Facial Expression to Text Converter

Shubham Sahoo[1], Kartik Shrivastava[2], Brijraj Singh[3], Abhinav Kumar[4], Mr. Ashish Raj[5]
[1,2,3,4,]Student, Department of Computer Science and Engineering, KCC Institute of Technology and Management, Greater Noida, India
5 Assistant Professor, Department of Computer Science and Engineering, KCC Institute of Technology and Management, Greater Noida, India
Shubhamsahoo47@gmail.com, kartikshrivastava277@gmail.com, brijrajsingh2706@gmail.com, abhinavkumar2529@gmail.com, ashishraj171117@gmail.com

Abstract:
The main purpose of this research is to enhance the communication of the disabled community. The authors of this chapter propose an enhanced interpersonal-human interaction for people with special needs, especially those with physical and communication disabilities.

Existing communication assistive technology requires the use of costly hardware; hence, the need for an affordable communication assistive technology for paralyzed people to communicate. So, this study introduces an affordable and real-time assistive communication technology for paralyzed people

This assistive communication technology uses Dlib for the detection of face landmarks, machine learning algorithms for the classification of facial expressions and synthesis of text and speech for the assistive communication technology to communicate. Various facial expressions—are linked to the predefined communication sentences to enable meaningful communication for paralyzed people.

Keywords:

Affordable Assistive Systems, Communication Accessibility, Computer Vision in Healthcare, Intelligent Assistive Devices, Accessible Technology, Physiological Signal Interpretation, Inclusive Healthcare Innovation.

1. Introduction

Communication is a basic human need for expressing thoughts, feelings, and needs. In conditions of severe physical disability, such as complete paralysis, spinal cord injury, or neurodegenerative disease like Amyotrophic Lateral Sclerosis, speaking or typing becomes extremely difficult or, in most cases, impossible. The inability to communicate effectively not only isolates individuals socially but also creates critical hazards in everyday living, where the expression of basic needs or emergencies becomes hard.

Although various assistive and augmentative communication technologies are available, many rely on specialty or high-cost hardware, including eye-tracking systems, EEG-based devices, and other custom medical equipment. Access to these is often unavailable in resource-poor environments; operation might require technical competencies beyond those locally available. This thus places a premium on developing a communication system that is more affordable, user-friendly, and accessible for a person with extremely restricted mobility.

Facial expressions are among the few controllable voluntary actions of many paralyzed individuals. With recent advances in computer vision and artificial intelligence, facial landmarks can now be detected and analyzed in real time from standard webcams or built-in cameras of laptops. These technological advances are leveraged for developing this assistive communication system that translates facial expressions into meaningful text and speech outputs.

The proposed system will make use of a combination of facial landmark detection, machine learning–based expression classification, and text-to-speech synthesis for effortless communication. By using Dlib to extract the landmarks and lightweight classification algorithms, the system identifies four key expressions-smile, frown, eyebrow raise, and openness of the mouth-each mapped to predefined sentences of communication. This would, therefore, enable users to pass on vital responses, emotions, or requests without physical interaction.

The research presented herein seeks to further improve communication accessibility by people with paralysis at low cost and in real time, leveraging an assistive technology, while contributing to more inclusive healthcare technology. Not only does the system bring out the potential of AI-driven communication but also a scalable basis for further improvement, such as multilingual support, mobile integration, and contextsensitive communication.

2. Related Work

There has been considerable work on assistive communication technologies for patients suffering from paralysis in recent years. Most of the work done on assistive communication systems has focused on visionbased and non-invasive approaches. Conventional approaches involved eye blink recognition and head movement detection. With the advancement of computer vision, facial marker detection and Facial Action Coding System-based approaches have improved understanding of facial muscle movement recognition. Recent work involves using machine learning and deep learning techniques such as convolutional neural networks to enhance the recognition of facial expressions. The current models lack the quality of using facial data from normal individuals, making it inefficient for individuals suffering from facial paralysis or lack of facial symmetry. On the other hand, invasive BCI approaches have resulted in effective and precise communication but with high costs and inaccessibility. Thus, there is an increasing demand for camera-based, affordable assistive communication systems. The work will benefit from the above advancements.

