



FACIAL EMOTION RECOGNITION USING DEEP LEARNING AND BI-LSTM

¹K.MADHUSRI,²Mrs.A.SIVA NAGA DURGA,³S. NIHARIKA,⁴P. BALAHARINI,⁵G. VYSHNAVI

¹Associate Professor of ECE Dept, Rise Krishna Sai Prakasam Group Of Institutions,Valluru,AP

^{1,3,4,5}Students, ECE Dept, RISE Krishna Sai Prakasam Group of Institutions,Valluru,AP

¹madhukotte630@gmail.com,²sivadurga3@gmail.com,³somineniniharika18@gmail.com,⁴harinipalna
ti12@gmail.com,⁵gavadagatlavyshnavi@gmail.com

ABSTRACT

Recognizing facial expressions is crucial component in the development of Human-Computer interaction (HCI) since it makes it possible for the machine to understand and respond to human emotions. The deep learning techniques for FER presented in this article are based on the fusion of a bidirectional long short-term memory network and a convolutional neural network. These methods are meant to be applied in order to create strong emotion classification systems by utilizing temporal and spatial factors. The proposed method takes into account the JAFFE and CK+ datasets. The five basic emotion classes—happy, sad, surprised, angry, and fear—are represented by the images in both sets. The process of preprocessing begins with the detection of faces through the utilization of Multi-task Cascaded Convolutional Networks (MTCNN), followed by the normalizing and scaling of the faces in order to maintain consistency in the input dimensions. Conversely, the CNN component is in charge of retrieving spatial properties from the images, while the BI-LSTM component is in charge of capturing temporal dynamics that describe face expressions across time. It is clear that such a combination architecture will lead to an increase in recognition accuracy because both types of features are integrated into a single structure. Its accuracy significantly increased when it was trained on the pre-processed datasets for a total of 25 epochs. The model's accuracy during training and validation was 95.2% and 93.1%, respectively, indicating that it can effectively generalize to data that it has never encountered before. The problems related to illumination, position, and the fluctuating intensity of expressions are all efficiently addressed by the combination of CNN and Bi-LSTM.

Keywords— Emotion Classification, Spatial Feature Extraction, Temporal Modelling, Face Detection

I INTRODUCTION

In the era of digital technology, the search for more natural and intuitive Human-computer interaction (HCI) has advanced significantly as a result of user interfaces. Among the different approaches Lately, facial emotion recognition (FER) has become a transformational methodology. By making it possible for computers to understand and react to human emotions, FER has the capacity to transform how people engage with technology, making more compelling, sympathetic, and successful interactions [1]. This work aims to explore the application of deep learning techniques, specifically Bi-directional Long Bi-LSTM (Short-Term Memory) networks, in order to FER. Our goal is to develop an accurate and dependable emotion detection system using the Japanese Female Facial Expression (Jaffe) dataset and the CK+ dataset, which is shown in Figure Because they provide important insights into the emotional state of the person being

spoken with, a person's facial expressions are a crucial part of human communication [2]. Conventional human-computer interaction (HCI) systems often rely on direct user input, including typing on the keyboard or clicking the mouse. keyboard, this can be a time-consuming and uncomfortable procedure. Real-time inference of user emotions is made possible by the integration of user emotion recognition (FER) into these systems, allowing for more flexible and accommodating relationships [3]. This capacity is particularly helpful in a range of applications, including virtual reality, gaming, mental health monitoring, customer service.



SAMPLE PICTURES:



II LITERATURE SURVEY

Recent advancements in emotion recognition have increasingly focused on deep learning techniques due to their ability to automatically learn complex patterns from data.[4]Zhentao Huang and Ya hong Ma (2023) proposed a CNN–Bi-LSTM model with an attention mechanism for EEG-based emotion recognition. In this approach, convolutional neural networks (CNN) are used to extract spatial features from EEG signals, while Bi-LSTM networks capture temporal dependencies in brain signal sequences. The attention mechanism further enhances performance by focusing on the most relevant signal segments, leading to improved emotion recognition accuracy. However, despite its effectiveness, the model requires large-scale and high-quality EEG datasets, and the data collection process is often expensive and time-consuming. In contrast to physiological signal-based approaches, vision-based methods have gained popularity due to their non-intrusive nature.[5]Hafiz Arslan Ramzan and Ahmed Sohaib (2024) presented a deep learning-based facial emotion recognition system that automatically identifies human emotions from facial expressions. Their work demonstrates that deep neural networks can effectively learn discriminative facial features, resulting in high emotion recognition accuracy across different expressions. Nevertheless, the system's performance is highly dependent on large and diverse facial image datasets, and it may face challenges such as illumination variations, occlusions, and pose changes. Similarly,[6]Ng Chin Kit and Chee-Pun Ooi (2023) proposed a

