A RealTime Chat Application Using a Deep Learning Based Facial Emotion Recognition System For Effective Communication

Nikhil R Nambiar [1]
Student, Department of Computer Science Engineering,
T. John Institute of Technology,
Bengaluru, Karnataka,
nikhil.4002.50.82@gmail.com

Vilas R Naik [2]
Student, Department of Computer Science Engineering,
T. John Institute of Technology,
Bengaluru, Karnataka,
vilasnaik050804@gmail.com

Shivam Singh [3]
Student, Department of Computer Science Engineering,
T. John Institute of Technology,
Bengaluru, Karnataka,
1singhshivam2003@gmail.com

Naveen Kumar S [4]
Student, Department of Computer Science Engineering,
T. John Institute of Technology,
Bengaluru, Karnataka,
itsmenaveen404@gmail.com

***Abstract* - In order to improve the emotional expressiveness of text based communication platforms, this paper introduces a realtime facial emotion recognition system. The suggested system computes geometric features like mouth width, eye openness, eyebrow slant, and facial symmetry using MediaPipe's Face Mesh model, which extracts 468 facial landmarks. The CK+(Cohn-Kanade Plus) dataset is used to train a deep learning model for emotion classification using these normalized landmark based features. The model successfully identifies the seven main human emotions which are happiness, surprise, anger, contempt, disgust, sadness, and fear, and has an accuracy rate of 85.28%. When this system is integrated into any chat application, it automatically detects and displays the user's emotional state during live conversations as corresponding emojis. By encouraging more organic, sympathetic, and context aware interactions and preserving lightweight performance appropriate for realtime deployment, this method closes the emotional gap in digital communication.**

Keywords - Chat applications, CK+ dataset, deep learning, emotion recognition, facial landmarks, MediaPipe, neural networks, realtime systems.

I. INTRODUCTION

Online communication has become a vital component of everyday life in the current digital era. People can easily stay connected around the world thanks to chat apps, social media, and instant messaging platforms. Notwithstanding these benefits, digital communication is devoid of emotional expressions, a crucial element of human interaction. Emotions are naturally expressed in face-to-face communication through tone of voice, gestures, and facial expressions; these elements are essential for determining empathy and intent. On the other hand, text based communication eliminates these indicators, which causes misunderstandings and a rift in users' emotions. A primary goal of human-computer interaction(HCI), this limitation drives the need for systems that can bring emotional depth back to digital interaction.

The study and development of interactive systems that enable efficient communication between people and machines is known as human computer interaction, or HCI. It places a strong emphasis on developing intelligent, flexible, and behavior responsive technology. A key component of HCI is emotion recognition, which gives computers the ability to sense and understand emotional states, enhancing the naturalness and engagement of digital communication. Because facial expressions directly reflect psychological states, facial emotion recognition(FER) in particular is one of the most accurate techniques for interpreting human emotions. Traditional text based communication can be changed into emotionally intelligent interaction by integrating FER into chat systems.

The suggested system offers a realtime facial emotion recognition function that can be added as an add-on module to chat programs. In order to compute and normalize geometric and distance based features like eye openness, eyebrow slant, and mouth aspect ratio, 468 facial landmarks are taken from the user's face using a webcam and MediaPipe's Face Mesh. The Cohn Kanade Plus(CK+) dataset is used to train a deep learning model that uses these features to classify seven different emotions which are happiness, surprise, anger, contempt, disgust, sadness, and fear.

When incorporated into a chat program, the system recognizes the user's emotion during a message and instantly displays it as an emoji next to the message. This bridges the gap between human emotions and computer mediated communication by improving emotional awareness, user engagement, and empathy in digital conversations.

II. LITERATURE SURVEY

For more than 20 years, facial emotion recognition(FER) has been a prominent field of study in affective computing and computer vision. It seeks to automatically identify and categorize human emotions from facial expressions, offering vital input for systems that are emotionally intelligent and perceptive. By facilitating automatic feature extraction and hierarchical representation learning from massive facial datasets, recent developments in deep learning have significantly enhanced FER performance. Convolutional neural networks(CNNs) and attention based models have emerged as the most popular approaches for emotion analysis in recent years, according to Li and Deng's [1] thorough review of deep learning approaches for facial expression recognition.

Deep neural networks for reliable emotion detection in a variety of poses and lighting conditions were investigated in earlier studies, such as those by Mollahosseini et al. [2]. However, the deployment of these image based models in realtime systems is limited because they frequently necessitate substantial computation and sizable annotated datasets. Recent research has investigated geometric and landmark based methods that employ particular facial feature points rather than complete images in order to overcome these limitations. Patel and Mehta [4] showed how realtime emotion detection with a much lower computational overhead could be accomplished by combining machine learning techniques with facial landmarks.

