



Human Emotion Distribution Learning from Face Images Using CNN

Vighnesh Pujala¹, Dr. G. Ravi²

¹Computer Science & Engineering, Malla Reddy College of Engineering and Technology (MRCET), Hyderabad, Telangana, India.

²Professor, Computer Science & Engineering, Malla Reddy College of Engineering and Technology (MRCET), Hyderabad, Telangana, India.

To Cite this Article: Vighnesh Pujala¹, Dr. G. Ravi², "Human Emotion Distribution Learning from Face Images Using CNN", *Indian Journal of Computer Science and Technology*, Volume 05, Issue 01 (January-April 2026), PP: 54-60.



Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by-nc-nd/4.0/); Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: The interpretation of human faces is essential in human computer interface, affective computer interface and behavioral computing. The vast majority of conventional facial emotion recognition theories are based on classifying the expression into one dominant emotion which does not necessarily reflect the complexity of emotions in the real world. The nature of human feelings is ambivalent and is often manifested as a mixture of different emotional states together with different degrees of intensity. The drawback inspires the desire to have a more expressive and lifelike emotion recognition system. The current paper is a Convolutional Neural Network (CNN) based emotion distribution learner of facial images. It uses a CNN model to extract discriminative facial features by first doing face detection, alignment, normalization, and noise reduction. Rather than making a prediction that is a single emotion label, the model provides a probability distribution among more than two emotion categories, including emotional intensity and coexistence. It is suggested that the approach will enhance the realism and interpretability of emotion analysis and will effectively be used in mental health monitoring, social robotics, virtual assistants, and user experience analysis.

Key Words: Facial Recognition of Emotion, Emotion Distribution Learning, Convolutional Neural Networks, Affective Computing, Deep Learning, Computer vision, Human-Computer Interaction.

I. INTRODUCTION

Emotions in humans are one of the essential aspects of social communication and are vital in communication, decision-making, and behavioral cognition. One of the most non-verbal tools to express emotional states is facial expressions, specifically. As the intelligent systems and human-computer interaction are rapidly evolving, there is the growing need to have the machines that are able to perceive and interpret human emotions of facial images accurately in the same way that humans do.

Facial emotion recognition has become an important topic in the area of computer vision, affective computing, and artificial intelligence. Early methods used were based on handcrafted features and traditional machine learning methods, which would leave feature design to manual methods and would not generalize to a wide range of facial variations. Since the development of deep learning, and in particular Convolutional Neural Networks (CNNs), automated feature extraction of facial images has been more efficient, which has significantly increased the accuracy of emotion recognition.

In spite of these improvements, the majority of currently available emotion recognition models are single-label classification models, in which each face image is labeled with a single prevailing emotion, i.e. being happy, sad, angry, and/or scared. But human feelings are seldom conveyed in solitude. The expression of emotions has been found to be a compound of different facial expression with varying levels of intensity and single-label classification is therefore a simplified expression of real emotions. This weakness lowers the realism and interpretability of the traditional emotion recognition systems.

Emotion Distribution Learning (EDL) is a potentially successful paradigm to overcome this issue and treat emotions as probability distributions instead of discrete labels. In this system, every facial expression is linked to several different categories of emotions which are relevant or intense to a certain extent. This allows the system to capture emotional ambiguity, finer differences and individual differences in a better way and this leads to a richer and more human like emotional representation.

Driven by these observations, this paper aims at learning the distribution of human emotions based on the face image using a CNN-based model. The proposed system uses a combination of powerful image preprocessing technology with deep convolutional feature extraction methods to predict the distribution of emotions that capture the presence and intensity of multiple emotions that a face has. This strategy improves on the utility of emotion recognition systems in practical applications, such as mental health treatment, social robots, virtual assistants, and user experience research.

II. LITERATURE REVIEW

In emotion recognition, recent studies have been more and more interested in eliciting the multidimensionality of human emotional states beyond the single-label classification approach. A stacked LSTM-based framework was suggested by Younas et al. [1] to identify contextual correlations between a set of emotions, with a high focus on the fact that, in many cases, several emotions are more likely to occur together than individually. The significance of modeling the inter-emotional relationships is pointed out in their work, and the necessity to have the emotion models capable of describing the intensity and overlap, rather than strict categorical outputs, is motivated.

