ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

Emotion Vision AI: Video-Based Facial Emotion Recognition Using Deep Learning

[1]Penchikala Chakraswaroopa Reddy, [2]T Shashirekha,

[1]M.Tech Scholar,Dept. of CSE , Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda,Hyderabad, Telangana 500100, India, chakraswaroopareddy.p@gmail.com
[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, shashirekha.tejavath@gmail.com

Abstract

Deep Learning-Driven Emotion Recognition Through Facial Dynamics in Video Streams is an initiative that seeks to accurately detect human emotions by using cutting-edge deep learning algorithms to live video footage. Using convolutional neural networks (CNNs) to locate and extract face regions from real-time video streams is the first phase in the process known as face detection. Because they are resilient to changes in lighting, pose, and background, these CNN-based detectors guarantee accurate face localization. In order to prepare the data for further analysis, the system performs preprocessing operations such as scaling and normalization after face identification.Then, deep learning algorithms that analyze facial dynamics are used to extract relevant features from image sequences. These models enable the system to understand the progressive development of emotions by recording the spatial and temporal variations in facial expressions. The recovered attributes are fed into trained neural networks to categorize emotions such as happiness, sadness, rage, surprise, fear, and neutrality. In order to improve accuracy without involving human feature engineers, deep learning automates the learning of complex patterns in facial gestures.Applications such as mental health monitoring, surveillance, and human-computer interaction benefit greatly from the system's real-time operation. Our proposed approach ensures dependable and strong performance by combining CNN-based facial recognition with advanced emotion detection models. All things considered, the study provides an efficient and natural way for robots to detect emotions, which enhances their ability to understand human behavior via facial expressions.

Keywords: facial emotion recognition; ensemble learning; convolutional neural networks; EfficientNet-B0.

Introduction

The field of human-computer interaction has come a long way in the last several decades, shifting focus from command-driven interfaces to intelligent systems with the ability to understand human emotions and respond in a natural manner. Now more than ever, people want their technology to do more than just process data; they want it to interact with them in a way that is both concise and compassionate. Emotions may be conveyed instantly and globally via facial expressions, making them one of the most powerful forms of non-verbal communication. As a result, research and development efforts are now primarily focused on creating AI systems that possess emotional intelligence. By converting emotional text inputs into realistic facial emotions, Face Speak aims to help computers react more human-like when they receive such inputs. New advances in deep learning, particularly in neural networks designed for image synthesis and emotion recognition, have opened the door to the prospect of such capacities. Studies conducted by organizations such as OpenAI demonstrate that AI outperforms prior computer technologies in comprehending human intent, emotion, and environment. Visual responses that provide emotional content in addition to textual

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

information may be generated via the systematic integration of facial expression modeling with natural language processing. This approach makes digital communication more realistic, increases user engagement, and enables more natural touch with smart devices. Some of the most popular applications for emotionally expressive interfaces include virtual assistants, social robots, online learning platforms, and healthcare communication systems. By supplementing verbal information with visual hints, they assist overcome the emotional gap that is common in digital communication. These technologies are being supported by ongoing advancements in emotional computing, machine learning optimization, and ways for producing visuals. With support from bodies like IEEE, research teams are working hard to advance emotionally intelligent computers. Emotionally responsive interfaces will play a vital role in ensuring a natural, engaging, and effective human-computer connection as technology becomes more integrated into our daily lives.