With its extensive collection of labeled facial expressions, the Cohn-Kanade Plus(CK+) dataset, first presented by Lucey et al. [3], has emerged as a benchmark dataset for FER model evaluation. The integration of FER into Human Computer Interaction(HCI) applications was further highlighted by Verma and Gupta [5], who provided examples of how emotion aware systems can improve user engagement and communication efficacy. High fidelity landmark detection is now more accessible for realtime emotion analysis thanks to the creation of lightweight frameworks like Google Research's Face Mesh for MediaPipe [7].

A balanced strategy that guarantees both computational efficiency and accuracy for realtime chat based emotion recognition is offered in this study by utilizing the CK+ dataset [3], MediaPipe landmarks [7], and a deep learning framework [6].

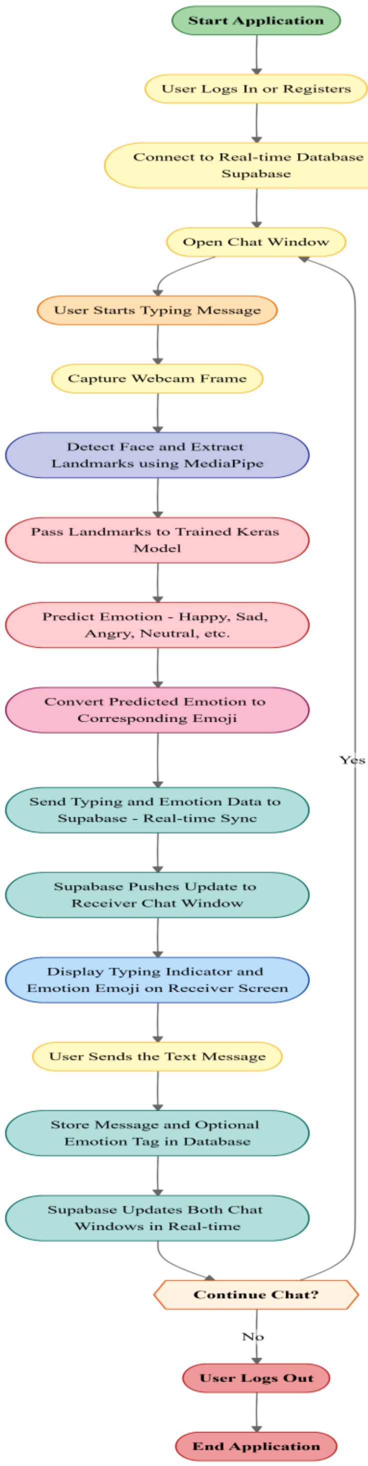
III. METHODOLOGY

In order to identify and show user emotions in real time within a chat application, the suggested facial emotion recognition system combines deep learning and computer vision techniques. Facial landmark extraction, feature computation, emotion classification, and realtime integration into the chat interface are some of the stages that make up the system. Because the entire process is modular, it can be easily implemented on any communication platform.

A. Overview of the System

A realtime chat application built with Supabase as the database for instant data synchronization has a modular extension in the form of the suggested architecture. Because the module functions independently of the main chat logic, the emotion recognition pipeline can run concurrently without compromising latency or message delivery. The architecture is made up of three main parts: (1)Frontend capture module, which connects to the user's webcam and streams frames to the CPU. (2)Emotion detection engine, which uses MediaPipe to extract facial landmarks and feeds the features into a deep learning model for emotion prediction, and (3)Chat interface integration layer, which associates the detected emotion with a corresponding emoji that is dynamically displayed next to the user's message.

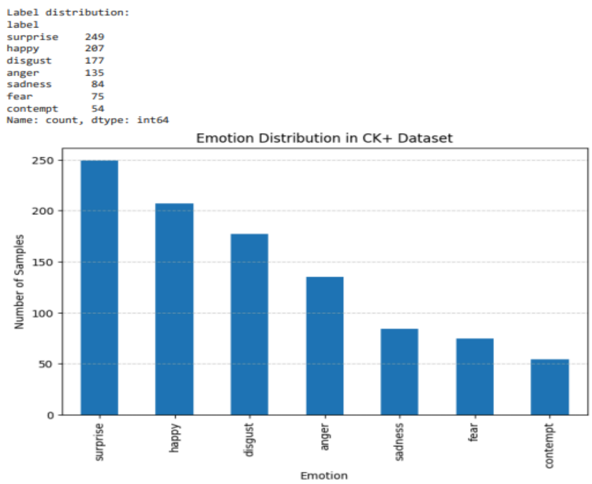
The emotion recognition feature can be implemented on desktop, mobile, and web platforms thanks to its modular architecture. In order to maintain low latency and realtime responsiveness, the system's design guarantees asynchronous processing, which means that emotion detection operates in parallel with chat communication. Additionally, the system is computationally efficient due to the use of lightweight geometric features instead of pixel based inputs.