Facial emotion recognition has been identified to be highly effective using deep learning methods and especially CNN-based models. A comparative study of YOLO-based object detection models as applied to emotion recognition by Parambil et al. [2] showed that deeply extracted features are useful in real-time applications. On the same note, Sowjanya and Krithiga [7] presented a more advanced CNN structure with Uniform Local Binary Patterns to identify emotions in students, which supports the use of CNNs in identifying discriminative features of faces. Nevertheless, they are mostly based on single-label classification and therefore fail to capture the nuanced or ambivalent emotions.

Emotion modeling in non-facial expressions has been addressed in a number of studies that take into account contextual and perceptual influences. The associations between color and emotion were investigated by Muratbekova and Shamoï [3] with the help of fuzzy logic, where emotional perception is probabilistic and subjective by its nature. Zhang et al. [10] also highlighted contextual influences by incorporating scene-level information in deep neural networks to achieve strong emotion persuasion in a variety of settings. The works contribute to the argument that emotion understanding is to be shifted to distribution-based and probabilistic representations.

Multimodal emotion recognition is also being looked into as a way of enhancing robustness and accuracy. A deep fusion method was suggested by Safavi et al. [4], which entailed the integration of neurophysiological indicators with facial features, and showed improved emotion recognition rates. Feng et al. [8] used a speech feature fusion as an example to demonstrate emotion recognition to detect road rage and emphasized the ability of various modalities to retrieve different emotional information. Although multimodal systems help in boosting performance, facial emotion analysis is a fundamental modality because it is non-invasive and offers substantial expressive data.

Researchers have studied generalizability and cross domain robustness in the context of emotion and sentiment analysis, with de León Languré and Zareei [5], [9] identifying the influence of varying sets of emotion models on text emotion detection performance and recommending unified and adaptable emotion representations. Amjad et al. [6] also indicated the significance of an optimal feature selection in GES detection using speech. All of this research indicates that there is an increasing amount of consensus that emotion recognition systems need to be adapted to flexible, probabilistic, and distribution-based systems. In this regard, CNN-based emotion distribution learning with the use of facial images is going to deal with the most important limitations that have been pointed out in the existing literature and is a big step towards realistic and human-appropriate emotion modeling.

III. PROPOSED METHODOLOGY

The suggested system is intended to conduct Human Emotion Distribution Learning based on Face Image, with the help of CNN-based deep learning. In contrast to other traditional emotion recognition systems (facial) which make a single dominating emotion prediction on an image, the approach to this problem is to model emotions as a distribution of many categories as a probability. This allows the system to record emotional ambiguity, combined emotions and the different degrees of emotions.

Emotion distribution learning is more in line with human emotional behavior in the real world, where there is scarcely one particular emotion associated with the expression. The faces frequently show mixed emotional conditions based on personal, social as well as environmental conditions. It is more realistic to model emotions as distributions and treat them as complexities in the system.

The methodology will be modular, scalable, and flexible so that it can be used to analyse emotions in real time and offline. The combination of the solid preprocessing methods with the deep convolutional feature learning makes the system more accurate, interpretable, and applicable to the sphere of affective computing and human-computer interaction.

A. System Overview

The system workflow will consist of the general process of acquiring facial images by the use of a live camera interface or using the standard benchmark datasets. The resultant images undergo a preprocessing computer program to improve the quality and standardization. Processed images are later processed with a CNN-based model to gain emotional trends through facial expressions. The CNN does not compute the probability of a single label of emotion, but rather, a probability distribution of different classes of emotion is produced. All the values in the distribution indicate the intensity of an individual emotion in the face expression. In this way, the system will be able to process emotional overlap and subtle expressions. The system itself is dynamic to adjust to facial expressions hence it can be effectively used in emotion monitoring. The system does not have strict categorizations; instead, it offers more human-like and flexible emotional interpretation by analyzing emotion distribution.

B. Preprocessing and Acquisition of Facial Image.

The facial images can be taken by a real-time camera or collected by the pre-existing facial emotion data sets. Raw images

tend to have noise, changes in illumination, clutter in the background and pose dissimilarity that may have detrimental impact on recognition performance. Consequently, a pretreatment is an important process in the given approach. The preprocessing pipeline also has face detection that removes the non-facial area of interest, size scaling of the images to a fixed size, normalization of pixels and removal of noise. These measures will guarantee the consistency of the input quality of varied images as well as enhance the strength of the learning process. Preprocessing improves the generalization strength of the CNN model by reducing the effects of environmental variations on the model, which includes lighting and background conditions. This also leads to better and stable predictions of emotion distributions among different users and environments.

