 1 FER 2013 Dataset

4.2. Overall Performance of ResNet18

The training loss decreased from 1.63 to 0.53 in more than 25 epochs, and the validation loss also decreased from 1.21 to 0.53. The model is getting more accurate and reaches an accuracy of 80.9% in the last epoch. These training versus validation loss curves are indicative of effective learning and good generalization. The validation loss converges around 0.5, which means that there is not much overfitting. Table 1 shows the Performance in ResNet18.

Table 1 ResNet18 Performance

Case	Precision	Recall	Specificity
1	0.801370	0.757282	0.982096
2	0.428571	0.250000	0.998301
3	0.620690	0.439024	0.998441
4	0.790698	0.545455	0.996093
5	0.930977	0.921815	0.976272
6	0.832130	0.887563	0.896667
7	0.729697	0.695150	0.964211
8	0.855603	0.860238	0.978296

Overall Accuracy: 0.8431731717627167

This would extend to 95 epochs in Phase 2, where the model is still fine-tuning its performance. Here, the training loss decreases more, dropping as low as 0.10, while the validation loss has reached as low as 0.50. The smooth rise of accuracy beyond 84% shows that again some effective learning is taking place with enhanced class discrimination capability. That certainly confirms that the model is indeed effective in learning the facial features required. Figure 2 represents the CAM HeatMap.

Figure 2 CAM HeatMap (ResNet18)

4.3. Overall Performance of ResNet34

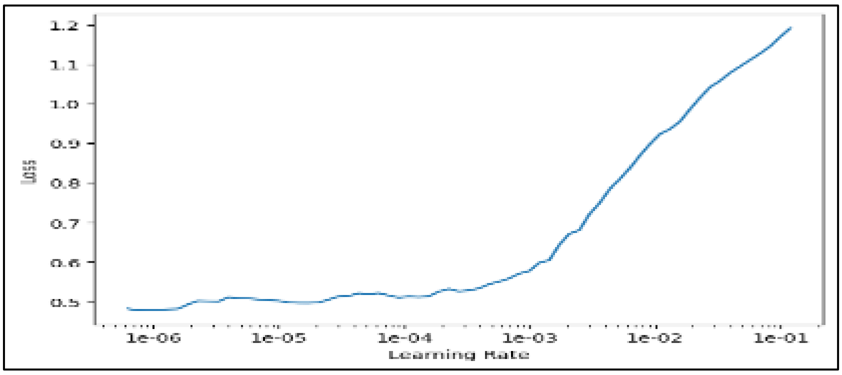

Figure 3 Learning Rate (ResNet34) [Project Implementation]

It's one of the most important intermediate steps in training, as the Learning Rate Finder graph will allow finding an optimum learning rate that the model could learn without overshooting or underfitting at as shown in Figure 3. In this case, an optimal learning rate is when a loss has consistently been declining. It's going to be most likely between le-04 and le-03. This will enable better training of the ResNet34 model by striking a balance between fast learning and stability.

Table 2 ResNet34 Performance

Case	Precision	Recall	Specificity
1	0.790924	0.789644	0.980090
2	0.344828	0.277778	0.997309
3	0.714286	0.487805	0.998866
4	0.760563	0.577540	0.995080
5	0.924878	0.928923	0.973804
6	0.847248	0.871390	0.909333
7	0.715777	0.712471	0.960680
8	0.866081	0.854821	0.980240

Overall Accuracy: 0.8445822178385233

The performance of ResNet34 on FER2013+ seems generally good: in the beginning, the training and validation losses decreased, which is a reflection that it learned effectively as shown in Table 2. Within 94 epochs, the model reached high accuracy, with 84.52% accuracy and also validation developing very well, with increasing trend to 80.8% after the first 24 epochs.

4.4. Performance of ResNet50

The performance of the ResNet50 [23] model, which will be demonstrated in the training data and loss curves, and which indeed depicts a well-optimized training with evident improvements on both the training and validation metrics.

Training loss steadily dropped from 1.50 to 0.33 for 25 epochs, with a constant validation loss, following a similar pattern all the way to its minimum value of 0.46; towards the last, this showed slight increments indicating overfitting. The accuracy was between 59% and 84%. However, in general, analysis of the learning rate has shown that it is between le-02 and le-01 [24] where the performance tends to be best, mainly because such a value allows striking a balance between problems related to underfitting and overfitting. Lastly, the model also demonstrated effective learning in Phase 2; training loss decreased to about 0.007 by the 94th epoch even as fluctuations in validation loss do hint at overfitting at times. It kept improving and reached around 85%, but this large fluctuation in the validation loss indeed suggests that some other techniques, such as regularization or early stopping, would potentially improve the performance. Overall, ResNet50 [25] is robust in terms of learning with a well-tuned learning rate but surely has scope for improvement in handling over fitting with more fine-tuning.

