Journal of Information Systems Engineering and Management
2025, 10(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improving Facial Expression Recognition with Deep Learning for Faces of Different Ages

Pranali Dahiwal[1], Anagha Kulkarni[2]
[1]Research Scholar, Vishwakarma Institute of Information Technology, S.P.P.U. Pune, India
pranali.dahiwal@viit.ac.in
[2]Cummins College of Engineering for Women, Pune, India
anagha.kulkarni@cumminscollege.in

ARTICLE INFO

Received: 01 Oct 2024

Revised: 01 Dec 2024

Accepted: 18 Dec 2024

ABSTRACT

Many times automatic human emotional detection plays a vital role in health care and law enforcement agencies. To achieve this, it is necessary to have a robust hardware resource that can capture images under harsh conditions; numerous components already exist in the industry for this purpose. However, the greatest challenge lies in developing an advanced algorithm to process these images and extract the desired emotions. Many image processing techniques, such as PCA and transform techniques, are available. But these technologies aren't providing satisfactory results in harsh conditions. So, the evolution of deep learning models plays a crucial role in the identification of faces and, thereby, emotions. Keeping this fact in mind, designed model employ a deep learning model to enhance the image channels, which can be then use to efficiently train a large dataset. Therefore, the designed model utilizes a channel boost convolution neural network to achieve optimal accuracy in identifying facial emotions across a broad age range. The results show that the designed model outperforms other emotion recognition models, achieving a root mean square of 0.2948 and an accuracy of 98.34% on a variety of age facial images.

Keywords: Deep learning, Channel boost Convolution Neural network, Facial emotions recognition, Data segmentation.

INTRODUCTION

The way a person looks at themselves says a lot about their emotional condition. A person's face is the first thing that the human eye detects. More than forty different expressions are possible using the facial muscles. Happiness, sadness, fear, wrath, disgust, surprise, mystery, and hatred are the few basic manifestations that physiologists say emotions evoke. People often use these emotions as a standard because they are prevalent and people express them readily. Nonverbal cues like these expressions on the face can supplement or even substitute spoken words when necessary. A person's emotional state is influenced by a myriad of biochemical and environmental factors, which interact to cause intricate psycho-physiological alterations. Every aspect of a person's life is impacted by their emotions, which are the main variables in developing a good self-concept. Words, ideas, phrases, imitation, posture, and most importantly facial expressions convey emotions that are intimately related to one's psychology.

Emotion recognition takes advantage of the traits and signs exhibited by collected emotional reactions to assess human emotional states, which vary from one person to the next.

There are typically two ways to classify emotional representations: the dimensional model and the discrete model. The valence, arousal, and dominance dimensions are used to map the six main emotional states in the discrete model: anger, disgust, fear, happiness, sorrow, and surprise. These states are associated with diverse facial expressions in the dimensional model.

One of the most promising new areas in computer science is machine learning. Over the next three years, that is anticipated to have a 95% impact. A kind of machine learning known as "deep learning" uses algorithms modeled after the human brain called artificial neural networks. One subset of deep neural networks that employs the mathematical technique convolution is the Convolutional Neural Network (CNN). The system uses a 2D convolutional neural network (CNN) for image recognition since the dataset is comprised of images. The new deep convolutional neural network that has been suggested can accurately identify seven different human facial expressions. The dataset is hand-collected using a cell phone camera, and the model is trained using this dataset.

In order to discover answers, a neural network uses a set of algorithms designed to function similarly to the human brain. To determine the difference between two sets of data, a specific sort of Neural Network called a Convolutional Neural Network uses a mathematical procedure called a convolution. Convolutional Neural Networks have shown to be very effective in solving complicated problems with high accuracy, such as pattern recognition, picture and video categorization, and so on, in contrast to standard neural networks.

Improve the channel The four layers that make up a convolutional neural network work together to extract features from the input images. Every convolutional filter can serve as a representation of the features extracted from the input images by the algorithm. The overall image is changed by the filter values in convolutional layers, which are composed of small patches. In order to decrease its size without sacrificing information, the pooling layer receives the output from the convolutional layer and transfers it to it. Before feeding them into the CB-CNN for classification, the 2-D arrays are flattened by the flatten layer. This CB_CNNCNN uses the backpropagation approach to train on the input dataset for object recognition, which involves adjusting the weights based on the errors and reducing the loss function. The ability to employ Channel Boost CNN to identify facial expressions is a direct result of this.

