Human Emotions : Enhancing facial emotions through facial expressions using Deep Learning.

Under Guidance of **Dr. Jeevana Jyothi Pujari**

Our Team :
(Team 13)
Mandadi Pavan Sai - 23BCE9626
Raya Raghuvaran - 23BCE9386
Shaik Rajabhaigari Mahaboob - 23BCE9426
Kommudasari Manoj Kumar - 23BCE20058

Abstract :

This research presents an effective deep learning approach for facial emotion recognition using the Kaggle "FERAC" dataset. The proposed system uses convolutional neural networks (CNNs) to accurately classify facial expressions into different emotional categories such as happiness, sadness, anger, surprise, disgust, fear, and neutral. Preprocessing techniques including image normalization and data augmentation were applied to improve model robustness and generalization.

Through extensive training and evaluation on the benchmark dataset, the model demonstrated competitive accuracy and reliability despite challenges such as class imbalance and subtle variations in expressions. The study underscores the significance of deep neural architectures and advanced training strategies in enhancing facial emotion recognition performance, with potential applications in human-computer interaction, behavioral analysis, and affective computing.

Introduction :

The ability of detecting human emotions is essential for creating better interactions between humans and machines. Facial expressions are globally recognized as primary indicators of human emotions and have been extensively studied in psychology and computer vision.

With the use of deep learning techniques, especially convolutional neural networks, automated facial recognition has achieved significant advancements over traditional machine learning methods in accuracy and robustness. This paper focuses on leveraging deep learning to detect facial emotions, aiming to enhance the reliability and applicability of emotion detection systems in real-time scenarios. The research investigates model architectures, preprocessing methods and dataset utilization to develop a system capable of recognizing diverse emotions effectively.

Literature Survey :

Recent studies in facial emotion recognition (FER) systems have used a range of image processing and machine learning approaches to improve classification accuracy and adaptability across different datasets.

- In 2022, a research work proposed by Sowmiya R and team overcomes the drawback of earlier approaches in recognizing emotions from different angles by employing a DenseNet-169 as the underlying network and a deep Con-volutional Neural Network to determine face emotions. Even though the images were taken from various perspectives, their work include difficulty in distinguishing between genuine and fake emotions, and reduced accuracy when dealing with low-resolution or real-time images.

- In 2020, Smith and Varun proposed CNN model for emotion classification. Their method initially involved using time-frequency representation to convert EEG data into pictures. These images were then processed by CNN models built on architectures such as AlexNet, ResNet50 and VGG16. The lack of experimentally tuned parameters for signal processing and classification, however is a significant drawback of their work.

- In 2021, Tuncer, Dogan and Subasi proposed a novel method for fractal pattern extraction as features in an automated emotion identification system that utilizes EEG signals. By integrating Fractal Feature Patterns(FFP) with the Tunable Q-factor Wavelet Transform(TQWT), they developed a multi-level feature generation framework. During the feature selection stage, a novel iteration selector was employed to optimize the process. The model displayed exceptional performance when tested on 14-channel sensory EEG inputs using Linear Discriminant Analysis(LDA), K-Nearest Neighbour(KNN), and Support Vector Machine(SVM) classifiers. However, a limitation of this approach lies in the need for further enhancement of the fractal-based deep feature generator.

- In 2024, jixiang and Jianxin Peng research shown the complex nature of emotions, coupled with variations in how they are perceived and processed by different individuals, continues to pose challenges.

Research has shown that certain emotions such as Fear, may be processed differently compared to others, leading to unique patterns in their recognition and classification. This distinction highlights the need for robust models capable of capturing these subtle variations in emotion perception.

- As machine learning and deep learning progressed, diverse methods emerged for emotion detection in facial images or videos. CNN algorithms, for instance, demonstrated remarkable potential for real-time video emotion detection in human faces, achieving high accuracies reaching up to 97.23 % in the study of Monisha et al. (Monisha, Yogashree, Baghyalaksmi & Haritha, 2023). The accuracy correlated with specific facial expressions portrayed by participants during testing, showcasing the algorithm's sensitivity to emotional nuances.

