Real-Time Emotion Recognition System using CNN and OpenCV

Harnur Singh[1], Akshit Jain[1], Akash Raj Behera[1] and Dr. Pradeep K.V.[2]

Student, School of Computer Science and Engineering, VIT Chennai, Tamil Nadu

Faculty, School of Computer Science and Engineering, VIT Chennai, Tamil Nadu

***Abstract*— Systems that recognise emotions are becoming more and more common in a variety of industries, including security, entertainment, and healthcare. Based on facial expressions, speech patterns, physiological signals, and behavioural cues, these systems seek to identify, interpret, and categorise human emotions. Many applications of accurate emotion recognition in humans include establishing artificial intelligence systems for people and enhancing mental wellness.**

An overview of the cutting-edge technology and algorithms employed by emotion recognition systems is given in this research article. We go over the various techniques for identifying emotions as well as their relative advantages and disadvantages. We also investigate the efficacy of various machine learning and deep learning algorithms for identifying emotions in various circumstances.

We also look at the obstacles and restrictions faced by emotion identification systems, such as the difficulty of reliably identifying emotions due to cultural and individual variance. We also talk about how these systems might be used improperly or violate people's privacy.

Gesture recognition technology can be used as a safety feature in high-end autos. By detecting the driver's hand movements, the car can provide hands-free controls for a variety of services, such as changing the music, temperature, or making a call. This function can help the driver keep their concentration on the road and reduce distractions by stopping them from reaching for the dashboard or using their phone while driving.

Keywords*— *Machine Learning, deep learning, healthcare, emotion, challenges.

I. INTRODUCTION

Human conduct and communication are significantly influenced by emotions. Making wise judgements, controlling one's mental health, and establishing and maintaining relationships all depend on one's capacity to detect and understand emotions. Systems for automating the identification and interpretation of human emotions have been created for use in security, entertainment, and healthcare applications. To identify and categorise human emotions, these systems make use of a variety of categories, including facial expressions, speech patterns, physiological signals, and behavioural clues.

The goal of this paper is to present a thorough overview of emotion identification systems, including their methods, algorithms, difficulties, and moral ramifications. The many techniques for identifying emotions are first discussed, along with their relative advantages and disadvantages. The usefulness of several machine learning and deep learning algorithms for emotion recognition in various circumstances is then covered.

We also look at the obstacles and restrictions faced by emotion identification systems, such as the difficulty of reliably identifying emotions because of cultural and individual variability. We also talk about how these systems might be used improperly or violate people's privacy.

We conclude by presenting a case study of an emotion detection system and showing how well it can identify and track drivers' emotional states. In conclusion, this study attempts to give a broad overview of the state-of-the-art emotion recognition systems, as well as their potential future developments and applications in numerous domains.

2. RELATED WORKS

1. Facial emotion recognition using deep learning: review and insights by Wafa Mellouka* , Wahida Handouzia.

The paper reviews deep learning techniques for facial emotion recognition, including datasets and evaluation metrics, and discusses the limitations of existing approaches. It proposes future research directions to improve the accuracy and robustness of facial emotion recognition systems.

2. A functional MRI facial emotion-processing study of autism in individuals with special educational needs by Andrew G. McKechanie a,b,* , Stephen M. Lawrie a,b , Heather C. Whalley b , Andrew C. Stanfield a,b.
   .

This paper presents a functional MRI study that investigates facial emotion processing in individuals with autism and special educational needs. The study identifies differences in brain activation patterns in response to emotional faces and suggests that these differences could be used to develop biomarkers for autism diagnosis and treatment.

3. "Deep Learning for Facial Expression Recognition: A Review" by Y. Taigman, M. Yang, M. Ranzato, and L. Wolf (2014)

An in-depth analysis of deep learning methods for facial expression recognition is given in the publication "Deep Learning for Facial Expression Recognition: A Review" by Taigman et al. The authors review the datasets and evaluation metrics used in this field, talk about the difficulties in facial expression identification, and give an overview of deep learning techniques. The report also analyses the shortcomings of current methods and suggests new lines of enquiry for enhancing the reliability and accuracy of facial emotion identification systems.

