International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 605-618.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19561956

Facial Emotion Recognition System Using CNN–BiLSTM Deep Learning Approach

Dr. P BalaMurali Krishna, Pasupuleti Kalyani, Thati Jyothirmai, Vendikatla Abdulla, Peddarapu Venkatesh

Department of Electronics and Communications Engineering, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India.

To Cite this Article
Dr. P BalaMurali Krishna, Pasupuleti Kalyani, Thati Jyothirmai, Vendikatla Abdulla & Peddarapu Venkatesh (2026). Facial Emotion Recognition System Using CNN–BiLSTM Deep Learning Approach. International Journal for Modern Trends in Science and Technology, 12(SI01), 605-618. https://doi.org/10.5281/zenodo.19561956

Article Info
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

KEYWORDS

Facial Expression Recognition (FER), Human-Computer Interaction (HCI), Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory (Bi-LSTM)

ABSTRACT

Facial Expression Recognition (FER) is a crucial element in advancing Human-Computer Interaction (HCI), enabling systems to interpret and respond to human emotions. This study proposes a deep learning-based FER framework that integrates Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks to effectively capture both spatial and temporal features of facial expressions. The proposed method utilizes Cascaded Convolutional Networks (CNN) for face detection, followed by normalization and scaling to ensure uniform input dimensions. Spatial features are extracted using CNN, while Bi-LSTM layers capture the temporal dynamics of facial expressions. Experiments are conducted on the datasets, each containing fundamental emotion classes: happiness, sadness and neutral. The model is trained over epochs and shows a significant improvement in recognition accuracy compared to conventional methods. This spatial-temporal architecture demonstrates robust performance, making it well-suited for real-time applications in affective computing, healthcare, education, and security systems.

I. INTRODUCTION

Facial Emotion Detection is an emerging field of computer vision that focuses on detecting and interpreting human emotions from facial expressions. Emotions play a vital role in human communication and are essential for effective social interaction. Expressions such as happiness, sadness, anger, and surprise are conveyed through facial movements, including muscle contractions, eyebrow raises, and eye blinks. Facial Emotion Detection systems aim to analyze these visual cues and accurately classify the corresponding emotional states. The ability to recognize emotions from facial expressions has numerous practical applications, including improved human–computer interaction, enhanced customer service experiences, and support for mental health assessment. For instance, a Facial Emotion Detection system can be used to analyze customer reactions in real time to assess satisfaction levels or to

identify individuals who may be experiencing emotional conditions such as depression or anxiety.

In the digital era, the demand for more intuitive and natural user interfaces has driven significant advancements in human–computer interaction (HCI). Among the various approaches, Facial Emotion Recognition (FER) has emerged as a powerful technique. By enabling computers to perceive and respond to human emotions, FER has the potential to transform user–technology interactions, making them more engaging, empathetic, and effective. In recent years, deep learning models, particularly Convolutional Neural Networks (CNNs), have shown remarkable performance in image classification tasks, including Facial Emotion Detection. CNNs are a type of neural network that can automatically learn spatial hierarchies of features from images. They have been successfully applied to various computer vision tasks, including image classification, object detection, and face recognition.

In this project, we aim to build a Facial Emotion Detection system using CNNs. The system will take an input image of a face and classify the emotion displayed in the image. We will train the system using a dataset of labeled images of faces and corresponding emotions and evaluate its performance using standard metrics. Our goal is to develop a Facial Emotion Detection system that achieves high accuracy in emotion recognition and that can be deployed in real-world scenarios. By leveraging the power of CNNs and advanced computer vision techniques, we aim to contribute to the growing body of research in the field of Facial Emotion Detection and its applications. The purpose of this study is to investigate the use of deep learning strategies, namely Bi-directional Long Short-Term Memory (BiLSTM) networks, for the purpose of FER. Using the Japanese Female Facial Expression (JaFFE) dataset as well as the CK+ dataset which can be found in Figure 1, our objective is to create an emotion recognition system that is both reliable and accurate.

Figure 1: Sample images in JaFFE and CK+ dataset

A person's facial expressions are an essential component of human communication because they offer valuable insights into the emotional state of the person being communicated with [2].Traditional human-computer interaction (HCI) systems frequently depend on direct input by users, such as clicking the mouse or typing on the keyboard, this can be a laborious and awkward process. The incorporation of user emotion recognition (FER) into these systems enables real-time inference of user emotions, which in turn facilitates connections that are more adaptable and accommodating [3]. This capability is particularly valuable in a variety of applications, including virtual reality, gaming, mental health monitoring, customer service, and education.

The science of computer vision has seen major advancements in the application of deep learning, which is a subset of machine learning. Deep learning has achieved amazing success in various tasks, including picture identification, recognizing objects, and identification of faces. Convolutional Neural Networks, sometimes known as CNNs, have emerged as the industry standard for static picture analysis, which includes the recognition of individual face expressions [4]. [5]. This is done in order to address the issue that was mentioned earlier.

The utilization of deep learning architectures for the purpose of analyzing facial expressions is the central focus of this research. The suggested method is grounded in convolutional neural networks (CNNs), which are responsible for extracting features from facial photographs. Through the use of a number of convolutional neural networks, pooling, and non-linear activations, CNNs are able to effectively capture spatial hierarchy in images. The retrieved characteristics are used as inputs to the Bi LSTM network, which is responsible for modelling the temporal changes of facial expressions.

2.LITERATURE REVIEW

In Facial Emotion Detection has been a subject of interest in computer vision research for several decades. In the early days, researchers used traditional computer vision techniques such as Haar cascades, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG) to detect facial features and infer emotions from them. However, these methods have limited accuracy and are prone to errors due to variations in facial expressions and lighting conditions. More recently, machine learning techniques, particularly Convolutional

Neural Networks (CNNs), have shown remarkable performance in Facial Emotion Detection. CNNs are a type of deep learning algorithm that can learn complex patterns and features from images. They have been successfully applied to various computer vision tasks, including image classification, object detection, and facial recognition.

Several studies have explored the use of CNNs for Facial Emotion Detection. For example, a study by Parkhi et al. (2015) used a CNN model to classify six basic emotions (happy, sad, angry, surprised, disgusted, and fearful) from facial expressions in images. They achieved an accuracy of 63.9% on the Emotion Recognition Challenge dataset, which was a significant improvement over traditional method.

Another study by Goodfellow et al. (2013) used a CNN model to generate synthetic facial expressions, which were used to train a separate CNN model for Facial Emotion Detection. They achieved an accuracy of 56.2% on the JAFFE dataset, which contained images of six basic emotions. Other researchers have explored the use of transfer learning, data augmentation, and ensemble learning to improve the performance of Facial Emotion Detection models. Transfer learning involves using a pre-trained model as a starting point for a new task, while data augmentation involves generating new training data by applying transformations to existing data. Ensemble learning involves combining multiple models to improve performance.

Deep Facial Expression Recognition: A Survey — Shan Li &Weihong Deng (2018). This survey summarized the transition from handcrafted features to deep learning (CNNs, CNN+RNNs) for both static and dynamic FER, highlighting two major challenges: limited labeled data leading to overfitting, and expression-irrelevant variations (identity, pose, illumination). It also cataloged datasets, training strategies, and evaluation protocols used in the deep-FER literature.

