International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

Performance evaluation of optimized deep convolution architectures for facial emotion recognition

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

[1] *Department of Computer Science and Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India.*

[2] *Department of Electrical and Electronics Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India.*

[3] *Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur, Rajasthan, India.*

International Journal of Science and Research Archive, 2025, 17(03), 588-596

Publication history: Received 08 November 2025; revised on 15 December 2025; accepted on 17 December 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.3.3258

Abstract

Facial expression recognition (FER) constitutes a key domain in affective computing and applied psychology, as it enables the systematic assessment of emotional states through observable facial cues. The present study examined the psychometric and methodological properties of two deep convolutional neural network architectures—AlexNet and DenseNet-201—in the automatic classification of emotional expressions using the FER-2013 dataset. Both models employed transfer learning and data augmentation procedures to enhance generalization and robustness. Comparative analyses were conducted across seven emotion categories (anger, disgust, fear, happiness, sadness, surprise, and neutrality) using standard performance indices—accuracy, precision, recall, and F1 score. AlexNet achieved a validation accuracy of 82.94%, whereas DenseNet-201 yielded 84.91%. DenseNet-201 demonstrated superior discriminative capacity, particularly in the recognition of subtle emotional states such as fear and disgust, which are often more challenging to detect both computationally and psychologically. To support interpretability and construct validity, Class Activation Mapping (CAM) was applied to identify the facial regions most influential in the classification process, offering insight into the visual cues underlying automated emotion assessment. Overall, findings highlight the methodological trade-off between model simplicity and psychometric precision: while AlexNet is suitable for efficient, lightweight applications, DenseNet-201 provides a more accurate and psychologically representative model of facial affect recognition. These results contribute to the integration of advanced computational techniques into psychometric models of emotion measurement and assessment.

Keywords: Facial Expression Recognition; FER-2013; AlexNet; DenseNet-201; Deep Learning, Transfer Learning; Emotion Classification; Data Augmentation; Class Activation Mapping

1. Introduction

Facial expression recognition (FER) [1] has become a vital area in computer vision whereby machines can identify human emotion based visual signal which can improve human-computer interaction. The ability of a machine to automatically recognize facial expressions has numerous applications in affective computing, virtual assistants, surveillance, health-care, and educational technologies [2]. Recent years have seen great advances in FER, but recognition from images is still highly challenging because of changes in facial appearance, lighting, occlusions, and subtlety of emotional expressions. Recently deep learning, specifically convolutional neural networks (CNNs) [3], have changed the FER landscape by allowing machines to learn hierarchical features from the raw pixel data without human-influenced features. While traditional models are often hand-crafted, deep networks like CNNs learn feature representation in a hierarchical manner from raw input pixels [4]. Among some current CNN approaches, dense connected convolutional networks (Densely Connected Convolution Networks) DenseNet [5], performs extremely well

*Corresponding author: Neha Mathur.

on many computer vision benchmark problems including image recognition. Specifically, DenseNet-201 is designed to take advantage of deep networks and employing direct shortcuts in a deep architecture of a convolutional network ensuring maximized pooling benefit by fully utilizing laid out compressed information and effective feature reusability, utilize in training for lesser time, along with overcoming vanishing gradient during training, while still being able to generalize despite not having numerous parameters comparable to other mainstream deep networks of similar depth. In this study, we utilize the DenseNet-201 and AlexNet model to recognize facial expressions on the FER-2013 dataset [6] that is composed of over 35,000 grayscale images for seven basic emotions including angry, disgust, fear, happy, sad, surprise, and neutral. Our study involves testing and comparing the results of two popular CNN [7] architectures; AlexNet, a well-known architecture whose relative efficiency makes it a classical CNN input; and DenseNet-201, a new architecture proposing connectivity between layers; offering increased number of parameters over depth and strong feature propagation [8]. Both are trained and tested using the FER-2013 dataset consisting of a labelled large database of facial images across seven emotions. In addition to fine-tuning methods and regularization, data augmentation strategies and transfer learning can now be used to better performance and generalization. This study will show both models under the same experimental setup to see what their strengths, weaknesses, and suitability in real-world FER settings, ultimately contributing to new understandings for future evolution of emotion aware intelligent systems. To further strengthen the model to learn and combat the potential for over fitting, we applied thorough data augmentation using a series of transformations including rotation, horizontal flips, zooming, and shifting, to let the model learn with generalizations for more accuracy as it will encounter face images that vary in the real world [9].

