Emotion Detection using Enhanced CNN Model Through Virtual Assistance

B. Yasaswini[1], Kamaluru Mahammad[2], T. Supriya[3], P Venkatakrishna[4] and V Sai Shanth Kumar[5]

{ saiyasaswinib@gmail.com[1], mahammadk@mits.ac.in[2], asupriya499@gmail.com[3], ponnavenkatakrishna02@gmail.com[4], saishanth094@gmail.com[5] }

UG Student, Department of Artificial Intelligence, Madanapalle Institute of Technology & Science, Madanapalle 517325, Andhra Pradesh, India[1, 3, 4, 5]

Assistant Professor, Department of Artificial Intelligence, Madanapalle Institute of Technology & Science, Madanapalle 517325, Andhra Pradesh, India[2]

Abstract. A key aspect of human-computer interaction is emotion detection, where systems can detect and respond to the emotional state of users. A better CNN model for emotion detection from webcam-captured facial expressions is proposed in this paper. Our approach employs a proprietary Convolutional Neural Network (CNN) trained on the FER2013 dataset, unlike traditional models based on pretrained architectures. The model incorporates a virtual assistant that engages with people according to the emotions it detects and divides emotions into seven categories. Additionally, the system uses the Google Maps to recommend the closest psychiatrist in situations involving negative emotions like fear, sad or anger. Real-time emotional support and improved emotion recognition accuracy are the goals of the suggested system.

Keywords: Emotion detection, Convolutional Neural Network (CNN), Virtual Assistance, Facial Expression Recognition, Google maps

1 Introduction

An important domain of artificial intelligence is emotion recognition, so machines can easily interpret and respond to human emotions. Recently, emotion detection technology has attracted much attention owing to the growing need of human-computer interfaces systems and mental health care. Using emotions and face as the input feature this paper, "An Intelligent System of Emotion Detection Utilizing an Enhanced CNN Model with Virtual Assistant," aims to build a smart system to detect the emotions and provide a customized virtual assistant. Use Cases for Emotion-based AI Emotion-based AI has wide ranging applications in social media, healthcare, education, customer support, and many more. But many of today's systems continue to rely on decades-old methods that aren't very accurate or responsive in real time. In this project, a proprietary Convolutional Neural Network (CNN) model trained on the FER2013 dataset is applied to enhance emotion recognition for the classification of facial expression into seven emotions: Happy, Surprise, Neutral, Disgust, Sad, Anger, and Fear.

Its capability to recognize emotions in realtime through a webcam allows it to monitor users' emotional states dynamically. Recognized emotion will be the trigger condition and the chatbot chats with the user giving personalized encouragement and sympathy based on the user's mood. Personalised content delivery has been found to enhance user engagement and well-being in the context of recommender systems of mental health platforms. Inspired by these findings, our app leverages Google Maps to give local psychiatrists as a recommendation if emotions such as sadness, anger, or fear are detected, and ensure that users get professional help when they need

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357817

it. Customized AI conversations Improve user engagement and mental health outcomes from content recommendation systems in web-based mental health intervention research. Buoyed by these propositions, our study combines AI-driven chabot answers and deep learning technology for facial expressions to build an end-to-end virtual MH assistant.

Moreover, the chatbot interface is designed to be context-aware and emotionally intelligent, ensuring that the emotional state of the interactant is carried over in the responses. Empathetic interactions are important for lay therapy in digital mental health support according to research on conversational AI. Using the Google Maps for the psychiatrist suggestion the utility of the system is enhanced. AI-powered digital mental health applications, according to digital health research, should providers when required, and also offer virtual support. This project bridges the gap from AI-based emotional support to actual mental health interventions by identifying mental health providers in the user's area automatically.

2 Related works

Emotion recognition by CNNs has become more popular with the growth of deep learning, allowing accurate facial expression identification. Chatbot integration improves mental health care and personalized user experience.