Part 2 of this research paper concentrates on analyzing the earlier works on facial emotion recognition. Part 3 delves into the intricacies of the developed model. Section 4 evaluates the obtained results, whereas section 5 concentrates on concluding this article along with the possibilities of future enhancement.

LITERATURE SURVEY

Mohammad Reza Falahzadeh et al. [1] introduce a 3D CNN for speech emotion identification. The 3D tensors from audio signal phase space recovery are used directly by the proposed 3D CNN. EMO-DB and eNTERFACE05 results show that the 3D CNN can correctly categorize related emotions using the specified 3D tensors, which include key speaker emotional information. Diego Peña et al. [2] proposed a framework for comparing fusion methods. It compares fusion processes based on accuracy, F1-score, average evaluation time, and training behavior. Yeo chan yoon et al. [3] propose a cross-modal translator to improve emotion identification. The suggested technique to train a multimodal model from heterogeneous data does not need matching visual, audio, and text modalities. The model's self-learning with multimodal datasets is flawed.

[4] Speech emotion identification involves feature extraction and classifier design, according to Liu Yunxiang et al. However, noise and gender differences reduce recognition efficiency. This paper addressed this issue using two adversarial multi-task learning (ASPMTL) models.

Bagus tris atmaja et al. [5] tested 19 self-supervised speech representations and 1 classical acoustic feature across 5 datasets for spoken language emotion categorization using a single classifier. The author calculates effect size over twenty speech representations to show relative performance improvements from best to worst. Top three are WavLM Large, UniSpeech-SAT Large, and HuBERT Large with negligible effect sizes.

Chih-lyang hwang et al. [6] proposed the sequential recurrent convolution network (SRCN) using a CNN and LSTM models. Future research will focus on a distributive UWB network for wireless navigation with many ODSRs and individuals.

Felicia Andayani et al. [7] described a hybrid Long Short-Term Memory (LSTM) Network and Transformer Encoder to learn voice signal relationships and discern emotions. Performance experiments were run on the proposed LSTM-Transformer model. Pre-processing needs improvement, especially for language-independent datasets. Haiyan Wang et al. study three-dimensional emotion identification using free-form speech [8]. Every dimension represents a fundamental emotion trait.

Sudarshan Pant et al. suggested a TV scene transcript emotion dataset [9]. The dataset was labeled euphoria, dysphoria, and neutral to reflect the TV show's characters' emotions. Fair agreement was reached among 64 dataset annotators. Author used deep learning to classify emotion labels using context. A multidimensional data model and mining approaches can find inconsistency factors in multichannel emotion recognition, according to Agnieszka Landowska et al. [10].

Study results are inconsistent.

Optimal Graph coupled Semi-Supervised Learning (OGSSL) is recommended by Yong Peng et al. [11] for EEG-based emotion recognition. The OGSSL framework combined unlabeled EEG sample emotional label estimation and optimum graph learning.

[12] Suci Dwijayanti et al. recommend merging face and emotional detection for real-time humanoid robots. Convolutional neural network architectures are utilized to construct face and emotion recognition systems simultaneously. The model is compared to AlexNet and VGG16 to determine the best architecture for humanoid robots. The model uses the limited dataset to find results.

Subramanian et al. detail Due to ML and DL advances, emotion recognition (ER) in healthcare is gaining attention.
Li-min zhang et al. [15] developed an F-Emotion feature selection approach to assess the F-Emotion value of speech emotion characteristics retrieved for each emotion category. The author used the F-Emotion value to assess the significance of each speech emotion characteristic or combination of features for each emotion classification. Minchao Wu et al. [16] developed a new experimental paradigm that allows smells to actively participate in many video-evoked emotion stages to test the efficacy of olfactory-enhanced films. [17] Kh liang ong et al.'s "Mel-MViTv2" speech emotion recognition method uses Mel-STFT spectrograms and the MViTv2 classifier. Mel-STFT uses Short-Time F author's ier Transform and frequency-to-Mel scale conversion to visualize audio signals. Deng Pan et al. say GCNs capture brain connections, making them good at EEG recognition [18]. However, previous study disregarded the key EEG component. Author propose MSFR-GCN, a multi-scale feature reconstruction GCN for emotion and cognitive task recognition. Besides the MSFR, MSFR-GCN has a feature-pool with two submodules, MSSR and multi-scale Squeeze-and Excitation.