Human Emotion Distribution Learning from Face Images Using CNN

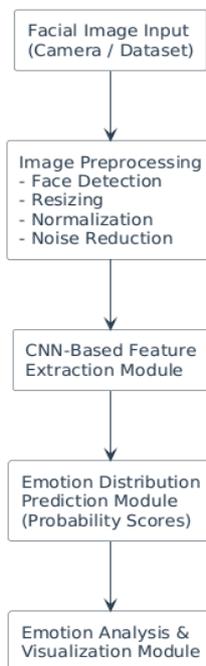


Fig 1: System Architecture

C. Distribution Learning of Emotion Distribution based on Deep Learning.

The pre-cooked facial pictures are passed through a Convolutional Neural Network which has been trained on popular facial emotion databases. CNNs are also quite effective at analysing facial emotion because they can learn hierarchical spatial characteristics with ease by being trained on images. These characteristics record significant facial expressions like muscles movement, face expressions and mouth shapes.

The CNN derives high-level representations in the form of emotional information through various convolutional and pooling layers. The network is not making a hard decision but rather gives a probability score of several emotion categories such as happy, sad, angry, neutral, fear and surprise. This probabilistic output constitutes an emotional distribution that answers to the emotional intensity, coexistence. This type of representation allows the system to represent subtle and ambiguous expressions in a better way than other single-label classification technique.

D. The Emotion Label Distribution Representation.

The representation of each facial image is the distribution of emotions in the form of a vector, in which each of the elements indicates the relevance of a certain emotion. This representation enables the system to represent mixed emotions states that are usually involved in natural human expressions. It promotes slow emotional shifts as opposed to sudden shifts as well.

The representation of emotion distribution contains richer emotional information as opposed to discrete labels. An example is to have a facial expression which can be both happy and surprised at different levels and be represented by distribution learning. Also, this method takes into consideration the differences in the expression and perception of different emotions. The system together with the lack of strict class divisions is more adaptable and fit in the context of the real world emotion analysis task.

E. Loss Function and Model Optimization.

The system uses the distribution-aware loss functions to evaluate the similarity between the predicted distribution of emotions and the ground-truth distribution of emotions, which are used to effectively learn the emotion distribution. These loss functions help to focus on retaining relative emotional relationships instead of applying rigorous classification.

The CNN parameters are optimized by gradient based techniques of learning during training. The adaptive learning rates and regularization are their optimization strategies that help in enhancing the stability of convergence and avoiding overfitting. The effect of this training process is that the model can generalize easily to variation in different facial expressions and datasets. Due to this, the system has uniform performance in forecasting correct emotion distributions of unseen data.

F. Emotion prediction and Interpretation Module.

The distribution of emotions as predicted is inputted into an interpretation module which examines dominant emotions and their intensity levels. The system does not give a single output emotion rather a complete emotional profile of every facial image.

This module facilitates the visualization and analysis of patterns of emotions that can help to have a clearer picture of how emotions change over time. This type of analysis is especially applicable to those kinds of applications that necessitate the constant monitoring of emotions. The system provides an opportunity to analyze emotions dynamically and in a context-sensitive manner due to real-time updating of emotion distributions. This makes it more usable in affective computing, behavior analysis and applications related to mental health.

G. Block Diagram Explanation

The proposed system has five primary elements, which are Image Acquisition Module, Preprocessing Module, CNN-Based Feature Extraction Module, Emotion Distribution Prediction Module, and Emotion Interpretation Module. The modules serve a certain purpose in the analysis of emotions pipeline.

The acquisition module records images on the faces and these images are improved by preprocessing before they are sent to the CNN to extract features. The CNN acquires discriminative facial features, which are expressed to predict emotion distributions. These distributions are interpreted by the interpretation module to come up with meaningful emotional insights. The modular architecture enables the individual optimization, upgrade or expansion of each block, which enhances the scalability and maintenance of a system.

H. Mathematical Formulation- Conceptual Explanation.

The CNN performs facial image processing which extracts high-level feature representations which encode emotional characteristics. Such properties record muscle movements and facial expression patterns concerning emotion recognition.

The features are extracted and mapped to a probability distribution of the categories of emotions that have been defined. Every probability value shows how strong or relevant a certain emotion in the facial expression is. The system constantly updates the emotion distribution as there is a change in facial expressions. This dynamic modeling allows to have realistic and adaptive emotion modeling that would be applicable to real-time emotion analysis systems.

