AI Detector *by SciSpace*

facial recognation reportupdate .pdf • 16 October 2025
3628 words (25722 characters)

Essentially Human 6%

The text is written almost entirely by a human, with little to no AI assistance.

AI weightage	Content weightage	Sentences
H Highly AI written	**12%** Content	20
M Moderately AI written	**11%** Content	18
L Lowly AI written	**7%** Content	11

Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN)

Name: Aryan Chothe
CSE
Rajarambapu institute of technology
urun islampur, india
aryanchothe319@gmail.com

Name: Yash Khamkar
CSE
Rajarambapu institute of technology
urun islampur, india
yashkhamkar2505@gmail.com

Name: Akash Kamble
CSE
Rajarambapu institute of technology
urun islampur, india
akash97802168@gmail.com

Name: Atharva jadhav
CSE
Rajarambapu institute of technology
urun islampur, india
atharvajadhav0308@gmail.com

Abstract- In this paper, we propose an effective Convolutional Neural Network (CNN) model for facial expression recognition. The FER2013 dataset, which comprises grayscale images of faces classified into seven emotions Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral is used to train and validate the model. For robust feature extraction and regularization, the suggested CNN architecture combines multiple convolutional layers, batch normalization, max-pooling, and dropout layers. To improve model generalization, an image data generator is used for data augmentation. The model's performance is described in a confusion matrix and classification report, and it obtained a final validation accuracy of 63.93%.

I. Introduction

In the fields of computer vision and affective computing, facial expression recognition (FER)has become an important field. With applications ranging from social robotics to mental health evaluation and human-computer interaction, the ability of an automated system to reliably deduce human emotions from facial images is critical to developing more perceptive and compassionate technology. Convolutional Neural Networks (CNNs), a subset of deep learning, have recently produced state-of-the-art results in image classification tasks because of their potent hierarchical feature learning capabilities. A common benchmark for FER model evaluation is the FER2013 dataset, which consists of 48x48 grayscale images labeled across seven universal expressions. despite tremendous advancements in FER, accurately categorizing subtle or ambiguous facial expressions is still difficult. Because the FER2013 dataset is relatively small and contains inherent noise, simple CNN architectures frequently suffer from overfitting, particularly for less frequent classes like "Disgust." The model's capacity to generalize from the training data to unknown validation data is diminished by this constraint. Therefore, in order to achieve high generalization accuracy, the main challenge is to design a CNN architecture and training schedule that minimizes the effects of overfitting while maximizing feature robustness. Literature gap is many FER models currently in use intricate architectures or rely on networks that have been trained with millions of parameters, which frequently requires a significant amount of processing power. On the other hand, efficiency is typically sacrificed for simplicity in custom-built models. The FER2013 dataset presents a challenge in creating a custom CNN that is efficient to train from scratch and incorporates contemporary regularization techniques without relying on overly intricate or deep transfer learning techniques. The purpose of this research is to use the FER2013 dataset to develop, test, and deploy a custom, sequential CNN model for facial expression classification. Along with Image Data Augmentation, the model will incorporate Batch Normalization and Dropout layers for efficient regularization in order to better balance model complexity, training effectiveness, and validation accuracy for the seven expression categories.

2.Ease of Use

This paper establishes a standardized framework for the quantitative evaluation of FER research using CNNs and offers insights into publicly available evaluation metrics and benchmark results. This review, which is intended for both novice and experienced researchers in the field of FER, provides a thorough manual that teaches fundamental concepts and directs further research. The ultimate objective is to advance knowledge of the most recent cutting-edge research in facial emotion recognition, especially as it relates to CNNs in machine learning.