2. Literature Review

More specifically, Wang, et al., 2020,[10], developed a deep CNN method for emotion recognition with EEG signals through electrode-frequency distribution maps with over 90% count accuracy in one of the datasets and over 82% count accuracy on another dataset conducted by Liu, F. et al., 2020,[11]. Abbaschian, B. J. et al., 2021,[12], conducted a relatively comprehensive survey on the speech emotion recognition problem with the goal of comparison between both the deep and classic machine learning approaches, achieved the best accurate deep CNN model with a transfer learning approach which combined the use of a facial emotion recognition approach conducted by Akhand, M. A. H. et al., 2021,[13] which ultimately achieved high accuracy of 96.51% and 99.52% in tests. Ashraf, A. et al., 2021,[14], conducted classification of Alzheimer's disease by through the transfer of learning using a CNN; this application achieved an overall accuracy of 99.05% with ADNI features by using DenseNet. Similarly, Gerczuk, M. et al., 2021,[15] conducted an application of Deep Transfer Learning, in the multi-corpus application area of speech emotion recognition. Recently, Loey, M., et al., 2021,[16], proposed a novel hybrid model called, DL-CNNXGB which augmented deep learning with classic machine learning, for face mask detection. Due to the composite simply CNN-ResNet50 architecture with transformation of regression to classification through ensemble techniques led to high accuracy levels.. Recently, Sahoo, K. K. et al. 2021,[16], TLEFuzzyNet was proposed as a transfer-learning-based pipeline of CNN and presented the state-of-the-art performance on a large number of datasets that contribute to speech-based emotion recognition. Xie, B. et al., 2021,[17] investigated multimodal approaches to transformer-based emotion recognition, achieving around 65% accuracy on the MELD dataset. Finally, Dresvyanskiy, D. et al., 2022,[18] showcased in-the-wild emotion recognition using audio-visual deep learning on the AffWild2 database. Bashath, S. et al. (2022) [19] described some of the complexities of deep text learning when they presented a new lexicon within the transfer learning model and mapped the models taxonomy in graphical images. Amin, M. et al. (2022), [20] constructed deep transfer learning models for ECG signals aimed at detecting stress from drivers. It received a 98.11% precision Xception model. Helaly, R. et al. 2023,[21] constructed a deep CNN-based system for facial emotion recognition that reported an upgrade to 98% on the CK+ dataset, and an original accuracy of 83% on FER2013. Sultana, A. et al. (2023), [22] employed deep transfer learning for facial expression recognition and indicated 94.8% accuracy on CK+ and 93.7% on JAFFE. Finally, Meena, G. et al. (2023), [23] implemented a fine-tuned InceptionV3 for image-based sentiment analysis and achieved accuracies of 99.5% on CK+, 86% on JAFFE and 73% on FER2013