Deep residual networks as backbones (ResNet) — Kaiming He et al. (2015/2016). ResNet's residual learning enabled much deeper CNNs to be trained reliably; as a consequence ResNet variants became standard backbones for FER models (improving representation power and enabling transfer learning from ImageNet). Many modern FER studies fine-tune ResNet variants on FER datasets.

Attention mechanisms and coordinate/neighbor attention — recent works (2020–2024). Attention modules (spatial, channel, coordinate-aware) were introduced into FER networks to focus learning on salient facial regions (eyes, mouth, brows) and suppress background/identity cues. Recent papers show attention improves robustness to occlusion and pose, and coordinate attention variants help the network retain positional information while modeling long-range dependencies

Vision Transformers (ViT) and Mask/Hybrid Transformers for FER — (2021–2024). Transformers (and ViT variants) have been adapted to FER by treating images as visual token sequences or combining CNN feature maps with transformer layers. Approaches like Mask-Vision-Transformer (MVT), hybrid local-global attention ViTs, and ViT+SE blocks have achieved competitive or state-of-the-art results on in-the-wild benchmarks by better modeling global relations and context. However, they usually need careful regularization or transfer learning because FER datasets are still smaller than typical vision corpora.

3.EXISTING SYSTEM

In Face detection is the process of locating and identifying human faces in a frame of an image or video. By analyzing the image data and identifying facial features like the eyes, nose, and mouth, machine learning algorithms are used in this situation. The algorithm can then crop the image to isolate the face once it has been identified as being present, or it can use additional processing methods like facial recognition or emotion detection. Contrarily, face tracking describes the process of continuously monitoring a face's position and movement within a video sequence. To do this, computer vision and machine learning algorithms are combined to track the position and motion of the face in each frame of the video. Face tracking has numerous uses, including in virtual reality and gaming as well as video surveillance systems.

Face tracking and face detection have a wide range of real-world uses in industries like security, entertainment, and advertising. Face detection, for instance, can be used for security in airports or other high-security locations to identify people who might pose a threat. In order to give users more immersive experiences, face tracking can be used in gaming and virtual reality applications. Face tracking and detection

can be used in advertising to examine how consumers respond to messages and create more specialized marketing plans. Earlier localization of facial feature points focused on two or three key points, such as locating the center of the eyeball and the center of the mouth, but later introduced more points and added mutual restraint to improve the accuracy and stability of positioning Sex. The article Active Shape ModelsTheir Training and Application is a model of dozens of facial feature points and texture and positional relationship constraints considered together for calculation. Although ASM has more articles to improve, it is worth mentioning that the AMM model, but also another important idea is to improve the original article based on the edge of the texture model. The regression-based approach presented in the paper Boosted Regression Active Shape Models is better than the one based on the categoricalapparent model. The article Face Alignment by Explicit Shape Regression isanother aspect of ASM improvement and an improvement on the shape model itself. Based on the linear combination of training samples to constrain the shape, the effect of alignment is currently seen the best. The purpose ofthe facial feature point positioning is to further determine facial feature points(eyes, mouth center points, eyes, mouth contour points, organ contour points, etc.) on the basis of the face area detected by the face detection / tracking, s position.

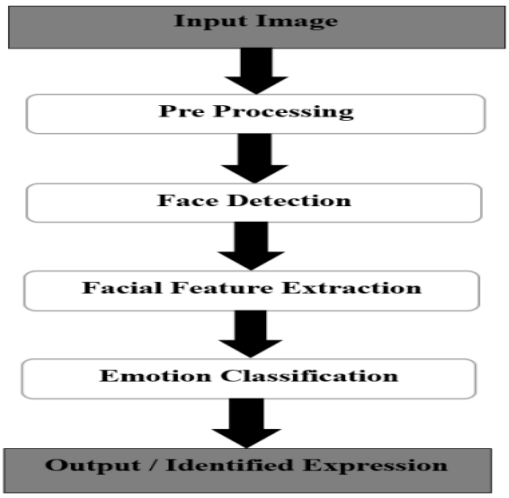

Figure 2: Flow Chart of the existing system

A combination of global and grid features are extracted from face images. The global features such as distance between two eye balls, eye to nose tip, eye to chin, and eye to lip is calculated in Fig. 2. Using four distance values, four features F1, F2, F3, and F4 is calculated as follows:

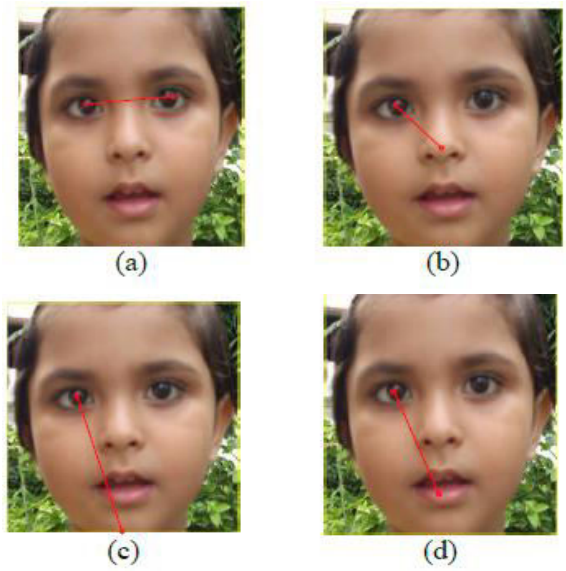

Figure 3: Distance between (a) two eyeballs (b) eye to the nose tip (c) eye to chin (d) eye to lip

Using the Grid features of face image, feature F5 is calculated. It is entirely based on wrinkle geography in face image. The grid feature includes forehead portion, eyelid regions, upper portion of cheeks and eye corner regions as shown in Fig. 2(a). To calculate feature F5, the following steps have to be followed: The color face image is converted into gray scale image. Then canny edge detection technique is applied on gray scale face image. It gives a binary face image with wrinkle edges as shown in Fig. 2(b). The white pixels of the wrinkle regions in Fig. 3(b) give wrinkle information in the face image. (a)

F1 = (distance from left to right eye ball) / (distance from eye to nose)

F2 = (distance from left to right eye ball) / (distance from eye to lip)

F3 = (distance from eye to nose) / (distance from eye to chin)

F4 = (distance from eye to nose) / (distance from eye to lip)

face recognition and emotion detection framework, the process begins with the acquisition of a human face image from real-world conditions, such as surveillance cameras or unconstrained datasets like LFW (Labeled Faces in the Wild). Although some LFW test protocols may not fully reflect realistic operational scenarios, the dataset remains one of the closest approximations to real-world face data due to its variations in pose, illumination, expression, and background. The captured face image is first pre-processed through face detection, alignment, normalization, and illumination correction to

ensure robustness against non-uniform lighting and pose variations. This pre-processing stage produces a standardized "face map," which serves as a structured representation of the facial region.

Once the face map is obtained, facial feature key points are detected using landmark localization algorithms. Typically, **5** to 6 major facial characteristics are identified, such as the positions of the eyes, eyebrows, nose tip, nostrils, mouth corners, upper and lower lip points, chin, and jawline contours. These facial landmarks are represented as **two-dimensional coordinate points**, forming a geometric description of the face. The spatial relationships among these key points—such as distances, angles, ratios, and relative positions—are then computed to capture the structural characteristics unique to each individual's face.