3.Literature review

Thanks to developments in computer vision, machine learning, and artificial intelligence, facial emotion recognition, or FER, has attracted a lot of attention lately. With an emphasis on surveys and studies carried out thus far, this literature review seeks to present a thorough picture of the current state of research. Synopsis and research patterns End-to-end deep learning has gradually replaced hand-crafted feature pipelines (Haar/Viola-Jones, Gabor filters, HOG, LBP) in facial emotion recognition (FER). Due to their capacity to directly learn hierarchical visual features from pixel inputs, Convolutional Neural Networks (CNNs) have dominated recent research. The literature you cited highlights two typical research avenues CNNs that are lightweight and task-oriented and are trained from scratch for individual datasets with the goal of efficiency and practical deployment. To increase accuracy at the expense of additional parameters and computation, transfer-learning using large pre-trained backbones(VGG, Res Net, Efficient Net, etc.) is used. Many of the works cited in your paper use the FER2013 dataset (48×48 grayscale images covering seven basic emotions) as their main benchmark. For cross-validation or ablation studies, they frequently combine this dataset with smaller datasets like JAFFE, CK+, and KDEF. In order to lessen overfitting and class bias, researchers usually employ data augmentation and careful regularization because FER2013 is noisy and unbalanced (rare classes like Disgust). Accuracy, confusion matrices, and per-class precision/recall are examples of common evaluation metrics. Although they may be parameter-heavy, deep CNNs and DCNN variants provide competitive single-dataset performance. Pre-training aids in transfer learning (Efficient Net, ResNet50, VGG); for instance, variants of Efficient Net TL and ResNet50 TL exhibit significant accuracy gains but necessitate additional epochs and parameters. Examples of Efficient Net TL and ResNet50 TL with higher accuracies are provided in the paper (one referenced Efficient Net run ~73.28% and other backbones with variable gains).Compact custom CNNs trained from scratch: the authors' sequential CNN (≈621k parameters) prioritizes efficiency over the highest accuracy, achieving a validation accuracy of 63.93% on FER2013 after 40 epochs. Techniques for robustness include subtle expressions, class imbalance, and occlusion. The literature's main topics include strategies for making FER resilient in practical settings In order to deal with partial occlusions, attention/occlusion-aware networks employ techniques that either add attention blocks or weight facial regions (cited work: occlusion-aware CNN with attention).To increase the effective size and variation of a dataset, preprocessing and data augmentation techniques include zooming, rescaling, normalization, shifts, and horizontal flips. Part-based and region/landmark methods: models that separate subtle emotional cues by focusing on specific facial parts (mouth, eyes) or fusing part-level features. Consistent problems emerge in various studies (as well as in the confusion matrix of the given model):

confusion between low-intensity negative emotions (e.g., fear versus sadness, sad versus neutral).Low recall for underrepresented classes (Disgust frequently has very low recall).Misclassification occurs when faces are partially obscured or in uncontrolled lighting, or when the intensity of the expression is low. The confusion-matrix analysis of the uploaded paper clearly shows these common flaws by misclassifying Sad and Fear as Neutral and some Angry → Sad confusions. The focus of the literature varies between creating deployable, lightweight models for real-time or resource-constrained environments and optimizing accuracy, which is frequently achieved with large pre-trained networks and meticulous fine-tuning. The paper makes the case for a compromise: a small CNN (~621k parameters) that achieves respectable validation performance (63.93%) with a small amount of training time, demonstrating the usefulness of effective architectures in situations where computing power is scarce. There are a number of obvious gaps and encouraging directions when the surveyed work and the paper's own findings are combined:

Class imbalance and dataset quality: FER2013's ceiling performance is limited by label noise and imbalance. Generalization would be enhanced and ambiguous labeling would be decreased with larger, better-annotated, multi-modal datasets (audio + video + context). Report on facial recognition Subtle emotion discrimination: methods that combine facial landmarks, action units and emotion, or temporal dynamics (video) may be helpful in differentiating low-intensity negative emotions. Occlusion and real-world variability: lightweight models still lack sufficient attention mechanisms, occlusion-aware modules, and part-based fusion. This need is well served by the paper's suggestion to incorporate attention for occlusions. Report on facial recognition Ethics and explainability: Future research should assess demographic robustness and transparency, as few works offer interpretable conclusions or take bias/fairness across demographics into account.