[1] Pathirana et al. proposed a reinforcement learning approach using multimodal emotion perception to provide personalized mental well-being care. Their framework improved user engagement and therapeutic effectiveness with adaptive response based on emotion recognition. [2] Almulla designed a multimodal emotion recognition system with CNNs and integrated facial expressions, speech, and gestures. The combination enhanced detection precision and resilience in real-world applications. [3] Kumar et al. solved data imbalance in facial emotion recognition with a CNN model, enhancing reliability and accuracy in recognizing various emotional states. [4] Agung et al. employed CNNs on the Emognition dataset to identify facial emotions from still images, showing enhanced accuracy for real-world emotion-aware systems. [5] Bhagat et al. employed CNNs for emotion recognition using facial expressions, facilitating automated systems that can correctly identify under different conditions. [6] Jagtap et al. emphasized boosting e-learning through facial emotion recognition to aid student engagement assessment and modify teaching accordingly. [7] Shahzad et al. presented a hybrid CNN-based framework for real-time FER, enhancing robustness and accuracy using feature fusion methods. [8] Aly et al. presented a deep learning model for facial expression recognition for e-learning, optimizing detection for enhancing virtual engagement. [9] Sowmya et al. suggested an optimized CNN model for improved emotion detection, advancing machine learning-based emotional intelligence systems. [10] Akhand et al.employed transfer learning for CNNs in facial emotion recognition, reliably enhancing detection performance from pre-trained networks. [11] Abubakar et al. employed reinforcement learning to create an intelligent mental health chatbot, improving personalization and therapy efficacy. [12] Ajansondkar et al. created an emotional support conversational agent with the focus on enhancing user happiness through empathetic interactions guided by AI. [13] Li et al. did a systematic review of AI conversational agents for mental health, with their effectiveness proven in enhancing therapeutic results as well as emotional well-being. [14] Al-Hasan et al. discussed trends in recommender systems and GPT-based chatbots, predicting future advancements in emotionally aware AI systems. [15] Remountakis et al. investigated ChatGPT and persuasive technology in personalized messaging and proposed wider applications for improving the user experience of AI-based dialogue systems.

3 Methodology

3.1 Dataset and Preprocessing

The dataset consists of grayscale facial images of seven emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Images are from the FER2013 dataset, presented in the ICML 2013 Challenges in Representation Learning, and taken under natural conditions with various lighting, occlusion, and pose variations. The dataset contains 35,887 images divided into Training (28,709), Public Test (3,589), and Private Test sets.

Preprocessing involves:

1. **Image Resizing:** All images are resized to a consistent 48×48-pixel resolution for standardization of input sizes.
2. **Grayscale Normalization:** Normalization of pixel values to a fixed value range to provide consistent contrast between images.
3. **Class Balancing:** Methods like oversampling and undersampling are used to correct any remaining class imbalance and provide fair training across emotion classes.
4. **Dataset Splitting:** Split the dataset into training, validation, and test sets to ensure proper evaluation and prevent overfitting when the model is being trained.

3.2 Model Architectures

Custom deep learning-based CNN model was designed for real-time facial emotion detection, integrating convolutional layers for feature extraction and a chatbot driven virtual assistance system for personalized user interaction. Fig. 1 shows the Model Architecture.

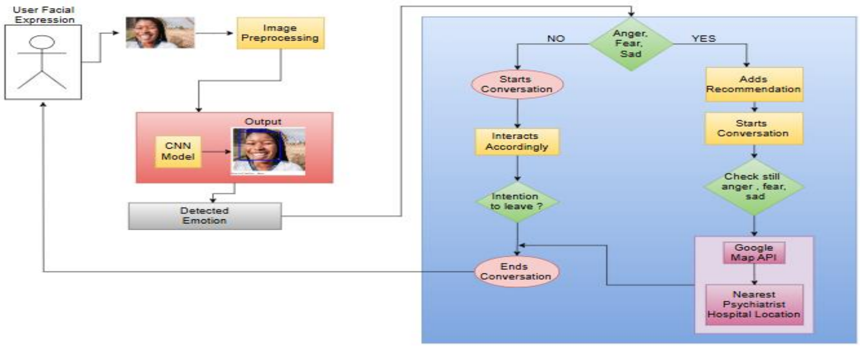

Fig. 1. Model Architecture.