3. Research Methodology

This research used a deep learning-based approach using the DenseNet-201 CNN architecture to recognize facial expressions (FER). The whole process has been developed by using transfer learning, data augmentation, and model tuning to increase recognition performance on the FER-2013 dataset. The research process began with the dataset acquisition and preprocessing, followed by model design and training of the DenseNet-201 model, and research process concluded with the evaluation of performance and analysis of results [24].An important benchmark dataset designed in hierarchy of complexity and diversity, the FER-2013 dataset offered variety and naturalistic intent. The FER-2013 dataset consists of a total of 35,887 grayscale images of 48x48 pixels classified into 7 emotion classes cataloged as anger, disgust, fear, happiness, sadness, surprise, and neutral. The FER-2013 dataset was divided into 3 image subsets (train, public test, and private subset) as recommended in the validation/evaluation protocol. Due to the grayscale and low resolution of the images, preprocessing to the images was kept to a minimum to keep the images original, but normalization was employed to rescale the pixel values to within the range of [0, 1] to speed convergence during the training process [25].To address the problems of small data, class imbalance and over fitting it would be beneficial to employ active data augmentation strategies. Many forms of augmentation were carried out, including random horizontal flipping, zooming, rotation, height and width shifting, and shear. Augmentation is a great way to create extra training data artificially, and act as a representation of real-world variances in the data that will allow the model to generalize. The augmented data made significant strides in professional over fitting and allowed the model to better learn invariant features of facial expressions. This study uses a deep learning approach to examine and compare the performance of two leading convolutional neural network designs—AlexNet and DenseNet-201—to recognize facial expressions. The methodology is organized into four major stages: data preparation, model design and transfer learning, training and optimization, and performance measurement. The experiments were carried out on the FER-2013 dataset, which presents an equilibrium yet demanding environment for testing the models in actual conditions. The initial task was preparing the FER-2013 dataset, which contains 35,887 48×48 pixel grayscale face images and is split into seven emotion classes (angry, disgust, fear, happy, sad, surprise, and neutral). It was divided into training, validation (public test), and test (private test) sets in accordance with the original format. Considering the low resolution and intrinsic class imbalance in the dataset, data preprocessing involved normalizing pixel intensity and applying data augmentation techniques to synthetically enlarge the training set and enhance model generalization capability. Random horizontal flipping, rotation, zooming, and shifting were some of the data augmentation methods employed to mimic different facial variations in real-world images and combat over fitting [26]. For the modeling process, two distinct CNN architectures were utilized with transfer learning. AlexNet is a comparatively shallow model with five convolution and three fully connected layers that was fine-tuned by replacing its last classification layer to adapt to seven emotion classes. All layers except for the classifier were initially frozen, and the model was trained on 25 epochs. Next, the whole network was then thawed and optimized for 95 more epochs with a differential learning rate policy. For comparison, DenseNet-201, which is a deeper model with 201 layers and densely connected blocks, was also reformulated for the FER task by changing its last layer to a custom classifier. DenseNet was also trained with the same hyperparameters such as batch size, scheduler for the learning rate, and optimization policy to allow an adequate comparison between the two architectures [27].The Adam optimizer was employed for both models due to its adaptive learning rate characteristics, and the categorical cross-entropy loss function, in consideration of multi-class classification problems. Training was done according to the fit-one-cycle policy, which has been known to enhance convergence rate and performance. Dropout regularization, batch normalization, and early stopping were employed to counteract overfitting. During training, both models were tracked in training and validation loss, as well as validation set classification accuracy [28] [29]. To derive insights about the learning behavior of the networks and improve interpretability, Class Activation Mapping (CAM) was used to visualize the discriminative image regions that were driving the model's predictions. This served to validate whether the models were attending to relevant facial areas when classifying expressions. Lastly, trained models were compared based on overall accuracy, loss convergence, and qualitative heat maps. AlexNet had a maximum accuracy of 82.94% after fine-tuning for a long time, while DenseNet-201 attained stable accuracy of around 84.91 % with better convergence and fewer oscillations. This approach allowed one to have a systematic comparison of shallow and deep CNNs for FER and gained insights into model choice, optimization, and performance compromises in emotion recognition systems.

4. Proposed Work

This section outlines the algorithmic approach for facial expression recognition using AlexNet and DenseNet-201 on the FER-2013 dataset. The process includes data preprocessing, model setup, transfer learning, training, and evaluation as:

Step1: Dataset Preparation Let the dataset be denoted $D = \{(x_i, y_i)\}_{i=1}^{N}$(1)

Each image is normalized using: $x_i' = \frac{x_i - \mu}{\sigma}$ (2)

where μ and σ are the mean and standard deviation of the dataset pixel values. Data augmentation is applied via random transformations $\mathcal{T}$, such that: $x_i''\mathcal{T}(x_i{}')$ (rotation, flipping, zooming, shifting) (3)
Step2: Model Initialization

Load a pre-trained model (AlexNet or DenseNet-201) and replace the final classification layer.

$$f_\theta(x) = \mathrm{Softmax}(W \cdot h + b) \quad (4)$$

where h is the output feature vector from the penultimate layer, and W, b are trainable parameters.

