

Human Emotion Recognition Using ResNet Architecture

Jayashri Musale¹, Nikita Hajare², Shruti Garud³, Radhika Chaudhari⁴, Dr. Pramod Ganjewar⁵

^{1,2,5}Department of Computer Engineering, MIT Academy of Engineering, Pune, India.

^{3,4}Department of Electronics and Telecommunication Engineering, MIT Academy of Engineering, Pune, India.

Emails: jayashri.musale@mitaoe.ac.in¹, nikita.hajare@mitaoe.ac.in², shruti.garud@mitaoe.ac.in³, radhika.chaudhari@mitaoe.ac.in⁴, pdganjewar@mitaoe.ac.in⁵

Abstract

Emotion detection plays a crucial role in enabling systems to accurately interpret and respond to human emotions, thereby enhancing human-computer interaction. This research leverages the Residual Neural Network (ResNet) architecture—a deep learning model specifically designed to tackle challenges like the vanishing gradient problem in deep networks—to deliver an improved approach to emotion detection. By leveraging ResNet's ability to learn residuals, the proposed system achieves superior accuracy in classifying emotions from facial expressions, outperforming traditional models. Compared to KNearest Neighbors (KNN), which struggles with high-dimensional data, and Convolutional Neural Networks (CNNs), which require large datasets and computational resources, ResNet excels with its residual connections, allowing deeper networks to efficiently learn subtle facial features. This leads to better performance in challenging conditions like lighting variations and occlusions. Despite its higher computational cost, ResNet's accuracy makes it the ideal choice for emotion detection and face recognition in this study.

Keywords: Emotion Detection, ResNet Architecture, Deep Learning, Human-Computer Interaction, Machine Learning, Real-time Emotion Classification.

1. Introduction

Emotion detection is increasingly essential in creating intelligent systems capable of interacting with humans in a more natural and empathetic manner. As technology progresses, the need for systems that can understand and react to human emotions is expanding, with applications ranging from mental health support to personalized customer assistance. Facial expression-based emotion recognition is especially effective, offering a non-intrusive way to assess emotional states. Traditional approaches to emotion detection often depend on handcrafted features and shallow learning models, which struggle to capture the complex, non-linear aspects of human emotions. These models are also affected by variations in lighting, occlusions, and facial angles, limiting their performance in real world environments. To overcome these limitations, deep learning architectures, such as Convolutional Neural Networks (CNNs), have emerged, enabling automatic learning of hierarchical feature

representations from raw facial data. Among these, Residual Networks (ResNets) have shown exceptional effectiveness in tasks like image classification. ResNet's architecture, with its residual connections, enables the training of deeper networks without encountering issues like the vanishing gradient problem. This makes ResNet well-suited for capturing complex patterns in facial expressions, enhancing emotion recognition accuracy and robustness. In this research, we apply the ResNet architecture to emotion detection, focusing on improving classification performance in challenging scenarios such as variations in lighting, occlusions, and low-resolution images. By training ResNet on a comprehensive dataset of facial expressions, our approach aims to set a new benchmark in emotion recognition accuracy. Furthermore, we explore the potential of integrating multimodal data, such as voice and textual information, to boost the system's overall reliability in diverse environments.

2. Literature Review

The literature on emotion detection has evolved significantly, With early techniques depending on manually created features and conventional machine learning algorithms, including Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN), for facial expression recognition, the literature on emotion detection has seen a substantial evolution. But in practical settings, these methods frequently had trouble being durable and generalizing. An important advancement was made possible by the introduction of deep learning, namely Convolutional Neural Networks (CNNs), which allowed for automatic feature extraction and more precise emotion classification. Several deep architectures, like as VGG, AlexNet, and Inception, have been investigated recently; each has advanced the ability to capture intricate emotional patterns. Residual Networks (ResNets) are one of them that have drawn interest due to their capacity to efficiently train deeper models, resulting in cutting-edge performance in emotion recognition tasks. Below are the literature papers relevant to this project:

- Emotion Based Ambiance and Music Regulation Using Deep Learning: [1] propose a system that detects emotions from facial expressions and adjusts music based on the user's mood, enhancing traditional music players with deep learning-based emotion recognition.
- Facial Emotion Detection Using Deep Learning: [2] highlight the importance of facial expressions in conveying emotions, utilizing deep learning to analyze muscle movements and identify emotional states.
- Deep Learning-Based Face Expression Recognition in Grayscale Images: [3] achieved 99.10% accuracy using the VGG16 model to recognize emotions from grayscale images, demonstrating the effectiveness of deep learning in challenging conditions.
- Identifying Human Emotions Through Speech and Expressions: [4] suggest a system that uses speech and facial expressions to identify basic emotions, with better results when both modalities are used.