- Durga and Rajesh (Durga & Rajesh, 2022) implemented transfer learning with a 2D-ResNet architecture to rectify inaccuracies in conventional facial expression detection, especially when true expressions were misclassified as others, particularly happy expressions being mistaken for disgust, exacerbated by facial coverings like masks. Their research exhibited superior performance metrics, achieving 99.3 % accuracy, 99.12 % recall, 0.98 % F1 score, and 99.16 % sensitivity from the confusion matrix.

- The utilization of the Bidirectional Long-Short Term Memory (BiLSTM) algorithm further refined emotion classification by employing different datasets for training and testing. Joshi et al. (Joshi, Ghongade, Joshi & Kulkarni, 2022) explored this approach with the SEED and DEAP datasets, showcasing performance enhancements of up to 8.2 % in recognition accuracy. This research emphasized the importance of optimizing deep neural network parameters for improved emotional recognition. Moreover, the integration of artificial intelligence and cognitive science showcased through hybrid algorithms like CNN-LSTM models, as proposed by Mishra et al. (Mishra et al., 2022), attained impressive training and validation accuracies of 95.93 % and 81.42 %, respectively. This hybrid approach outperformed other models, exhibiting superior generalizability in facial emotion identification.

- In artificial intelligence, especially within deep learning, alongside custom architectures, leveraging pre-existing ones through transfer learning has proven effective in detecting emotions through facial expressions. Akhand et al. (Akhand, Roy, Siddique, Kamal & Shimamura, 2021) utilized eight pre-trained Deep CNN (DCNN) models—VGG-16, VGG-19, ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, and DenseNet-161—achieving high accuracy across various emotional expressions using the Karolinska Directed Emotional Faces (KDEF) Dataset and Japanese Female Facial Expression (JAFFE). Emotions studied included afraid, angry, disgust, happy, neutral, sad, and surprised.

- Ullah et al. (Ullah, Qi, Hasan & Asim, 2022) introduced a CNN-RNN hybrid model for face emotion recognition using image datasets like the MMI Facial Expression Database and Japanese Female Facial Expression (JAFFE). This custom architecture not only enhanced overall accuracy compared to the proposed CNN model but also reduced false detections. When applied to a different image dataset, Posed and Non-posed Facial Expression (PNFE), the same hybrid model achieved a high accuracy of 88.54 % while identifying six basic expressions: anger, disgust, fear, happiness, sadness, and surprise.

- In the realm of computer vision, emotion detection through facial expressions has utilized models like the Improved Deep CNN and the hybrid RNN+Bi-GRU. Ullah et al. (Ullah et al., 2022), employing image datasets from the JAFFE database and the CK+ dataset, detected seven expressions: anger, disgust, fear, happiness, neutral, sadness, and surprise. This process involved three operational phases: super resolution, facial recognition, and classification, showcasing superiority over traditional FER methods.

Methodology :

The proposed system for emotion detection is built using deep learning algorithms. We used MobileNetV2, VGG19, ResNet50 and AlexNet model architectures to train our system and after getting results, we finalize the model architecture with better results.