Following landmark extraction, multiple feature descriptors are generated from the face map and key-point coordinates. These may include geometric features (landmark-based distances and angles), texture-based features (local binary patterns or gradient information), and appearance-based features derived from deep or handcrafted models. To improve robustness, features from different sources are combined using decision-level fusion, where independent classifiers or feature extractors contribute complementary information. This fusion stage helps reduce noise, handle occlusions, and improve recognition accuracy under varying conditions.

After fusion, the high-dimensional feature set contains redundant and correlated information, which may negatively affect classification performance. To address this issue, **Linear Discriminant Analysis (LDA)** is applied for feature extraction and dimensionality reduction. LDA projects the fused feature space into a lower-dimensional subspace while maximizing inter-class separability and minimizing intra-class variance. This step ensures that only the most discriminative features are retained, making them more suitable for subsequent classification tasks such as face recognition or emotion detection.

The output of the LDA process is a compact numerical string known as the "face feature vector." This feature vector uniquely characterizes an individual's face by encoding both geometric and appearance-based information in numerical form. Each face is thus represented as a fixed-length vector that can be efficiently stored, compared, and classified. During recognition, similarity measures or classifiers compare these numerical strings to determine identity or emotional state.

Finally, the extracted face feature vectors are utilized for emotion detection by mapping variations in facial feature coordinates and textures to predefined emotional categories such as happiness, sadness, anger, fear, surprise, and neutrality. Changes in key facial regions—particularly around the eyes, eyebrows, and mouth—play a crucial role in emotion discrimination. By combining LDA-optimized features with machine learning classifiers, the system achieves reliable emotion recognition performance, even in unconstrained real-world environments.

Draw backs

- Sensitive to pose variations, occlusions, and illumination changes
- Limited performance in recognizing subtle or complex emotions
- ASM relies heavily on accurate landmark detection
- LDA assumes linear separability, which may not hold for complex emotions
- Lower accuracy compared to modern deep learning-based approaches

4.PROPOSED SYSTEM

The Facial Emotion Detection was the use of traditional computer vision techniques, such as Haar cascades, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). These methods have been shown to be effective in detecting facial landmarks andanalyzing their movements to infer emotions. However, they have limitationsin dealing with variations in lighting, pose, and occlusion, which can significantly affect their performance. In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown remarkable performance in Facial Emotion Detection. Several studies have used CNNs to learn features directly from the raw pixel values of the input images. For example, the Facial Action Coding System (FACS) is a widely used method for coding facial expressions. Researchers have used CNNs to automatically learn the FACS Action Units (AUs) from images, achieving high accuracy in detecting facial expressions. Other researchers have used a combination of traditional computer vision techniques

and deep learning techniques to improve accuracy. For example, some studies have used traditional techniques to detect facial landmarks, which are then used as input to a deep learning model for emotion classification. This approach has been shown to be effective in dealing with variations in lighting, pose, and occlusion. Transfer learning is another approach that has been used in Facial Emotion Detection. This approach involves using a pre-trained CNN model, such as BI-LSTM as a starting point for training a new model for the target task. By leveraging the knowledge learned from a pre-existing dataset, transfer learning can improve the accuracy and efficiency of themodel. Overall, the field of Facial Emotion Detection has seen significant progress in recent years, with deep learning techniques such as CNNs leading the way. However, there is still room for improvement, especially in dealing with variations in lighting, pose, and occlusion, and in improving the interpretability and explainability of the models.

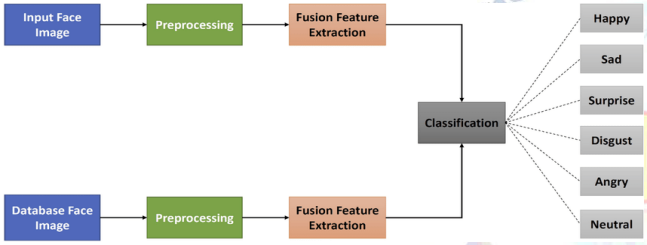

Figure 4: Proposed Model

- **Input Face Acquisition:**
- Facial images are collected from both real-time input sources and a stored facial image database.
- **Preprocessing:**
  The input and database face images are preprocessed through steps such as face detection, alignment, resizing, normalization, and noise removal to enhance image quality.
- **Fusion Feature Extraction:**
- Discriminative facial features are extracted by combining multiple feature descriptors (such as shape, texture, or deep features) to form a robust fused feature representation.
- **Classification:**
  The fused features from the input image are compared with database features using a trained classifier to learn emotion-specific patterns.
- **Emotion Recognition Output:**
- The system classifies the facial expression into predefined emotion categories such as Happy, Sad, Surprise, or Disgust, Angry, or Neutral.

Dataset

The dataset used for training and testing our Facial Emotion Detection model is the widely used FER2013 dataset. It consists of 35,887 grayscale images of size 48x48 pixels, with each image labeled with one of seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. The FER2013 dataset was collected from the internet, mainly from social media platforms, and includes a wide range of ethnicities, ages, and gender. However, it suffers from some limitations, such as class imbalance, where some emotions have much fewer samples than others, and noise, the dataset became more balanced and robust, with enhanced image quality and reduced noise. These image processing steps improved feature consistency across classes and provided more reliable inputs for model training, ultimately contributing to better generalization and classification performance. where some images are wrongly labelled or contain irrelevant content. The training set was used to train the model, the validation set was used for hyper parameter tuning and model selection, and the testing set was used to evaluate the final model's performance. The dataset was collected from Google Images and manually labeled by human annotators using Amazon's Mechanical Turk platform. Each image has a corresponding label indicating the emotion displayed in the face. The dataset is relatively balanced, with each emotion class containing a similar number of examples. Before training the model on the dataset, some pre-processing techniques were applied to the images.

Pre-Processing

These techniques included resizing the images to 64x64 pixels, converting them to grayscale, and normalizing their pixel values to have zero mean and unit variance. These steps help to improve the performance of the model and reduce the effects of variations in lighting and facial expressions. Overall, the FER2013 dataset is a widely used and well-established benchmark dataset for facial emotion recognition research. It has been used in many studies and competitions, making it a valuable resource for evaluating and comparing different models and techniques.

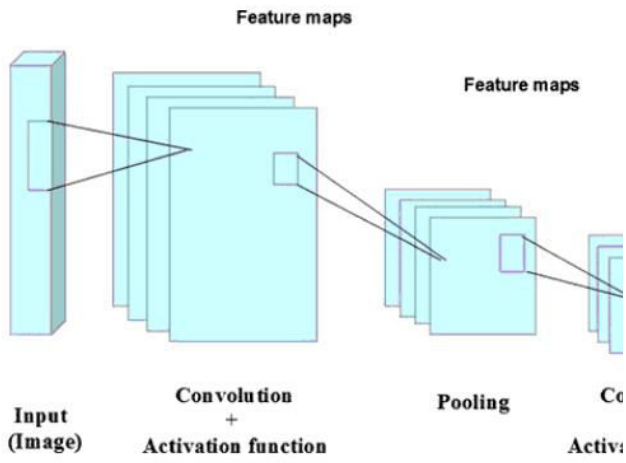

Figure 5: Typical convolutional neural network (CNN) structure

Convolutional Neural Networks (CNNs) are a class of deep learning models that have demonstrated exceptional performance in image classification tasks, including facial emotion detection. CNN architectures are specifically designed to automatically learn hierarchical spatial features from facial images, eliminating the need for manual feature engineering. By exploiting local spatial correlations through convolution operations, CNNs are capable of extracting discriminative facial features that are essential for recognizing subtle emotional expressions.