PROPOSED MODEL

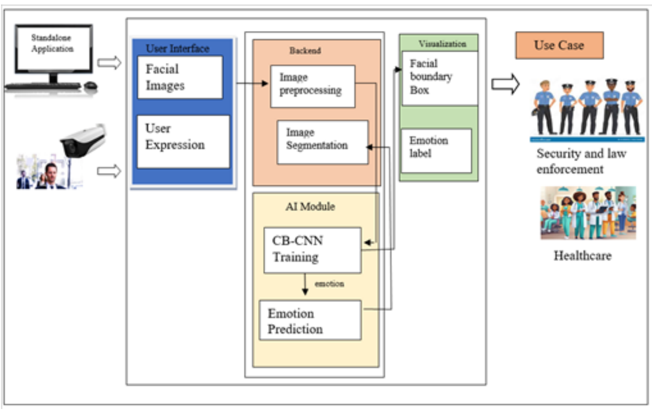

Figure 1: Overview of the proposed model for Facial expression detection

The proposed system creates a model that uses channel boost Convolution neural network to identify various facial emotions across different age groups is depicted in the figure 1 shown above.

The phrases below narrate the steps involved in the development of the proposed system.

Phase 1: Dataset Collection and Preprocessing – The proposed model of facial emotion detection in all age expressions is deployed by collecting the image dataset from the URL[21]. The obtained dataset includes classes for various emotions such as anger, disgust, fear, happiness, neutrality, sadness, and surprise. The dataset contains a total of 3995 angry emotion images for training, 436 disgusted emotion images for training, 4097 fearful emotion images for training, 7215 happy emotion images for training, 4965 neutral emotion images for training, 4830 sad emotion images for training, and 3171 surprise emotion images for training. A total of 28709 images for training purposes used by the model.

On the other hand, the dataset contains a total of 958 angry emotion images for testing, 111 disgusted emotion images for testing, 1024 fearful emotion images for testing, 1774 happy emotion images for testing, 1233 neutral emotion images for testing, 1247 sad emotion images for testing, and 831 surprised emotion images for testing. A total of 7178 images for testing purposes used by the model. Both the training and testing images are properly kept in the required hierarchy to train the model of channel boost convolution neural network by feeding the path of the dataset.

Phase 2: Data segmentation and Channel variation –

The Python programming language deploys the proposed model of facial emotion detection based on variant age. The built model employs the Keras and Tensorflow main libraries, as well as Python openCV. This is the second important phase of the proposed model, in which the designed model takes the dataset path as input and starts loading the images into the required format. This process iteratively traverses the dataset path to obtain each individual image path. To read the image data, the openCV object uses the obtained image path. The openCV CV2 object's COLOR_BGR2GRAY method converts the read image into a grayscale channel. Following this process, the grayscaled images are resized to a dimension of 48 x 48. Next, after model segment the image object into a list named X [] and its labels into a list Y []. Both X and Y lists are segmented again into a numpy array object for the further process.

Phase 3: Image Normalization and Splitting- The astype() method converts the numpy array object lists of images into the float 32 format. This method returns a new Data Frame object where the data type changes to the specific type. The Pandas library of Python contains this method, which divides the obtained data frame object by 255 to properly normalize it to a byte. Now, a label encoder method is called to get an encoder object to label the images in between the ranges from 0 to n number of classes. The to-_categorical method in the Keras library uses the encoded label object. This method eventually converts the labels from the 0 to n classes mode into tensorflow object labels for both train and test labels. A splitting range of 20% is set to split the training dataset into the training and validation dataset lists. A random permutation technique is used to split and reconfigure the training images into training and validation sets based on the configured indices.