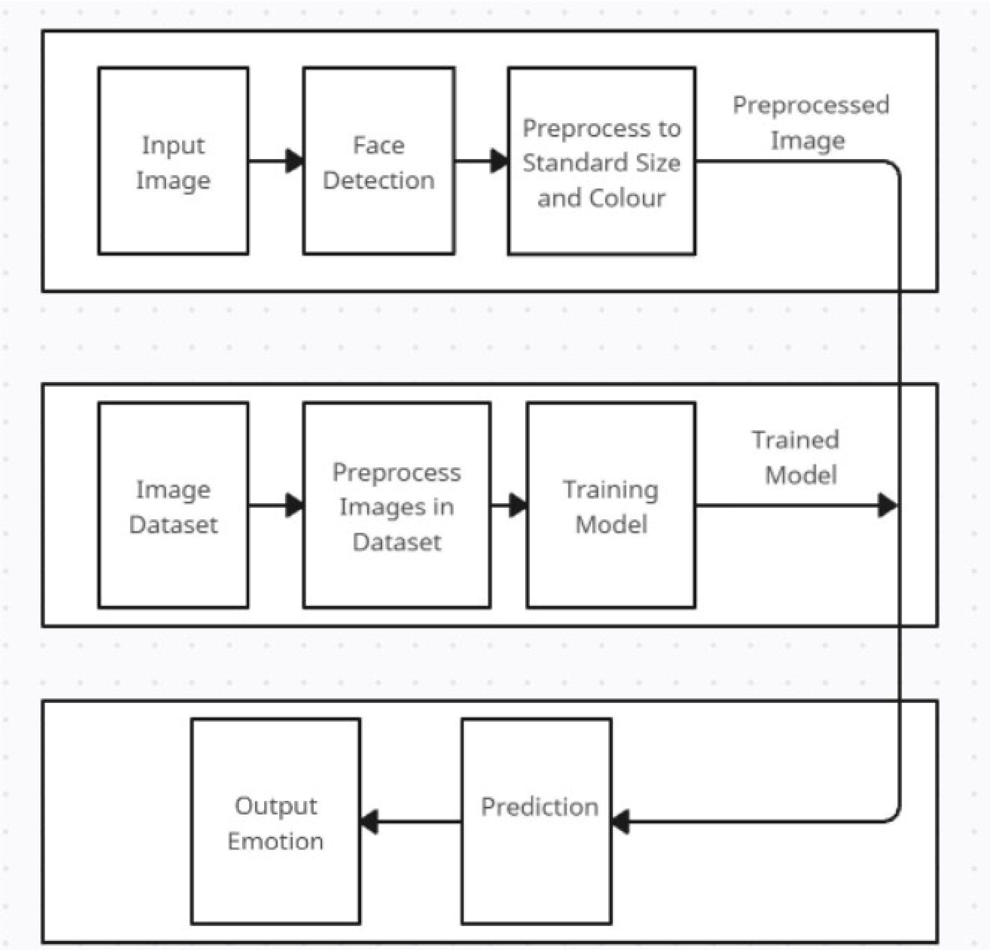

Figure-1

Fig-1 shows the algorithm of this system. It shows how our system works and how the emotion is classified from the input image.

1. Data Collection and preprocessing

We used the FERAC dataset taken from Kaggle. It consists of 770 images with different emotion categories.

a) Image resizing : All images in the dataset are resized into standard 224 x 224 pixels, which co-opreate with the requirements of the deep learning models.
b) Pixel Normalization : Pixel values are scaled between 0 and 1 to standardize input data, ensuring the stability of model training.
c) Data Augmentation : Different techniques like rotation, flipping horizontally, zooming, width and height shifting are used to improve variability and enabling the model for better identification of emotion.

2. Deep Learning Model Architecture :

The proposed approach uses four model architectures individually and then we finalize the model with better results for further predictions.

a) MobileNetV2 Architecture :

MobileNetV2 is a highly efficient convolutional neural network architecture designed for mobile and embedded vision applications. It is developed by researchers at Google, MobileNetV2 improves upon its previous version, MobileNet V1, by providing better accuracy and reduced computational complexity.

b) VGG19 Architecture :

VGG-19 is a deep convolutional neural network with 19 weight layers, comprising 16 convolutional layers and 3 fully connected layers. The architecture follows a straightforward and repetitive pattern, making it easier to understand and implement.

c) ResNet50 Architecture :

ResNet (Residual Network) is a deep convolutional neural network (CNN) architecture that uses "skip connections" to address the problems of vanishing gradients and performance degradation in very deep networks

d) AlexNet Architecture :

AlexNet was developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton and famously won the 2012 ImageNet Large Scale Visual Recognition Challenge(ILSVRC). AlexNet consists of 8 layers, 5 convolutional layers and 3 fully connected layers, ending with a 1000-way softmax output layer for computer vision.