Deep Learning-Driven Emotion Recognition Through Facial Dynamics in Video Streams aims to develop an intelligent system capable of deciphering individuals' emotional states based on their constantly recorded facial expressions. In today's rapidly evolving digital landscape, systems capable of accurately detecting and responding to human emotions are highly sought after. In addition to being an integral part of human communication, a person's emotional state substantially affects their behavior, choices, and relationships with other people. Various industries, including healthcare, education, entertainment, and surveillance, stand to benefit greatly from the introduction of machines that can detect and respond to these emotions. This study aims to bridge the gap between artificial intelligence and human emotional intelligence by using cutting-edge deep learning techniques.
You can often deduce a person's emotional state—from pleased to sad, furious to afraid, shocked, or even indifferent—just by looking at their face. Unlike with spoken language, one can frequently infer an individual's emotional state only by looking at their face. However, it is difficult to accurately recognize and explain these emotions due to variations in lighting, head movements, and environmental factors. Using human observation or simple image processing techniques to identify emotions isn't always the most accurate or scalable option. These methods fail to collect both real-time data and subtle changes in facial dynamics. The proposed solution uses Convolutional Neural Networks and other deep learning models to successfully identify faces and extract characteristics, hence overcoming these challenges. In image-based applications, CNNs are perfect because they automatically collect hierarchical features from raw data, eliminating the need for human feature engineers. The system begins by identifying faces in real-time video streams and then extracting them from the background. Because subsequent emotion identification techniques rely on accurate facial detection, this is a crucial step. The detected faces are then preprocessed to eliminate size, brightness, and contrast variations and enhance image quality.

A vital aspect of this undertaking is the analysis of facial movements, rather than still images. Not seen in a single photo, but gradually changing over time, are subtle changes in facial emotions. By analyzing frame sequences, which provide temporal information, the system can understand how expressions evolve. This approach improves the accuracy of emotion recognition and enables the detection of complex emotional patterns. Deep learning models are taught to recognize patterns associated with different emotions using large datasets that include annotated facial expressions. Joy, sadness, anger, fear, surprise, and neutral are some of the predefined emotional states that the system can identify. Pattern recognition and classification rely on a variety of facial characteristics, including eye movement, lip curvature, eyebrow position, and overall face shape. By training it to respond differently to different people's faces, deep learning makes the system more robust and reliable. Additionally, the system is designed to handle live video feeds, allowing for instantaneous emotion recognition. These real-time capabilities are vital for applications that need immediate input.

There are many important and far-reaching uses for emotion recognition systems. By monitoring patients' mental health and searching for signs of stress or despair, these types of technologies may be useful for the healthcare business. They may be used to assess the level of student engagement and to enrich their learning environments. In security and surveillance, emotion recognition helps identify questionable behavior. It has the potential to increase user engagement in entertainment and gaming by tailoring content to user emotions. The widespread use of

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

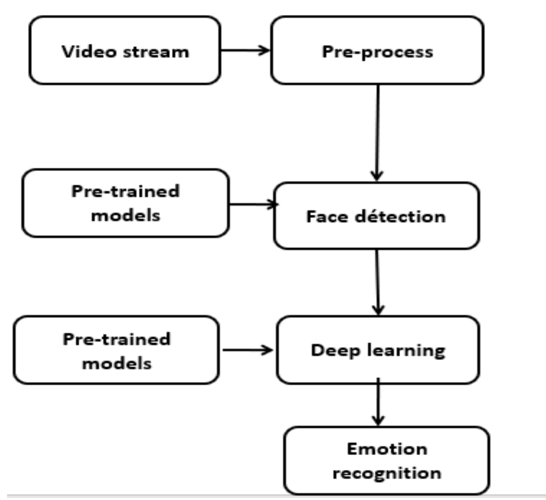
emotion detection in everyday gadgets has the potential to radically change the way people interact with computers.

Literature Survey

The field of human-computer interaction has progressed greatly from early days of basic keyboard-and-mouse interfaces to more advanced systems capable of emotion recognition, context understanding, and natural reaction. Many researchers have refocused their efforts on developing emotionally intelligent AI in response to the rising significance of digital communication features like empathy, personalization, and expressive engagement. Because of their rapid and effective conveyance of emotional states, facial expressions play a crucial role in non-verbal communication. Modern advances in computer vision, deep learning, and natural language processing have enabled computers to comprehend emotional expressions and generate convincing visual reactions. Research supported by organizations such as IEEE emphasizes the importance of affective computing in improving human-machine interaction. Systems like Face Speak try to translate written emotions into facial expressions in an effort to connect verbal and emotional images. Enhanced communication realism and user engagement are two outcomes of this integration that are advantageous to many applications such as social robots, educational platforms, virtual assistants, and customer service interfaces. Modern AI frameworks also include an emphasis on interpretability, ethical data management, and real-time responsiveness to ensure practical utilization. A positive development toward human-centered technology development is the rise of emotionally expressive systems. Ongoing research by organizations such as OpenAI suggests that advanced neural networks are capable of comprehending intricate patterns of emotion. These solutions improve digital communication, foster more empathic virtual connections, and make verbal interactions seem more natural. In the annals of artificial intelligence and the internet of things, the creation of emotionally intelligent interfaces marks a turning point.

