KEC Conference 2025

ISSN: 3021-940X (Print)

DOI: https://doi.org/10.3126/injet.v2i2.78596

Human Emotion Detection and Face Recognition System

Bikash Kumar Jha[1], Bharat Paudel[2], Adarsh Mishra[3], Aabik Maharjan[4, *], Pralhad Chapagain[5]

[1]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, Nepal, bikashzha2058@gmail.com*

[2]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, Nepal, bharatpaudel1010@gmail.com*

[3]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, Nepal, adarshyoyo170@gmail.com*

[4]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, Nepal, aabeekmhj@gmail.com*

[5]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, Nepal, pralhadchapagain@kec.edu.np*

Abstract

This study presents an integrated Human Emotion Detection and Face Recognition system, combining computer vision and deep learning to perform real-time facial analysis. The system processes live video to recognize individuals and classify emotions into seven categories (angry, disgust, fear, happy, neutral, sad, surprise) using a Convolutional Neural Network (CNN) trained on the augmented and filtered FER-2013 datasets. Face recognition is achieved through OpenCV's Haar Cascade for detection and SVM (Support Vector Machine)/KNN (K-Nearest Neighbor) for matching facial features. The system pre-processes the image data which includes grayscale conversion for the optimal CNN and SVM input. The system features an interactive interface with secure authentication, real-time overlays for emotion and identity visualization, and dynamic thresholding to enhance accuracy. Moreover, the system generates dataset of face and emotion detected in CSV file format and generates chart accordingly. The test accuracy obtained from custom CNN model is 74.78%. This project offers significant opportunities for future research, as it intersects with a variety of fields including AI, computer vision, healthcare, education, human behavior, security, and ethics.

Keywords: Computer Vision, OpenCV, FER-2013, CNN, SVM

1. Introduction

Emotions result from the interplay of mental and physiological states, are fundamental to human experience and are expressed through behaviors, thoughts, and physical cues such as facial expressions. The most common purpose of the analysis of facial expressions is to determine the current emotion of a particular person (Bialek et al., 2023). The features displayed on emotion images were derived with a CNN, and these emotional features were visualized to determine the facial landmarks that contains major information (Huang et al., 2023). These landmarks act as physiological markers, bridging the gap between raw visual data and interpretable affective states. Human facial expressions are linked to emotions, and learning to depict them is an important step for creating seamless human to machine communication (Babajee et al., 2020). Facial emotion recognition (FER) has gained significant attention in fields such as artificial intelligence, human-computer interaction, and psychological research. By leveraging deep learning techniques, particularly Convolutional Neural Networks (CNNs), FER systems can effectively analyze micro expressions and subtle facial variations to classify emotions. Monitoring students' emotions during classroom sessions and exams can provide useful information on their engagement, focus, and stress levels in educational settings. Implementing an emotion detection and facial recognition system in schools and colleges can assist educators in identifying pupils who may be having difficulty learning or are anxious during tests. This can help improve remote learning situations by providing feedback on students' emotional responses, allowing teachers to tailor their instruction accordingly.

*Corresponding Author

2. Related Works

Facial emotion recognition has become a critical area of research with applications in human-computer interaction and mental health assessment. This section reviews key research contributions that have advanced the field, focusing on the use of Convolutional Neural Networks (CNNs) and other machine learning algorithms for facial emotion recognition.

In (Kwong et al., 2018), the effectiveness of various machine learning algorithms in recognizing human emotions through facial expressions was investigated. They experimented with traditional classifiers such as Support Vector Machines (SVM) and Decision Trees, as well as contemporary deep learning methods like Convolutional Neural Networks (CNNs). The study aimed to determine the optimal combination of feature descriptors and machine learning models for accurate emotion recognition. The authors utilized datasets containing labeled facial images to train and evaluate their models, assessing performance using metrics such as accuracy, precision, recall, and F1-score. Their findings contribute to advancements in emotion detection technology, with potential applications in areas like human-computer interaction and psychological analysis. In (Jaiswal et al., 2020), an artificial intelligence system to detect human emotions through facial expressions was developed. Their approach involves three primary steps: face detection, feature extraction, and emotion classification. They proposed a deep learning architecture based on Convolutional Neural Networks (CNNs) for this task (Jaiswal et al., 2020). The performance of their model was evaluated using two datasets: the Facial Emotion Recognition Challenge (FERC-2013) and the Japanese Female Facial Expression (JAFFE) dataset. The model achieved accuracies of 70.14% on FERC-2013 and 98.65% on JAFFE, indicating its effectiveness in recognizing emotions from images.