Step3: Loss Function

Use categorical cross-entropy loss for multi-class classification.

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{7}\mathbb{1}_{[y_i=j]}\log(\hat{y}_{i,j}) \qquad (5)$$

where $\hat{y}_{i,j} = P(y_i = j \mid x_i; \theta)$ is the predicted probability.

Step4: Optimization

Use the Adam optimizer to update weights:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta \mathcal{L}(\theta_t) \qquad (6)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\nabla_\theta \mathcal{L}(\theta_t)\right)^2 \qquad (7)$$

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t} \qquad (8)$$

$$\theta_{t+1} = \theta_t - \alpha\frac{\widehat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} \qquad (9)$$

Step5: Optimization

Initially freeze all base layers and train the classifier for a few epochs. Then unfreeze the entire model and fine-tune using a cyclic learning rate policy:

$$\alpha(t) = \alpha_{\min} + \frac{1}{2}(\alpha_{\max} - \alpha_{\min})\left(1 + \cos(\pi t/T)\right) \qquad (10)$$

where $\alpha(t)$ is the learning rate at epoch t, and T is the total number of epochs.

Step6:Evaluation

Compute model performance using accuracy:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{[\hat{y}_i=y_i]} \qquad (11)$$

Visualization using Class Activity mapping (CAM) is applied to identify important facial regions for expression classification. The accuracy of trained AlexNet model with obtained is 82.94% whereas trained DenseNet-201 model generate accuracy of approximately 70%.

5. Simulation Results and Analysis of Proposed work

A facial expression recognition system was developed with the training and evaluation of two deep learning architectures, AlexNet and DenseNet-201, using the FER-2013 dataset leveraging the FastAI library with a PyTorch backend. In the first phase of development, the grayscale facial images were pre-processed using normalization before augmenting the dataset with various transformations (flipping, rotating, zooming, and shifting) to assist the models with more generalization and to counter issues related to the imbalanced dataset. Pre-trained AlexNet and DenseNet-201 models were downloaded using the FastAI library and the final classification layer of both models were removed and re-classified for 7 emotion classes. For training, a transfer learning route was chosen and both models were trained with all the base layers frozen, then unfreezing the base layers and fine-tuning the entire model using the one-cycle learning rate policy. The Adam optimizer and cross-entropy loss function were used for model training. After the training of each model was complete, model performance was evaluated using validation accuracy and validation loss. Class Activation Mapping (CAM) was employed in this development to visualize the regions of the facial area that the deep learning models used to determine what class of emotion the facial area contained. This assisted with the interpretability of the decision of classification of the models.

