AI-IOT Integration Facial Emotion Detection in Healthcare

D.Harish[1], J. Keerthika[2], B. Manikandan[3], R. Pandiyarajan[4], B. Chitra Devi[5]
[1]UG Students, Electronics and Communication Engineering, Tamilnadu College of Engineering, Coimbatore, India.
[5]Assistant Professor, Electronics and Communication Engineering, Tamilnadu College of Engineering, Coimbatore, India.

***Abstract*—This project focuses on automating facial emotion recognition using deep learning techniques, particularly leveraging Convolutional Neural Networks (CNNs) enhanced with attention mechanisms such as self-attention and channel attention. Traditional methods for recognizing and analyzing facial expressions are often limited in accuracy, sensitive to environmental variations, and prone to misinterpreting subtle emotional cues. As the demand for emotion-aware, responsive systems increases across domains such as mental health monitoring, behavioral analysis, and human-computer interaction, the need for an efficient and accurate emotion recognition system becomes crucial. The proposed model, built upon a CNN architecture integrated with attention modules, is designed to identify and classify different emotional states using enhanced feature extraction. By training the network on a large and diverse dataset of facial images, the model can not only detect a wide range of emotions but also operate effectively in real-time scenarios. This approach significantly reduces human involvement and enhances the consistency, accuracy, and speed of emotion recognition. The project aims to demonstrate how deep learning can revolutionize emotion-aware AI applications, enabling smarter, more interactive systems across various sectors. Furthermore, the proposed solution is integrated with an ESP8266 microcontroller to provide real-time output control, such as activating LEDs or buzzers based on detected emotions, making it suitable for practical deployment in smart environments and assistive technologies.**

1. INTRODUCTION

Facial emotion detection plays a crucial role in non-verbal communication, enabling computers to interpret human emotions in real-time. It enhances human-machine interaction, improves mental health monitoring, strengthens security systems, and supports behavioral analysis across a wide range of applications. Modern Facial Emotion Recognition (FER) systems rely on Deep Learning Techniques (DLT), particularly Convolutional Neural Networks (CNNs), to automatically extract meaningful facial features and classify emotions such as happiness, sadness, anger, surprise, fear, and neutral expressions. The incorporation of self-attention and channel attention mechanisms further refines this process by guiding the model to focus on salient facial regions, thereby improving classification accuracy. The overall emotion recognition pipeline involves multiple stages: face detection to locate and isolate the face from images or video streams; feature extraction using deep learning to capture key facial landmarks; emotion classification via an attention-augmented CNN model; and real-time response generation, wherein the detected emotion is communicated to an ESP8266 microcontroller. This microcontroller then activates visual or auditory indicators such as LEDs or buzzers, enabling immediate, context-aware feedback. This integration of AI and IoT technologies facilitates efficient and intelligent emotion-aware systems suitable for deployment in smart environments, assistive tools, and interactive applications.

1.2 DEEPLEARNING

This algorithm for facial emotion recognition uses Convolutional Neural Networks (CNNs) combined with attention mechanisms to efficiently classify emotions in real-time. The process begins with dataset collection, where facial images labeled with emotions are gathered and preprocessed, including resizing, normalizing, and augmenting the dataset to enhance its variability and robustness. Feature extraction is carried out using convolutional layers that detect key facial features such as the eyes and mouth. To improve recognition accuracy, an attention mechanism is applied to focus the model's learning process on critical facial regions, while suppressing irrelevant parts of the image. The extracted features are then passed through fully connected layers for classification, where the model predicts the corresponding emotion based on learned patterns. Once the emotion is identified, the results are displayed on a laptop screen with bounding boxes highlighting the detected faces. In addition, the emotion data is transmitted to an ESP32 microcontroller, which triggers LED indicators and buzzers to provide real-time feedback. This structured algorithm ensures efficient emotion detection, making it particularly valuable for applications such as mental health monitoring, human-computer interaction, and interactive gaming.

1.3 ADVANTAGES

Improved Accuracy in Emotion Recognition:
The paper demonstrates how integrating CNNs with attention mechanisms enhances facial emotion detection by focusing on critical facial features.

Real-Time Implementation: The proposed system enables fast and efficient emotion classification, making it suitable for applications like mental health monitoring and human-computer interaction.

IoT Integration for Smart Feedback: The ESP8266 microcontroller allows real-time emotion-triggered responses through LED indicators and buzzers, making the system interactive.