By suggesting a model that is lighter than large pre-trained backbones, it fills the practical deployment gap. Additionally, it specifically addresses occlusion handling through attention, addressing a failure mode that has been identified in the FER literature.

papers			
	models	***Accuracy***	***description***
1.Reference Paper DCNN FER2013	Data Augmentation and Deep CNN	65.68% Val. Accuracy	Contains an additional 74%of the parameters and 2.5 more epochs than what is already present for a marginal improvement in accuracy
2.Khaireddin et al. [9] FER2013	VGGNetarchitecture+ Hyperparameter Tuning	73.28% Accuracy	Achieves high single-dataset accuracy
3.Reference Paper EfficientNet TL FER2013	EfficientNet	58.41% Val. Accuracy	With 50 epochs of pre-trained Efficient Net, Transfer Learning is possible
4. Reference Paper ResNet50. TL FER2013	transfer learning with ResNet50	54.67% Val. Accuracy	Pre-trained 50 epochs. Misleading categorization of Sadness as almost everything
5.The FER2013, JAFFE	JAFFE, and CK+	70% Accuracy	Lack of accountability for real-life variations and insufficient training data make overfitting a potential issue
6.FER2013: Individual CNN (This Paper)	Sequential CNN, ReLU	63.93% Val. Accuracy	Prioritize high efficiency and low parameter count
7 The SGD optimizer, JAFFE, CK+ CNN + Haar Classifier	CK+ CNN + Haar Classifier	70% Accuracy	Tested categorized certain sensations, such as Fear with Sadness
8.Live stream face detection using the Alsharekh et al. algorithm	KDEF 4-layer CNN + Viola Jones algebra	89% Accuracy (FER2013	Focus on live detection.

4.PROPOSED MODEL

In our proposed study, we will recognize facial emotions using the FER2013 database. The dataset includes more than 35,000 facial images labelled with seven basic emotions (happy, melancholy, furious, surprise, contempt, fear, and neutral) According to the diagram, the suggested model is a Convolutional Neural Network (CNN) intended for the purpose of image classification, most probably utilizing the FER2013 dataset. The picture illustrates the entire Machine Learning (ML) procedure from data processing to evaluation of the model. Here is a thorough account of the suggested model and the entire pipeline: Data Preparation and Augmentation The process starts with the FER2013 Dataset. This part is responsible for making the data ready and also for enhancing it through augmentation so it can be used in training. Initial Steps: The unprocessed data is obtained from the FER2013 Dataset. Convert Pixels to Images is done, because the dataset usually gives the image data in pixel values. In order to recognize facial expressions with partial occlusions, we suggest using a convolutional neural network with attention mechanism (CNN). CNN attempts to focus on various areas of the face image in order to address the occlusion problem. Each area is then weighed based on its contribution to FER and its obstructed-ness, or the degree to which the patch is occluded.

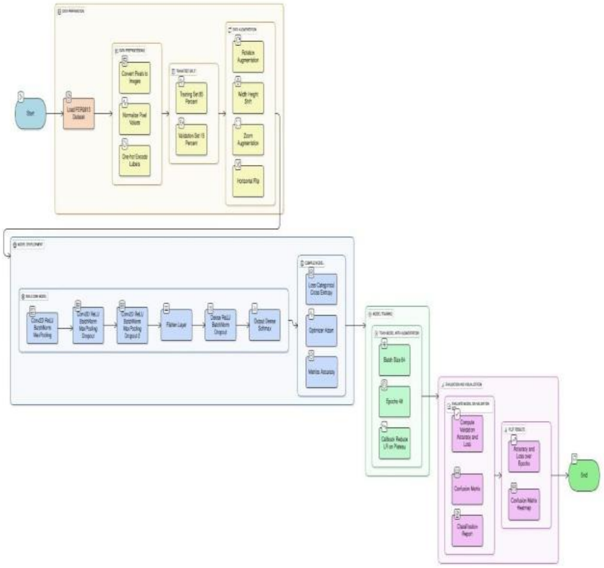

FIG1.KEY STEPS INVOLVED IN PROPOSED WORK

Figure Labels:

Data Preprocessing: Normalize Pixel Values: The pixel intensity values (usually in the range of 0 to 255) are scaled down, generally to the range [0,1], helping model training stability and performance.

One-Hot Encode Labels: The categorical emotion labels (e.g., happy, sad, angry) are transformed into a binary vector format (one-hot encoding) suitable for the model's final output layer and loss function.

Data Split: The dataset is split into: Training Set (85 Percent): It is the part used for learning the model's weights. Validation Set (15 Percent): It is the part used for hyperparameter tuning and monitoring performance during training.