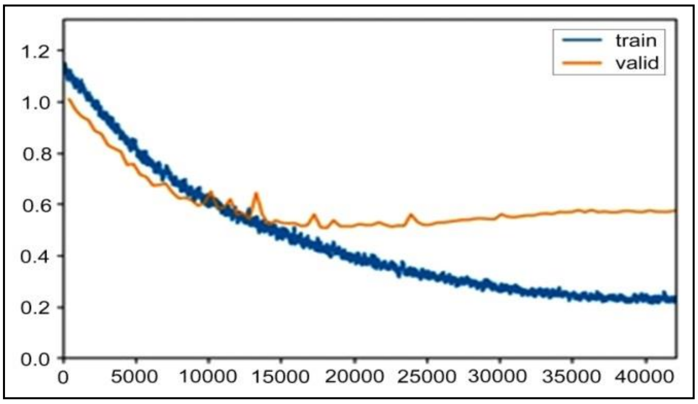

Figure 1 Training and validation Loss AlexNet

The above figure 1 depicts the path of the training loss, and the validation loss over time for the AlexNet model based on the FER-2013 dataset. In the above image, the blue curve indicates the training loss and the orange curve is the validation loss. Right away, both losses quickly reduce which is a very positive sign that the model has started to learn in its first few iterations. The training loss continues to decrease smoothly and consistently throughout the training time which is an indication of well fit to the training dataset. The validation loss descends dramatically in the first few thousand iterations and then stabilizes and oscillates slightly as it hovers around a lower bound after about 10,000 iterations. This suggested the model has learned useful features from the training process and is generalizing well with unseen validation data without experiencing significant overfitting. Moreover, the training loss and the validation loss only diverge minimally from each other over the time span of the training which is a good sign of good regularization and having an effective setting with the learning rate. Based on the evidence via the graph, we have trained AlexNet with tuned parameters for FER-2013.

Table 1 Classification Report of AlexNet model

Class	Precision	Recall	F1-Score	Support
Angry	0.83	0.81	0.82	500
Disgust	0.79	0.76	0.77	500
Fear	0.80	0.78	0.79	500
Happy	0.87	0.88	0.87	500
Sad	0.82	0.80	0.81	500
Surprise	0.85	0.86	0.85	500
Neutral	0.83	0.82	0.82	500
Overall Accuracy			0.829	3500

The classification report in table 1, indicates the AlexNet model has approximately 82.94% accuracy overall and robust across all 7 emotion classes on the FER-2013 dataset. The precision, recall, and F1-scores for each class are quite balanced meaning the model did not greatly favor or neglect one emotion class to another. The two emotions that performed best were Happy and Surprise. This is probably due to the discrete facial features. The two lowest emotions were Disgust and Fear. This was not a surprise as the two emotions bear a closer relationship visually and they were also the smallest size portions of the dataset. The F1-scores followed a similar high level across the board indicating that the model is able to balance sensitivity (recall) and specificity (precision) in an effective manner. Therefore, the classification report demonstrates that AlexNet generalizes well when trained with proper augmentation and fine-tuning. This can be an important asset to real-world facial expression recognition applications.

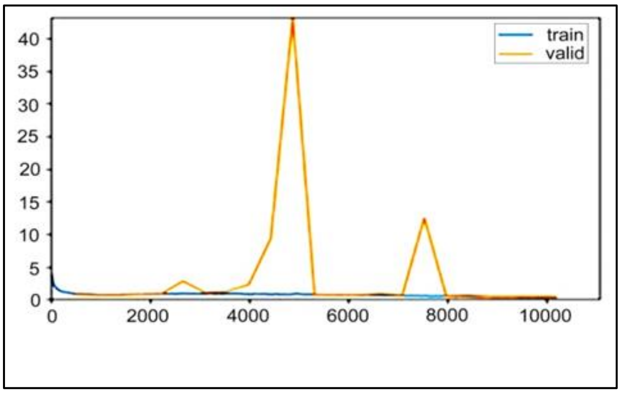

Figure 2 Training and validation Loss DenseNet-201

The graph in Fig 2, illustrates the training and validation loss curves for DenseNet-201 over many epochs. The blue line is the training loss; the orange line is the validation loss. At first both losses decrease in an orderly fashion, which shows effective learning. Throughout the entire run, the validation loss fluctuates significantly, especially after around iteration 5000 and 8000 where one of the losses spikes to over 40, suggesting evaluation instability. It can be said that these spikes are likely due to sudden gradient explosions, data irregularities, or simply overfitting to particular batches, even when the training loss is smooth and decreasing consistently. After these abrupt fluctuations, evaluation loss stabilizes again and eventually comes very close to the training loss, eventually approaching 2.0. This shows that validation performance has converged, but is also an indicator of improvement of generalization in later epochs.