While previous research in emotion detection and face recognition has largely focused on emotion classification using Convolutional Neural Networks (CNNs) and various machine learning models such as Support Vector Machines (SVM) and Decision Trees, the Human Emotion Detection and Face Recognition System enhances these approaches by integrating both emotion recognition and real-time face identification into a single system. The system leverages CNNs for emotion classification and Haar Cascade for efficient face detection, ensuring high accuracy in real-world scenarios. In addition, SVM and KNN algorithms are employed for robust face recognition, allowing the system to identify individuals based on facial features. A key feature of the system is its ability to operate in real-time, with an interactive user interface that processes live video feeds, detects emotions like happiness, sadness, fear, and surprise, and recognizes faces simultaneously.

Beyond simple detection, the system generates detailed logs of recognized faces and emotions, which are stored in CSV files and visualized through charts for further analysis. This makes the system highly valuable for applications in educational supervision, where it can be used to monitor and analyze students' emotional states during lessons or examinations. The system’s ability to combine emotion detection, face recognition, and data logging sets it apart from traditional emotion recognition systems, offering both accurate results and extensive functionality for ongoing analysis and study.

Previous research has extensively explored emotion detection and face recognition as independent tasks, often relying on computationally intensive models or controlled laboratory conditions. Our work addresses this limitation by integrating both tasks into a lightweight, real-time system optimized for edge deployment. Using Haar Cascade for efficient face detection combined with a custom CNN for emotion classification (74.78% accuracy on FER-2013) and SVM/KNN for recognition, we achieve practical performance on resource-constrained hardware. The system further incorporates automated data logging and visualization, targeting educational applications where existing solutions remain impractical. This unified approach demonstrates how real-world affective computing can balance accuracy and efficiency.

3. Methodology

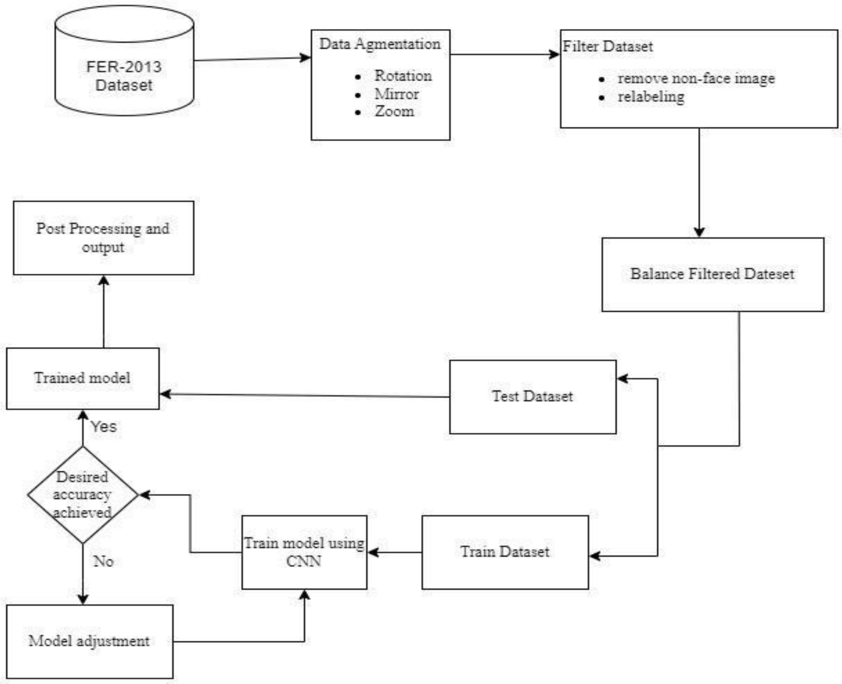

Figure 1. Block Diagram for Emotion Detection System

The system begins with input images subjected to preprocessing steps like grayscale conversion and resizing to standardize inputs. Data augmentation (rotation, mirroring, zoom) expands dataset diversity, while filtering removes non-face images and corrects labels. A balanced dataset is created to address class imbalance, ensuring unbiased training. A Convolutional Neural Network (CNN) is trained on this refined data, with iterative adjustments to improve accuracy. Though preliminary results are achieved, the figure notes that target accuracy remains unmet, highlighting the need for further optimization. The pipeline emphasizes robust data preparation and deep learning for emotion recognition.