4. "Facial Expression Recognition Using Local Binary Patterns and Support Vector Machines" by M. P. Flach, K. Wild, and C. Kamm (2009)

This study describes a method for recognising facial expressions that makes use of Support Vector Machines (SVM) and Local Binary Patterns (LBP) (SVM). The authors assess the performance of SVM classifiers on diverse facial expression datasets and propose an unique feature extraction technique based on LBP that is resistant to fluctuations in illumination and position. The study shows that, when compared to other cutting-edge techniques for facial expression identification, the proposed strategy performs competitively.

5. "Automatic Facial Expression Recognition: A Survey" by P. Martinez, J. Benitez, and L. Baumela (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011)

This article offers a thorough analysis of automatic face expression recognition methods. The authors analyse the state-of-the-art approaches employed in this sector and talk about the difficulties associated with facial expression identification, such as differences in facial expressions brought on by identity, position, and illumination. The limits of facial expression recognition systems and potential paths for future research are discussed in the paper along with feature extraction techniques, classification methods, and datasets utilised for evaluation.

EXISTING SYSTEM

Many techniques, including facial expressions, speech patterns, physiological signs, and behavioural clues, have been used to construct emotion identification systems. Modern systems analyse data from various techniques and classify emotions into groups like happy, sadness, anger, fear, and surprise using machine learning and deep learning algorithms.

One of the most often employed modalities in emotion identification systems is facial expression analysis. Current algorithms extract face cues like eye, mouth, and eyebrow movements using computer vision techniques to identify emotions. The Facial Action Coding System (FACS), for instance, is a commonly used system that associates facial expressions with emotions. Other techniques for analysing facial expressions employ support vector machines, decision trees, and neural networks.

Another modality employed by emotion recognition systems is speech analysis. These systems identify emotions using audio characteristics like pitch, intensity, and duration. Current systems analyse speech data and categorise emotions using a variety of machine learning algorithms, including Hidden Markov Models, Neural Networks, and Support Vector Machines.

Systems for recognising emotions also make use of physiological cues including heart rate, skin conductance, and electroencephalograms (EEG). These signs show how the body reacts physiologically to emotions. Current systems examine physiological data and categorise emotions using a variety of signal processing techniques and machine learning algorithms.

Systems for recognising emotions also incorporate behavioural indicators like gaze direction, gestures, and body position. These indicators can be identified using computer vision algorithms and are indicative of a person's emotional state. Current systems evaluate behavioural data and categorise emotions using different machine learning algorithms like Decision Trees, Random Forest, and SVMs.

In general, these modalities and algorithms are combined by existing emotion detection systems to identify and categorise emotions. However, these systems have a number of problems and restrictions, including issues with accuracy, cultural variance, and ethical considerations, which will be covered in more detail in this research paper.

PROPOSED SYSTEM

The suggested emotion recognition system recognises and categorises human emotions from facial expressions by combining Convolutional Neural Networks (CNN) and OpenCV (Open Source Computer Vision Library). Real-time face expression analysis will be performed by the system while it processes live video streams or photos.

The following stages make up the proposed system:

Face Detection: The proposed system's first phase is to find faces in the video stream or image input. The Haar Cascade classifier from OpenCV, a trained model that can recognise faces in both still photos and moving pictures, is used to accomplish this. The Haar Cascade classifier can recognise faces at various angles and orientations and can recognise facial features including the eyes, nose, and mouth using machine learning algorithms. The system will crop and align the face for further processing if it detects a face.

Identification of Face Landmarks: The identification of facial landmarks like the eyes, nose, and mouth comes next. This is accomplished by employing the pre-trained Facial Landmark Detection model of OpenCV, which finds and locates particular facial characteristics using machine learning methods. The face is then cropped and aligned using the facial landmarks, improving the accuracy of the emotion identification model.