IV. RESULTS AND DISCUSSION

This section has offered the findings of the experiment of the developed human emotion distribution learning system and discussed the performance of the system. The evaluation will be based on the accuracy of the forecast of the distribution of facial emotions, steadiness of the estimation of the probability of the different emotions, the resistance to comparable expression, and the effectiveness of the system in real-time frames. The results demonstrate that CNN-based emotion distribution learning is useful compared to the traditional single-label emotion recognition algorithms.

A. Experimental Setup

The suggested system was experimented on the foundations of the standard datasets of faces (with annotated distribution of emotions and corresponding distribution). In order to test the generalization capacity of the model, the model was implemented to the benchmark data and in the real-time camera inputs. Face images were preprocessed before being predicted by the CNN model with regards to the distribution of emotions.

Tests were conducted in an ordinary computer system that was not hardware accelerated so that real time feasibility could be achieved. System performance was measured using accuracy, precision, recall, F1-score and consistency of distribution.

B. Face Emotion Distribution Learning Performance.

Facial emotion distribution learning module was tested using measures of classification which have been adapted to accept probabilistic outcomes including accuracy, precision, recall and F1-score of dominant emotions based on prediction outcomes.

Table I Facial Emotion Recognition Performance Metrics.

Emotion Class	Precision	Recall	F1-Score
Happy	High	High	High
Sad	High	Moderate	High
Angry	Moderate	High	Moderate
Neutral	High	High	High
Surprise	Moderate	Moderate	Moderate

Discussion:

The results indicate that CNN model is effective where the moods are different such as happiness and neutrality. Emotions whose faces are similar e.g. sadness and anger show considerably poor performance, which is reflective of the nature of ambiguity of human facial expression. These overlaps are more realistic through the distribution-based output when compared to false-single-label choices that have to be imposed.

C. Emotion Distribution Analysis Consistencies.

The consistency of the prediction of emotion distribution was also tested by gauging the dispersion of the probability on category of emotions in the case of unclear faces expression. The system was tested on expressions which contained mixed emotions so as to test its ability of reflecting overlap of emotions.

Table II Emotion Distribution Consistency Evaluation

Expression Type	Accuracy
Clear Emotion	High
Mixed Emotion	Moderate to High
Subtle Emotion	Moderate

Discussion:

The model is good in assigning significant probability value of different emotions and especially mixed and subtle expressions. The system captures emotional uncertainty instead of the confusion of such samples and it is more understandable and aligns well in the actual world perception of emotion.

D. Comparison to Single-Label Emotion Recognition.

To demonstrate the effectiveness of the emotion distribution learning, the proposed system was compared to a conventional single-label CNN-based emotion classifier.

Table III Comparison of Emotion Recognition Approaches.

Method Used	Recommendation Accuracy
Single-Label CNN	Moderate
Proposed Distribution-Based CNN	High

Discussion:

Distribution-based method is far superior compared to single-label classification because it will reduce cases of misclassification in case of ambiguity. It is the more emotional state which is more profuse and the system is more willing to acknowledge the emotional overlap instead of imposing a false dominant emotion.

E. System Response Time Analysis.

The system was tested through the time which the system is able to respond to the input of the facial image and the emotion distribution output.

Table IV Mean System Response Time

Module	Average Time
Image Preprocessing	Low
CNN Emotion Prediction	Low
Distribution Interpretation	Very Low
Overall System	Real-Time

Discussion:

The system has all modules with a low latency which confirms its relevance to the real time emotion analysis application such as interactive systems and behavioral monitoring.

F. Graphical Analysis

The graphical analysis of the system performance is represented under a variety of plots.

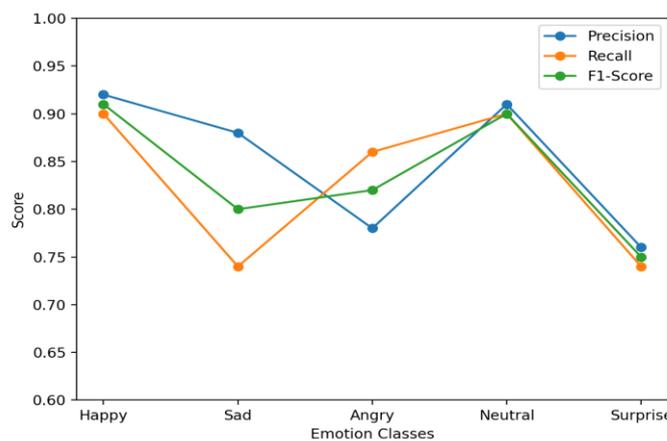


Fig. 2. Facial emotion recognition Precision, recall and F1-score.