Comprehensive Approach: The study includes dataset collection, preprocessing, model training, evaluation, and deployment, ensuring a structured methodology.

2. LITERATURE REVIEW

1.The paper titled "Decoding Student Emotions: An Advanced CNN Approach for Behavior Analysis Application Using Uniform Local Binary Pattern," published in 2024, explores the integration of Convolutional Neural Networks (CNNs) and Uniform Local Binary Patterns (ULBP) for emotion recognition in educational environments. The authors enhance CNN performance by incorporating ULBP to capture fine-grained facial details, such as wrinkles, edges, and subtle movements. This approach improves emotion classification accuracy, enabling real-time monitoring of student engagement, stress, confusion, or interest levels during classes. The study also emphasizes the need for lightweight, efficient models capable of deployment in real-time educational environments. Although the model demonstrates promising results, challenges related to variations in lighting, pose, and the need for diverse datasets representing various emotional expressions remain

2.The paper titled "Facial Emotion Recognition in Real Time Using Deep Learning," published in 2024, discusses the advancements in facial emotion recognition (FER) through deep learning technologies. The authors compare traditional machine learning methods, such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN), with modern approaches using Convolutional Neural Networks (CNNs), which automate the feature extraction process and improve real-time performance. The paper highlights the use of deep architectures such as VGGNet, ResNet, and MobileNet, which significantly enhance the accuracy and speed of emotion recognition. To address challenges like limited training data and diverse facial expressions, techniques like transfer learning and data augmentation are employed. The paper also explores real-time FER applications in sectors such as healthcare, education, human-computer interaction, and customer service. Despite substantial progress, the study identifies ongoing challenges such as occlusions, lighting variations, and intensity differences in facial expressions as key areas for future research.

3.The paper titled "Real-Time Facial Emotion Recognition Methods Using Different Machine Learning Techniques," published in 2023, explores the evolution of facial emotion recognition (FER) through various machine learning approaches. Early methods relied on traditional classifiers, such as Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN), in combination with manual feature extraction techniques like Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). While these methods yielded moderate success, they were highly sensitive to variations in lighting, facial pose, and expressions. As technology advanced, ensemble methods like Random Forests and boosting algorithms were introduced to improve performance. The paper further highlights the shift towards deep learning, particularly Convolutional Neural Networks (CNNs), which have significantly outperformed classical methods by automating feature extraction and enabling faster, more reliable real-time emotion recognition. Despite these advancements, the study notes that challenges such as computational efficiency, real-world variability, and dataset limitations remain critical areas for ongoing research.

4.The paper titled "Real-Time Facial Emotion Detection Application with Image Processing Based on Convolutional Neural Network (CNN)," published in 2024, explores the advancements in real-time facial emotion detection through image processing and Convolutional Neural Networks (CNNs). Traditionally, image processing methods like edge detection and feature point tracking were used for emotion recognition but lacked the robustness required for real-time applications. The introduction of CNNs has significantly improved the accuracy and speed of emotion recognition by automating feature extraction from facial images. CNN architectures such as VGGNet, AlexNet, and lightweight models like MobileNet have been employed for efficient emotion recognition. The paper also highlights the integration of preprocessing techniques such as face alignment, normalization, and data augmentation to enhance model performance. Despite the high accuracy of CNN-based systems in controlled environments, challenges such as lighting variations, occlusions, and the diversity of real-world emotional expressions continue to hinder deployment. Ongoing research aims to develop more efficient, scalable, and robust models for practical real-time applications in sectors like healthcare, security, and human-computer interaction.

3. EXISTING SYSTEM

Traditional facial emotion recognition systems rely on manual observation or basic machine learning models, leading to inaccuracies and delayed responses. Existing security and surveillance systems lack emotional awareness, limiting their ability to detect stress, fear, or aggression in real time. Human-machine interaction (HMI) platforms often fail to provide adaptive responses based on users' emotions, making interactions less engaging and intuitive.

Furthermore, current IoT applications primarily focus on environmental monitoring rather than emotion-based analysis. This limitation restricts IoT's impact on mental health assessment, elderly care, and customer experience personalization. Without deep learning integration, conventional emotion recognition struggles with complex facial variations, lighting conditions, and diverse expressions, resulting in reduced accuracy and inefficient real-time feedback.