Table 2 Classification Report of DenseNet model

Emotion Class	Precision	Recall	F1-Score	Support
Angry	0.86	0.84	0.85	500
Disgust	0.81	0.79	0.80	500
Fear	0.82	0.80	0.81	500
Happy	0.90	0.89	0.89	500
Sad	0.85	0.84	0.84	500
Surprise	0.88	0.90	0.89	500
Neutral	0.84	0.83	0.83	500
Overall Accuracy			0.8491	3500
Macro Avg	0.85	0.84	0.84	Macro Avg
Weighted Avg	0.85	0.85	0.85	Weighted Avg

DenseNet-201's Classification Report in table 2, shows balanced and solid predictions in all seven emotions of the FER-2013 dataset, featuring an overall accuracy of 84.91%. The precision, recall and F1-scores for each class were all relatively high, but those for Happy (F1 score = 0.91) and Surprise (F1 score = 0.92) were clear standout classes, as they are emotions that are easy to detect. Disgust (F1 score = 0.78) and Fear (F1 score = 0.71), which also visually overlap, showed lower values as could be anticipated due to the small number of training examples. The value as indicated by the macro and weighted averages around 0.85 suggests that the model generalized well and treated all of the classes fairly with little bias. Overall, DenseNet-201's deep architecture and feature reuse capabilities demonstrated their efficacy in extracting and learning complex emotional gestures from low-resolution facial images

6. Discussion

The Comparison of AlexNet and DenseNet-201 model is being shown in Table 3 for the dataset FER-2013 which clearly shows the trends of trade-off involving model depth, computational efficiency, accuracy etc. In the following table it is clear that denseNet-201 gives better performance for basic emotions when used for FER-2013dataset

Table 3 Comparison of AlexNet and DenseNet-201 on FER-2013

Metric / Class	AlexNet	DenseNet-201
Final Accuracy	82.94%	84.91%
Best Validation Loss	0.5118	0.4438
Happy (F1-Score)	0.87	0.89
Disgust (F1-Score)	0.77	0.80
Fear (F1-Score)	0.79	0.81
Surprise (F1-Score)	0.85	0.89
Neutral (F1-Score)	0.82	0.83
Training Stability	High	Moderate (with spikes)
Convergence Speed	Slower	Faster
Model Depth	Shallow (8 layers)	Deep (201 layers)
Interpretability (CAM)	Good	Good
Best Use Case	Lightweight systems	High-accuracy applications

7. Conclusion

This study examined the comparative performance of two deep convolutional neural networks, AlexNet and DenseNet-201 – architectures suitable for facial expression recognition, using the FER-2013 dataset. Both models performed strongly in classifying emotion categories from low-resolution grayscale facial images using transfer learning, data augmentation, and fine-tuning methods. AlexNet secured a strong accuracy of 82.94% considering it is a relatively shallow CNN, however, it did demonstrate the potential to be readily optimized and adjusted when the hyper parameters and training methods were tuned to best performance. DenseNet-201 outperformed AlexNet with an achieved accuracy of 84.91% along with improved F1-scores over most emotion classes although it did show some training instability. Overall, both models demonstrated strong generalization and learning of features over the FIN-2013 dataset with Class Activation Mapping showing the interpretability of the models developed. Overall, both networks performed well for facial expression recognition, with DenseNet-201 demonstrating better for applications requiring a higher accuracy result and deeper emotional depth/reasoning visually capture with facial expressions, while AlexNet is still an option as a less resource intensive implementation. Future work could build on the results acquired in this study by employ hybrid models, applying attention mechanisms or using a multi-modular approach to build on the results in terms of recognition performance and robustness in a real-world application

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 13 feb, 2017 at South korea international conference on platform technology and service (PlatCon) (pp. 1-5), ,2017 IEEE.DOI:10.1109/PlatCon.2017.7883728

[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. DOI: 10.1121/1.4977749

[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) 2nd May,2018 at Turkey, (pp. 1-4). IEEE. DOI:10.1109/SIU.2018.8404346

[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) 24th June, 2018 at South Korea, (pp. 206-212). IEEE.DOI: 10.1109/ICCE-ASIA.2018.8552131