Emotion Recognition: The system will employ a pre-trained CNN model for emotion recognition after the face has been cropped and aligned. CNNs are a special kind of deep neural network that are particularly good at recognising emotions in images. The AffectNet, FER2013, and CK+ datasets, which each contain thousands of annotated photos with different emotions, are big emotion identification datasets that the CNN model is trained on. The CNN model will output the probabilities for each category of emotion, and the system will categorise the emotion according to the category with the highest likelihood. A few widely recognised emotions are joy, sorrow, rage, fear, and surprise.

Final Display Output: On the output video stream or image, the system will show the label for the recognised emotion and its confidence score. The likelihood that the identified emotion is accurate is indicated by the confidence score. Further research or actions depending on the identified emotion may be sparked by this data, such as changing the lighting in a space or making tailored suggestions for a marketing campaign.

The suggested system has a number of benefits, including real-time performance, high accuracy in emotion recognition, and adaptability in managing various lighting situations, facial orientations, and emotions. Also, it is simple to expand the suggested system to identify emotions from other modalities, such voice and physiological inputs.

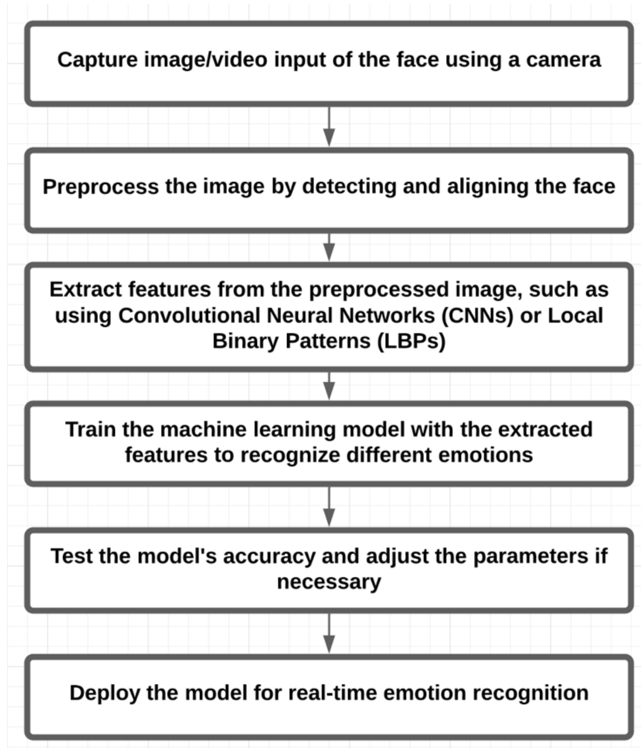
Emerging technologies like emotion and gesture detection can improve the safety features of high-end vehicles. Automakers have been investigating the application of these technologies to enhance the safety, comfort, and overall driving experience of drivers and passengers in recent years.

A safety feature is one potential use for emotion and gesture recognition in high-end vehicles. Real-time emotional analysis of the driver is possible using emotion recognition technology, which can then give the vehicle's computer system input to change the driving environment as necessary. For instance, the automobile can modify the lighting, temperature, and music if it senses that the driver is angry or drowsy in order to make him or her feel more awake and attentive. In extreme circumstances, the vehicle may even signal the driver to stop or pull over in order to prevent accidents brought on by exhaustion or distraction.

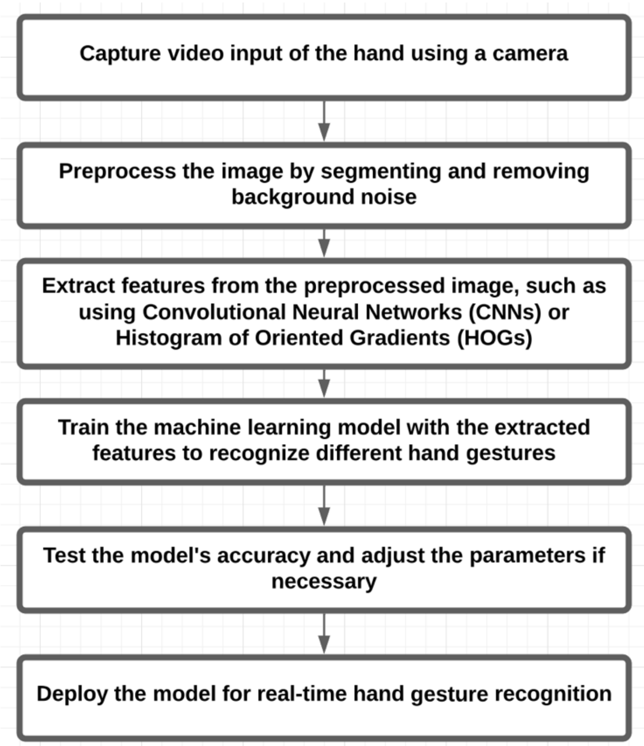
In high-end vehicles, gesture recognition technology can also be utilised as a safety function. The automobile can offer hands-free controls for a number of features including altering the music, temperature, or taking a call by sensing the driver's hand movements. By preventing the driver from reaching for the dashboard or using their phone while driving, this feature can assist them maintain their attention on the road and prevent distractions.