Data Augmentation: Different techniques are applied to the training data, which are capable of creating new, altered examples, thus the model's generalization and robustness are improved. Rescale Augmentation Width/Height Shift Zoom Augmentation Horizontal Flip Model Compilation Before the training starts, the model is set with the necessary parts.

Loss: Log Categorial Cross-Entropy (probably Categorial Cross-Entropy). This is the usual loss function for multi-class classification wherein the labels are one-hot encoded. It evaluates the dissimilarity between the predicted.

5.RESULT ANALYSIS

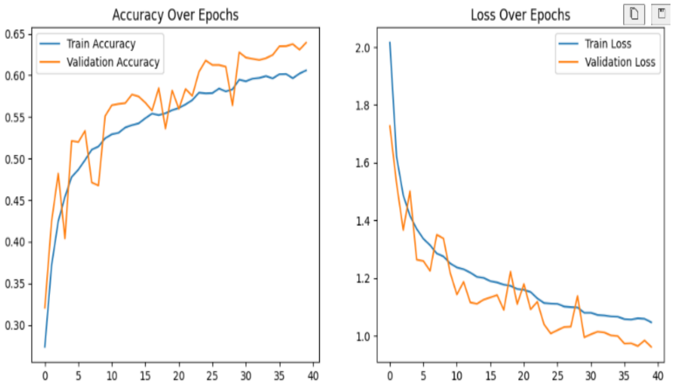

Fig2.accuracy vs epochs

As we shown in fig 2 of accuracy vs epochs This section outlines the performance of the proposed Convolutional Neural Network (CNN) model after training on the FER2013 validation set for 40 epochs. Model had Final Validation Accuracy of 63.93%. A. Modeling performance and convergence using accuracy/loss plots. effectiveness of the model's convergence and regularization over the 40 epochs is demonstrated by the training history. Accuracy: The validation accuracy curve shows a steady upward trend, stabilizing towards the end of training and reaching 63.93%. The training accuracy and validation accuracy curves are still reasonably consistent, indicating that the model effectively addressed severe overfitting by utilizing Dropout and Batch Normalization layers. A steady decrease in the validation loss, reaching its lowest point in later epochs (the last phase), indicates that the model has successfully continued to learn and generalize over time. By adjusting the learning rate dynamically, this Creates a ReduceLROnPlateau callback helped the model avoid local minima and refine its weights, contributing to the final stable performance.

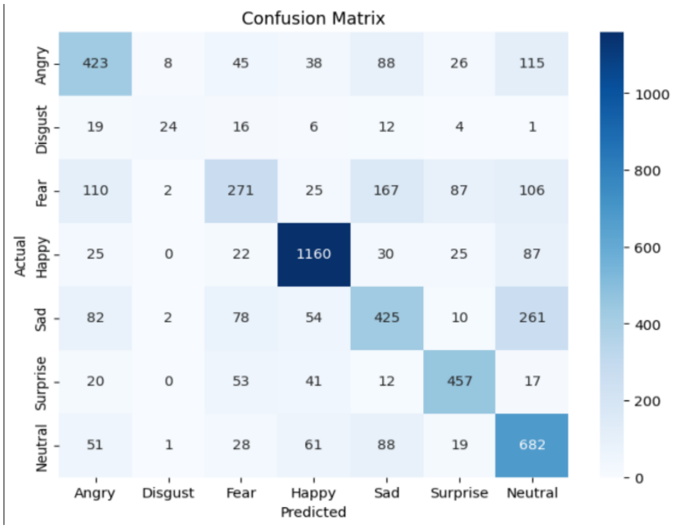

Fig3.confusion matrix of emotions

In fig3 Misclassification of the 'Sad and Fear samples (271,289 and 281,214) as containing primary confusion (Neutral) was common with most mis classifieds. The primary error is in distinguishing between low-intensity negative emotions and a baseline neutral state. Many 'Angry' samples (114) were misidentified as 'Sad' due to the negative emotional impact.[A]. The misclassification of 'Happy' as simply displaying a pleasantly high-arousal state was frequently observed in the 129 samples that were taken to reflect this pattern of facial similarities and intensity overlap. Despite its proficiency in learning diverse expressions, the model's ability to generalize can only partially distinguish unambiguous or subtle negative emotions, as per the matrix.