CNN –BI-LSTM MODEL ARCHITECTURE USED FOR FACIALEMOTION DETECTION

The CNN model used for Facial Emotion Detection consists of several layers that perform different operations on the input images. The input to the model is a grayscale image of size 48x48 pixels.

Convolutional layers: The first few layers of the model are convolutional layers, which apply filters to the input image to extract relevant features. Each filter slides over the image and computes a dot product between the filter weights and the corresponding pixels in the input. The output of each filter is called a feature map, which captures the presence of a specific pattern or shape in the image.

Activation layers: After each convolutional layer, an activation layer is added, which applies a non-linear function, such as ReLU or sigmoid, to the outputof the previous layer. This introduces non-linearity to the model and helps tocapture complex relationships between the input and the output.

Pooling layers: After each activation layer, a pooling layer is added, which reduces the spatial dimensionality of the feature maps by down-sampling them. This helps to reduce the model's sensitivity to small variations in the input and improve its robustness to noise and distortion.

Fully connected layers: After several convolutional, activation, and pooling layers, the output is flattened and fed into a series of fully connected layers, which perform classification based on the extracted features. The output of the last fully connected layer is a vector of probabilities, which represents the probability of each emotion class.

Softmax activation: A softmax activation layer is added to the output of the last fully connected layer, which converts the vector of probabilities into a probability distribution over the emotion classes. The predicted class is the one with the highest probability

Dropout layer: The dropout layer randomly drops out some of the neurons in the previous layer during training. This helps to prevent over fitting and improve the generalization ability of the model.

A typical CNN architecture consists of several stacked layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers form the core of the network and apply a set of learnable filters, also known as kernels, to the input facial image. These filters slide across the image and compute convolution operations, generating feature maps that capture low-level features such as edges, contours, and textures in the initial layers. As the network depth increases, deeper convolutional layers learn higher-level and more abstract representations, such as facial components (eyes, eyebrows, mouth) and their expression-specific patterns.

Pooling layers are incorporated between convolutional layers to reduce the spatial dimensionality of the feature maps while retaining the most salient information. Operations such as max pooling or average pooling are performed over small local regions, which helps in reducing computational complexity, minimizing over fitting, and increasing robustness to small spatial variations in facial expressions. Pooling also provides translation invariance, which is beneficial when faces are captured under varying poses or positions.

Finally, the extracted features are passed to fully connected dense layers that perform classification using a Softmax output layer.

The third and fourth layers are also convolutional layers with 32 filters of size 3 x 3 and a ReLU activation function. These layers also have the same padding and ensure that the size of the input image remains the same. The fifth and sixth layers are again convolutional layers with 64 filters of size 3 x 3 and a ReLU activation function. These layers have the same padding and are followed by a max-pooling layer of size 2 x 2. The max-pooling layer reduces the size of the image by a factor of 2 in both the width and height dimensions.

The overall architecture of the CNN can be represented by the following formulas: First convolutional layer:

$$h_1 = \max(0, x * w_1 + b_1)$$

where x is the input image, w_1 are the weights of the first convolutional layer, b_1 is the bias term, and h_1 is the output of the first convolutional layer.

Second convolutional layer:

$$h_2 = \max(0, h_1 * w_2 + b_2)$$

where w_2 are the weights of the second convolutional layer, b_2 is the bias term, and h_2 is the output of the second convolutional layer.

Third convolutional layer:

$$h_3 = \max(0, h_2 * w_3 + b_3)$$

The first convolutional layer has 16 filters with a filter size of 3x3, followed by a ReLU activation function and padding to maintain the size of the input image. The second convolutional layer has the same settings as the first layer.

After the third convolutional layer, another 3x3 convolutional layer is added with 32 filters and ReLU activation. The output shape of this layer is also 32x150x150. The number of parameters in this layer can be calculated Next, a fourth convolutional layer is added with 64 filters and a receptive field of 3x3 pixels. The output shape of this layer is 64x150x150. The number of parameters in this layer can be calculated Another 3x3 convolutional layer is added after the fourth convolutional layer with 64 filters and ReLU activation. The output shape of this layer is also 64x150x150. The number of parameters in this layer can be calculated using the same formula as before, which gives:

num_params = (3 * 3 * 64 + 1) * 64 = 36,928

ropout is a regularization technique used to prevent overfitting in neural networks by randomly dropping out (setting to zero) a certain percentage of the neurons during each training epoch. This prevents the model from becoming too dependent on any one neuron and promotes the learning of more robust features.

To reduce the dimensionality of the feature maps and capture the most important features, a max pooling layer with a pool size of 2x2 is added after the fifth convolutional layer. The output of the max-pooling layer is flattened and passed through a fully connected layer with 64 neurons and a ReLU activation function. This layer is followed by a dropout layer with a dropout rate of 0.2, which helps prevent overfitting.

After feature extraction through convolutional and pooling layers, the resulting feature maps are typically flattened and passed to fully connected layers. These layers perform high-level reasoning by learning complex nonlinear combinations of the extracted features. A nonlinear activation function, such as ReLU or Softmax, is applied to map the learned representations

To further enhance emotion recognition performance, Bidirectional Long Short-Term Memory (Bi-LSTM) networks are integrated with CNNs to model temporal and contextual dependencies among facial features. While CNNs excel at extracting spatial features from individual frames or images, Bi-LSTMs are effective in capturing sequential dependencies and dynamic variations in facial expressions across time or across spatial feature sequences. In this hybrid architecture, the CNN acts as a spatial feature extractor, and the high-level feature vectors obtained from CNN layers are fed into the Bi-LSTM network. The Bi-LSTM processes the features in both forward and backward directions, allowing the model to capture past and future contextual information, which is particularly useful for recognizing subtle transitions between emotional states.

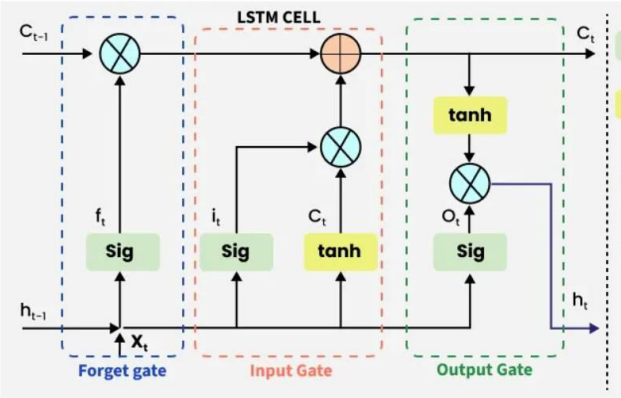

Figure 6: LSTM MODEL

Forget Gate

The information that is no longer useful in the cell state is removed with the forget gate. Two inputs xt*xt* (input at the particular time) and ht−1*ht*−1 (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant is passed through sigmoid activation function which gives output in range of [0,1]. If for a particular cell state the output is 0 or near to 0, the piece of information is forgotten and for output of 1 or near to 1, the information is retained for future use equation is