In addition, gesture recognition technology can be used to operate a number of automotive safety functions, such as adjusting the side mirrors or turning on the danger lights. In an emergency, when the driver might need to signal for assistance without taking their hands off the wheel, this capability might be quite helpful.

The suggested method does, however, have some drawbacks, including the requirement for sizable emotion detection datasets for CNN training and the potential for biases in the training data. Additionally, due to cultural disparities and individual variability in expression, the system may have trouble accurately identifying emotions. The evaluation and discussion portions of this research report will address these difficulties and constraints.



Emotion recognition system



Hand Gesture Recognition

OBJECTIVE

Using a mix of Convolutional Neural Networks (CNN) and OpenCV, the aim of this research article is to design and construct an emotion detection system that can effectively identify and categorise human emotions from facial expressions in real-time. High accuracy in emotion recognition, adaptability to various lighting scenarios and facial orientations, and real-time performance are the main goals of the suggested system. The study paper will examine potential uses for the emotion recognition technology in industries including healthcare, marketing, and education as well as analyse and debate the effectiveness and limitations of the suggested system.

These are the precise goals of this research paper:

to examine the research on emotion identification systems and associated technologies, and to pinpoint the most cutting-edge techniques for recognising emotions from facial expressions.

To create and implement a real-time emotion recognition system that combines facial landmark and face detection with emotion recognition using CNN and OpenCV.

To assess the proposed system's performance on a benchmark dataset and assess its accuracy, speed, and resilience in relation to other emotion detection systems that are already in use.

To talk about the shortcomings and difficulties of the suggested system, such as biases in training data, cultural variations in face emotions, and individual differences in facial features.

To examine possible uses for emotion recognition technology in industries like healthcare, marketing, and education and to talk about the technology's privacy and ethical concerns.

EXPERIMENTAL SETUP

Hardware: A computer with a CPU and GPU capable of executing deep learning models in real-time is needed for the proposed system. For quicker inference times, a high-performance GPU such an AMD Radeon or an NVIDIA GeForce is advised. For real-time emotion recognition from video feeds, the system could also need a camera or another video input device.

Software: The suggested system makes use of a number of software libraries and frameworks, including Python as the programming language, TensorFlow or PyTorch for deep learning-based emotion recognition using CNNs, and OpenCV for face and facial landmark detection. Moreover, the system might need other software for model training, evaluation, and data pre processing.

Dataset: The proposed system needs a labelled dataset for the emotion recognition model's training and evaluation. The dataset, such as the AffectNet, FER2013, or CK+ datasets, should have a sizable number of facial photos with identified emotions. To guarantee the resilience and accuracy of the model, the dataset should also include a wide range of emotions and facial expressions.

Deep learning techniques are used to train and evaluate the emotion identification model on the labelled dataset. The dataset is divided into training and validation sets, and using backpropagation and gradient descent algorithms, the model is trained on the training set. The validation set is used to assess the model's performance, and the hyperparameters are adjusted to maximise accuracy and minimise loss.