Table 3 ResNet50 Performance

Case	Precision	Recall	Specificity
1	0.814815	0.783172	0.983022
2	0.350000	0.194444	0.998159
3	0.628571	0.536585	0.998158
4	0.769784	0.572193	0.995369
5	0.917297	0.927829	0.970957
6	0.838332	0.882557	0.901778
7	0.733990	0.688222	0.965335
8	0.863089	0.853738	0.979754

Overall Accuracy: 0.8444413132309426

The performance of the ResNet50 [23] model, which will be demonstrated in the training data and loss curves, and which indeed depicts a well-optimized training with evident improvements on both the training and validation metrics. Training loss steadily dropped from 1.50 to 0.33 for 25 epochs, with a constant validation loss, following a similar pattern all the way to its minimum value of 0.46; towards the last, this showed slight increments indicating overfitting. The accuracy was between 59% and 84%. However, in general, analysis of the learning rate has shown that it is between le-02 and le-01 [24] where the performance tends to be best, mainly because such a value allows striking a balance between problems related to underfitting and overfitting. Lastly, the model also demonstrated effective learning in Phase 2; training loss decreased to about 0.007 by the 94th epoch even as fluctuations in validation loss do hint at overfitting at times. It kept improving and reached around 85%, but this large fluctuation in the validation loss indeed suggests that some other techniques, such as regularization or early stopping, would potentially improve the performance. Overall, ResNet50 [25] is robust in terms of learning with a well-tuned learning rate but surely has scope for improvement in handling over fitting with more fine-tuning

5. Discussion

This clearly indicates the trends of a trade-off involving model depth, computation efficiency, and accuracy. For example, ResNet18 provides a very shallow architecture, and hence, its learning curve turns out to be very smooth, and the losses in training and validation decay every iteration, resulting in an accuracy of 80.9%. On the other hand, ResNet34 [26] increased with accuracy but showed an overfitting problem after about 20 epochs and seems to require even more careful regularization. ResNet50 benefits from a deeper architecture and does pretty well, with even 85% accuracy in Phase 2 but shows signs of overfitting. It especially has the validation loss swing wildly. The deepest architecture out of the compared models is ResNet101. It generally performed well: training losses drop noticeably; accuracy goes over 85%. From the analysis above of ResNet18, ResNet34, and ResNet50, it was concluded that, given the case, relative strengths were observed to depend on the adopted metrics. Regarding ResNet18, accuracy came out a little lower compared with others at 0.8432 while the same network beats the rest with regard to some precision cases such as Case 5, attaining a value of 0.930977. Meanwhile, it exhibits very high specificity values for all the cases, especially for Case 2 at 0.998301. ResNet34 had relatively a higher overall accuracy at 0.8446 and really well performed on the precision and recall metrics, especially in Case 5 where it had great precision at 0.924878 and recall of 0.928923. ResNet50, with an overall accuracy of 0.8444, had equal performances on all the metrics. It reaches the highest value for precision on Case 1 (0.814815) and Case 5 (0.917297), with high specificity in most of the cases; for example, Case 2 achieves

0.998159. In summary, ResNet34 has the strongest performance at a slightly higher overall accuracy and with equal well-balanced performance at all precision, recall, and specificity levels.

6. Conclusion and Future Scope

According to performance evaluation, the improvement in the model's depth level, the respective accuracy has enhanced; however, at the cost of the overall computational efficiency because the computational power of ResNet18, ResNet34, ResNet50, and ResNet101 to identify facial expressions are reduced. The complicated architecture of ResNet50 and ResNet101 gives it a higher accuracy together with higher capability of achieving more complex features of emotions. However, they have had more critical overfitting and heavy computation, which can be reflected in the drastic fluctuation of validation loss. ResNet18 and ResNet34 do show better balanced trade-offs between the two but also exemplify their insufficiency for cases with nuanced emotional expressions and across longer epochs for generalization. Again, this shows that additional optimization techniques are also important by including dynamic learning rates and also regularization methods with which the strength of the robustness of this model can improved and enhanced.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Badshah, A. M., Ahmad, J., Rahim, N., & Baik, S. W. (2017, February). Speech emotion recognition from spectrograms with deep convolutional neural network. In *2017 international conference on platform technology and service (PlatCon)* (pp. 1-5). IEEE.

[2] Ntalampiras, S. (2017). A transfer learning framework for predicting the emotional content of generalized sound events. *The Journal of the Acoustical Society of America*, *141*(3), 1694-1701.

[3] Abanoz, H., & Çataltepe, Z. (2018, May). Emotion recognition on static images using deep transfer learning and ensembling. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

[4] Lee, M. K., Kim, D. H., Choi, D. Y., & Song, B. C. (2018, June). Deep transfer learning for emotion recognition networks. In *2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)* (pp. 206-212). IEEE.