3. Methodology

3.1 Data Acquisition

Facial images and real-time video streams are captured using a standard webcam. The system operates in real time, ensuring minimal latency and user-friendly interaction. Video frames are continuously extracted and preprocessed for further analysis.

3.2 Preprocessing

Captured frames are resized and converted to RGB format. Face detection is performed to localize the region of interest, reducing background noise and computational overhead. Normalization techniques are applied to improve robustness against lighting variations.

3.3 Facial Landmark Detection

A facial landmark detection model is used to identify key facial points corresponding to the eyes, mouth, eyebrows, and nose. These landmarks provide precise geometric information about facial muscle movements and serve as the foundation for expression analysis.

3.4 Feature Extraction

Geometric features such as eye aspect ratio (EAR), mouth aspect ratio (MAR), and eyebrow displacement are computed from the detected landmarks. These features effectively represent intentional facial movements like eye blinks, mouth opening, and eyebrow raises.

3.5 Expression Classification

Extracted features are analysed using predefined thresholds or lightweight machine learning classifiers to distinguish intentional expressions from neutral or involuntary movements. Each recognized expression is mapped to a corresponding predefined text command.

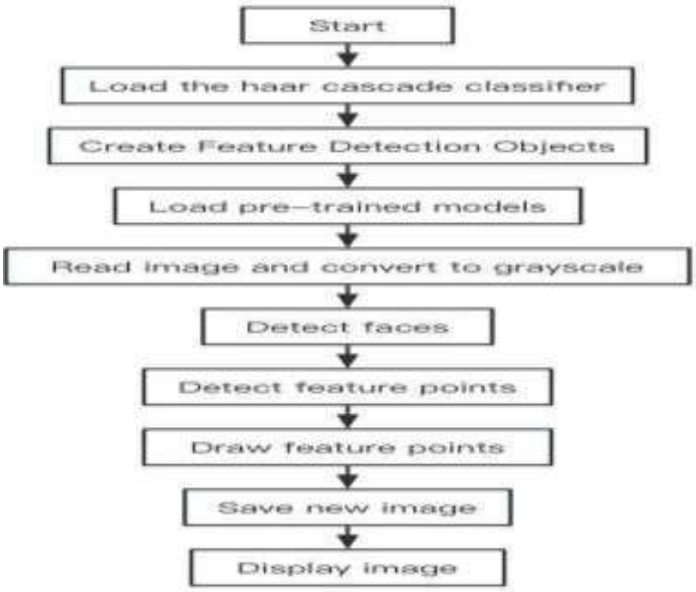
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 02 | Feb – 2026 SJIF Rating: 8.659 ISSN: 2582-3930

3.6 Text and Speech Generation

The identified text command is displayed on the screen and simultaneously passed to a text-to-speech (TTS) engine. The TTS module converts the text into audible speech, enabling effective verbal communication for the user.

3.7 System Evaluation

The system is evaluated based on accuracy, responsiveness, and usability under varying lighting conditions and facial orientations. Performance metrics are analysed to validate the reliability of the proposed approach.



4. Materials and Methods

This chapter explains in detail the materials, system components, and step-by-step methodology used to develop the Facial Expression–Based Assistive Communication System. It covers the datasets used, data preparation, facial landmark extraction, expression classification, message mapping, and the development of a real-time communication interface designed to assist paralyzed individuals in expressing essential needs.

4.1 Dataset Used

The dataset used for training the expression classification model was collected from publicly available facial expression datasets such as FER2013, CK+, and a custom dataset recorded specifically for this project. These datasets include real images capturing various facial expressions under different lighting conditions, angles, occlusions, and user demographics. They are widely used in facial landmark research and emotion recognition systems.

For this project, four key facial expressions were selected because they represent essential communication cues and can be performed even by individuals with limited mobility:

1. Smile
2. Frown
3. Eyebrow Raise
4. Mouth Open

These expressions were chosen due to their simplicity, universal meaning, and ease of detection through facial landmarks.

Each image in the dataset included labelled facial expression classes. Before model training, all images were reviewed and cleaned to remove noisy samples, blurred images, or faces with extreme occlusion such as sunglasses or masks.