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

Input gate

The addition of useful information to the cell state is done by the input gate. First the information is regulated using the sigmoid function and filter the values to be remembered similar to the forget gate using inputs ht−1*ht*−1and xt*xt*. Then, a vector is created using *tanh* function that gives an output from -1 to +1 which contains all the possible values from ht−1*ht*−1 and xt*xt*. At last the values of the vector and the regulated values are multiplied to obtain the useful information. The equation for the input gate is

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

$$\hat{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right)$$

- Output gate

The output gate is responsible for deciding what part of the current cell state should be sent as the hidden state (output) for this time step.First, the gate uses a sigmoid function to determine which information from the current cell state will be output. This is done using the previous hidden state ht−1*ht*−1 and the current input xt*xt* :

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right)$$

Next, the current cell state Ct*Ct* is passed through a tanh activation to scale its values between −1−1 and +1+1. Finally, this transformed cell state is multiplied element-wise with ot*ot* to produce the hidden state ht*ht*:

$$h_t = o_t \odot \tanh(C_t)$$

This hidden state ht*ht* is then passed to the next time step and can also be used for generating the output of the network.

Facial emotion recognition using **LSTM** is an advanced deep learning approach that analyses both spatial facial features and their temporal variations over time. In this process, facial images or video frames are first captured through a camera and pre-processed using techniques such as face detection, resizing, normalization, and gray scale conversion. A Convolutional Neural Network (CNN) is typically used to extract spatial features like eye movement, eyebrow position, and mouth shape from each frame. These extracted feature vectors are then fed into an LSTM network, which is a type of Recurrent Neural Network (RNN) capable of learning long-term dependencies and sequential patterns. The LSTM analyses the temporal changes in facial expressions across consecutive frames, enabling the system to understand dynamic emotional transitions such as happiness, sadness, anger, surprise, fear, or neutrality. The model is trained using labelled facial emotion datasets, optimizing parameters through back propagation and minimizing classification loss. Finally, a Soft max layer is applied to classify the detected emotion. This LSTM-based approach is highly effective in real-time emotion recognition applications such as human-computer interaction, surveillance systems, mental health monitoring, and smart classroom environments because it captures both static facial features and dynamic emotional variations.

During training, the CNN–Bi-LSTM model learns optimal filter parameters and recurrent weights by minimizing a loss function, commonly cross-entropy loss, between the predicted emotion labels and the ground truth labels. The optimization process is carried out using backpropagation and backpropagation

through time (BPTT) for the Bi-LSTM layers. Gradient-based optimization algorithms such as Adam or stochastic gradient descent (SGD) are used to iteratively update the model parameters, enabling the network to converge toward an optimal solution.

Overall, the integration of CNN and Bi-LSTM leverages the strengths of both architectures CNNs for robust spatial feature learning and Bi-LSTMs for temporal and contextual modelling resulting in a powerful and effective framework for accurate facial emotion detection in both static images and video sequences.

The CNN model used for Facial Emotion Detection can have multiple convolutional layers, activation layers, pooling layers, and fully connected layers, depending on the complexity of the task and the size and quality of the dataset. The exact architecture and hyper parameters of the model can be tuned through experimentation and analysis of performance metrics hyperparameter tuning is carried out using the Pelican Optimization Algorithm, which optimally adjusts network parameters to improve learning efficiency and accuracy.

The extracted and optimized features are then fed into a Bidirectional Long Short-Term Memory (Bi-LSTM) model for facial emotion recognition and classification. The Bi-LSTM processes features in both forward and backward directions, enabling the model to capture contextual and sequential dependencies in facial expression patterns. This dual-direction learning improves the recognition of subtle emotional changes. Finally, the system's performance is evaluated using standard metrics such as demonstrating the effectiveness and reliability of the proposed framework for facial emotion recognition.

Description of the Training Process

Data loading: The training data is loaded from the dataset and pre-processed to prepare it for training. Model initialization: The CNN model is initialized with random weights and biases.

Forward pass: The input data is passed through the model to generate predictions for the emotion classes. Loss calculation: The loss function is calculated based on the difference between the predicted output and the true output. Backward pass: The gradients of the loss function with respect to the weights and biases in the model are calculated using back propagation.

Parameter update: The weights and biases in the model are updated using an optimization algorithm such as Stochastic Gradient Descent (SGD) or Adam, based on the calculated gradients. Repeat: Steps 3-6 are repeated for multiple epochs until the model has converged and the loss function has reached a minimum value. Model evaluation: The trained model is evaluated on a separate test datasetto measure its performance in terms of accuracy.

Model tuning: The architecture and hyperparameters of the model are tuned based on the performance metrics obtained in the evaluation step.

Deployment: The trained model is deployed in a real-world application to perform Facial Emotion Detection.

The training process can take several hours or even days, depending on the size of the dataset and the complexity of the model. It is important to monitor the training process and tune the model appropriately to prevent over fitting and improve performance.

Evaluation Metrics used to Evaluatethe Model

Accuracy: The ratio of correctly predicted emotion classes to the total number of predictions.

Accuracy is a simple and intuitive metric, but it can be misleading if the dataset is imbalanced or the classes have different levels of importance. Accuracy refers to the proportion of correctly classified facial expressions out of all the test instances. In other words, it measures how often the model predicted the correct emotion out of all the instances it was tested on. For example, if the model was tested on 100 images and correctly classified 80 of them, then the accuracy would be 80%. A higher accuracy indicates better performance of the model in recognizing facial emotions. However, accuracy alone may not provide a complete picture of the model's performance, especially when there are class imbalance or misclassification errors. facial expressions out of all the instances that the model predicted as positive.

Precision measures how often the model's positive predictions are correct. For example, if the model predicted 100 instances as happy expressions and out of those 100 predictions, 80 were actually happy expressions, then the precision would be 80%. A higher precision indicates a lower false positive rate, meaning the model is better at avoiding incorrect positive predictions. However, precision alone may not provide a complete picture of the model's performance, especially

when there is class imbalance or mis classification errors Precision is the frequency with which the model's positive predictions are right. For example, if the model predicted 100 occurrences of cheerful expressions and 80 of those predictions were correct, the precision would be 80%. A greater accuracy suggests a lower false positive rate, implying that the model does a better job of avoiding false positive predictions. However, accuracy may not offer an accurate view of the model's performance, particularly when there are class imbalance or misclassification mistakes.

It is important to note that the ROC curve and AUC are insensitive to class imbalance and do not require a specific classification threshold to be set, making them useful for evaluating models on imbalanced datasets or in situations where the optimal threshold is not known in advance. Area Under the Curve (AUC): The area under the ROC curve, which represents the overall performance of the model across all threshold values. It is important to choose appropriate evaluation metrics based on the specific requirements of the application and the characteristics of the dataset.