3.2.1 Custom CNN Model

The Emotion Detection Module is constructed around a deep Convolutional Neural Network (CNN) designed for facial expression recognition using 48×48 grayscale images. In order to learn both high-level and low-level spatial features, the network's architecture consists of nine convolutional layers organized in increasingly deeper blocks. Following the extraction of base features by two Conv2D layers with 64 filters and 33 kernels, MaxPooling reduces the spatial dimensions to 24×24×64. The convolution and pool steps remain the same in the second convolutional block, which doubles the number of filters to 128. Later on, the network uses three Conv2D layers with 256 filters, keeping spatial resolution intact with proper padding and using MaxPooling to downsize to 6×6×256. The last block uses three additional Conv2D layers with 512 filters, which further increases feature granularity, and then uses MaxPooling to get a tight 3×3×512 representation. Each convolutional operation is followed by ReLU activation and Batch Normalization to improve gradient flow and non-linearity. To avoid overfitting, dropout layers are added strategically after critical blocks. The generated feature maps are passed through two fully connected Dense layers of 512 and 256 neurons, respectively, and flattened into a vector of size 4068. A final Dense layer using softmax activation offers the emotion categorization into seven categories: Angry, Disgust,Fear, Happy, Sad, Surprise, and Neutral. Real-time emotion recognition systems would highly benefit from this strong CNN model's capability of achieving regularization, feature abstraction, and depth at an acceptable level. Fig. 2 shows the Custom CNN Model Architecture.

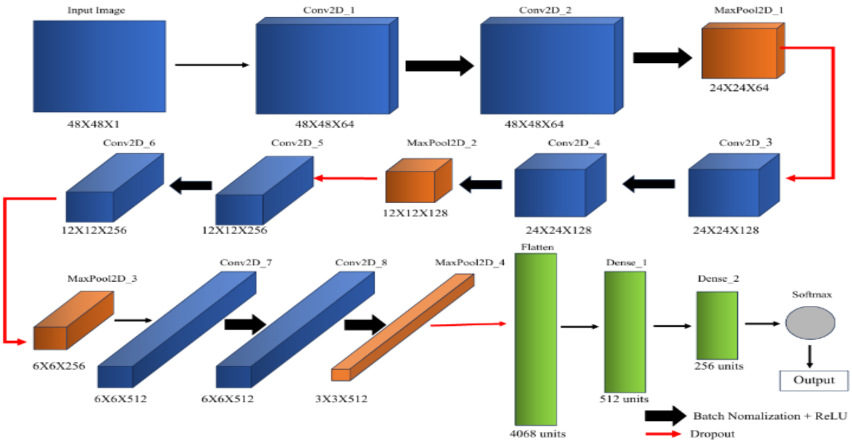

Fig. 2. Custom CNN Model Architecture.

3.2.2 Chatbot Interaction

The chatbot module is based on the Gemini conversational language model API, which uses Transformer-based models to support context-dependent and emotion-aware dialogue generation. The module takes the predicted emotion as an input and dynamically controls the prompt format to produce empathetic and responsive answers. Emotion classes are categorized by severity: non-critical (Happy, Surprise, Neutral, Disgust) and critical (Sad, Fear, Angry). For non-critical moods, the system builds lean prompts to encourage or boost positivity. For critical moods, prompt templates are personalized to ensure empathy, motivational encouragement, and psychological reassurance. The API for the chatbot is reached through RESTful calls, with the response output rendered on the frontend in the form of an actual-time chat interface. Every interaction is designed asynchronously to permit smooth user experience. This module connects emotion detection based on deep learning with natural language processing to provide personalized emotional support.

3.2.3 Psychiatrist Recommendation

For users going through critical emotions (Sad, Fear, Angry), the system triggers a geolocation-aware psychiatrist recommendation engine. The API provides a ranked list of suitable healthcare providers. The results are rendered in the frontend as an interactive map with clickable markers, enabling users to navigate and access proximal professional help effortlessly. The recommendation engine is conditionally activated only for negative high-risk emotion predictions to prevent redundant exposure and maintain user attention. This module increases the real-world applicability of the system by marrying emotional analysis with actionable, location-based healthcare advice.

3.2.4 Algorithm

1. Image Capture: Capture real-time facial images via a webcam for live emotion recognition.
2. Preprocessing: Identify and crop faces with OpenCV, then resize and normalize for uniformity.
3. Emotion Classification: Classify facial expressions using a custom CNN model trained on the FER2013 dataset.
4. Emotion Identification: Identify the expression as one of seven emotions: Happy, Sad, Angry, Fear, Surprise, Disgust, or Neutral.
5. Chatbot Integration: Integrate with the Gemini to facilitate emotion-aware chatbot interactions.
6. Response Generation: Provide context-sensitive chatbot responses according to the detected emotion.
   Decision Flow:
   - If emotion is positive (Happy, Surprise, Neutral, Disgust): end chatbot interaction.
   - If emotion is negative (Sad, Angry, Fear): proceed with emotional support conversation.
7. Psychiatric Support: Suggest consulting a psychiatrist for negative emotions by locating the nearest mental health center through Google Maps integration.
8. Deployment: Deploy the whole system in a real-time setting for ongoing emotion detection and support.