B. Description of the Dataset(CK+ dataset)

The suggested deep learning model was trained and assessed using the Cohn-Kanade Plus(CK+) dataset [3]. CK+, which features high resolution photos depicting a range of emotional states, is a recognized standard in the field of facial expression recognition. Annotated with emotion labels like happiness, surprise, anger, contempt, disgust, sadness, and fear, it consists of 593 video sequences from 123 subjects. Action Units, which are used to record minute facial muscle movements necessary for precise emotion classification, are also included in the dataset.

From a neutral expression to a peak emotional state, each CK+ sequence advances. The pictures were preprocessed by being resized for consistency and converted to grayscale. The MediaPipe's Face Mesh was then used to extract landmarks from each image, yielding 468 unique coordinate points per face. These landmarks formed the basis for classifying emotions by providing the input for the computation of geometric features.



C. Detecting Facial Landmarks with MediaPipe

In order to extract the geometric features that characterize emotional expressions, facial landmark detection is an important step. The suggested system makes use of MediaPipe's Face Mesh, a realtime framework that offers 468 high fidelity 3D facial landmarks and was created by Google Research [7]. These landmarks allow for in depth geometric analysis by precisely capturing the structure of facial features like the mouth, jawline, eyes, and eyebrows.

The webcam stream is processed frame by frame as the user engages with the chat application. To meet MediaPipe's processing needs, the captured image is first converted from BGR to RGB format. Then, even in the presence of changes in lighting, angle, and facial orientation, the Face Mesh model recognizes facial landmarks. In order to preserve scale invariance across various users and distances from the camera, these coordinates are normalized.

D. Extraction of Features

Following the detection of facial landmarks, geometric and ratio based features that accurately depict emotional states are extracted. The system calculates different aspect ratios and Euclidean distances between important facial points that correspond to important areas such as jaw, eyes, mouth, and eyebrows. To capture changes in facial muscle positions that correspond to various emotions, for example, the mouth aspect ratio(MAR) and eye aspect ratio(EAR) are computed.

Parameters like mouth width, mouth height, eye openness, jaw drop, symmetry of mouth corners, eyebrow slant, eyebrow raise, and nose-mouth distance are among the features that were extracted. To ensure uniformity across various face sizes and orientations, these 14 features are normalized by dividing each by the inter-ocular distance, or the distance between the outer corners of the eyes.

A condensed and discriminative feature set for emotion recognition is offered by this geometric representation. The method drastically lowers dimensionality and gets rid of unnecessary data by avoiding direct pixel-level image processing. The neural network is then fed the resultant feature vector in order to classify it. This design is perfect for real-time applications where responsiveness is essential because it strikes a balance between interpretability and computational efficiency.

E. Architecture of the Deep Learning Model

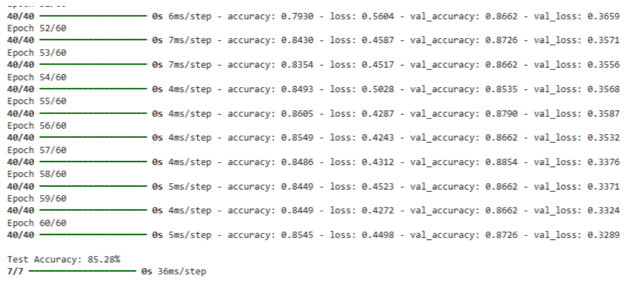
A fully connected feed forward neural network constructed with TensorFlow and Keras is used to implement the deep learning model. The 14 extracted features are represented by the input layer of the architecture, which is followed by two hidden layers and an output layer. To avoid overfitting, a dropout layer is placed after the first hidden layer, which has 64 neurons with Relu (Rectified Linear Unit) activation. To further improve generalization, a dropout layer comes after the second hidden layer, which has 32 neurons.