3. Model Training :

All four convolutional neural network (CNN) architectures were trained separately on the same emotion detection dataset to ensure a fair and consistent evaluation of their performance. Each model was trained for 150 epochs using identical training configurations, including the same training–validation split, data preprocessing pipeline, and hyperparameter settings, except for architectural variations. The models were optimized using the categorical cross-entropy loss function and the Adam optimizer, which provided efficient convergence for the multi-class emotion classification task. To prevent overfitting and enhance generalization, data augmentation techniques such as random rotation, horizontal flipping, and zooming were applied during training. Additionally, early stopping and model checkpointing mechanisms were used to monitor validation accuracy and save the best-performing weights. Training was conducted on a high-performance GPU environment to speed up the process and handle the computational load of 150 epochs per model. The training curves of accuracy and loss were analyzed to confirm model convergence and stability across all architectures.

4. Formulae :

- Inverted Residual Block:
  Each block expands channels by factor t:
  a. Expansion :
  $D\_exp = t \times D\_in$
  Where t is the expansion factor (commonly 6), D_in is the input depth.
  b. Depthwise Convolution:
  $Parameters = K \times K \times D\_exp$
  Where K is the kernel size (typically 3).
  c. Projection (Pointwise Conv):
  $Parameters = D\_exp \times D\_out$
- Total Parameters per Inverted Residual Block :
  $Block\ Params = D\_in \times (t \times 1 \times 1) + (K \times K \times D\_exp) + (D\_exp \times D\_out)$
- Convolution Layer Output Size :
  $O = (I - K + 2P)/S + 1$
  Where O is the output size, I is input size, K is kernel size, P is padding, and S is stride.
- Number of parameters in a Conv Layer:
  $Parameters = (K \times K \times C\_in) \times C\_out + C\_out$

Results :

The system was trained on all four architectures seperately on FERAC dataset after preprocessing and data augmentation. The results of four architectures are below.

a) ResNet50 Model Architecture Results :

Fig-2 shows the graph comparision between Train accuracy and Validation accuracy of ResNet50 model. The training and validation accuracy of ResNet50 model are 61.34 and 60.00% respectively.

Fig-3 shows theTraining and validation loss for ResNet50 model.

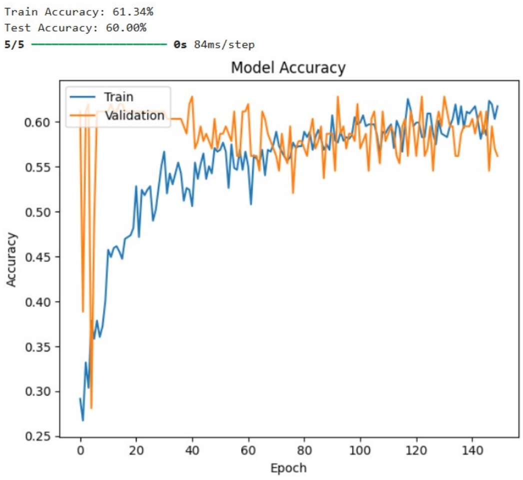

Figure - 2

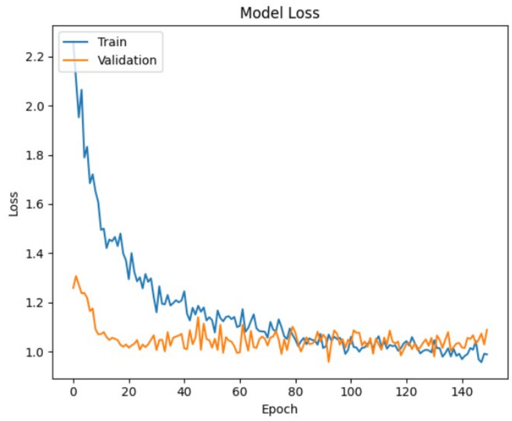

Figure - 3

b) MobileNetV2 Model Architecture Results :

Fig-4 shows the graph comparision between Train accuracy and Validation accuracy of MobileNetV2 model. The training and validation accuracy of MobileNetV2 model are 88.29% and 74.19% respectively.
Fig-5 shows theTraining and validation loss for MobileNetV2 model.