facial emotion recognition framework using deep learning-based face detectors and classifiers. The system first detects faces from input images and then classifies emotions, thereby automating the entire recognition pipeline. This approach simplifies emotion analysis and improves efficiency, making it suitable for real-world applications. However, it requires extensive labeled datasets for training and may suffer from reduced accuracy in unconstrained environments.[7]Panel Aziguli Wulamu (2025) introduced an enhanced multi-modal emotion recognition approach using feature-level fusion. By combining features from multiple data sources, such as facial expressions and other modalities, the system achieves higher accuracy and robustness compared to single-modal approaches. Despite these advantages, the model is computationally complex and requires synchronized multi-modal data, which increases system design and implementation challenges. Furthermore,[8] Saiyed Umer and Ranjeet Kumar Rout (2022) proposed a deep fusion-based facial expression recognition system aimed at enhancing human–computer interaction (HCI). Their approach integrates deep learning models to recognize emotions in real time, enabling more natural and intelligent interaction between humans and machines. While the system demonstrates real-time performance and improved interaction quality, it involves complex model architecture and higher computational requirements.

III EXISTING SYSTEM

In order to ensure that the system learns from a variety of expressions and conditions, the current facial emotion recognition system starts with data acquisition, where facial images or video frames are gathered using cameras, sensors, or publicly available datasets. After that, the data is preprocessed to improve quality and accuracy. This includes eliminating noise, resizing images, adjusting brightness and contrast, and occasionally converting images to grayscale.[9]

After preprocessing, face detection is performed to locate and isolate the facial



region from the background using algorithms such as Haar Cascade, HOG, or deep learning-based methods, ensuring that only relevant facial information is analyzed. Once the face is detected, facial feature extraction is carried out to identify important features such as the eyes, nose, mouth, eyebrows, and facial landmarks that reflect subtle muscle movements associated with emotions.[10]

Once the face is detected, facial feature extraction is carried out to identify important features such as the eyes, nose, mouth, eyebrows, and facial landmarks that reflect subtle muscle movements associated with emotions.[11]

These features are then sent to the emotion classification stage, where machine learning or deep learning algorithms like Support Vector Machine (SVM) or Convolutional Neural Networks (CNN) analyze the patterns and classify the facial expressions into emotions such as happy, sad, angry, fear, and surprise.[12]

Finally, the system produces the output by displaying the recognized emotion, sometimes along with a confidence score, enabling its use in applications like human-computer interaction, mental health monitoring, smart surveillance, and behavior analysis, thereby improving communication between humans and intelligent systems.[13]

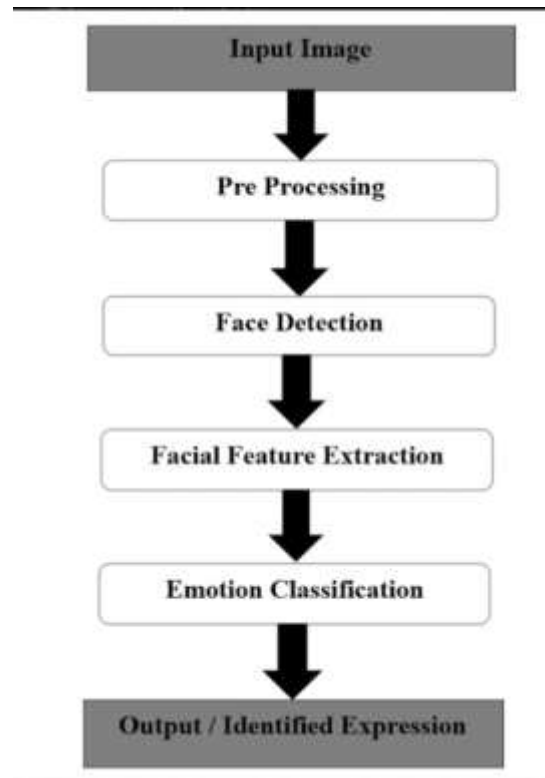


Fig.1 Block Diagram

IV RESEARCH METHODOLOGY

Ten Japanese female models captured 213 images of expressive facial expressions for the Japanese Female Facial Expression (JaFFE) collection. Each image is given one of the following six basic emotions: joy, grief, astonishment, frustration, disgust, and horror. Each image also includes a statement of balance. For training and assessment, the annotated dataset offers trustworthy ground truth labels. JaFFE is very useful for research on East Asian face expressions and features.[14]