The output layer creates probability distributions for each of the seven emotion classes using a softmax activation function. The sparse categorical cross entropy loss function is used to efficiently handle multiclass classification, and the model is trained using the Adam optimizer, which offers adaptive learning rates for quicker convergence. Class weighting was utilized to address data imbalance across emotion categories during the 60 epochs of training, which had a batch size of 16. Accuracy and efficiency are balanced in this architecture. With an average test accuracy of 85.28%, the trained model proved its dependability in identifying emotional states using basic geometric features as opposed to computationally costly image representations.

F. Metrics for Training and Assessment

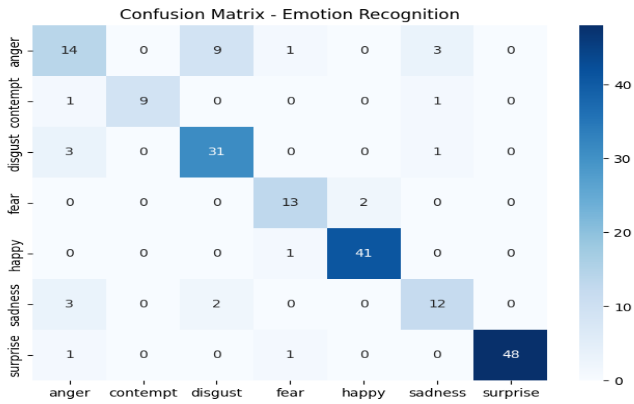
To ensure stratification across all emotion categories, the model was trained using an 80:20 train test split. The training dataset was further split into training and validation sets in an 80:20 ratio to improve generalization. StandardScaler, which standardizes features by taking the mean and scaling them to unit variance, was used to apply data normalization. To reduce bias toward dominant emotion classes, class weights were calculated.

A confusion matrix was used to visualize the classification results, and important metrics like accuracy, precision, recall, and F1 score were used to assess the model's performance. The confusion matrix showed that complex emotions like contempt and disgust were moderately accurately detected, while happiness and surprise were detected with high accuracy. To track training progress and identify possible overfitting, the accuracy and loss curves were plotted.



```
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - accuracy: 0.7930 - loss: 0.5604 - val_accuracy: 0.8662 - val_loss: 0.3659
Epoch 52/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.8430 - loss: 0.4587 - val_accuracy: 0.8726 - val_loss: 0.3571
Epoch 53/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step - accuracy: 0.8354 - loss: 0.4517 - val_accuracy: 0.8662 - val_loss: 0.3556
Epoch 54/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.8493 - loss: 0.5028 - val_accuracy: 0.8535 - val_loss: 0.3568
Epoch 55/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.8605 - loss: 0.4287 - val_accuracy: 0.8790 - val_loss: 0.3587
Epoch 56/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.8549 - loss: 0.4243 - val_accuracy: 0.8662 - val_loss: 0.3532
Epoch 57/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.8486 - loss: 0.4312 - val_accuracy: 0.8854 - val_loss: 0.3376
Epoch 58/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.8449 - loss: 0.4523 - val_accuracy: 0.8662 - val_loss: 0.3371
Epoch 59/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 4ms/step - accuracy: 0.8449 - loss: 0.4272 - val_accuracy: 0.8662 - val_loss: 0.3324
Epoch 60/60
40/40 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.8545 - loss: 0.4498 - val_accuracy: 0.8726 - val_loss: 0.3289

Test Accuracy: 85.28%
7/7 ━━━━━━━━━━━━━━━━━━━━ 0s 36ms/step
```

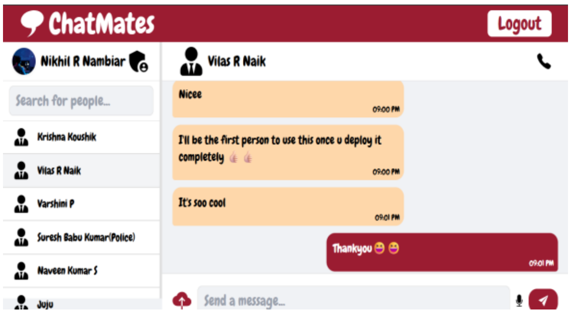


G. Integrating the model into Chat Application

Integrating the deep learning model that has been trained into the realtime chat system is the last step. The chat application manages user authentication, message storage, and data streaming. It was developed with the help of contemporary web technologies and is powered by Supabase for database synchronization. A frontend extension that uses browser API's to communicate with the user's webcam is used to implement the emotion recognition feature.