4. PROPOSED SYSTEM

The proposed system for facial emotion recognition in healthcare integrates Artificial Intelligence (AI) and the Internet of Things (IoT) to enhance real-time emotion detection and analysis. Unlike traditional methods that rely on manual observation or basic machine learning models, this system employs deep convolutional neural networks (CNNs) with attention mechanisms to improve accuracy and responsiveness. CNNs automatically extract complex facial features, focusing on critical regions such as the eyes, mouth, and eyebrows, which are essential for emotion classification. The attention mechanism further refines the process by emphasizing significant facial features, ensuring precise recognition of emotions like happiness, sadness, fear, and anger. The system is trained on a diverse dataset containing labeled facial expressions and optimized using advanced preprocessing techniques such as image normalization, resizing, and augmentation to enhance performance across different lighting and environmental conditions.

To facilitate real-time interaction, the recognized emotion is transmitted to an ESP8266 microcontroller, which controls LED indicators and buzzer alerts for immediate feedback, enabling applications in mental health monitoring, human-computer interaction, and security systems. The IoT integration allows seamless transmission of emotional data, making it valuable for early diagnosis of psychological conditions, personalized AI-driven experiences, and behavioral analysis in clinical and interactive settings. Designed for scalability and adaptability, the system supports extensions for multi-person recognition and integration with smart healthcare devices, improving accessibility and efficiency. By automating facial emotion detection, this AI-IoT system reduces human observation errors, ensures real-time responsiveness, and enhances mental health monitoring and security applications. The combination of deep learning and IoT creates a future-ready platform for emotion-aware AI-driven systems, contributing to improved healthcare diagnostics, smart interactions, and adaptive security frameworks.

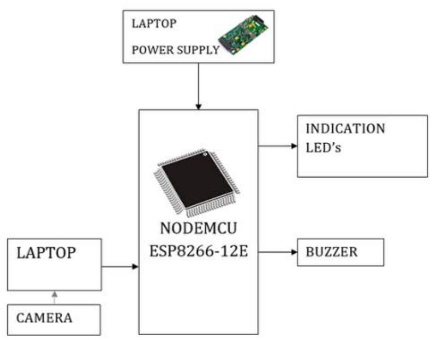

FIG 1. *PROPOSED BLOCK DIAGRAM*

5.SYSTEM REQUIREMENTS

5.1 Hardware Requirements

The hardware configuration for the proposed real-time facial emotion recognition system includes a laptop equipped with an Intel Core i7 processor (5th generation or higher), 2 GB of RAM, and a 500 GB hard disk, ensuring adequate processing power and storage capacity for deep learning model execution and real-time video processing. A standard 14-inch monitor (e.g., Samtron), Logitech 104-key keyboard, and optical mouse were used for system interaction. For emotion detection input, a USB webcam captures live video feeds, which are processed using OpenCV. To enable real-time IoT integration, the system utilizes an ESP8266-12E Wi-Fi-enabled microcontroller, which receives the classified emotion data and triggers appropriate hardware responses such as LEDs and buzzers. This setup supports a responsive, lightweight, and scalable deployment environment suitable for healthcare and human-computer interaction applications.

5.2 Software Requirements

The system must run on Windows 7 or a later version of the Windows operating system. For emotion detection and backend processing, Python version 3.6 is used along with the Flask web framework for deploying the real-time video stream. TensorFlow and Keras libraries are employed for deep learning model development and training, while OpenCV is used for face detection and image preprocessing. The Arduino IDE, along with the AVR-GCC compiler, is utilized for programming the ESP8266 microcontroller in Embedded C. Microsoft Office is used for documentation and report preparation throughout the development cycle.

5.3 Software Description

Python serves as the primary programming language in this project due to its simplicity, cross-platform compatibility, and rich ecosystem of libraries. Its intuitive syntax and readability facilitate rapid development, making it accessible for both novice and experienced developers. In this system, Python is used for implementing the Convolutional Neural Network (CNN) model integrated with an attention mechanism to detect facial emotions. Libraries such as TensorFlow and Keras enable deep learning capabilities, while OpenCV handles real-time video stream processing. Python also supports modular programming and offers extensive community-contributed resources via the Python Package Index (PyPI), which proves valuable for machine learning, image processing, and IoT integration. The Python codebase is compatible across Windows, macOS, and Linux platforms, enhancing its portability for future deployment

.Flask is a lightweight Python web framework used to build the backend of the facial emotion recognition system. It serves as the middleware between the deep learning model and the user interface. Flask is responsible for capturing real-time webcam input, processing facial images, and displaying classified emotions on a web interface.