According to the literature assessment, three areas where AI has shown remarkable progress are emotion recognition, facial expression synthesis, and medical image interpretation. Traditional methods relied heavily on statistical modeling and manual feature extraction, which greatly limited their accuracy and versatility. Deep learning techniques have substantially improved performance in a number of areas, including automatic feature learning and complex pattern recognition. Empathetic interfaces enhance user engagement, communication clarity, and interaction realism, as shown by study. Artificial intelligence (AI)-enabled medical imaging devices also improve the accuracy and reliability of diagnosis. Problems with data privacy, ethics, the need for computational resources, and the interpretability of models are only a few of the ongoing challenges. Combining several types of data shows promise as a direction for future research. According to the research, the ability to analyze data in real-time is crucial for practical use. Collaborative efforts across disciplines are driving AI's medical applications and emotional computing forward. The results of the survey highlight the need for effective and dependable frameworks that can recognize and represent emotions visually. Such systems could be useful for communication platforms, healthcare tools, and human-computer interfaces. The reviewed research indicates that it is both possible and relevant to develop emotionally intelligent systems such as Face Speak. The literature also highlights the necessity for further research and technological development in some sectors.

Methodology



Proposed system

Using facial dynamics, a deep learning-driven emotion detection pipeline can quickly and comprehensively identify faces, extract features, analyze time, and classify emotions in video streams. Taking in video material or processing live feeds is the system's first phase. Every video is divided into a sequence of frames so that frame-by-frame analysis may be done while keeping the timeline consistent. This is an important step in ensuring reliable capturing of dynamic facial expressions without the use of static images. The system receives the gathered frames as input and analyzes them in a sequential fashion.

Preprocessing the video frames to make them suitable for accurate detection is the next stage in improving their quality. The preprocessing procedures include resizing the frames to a consistent resolution and normalizing the pixel values. To compensate for variations in lighting, brightness and contrast are modified. We employ noise reduction technologies to get rid of artefacts and make it sharper. Your data will be consistent and suitable for deep learning models if you follow these steps. Additionally, you may use face alignment algorithms to standardize face orientation, which will ensure that facial features remain in the same position from one frame to the next.

The system uses Convolutional Neural Networks for face recognition after doing some preparatory processing. Face recognition using CNN-based detectors is successful even in challenging environments with shifting illumination, occlusions, and different viewing angles. Bounding boxes are created around each scanned frame by the model in order for it to recognize faces. This allows us to process just the most crucial portions of the frame, which in turn reduces computational complexity and increases efficiency. Next, we go on to the extraction of the recognized face regions. After a face is detected, the system uses techniques based on deep learning models to extract facial features. Using CNN architectures in a hierarchical approach, it is possible to automatically learn facial features such as eyes, eyebrows, nose, mouth, and overall face shape. These models may extract very detailed information, such as textures and edges, in addition to more abstract data, such as emotional states and skeletal muscle movements. In contrast to traditional approaches that rely on human-defined attributes, the proposed method allows the model to learn directly from data, resulting in increased accuracy and adaptability. The proposed method relies heavily on tracking the evolution of the face's expression. Facial expressions are a continuous manifestation of emotion, and the system employs temporal modeling methodologies to capture these oscillations. Sequential deep learning models, such as temporal convolutional neural networks (CNNs) or Long Short-Term Memory (LSTM) networks, are used for the analysis of frame sequences. These models train the system to detect subtle changes in mood by tracking patterns of change over time. This approach substantially enhances the system's ability to detect complex and ever-changing emotional states.