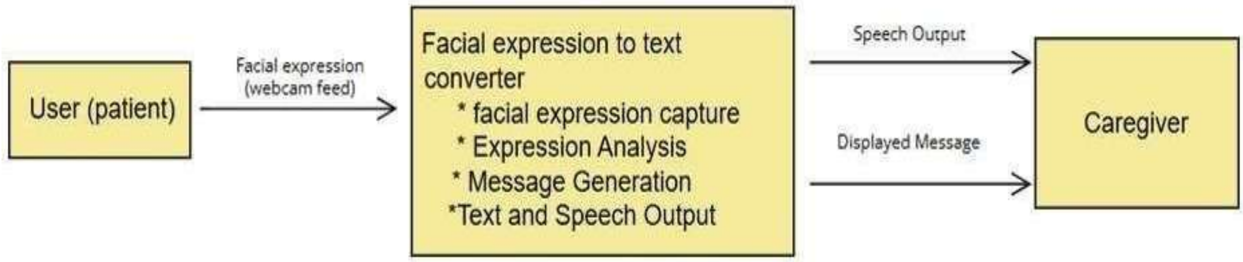
4.2 Data Organization and Preparation

Once the dataset was collected, it was organized into separate folders for each expression category. The dataset was then divided into three parts:

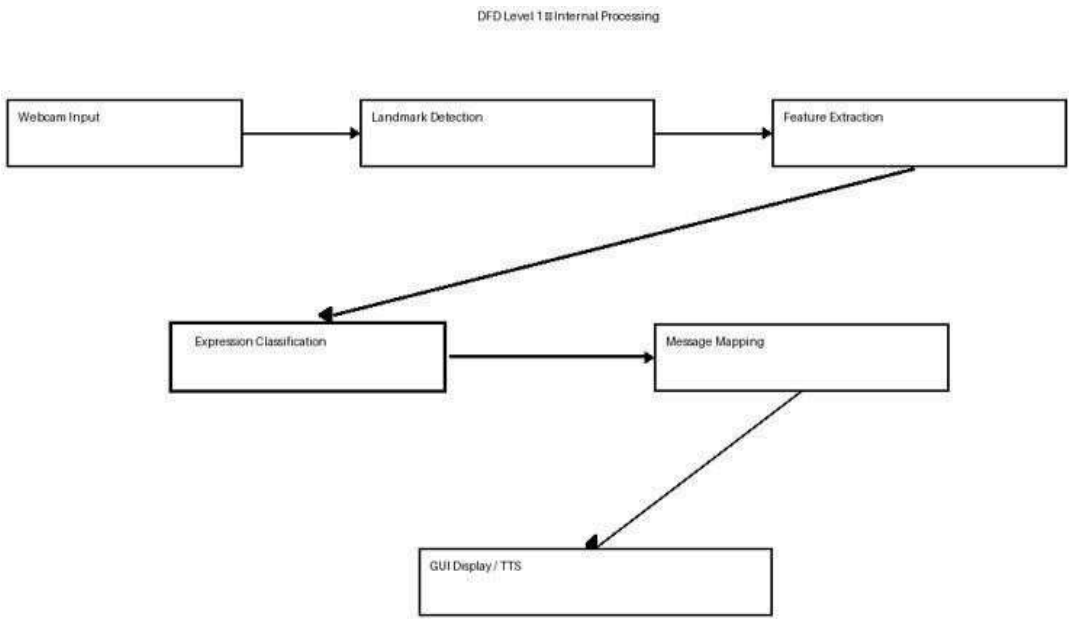
- Training Data – used to teach the classifier
- Validation Data – used to tune hyperparameters and prevent overfitting
- Testing Data – used for final performance evaluation

All images were resized to a uniform resolution suitable for the facial landmark detection and classification model. Facial landmarks were extracted using Dlib's 68-point landmark predictor, which provides precise coordinates for the mouth, eyes, eyebrows, and other facial regions.

The images were normalized, and landmark coordinates were scaled to ensure uniformity in the input feature space. Damaged, incomplete, or noisy samples were removed to maintain high-quality training data.