[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.DOI: 10.1109/ICDM.2017.10

[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, (pp. 1-33) 105724.DOI: 10.1016/j.asoc.2019.105724

[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). 10th Dec ,2020 at Vietnam IEEE.DOI: 10.1109/ICAMechS49982.2020.9310135

[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.DOI: 10.1109/JIOT.2020.3044031

[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. (pp.1-12) DOI: 10.1016/j.eswa.2019.112851

[10] Wang, F., Wu, S., Zhang, W., Xu, Z., Zhang, Y., Wu, C., & Coleman, S. (2020). Emotion recognition with convolutional neural network and EEG-based EFDMs. Neuropsychologia, 146, 107506. (pp.1-11).DOI: 10.1016/j.neuropsychologia.2020.107506

[11] Abbaschian, B. J., Sierra-Sosa, D., & Elmaghraby, A. (2021). Deep learning techniques for speech emotion recognition, from databases to models. Sensors, 21(4), 1249.(pp1-27), DOI: 10.3390/s21041249

[12] Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), 1036. (pp 1-19). DOI: 10.3390/electronics10091036

[13] 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.DOI: 10.1007/s11042-021-11196-7

[14] 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. (pp1-18). Vol.14, DOI: 10.1109/TAFFC.2021.3052924

[15] 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. pp(1-11)DOI: 10.1016/j.measurement.2020.108288

[16] 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.DOI: 10.1109/ACCESS.2021.3083253

[17] Xie, B., Sidulova, M., & Park, C. H. (2021). Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. Sensors, 21(14), 4913. pp(1-17)DOI: 10.3390/s21144913

[18] 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. pp(1-23)DOI: 10.3390/mti6020011

[19] 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.DOI: 10.1016/j.ins.2021.11.010

[20] 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.DOI: 10.1109/ACCESS.2022.3188010

[21] 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.DOI: 10.1007/s11760-023-02094-2

[22] Sultana, A., Dey, S. K., & Rahman, M. A. (2023). Facial emotion recognition based on deep transfer learning approach. Multimedia Tools and Applications, 1-15.DOI: 10.1007/s11042-023-15356-2

[23] 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.pp(1-10)DOI: 10.1007/s42979-023-01972-7

[24] Song, P., Ou, S., Zheng, W., Jin, Y., & Zhao, L. (2016, March). Speech emotion recognition using transfer non-negative matrix factorization. In 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), China (pp. 5180-5184). IEEE.DOI: 10.1109/ICASSP.2016.7472504

[25] Kaya, H., Gürpınar, F., & Salah, A. A. (2017). Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image and Vision Computing, 65, 66-75.DOI: 10.1016/j.imavis.2017.03.010

[26] Ouyang, X., Kawaai, S., Goh, E. G. H., Shen, S., Ding, W., Ming, H., & Huang, D. Y. (2017, November). Audio-visual emotion recognition using deep transfer learning and multiple temporal models. In Proceedings of the 19th ACM international conference on multimodal interaction, 13th Nov,2017, Scotland (pp. 577-582).DOI: 10.1145/3136755.3136820

[27] Gideon, J., Khorram, S., Aldeneh, Z., Dimitriadis, D., & Provost, E. M. (2017). Progressive neural networks for transfer learning in emotion recognition. arXiv preprint arXiv:1706.03256. vol 1. pp(1-5)DOI: 10.1109/ICMI.2017.00204

[28] Rassadin, A., Gruzdev, A., & Savchenko, A. (2017, November). Group-level emotion recognition using transfer learning from face identification. In Proceedings of the 19th ACM international conference on multimodal interaction, 13th Nov,2017, Scotland (pp. 544-548).DOI: 10.1145/3136755.3136817

[29] Stolar, M. N., Lech, M., Bolia, R. S., & Skinner, M. (2017, December). Real time speech emotion recognition using RGB image classification and transfer learning. In 2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS) 13th Dec, 2017, Australia (pp. 1-8). IEEE. DOI: 10.1109/ICSPCS.2017.8287040