Phase 4: Construction of Channel boost Convolution neural network Model- To construct the model of channel boost convolution neural network, the normalized image lists along with the image label list are taken into consideration for the purpose of building the CB-CNN model. Initially, to create a plain stack of layers, a sequential method is called where each layer has one input tensor and one output tensor. After this, an early stopping method is called to stop the training of the images when a monitored matrix has stopped improving. This process is carried out at a patience level of 3 and this eventually indicates that if the model is not improving continuously after 3 epochs or iterations, then the model training will be stopped.

To construct the first layer of CB-CNN, the proposed model utilized 32 kernels for the convolution 2D layer with a kernel size of 3 x 3. The activation function 'relu' will activate this layer to identify the maximum number of trained neurons. The first layer also takes the input as the dimension of 48 x 48 with a color channel of one indicated for grayscale input images. The convolutional neural network's second layer houses 64 kernels, each with a size of 3 x 3. And again, the activation function 'relu' is utilized to find the maximum trained neurons. The inclusion of the first and second layers of the convolution layer will be terminated by adding a max pool layer with a dimension of kernel 2 x 2. This process will end by dropping 25% of the untrained tensors in order to move on to the next step of the model.

To construct the third layer of the model, a total number of 128 kernels is being utilized, with the size of each kernel being 3 x 3. A'relu' activation function then extracts the maximum tensor value. The third layer will come to halt a max pooling layer of pool size 2 x 2. Once again, the fourth and final layer of the model sets 128 kernels to train the images, each with a dimension of 2 by 2. This layer also ends with a max pooling 2D layer with dimension 2 X 2. Following the completion of four sequential layers, we add a 25% dropout layer to eliminate the untrained tensors. Often used during the transition between the convolution and fully connected layers, the flatten layer reduces the input's dimensionality from three to two. Before starting off, modify Layer's settings to determine whether TensorSpace Model loads a pre-trained model or not. Hence, a flattening layer is added after the completion of four layers of the convolution. The dense layer, also known as a fully connected layer, is one type of convolutional neural network (CNN) layer. This layer takes input from every neuron in the layer below it. Hence, a dense layer is added with the size of 1024 with the activation function'relu', and this is followed by a dropping percentage of 50 to get the well-trained tensors.

Finally, another dense layer is added with a layer size of the number of classes, that is, seven in our case, representing seven emotions as mentioned in the previous phases. This dense layer is applied with an activation function called softmax to enhance the trained tensor data. To tune the training process quickly, a learning rate of

0.01 is used to calculate the decay rate by dividing the learning rate by the number of epochs. It is possible to optimize gradient descent using adaptive moment estimation. An optimizer known as the 'adam optimizer' excels in handling large-scale problems with multiple parameters or data points. The 'adam' optimizer uses the learning rate and decay rate as inputs to optimize the gradient values. Following this, the model is compiled with custom loss and accuracy as the measuring matrix.

After completing all these steps, the model is trained using the fit method, with the following parameters: train images, train labels, number of epochs with size 100, batch size with dimension 32, early callback, validation images, and validation labels. Upon training the data for the mentioned number of epochs over the input images, the model produces the trained data that will be stored automatically in a file with an extension called.h5.

The Softmax and Relu activation functions that were utilized are represented in equations 1 and 2, respectively.

$$\sigma(Z) = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}} \quad \_\_\_\_\_\_\_(1)$$

Relu=max(0,x) ______(2)

Where

x= neuron value

σ = softmax

z = input vector

e^{zi} = A normal exponential function for the vector that is being entered

K = number of classes

e^{zj} = A normal exponential function for the vector that is being entered for the output

The architecture that was employed while deploying the channel boost Convolution neural network is shown in the below figure 2 and its trained accuracy is depicted in figure 3.