3.2.5 Training and Evaluation

The training pipeline is in a well-defined sequence:

1. Data Preparation: FER2013 dataset is normalized and augmented with operations such as rotation, flip, and zoom to increase variability.
2. Feature Extraction: A specialized CNN model extracts discriminative facial features from images to identify emotional hints.
3. Emotion Classification: The model is trained to classify images into seven classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.
4. Hyperparameter Tuning: Adam optimizer is employed to tune learning rate, batch size, and dropout rates for better convergence. Table 1 shows the Performance Metrics for Emotion Classification using CNN Model.
5. Performance Metrics: Accuracy, precision, recall, F1-score, and confusion matrix are employed to measure the model's classification performance.

Table 1. Performance Metrics for Emotion Classification using CNN Model.

Emotion	Accuracy	Precision	Recall	F1-Score
Anger	97	79	95	87
Fear	80	78	81	79
Sad	91	83	89	87
Neutral	62	86	63	72
Happy	98	85	96	91
Surprise	80	81	78	81
Disgust	62	75	60	68

4 Results and Evaluation

4.1 Emotion Detection

The suggested emotion detection model is tested on the FER2013 dataset in order to determine facial expressions against seven emotion categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The model performance was measured against common classification evaluation metrics such as precision, recall, F1-score, and overall accuracy.

4.2 Chatbot Interaction

Once an emotion is identified, the system invokes a Gemini powered chatbot that interacts with users in appropriate, emotionally supportive conversations. The conversation is tailored to provide comfort, encouragement, or support depending on the user's emotional condition.

- For neutral or positive emotions (Happy, Surprise, Neutral, Disgust), the chatbot provides affirmations and maintains light-hearted conversation.
- For negative emotions (Sad, Fear, Angry), the chatbot adopts a friendly tone and makes an effort to motivate the user.

The chatbot dynamically changes based on inferred emotion and creates human-like empathetic responses in order to enhance user trust and emotional bonding. Fig. 5 shows the Displaying nearby Psychiatrist Hospital by Google Maps.

4.3 Psychiatrist Recommendation System

For purposes of mental well-being assistance over virtual advice, the system uses Google Maps API to suggest the closest psychiatrist or mental health clinic according to the user's location. Upon detecting Fear, Sad, or Anger, the chatbot first engages with the user. When the negative state continues or is verified, the system initiates a real-time location-based lookup to suggest authenticated psychiatric experts or hospitals in proximity. This aspect brings a real-world, practical layer to the application that connects emotional detection to mental health assistance. Fig. 3 shows the Input Happy Emotion with Chatbot Interaction. Fig. 4 shows the Input Sad Emotion with Chatbot Interaction.

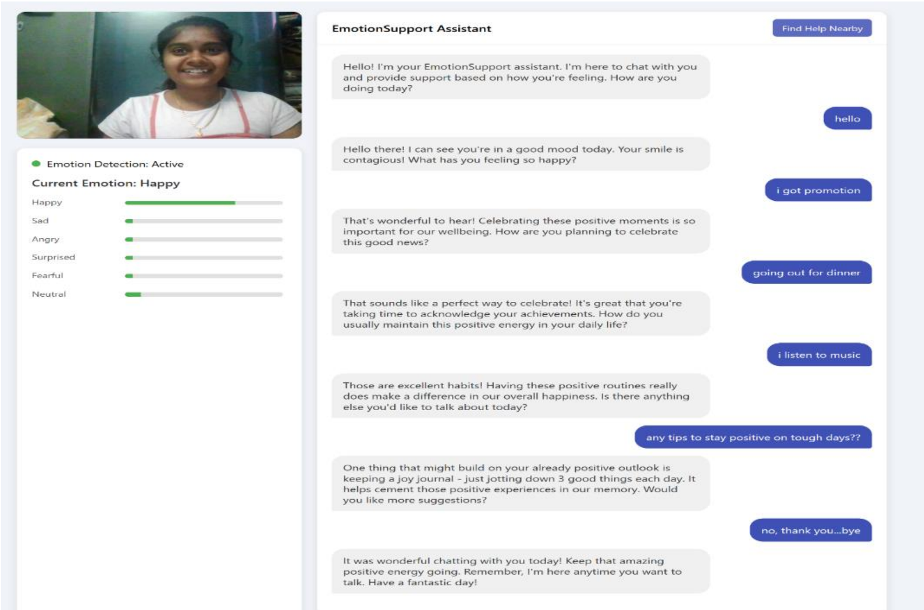

Fig. 3. Input Happy Emotion with Chatbot Interaction.