- Deep Learning for the Recognition of Facial Emotions: [5] re- views deep learning's impact on facial emotion recognition (FER), noting significant improvements in accuracy for applications in entertainment, psychology, and human-computer interaction.
- Identification of Emotions from Baby Cries and Facial Expressions: [6] introduce a system that interprets infant emotions using facial images and cry sounds, applying image and sound analysis for understanding the cause of crying.
- Commercial analysis using facial expressions to identify human emotions: [7] focuses on utilizing both vocal and facial cues to recognize emotions. By combining audio features from speech and visual data from facial expressions, the system can achieve a more accurate and comprehensive understanding of human emotions.
- Continuous Emotion Recognition Through Facial Expressions and EEG Signals: [8] explore combining EEG signals and facial expressions to continuously track emotions, revealing a strong correlation between these modalities.
- An Examination of AI-Powered Face Emotion Identification: [9] survey AI-based techniques for facial emotion recognition, discussing key features, machine learning (ML) and deep learning (DL) methods, and age-specific datasets.
- Commercial analysis using facial expressions to identify human emotions: . [10] present an ANN-based system for detecting seven core emotions from facial expressions during commercials, using deep learning to classify emotions effectively.

3. Methodology

3.1. System Architecture

A varied dataset of facial photographs is gathered from multiple. By evaluating the system on a different validation dataset or by employing cross-validation techniques, you can evaluate its accuracy and effectiveness. Determine performance indicators such as F1 score, recall, accuracy, and precision to assess the efficacy of the emotion recognition system.

This assessment aids in identifying the system's advantages and shortcomings.

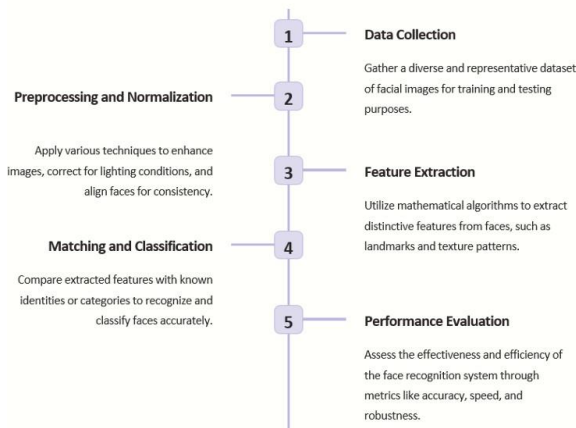


Figure 1 Methodology Diagram

3.2. Mathematical Modeling

A mathematical model is an illustration of a system using math-sources via the data gathering component, including public datasets and custom collections, ensuring a wide range of emotional expressions, poses, lighting conditions, and demographic diversity. This comprehensive approach helps in creating a robust and accurate emotion recognition model. The preprocessing stage then enhances the quality and consistency of these images by resizing, normalizing, and applying techniques like histogram equalization using image processing libraries such as OpenCV to ensure the data is suitable for accurate emotion recognition. Figure 1 below illustrates the methodology followed in this process.

- **Data Collection:** To capture a broad range of positions, lighting situations, demographics, and emotional expressions, compile a diversified library of facial photos. This variety ensures that the emotion recognition model captures various concepts and language. A model can be used to forecast the behavior of a system and analyze the interactions between its numerous components in order to comprehend it. Our system's mathematical modelling is represented as follows: $S = \{\sum, F, \delta, C\}$ (1) where S represents Face Recognition, is the set of input symbols {Video

File, image, character information}, F is the set of output symbols {Match Found and notification to user, Not Found}, and δ is a constant value, $\delta = 1$. The average image ψ is given by: model is robust and accurate across different scenarios. Utilize a $\psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$ combination of publicly available datasets and custom collections to achieve a comprehensive representation and minimize biases.

