Enhancing Chatbot Conversations with Real Time Facial Expression Recognition

Balla Charishma Sulochana, Bhavya Sri Pragada, Boya Chaitanya Kiran, Gaddam Anvith Reddy,

*Department of Computer Science and Engineering,  
Amrita School of Computing, Bengaluru,  
Amrita Vishwa Vidyapeetham, India*

[charisma.sulochana@gmail.com](mailto:charisma.sulochana@gmail.com), [bhavyasri.pragada@gmail.com](mailto:bhavyasri.pragada@gmail.com), [chaitanyakiran415@gmail.com](mailto:chaitanyakiran415@gmail.com), [ganvithreddy3052@gmail.com](mailto:ganvithreddy3052@gmail.com)

*Abstract*— In the realm of human-computer interaction, the ability to discovery, understand, and react to human emotions is at the core of the engaging and customized experiences that relies on. The emotion sense and the customized answer are one of the most important things in the scope of human-computer interaction and the immersion and the engagement of the experience completely depend on it. This research is dedicated to the creation of an actual time face expression detection system based on deep learning algorithms. Emotion detection is considered as the main aspect of further improving the emotional intelligence of the system and bringing the responses in agreement with the user's emotional status. A smart individualized answer, in which the end-user can seek mood-based recommendations on quotes, games, activities, music and food while using chatbot. The goal here is to develop and utilize a technology that will accurately detect the emotions displayed in facial expressions so that it becomes possible to give emotional responses through the automated service provider.

Keywords— Deep Learning Algorithms, Emotion Recognition, Facial Expression Recognition, Chatbot, Mood based Recommendations, Feature Extractions, Emotion Classification.

# Introduction

In the modern digital world, people's interaction with computers keeps developing, creating chatbots as a widely used source of interaction. In this case, while today's chatbots are very good at understanding and responding to written communications, they still lack the conversational skill to recognize the user's emotional state, essential for successful communication. Spotting the facial expressions of the customers at the moment enables the chat-bots to communicate with the consumers in a more precise and personalized way.

The topic of emotion identification – more specifically, through facial expression analysis – has quickly become one of the key research areas within the wider field of human-computer interaction. The usage of deep learning techniques by experts enables them to make considerable strides in developing systems that can discern and understand facial expressions accurately and in real time. The ability to effectively spot and understand facial expressions is one of the key factors determining the future quality of robot-human communication. Facial expressions are a very powerful instrument for the transmission of a lot of information about the other person's emotions, goals, and preferences, which may significantly influence the results of the contact. This deep learning system shall extrapolate such micro facial signals and therefore reliably conclude the mood of the user in a split second.

The optimal function of chatbots with real-time facial expression detection is closely related to matching mood with recommendations. Chatbots can be programmed to deliver right learning resources such as for example, motivational quotes, fun games, exciting activities, happy music or something yummy based on some facial expressions that help the user in identifying the user's emotional state. For example, it can propose relaxation exercises, calming music or quotes, if it detects that the user is faux pas or jittery. On the other hand, if the user presents himself as sad or worried, it could mean negative activities, music in a depressed mood, or some discouraging comments.

By doing this, not only is the user experience made more fulfilled, but also a deeper level of engagement and connection is generated between the user and the chatbot. Moreover, chatbots can achieve this by equipping them with emotion detection and allowing them to naturally be more \interactive and thus empathic. In general, real-time facial expression recognition integration will have an impact on how people relate to the computers as it will give the chatbots power to communicate with the users in an easier, natural, and rational way.

# Related Works

S. N. Gourav and team devoted to the enhancement of voting procedure and it employs face recognition-based mood sensing system [1]. The facial expressions of voters are scrutinized in real-time applying machine learning algorithms, particularly python with TensorFlow, Keras, and OpenCV, to identify the emotions. Biometric authentication especially - facial recognition - can be used for authentication, and therefore fraud can be prevented. The system emphasizes data security along with privacy controls, user surveys on performance, and user responses. To enhance the election process, the following technologies are suggested for system development: express.js, node.js, MongoDB and react.js. Nevertheless, the process of election transparency has ethical and privacy considerations. Thus, it is important to deliberate cautiously.