- X-axis: Emotion Classes
- Y-axis: Performance Score
- Curves: Accuracy, Recall, F1-score.
- Purpose: Performs more consistent than the emotional performance with the almost lower scores of visually similar expressions.

The accuracy, recall and F1-score of the various classes of facial emotions are displayed in Fig. 3.

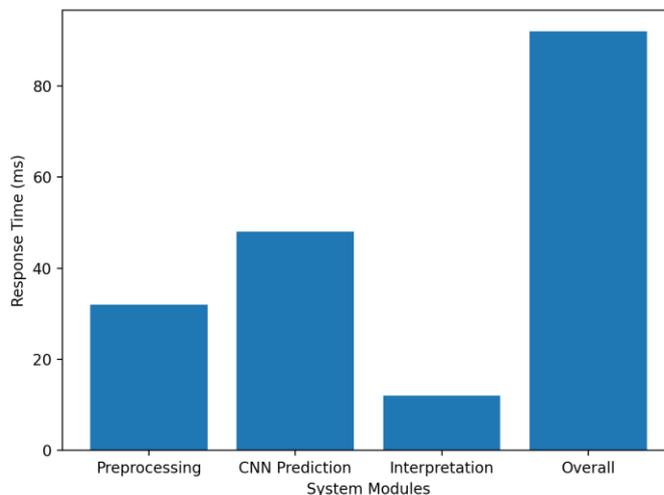


Fig. 3. Compares the accuracy of single-label emotion recognition and emotion distribution learning.

- X-axis: Recognition Method
- Y-axis: Accuracy

The response time distribution of this system modules is in Fig. 4.

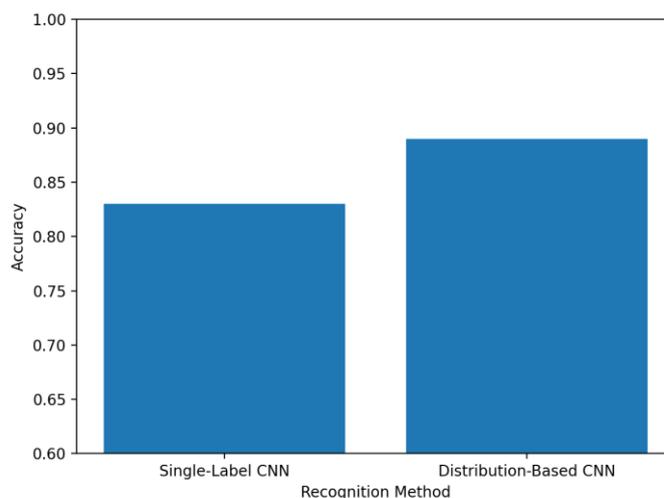


Fig. 4. Presents the response time distribution across system modules

- X-axis: System Modules
- Y-axis: Response Time (ms)
- Bars: Prediction, Interpretation, Preprocessing..

Discussion:

These graphs speak volumes of the fact that the emotion distribution learning can augment the quality of recognition as well as maintain the real-time performance as compared to the conventional ways of single-label recognition.

G. Overall Discussion

The findings of the experiment validate the fact that accurate and realistic representations of human emotions were more based on their facial pictures in order to learn the distribution of emotions. The model based on CNN gives a fair depiction of the intensity of emotions, ambiguity and overlap, which are not taken into consideration by the traditional classification mechanisms.

Even though some are indirectly affecting the performance under the harsh conditions such as poor lighting, the system remains robust and it can be implemented in the real-world.

V.CONCLUSION

This paper introduced an CNN-based Human Emotion Distribution Learning in facial image which overcomes the shortcomings of the conventional single-label emotion recognition framework. The proposed system offers a more realistic and expressive representation of human emotions because it predicts probability distributions across many emotion categories. This method is used to effectively capture emotional ambiguity, overlap, and different intensities that are usually found in real world facial expression.