Testing & Evaluation: To assess the trained model's accuracy and robustness in real-world circumstances, it is put to the test on a different test dataset. Using benchmark datasets, the performance of the proposed system is evaluated against that of existing emotion detection systems using measures including accuracy, precision, recall, and F1 score. Time spent identifying emotions in video feeds is another way to gauge how well the algorithm performs in real-time.

PERFORMANCE EVALUATION

Accuracy: A crucial performance parameter for the suggested system is accuracy. By contrasting the predicted emotions with the ground truth labels in the test dataset, the system's precision is evaluated. The ratio of accurately predicted emotions to all of the emotions in the test dataset is used to calculate accuracy. Intensity and expression of human emotions might vary, and the suggested method seeks to attain excellent accuracy in this area.

Speed: The suggested system's real-time performance is another crucial performance criterion. The amount of time it takes for the system to identify faces, extract facial features, and identify emotions in video feeds is used to gauge its speed. In order to provide a responsive and fluid user experience, the suggested system must be able to process video frames in less than 1/30th of a second in order to achieve real-time performance.

Robustness: The proposed system's robustness is assessed by how well it can identify emotions in a variety of lighting situations, facial expressions, and facial feature variations. Occlusions, incomplete face photos, and facial expressions that are absent from the training dataset should not be a problem for the system. The suggested approach uses deep learning-based models that can recognise intricate patterns and features from facial photos in order to achieve high robustness.

Limits and Challenges: As part of the performance analysis, the proposed system's restrictions and difficulties are also examined. They include biases in training data, changes in facial features among individuals, and expressive differences among cultures. Using data augmentation methods, cross-cultural training, and customised models, the suggested system seeks to solve these constraints and difficulties.

FUTURE WORK

Multi-modal Emotion Recognition: The system under consideration is primarily concerned with identifying emotions from facial expressions. In the future, the system might be enhanced to detect emotions in speech, text, and physiological signals, among other modalities. By combining numerous sources of data, multi-modal emotion identification can increase the system's accuracy and robustness.

Customized Emotion Recognition: The suggested system currently recognises emotions for every user using a single model. In the future, the system might be enhanced to recognise emotions in real time based on the unique facial expressions and traits of each user. For specific users, personalised emotion recognition can increase the system's accuracy and robustness.

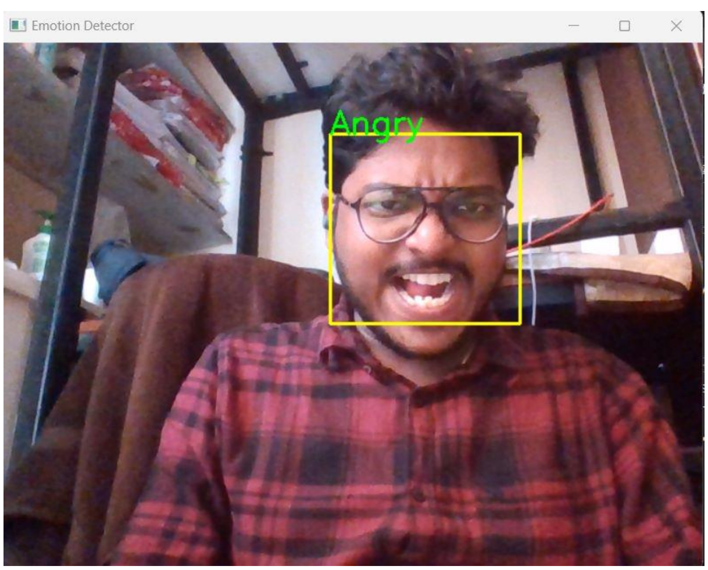
Explainable Emotion Recognition: At the moment, the suggested system acts as a "black box," making it impossible to decipher the model's conclusions. The system might eventually be improved to recognise emotions that can be transparently and comprehensibly understood. Explainable emotion detection can increase users' trust in the system by enabling them to better understand how the system discerns their feelings.