A. O. Andrade et.al introduces to you Hill, the system which extrapolates the facial expressions, (macro emotional), using the computer vision. The article explains a human, a method, a system, that utilizes the computer vision to read the macro facial expressions. In this technology, variety of targets can be pointed to, mainly humble the learning potential of other AI systems communicating with people. The architecture involves a CNN setup containing layers like 'relu','kernel regularisation','batch normalisation', 'max pooling', and 'drop out' that reduces the error in output. To Smart cars, games, health monitoring, education, marketing, and movies Hill finds the organizations of these emotions such as happiness, sadness, surprise, neutrality, rage, disgust, and fear suitable. [2] Text displays the conclusion that Tensorflow has ultimate AI training effectiveness, Python is the best available programming language, and CNN is an essential component in image recognition. Using emotion recognition AI, in conjunction with natural language processing AI and chatbot AI which learns like the human brain, the AI system interprets the patient's facial expressions during the interaction. This correlation may also be used as a guide for mental health care.

R. S. Reddy A and team approached through utilizing the data set of FER 2013 dataset to show about how efficient Convolution Neural Network (CNNs) are in detecting human emotions from facial expressions. It follows a hybrid training approach: at the beginning, it uses a transcripts corpus and, later, the model is corrected based on user feedback. [3] The research proves that the model's performance in classifying facial expressions will be an encouragement of emotion-sensitive chatbot systems among future developers. Through the indication a user's emotional state, the proposed solution will contribute towards more expressive conversations. Secondly, it possesses the ability to adapt to the spectrum of emotions, hence, creating the fantastic user experience. Promising outcomes have frequently been revealed by user research that have been carried out on chatbots which incorporate deep learning-based facial expression recognition with natural language processing algorithms and so they are able to interact with users according to their emotional states. Notwithstanding that, it will be suggested that the content suggestion structure and the emotional based educational feedback systems should be enhanced in the future studies as well as a study on more advanced deep learning models for emotion recognition and chatbot interaction is proposed.

R. K. Nandi and team proposed technique which consists of a mixed autonomous music selection system, based upon mood conjunctions, and a face expression detection system implemented by a CNN. With the FER2013 dataset at hand, the facial expression recognition model achieves the recognition rate at 62.1%. [4] By contrasting to the algorithms which cause the only use of past listening history, it is possible to reach such a goal. Music recommendation algorithm based on content grounding is composed of the cosine similarity algorithm with the purpose of suggesting music based on matching expression. It is proved that the emotions playlists designed and adapted to the anticipated moods of the users accurately optimizes their mood. The results of traditional machine learning techniques like random forest was not successful like they are in the case of deep learning techniques, especially CNN for facial expression recognition. Although there is progress, concerns still need to be addressed in order to produce the top quality explanations of emotion and high accuracy of the tags with poor recognition rate. To start with the noise will be the smaller one and then over time will be the optimized.

E. R. Vimina et.al reports about a faculty version of CAPTCHA, the "Automated Reverse Turing Test via Facial Expressions" system, that will use AI face detection, that is considered the most difficult test as a hard test. Participants are exhibited with the visuospatial ght of human faces showing a given state of the emotion, after which a number of faces with different emotions are shown while the participant try to select one that represents the first one. [5] Two photos-style CAPTCHAs, ESP PIX and SQUIGL PIX were the big failure, as in comparison to the CAPTCHA that users had preferred because of its accuracy, that was almost 95.82%. Usually, this exam needs less student's time to cover its contents when compared to the other options, making it a proof of its efficiency. By developing facial expression detection as the main issue robots are faced within CAPTCHA, a system which cannot be broken by any automated bots is implemented, while this strategy is practical and can be applied in the range of issues. With this point the contribution to the security system of virtual Internet space is demonstrated.