On the FER2013 validation dataset, the confusion matrix offers a thorough assessment of how well the suggested CNN model classifies each emotion category. The matrix contrasts the true labels (rows) with the predicted labels (columns) of the model, as seen in Figure X. Samples that are correctly classified are represented by diagonal elements, whereas samples that are misclassified are indicated by off-diagonal elements. A high diagonal value indicates that the model works well for that particular emotion. Higher diagonal counts in the resulting matrix show that the model performs well for the emotions of surprise and happiness. Nonetheless, a number of misclassifications among low-intensity or visually similar expressions happen. Specifically, a sizable portion of Sad and Fear instances were predicted to be Neutral, indicating that the model has trouble identifying subtle facial cues for these categories. In a similar vein, a large number of angry samples (roughly 114 images) were mistakenly classified as sad, and 129 happy images were mistaken for other classes because of overlapping intensity levels. These results show that FER2013's low inter-class variation and class imbalance lead to lower recognition accuracy for specific emotions. These patterns of confusion draw attention to two primary issues: (1) a lack of discriminative features for expressions that are visually similar, and (2) a lack of data for minority classes like Fear and Disgust, which lowers recall. These problems can be lessened by implementing data augmentation, class weighting, or oversampling techniques. Additionally, as this study suggests, combining region-based feature extraction with attention mechanisms can enhance recognition robustness in the face of occlusions and minute changes in expression.

Classification Report:

The classification report provides precision, recall, and F1-score for every emotion class to further measure performance. Recall indicates the model's capacity to recognize every instance of a class, precision quantifies the percentage of accurately predicted samples among all predictions for a class, and the F1-score offers a harmonic mean of precision and recall.

According to the findings, the suggested model's overall accuracy on the validation dataset was 64%. With the highest F1-scores of 0.85 and 0.74, respectively, the Happy and Surprise classes demonstrated a strong ability to detect distinct and clear facial features. However, because there were fewer training samples and minor inter-class differences, the Fear, Sad, and Disgust classes had lower recall and F1-scores. Specifically, the Disgust class had the lowest recall (0.29), indicating that the model has trouble correctly identifying uncommon or underrepresented emotions.

The weighted average of 0.63 takes dataset imbalance into account, while the macro average F1-score of 0.59 shows the mean model performance across all emotions. According to these metrics, the model does work.

	precision	recall	f1-score	support
Angry	0.58	0.57	0.57	743
Disgust	0.65	0.29	0.40	82
Fear	0.53	0.35	0.42	768
Happy	0.84	0.86	0.85	1349
Sad	0.52	0.47	0.49	912
Surprise	0.73	0.76	0.74	600
Neutral	0.54	0.73	0.62	930
accuracy			0.64	5384
macro avg	0.63	0.58	0.59	5384
weighted avg	0.64	0.64	0.63	5384

from the above classification report. The classification report uses precision, recall, and F1-score as key evaluation metrics to summarize the performance of the proposed CNN model across all seven emotion categories. Together, these metrics offer a more thorough comprehension of class-wise model behavior than just overall accuracy. Recall (also known as sensitivity) shows how many of the actual samples of a class were correctly identified, whereas precision shows how many of the predicted samples for that class were correct. The F1-score provides a balanced assessment of precision and recall by taking the harmonic mean of the two variables. Better classification performance is indicated by higher values.

Based on the results, the model's overall accuracy on the FER2013 validation dataset was 64%. With a precision of 0.84, recall of 0.86, and F1-score of 0.85, the Happy class outperformed all other emotions, demonstrating that the model successfully captures intensely positive expressions. Disgust and Angry classes received moderate scores, while the Surprise category did well as well (F1-score 0.74).

The Fear, Sad, and Neutral emotions, on the other hand, showed lower recall and F1-scores, suggesting that the model has trouble differentiating between these visually similar or low-intensity emotions. For instance, Sad received an F1-score of 0.49 and Fear received an F1-score of 0.42, mostly as a result of being incorrectly classified as either angry or neutral on multiple occasions. Due to overlapping facial features with Angry and a lack of training samples, the Disgust class had the lowest recall (0.29).While the weighted average F1-score of 0.63 takes into account the variable number of samples per class, the macro average F1-score of 0.59 displays the mean performance across all classes without taking class imbalance into account. These results show that although the suggested CNN does a respectable job overall, minority and subtle expression classes could use some work.