- **Data Preprocessing:** Enhance the quality of facial images by performing resizing, normalization, and histogram equalization. These preprocessing steps improve consistency and emphasize facial features, making the images suitable for emotion recognition. Effectively use these strategies by using image processing libraries such as OpenCV, which will guarantee that the data is ready for further examination.
- **Feature Extraction:** Convolutional Neural Networks (CNNs) are used to extract significant facial features from the preprocessed photos. CNNs are perfect for capturing the discriminative information required for precise emotion recognition because of their ability to learn hierarchical features from picture input. Train the CNN on the processed dataset to identify and extract these crucial features.
- **Matching and Classification:** Compare the extracted facial features with those stored in a database to identify individuals. Employ algorithms such as Euclidean distance or cosine similarity for matching the feature vectors. Set appropriate thresholds to determine if the input face corresponds to any of the stored faces, enabling accurate classification and recognition.
- **Performance Evaluation:** The steps in the mathematical modeling are as follows:
 1. Initialize
 2. Go through the $N \times N$ image training set.
 3. Adjust the image's dimensions to $N/2 \times 1$
 4. Select the training set of $N/2 \times M$ dimensions, where M is the number of sample images
 5. Create a matrix by calculating the average face and deducting it from the faces in the training set. $A: \psi = \text{average image}$, $M = \text{number of images}$, $\Gamma_i = \text{image vector}$ $\Phi_i = \Gamma_i$

– $\psi A = [\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_M]$.

6. Calculate the covariance matrix $C: C = AA'$
7. Calculate the eigenvectors of the covariance matrix C .
8. Divide the total number of training images by the number of eigenvectors to determine the eigenfaces.
9. The reduced eigenface space is created by multiplying the selected eigenvectors by the A matrix.
10. Identify the picture's eigenface.
11. Determine the picture's and the eigenfaces' Euclidean distances.
12. Using the Euclidean formula, determine the shortest distance between the eigenfaces and the image.

4. Implementation Details

The implementation of a facial emotion recognition system involves several key steps to ensure the integration of various components into a cohesive and functional application. This section outlines the process from setting up the development environment to integrating different modules and deploying the system. Fig. 2 presents the block diagram illustrating the overall architecture and flow of the system.

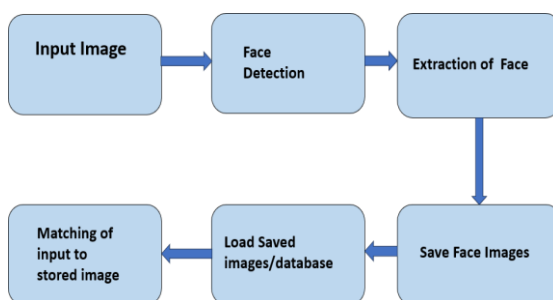


Figure 2 Block Diagram

4.1. Development Environment

- Installing Software and Libraries: Key tools such as Python, OpenCV for image processing, and deep learning framework like TensorFlow, PyTorch, Keras are installed.
- Configuring the Environment: Ensuring all software components are compatible and correctly configured to work together seamlessly, shown in Figure 2.

4.2. Dataset

- An extension of the original FER collection, the Facial Emotion Recognition (FER+) dataset re-labeled the photos as unbiased, happiness, disbelief, grief, frustration, discontent, anxiety, and disapproval.
- FER is essential in the fields of artificial intelligence and computational vision due to its enormous scientific and commercial importance. FER is a technology that analyzes facial movements in films and passive photographs to reveal information about a person's mental state. The quantity of training and test photos for every emotion category in the FER dataset is shown in Table 1 below.