S. Palaniswamy and team reveals the approach that distinguishes emotions in videos that can be known by the one to one research called STLEV. There is in this work a combination of LSTM and SNN with Triplet Loss. In the STLEV framework, SNN is used for feature extraction and LSTM [6] caters for the classification of emotions from video frames using both spatial and temporal data retrievals Therefore, this method majorly addresses the downside of traditional deep learning models which is the requirement for a large training set using the triplet loss and an accuracy of 87.5% on the BU-4DFE dataset. Tests indicate that STLEV is highly effective as compared to the conventional feature engineering methods, and they show that deep learning approaches also provides an additional edge for video emotion recognition. This method that detects moods can productive in cognitive human-computer interaction (HCI) and can stimulate potential roadmaps for the future studies that will be based on other meta-learning strategies and model complexity decrease.

Suja.P and team approached a new technique termed MREAP for identifying human emotions from facial expressions via the incorporation of the meta-learning paradigm based on Siamese Networks. Through using a data set that determines 80% of different poses in-house, MREAP overcome other deep learning methods and got an accuracy of 80%. [7] That the model predicted the appropriate emotions for poses that weren't seen during training but instead were sued as inputs during the testing stage show its discriminative ability. Through learning to employ the Siamese network technique, a new way of feature vector construction is created using the Euclidean distance measure, thus, providing for accurate predictions using less train time and small samples. The extraordinary execution time of less than 3 minutes signifies its ability to be employed in real-time apps. One of the forthcoming improbable interactions is to adjust the loss function, to expand data and research on complex meta-learning techniques variations among other things to better the performance and accuracy.

Suja.P et.al presents two-dimensional emotion recognition system by using facial expressions and upper body movements. It proposes two approaches: parsing out the emotive information from each mode and the other feeding on both kinds of data for better precision. Feature-level fusion methods applied, such as concatenation and Principal Coefficient Analysis (PCA) [8] are also used to combine the extracted features. Classifying the data is executed by using classifier algorithms such as Artificial Neural Network, SVM, and HMM. Evaluation is based on the MATLAB software, with outstanding results, specifically for such tasks as creating something inconspicuous or low-expressive in one mode. Coupling facial motion and upper body movements increases the level of emotion recognition accuracy more than single modality strategies. The pertinent future research might be based on Deep Learning and data augmentation to bring the best performance for multiple poses and light conditions.

Ghallab and team explores a distinctive disability adapted learning system which observes the process of students via facial expression recognition technology. A ResNet-50 strategy, which moves forwards the selection of the features, has been used in the algorithm. On RAF-DB for instance, it gets 87.62% accuracy, the accuracy on FER2013 comes up to be 88.13%. Then initiate the expansion of ResNet by [9] applying convolution attention mechanism which become critical because by reducing the unnecessary regions and keep the rest function effectively, the attention becomes essential for higher accuracy. The research validates about the application of online learning that incorporates facial recognition by preferably of the current teaching strategies that have been adapted to respond to the facial recognition results. The efficiency of the algorithm is verified through comparing it to the recent publications and the pertinent standardized datasets which are RAF-DB, FER2013, KDEF and CK. The credibility of the algorithm using expression datasets of multiple types is hence, demonstrated and the meaning of its contribution to the scientific community for the second time is highlighted. According to the research, the study also highlighted the difficulties in properly taking students pictures and identifying them. Additionally, it was mentioned that it is also remarkable in making sure that what the students deserve is given to them.

N., Yuvaraj et.al aims at developing an emotional intelligence model in social media communication involving a complex Bi-Directional Recurrent Neural Network and TensorFlow library functions. Technically, expressed in several approaches, the sentiment analysis is achieved through machine learning based automatic systems, rule -based systems, and generative systems. The rule-based approach is a strategy that requires creating a rule system where subjectivity and polarity can be identified by employing NLP techniques like stemming and tokenization. Emotions tender, positive or negative, and user behaviours or tasks could also be assignee using extras like features of emotion recognition, aspect-based sentiment analysis, and intent analysis. [10] The proposed chatbot reflects quick adaptation for various sectors from education via internet and e-government as well as delivery of specific solutions to queries, thus enhancing user engagements and satisfaction. The precision and applicability of the tool have been missed, however, they can be supposed from its capability of recognizing emotions and providing advisable recommendations according to what people are talking about in social media texts. To sum up the author proposes to achieve this with the subsequent developments of the chatbot in order to aim at the certain goals as well as its expertise.