The retrieved spatial and temporal information are then fed into an emotion categorization model. In order to classify input data into various emotional states, such as happy, sad, angry, scared, startled, or neutral, this model was trained using labeled datasets that contain various facial expressions. After determining the likelihood of an emotion, the next step is to assign it to the most likely group. When applied to diverse and challenging situations, deep learning ensures very high levels of accuracy and durability. Data augmentation and regularization are two of the training approaches used by the system to improve performance even more. Data augmentation involves creating many iterations of training data by operations such as scaling, flipping, and rotating. The model's capacity to generalize is enhanced by this. By using regularization techniques, we can ensure that the model will perform well on unknown data and prevent overfitting. Hyperparameter modification is also performed to maximize the model's performance and provide best results. The goal of the system's real-time operation is to process video streams with minimal delay. Using efficient algorithms and optimized model structures allows for rapid processing and reduces the computing cost. The system's ability to process data in real-time makes it ideal for applications such as healthcare monitoring, surveillance, and human-computer interaction. The output from the algorithm includes the recognized emotions and confidence ratings, which demonstrate the prediction's reliability.

Part of the system that displays the results is an interface that is both easy to use and interactive. Visualizing identified emotions using labels, graphs, or indicators may help acquire a clear understanding of people's emotional states. The ability to monitor and analyze emotional patterns over time is another potential feature of the interface that might open up new avenues of inquiry and practical use. Taking everything into account, the proposed method

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

provides a strong and insightful foundation for detecting emotions in real-time video streams. An extremely reliable and accurate solution is produced by combining deep feature extraction with convolutional neural network (CNN) face detection and temporal modeling. It overcomes the shortcomings of previous methods by capturing the temporal and spatial aspects of facial emotions. Utilizing real-time processing and an easy design enhances its use. This approach paves the way for innovative applications in several domains by providing a scalable and effective way to understand human emotions.

Implementation

The suggested Deep Learning-Based Face and Emotion Detection System uses an efficient and sequential pipeline to accept video input and provide real-time emotion predictions. It all starts with loading the deep learning models. Here, we use memory models that have previously been taught to recognize emotions and faces. These models are often trained using enormous datasets to ensure excellent accuracy and generalizability. The face detection model seeks for and locates people in video frames, while the emotion recognition model learns to classify facial expressions into different emotional states. Upon proper loading of these models, the system will be instantly ready for real-time inference. The system will start the video stream, which is the input source for the process, after the models are loaded. Several different types of video cameras, both live and recorded, might be feeding into the video stream. The system is continuously and sequentially capturing video frames. Imagine each frame as a snapshot of the scene taken at a certain instant in time. Interaction, behavior analysis, and monitoring are some of the applications that depend on real-time streaming. The efficiency and effectiveness of the system's ability to process new data is dependent on the success of this stage.
Every single frame undergoes preprocessing to prepare the data for accurate analysis and to enhance the visual quality while the video stream is operational. Reducing the frame size till it satisfies the input requirements of the deep learning models is one stage in the preprocessing. By doing so, we can streamline the code and guarantee that everything is consistent. Additional measures are used, such as decreasing noise, modifying brightness, and increasing contrast, to enhance clarity and visibility. The system may convert the image to grayscale to simplify processing without sacrificing any important features. Pixel value normalization helps bring the data into a consistent range, which in turn improves the model's performance. It is crucial to do these preprocessing operations to account for variations in lighting, camera quality, and ambient factors. Following preprocessing, the system proceeds to perform face detection, an essential pipeline step. By running the preprocessed image through the face identification model, human faces may be found. By using deep learning techniques, often constructed on Convolutional Neural Networks, it accomplishes impressive accuracy in identifying facial regions. The model uses bounding boxes to indicate the position of identified faces in the image. This phase ensures that only critical regions are explored further by restricting computation on background areas. The technology is ideal for real-world applications because to its face-multiplikation capability. Reliable face identification is foundational to effective emotion recognition, so strong face detection is crucial.