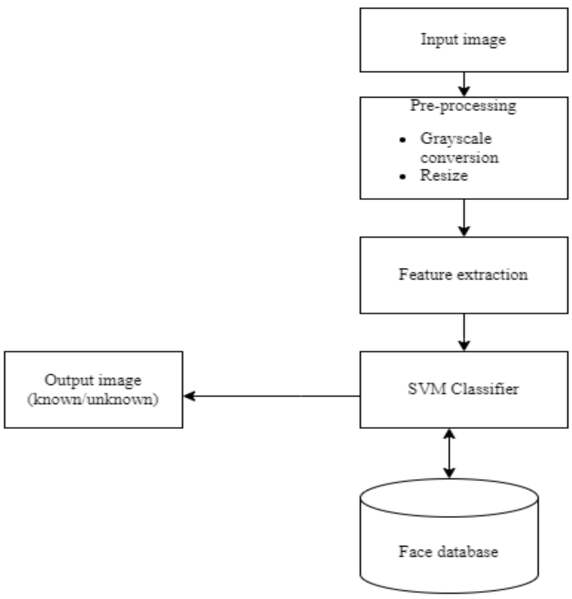

Figure 2. Block Diagram for Face Recognition System

Input images undergo preprocessing (grayscale conversion, resizing) to standardize inputs. Feature extraction isolates critical facial patterns (e.g., edges, textures) for analysis. These features are classified using a Support

Vector Machine (SVM) algorithm, which matches them against a face database to identify individuals. The system emphasizes efficiency, combining traditional image processing with machine learning for reliable recognition. This streamlined workflow transforms raw input into actionable results, prioritizing computational simplicity and accuracy in real-world applications.

3.1. Dataset Description

3.1.1. FER-2013

The FER-2013 dataset was selected over CK+ and AffectNet due to its optimal balance of size, diversity, and practicality for real-time applications. While CK+ contains only 593 posed images captured in laboratory-controlled conditions, and AffectNet offers an extensive 450,000 images that require substantial computational resources, FER-2013 provides 35,887 real-world images that represent a practical compromise. This dataset is sufficiently diverse for robust model training while remaining computationally lightweight. Additionally, its inherent class imbalance enabled us to effectively demonstrate data augmentation techniques that are crucial for real-world deployment. These characteristics make FER-2013 ideally suited for our portable, real-time system.
The data consists of 48x48 pixel grayscale images of faces. The faces have been au tomatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Dis gust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples. The public test set used for the leader board consists of 3,589 examples. The final test set consists of another 3,589 examples. The Disgust expression has the minimal number of images– 600, while other labels have nearly 5,000 samples each.