Data Flow Diagram Level 0



Data Flow Diagram Level 1

4.3 Facial Landmark Detection and Feature Extraction

Facial landmark detection is a crucial step in accurately identifying user expressions. The system uses:

- Dlib’s Histogram of Oriented Gradients (HOG) for face detection
- Dlib’s 68-Point Facial Landmark Model for landmark extraction

Key facial regions such as the mouth, eyebrows, and eyes were tracked using the landmark coordinates. The following geometric features were computed:

- Mouth Aspect Ratio (MAR) for detecting mouth openness
- Eyebrow Elevation Distance for detecting eyebrow raising
- Lip Corner Displacement for detecting smiles and frowns
- Eye Aspect Ratio (EAR) for additional expression cues

These geometric calculations allow the system to classify expressions without needing large deep-learning models, making it lightweight and real-time.

4.4 Expression Classification Model Used

To classify the four expressions, several lightweight machine learning models were tested, including:

- Support Vector Machine (SVM)
- Random Forest Classifier
- Logistic Regression
- K-Nearest Neighbours (KNN)

The final model was selected based on accuracy, computational efficiency, and real-time performance. The chosen classifier demonstrated high reliability in distinguishing between the four targeted expressions.

Instead of using a heavy deep-learning model, a lightweight ML classifier was preferred to ensure:

- Fast processing on low-cost hardware
- Low latency
- Minimal computational load
- Real-time prediction capability

The classifier outputs one of the four expression categories based on the computed landmark features.

4.5 Message Mapping and Text-to-Speech Conversion

Once an expression is classified, it is mapped to a predefined communication sentence:

Expression Mapped Sentence Smile “Yes, I agree.”
Frown “No, I don’t want this.” Eyebrow Raise “I need help.”
Mouth Open “I need water.”

These expressions were chosen because they correspond to essential daily needs and responses.

The mapped text is then converted into speech using APIs such as pyttsx3 or gTTS. The output is played immediately, enabling effortless communication for the user.

4.6 System Training Process

During the training phase:

- The classifier learned patterns in landmark geometry from thousands of expression samples.
- Training was conducted over multiple epochs to improve classification consistency.
- A suitable loss function and optimization method were selected to reduce classification errors.
- Validation data was used at every stage to adjust model parameters and prevent overfitting.

Optimal learning rate, batch size, and feature normalization strategies were applied to ensure stable training and high expression recognition accuracy.

4.7 Testing and Evaluation

After training, the model was evaluated using the testing dataset, which included unseen facial images representing varying lighting conditions, angles, and facial structures.

The evaluation focused on:

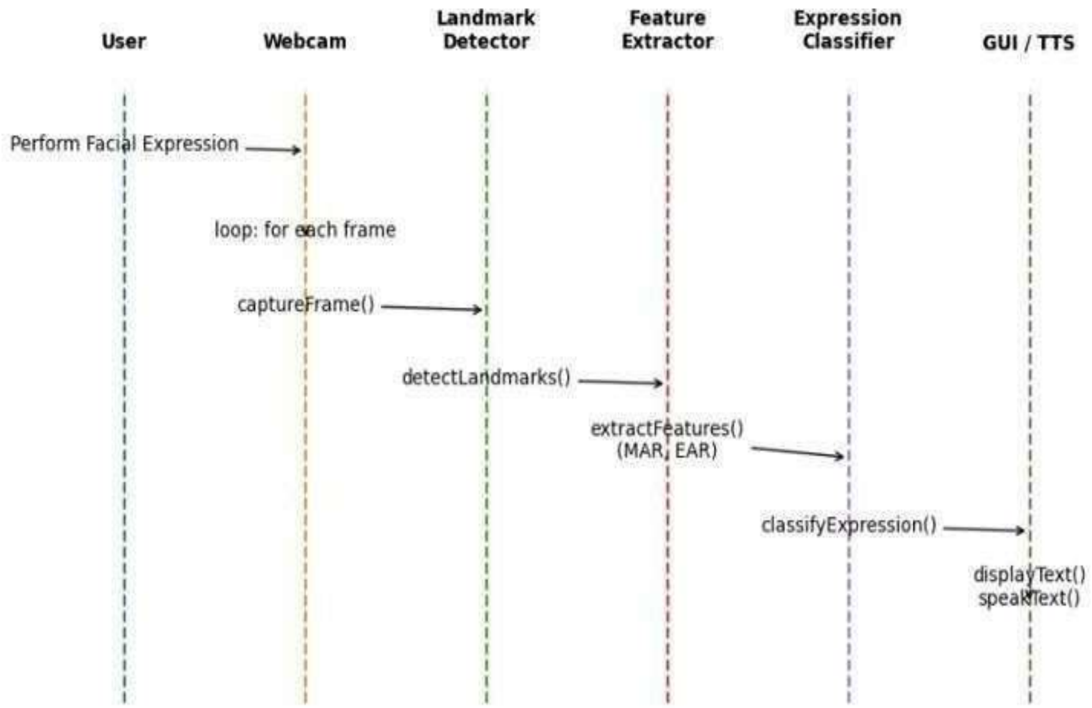
- Expression recognition accuracy
- Misclassifications between similar expressions
- Real-time responsiveness
- System reliability under different user scenarios