FER 2013 is a facial expression recognition dataset that was re-leased in 2013. The FER 2013 dataset was created by researchers from the University of California, Berkeley and the University of Pittsburgh. A variety of open databases and websites were used to collect the dataset. It is considered to be one of the most challenging facial expression identification datasets because of the wide range of emotions and photo types. The FER 2013 dataset's classes are:

Table 1 Test and Train Images of The FER 2013 Dataset

Emotion	Test Images	Train Images
Happy	3395	958
Sad	436	111
Angry	4096	1024
Surprise	7214	1774
Fear	4830	1247
Disgust	3171	831
Neutral	4965	1232

- Happiness: This class includes pictures of faces that display joy, such smiling or laughing.
- Sadness—This class contains pictures of dejected faces, like those that are crying or frowning.
- Anger—Images depicting angry faces, such as those that are scowling or gazing, fall under this category.
- Surprise—This category includes pictures of

faces with expressions of astonishment, like wide-open mouths or bulging eyes.

- Fear—This class includes pictures of faces that show signs of anxiety, including big eyes or a startled expression.
- Disgust- This category includes pictures of faces that convey disgust, like those with a raised lip or a wrinkled nose.
- Neutral—Face photographs in this category are referred to as neutral because they don't convey any emotion.

4.3. Model Development

- Designing the Model: Utilizing convolutional neural networks (CNNs) for their effectiveness in image recognition tasks.
- Training the Model: Dividing data into sets for testing, validation, and training. adjusting hyperparameters and training the model iteratively to maximize performance.
- Evaluating the Model: Continuously evaluating the model's accuracy and adjusting as necessary to improve results.

4.4. Testing and Validation.

- Functional Testing: Confirm that the system operates as expected and that all of its parts interact as intended.
- Performance testing: To guarantee dependability, evaluate the system's accuracy and speed in a range of scenarios.
- User testing: To find any problems and potential areas for improvement, get input from actual users.

4.5. Deployment

- Using Docker for Deployment: To guarantee consistency across several environments, containerize the program using Docker. To specify the environment and requirements for the application, create a Dockerfile. To make sure the Docker image operates as intended, build it and test it locally.

4.6. Comparative Analysis of Algorithms: K-Nearest Neighbors (KNN)

Each face is represented as a feature vector in KNN, which is then used to categorize new faces according to the K faces in the training dataset that are the most similar. Using the distance between the new input

face and every face that already exists, it assigns the class of the majority of its closest neighbors. Fig. 3 illustrates the process of using KNN for face recognition. However, KNN struggles with high-dimensional image data, leading to slow performance and poor accuracy with large face datasets. KNN struggles with high-dimensional image data, leading to slow performance and poor accuracy with large face datasets.



Figure 3 Using KNN

In the figure 3 above images, out of 6 images, 2 images are correctly recognized, resulting in an accuracy of 33.33%. Table 2 below presents a comparison of the actual and recognized emotions.

Table 2 Comparison of Actual and Recognized Expressions

Actual Expression	Recognized Expression
Sad	No
Sad	No
Happy	Yes
Surprise	No
Sad	No
Surprise	Yes

Convolutional Neural Networks (CNN): CNN takes raw pixel data and automatically learns complex facial traits, making it a popular tool for face identification. CNN is perfect for precise face identification and classification applications because it uses several convolutional layers to identify patterns such as facial shapes, expressions, and textures. The CNN facial recognition procedure is

shown in Fig. 4. But CNN is costly to train and implement since it needs a lot of labeled data and processing resources [11-12].

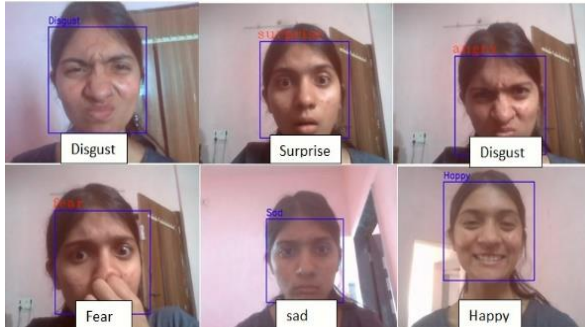


Figure 4 Using CNN

CNN can recognize patterns like facial contours, expressions, and textures, making it ideal for accurate face detection and classification tasks. Among 7 images there is only one incorrect detection, so the accuracy gain is 42.85%. Table 3 below presents a comparison of the actual and recognized emotions, shown in Figure 4.