The system records live frames and carries out facial landmark detection locally while a user is typing or interacting with the interface during a chat session. The trained model receives the extracted feature vector and makes a realtime emotion prediction. Next, an appropriate emoji is mapped to the predicted emotion and dynamically displayed next to the outgoing message. The emoji enhances emotional awareness by appearing as an expressive visual cue on the receiver's interface.

The emotion detection process operates asynchronously without interfering with message flow thanks to this smooth integration. By enabling users to sense one another's emotional states during conversations, the method fosters emotionally intelligent communication and increases user empathy and engagement in digital communication.



IV. Experimental Results and Analysis

The Cohn-Kanade Plus(CK+) dataset, which consists of pictures depicting seven different emotions - happiness, surprise, anger, contempt, disgust, sadness, and fear were used to train and assess the suggested facial emotion recognition system. After being extracted from MediaPipe's Face Mesh, the geometric features were combined into a feature dataset and split 80:20 between training and testing sets. Z-score normalization was used to standardize each feature in order to guarantee consistent model performance and uniform scaling. During training, class weights were used to reduce imbalances between emotion categories, especially for emotions like fear and contempt that have fewer samples.

With a test accuracy of 85.28%, the deep learning model showed a high degree of generalization across a variety of subjects and emotion types. With training and validation loss steadily declining over 60 epochs, the model converged smoothly during training, suggesting efficient learning and little overfitting. The confusion matrix showed that while visually subtle emotions like disgust and contempt occasionally showed misclassifications due to overlapping facial features, emotions like happiness and surprise were classified with high precision. The classification report also verified that most categories had balanced recall and precision.

Heatmaps and training curves plotted with Matplotlib and Seaborn were used to visually assess performance. Effective class discrimination and model stability were validated by these visualizations. The chat application's realtime testing also showed low latency, with an average emotion detection and display delay of less than 300 milliseconds, guaranteeing seamless integration during live chat.

Overall, the experimental findings confirm that a compact neural network combined with geometric facial features offers a dependable, portable, and effective realtime facial emotion recognition solution that can be implemented in interactive chat systems.

```
Classification Report:
              precision    recall  f1-score   support

       anger       0.64      0.52      0.57        27
    contempt       1.00      0.82      0.90        11
     disgust       0.74      0.89      0.81        35
        fear       0.81      0.87      0.84        15
       happy       0.95      0.98      0.96        42
     sadness       0.71      0.71      0.71        17
    surprise       1.00      0.96      0.98        50

    accuracy                           0.85       197
   macro avg       0.84      0.82      0.82       197
weighted avg       0.85      0.85      0.85       197
```

😮 - Surprise

V. Conclusion and Future Scope

In order to improve emotional expressiveness during digital communication, this paper presented a realtime facial emotion recognition system integrated into a chat application. A deep learning model trained on the CK+ dataset was used to process the geometric features that were taken from 468 facial landmarks using MediaPipe's Face Mesh. The model successfully identified seven main emotions with an accuracy of 85.28%. The system bridges the emotional divide in text based communication by adding an empathic layer to standard chat interactions by mapping these emotions to emojis.

Voice and text sentiment analysis can be added to the system in the future to enable multimodal emotion detection. Performance and scalability can be improved across a range of user environments with additional optimization using model pruning, lighting adaptation, and cross dataset training.

VI. Acknowledgment

We would like to sincerely thank our college T. John Institute of Technology's Department of Computer Science and Engineering for their unwavering support, direction, and encouragement during this research. We would especially like to thank Dr. Sudaroli, our project guide, and other faculty members for their insightful comments on system integration and facial emotion recognition.

VII. References

[1] S. Li and W. Deng, *"Deep facial expression recognition: A survey"* IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195–1215, 2022.

[2] A. Mollahosseini, D. Chan, and M. H. Mahoor, *"Going deeper in facial expression recognition using deep neural networks"* in Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1–10.

[3] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, *"The extended Cohn-Kanade (CK+) dataset: A complete dataset for action unit and emotion-specified expression"* in Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 2010, pp. 94–101.

[4] M. Patel and S. P. Mehta, *"Real-time emotion detection using facial landmarks and machine learning techniques"* International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 7, pp. 530–538, 2021

[5] G. Verma and K. Gupta, *"Integrating emotion recognition systems into human-computer interaction applications"* IEEE Access, vol. 9, pp. 145013–145025, 2021.

[6] F. Chollet, *Deep Learning with Python, 2nd ed*. New York, NY, USA: Manning Publications, 2021.

[7] Google Research, *"MediaPipe's Face Mesh: High-fidelity facial landmark detection"* Google AI Blog, 2020.