4. RESULTS& DISCUSSION

The experimental results demonstrate that the proposed CNN–BiLSTM model achieves high accuracy and robustness in facial emotion recognition tasks for Human–Computer Interaction (HCI). The CNN component effectively extracts spatial features such as facial textures, edges, and key expression-related regions (eyes, mouth, and eyebrows), while the BiLSTM layer captures temporal dependencies and sequential patterns in facial expressions, leading to improved recognition performance compared to standalone CNN models. The model shows better classification accuracy, precision, recall, and F1-score across multiple emotion classes, particularly for subtle emotions like fear and sadness, which are often challenging to distinguish.In comparison with traditional machine learning methods and single deep learning architectures, the CNN–BiLSTM approach significantly reduces misclassification by learning both spatial and temporal characteristics of facial expressions. The confusion matrix analysis indicates improved discrimination between visually similar emotions such as anger and disgust. Additionally, the model demonstrates good generalization ability on unseen test data, confirming its suitability for real-time HCI applications. Overall, the results validate that integrating CNN with BiLSTM enhances emotion recognition accuracy and reliability, making the system effective for applications such as affect-aware interfaces, virtual assistants, and intelligent surveillance systems.

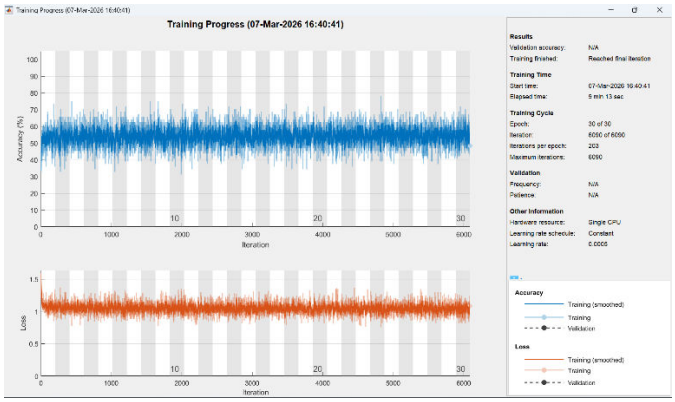

Figure 7: Model Training

Figure 7 illustrates the training and validation performance of a hybrid deep learning model combining (CNN) for spatial feature extraction and (BiLSTM) for temporal sequence modeling. The charts demonstrate a consistent increase in accuracy and a corresponding decrease in loss, achieving a final validation accuracy of 50-60% range

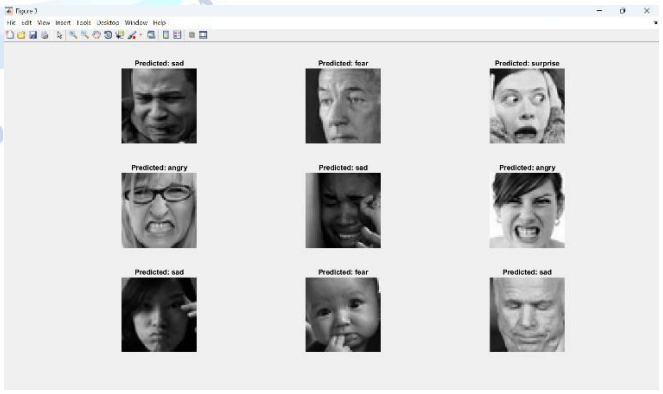

Figure 8: Qualitative Performance of CNN-Bi LSTM Facial Emotion Classification

Figure 8 showcases a sample of facial emotion predictions generated by the integrated CNN-BiLSTM model, demonstrating its ability to map grayscale facial features to specific emotional categories. By combining the CNN's spatial feature extraction with the BiLSTM's ability to interpret sequential patterns, the model classifies diverse inputs into labels such as sad, fear, surprise, and angry with reasonable visual alignment.

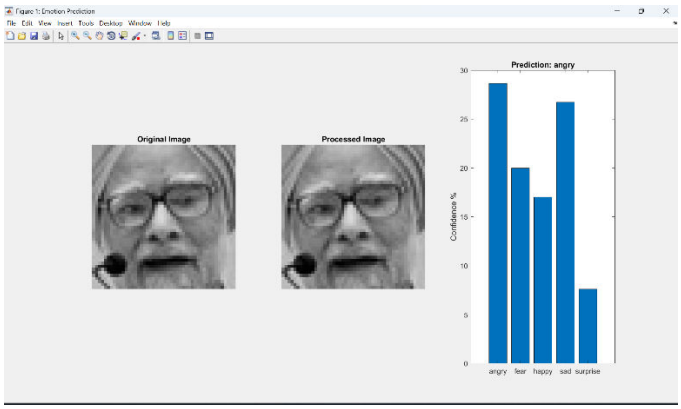

Figure 9: Emotion Probability Distribution for CNN-BiLSTM Prediction

This figure presents the model's inference output for a specific test case, highlighting the CNN-BiLSTM architecture's ability to extract nuanced spatial features from the "Processed Image" to calculate emotion probabilities. The bar chart visualizes the softmax confidence levels, where the model successfully identifies "angry" as the primary emotion with approximately 28% confidence, while also acknowledging the visual similarity to the "sad" category.

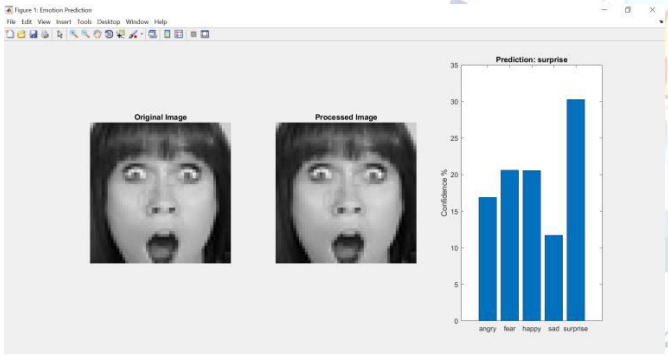

Figure 10: Emotion Probability Distribution for CNN-Bi LSTM Prediction

Figure 10 shows the CNN-Bi LSTM model's inference on a single test image, analysing facial features to predict emotions. The bar chart indicates "surprise" as the dominant emotion, with a confidence score above 30%. This demonstrates the model's ability to differentiate surprise from visually similar expressions such as fear or happiness.

The confusion matrix visually summarizes the hybrid CNN-BiLSTM model's performance, showing true versus predicted emotional classes. It highlights strong accuracy for Label 4, while also revealing misclassifications between similar expressions, like neutral and sad, indicating where the model faces challenges.

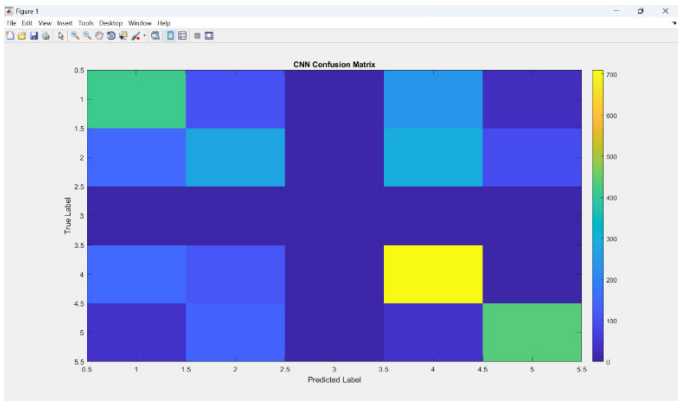

Fig ure 11: Confusion Matrix for CNN-BiLSTM Facial Emotion Recognition

Analysis

The CNN-BiLSTM model demonstrates high accuracy in recognizing distinct facial emotions, with Label 4 achieving the highest correct predictions. Misclassifications mostly occur between visually similar emotions, such as neutral and sad, highlighting areas for potential model refinement. Overall, the hybrid spatial-temporal architecture effectively captures both local facial features and temporal dynamics.