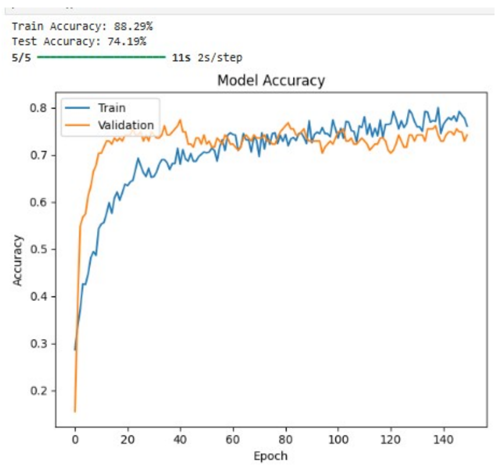

Figure - 4

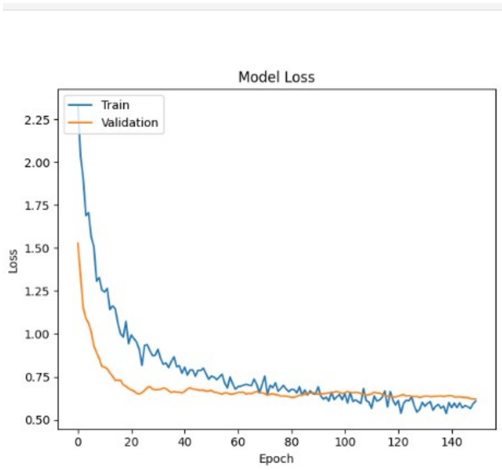

Figure - 5

c) VGG19 Model Architecture Results :

Fig-6 shows the graph comparision between Train accuracy and Validation accuracy of VGG19 model. The training and validation accuracy of VGG19 model are 82.28% and 67.74% respectively.
Fig-7 shows theTraining and validation loss for VGG19 model.

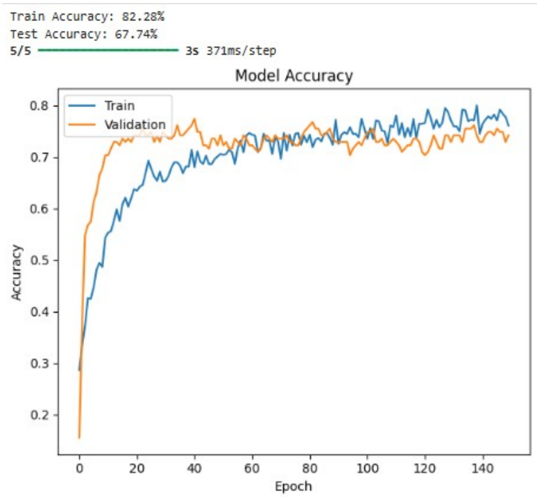

Figure - 6

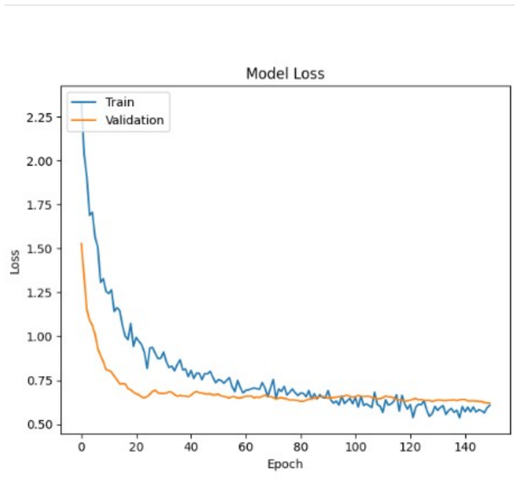

Figure - 7

d) AlexNet Model Architecture results :

Fig-8 shows the graph comparision between Train accuracy and Validation accuracy of AlexNet model. The training and validation accuracy of AlexNet model are 59.92% and 60.60% respectively.
The accuracy was never changed in this trained model, so we can conclude that it was never learning anything over iterations.
Fig-9 shows theTraining and validation loss for AlexNet model.