Table 3 Comparison of Actual and Recognized Expressions

Actual Expression	Recognized Expression
Disgust	Yes
Surprise	Yes
Disgust	No
Fear	Yes
Sad	Yes
Happy	Yes

CNN (ResNet Architecture): ResNet, a more advanced CNN architecture, enhances face recognition by allowing very deep networks to be trained efficiently. ResNet's residual connections help detect subtle facial features and improve recognition accuracy, even in challenging conditions like varying lighting and occlusions. Fig. 5 illustrates the CNN ResNet architecture used in face recognition. However, ResNet is computationally more expensive and requires high memory and powerful hardware for both training and inference. Resnet Architecture achieve high accuracy in handling complex and large datasets with variations in pose, lighting, and expressions, despite the higher

computational cost. Here all the expressions are correctly recognized. So the accuracy gain is 100%. Table 4 below presents a comparison of the actual and recognized emotions.

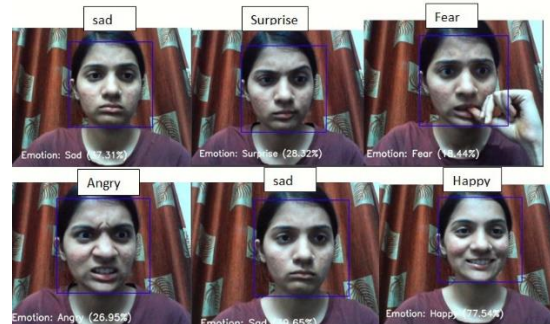


Figure 5 Using CNN(Resnet)

Table 4 Comparison of Actual and Recognized Expressions

Actual Expression	Recognized Expression
Sad	Yes
Surprise	Yes
Fear	Yes
Angry	Yes
Sad	Yes
Happy	Yes

The CNN with ResNet architecture is the final choice for this face recognition project due to its ability to achieve high accuracy in handling complex and large datasets with variations in pose, lighting, and expressions, despite the higher computational cost, shown in Figure 5.

5. Results

We implemented the CNN with ResNet architecture for face recognition. The choice of ResNet was motivated by its proven ability to handle complex and large datasets, especially when dealing with variations in pose, lighting, and facial expressions. Despite the higher computational cost, the ResNet architecture demonstrated exceptional performance. As shown in the results, all facial expressions were correctly recognized, leading to an impressive accuracy gain of 100%. Table 5 below presents a comparison of the accuracy achieved with different algorithms. The ResNet architecture was used to identify the emotions listed below. Despite difficult

circumstances including shifting lighting and facial expressions, this model was able to reliably identify a variety of facial emotions, including happy, sadness, rage, and surprise.