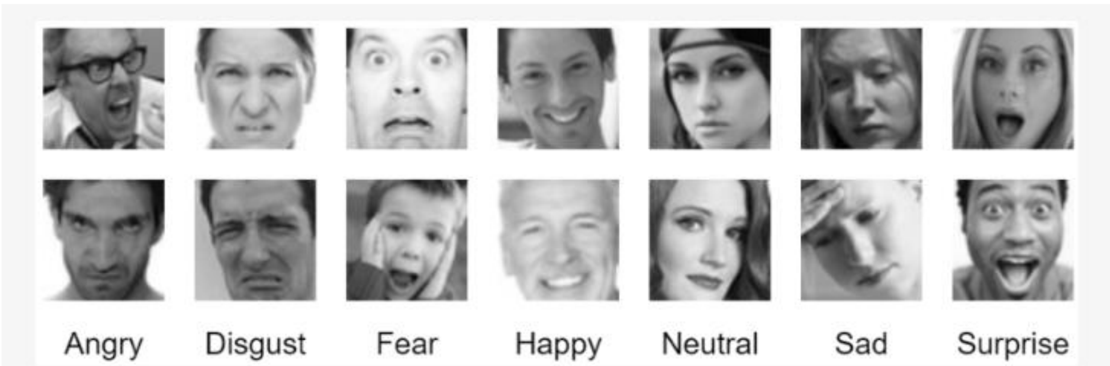

Figure 3. Selected examples of the original FER2013 dataset

3.2 Data Pre-Processing

3.2.1 Dataset Augmentation

The FER2013 dataset comprises 35,887 grayscale facial images (48 × 48 pixels), partitioned into three subsets: training (28,706 images), validation (3,585 images), and test (3,596 images). Each image depicts a single cropped face annotated with one of seven emotion classes: anger, disgust, fear, happiness, neutrality, sadness, or surprise. A notable limitation of the dataset is its significant class imbalance, with substantial variation in sample counts across emotion categories. To mitigate this imbalance and enhance model generalizability, three standard data augmentation techniques were systematically applied to the underrepresented classes.

1.Rotation

Images were rotated randomly within an angular range of -10° to +10° to simulate natural variations in head pose.

2.Horizontal Flipping

Images were mirrored along the vertical axis to preserve facial symmetry, while vertical flipping was excluded due to its incompatibility with upright facial orientation.

3.Zoom Augmentation

A random zoom factor between 1.1× and 1.2× was applied to simulate variations in facial scale and perspective.

3.2.2. Filtered FER-2013

In the filtered FER2013, the dataset was manually cleaned, in order to remove non-face images or those which were clearly not corresponding to any particular class. In the case of explicit mislabeling, the picture was subjectively relabeled to the proper category. The resulting dataset consisted of 34,140 images split into 3 sets: training—27,310, validation—3410, and test—3420.

Table 1. Number of training samples per class in particular datasets

	Original	Balanced original	Filtered	Balanced Filtered
Angry	3995	8000	3623	8000
Disgust	436	8000	442	8000
Fear	4097	8000	3711	8000
Happy	7215	8000	6816	8000
Neutral	4965	8000	4493	8000
Sad	4830	8000	5132	8000
Surprise	3171	8000	3093	8000

3.3 Algorithm Description

3.3.1. Convolutional Neural Network

A convolutional neural network (CNN) is a type of deep learning model specifically designed to process data with a grid-like structure, such as images. At its core, a CNN uses convolution operations, where small filters slide over the input data to capture local features like edges, textures, or simple shapes. These features are then combined in successive layers, often interspersed with pooling operations that reduce the spatial dimensions and help to highlight the most important information.

Convolutional neural networks layers:

- Convolution layer

  Convolutional layers form the cornerstone of convolutional neural networks (CNNs), serving as the primary mechanism for feature extraction from input data.In the early stages of the network, these filters typically identify basic, low-level features such as edges, lines, and corners. As the network deepens, the same convolutional process enables the extraction of progressively complex patterns, contributing to the overall understanding and abstraction of the data.

- Pooling Layer

  After convolutional layers extract feature maps, pooling layers are used to reduce their spatial dimensions, making the representations more compact and computationally efficient.The pooling operation works by dividing each feature map into distinct regions and then summarizing each region with a single value,commonly using either the maximum or the average value.

- Activation Function

A critical component of these networks is the activation layer, which introduces nonlinearity into the model.This step is vital because it transforms the linear outputs of the convolutional operations into nonlinear representations, allowing the network to learn and approximate intricate functions.

- Fully Connected Layer

Fully connected layers represent the final stage in many CNN architectures, where the local features extracted and aggregated by previous convolutional and pooling layers are transformed into a global feature vector. This vector encapsulates the overall content and structure of the input, and the fully connected layer uses it to calculate scores for each final category.

3.3.2. Chart Generation

The system begins by retrieving emotion data from its storage, such as CSV files, and then filters and aggregates this data based on user-defined preferences or specific time intervals. Using the processed data, it applies chart generation algorithms tailored to user requirements, transforming the raw numerical information into visual representations. Users have the flexibility to choose the most appropriate chart type—whether bar charts, line graphs,pi chart based on the nature of the data.

3.4. Model Description

3.4.1. Customized CNN

The network architecture is fine-tuned by adjusting layer configurations, activation functions, and regularization techniques to enhance classification accuracy. The model is trained on an augmented and pre-processed version of the FER-2013 dataset, ensuring improved generalization. Pre-processing steps such as grayscale conversion and normalization refine the input data, enabling the CNN to effectively extract spatial hierarchies from facial features. The architecture consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for final classification. Through iterative training and evaluation, the CNN achieves reliable emotion detection and is integrated into the system for real-world applications.