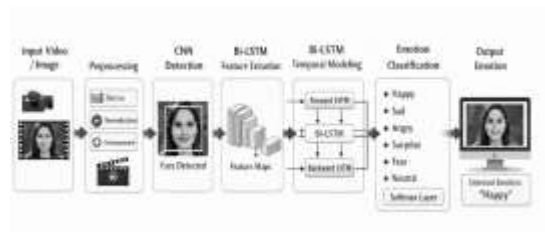


Fig.2 proposed system architecture

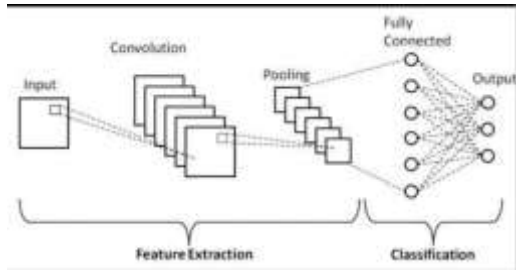


Fig.3 CNN(Convolution Neural Network)

Data Preprocessing:

Using a face identification technique, such as the Multi-task Cascaded Convolutional Networks (MTCNN) or the Haar Cascade classifier, identify and crop faces from the photos. This is referred to as face detection. Through the process of normalization, the pixel values of the photographs are normalized to a range of [0, 1] by dividing them by 255. To ensure that the CNN receives inputs with consistent proportions, it is necessary to resize all of the images to a predetermined size, such as 48 by 48 pixels.[15]

Feature Extraction using CNNs

When the facial photos have been processed beforehand CNNs are utilized in order to derive spatial information from the pictures. It is common for the architecture to be composed of several convolutional layers, which are subsequently followed with layers that pool. Accessible filtering are applied to the pictures provided by the convolutional layers, which results in the production of feature maps. The geographical dimensions of these feature maps are reduced by the layers used for pooling, which allows the most relevant features to be preserved while simultaneously minimizing the effort of the computations.[16]

Temporal Modeling using Bi-LSTM

In order to feed the features of maps that were recovered by CNN's algorithm into a BiLSTM system, they are first flatten. Bidirectional processing of the series of features is performed by the BiLSTM, which captures time dependence in both directions. At each particular time step, the result of the BiLSTM takes into account data gathered from previous and upcoming states.[17]

Emotion Classification

To generate the distribution of probabilities over the emotion classes, the last of the concealed states associated with Bi-LSTM are processed by passing them through a layer that is densely linked and fully connected together with a function that activates Softmax.[18]

V SOFTWARE DISCRIPTION

The proposed facial emotion recognition system is developed using MATLAB, which provides a powerful environment for image processing, deep learning, and data analysis. MATLAB offers built-in tools that simplify the design and implementation of intelligent systems, making it suitable for research and academic projects.[19]

The system utilizes the Image Processing Toolbox for tasks such as image acquisition, face detection, preprocessing, resizing, and normalization to enhance image quality. The Deep Learning Toolbox is used to design and train convolutional neural networks (CNN) for extracting important facial features. These extracted features are then passed to the Bidirectional Long Short-Term Memory (Bi-LSTM) network to capture temporal dependencies in facial expressions, improving the accuracy of emotion classification.[20]

MATLAB also supports dataset management, model training, validation, and performance evaluation through graphical tools and visualization plots. The software enables real-time emotion detection using webcam input and provides clear graphical outputs for analysis. Overall, MATLAB offers a user-friendly and efficient platform for developing a robust facial emotion recognition system that enhances human-computer interaction.[21]

VI RESULTS AND PERFORMANCE ANALYSIS

Figure 3 shows sample output images predicted by the proposed CNN model. The system preprocesses facial images and extracts feature to detect emotions. Images are resized to 48x48 for efficient model processing. The highest confidence score indicates the predicted emotion. Results show accurate classification of human emotions.[22]



Fig 4: Sample Classification Results

Figure 5 shows the training and validation accuracy and loss during CNN training. The training progress graph shows that model accuracy gradually improves across epochs. The loss curve slightly decreases, indicating better learning and reduced prediction errors. Overall, the results suggest the model is effectively training and improving performance.[23]