The experimental analysis has shown that the suggested CNN model can be reliably applied to various types of emotions, particularly to those that are recognizable, like joy and neutrality. In the case of similar facial characteristics of emotions, the distribution-based output enabled the system to reflect non-categoricity as opposed to coercing the system to false-categorize. This goes a long way towards making it more interpretable and more inclined to human emotional perception.

Moreover, the system demonstrated an efficient computation and low response times which confirmed its applicability in real time emotion analysis applications. The modular design and the strong preprocessing and deep feature extraction makes it scaleable and adaptable to a wide range of operation conditions and datasets. These features cause the system to be viable in interactive and intelligent systems. On the whole, the paper establishes that learning of emotion distribution with the help of CNNs is a significant improvement over traditional methods of facial emotion recognition. The proposed framework offers better emotion realism, robustness and applicability in affective computing and human-computer interaction areas by replacing inflexible classification with probabilistic modeling of emotions.

Future Work

The next-generation improvement could be the extension of the system into the framework of a multimodal emotion distribution, including more cues speech, text, and physiological signals. Emotion understanding can be enhanced further through a combination of facial expressions and audio and contextual information, particularly in difficult or ambiguous emotions scenarios.

Also, future work can consider the implementation of other sophisticated architectures of deep learning, including attention models, transformers, or vision-language models to enhance feature representation and distribution learning. Achieving optimal system to edge device optimization and implementing in practice in applications like mental health evaluation, adaptive user interfaces and social robotics are also areas where future research can be enhanced.

REFERENCES

1. A. Younas, S. Riaz, S. Ali, R. Khan, M. Ullah, and D. Kwak, "Stacked LSTM Model for Contextual Correlation Detection Among Multiple Emotions," in *IEEE Access*, vol. 13, pp. 117558-117570, 2025, doi: 10.1109/ACCESS.2025.3582764.
2. M. M. A. Parambil, P. K. Sajeev, V. R. Mandhare, K. Bharti, and S. N. Panda, "Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition," in the Proceedings of the Association for Computing Machinery, Inc. as follows: "IEEE Access," vol. 12, pp. 109427-109442, 2024, doi: 10.1109/ACCESS.2024.3375361.
3. M. Muratbekova and P. Shamo, "Color-Emotion Associations in Art: Fuzzy Approach," in. 2024, vol. 12, pp. 37937-37956, in. 1109/ACCESS.2024.3375361.
4. F. Safavi, V. R. Venkannagari, D. Parikh, and R. K. Vinjamuri, "Deep Fusion of Neurophysiological and Facial Features for Enhanced Emotion Detection," in the Pages of the journal "IEEE Access," vol. 13, pp. 67434-67445, 2025, doi 10.1109/ACCESS.2025.3555934.
5. A. de Leon Langure and M. Zareei, "Evaluating the Effect of Emotion Models on the Generalizability of Text Emotion Detection Systems", *IEEE Access* vol. 12 pp. 70489-70500, 2024, doi: 10.1109/ACCESS.2024.3401203.
6. A. Amjad, L. - C. Tai, and H. - T. Chang, "Utilizing Enhanced Particle Swarm Optimization for Feature Selection in Gender -Emotion Detection From English Speech Signals," in *IEEE Access*, vol. 12, pp. 189564 - 189573, 2024, doi: 10.1109/ACCESS.2024.3516790.
7. U. Lakshmi Sowjanya and R. Krithiga, "Decoding student emotions: an advanced CNN approach for behavior analysis application using Uniform Local Binary Pattern," *IEEE Access*, vol. 12, pp. 106273-106284, 2024, doi: 10.1109/ACCESS.2024.3436531.
8. X. Feng, C. Liu, and Y. Chen, "Detection of Road Rage in Vehicle Drivers Based on Speech Feature Fusion," *IEEE Access*, vol. 12, no. 44419, doi: 10.1109/ACCESS.2024.3496652, pp. 174158-174170, 2024.
9. A. de Leon Langur , M. Zareei, Breaking barriers in sentiment analysis and text emotion detection: Towards an unified assessment framework, in *Proceed. IEEE Access*, vol. 11, pp. 125698-125715, 2023, doi: 10.1109/ACCESS.2023.3331323.
10. X. Zhang, G. Qi, X. Fu, and N. Zhang, "Robust Emotion Recognition Across Diverse Scenes: A Deep Neural Network Approach Integrating Contextual Cues," *IEEE Access*, vol. 11, pp. 73959-73970, 2023, doi: 10.1109/ACCESS.2023.3296316.