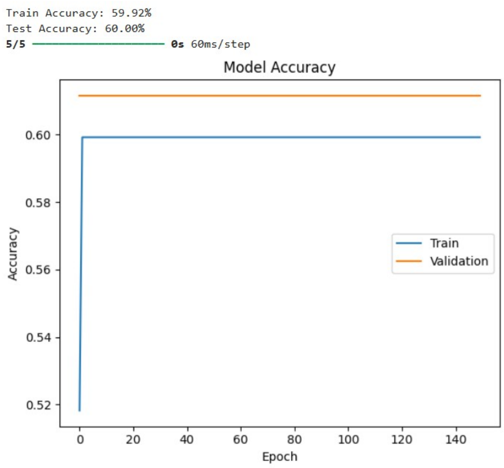

Figure - 8

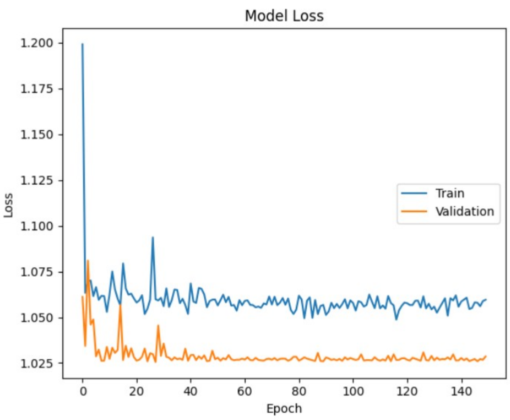

Figure - 9

e) **Comparision of all four models** :

Fig-10 shows the comparison between all four model architectures based on different metrics. We plotted a bar plot for this comparision and MobileNetV2 bar is coloured blue, VGG19 is coloured orange, AlexNet is coloured green and ResNet50 is coloured red.