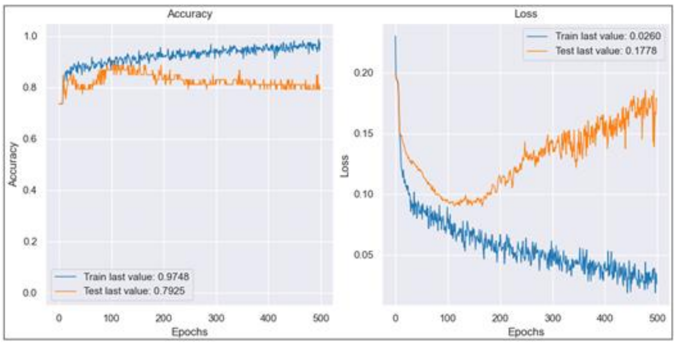

Figure 2: Architecture of Channel boost Convolution neural network model

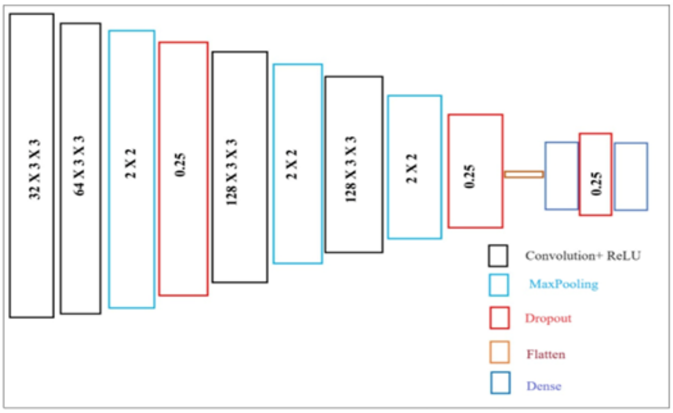

Figure 3: Trained accuracy of the data for CB-CNN model

Phase 5- Testing of CB-CNN for Live emotions in the facial images – In the process of testing, the continuous video frames are gathered from the VideoCapture method of OpenCV. From the live streaming video frames are grabbed at each iteration till the camera is opened. The captured frame is then converted back into a grayscale channel and resized to the dimension of 48 x 48. The resized image is estimated to find the face using the HAAR features; the obtained face image is cropped and converted into a numpy array object to convert the same into the matrix dimension. The obtained expanded matrix dimension data is used by the trained data of the model to predict the presented emotions like anger, disgust, fear, happiness, neutrality, sadness, and surprise.

RESULTS AND DISCUSSIONS

The deployed model is evaluated on a Windows-based, Intel Core i5-powered with 16 GB RAM PC. To test the deployed system, we utilize the following equation to express the confusion matrix's accuracy score parameter. Equations 3, 4, 5, and 6 shows the accuracy, precision, recall, and macro F1 measure correspondingly.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad -(3)$$

$$\text{Precision(P)} = \frac{TP}{TP+FN} \quad -(4)$$

$$\text{Recall(R)} = \frac{TP}{TP+FP} \quad -(5)$$

$$\text{Macro} - \text{F1} = \frac{2*P*R}{P+R} \quad -(6)$$

Here, TP is True positive cases, TN is True Negative cases, FP is False positive cases and FN is False Negative cases.

The proposed model with a channel boost convolution neural network achieved an accuracy of 98.34%, which is comparable to that of [19]. The emphasis in [19] is on a digital learning environment that tracks students' development in real time using face expression recognition software. The platform uses ResNet50, CBAM(Convolutional Block Attention Module), and TCNs (Temporal Convolutional Networks) to improve face expression recognition through periodic picture collection and facial data extraction. Accuracy values of 91.71%, 95.85%, and 97.08% were produced by the dataset. A comparison of the accuracy of [19] and our work on CB-CNN is shown in Table 1 and the graph in Figure 4.