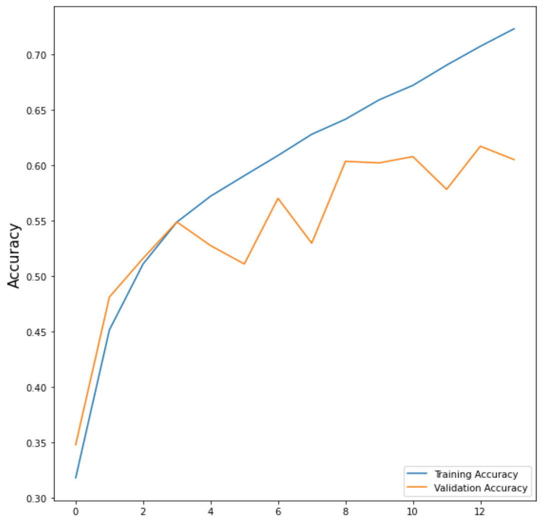

Figure 4. Training and Validation Accuracy

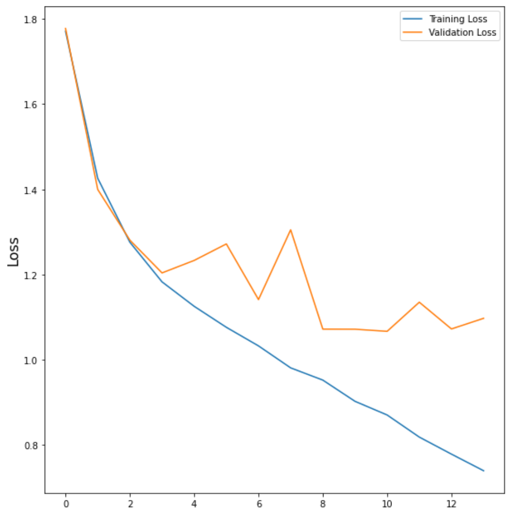

Figure 5. Training and Validation loss

3.4.2. Face Detection Using Haar Cascade

This method efficiently identifies facial structures in images or video streams by scanning input frames through multiple filtering stages. The Haar Cascade algorithm provides a lightweight and computationally efficient approach, making it suitable for real-time applications where rapid face detection is required.

3.4.3. Face Recognition Using SVM and KNN

Face recognition is performed using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), leveraging extracted facial features to identify individuals.

1. SVM for Face Recognition

   SVM constructs optimal decision boundaries to differentiate between faces. It maps input feature vectors to a higher-dimensional space, identifying an optimal hyperplane for classification. This approach ensures robust and accurate face recognition, even in varying lighting conditions and facial orientations.

2. KNN for Face Recognition

   KNN classifies detected faces based on similarity to stored feature vectors, assigning a label based on the majority vote among its nearest neighbors. This method provides an effective and interpretable approach to face recognition, balancing simplicity and accuracy.

4. Emotion Detection and Face Recognition Analysis Results

In the final phase, we developed an interactive user interface for the Human Emotion Detection and Face Recognition system, enabling seamless user interaction. The system processes live images in real-time, detecting faces and classifying emotions such as anger, disgust, fear, happiness, neutrality, sadness, or surprise. using a custom CNN model. Additionally, it automatically generates a dataset of detected faces and emotions in CSV format, providing structured data for further analysis. To enhance user engagement, the system also includes a dynamic chart generation feature, allowing users to visualize emotional trends through interactive graphs.

Figure 6. Emotion Detection and Face Recognition

5. Discussion

This study introduced a real-time Human Emotion Detection and Face Recognition system integrating computer vision and deep learning. Using a custom CNN trained on the augmented FER-2013 dataset, the system classifies facial expressions into seven categories with a test accuracy of 74.78%. Face recognition is performed using OpenCV's Haar Cascade for detection and SVM/KNN for feature matching. Pre-processing steps, including grayscale conversion, optimize data input. Additionally, the system features real-time overlays, secure authentication, and automatic dataset generation in CSV format with chart visualization. These results highlight the system's potential for applications in classroom, healthcare, and human behavior analysis.