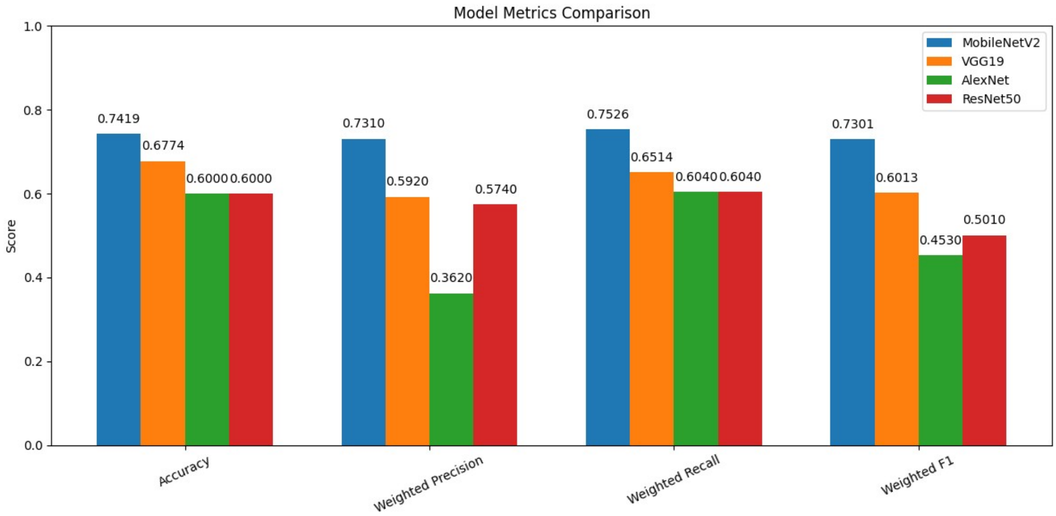

Figure - 10

- Accuracy defines the correctness of the model for predicting the output.
- Precision shows “Out of all instances predicted as positive, how many are actually positive?”.
- Recall shows “Out of all actual positive instances, how many did the model identified correctly?”.
- F1 Score is the harmonic mean of precision and recall, combining both metrics into a single value.

From the results obtained, MobileNetV2 model got better results than other three models based on all metrics calculated.

Conclusion and Future Work:

Our Emotion Detection System uses MobileNetV2 model that effectively classifies emotion of the person in Input image after preprocessing into standard size(224 x 224) pixels. We achieved 88.29% of Training accuracy and 74.19% of Validation accuracy for our MobileNetV2 model. These results prove the ability of the system to classify human emotion from an imput image. Our Future Work is to develop this model to real time detection by using web-cam of the device using OpenCV with pre-trained models like Haar Cascade Classifier.

References :

Kaggle for exploring datasets. https://www.kaggle.com/
https://ieeexplore.ieee.org/Xplore/home.jsp

https://www.sciencedirect.com/
Facial Emotion Recognition using Deep Learning published in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS). (DOC - 13-15 December 2022) https://ieeexplore.ieee.org/document/10029092
CNN based emotion Recognition published in IEEE Transactions on Neural Networks and Learning Systems. (DOP - 31 July 2020) https://ieeexplore.ieee.org/document/9153955
Fractal Pattern Feature generation function based emotion recognition method using EEG published in ScienceDirect in March 2021. https://www.sciencedirect.com/science/article/abs/pii/S0960077921000242?via%3Dihub
Multimodal Emotion Recognition Based on Facial Expressions published in IEEE Journal of Biomedical and health Informatics (DOP - 10 October 2024). https://ieeexplore.ieee.org/document/10602745

Enhanced automatic recognition of human emotions using machine learning techniques . https://www.sciencedirect.com/science/article/pii/S1877050923000200?via%3Dihub

A ResNet Deep Learning Based Facial Recognition Design For Future Multimedia Applications . https://www.sciencedirect.com/science/article/abs/pii/S0045790622006012?via%3Dihub

Deep BiLSTM neural network model for emotion detection using cross-dataset approach . https://www.sciencedirect.com/science/article/abs/pii/S1746809421010041?via%3Dihub

Facial emotion recognition using transfer learning in the deep CNN . https://www.mdpi.com/2079-9292/10/9/1036
Improved deep cnn-based two stream super resolution and hybrid deep model-based facial emotion recognition . https://www.sciencedirect.com/science/article/abs/pii/S0952197622004766?via%3Dihub

https://ieeexplore.ieee.org/document/10782924/
https://ieeexplore.ieee.org/document/10049334/
https://ieeexplore.ieee.org/document/10127369/
https://ieeexplore.ieee.org/document/11089747/
https://ieeexplore.ieee.org/document/9302866/
https://ieeexplore.ieee.org/document/11051988/
https://ieeexplore.ieee.org/document/9702451/
https://ieeexplore.ieee.org/document/10830488/
https://ieeexplore.ieee.org/document/1605389/
https://ieeexplore.ieee.org/document/6998925/
https://www.researchgate.net/publication/356066109_Human_Facial_Emotion_Detection_Using_Deep_Learning
https://ieeexplore.ieee.org/document/10298644/
https://ieeexplore.ieee.org/document/10953888/
https://www.deeplearningbook.org/contents/convnets.html
https://www.sciencedirect.com/science/article/pii/S2667305324000152#bib0041

"*Adaptive Computation and Machine Learning series*" book by Ian Goodfellow, Yoshua Bengio, Aaron Courville, published in 2015.