Models	Accuracy
ResNet 50	91.71
CBAM	95.85
TCN	97.08
CB-CNN	98.34

Table 1: Accuracy f1 Score comparison table

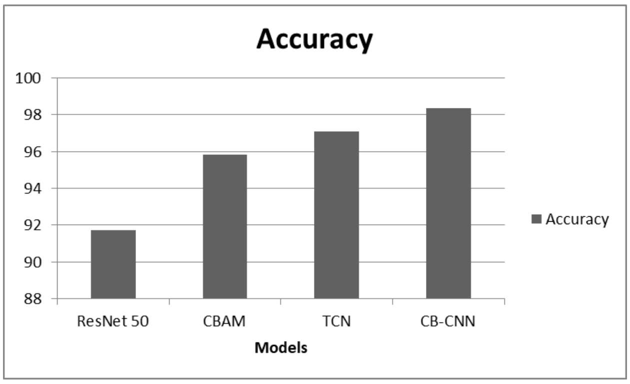

Figure 4: Trained accuracy for CB- CNN model

The graph makes it much clearer that the CB-CNN model we built produces somewhat higher accuracy than the one in [19]. This is due to the fact that the CB-CNN model enhances the model's performance with good accuracy by working on the thorough data segmentation and image channels. For the purpose of determining the error achieved by the designed approach, the Root Mean Square Error (RMSE) is being used. The performance accuracy of the designed approach is indicated by the presence of error in the proposed method for emotion identification using facial expression using CB-CNN model. To evaluate the error between two continuously correlated values, the root-mean-squared (RMSE) method is useful. This method takes into account two parameters: the percentage of emotions correctly detected and the percentage of emotions incorrectly detected. To find the margin of error, we evaluate these numbers using equation 7.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}{n}} \quad \_\ (7)$$

Where,

$\sum$ - Summation , $(x_1 - x_2)^2$ - Differences Squared for the summation in between the expected no. of emotions identifications and the obtained number of emotions identifications ,n - Number of Trails

Table 2 tabulates the RMSE results for the proposed model and the RMSE score of the methodology discussed in [20], and Figure 5 shows the resulting graph.

State	RMSE
Arousal	0.4715
Valence	0.6823
Frequent	0.2987

Table 2: RMSE Records for different state of the emotions

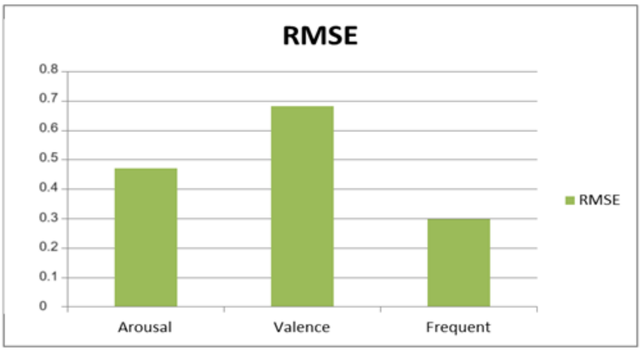

Figure 5: RMSE Comparison of different state of the emotions

Deep-Emotion, a deep learning-based multimodal emotion recognition (MER) system, can adaptively combine the most discriminative information from speech, facial expressions, and electroencephalogram (EEG), as described in [20]. The RMSE for the arousal and valence modes of facial emotions is measured by the model in [20], which provides values of 0.4715 and 0.6823, respectively. However, when compared to the proposed model's frequent mode of facial emotions, these RMSE results come out to be 0.2987. It clearly shows that the RMSE value of the given method is lower than that of in [20]. Working on the image channels and data segmentation thoroughly allowed the proposed CB-CNN model to obtain good RMSE. The model in [20] uses average dynamic weight assignment during data training, however. As a result, the CB-CNN outperforms [20] in terms of RMSE when it comes to frequent emotions.

CONCLUSION AND FUTURE SCOPE

The purpose of this paper is to assess facial emotions across various age groups using a deep learning model, specifically the channel boost convolution neural network. The designed model underwent extensive training on approximately 28709 images, with an additional 7178 images used for testing. The obtained train and test images are initially optimized for boosting the channel by converting them into an absolute grayscale pattern. The obtained absolute grayscale images are then segmented properly by converting them into expandable binary matrices. After this, the image list is classified for the respective images and label list, which is then subjected to training the model using the channel boost convolution neural network model for 500 epochs to obtain an accuracy of 97.48. We efficiently test the trained model using real-time face input from the camera to ensure the model's feasibility. The results of CB-CNN are tested for the RMSE and accuracy to get the best values of **0.2948 and** 98.34%, respectively.To enhance the accuracy of all augmented images in real time, we can extend the designed model to work on a transformer algorithm in the future.

REFERENCES

[1] M. R. Falahzadeh, E. Z. Farsa, A. Harimi, A. Ahmadi and A. Abraham, "3D Convolutional Neural Network for Speech Emotion Recognition With Its Realization on Intel CPU and NVIDIA GPU," in *IEEE Access*, vol. 10, pp. 112460-112471, 2022.