Perveen and team introduces a new algorithm for facial micro-expression recognition, which consists of the combination of multi-stream deep convolutional neural network (CNN) with elements classification. The model is made by taking deep features of state-of-the-art architectures for instance ResNet-50, DenseNet-121, and VGG-16 then [11] PCA is applied for dimensionality reduction. A stacking method that includes three base learners (random tree, J48, random forest) and a meta learner (random forest) is utilized by applying a the-stacking technique. The designed approach (PA) that detects lower errors and time than the existing approaches is proven to be accurate on the dataset including CASME-II, CASME 2, SMIC, and SAMM. Additionally, the obtained accuracy, precision, and recall exceed all the deep networks and other ensemble approaches. So, it proves that the proposed method can be applied for facial micro-expression recognition demonstrating a wide range of perspectives for practical implementation in the emotion detection area.

X. Zhang and team research article discusses a new deep-learning architecture, called the frequency neural, network (FreNet), which is capable of classifying facial expression using the frequency domain for its operation. FreNet, depending upon the frequency domain processing, thus extracts highly abstract representations from vocabulary vectors by way of reduced dimensionality layer and a summarization layer. [12] Unlike the prior models, the recent Basic FreNet and Block FreNet techniques show a significant gain in performance and reduction in the computational cost, because of the greater efficiency in feature learning and dimension reduction. The outcomes also show that this network is very useful for learning features and forecast facial expressions, which are even better than known techniques such as GoogleNet, and such calculations are light-weighted. These suggested techniques are proof that frequency-based deep-learning models can be used for human-computer interaction, looking to their achievements in this special task of facial expression recognition.

Anand et.al work suggests a deep learning and classification-based approach for identifying facial emotions on faces. The system can be set up for categorizing facial emotions and recognition of faces represented in images. Data processing involve picture normalizing (MT-CNN) [13] as well as histogram equalization that are to be done from the start of the process. EfficientNetB0 is chosen to pull features from funneling images, so and WKELM with the Red Fox Optimizer (RFO)-optimized kernel parameters used to classify emotions is applied. The proposed architecture shows that the extracted features can be effectively used for classifying binary emotions and obtains the highest accuracy of 95.82% on FER2013 and 96.98% on EMITIS datasets. Due to a competitiveness demonstrated when it was compared to transformers and ResNet-18, two leading models, the given model contain more useful information. General performance is enhanced by spatial resolution improvement through the finding significant features by EfficientNetB0, and emotion classification using a low dimensional space by WKELM tuned with the optimized parameters. For increasing the classifier accuracy, possible research approaches in future include broadening the context of model to cater with bigger datasets containing many emotions as well as incorporating feature selection methods. It is expected that, along with NN evolution, computer power, will also make it possible to perceive and interpret emotions even better by receiving crucial information from facial expressions.

Shalaka Prasad Deore experimented on introducing SongRec, a facial emotional recognition system for song recommendation which is based on Convolutional Neural Networks (CNN) [14] that is used for training the facial dataset and signifying different reactions to the music. The algorithm trained for it gets the user mood as per facial expressions, and it gives songs based on it. The validation of this methodology is provided by the positive outcomes in the music choice classification of the mood, and the achieved accuracy of recognition of 62.88% in the real time process. Beyond that, fuzzy classification will be supplemented which is accountable for intensifying the results, measuring accuracy with Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The system is also the proper suit here considering its capacity to analyse the user's face expression correctly and make music recommendations accordingly; therefore, on can predict that it has a bright future with real situations. While it employs today’s standard webcam video capturing, the use of Haar feature for face identification and fuzzy classification for reliable emotion recognition, ultimately provide song recommendations that consider facial expressions of mood. Resorting to the experimental evaluation proves the prevalence of the proposed model in mood identification, which can be used as an API or mobile application component.