6. Conclusion

This research successfully developed an integrated "Human Emotion Detection and Face Recognition" system, utilizing computer vision and machine learning to enable real-time emotion analysis and facial recognition. By employing a customized Convolutional Neural Network (CNN) trained on the filtered FER-

2013 dataset, along with Haar Cascade for face detection and SVM/KNN for classification, the system achieved robust accuracy in detecting and categorizing facial emotions. The system features real-time video processing, an interactive user interface, and automated dataset generation with visualization, offering a seamless user experience. Built on a Raspberry Pi, it provides a portable and scalable solution suitable for applications in sectors such as education, security, and behavioral analysis.

The system's portability could be further enhanced by transitioning to more powerful hardware or edge computing devices, enabling faster processing speeds and the ability to handle more complex tasks such as real-time audio analysis alongside facial recognition to detect emotions more holistically. By combining speech tone, pitch, and facial expressions, the system could offer a more accurate interpretation of a person's emotional state. Integrating natural language processing (NLP) techniques could allow for the recognition of emotions from text, adding another layer of interaction in chatbots or virtual assistant. The user interface could also be improved by adding features such as personalized feedback, allowing users to track their emotional states over time and receive recommendations based on their emotional trends. This would enhance the system's value in fields like mental health and personal well-being.

Acknowledgements

We would like to thank everyone who helped bring the "Human Emotion Detection and Face Recognition" project. We'd also want to thank the Department of Computer and Electronics Engineering at Kantipur Engineering College in Dhapakhel, especially Er. Sujin Gwaccha and Er. Gopal Karn, as well as their committed staff. We also value the advice supplied by our mentors and peers, as well as the dedication of our team members. We not only met our project objectives, but also expanded our learning experience. We sincerely thank everyone for their invaluable contributions and unwavering support during our journey.

References

Alrihaili, A., Alsaedi, A., Albalawi, K. and Syed, L., 2019. Music recommender system for users based on emotion detection through facial features. *IEEE*.

Babajee, P., Suddul, G., Armoogum, S. and Foogooa, R., 2020. Identifying human emotions from facial expressions with deep learning. *IEEE*.

Białek, B., Matiolański, A. and Grega, M., 2023. An efficient approach to face emotion recognition with convolutional neural networks. *Electronics*, 12(12), p.2707.

Feta, N.R., 2023. Comparison of KNN and SVM Algorithms in Facial Image Recognition Using Haar Wavelet Feature Extraction. *Jurnal Riset Informatika*, 5(3), pp.321–330. [Online] Available at: https://www.researchgate.net/publication/371876569_Comparison_of_KNN_and_SVM_Algorithms_in_Facial_Image_Recognition_Using_Haar_Wavelet_Feature_Extraction [Accessed 12 February 2025].

Hosur, V. and Desai, A., 2022. Facial emotion detection using convolutional neural networks. *IEEE*.

Huang, Z., Chiang, C., Chen, J., Wang, Y., Xie, Y. and He, H., 2023. A study on computer vision for facial emotion recognition. *Scientific Reports*, 13, p.8425.

Jaiswal, A., Raju, A.K. and Deb, S., 2020. Facial emotion detection using deep learning. *IEEE*.

Ji, L., Wu, S. and Gu, X., 2023. A facial expression recognition algorithm incorporating SVM and explainable residual neural network. *Signal, Image and Video Processing*, 17, pp.4245–4254. [Online] Available at: https://link.springer.com/article/10.1007/s11760-023-02657-1 [Accessed 27 February 2025].

Kwong, J.C.T., Garcia, F.C.C., Abu, P.A.R. and Reyes, R.S.J., 2018. Emotion recognition via facial expression: Utilization of numerous feature descriptors in different machine learning algorithms. *IEEE*.

Shahzad, H.M., Bhatti, S.M., Jaffar, A., Akram, S., Alhajlah, M. and Mahmood, A., 2023. Hybrid facial emotion recognition using CNN-based features. *Applied Sciences*, 13(9), p.5572. [Online] Available at: https://www.researchgate.net/publication/370458963_Hybrid_Facial_Emotion_Recognition_Using_CNN-Based_Features [Accessed 27 February 2025].