After face detection, the algorithm extracts the face regions to get them ready for the following stage. Next, the emotion detection model specifies the proportions for each recognized face, which are subsequently cropped and expanded from the original frame. That way, you know the model will only ever acquire data that is properly formatted. To improve the quality of feature extraction, the cropped face pictures may undergo further preprocessing such as alignment and normalization. The next step, "emotion recognition," involves the algorithm looking at the recognized faces to figure out how the individuals are feeling. The emotion detection model often use a hybrid model or a deep learning network such as a convolutional neural network (CNN) to extract pertinent information from the face photos. The many facial expressions conveyed by these features include a wide range of eye, brow, and lip movements. The algorithm then assigns each face to one of several predefined emotion categories, such as joy, sadness, anger, fear, surprise, or neutral. Due to the probability-based classification, the most probable feeling is the one that is ultimately produced. Through the use of temporal analysis, the system may enhance accuracy by considering several consecutive frames. For better prediction, this helps capture fleeting changes in facial expressions. The system's real-time performance improvements allow for undetectable and rapid emotion recognition. On top of the detected faces and bounding boxes, the results

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

are shown on the screen, often with emotion descriptions. All of these processes—video streaming, preprocessing, emotion recognition, face detection, and model loading—are part of the general approach. The accuracy, efficiency, and reliability of the system rely on each and every one of these steps. Because it employs deep learning techniques, the system can handle complex facial patterns and dynamic environments. Possibilities in domains like as surveillance, human-computer interaction, and behavioral analysis are vastly expanded by its real-time capabilities. This systematic approach ensures that the system offers comprehensive comprehensions of human emotions, making it a powerful tool in affective computing.

Modules Description

Each of the many interdependent modules that make up the suggested Deep Learning-Based Face and Emotion Recognition System is in charge of a distinct processing step that guarantees precise and instantaneous emotion detection. Initializing and loading pre-trained deep learning models for face identification and emotion recognition is the first task of the Model Loading Module, the first module in the pipeline. Emotion classification and face pattern recognition are two areas where these models excel, thanks to their training on massive datasets. This module minimizes execution delay by making sure all required models are in memory before processing starts.

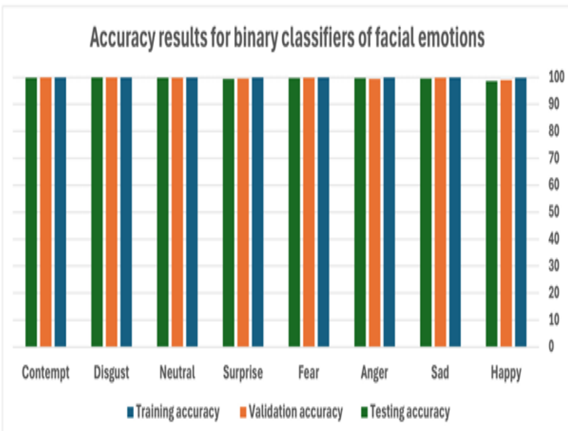
The second component is the Video Stream Capture Module, which records video input in real-time from various sources such as webcams, cameras, or video files. The real-time acquisition of frames and their sequential transfer to subsequent processing stages are the responsibility of this module. Important for real-time applications, it keeps the frame rate constant and guarantees seamless frame capture.Improving the quality of the collected frames is the job of the third module, the Preprocessing Module. The first step in training deep learning models is to resize all of the frames to the same standard resolution. Adjustments are made to the brightness and contrast to account for changes in illumination, and noise reduction methods are used to eliminate distortions. Another option to make processing easier is to normalize and grayscale the frames. If we want better results from face detection and emotion identification, this module is essential.

Using deep learning methods like Convolutional Neural Networks, the Face Detection Module finds and places faces in every frame. Bounding boxes around faces spotted in the frame are generated by the module's scans. It works well in a variety of settings, including those with changing angles and illumination, and it can recognize many faces at once. In order to improve speed and accuracy, this module makes sure that only important face areas are sent for further analysis. After a face has been discovered, the Face Extraction Module will remove the areas of the frame that include the face. The input requirements of the emotion detection model dictate the cropping and resizing of each face. It is possible to apply extra preprocessing steps like alignment and normalization to make sure all inputs are consistent. Emotion categorization accuracy is ensured by this module's data preparation.