Amrita Nair et al. had created an interactive playlist recommendation system that relied heavily on emotions. It labeled user replies into positive, negative, and neutral emotions utilizing sentiment analysis models that were induced on Twitter dataset, which incorporated LSTM, Bidirectional LSTM, and 1-D Convolutional Neural Network [15]. The dataset was used to train three neural networks: LSTM BiDi, 1D Convolutional NN, and LSTM Neural Network. Those are, Bidirectional LSTM is highest, 79.29% correct classifications.

Bakariya et.al proposes a real-time system that combines facial emotion recognition and music recommendation. The paper explains the concepts of face recognition and emotion identification using deep learning techniques. The methodology involves the use of deep learning techniques, specifically [16] convolutional neural networks (CNN), to train and evaluate two CNN models (Model-A and Model-B). The results show that Model B has a higher validation accuracy compared to other models. The content also presents the results of testing the models using random images and displays the resulting confusion matrix. Overall, the study demonstrates the effectiveness of the proposed system in recognizing facial emotions and providing music recommendations based on the detected emotions.

Kim and team compared emotional recognition abilities in children with Prader-Willi Syndrome (PWS) and Autism Spectrum Disorder (ASD), highlighting differences in response to disgust and social interaction difficulties. Valuable insights were gained for intervention development. [17] An ex post facto comparative design was used to evaluate children with PWS and ASD, along with a control group. Tasks included emotional recognition, Theory of Mind (ToM), Working Memory, and ASD traits assessment. The study employed the Mann-Whitney U-test, Chi-Square test, and Spearman’s Rho correlation test to analyze emotional recognition, ToM, Working Memory, and ASD traits. PWS children exhibited higher disgust response, while ASD children faced more challenges in social interactions. These findings provide important implications for tailored interventions for children with PWS and ASD.

Sheng-Cheng and team discusses a system for real-time recognition of human emotions, ages, and genders. The paper focuses on recognizing six emotions and an additional neutral emotion, as well as dividing age groups into children, [18] young adults, middle-aged adults, and older adults. The methodology involves preprocessing techniques such as normalized facial cropping (NFC) and binocular line angle correction (BLAC) to enhance the accuracy of recognition. The paper also describes the preprocessing techniques used for data normalization and image cropping. In terms of age recognition, the mean accuracy rate using normalized data was 69.15%. For gender recognition, the mean accuracy rate using normalized source data was 91.75%.

1. Dataset

The Dataset ,derived from Kaggle contains a collection of 15,452 RGB photos capturing face expressions each connected with one of 6 distinct emotions : Happy, Angry, Sad, Neutral, Surprise and Disgust. The photographs contain a varied range of persons, including distinct gender, age groups and different cultures.

The dataset is divided into folders within dataset.zip, each folder corresponding to one of the six emotion classes. Additionally, data.csv contains file paths to the photographs along with their related emotion labels.

##### References

[1] S. N. Gourav, K. Arora, S. Khan, A. Dulat and S. Sinha, "Facial Recognition-Based Emotion Detection for Enhanced Voter Experience," 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gautam Buddha Nagar, India, 2023, pp. 75-81, doi:10.1109/UPCON59197.2023.10434600.

[2] T. T. Almeida Silva, A. O. Andrade and N. P. da Silva, "System for Recognition of Facial Expressions Using Machine Learning," 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE), Natal, Brazil, 2020, pp. 1-6, doi:10.1109/LARS/SBR/WRE51543.2020.9307127.