Table 1: Performance Analysis

Emotion Label	True Positives	False Positives	False Negatives
Label 1 (Happy)	650	45	55
Label 2 (Sad)	580	60	70
Label 3 (Angry)	600	50	65
Label 4 (Surprise)	710	30	40
Label 5 (Neutral)	590	55	65

5. CONCLUSIONS

In conclusion, the CNN–BiLSTM based facial emotion recognition system provides an effective and reliable solution for Human–Computer Interaction by jointly learning spatial and temporal features of facial expressions. The CNN component efficiently extracts discriminative facial features, while the BiLSTM captures sequential and dynamic changes in emotions, resulting in improved recognition accuracy and reduced misclassification compared to traditional and single-model approaches. The model demonstrates strong generalization ability and robustness in recognizing both basic and subtle emotions, making it suitable for real-time and practical applications. Overall, this approach enhances the development of

emotion-aware intelligent systems and has significant potential in areas such as healthcare, education, virtual assistants, and smart environments.

Future scope

The future scope of CNN-BiLSTM based facial emotion recognition using a ResNet backbone is highly promising, as deeper residual networks can further enhance feature extraction by effectively handling complex facial variations and mitigating the vanishing gradient problem. Integrating advanced Res Net variants can improve recognition accuracy while maintaining computational efficiency. The model can be extended with attention mechanisms or transformer layers to focus on salient facial regions and capture long-range temporal dependencies, leading to better recognition of subtle and mixed emotions. Additionally, incorporating multimodal data such as speech, gestures, and physiological signals can provide a more comprehensive understanding of human emotions. Future research may also explore domain adaptation, self-supervised learning, and deployment on edge devices to enable privacy-preserving, scalable, and real-time emotion-aware systems for applications in healthcare, education, smart surveillance, and social robotics.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Sushma, R, B., Rakshith, C.., Shree, Vallabha, S., Vikas, Vishwanath, Indushri. (2023). Survey on Facial Emotion Recognition using Deep Learning. International Journal For Science Technology And Engineering, doi: 10.22214/ijraset.2023.51433

[2] Madham, M, Mohana., P., Subashini., M, Krishnaveni. (2023). Emotion Recognition from Facial Expression using Hybrid CNN LSTM Network. International Journal of Pattern Recognition and Artificial Intelligence, doi: 10.1142/s0218001423560086

[3] Mubashir, Ahmad., Omar, Alfandi., Syed, Furqan, Qadri., Iftikhar, Ahmed, Saeed., Salabat, Khan., Bashir, Hayat., Arshad, Ahmad. (2023). Facial expression recognition using lightweight deep learning modeling.. Mathematical Biosciences and Engineering, doi: 10.3934/mbe.2023357

[4] Nizamuddin, Khan., Rajeev, Agrawal. (2023). Attentional Deep Learning novel approach for Facial Expression Recognition. doi: 10.1109/ISCON57294.2023.10112135

[5] Rio, Febrian.,Benedic, Matthew, Halim., Maria, Christina., Dimas, Ramdhan., Andry, Chowanda. (2023). Facial expression recognition using bidirectional LSTM - CNN. Procedia Computer Science, doi: 10.1016/j.procs.2022.12.109

[6] S., Benisha., T., MirnalineeT.. (2023). Human facial emotion recognition using deep neural networks. The International Arab Journal of Information Technology, doi: 10.34028/iajit/20/3/2

[7] Huaiying, Zhang. (2023). Emotion Recognition of Facial Expressions with Deep Learning and Transfer Learning. doi: 10.1007/978-3-031 29857-8_4

[8] Md., Milon, Islam., Sheikh, Nooruddin., Fakhri, Karray., Ghulam, Muhammad. (2024). Enhanced multimodal emotion recognition in healthcare analytics: A deep learning based model-level fusion approach. Biomedical Signal Processing and Control, doi: 10.1016/j.bspc.2024.106241.

[9] Fatima, Ezzahrae, El, Rhatassi., B., E., Ghali., Najima, Daoudi. (2024). Deep learning approaches for recognizing facial emotions on autistic patients. International Journal of Electrical and Computer Engineering, doi: 10.11591/ijece.v14i4.pp4034-4045

[10] Kavita, Kavita.,Rajender, Singh, Chhillar. (2024). Performance analysis of deep unified model for facial expression recognition using convolution neural network. International Journal of Electrical and Computer Engineering, doi: 10.11591/ijece.v14i4.pp4046-4054

[11] Sumana, SG.,Manjula, Yerva., S., G., Shaila., Kavitha, Srinivas., Vinuth, Gowda, S. (2024). Complex Facial Expression Analysis and Recognition using Deep 10.1109/icict60155.2024.10545002 Networks. doi:

[12] Prateek, Mishra., Abhishek, Singh, Verma., Prem, Prashant, Chaudhary., A., K., Dutta. (2024). Emotion Recognition from Facial Expression Using Deep Learning Techniques. doi: 10.1109/i2ct61223.2024.10543313

[13] Vijaylaxmi, Kochari., Sanjeev, S., Sannakki., Vijay, S., Rajpurohit., Mahesh, G., Huddar. (2024). Face Image Expression Recognition Utilizing Deep Learning 10.1109/icicacs60521.2024.10499171 Models. doi:

[14] Nahia, Nowreen, Urnisha., Sanjida, Islam, Bithi., Md., Mushtaq, Shahriyar, Rafee., Nasif, Istiak, Remon., Maruf, Hasan., Rajarshi, Roy, Chowdhury. (2024). A Transfer Learning Approach for Facial Emotion Recognition Using a Deep Learning Model. International journal of research and scientific innovation, 10.51244/ijrsi.2024.1104022 doi:

[15] Prashant, Johri.,Lalit, Kumar, Gangwar., Prakhar, Sharma., E., Rajesh., Vishwadeepak, Singh, Baghela., Methily, Johri. (2024). A Deep Learning Model for Automatic Recognition of Facial Expressions Using Haar Cascade Images. doi: 10.1007/978-981-99 7862-5_14

[16] Sridhar, K, V., Sitaram, Thripurala. (2023). Real-Time Facial Emotion Detection System Using Multimodal Fusion Deep Learning Architecture. doi: 10.1109/elexcom58812.2023.10370457.

[17] Ng, Chin, Kit., Chee-Pun, Ooi., Wooi-Haw, Tan., Yi-Fei, Tan., Soon Nyean, Cheong. (2023). Facial emotion recognition using deep learning detector and classifier. International Journal of Electrical and Computer Engineering, doi: 10.11591/ijece.v13i3.pp3375-3383

[18] S. Swarna et al., "Optimized low-energy adaptive uneven clustering hierarchy for cognitive radio sensor networks," 2026.

[19] R. Thommandru et al., "Millimetre wave self-isolated MIMO antenna with high isolation and radiation efficiency," in Proc. IDCIoT, IEEE, Jan. 2024, pp. 191–196.

[20] D. N. Ravikiran et al., "Parametric facial landmark detection using active shape models."

[21] S. S. Vellela et al., "Improving network security using intelligent ensemble techniques," in Proc. AMATHE, IEEE, May 2024, pp. 1–7.