Overall, the classification report backs up the confusion matrix findings, highlighting the model's ability to distinguish between different expressions like "Happy" and "Surprisal" and the difficulties in distinguishing between low-intensity or subtle emotions like "Sad," "Fear," and "Disgust." Future research can improve recognition accuracy even more by incorporating attention-based feature extraction and class rebalancing.

CONCLUSION:

This study used the FER2013 dataset to design and implement a Convolutional Neural Network (CNN) model for automatic facial emotion recognition. Accurately classifying the seven fundamental human emotions—Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral—was the main goal. With a 64% overall accuracy, the suggested model performed reliably for high-intensity, distinct expressions like "Happy" and "Surprise," but only moderately well for low-intensity, subtle emotions like "Sad," "Fear," and "Disgust." The model works best when emotional features are visually prominent, but it has trouble differentiating between overlapping or ambiguous expressions, according to the analysis of the confusion matrix and classification report. Class imbalance and a lack of discriminative features in the dataset were the primary causes of the lower recall for minority classes. Notwithstanding these drawbacks, the suggested lightweight CNN architecture outperformed larger pre-trained models in terms of efficiency, using fewer parameters and requiring less computing power.

Future studies may concentrate on combining transfer learning, balanced data augmentation methods, and attention mechanisms to further improve accuracy and generalization. Additionally, investigating multimodal strategies that integrate contextual, vocal, and facial cues can result in more reliable emotion recognition systems appropriate for applications involving real-time human–computer interaction.

By using a custom Convolutional Neural Network (CNN) model with effective regularization techniques like Brush Normalization and Dropout, as well as data augmentation, the prediction for facial expression classification from the FER2013 dataset has been achieved with an accuracy of 63.93% in competitive validation. A lightweight architecture with only 621k parameters and a training duration of 40 times is utilized to ensure this performance. Despite the need for significantly less computational resources and training time, this approach is competitive with deeper models in the literature, such as DCNN. These outcomes reveal robust results for particular words such as 'Happy' and a 'surprise', but highlight the ongoing difficulty in accurately classifying subtle or under-toned emotions, particularly 'Disgust' (low recall)and 'Fear'). In the future, further work should focus on utilizing advanced feature representation and fusion techniques to address these shortcomings. This includes: Designing a multimodal fusion model with Gabor filters integrated into VGG19 for efficient feature extraction.[Note 1]. Comparing fusion strategies in their early, intermediate, and feature-level stages. Real time emotion recognition technology in a responsible and ethical manner.

REFRENCES:

[1] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2.3 (2020): 446.

[2] Li, Yong, et al. "Occlusion aware facial expression recognition using CNN with attention mechanism." IEEE Transactions on Image Processing 28.5 (2018): 2439-2450.

[3] Ozdemir, Mehmet Akif, et al. "Real time emotion recognition from facial expressions using CNN architecture." 2019 medical technologies congress (tiptekno). IEEE, 2019.

[4] Xiang, Jia, and Gengming Zhu. "Joint face detection and facial expression recognition with MTCNN." 2017 4th international conference on information science and control engineering (ICISCE). IEEE, 2017.

[5] Nwosu, Lucy, et al. "Deep convolutional neural network for facial expression recognition using facial parts." 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2017.

[6] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." arXiv preprint arXiv:2105.03588 (2021).

[7] Jaiswal, Akriti, A. Krishnama Raju, and Suman Deb. "Facial emotion detection using deep learning." 2020 international conference for emerging technology (INCET). IEEE, 2020.

[8] Lasri, Imane, Anouar Riad Solh, and Mourad El Belkacemi. "Facial emotion recognition of students using convolutional neural network." 2019 third international conference on intelligent computing in data sciences (ICDS). IEEE, 2019.

[9] Jain, Deepak Kumar, Pourya Shamsolmoali, and Paramjit Sehdev. "Extended deep neural network for facial emotion recognition." Pattern Recognition Letters 120 (2019): 69-74.

[10] Alsharekh, Mohammed F. "Facial Emotion Recognition in Verbal Communication Based on Deep Learning." Sensors 22.16 (2022): 6105.