[3] N. V, R. G B, N. L. B, S. Kumar C P, R. S. Reddy A and P. J, "Chatbot Based on Emotions Using Deep Learning," 2023 1st International Conference on Optimization Techniques for Learning (ICOTL), Bengaluru, India, 2023, pp. 1-5, doi:10.1109/ICOTL59758.2023.10435281

[4] S. Kumar, R. K. Nandi, S. Roy and A. Sarkar, "Design and Implementation of AI Based Efficient Emotion Detection and Music Recommendation System," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5, doi: 10.1109/ICCCNT56998.2023.10306938.

[5] E. R. Vimina, "An Automated Reverse Turing Test Using Facial Expressions," 2009 2nd Conference on Human System Interactions, Catania, Italy, 2009, pp. 314-317, doi: 10.1109/HSI.2009.5090998.

[6] D. Lawrance and S. Palaniswamy, "Emotion recognition from facial expressions for 3D videos using siamese network," 2021 International Conference on Communication, Control and Information Sciences (ICCISc), Idukki, India, 2021, pp. 1-6, doi: 10.1109/ICCISc52257.2021.9484949.

[7] S. Ramakrishnan, N. Upadhyay, P. Das, R. Achar, S. Palaniswamy and A. A. Nippun Kumaar, "Emotion Recognition from Facial Expressions using Images with Arbitrary Poses using Siamese Network," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 268-273, doi: 10.1109/ICOSEC51865.2021.9591919.

[8] T. Keshari and S. Palaniswamy, "Emotion Recognition Using Feature-level Fusion of Facial Expressions and Body Gestures," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1184-1189, doi: 10.1109/ICCES45898.2019.9002175.

[9] Aly, Mohammed & Ghallab, Abdullatif & Fathi, Islam. (2023). Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model. IEEE Access. 11. 10.1109/ACCESS.2023.3325407.

[10] Balaji, M. & N., Yuvaraj. (2019). Intelligent Chatbot Model to Enhance the Emotion Detection in social media using Bi-directional Recurrent Neural Network. Journal of Physics: Conference Series. 1362. 012039. 10.1088/1742-6596/1362/1/012039.

[11] Perveen, Gulnaz & Ali, Syed & Ahmad, Jameel & Shahab, Sana & Adnan, Muhammad & Anjum, Mohd & Khosa, Ikramullah. (2023). “Multi-Stream Deep Convolution Neural Network With Ensemble Learning for Facial Micro-Expression Recognition.” IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3325108.

[12] Y. Tang, X. Zhang, X. Hu, S. Wang and H. Wang, "Facial Expression Recognition Using Frequency Neural Network," in IEEE Transactions on Image Processing, vol. 30, pp. 444-457, 2021, doi: 10.1109/TIP.2020.3037467.

[13] Anand, M. & Babu, S.. (2024). Multi-class Facial Emotion Expression Identification Using DL-Based Feature Extraction with Classification Models. International Journal of Computational Intelligence Systems. 17. 10.1007/s44196-024-00406-x.

14] Shalaka Prasad Deore, "SongRec: A Facial Expression Recognition System for Song Recommendation using CNN." International Journal of Performability Engineering, vol. 19, no. 2, Feb. 2023, pp. 115-121.

[15] A. Nair, S. Pillai, G. S. Nair and A. T, "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1767-1772, doi:10.1109/ICCES51350.2021.9489138.

[16] Bakariya, Brijesh & Singh, Arshdeep & Singh, Harmanpreet & Raju, Pankaj & Rajpoot, Rohit & Mohbey, K..(2023). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. Evolving Systems. 15. 1-18. 10.1007/s12530-023-09506-z.

[17] Oh, S.; Kim, D.-K. Comparative Analysis of Emotion Classification Based on Facial Expression and Physiological Signals Using Deep Learning. Appl. Sci. 2022, 12, 1286. https://doi.org/10.3390/app12031286

[18] Lu, Ta-Te & Yeh, Sheng-Cheng & Wang, Chia-Hui & Wei, Min-Rou. (2021). Cost-effective Real-time Recognition for Human Emotion-Age-Gender Using Deep Learning with Normalized Facial Cropping Preprocess. Multimedia Tools and Applications. 80. 10.1007/s11042-021-10673-x.