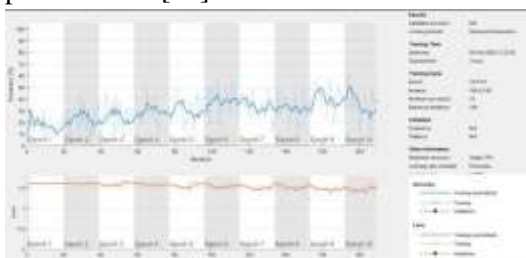


Fig.5 Training Accuracy and Loss Curve

VII CONCLUSION

Using deep learning and temporal modeling with Bi-LSTM, this project effectively created a MATLAB-based facial expression identification system to improve human-computer interaction. In order to accurately classify emotions, the system uses a Convolutional Neural Network to extract face features and a Bidirectional LSTM network to describe temporal fluctuations in facial expressions. Effective model creation, training, and performance assessment were made possible with MATLAB's Deep Learning and Computer Vision Toolboxes. The suggested system illustrates the potential of emotion-aware intelligent systems for interactive and real-time applications and shows dependable emotion recognition from image and video data.[24]

REFERENCES

1. A. Wulamu, Y. Wu, and X. Liu, "Enhanced Multi-Modal Emotion Recognition Using Feature-Level Fusion," published on 15 December 2025.
2. M. Wang, "Research on Facial Emotion Recognition Based on Deep Learning," published on 29 September 2025.
3. P. Karthikeyan and K. Kirutheesvars, "Facial Emotion Recognition for Enhanced Human-Computer Interaction Using Deep Learning and Temporal Modelling with Bi-LSTM," published on 05 April 2024.
4. Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463-474. <https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474>
5. Prasad Rongali, L., & Kumar Budda, G. A. (2020). Architecting Scalable and Secure Infrastructure for DevOps, Patterns and Practices for Cloud, Hybrid, and On-Premise Environments. International Journal of Enhanced Research In Science Technology &



- Engineering, 09(10), 32–40.
<https://doi.org/10.55948/ijerste.2020.1004>
6. Bhagwat, V. B. (2025). Simplifying Payroll Balance Conversions in Payroll Systems Implementation through the Use of Generative AI.
 7. Todupunuri, A. (2025). Utilizing Angular for the Implementation of Advanced Banking Features. SSRN Electronic Journal.
<https://doi.org/10.2139/ssrn.5283395>
 8. Ganji, M. (2025). Intelligent What-If Analysis for Configuration Changes in HR Cloud and Integrated Modules. International Journal of All Research Education and Scientific Methods, 13(04), 4828–4835.
<https://doi.org/10.56025/ijaresm.2025.1304254828>
 9. Vikram, S. (2025). Model-Centric Data Validation: A Feedback-Loop Approach to Dynamic Quality Control. 2025 3rd World Conference on Communication & Computing (WCONF), 1–7.
<https://doi.org/10.1109/wconf64849.2025.11233540>
 10. Mallick, P. (2020). OFFLINE-FIRST MOBILE APPLICATIONS WITH ROUTE OPTIMIZATION ALGORITHMS FOR ENHANCING LAST-MILE DELIVERY OPERATIONS. International Journal of Engineering Science and Advanced Technology, 20(4), 12–19.
<https://doi.org/10.64771/ijesat.2020.v20.i04.pp12-19>
 11. D. Harianto and S. Filbert, “Analyzing Customer Satisfaction Through Face Emotion Recognition: A Comparative Study of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM),” published on 01 October 2024.
 12. The Future of Conversational AI in Banking: A Case Study on Virtual Assistants and Chatbots*: Exploring the Impact of AI-Powered Virtual Assistants on Customer Service Efficiency and Satisfaction. (2024). International Research Journal of Economics and Management Studies, 3(10).
<https://doi.org/10.56472/25835238/irjems-v3i10p124>
 13. N. C. Kit, C.-P. Ooi, W. Tan, Y.-F. Tan, and S. Cheong, “Facial Emotion Recognition Using Deep Learning Detector and Classifier,” published on 01 June 2023.
 14. Z. Huang, Y. Ma, and R. Wang, “A Model for EEG-Based Emotion Recognition: CNN-Bi-LSTM with Attention Mechanism,” published on 22 July 2023.
 15. A. Rehman Khan, “Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods,” published on 25 May 2022.
 16. S. Umer and R. K. Rout, “Human–Computer Interaction Using Deep Fusion Model-Based Facial Expression Recognition System,” published on 21 October 2022.
 17. M. Kalpana Chowdary and D. Jude Hemanth, “Deep Learning-Based Facial Emotion Recognition for Human–Computer Interaction Applications,” published on 22 April 2021.
 18. H. A. Ramzan, A. Sohaib, and S. Ramzan, “Facial Emotion Recognition Using Deep Learning,” published on 19 March 2024.