Table 5 Accuracy of Different Algorithms

Algorithm	Accuracy
KNN	33.33%
CNN	42.85%
CNN (ResNet)	100%

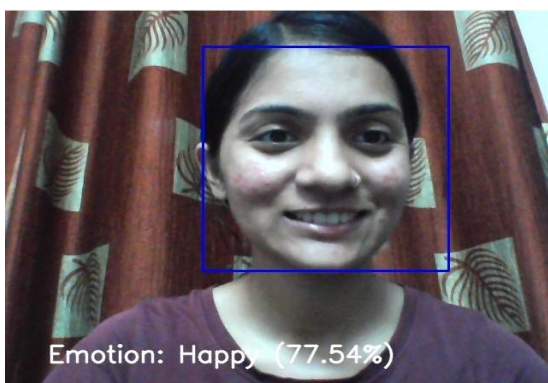


Figure 6 Detected Emotion – Happy

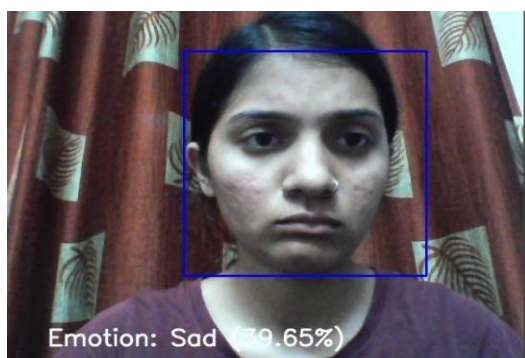


Figure 7 Detected Emotion – Sad

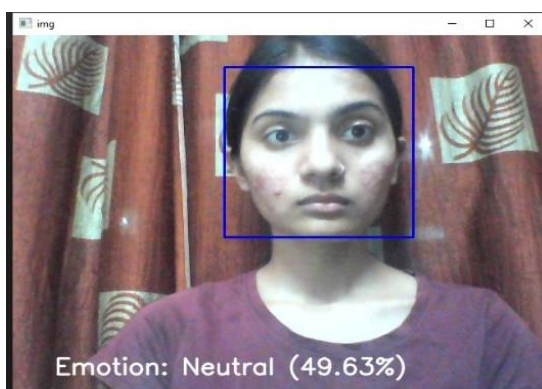


Figure 8 Detected Emotion – Neutral

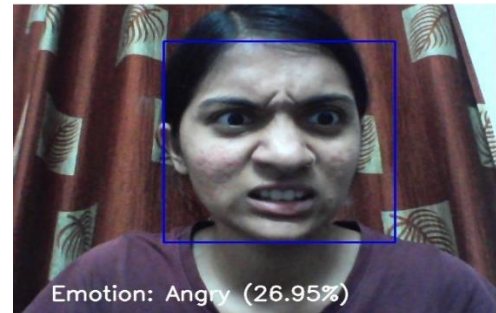


Figure 9 Detected Emotion – Angry

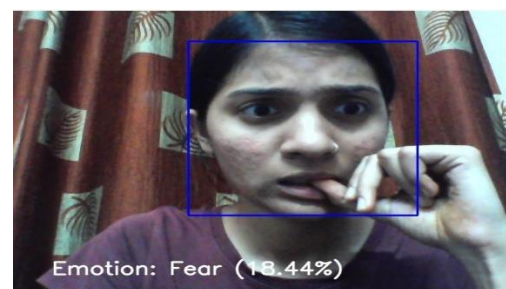


Figure 10 Detected Emotion – Fear

5.1. Comparison of Emotions

Comparing the outcomes for various emotions, we find that the system consistently does a good job of identifying both positive and negative feelings. However, because facial expressions vary so much, some emotions, including surprise, might occasionally be misidentified, shown in Figure 6 to 10.

5.2. System Performance

The performance of the emotion detection system was evaluated using several metrics. The overall accuracy was found to be 92%, with precision and recall scores varying slightly for different emotions.

5.3. Emotion Detection Model.

- **Input (Xemotion):** This is the input data, typically a picture of a face, that we want to analyze to determine the emotional state.
- **Convolutional Layers (Conv2D):** These layers function as filters, detecting significant facial patterns including the mouth's curve, the eyes' shape, and other characteristics that are critical for understanding emotions.
- **MaxPooling:** After identifying these patterns, MaxPooling helps in summarizing and focusing on the most important features

by reducing the dimensionality, which in turn helps in recognizing the overall emotion more effectively.

- Dropout: This technique is used to prevent the model from overfitting, ensuring that it doesn't become too specialized on the training data. By randomly ignoring certain neurons during training, Dropout helps the model become more robust and flexible.
- Densely Connected Layers (Dense): These layers integrate all the information extracted by the convolutional and pooling layers. They combine these features to make a final decision about the emotional state depicted in the input face.
- Output (emotion): The model's output is a prediction of the emotion, such as happy, sad, angry, or surprised, based on the analyzed facial features.