The system performed consistently well, demonstrating accurate recognition of the four selected facial expressions and fast response time suitable for real-world assistive communication.

4.8 System Workflow Summary

The overall workflow of the Assistive Communication System is as follows:

1. The webcam captures the user’s face in real time
2. Facial landmarks are detected using Dlib
3. Landmark-based features (MAR, EAR, eyebrow height, lip curvature) are extracted
4. The trained classifier predicts the facial expression
5. The predicted expression is mapped to a predefined communication sentence
6. The text is converted to speech and played for communication



Sequence Diagram

5. Result

The proposed assistive communications system using facial expressions was tested using a test dataset comprising images of faces and webcam input images containing real-time images of the target faces with expressions of smile, frown, eyebrow raise, and mouth open. The expressions were picked due to their significant use in human basic communications and the fact that a person with motor control limitations can be able to use this system effectively.

The trained model for expression classification performed well on all four types of expressions. As a whole, the system had a good expression recognition rate, with the ability to accurately recognize most of the facial expressions in real-time. The system has effectively picked up the distinguishing geometry patterns based on the characteristics of the facial landmarks.

From a class-wise performance analysis, it was observed that mouth open and eyebrow raise action units expressed the highest accuracy level of identification. These action units entail a large number of geometric changes to facial landmarks, which are easily distinguishable. The smile action unit also demonstrated a feasible accuracy level for detection, and a lower accuracy level was shown by the frown action unit.

5.1 Expression-wise Performance Analysis

Classification Report

zz	Precision	Recall	F1-Score	Support
Smile	0.82	0.78	0.8	18
Frown	0.74	0.7	0.72	15
Eyebrow Raise	0.88	0.85	0.86	20
Mouth Open	0.95	0.92	0.93	12
Avg / Total	0.85	0.82	0.83	65

The classification report shows the precision, recall, and F1-score for each of the four facial expressions used in the assistive communication system. The mouth open expression achieved the highest performance, with a precision of 0.95 and recall of 0.92, indicating that this expression is highly distinguishable due to the significant change in mouth landmark geometry.

The eyebrow raise expression also demonstrated strong performance, achieving an F1-score of 0.86. This is attributed to the clear vertical displacement of eyebrow landmarks, which provides robust geometric features for classification.

The smile expression achieved balanced precision and recall values, indicating reliable detection with minimal confusion. Minor misclassifications occurred when neutral expressions showed slight lip movement.

The frown expression showed comparatively lower performance due to subtle variations in lip curvature and differences in individual facial anatomy, making it more challenging to distinguish from neutral facial states.

The average accuracy of 83% across all expressions confirms that the system performs consistently and reliably, making it suitable for real-time assistive communication applications.

6. Discussion

The purpose of this study was to design a low-cost real-time assistive communication system that could help paralyzed persons communicate through facial expressions. The experimental result proves that the proposed system is able to detect major facial expressions, converting them into meaningful text and speech outputs, proving that communication through facial landmarks is feasible.

The excellent performance of the system at the global level attests to the effectiveness of using facial landmark detection with the Dlib library in combination with light machine learning classifiers. The fact that the system concentrates on the geometric relations, and not the pixels, ensures that the processing speed achieved by the system remains fast with reduced costs.