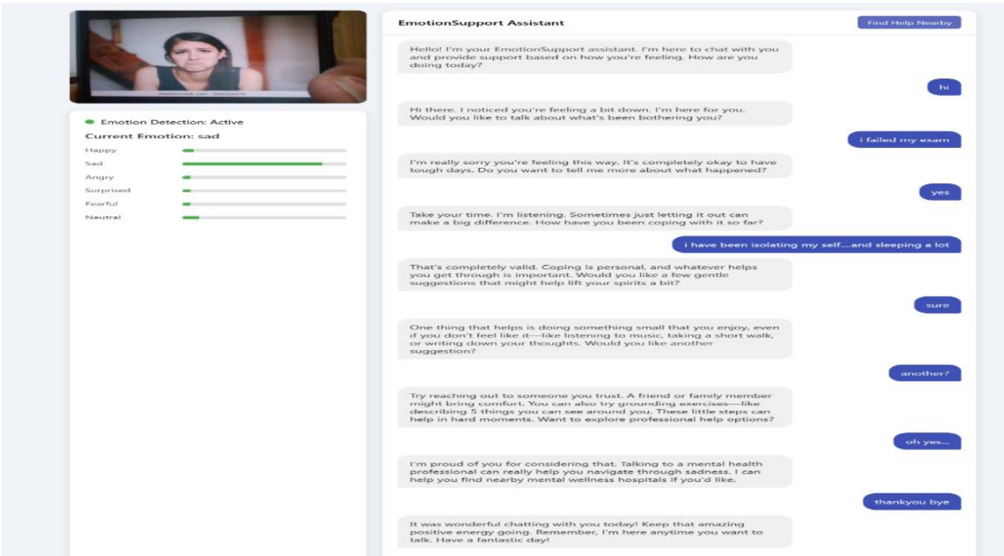

Fig. 4. Input Sad Emotion with Chatbot Interaction.

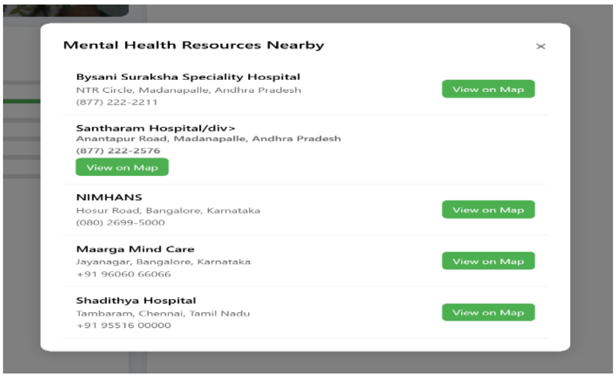

Fig. 5. Displaying nearby Psychiatrist Hospital by Google Maps.

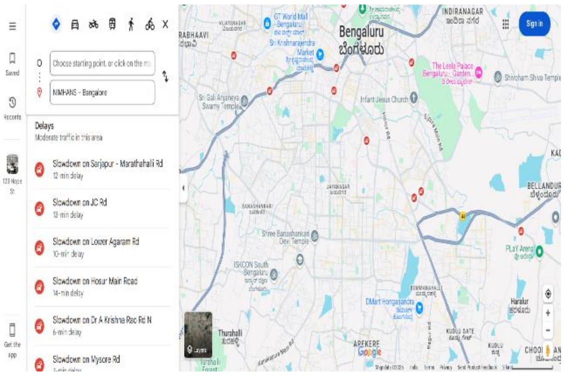

Fig. 6. Navigates to the nearby hospital.

5 Conclusion and Future Scope

Using a custom Convolutional Neural Network (CNN), an advanced chatbot, and a geolocation-based psychiatrist referral module, the proposed system is presented as a real time emotion recognizer and mental health support system. The CNN trained on the FER2013 dataset had an 81.4% average accuracy in efficiently categorizing seven emotions: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Fig. 6 shows the Navigates to the nearby hospital.The model's stability is demonstrated by the high precision and recall of the happy and angry emotions. A context-aware chatbot that responds emotionally to the identified emotion and provides real-time care is another feature of the system. For negative emotions like Sad, Fear, or Angry, the chatbot begins soothing conversation, and the system utilizes Google Maps to suggest close-by mental health professionals, thereby increasing access to proper care.