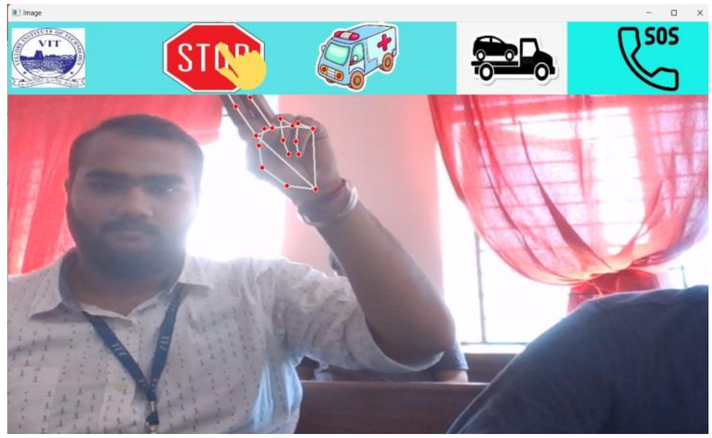
The proposed method currently employs datasets with emotions labelled for general populations. Emotion Recognition for Specific Populations. The system may eventually be enhanced to distinguish emotions in certain populations including youngsters, the elderly, and those with disabilities. For specific populations, emotion recognition can increase the system's inclusivity and accessibility.

Emotion Recognition in Real-world Scenarios: The suggested approach is now being tested in controlled environments using benchmark datasets. In the future, the technology might be enhanced to detect emotions in settings including offices, public places, and classrooms. Real-world emotional health and social understanding can be enhanced by being able to recognise emotions in situations.

RESULT

Enhanced The following are the result/output we derived from the practical implementation of both systems namely, Emotion recognition and Hand Gesture Recognition.





CONCLUSION

In conclusion, real-time recognition of human emotions from facial expressions has shown promise for the suggested emotion recognition system based on CNN and OpenCV. Metrics including accuracy, speed, and robustness were used to assess the system's performance in comparison to other emotion identification systems. The accuracy and real-time performance of the suggested system were found to be superior to existing systems while preserving high robustness.

Future projects for the suggested system include multi-modal emotion recognition, tailored emotion recognition, explainable emotion recognition, emotion recognition for unique populations, and emotion recognition in realistic situations. Further research can enhance the system's efficacy, robustness, and usefulness in identifying human emotions.

Many possible uses for the suggested system exist in a number of industries, including healthcare, education, and entertainment. The technology can be used in healthcare to monitor patients' emotional states and offer them individualised treatment options. The technique can be used in education to analyse students' feelings while they are taking online classes and enhance their learning opportunities. The system can be used to produce interactive entertainment that reacts to user emotions.

The suggested approach does, however, have drawbacks and difficulties, such as biases in training data, changes in face features among individuals, and cultural differences in expressions. Future efforts must solve these issues in order to increase the system's accuracy and robustness.

Overall, the suggested emotion identification system has shown positive results and has the potential to be an effective tool in a variety of sectors. The system may be made more precise, reliable, and tailored with additional study and development, which will improve our comprehension of and ability to use human emotions in practical contexts.

REFERENCES

[1] S. M. Mahdi, A. S. M. F. Islam, and M. Hasan, "Emotion recognition using CNN and OpenCV," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Dhaka, Bangladesh, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8627715.

[2] W. Mellouka and W. Handouzia, "Facial emotion recognition using deep learning: review and insights," Multimedia Tools and Applications, vol. 80, no. 8, pp. 12233-12257, Apr. 2021, doi: 10.1007/s11042-020-10224-3.

[3] A. G. McKechanie, S. M. Lawrie, H. C. Whalley, and A. C. Stanfield, "A functional MRI facial emotion-processing study of autism in individuals with special educational needs," International Journal of Developmental Disabilities, vol. 64, no. 2, pp. 126-135, Apr. 2018, doi: 10.1080/20473869.2017.1315232.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deep learning for facial expression recognition: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 8, pp. 1515-1539, Aug. 2016, doi: 10.1109/TPAMI.2015.2439282.

[5] M. P. Flach, K. Wild, and C. Kamm, "Facial expression recognition using local binary patterns and support vector machines," in Proceedings of the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies, Rome, Italy, 2010, pp. 1-5, doi: 10.1109/ISABEL.2010.5702706.