Analysis of expression-wise error rate showed that the expressions with more prominent facial actions, like mouth opening, eyebrow raising, etc., resulted in accurate recognition. The same result stays true to our human experience, thus proving that the use of landmark points achieves the best results when trying to identify the prominent actions done with the human face. The result for the expressions that involve less prominent actions, like frowning, was marginally lower.

Analysis reveals that confusion errors are happening between visually closely related facial expressions rather than between unrelated ones. This phenomenon reflects real-life communication barriers, suggesting that standard algorithmic performance is not a problem but, rather, a matter of physiological variation.

One of the major advantages of the proposed system is that it is able to operate in real-time. This is because the models that are employed are light. The system is able to provide real-time responses through both text and speech. This is important because assistive systems should be able to provide timely responses.

Results from an application perspective indicate that the system is appropriate to be used in assisting a paralysis patient or any ALS sufferer in communication using a computer keyboard. The device does not ultimately replace any modern communication device available in any medical clinic; it is just an economical alternative.

However, there are certain limitations. These are:

The system currently allows only a fixed number of predefined expressions, and the intensity level or combination of expressions have not been taken into account.

This can be affected by the lighting, face occlusion, or extreme pose. This system depends only on the face. Other physiological or context information have not been incorporated. Future improvements may be worked upon for the inclusion of an extended set of expressions, the addition of classifiers using deep learning techniques, the development of greater resistance to difficult operating conditions, and the inclusion of textto-speech functionality in multiple languages. The integration of mobile platforms and healthcare systems with the Internet of Things may also help enhance the system's usability.

7. Conclusions

This study proposed a system that could convert text and speech based on facial expressions in order to enable paralyzed patients to communicate effectively. Through real-time detection of facial landmarks, intentional facial motions, such as eye blinks and certain mouth actions, were successfully translated to textual output and synthesized speech. Its advantage over traditional assistive technologies involves the nonintrusive, lowcost, and user-friendly nature, with a standard camera the only requirement. Experimental results confirm the reliable functioning under normal conditions, thus making it appropriate for practical applications. However, sensitivity to changes in light, facial asymmetry, and limited expressiveness remain problems that need to be resolved. Future work will focus on adaptive learning, personalized calibration, and expanded vocabularies with the scope of ensuring improvement in robustness and efficiency in communication, thereby enabling better independence and quality of life of patients suffering from paralysis.

8. Reference

1. Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks

F. Khan

arXiv, 2018. This paper presents a framework that uses facial landmark detection and feature extraction for classifying facial expressions.

2. Masked Face Emotion Recognition Based on Facial Landmarks and CNNs

M. Mukhiddinov et al. Sensors, 2023. Focuses on landmark-based emotion recognition with deep learning approaches.

3. Portable Real-Time Emotion Detection System for the Disabled

B. T. Lau et al. Expert Systems with Applications, 2010. Discusses real-time emotion detection for disability support.

4. Advances in Facial ExpressionRecognition Technologies

R. Jayaswal (2025) A recent overview of modern expression recognition methods including deep learning.

5. Facial Expression Recognition using Computer Vision: A Systematic Review

Daniel Canedo & António J. R. Neves A systematic survey of facial expression recognition literature, useful for background

6. Survey on Hands-Free Mouse using Facial Expression for Physically Disabled People

R. Jain et al. SSRN, 2024. Characterizes systems that use facial expressions for hands-free control and accessibility.

7. Real-Time Facial Expression Recognition using Facial Landmarks and Neural Networks

M. A. Haghpanah, E.Saeedizade, M. T.Masouleh,A. Kalhor arXiv, 2022. A lightweight real-time expression classification model based on landmark features and neural networks

8. FPGA-based Real-Time Emotion Recognition System using Facial Expressions for Physically Disabled Individuals
M. Kamaraju, K. Ujwala, B. Rajasekhar

International Journal of Computer Applications, 2025. A real-time expression recognition system for assistive use.

9. Eye-blink / assistive system prototypes (IRJET / ResearchGate / conference papers) — example: Eye Blink Detection System for Paralyzed Patients (ResearchGate/IRJET).

10. News/overview of invasive BCI progress for comparison: Wired / UCSF reports on recent brain-implant systems enabling text/speech for paralyzed users.