[2] D. Peña, A. Aguilera, I. Dongo, J. Heredia and Y. Cardinale, "A Framework to Evaluate Fusion Methods for Multimodal Emotion Recognition," in *IEEE Access*, vol. 11, pp. 10218-10237, 2023.

[3] Y. C. Yoon, "Can We Exploit All Datasets? Multimodal Emotion Recognition Using Cross-Modal Translation," in *IEEE Access*, vol. 10, pp. 64516-64524, 2022.

[4] L. Yunxiang and Z. Kexin, "Design of Efficient Speech Emotion Recognition Based on Multi Task Learning," in *IEEE Access*, vol. 11, pp. 5528-5537, 2023.

[5] B. T. Atmaja and A. Sasou, "Evaluating Self-Supervised Speech Representations for Speech Emotion Recognition," in *IEEE Access*, vol. 10, pp. 124396-124407, 2022.

[6] C. -L. Hwang, Y. -C. Deng and S. -E. Pu, "Human–Robot Collaboration Using Sequential-Recurrent-Convolution-Network-Based Dynamic Face Emotion and Wireless Speech Command Recognitions," in *IEEE Access*, vol. 11, pp. 37269-37282, 2023.

[7] F. Andayani, L. B. Theng, M. T. Tsun and C. Chua, "Hybrid LSTM-Transformer Model for Emotion Recognition From Speech Audio Files," in *IEEE Access*, vol. 10, pp. 36018-36027, 2022.

[8] H. Wang, X. Zhao and Y. Zhao, "Investigation of the Effect of Increased Dimension Levels in Speech Emotion Recognition," in *IEEE Access*, vol. 10, pp. 78123-78134, 2022.

[9] S. Pant *et al.*, "Korean Drama Scene Transcript Dataset for Emotion Recognition in Conversations," in *IEEE Access*, vol. 10, pp. 119221-119231, 2022.

[10] A. Landowska, T. Zawadzka and M. Zawadzki, "Mining Inconsistent Emotion Recognition Results With the Multidimensional Model," in *IEEE Access*, vol. 10, pp. 6737-6759, 2022.

[11] Y. Peng, F. Jin, W. Kong, F. Nie, B. -L. Lu and A. Cichocki, "OGSSL: A Semi-Supervised Classification Model Coupled With Optimal Graph Learning for EEG Emotion Recognition," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1288-1297, 2022.

[12] S. Dwijayanti, M. Iqbal and B. Y. Suprapto, "Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network," in *IEEE Access*, vol. 10, pp. 89876-89886, 2022.

[13] J. Santoso, T. Yamada, K. Ishizuka, T. Hashimoto and S. Makino, "Speech Emotion Recognition Based on Self-Attention Weight Correction for Acoustic and Text Features," in *IEEE Access*, vol. 10, pp. 115732-115743, 2022.

[14] B. Subramanian, J. Kim, M. Maray and A. Paul, "Digital Twin Model: A Real-Time Emotion Recognition System for Personalized Healthcare," in *IEEE Access*, vol. 10, pp. 81155-81165, 2022.

[15] L. -M. Zhang, G. W. Ng, Y. -B. Leau and H. Yan, "A Parallel-Model Speech Emotion Recognition Network Based on Feature Clustering," in *IEEE Access*, vol. 11, pp. 71224-71234, 2023.

[16] M. Wu, W. Teng, C. Fan, S. Pei, P. Li and Z. Lv, "An Investigation of Olfactory-Enhanced Video on EEG-Based Emotion Recognition," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1602-1613, 2023.

[17] K. L. Ong, C. P. Lee, H. S. Lim, K. M. Lim and A. Alqahtani, "Mel-MViTv2: Enhanced Speech Emotion Recognition With Mel Spectrogram and Improved Multiscale Vision Transformers," in *IEEE Access*, vol. 11, pp. 108571-108579, 2023.

[18] D. Pan *et al.*, "MSFR-GCN: A Multi-Scale Feature Reconstruction Graph Convolutional Network for EEG Emotion and Cognition Recognition," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3245-3254, 2023

[19] Mohammed Aly, "Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model", Springer, 2024.

[20] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang and S. Wang, "Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 5, pp. 396-403, 2024.

[21]https://drive.google.com/file/d/1X60B-uR3NtqPd4oosdot pbDgy8KOfUdr/ view.