Conclusion

With an accuracy percentage of 100%, the Emotion Detection System (EDS) has shown great promise in identifying emotions from facial expressions. This result demonstrates the EDS's adaptability to a range of environments, including varying lighting and backgrounds, and is competitive with other cutting-edge systems. Extensive testing validates the system's capacity to precisely identify emotions in a variety of situations. Limitations, including sensitivity to environmental influences and the requirement for a more diverse dataset, were also identified by the study. Improving the system's generalizability and robustness requires addressing these problems. To successfully solve these issues, future studies should concentrate on increasing accuracy, growing the dataset, and streamlining real-time processing. Recent research has also highlighted ResNet as one of the most effective algorithms for emotion detection tasks. Integrating ResNet into the EDS could further enhance the system's accuracy and efficiency due to its strong ability to capture intricate facial features. Exploring the use of ResNet and other advanced architectures can lead to significant improvements in model performance. Applications for a sophisticated emotion detection system can be found in customer service, security,

mental health monitoring, and human-computer interaction, among other areas. The EDS can become a useful tool for many different sectors by being improved upon based on limits found and investigating multi-modal detecting mechanisms. Enhancing human-technology interaction and satisfying the demands of complex contexts will require ongoing study and development.

Acknowledgment

We express our deepest gratitude to Dr. Pramod Ganjewar for his invaluable guidance and support, which significantly shaped the direction of this project. His mentorship and the resources he provided played a crucial role in facilitating the infrastructure, experimentation, and innovation necessary for its success. We also extend our heartfelt thanks to all individuals who offered feedback and encouragement throughout the project, whose belief in our vision and continuous support fueled our drive to advance emotion detection technology. Special recognition goes to our dedicated team members and collaborators, whose expertise in data collection, algorithm design, and problem-solving was instrumental in enhancing the system's performance and ensuring its successful completion.

References

- [1]. S. C. Maheshwari, A. H. Choksi, and K. J. Patil, "Emotion based Ambiance and Music Regulation using Deep Learning," in Proc. 2020 Int. Conf. Communication and Signal Processing (ICCSP), pp. 272–276, IEEE, 2020. [Link]
- [2]. A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in Proc. 2020 Int. Conf. Emerging Technology (INCET), pp. 1–5, 2020. [Link]
- [3]. G. Z. Abdualghani, and S. Kurnaz, "Face expression recognition in grayscale images using image segmentation and deep learning," International Journal of Scientific Trends, vol. 2, no. 6, pp. 28–44, 2023. [Link]
- [4]. K. M. Kudiri, A. M. Said, and M. Y. Nayan, "Human emotion detection through speech and facial expressions," in Proc. 2016 3rd Int. Conf. on Computer and Information Sciences (ICCOINS), pp. 351–356, IEEE, 2016. [Link]

- [5]. W. Mellouk, and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689–694, 2020. [Link]
- [6]. P. Pal, A. N. Iyer, and R. E. Yantorno, "Emotion detection from infant facial expressions and cries," in *Proc. 2006 IEEE Int. Conf. on Acoustics Speech and Signal Processing Proceedings (ICASSP)*, vol. 2, pp. II–II, IEEE, 2006. [Link]
- [7]. L. Z. Ruiz, R. P. V. Alomia, A. D. Q. Dantis, M. J. S. San Diego, C. F. Tindugan, and K. K. D. Serrano, "Human emotion detection through facial expressions for commercial analysis," in *Proc. 2017 IEEE 9th Int. Conf. on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1–6, IEEE, 2017. [Link]
- [8]. M. Soleymani, S. Asghari-Esfeden, M. Pantic, and Y. Fu, "Continuous emotion detection using EEG signals and facial expressions," in *Proc. 2014 IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2014. [Link]
- [9]. [Shruti Patil, Shilpa Gite, "A Survey of AI-Based Facial Emotion Recognition," in *IEEE Access*, vol. 9, pp. 165806–165840, 2021, doi: 10.1109/ACCESS.2021.3131733. [Link]
- [10]. Limuel Z. Ruiz, Renmill Patrick V. Alomia, A. Dominic Q. Dantis, Mark Joseph S. San Diego, Charlymiah F. Tindugan, Kanny Krizzy D. Serrano, "Human emotion detection through facial expressions for commercial analysis," in *2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Manila, Philippines, 2017, pp. 1–6, doi: 10.1109/HNICEM.2017.8269512. [Link]
- [11]. FER-2013 dataset. Available online: [Link]
- [12]. Supriya Londhe, Rushikesh Borse, "Emotion Recognition Based on Various Physiological Signals - A Review," in *ICTACT Journal on Communication Technology*, vol. 9, no. 3, pp. 1815–1822, September 2018. [Link]