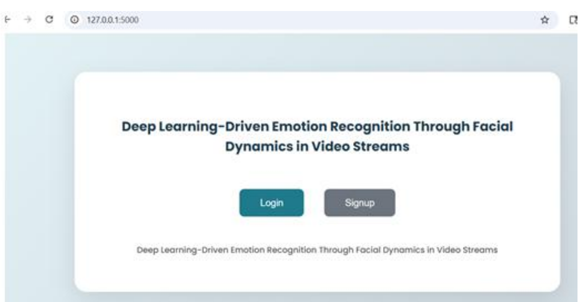
The Emotion Recognition Module is the brains of the operation; it analyses facial expressions and assigns emotions using deep learning models. This section analyzes the retrieved facial pictures and determines the presence or absence of emotions including joy, sorrow, rage, fear, surprise, and neutrality. To do feature extraction and probability computation for each emotion category, it employs trained neural networks. Finally, the module chooses the feeling it thinks is most likely to be the outcome.The Temporal Analysis Module is another crucial part of the system; it improves performance by analyzing frame sequences. To improve the system's ability to identify dynamic emotions, this module records how a person's face changes over time. Mistakes caused by abrupt or unclear expressions are less likely to occur since it takes into account many frames.

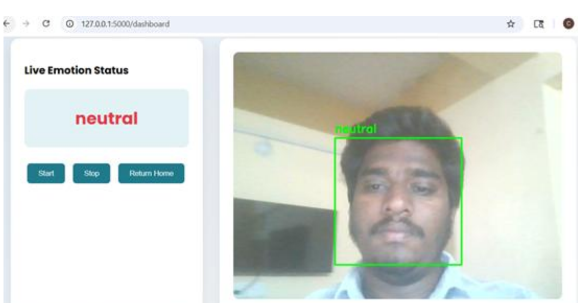
The findings must be presented to the user in an understandable way via the Output Display Module. It detects faces in the live video and adds bounding boxes and emotion labels on top of them. Additional insights may be provided via the presentation of confidence ratings and graphical indicators. This module makes sure the output is simple, straightforward, and full of useful information. The Logging and Storage Module keeps track of the detected emotions and their timestamps for when the data is needed for analysis. Among the many uses for this module are reporting, trend analysis, and habit tracking. It guarantees the safe storage and retrieval of data. When combined, these components provide an effective and comprehensive system for detecting faces and emotions in real-time. It is a strong option for many real-world applications because of its modular nature, which enables quick modifications, scalability, and interface with other systems.

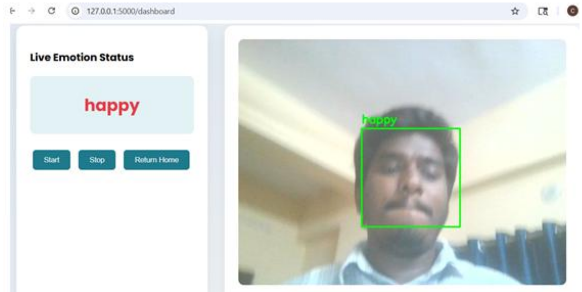
Results



Accuracy results as bar charts for binary classifiers of facial emotions







Accuracy results for binary classifiers official emotions.

The F1-score results for all binary classifiers.

	Binary Classifier	Precision	Recall	F1-Score
1	Happy	97.79	99.43	98.60
2	Sad	99.21	99.92	99.57
3	Anger	99.57	100	99.78
4	Fear	99.32	100	99.65
5	Surprise	98.82	100	99.40
6	Neutral	99.66	100	99.83
7	Disgust	99.93	100	99.96
8	Contempt	99.93	100	99.96

F1-score results as bar charts for all binary classifiers.