[5] Banerjee, D., Islam, K., Xue, K., Mei, G., Xiao, L., Zhang, G., ... & Li, J. (2019). A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. *Knowledge and Information Systems*, *60*, 1693-1724.

[6] Hung, J. C., Lin, K. C., & Lai, N. X. (2019). Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing*, *84*, 105724.

[7] Yokoo, K., Atsumi, M., Tanaka, K., Haoqing, W. A. N. G., & Lin, M. E. N. G. (2020, December). Deep learning based emotion recognition iot system. In *2020 International Conference on Advanced Mechatronic Systems (ICAMechS)* (pp. 203-207). IEEE.

[8] Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S. U., & Rodrigues, J. J. (2020). LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19. *IEEE Internet of Things Journal*, *8*(23), 16863-16871.

[9] Ahmad, Z., Jindal, R., Ekbal, A., & Bhattachharyya, P. (2020). Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding. *Expert Systems with Applications*, *139*, 112851.

[10] Ashraf, A., Naz, S., Shirazi, S. H., Razzak, I., & Parsad, M. (2021). Deep transfer learning for alzheimer neurological disorder detection. *Multimedia Tools and Applications*, 1-26.

[11] Gerczuk, M., Amiriparian, S., Ottl, S., & Schuller, B. W. (2021). Emonet: A transfer learning framework for multi-corpus speech emotion recognition. *IEEE Transactions on Affective Computing*.

[12] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, *167*, 108288.

[13] Sahoo, K. K., Dutta, I., Ijaz, M. F., Woźniak, M., & Singh, P. K. (2021). TLEFuzzyNet: Fuzzy rank-based ensemble of transfer learning models for emotion recognition from human speeches. *IEEE Access*, *9*, 166518-166530.

[14] Xie, B., Sidulova, M., & Park, C. H. (2021). Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. *Sensors*, *21*(14), 4913.

[15] Dresvyanskiy, D., Ryumina, E., Kaya, H., Markitantov, M., Karpov, A., & Minker, W. (2022). End-to-end modeling and transfer learning for audiovisual emotion recognition in-the-wild. *Multimodal Technologies and Interaction*, *6*(2), 11.

[16] Bashath, S., Perera, N., Tripathi, S., Manjang, K., Dehmer, M., & Streib, F. E. (2022). A data-centric review of deep transfer learning with applications to text data. *Information Sciences*, *585*, 498-528.

[17] Amin, M., Ullah, K., Asif, M., Waheed, A., Haq, S. U., Zareei, M., & Biswal, R. R. (2022). ECG-Based Driver's Stress Detection Using Deep Transfer Learning and Fuzzy Logic Approaches. *IEEE Access*, *10*, 29788-29809.

[18] Helaly, R., Messaoud, S., Bouaafia, S., Hajjaji, M. A., & Mtibaa, A. (2023). DTL-I-ResNet18: facial emotion recognition based on deep transfer learning and improved ResNet18. *Signal, Image and Video Processing*, 1-14.

[19] Sultana, A., Dey, S. K., & Rahman, M. A. (2023). Facial emotion recognition based on deep transfer learning approach. *Multimedia Tools and Applications*, 1-15.

[20] Meena, G., Mohbey, K. K., Kumar, S., Chawda, R. K., & Gaikwad, S. V. (2023). Image-based sentiment analysis using InceptionV3 transfer learning approach. *SN Computer Science*, *4*(3), 242.

[21] Taru, Sneha, et al. "Emotion Recognition System from Facial Expressions Using Machine Learning." Journal of Artificial Intelligence and Capsule Networks 7.2 (2025): 125-136.

[22] Pradesh, Uttar, and Bodh Gaya. "Facial emotion recognition using neural networks." Journal of Information & Optimization Sciences ISSN 0252-2667 (Print), ISSN 2169-0103 (Online) Vol. 44 (2023), No. 6, pp. 1115–1126

[23] Venkatraman, S., Sharma, V., Malarvannan, S. and Narendra, M.,"Multimodal emotion recognition using audio-video transformer fusion with cross attention". arXiv preprint arXiv:2407.18552, 2024

[24] Aktürk, K. and Keçeli, A.S., Deep operational audio-visual emotion recognition. Neurocomputing, 588, p. 127713, 2024

[25] Fan, S., Jing, J. and Wang, C., "Audio-Visual Learning for Multimodal Emotion Recognition". Symmetry, 17(3), p.418, 2025

[26] Sun, Z., Liu, H., Li, H., Li, Y. and Zhang, W., AVERFormer: "End-to-end audio-visual emotion recognition transformer framework with balanced modal contributions". Digital Signal Processing, 161, p. 105081. 2025.