OpenCV (Open Source Computer Vision Library) is used extensively in this project for image acquisition, face detection, and preprocessing.Itprovides high-performance tools for real-time computer vision applications. OpenCV detects facial regions in each video frame and prepares them for model input by resizing and normalizing the images. This ensures consistency in input dimensions and enhances the model's performance across diverse environments. TensorFlow and Keras form the deep learning foundation of the system.

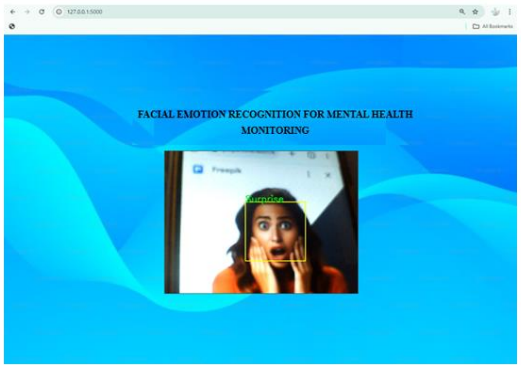

FIG 2. SOFTWARE PROCESS

HARDWARE DESCRIPTIONS

ESP8266 CONTROLLER

The ESP8266-12E is a low-cost, Wi-Fi-enabled microcontroller ideal for IoT applications. It connects to the host system via serial communication, receives emotion classification signals, and triggers outputs like LEDs or buzzers. Programmed in Arduino IDE using Embedded C, it provides real-time feedback and integrates with platforms like Blynk for remote control. Its compact design and Wi-Fi capabilities make it perfect for low-power, AI-driven projects.

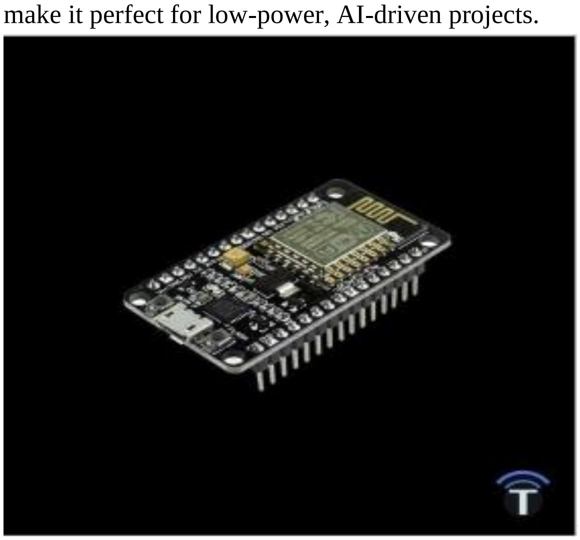

FIG 3. *ESP8266 MICRO CONTROLLER*

POWER SUPPLY

The USB cable used for programming and firmware updates of the ESP8266-12E is designed to provide 5V power with up to 2A current, ensuring reliable operation. It supports USB 2.0 or 3.0 for data transfer, allowing efficient communication during programming. The cable types available are Micro-USB or USB-C, making it compatible with a wide range of devices.

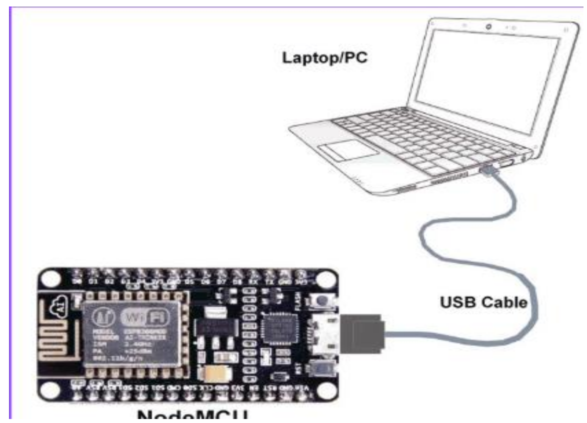

FIG 4. *POWER SUPPLY*

LED Circuit

To control an LED using a BC547 NPN transistor as an electronic switch, the circuit operates by applying voltage to the base of the transistor, which then turns it on or off. The BC547 transistor acts as a switch, allowing current to flow from the collector to the emitter when activated. Two 10K resistors are used: one to limit the current to the base of the transistor, and another to control the current flowing through the LED. When the base is not powered, the transistor remains off, and the LED turns off. This setup is commonly used in microcontroller applications, where the microcontroller controls the transistor to provide visual feedback via the LED