The suggested framework suggests a few ways to improve it in the future to make it better and more applicable to real world situations. Accuracy and resilience would be improved when sophisticated transfer learning models like VGGFace, ResNet, or EfficientNet were incorporated. These models could be used in a variety of lighting and facial expression scenarios. Integrating facial expressions with voice, speech sentiment, or physiological signals like heart rate in multimodal emotion recognition can improve reliability. Implementing the system as a mobile or web application would enhance accessibility. Users may be able to track trends in mental health with the help of features like emotional history tracking, and an emergency alert mechanism may notify caregivers of distress. In addition, incorporating this system into clinical and educational settings may facilitate early diagnosis, awareness of mental health, and intervention, all of which contribute to emotional well-being in a variety of real-world contexts.

References

[1] Pathirana, A., Rajakaruna, D.K., Kasthurirathna, D., Atukorale, A., Aththidiye, R. and Yatipansalawa, M., 2024. A Reinforcement Learning-Based Approach for Promoting Mental Health Using Multimodal Emotion Recognition. Journal of Future Artificial Intelligence and Technologies, 1(2), pp.124-142.

[2] Almulla, M.A., 2024. A multimodal emotion recognition system using deep convolution neural networks. Journal of Engineering Research.

[3] Kumar, T.A., Aashrith, M., Vineeth, K.S., Subhash, B., Reddy, S.A., Alam, J. and Maity, S., 2024, April. Enhancing Facial Emotion Level Recognition: A CNN-Based Approach to Balancing Data. In International Conference on Advances in Information Communication Technology & Computing (pp. 381-395). Singapore: Springer Nature Singapore.

[4] Agung, E.S., Rifai, A.P. and Wijayanto, T., 2024. Image-based facial emotion recognition using convolutional neural network on emognition dataset. Scientific reports, 14(1), p.14429.

[5] Bhagat, D., Vakil, A., Gupta, R.K. and Kumar, A., 2024. Facial emotion recognition (FER) using convolutional neural network (CNN). Procedia Computer Science, 235, pp.2079-2089.

[6] Jagtap, M.S., Marne, M.A., Sheikh, M.A., Potdar, M.V. and Chate, P., Design and Analysis of Unlocking Student Emotions: Enhancing E-Learning with Facial Emotion Detection.

[7] Shahzad, H.M., Bhatti, S.M., Jaffar, A., Akram, S., Alhajlah, M. and Mahmood, A., 2023. Hybrid facial emotion recognition using CNN-based features. Applied Sciences, 13(9), p.5572.

[8] Aly, M., Ghallab, A. and Fathi, I.S., 2023. Enhancing facial expression recognition system in online learning context using efficient deep learning model. IEEE Access, 11, pp.121419-121433.

[9] Sowmya, B., Alex, S.A., Kanavalli, A., Supreeth, S., Shruthi, G. and Rohith, S., 2024. Machine learning model for emotion detection and recognition using an enhanced convolutional neural network. Journal of Integrated Science and Technology, 12(4), pp.786-786.

[10] Akhand, M.A.H., Roy, S., Siddique, N., Kamal, M.A.S. and Shimamura, T., 2021. Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), p.1036.

[11] Abubakar, A.M., Gupta, D. and Parida, S., 2024. A reinforcement learning approach for intelligent conversational chatbot for enhancing mental health therapy. Procedia Computer Science, 235, pp.916-925.

[12] Ajansondkar, A.S., Bachate, S.A., Bhange, G.S., Gwalvanshi, Y.S. and Take, A.T., Developing an Intelligent Conversational Agent for Mental Health Support.

[13] Li, H., Zhang, R., Lee, Y.C., Kraut, R.E. and Mohr, D.C., 2023. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digital Medicine, 6(1), p.236.

[14] Al-Hasan, T.M., Sayed, A.N., Bensaali, F., Himeur, Y., Varlamis, I. and Dimitrakopoulos, G., 2024. From traditional recommender systems to gpt-based chatbots: A survey of recent developments and future directions. Big Data and Cognitive Computing, 8(4), p.36.

[15] Remountakis, M., Kotis, K., Kourtzis, B. and Tsekouras, G.E., 2023. Using ChatGPT and persuasive technology for personalized recommendation messages in hotel upselling. Information, 14(9), p.504.