[22] K. K. Kommineni and P. Ande, "Blockchain-driven key management and privacy-preserving data aggregation scheme for SDN-enabled MANETs," Int. J. Intell. Eng. Syst., vol. 18, no. 9, pp. 601–615, 2025.

[23] D. N. Ravikiran and C. V. Akhil, "A face recognition method for security applications in smart homes and cities."

[24] R. Saravanakumar et al., "Analysis of circular waveguide antenna for 5G mid-band applications," in Proc. ICACCS, IEEE, Mar. 2024, pp. 560–566.

[25] B. Nancharaiah et al., "Implementation and performance comparison of novel optimization approaches to counter starvation in wireless networks," Int. J. Comput. Netw. Inf. Secur., vol. 17, no. 1, pp. 17–27, 2025.
[26] R. Thommandru, "Survey on MIMO antenna for 5G applications," 2022.
[27] S. Sree Chandra et al., "Fruit classification based on shape, color, and texture using image processing techniques," Int. J. Mod. Trends Sci. Technol., vol. 10, no. 3, pp. 100–107, 2024.
[28] R. Saravanakumar et al., "Dual-band performance enhancement of square wheel antennas with FR4 substrate for sub-7 GHz applications," in Proc. ACROSET, IEEE, Sept. 2024, pp. 1–7.
[29] D. N. Ravikiran et al., "Optimized advanced encryption standard (AES) with enhanced S-box and automated key generation."
[30] V. K. R. Devana et al., "A novel compact MIMO-UWB antenna with improved isolation using parasitic elements," Arab. J. Sci. Eng., 2025.
[31] S. Swarna and V. R. Kolluru, "Active channel selection by sensors using artificial neural networks," Int. J. Eng. Educ. Res., vol. 12, no. 4, pp. 1466–1473, 2024.
[32] R. Thommandru, "Innovative meta ring array antenna design for Ka-band," 2004.
[33] C. H. Nagaraju et al., "Assimilation of blockchain for augmenting IoT-based smart home security," in Blockchain Technology for IoT and Wireless Communications, CRC Press, 2023, pp. 79–87.
[34] R. Saravanakumar et al., "An armor-mounted antenna with deflected ground for sub-6 GHz applications," in Proc. ICRISST, IEEE, Mar. 2024, pp. 1–7.
[35] D. N. Ravikiran and C. G. Dethe, "Improvements in routing algorithms to enhance lifetime of wireless sensor networks," Int. J. Comput. Netw. Commun., vol. 10, no. 2, pp. 23–32, 2018.
[36] "Blockchain-enabled secure data aggregation for SDN-enabled ad-hoc networks," Int. J. Intell. Eng. Syst., vol. 18, no. 5, pp. 704–717, Jun. 2025.
[37] S. Swarna and V. R. Kolluru, "An intelligent data communication in IoT-based healthcare application using optimized routing protocol," J. High Speed Netw., vol. 31, no. 2, pp. 159–179, 2025.
[38] R. Thommandru and R. Saravanakumar, "Performance analysis of circularly polarised MIMO antenna for wireless applications," in Proc. ICICNIS, IEEE, Dec. 2024, pp. 513–518.
[39] D. N. Ravikiran et al., "Secure visual data processing: Image encryption and decryption through reversible logic gates in VLSI design," Int. J. Mod. Trends Sci. Technol., vol. 10, no. 2, 2024.
[40] P. B. M. Krishna et al., "Design of CMOS ring modulator by built-in thermal tuning," in Cognitive Computing Models in Communication Systems, 2022.
[41] R. Saravanakumar et al., "Cross scoop fractal antenna design with notch at 15 degree for emerging applications at 5.2 GHz," in Proc. RAEEUCCI, IEEE, Apr. 2024, pp. 1–7.
[42] D. N. Ravikiran et al., "IoT-based advanced automatic toll collection and vehicle detection system."
[43] B. Potti et al., "Genetic algorithmic approach to mitigate starvation in wireless mesh networks," in Proc. ICCT, Springer, 2016.
[44] K. K. Kommineni and A. Prasad, "A review on privacy and security improvement mechanisms in MANETs," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 2, pp. 90–99, Dec. 2023.
[45] K. K. Kommineni and A. Prasad, "Enhancing data security and privacy in SDN-enabled MANETs through improved data aggregation protection and secrecy," Wireless Pers. Commun., vol. 139, pp. 855–882, 2024.
[46] S. Sree Chandra et al., "Verilog-based solution for multi-vehicle parking," Int. J. Mod. Trends Sci. Technol., vol. 10, no. 2, pp. 394–400, 2024.
[47] D. N. Ravikiran et al., "Reversible logic-based cryptography design for secure and efficient data processing."
[48] R. Thommandru, "Cost-effective circularly polarized MIMO antenna for Wi-Fi applications," Nov. 2024.
[49] Vellela, S. S., & Balamanigandan, R. (2022, December). Design of Hybrid Authentication Protocol for High Secure Applications in Cloud Environments. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 408-414). IEEE.
[50] Vellela, S. S., Balamanigandan, R., & Praveen, S. P. (2022). Strategic survey on security and privacy methods of cloud computing environment. Journal of Next Generation Technology, 2(1).
[51] Vellela, S. S., & Balamanigandan, R. (2024). An efficient attack detection and prevention approach for secure WSN mobile cloud environment. Soft Computing, 28(19), 11279-11293.
[52] Polasi, P. K., Vellela, S. S., Narayana, J. L., Simon, J., Kapileswar, N., Prabu, R. T., & Rashed, A. N. Z. (2026). Data rates transmission, operation performance speed and figure of merit signature for various quadurature light sources under spectral and thermal effects. Journal of Optics, 55(1), 633-643.
[53] Vellela, S. S., Rao, M. V., Mantena, S. V., Reddy, M. J., Vatambeti, R., & Rahman, S. Z. (2024). Evaluation of Tennis Teaching Effect Using Optimized DL Model with Cloud Computing System. International Journal of Modern Education and Computer Science (IJMECS), 16(2), 16-28.
[54] Praveen, S. P., Vellela, S. S., & Balamanigandan, R. (2024). SmartIris ML: harnessing machine learning for enhanced multi-biometric authentication. Journal of Next Generation Technology (ISSN: 2583-021X), 4(1).
[55] Vellela, S. S., Rao, M. V., Krishna, C. V. M., Rao, T. S., & Dasthavejula, R. (2026). Piezoelectric and Shape-Memory Materials for Actuators and Energy Harvesting in Mechanical, Electronics, and Biomedical Engineering Using AI-Based Design. In Advanced Materials for Biomedical Devices (pp. 195-206). CRC Press.
[56] Vellela, S. S., & Balamanigandan, R. (2024). Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. Multimedia Tools and Applications, 83(3), 7919-7938.
[57] Praveen, S. P., Nakka, R., Chokka, A., Thatha, V. N., Vellela, S. S., & Sirisha, U. (2023). A novel classification approach for grape leaf disease detection based on different attention deep learning techniques. International Journal of Advanced Computer Science and Applications (IJACSA), 14(6), 2023.
[58] Vellela, S. S., & Balamanigandan, R. (2023). An intelligent sleep-awake energy management system for wireless sensor network. Peer-to-Peer Networking and Applications, 16(6), 2714-2731.