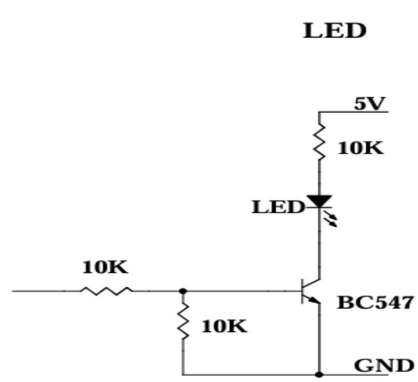

FIG 5: LED

BUZZER

A buzzer is an essential component widely used in electronic systems to provide audible alerts for alarms, household appliances, security systems, and vehicles. They are designed to notify users of important events, warnings, or faults through sound. Buzzers come in two main types: electromechanical and piezoelectric. These characteristics make buzzers ideal for providing clear and attention-grabbing alerts in environments such as smoke detectors, doorbells, washing machines, and car warning systems.

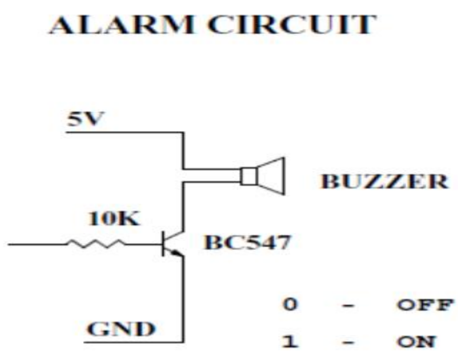

FIG 6: BUZZER

6. RESULT ANALYSIS

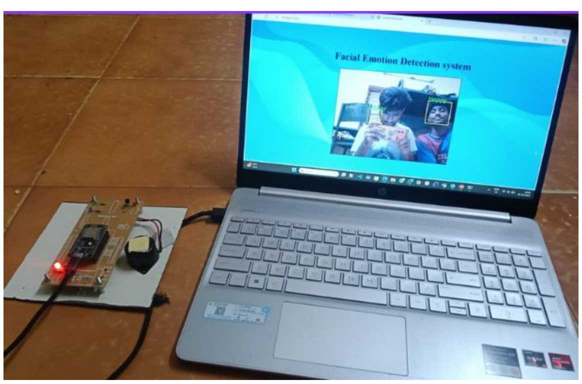

FIG *7. RESULT*

7. CONCLUSION

In this project, we developed a system that detects human emotions based on facial expressions using Convolutional Neural Networks (CNN) enhanced with an attention mechanism. The attention mechanism enables the model to focus on the most significant facial features, improving the accuracy and reliability of emotion detection. We integrated this system with an ESP32 microcontroller, which uses the identified emotion to control a light and a buzzer, demonstrating how emotions can trigger real-world actions. The system performs effectively in various environments and with diverse individuals, making it applicable to fields like mental health care, customer service, and smart robotics. This project showcases the potential of combining Artificial Intelligence (AI) with hardware to create smart systems capable of understanding and responding to human emotions in real-time.

REFERNECES

[1] Decoding Student Emotions: An Advanced CNN Approach for Behavior Analysis Application Using Uniform Local Binary Pattern" by Uppanapalli Lakshmi Sowjanya and R. Krithiga was published in IEEE Access in August 2024.

[2] Facial Emotion Recognition in Real Time Using Deep Learning" by Dr. Laxmi Math was published in the International Research Journal of Engineering and Technology (IRJET), Volume 11, Issue 7, in 2024.

[3] Real Time Fcial Emotion Recognition Methods using Different Machine Learning Techniques" is authored by K.R. Prabha, B. Nataraj, R. Ajaydevan, S. Kabilan, and V. Muthuselvam. It was presented at the 3rd International Conference on Smart Electronics and Communication (ICOSEC) in 2023.

[4] Real-Time Facial Emotion Detection Application with Image Processing Based on CNN" was authored by Jonathan Cristiano Rabika, Muhammad Rafi' Rusafni, 2024

[5] Fire and Smoke Detection using Yolo Algorithm, International Journal of Modern Engineering Research, Vol.14, Issue.3, May – June 2024.