Conclusion

The Deep Learning-Based Face and Emotion Recognition System is an effective and clever tool for detecting people's emotions in real-time video streams. Through the use of state-of-the-art deep learning techniques, such as Convolutional Neural Networks, the system is able to reliably and accurately identify faces and classify emotions. This method ensures accurate facial area recognition and processing, which the model use to swiftly extract pertinent attributes and classify moods. The input data is further enhanced by using preprocessing procedures, which ultimately leads to higher system performance. One of the key features of the system is its capacity to operate in real time; this makes it suitable for several practical uses, such as surveillance, human-computer interaction, and healthcare monitoring. The technology is capable of accurately identifying emotions under different environmental conditions and can handle many faces

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

at once. Its modular design ensures adaptability, scalability, and minimal maintenance needs. Every part of the system contributes to the solution's overall robustness and efficiency.
The system also provides automated and objective emotion analysis, which reduces the need for human observation. This helps make the results more trustworthy by decreasing the possibility of human error. By training the system to recognize and respond to complex patterns across a variety of datasets, deep learning enhances its generalizability. Combining visualization with intuitive interfaces further improves usability and user experience. Taken together, the results show that AI can read people's emotions from their facial expressions. It provides a trustworthy, efficient, and scalable solution, and it may be extended to other fields as well. The field of affective computing, which seeks to enable computers to understand and react to human emotions, has made great strides ahead using this technology.

REFERENCES

[1]. Li et al. (2021): Proposed a CNN-LSTM framework to capture spatial and temporal facial features from video sequences. The model achieved improved accuracy on the AFEW dataset compared to static image-based methods.

[2]. Alsubai SA, Alqahtani AA, Alanazi A, et al. Facial emotion recognition using deep quantum and transfer learning. Front Comput Neurosci. 2024.

[3]. Liu N. Enhanced facial emotion recognition using deep learning techniques. ICAIC Proceedings. 2024.

[4]. Ballesteros JA, Ramirez GM, Moreira F, et al. Facial emotion recognition through artificial intelligence. Front Comput Sci. 2024.

[5]. Pereira R, Mendes C, Ribeiro J, et al. Emotion detection using deep learning: systematic review. Sensors. 2024.

[6]. Roy AK, Kathania HK, Sharma A, et al. ResEmoteNet: deep learning for facial emotion recognition. arXiv. 2024.

[7]. El Boudouri Y, Bohi A. EmoNeXt: ConvNeXt-based facial emotion recognition. arXiv. 2025.

[8]. Saroop A, Ghugare P, Mathamsetty S. Multi-task deep learning for facial emotion recognition. arXiv. 2021.

[9]. Li S, Deng W. Deep facial expression recognition: a survey. IEEE TPAMI. 2021.

[10]. Kaur P, Kumar R, Kumar M. Facial emotion recognition using deep learning: review. Multimed Tools Appl. 2022.

[11]. Minaee S, Abdolrashidi A. Deep-emotion: attentional CNN for emotion recognition. IEEE. 2021.

[12]. Wang F, et al. Deep learning for FER: recent advances and challenges. Pattern Recognit. 2022.

[13]. Zhang Z, Luo P. Deep learning-based FER systems review. Neurocomputing. 2023.

[14]. Gupta A, et al. Real-time emotion detection using CNN and OpenCV. IEEE Access. 2023.

[15]. Verma GK, Tiwary US. Multimodal emotion recognition: survey. IEEE Trans Affect Comput. 2021.

[16]. Kollias D, Zafeiriou S. AffectNet dataset and FER benchmarking. IEEE. 2021.

[17]. Mollahosseini A, et al. AffectNet: large-scale facial expression dataset. IEEE. 2020.

[18]. Goodfellow I, Bengio Y. Deep learning for vision-based emotion detection. MIT Press. 2021.

[19]. Chen J, et al. Transformer-based emotion recognition models. IEEE Access. 2023.

[20]. Zhang et al. (2021): Introduced a 3D CNN model for spatiotemporal emotion recognition in video streams. The system demonstrated strong performance on CK+ and RAF-DB datasets.