AffectAI: An Emotion-Aware conversational AI for Human-Computer Interaction

1st Ms. Rajvee Sakariya
Dept. of Computer Science & Engineering
rajveeepatel777@gmail.com

2nd Mr. Om Sakariya
Dept. of Information Technology
omsakariya7373@gmail.com

3rd Ms. Rauki Yadav
Dept. of Computer Science & Engineering
rauki.yadav@bmusurat.ac.in

Abstract— In the present digital environment, conversational agents are becoming more and more common but they usually do not know how to interpret and react to human emotions, contributing to robot-like communication. The current study introduces AffectAI, an effective conversational system that will close this emotional intelligence gap. The holistic AffectAI system envisions a multimodal system, with the text sentiment analysis, speech emotion recognition, and facial expression detection to identify the affective state of a user. The conversational engine uses a dialogue manager which is a reinforcement learning agent that is capable of dynamically adjusting its tone, empathy level, and response content depending on the emotion that has been detected, and the context of the dialogue. The paper describes the design, development, and testing of the underlying visual element of this system, a real-time facial emotion recognition module. Based on a Convolutional Neural Network (CNN) trained using TensorFlow and optimized to execute on the Android platform through TensorFlow Lite, this module is implemented on the Android platform. Experimental analyses prove that this element can show a higher accuracy in emotion detection and user interaction, creating a strong paradigm of building more empathetic, context-aware, and human-like conversational agents to work in the mental health, education, and customer service areas.

Keywords—face recognition, face detection, human-computer interaction, and real-time emotion detection

I. INTRODUCTION

Conversational artificial intelligence is now a vital aspect of the digital system today, which assists the user with the applications of the virtual assistant, health services [9], education, and customer care[12]. Developments in the fields of deep learning and natural language processing have contributed to the ability of conversational agents to process user queries and produce relevant responses to the query [11]. Yet, the majorities of the current systems are unemotional and concentrate mostly on completing the tasks but do not pay much attention to the emotional condition of the user. This leads to a situation where interactions are bereft of intimacy and do not work well in emotionally sensitive situations. Even human communication is not devoid of emotions, and facial expression is one of the important non-verbal clues of the affective state[9]. The identification of these cues plays a very vital role in developing conversational systems that are viewed as compassionate and trusting. Affective computing is meant to address this gap and allow machines to detect and react to human emotions [3],[11]. Recent developments in Convolutional Neural Networks (CNNs) have enhanced the performance of facial emotion recognition [6], but most solutions are tested offline and lack optimism to perform in real-time related to behaviors [4] on mobile devices.

In addition, a lot of conversational agents utilize the use of text-based sentiment analysis only, which gives only a partial picture of user emotions. Non-verbal messages, especially facial expressions can deliver information that is emotional and not presented in the text. This shows why more practical conversational systems involving emotion recognition and adaptive response generation are required to be efficient and without compromising user privacy.

In dealing with these issues this paper introduces AffectAI, an emotion aware conversational system that integrates real-time facial emotion recognition through the use of cellular conversational systems. The given system consists of a CNN-based model optimized with the help of TensorFlow Lite and applied in on-device mode to Android-based systems. This research aims at creating an effective emotion recognition model that can be deployed in real time, the integration of emotion awareness into converse responses, and measuring the functionality of the system and its usage. AffectAI will increase the naturalness and quality of human computer interaction, using affective intelligence as one of its tools.

II. LITERATURE SURVEY

Recent years have seen the popularity of emotion recognition and affect aware systems because of their capability to improve Human Computer Interaction. [1],[14]. Initial studies in emotion-recognition had typically used manually constructed features based on face-image or speech-wave samples, then some standard machine-learned classifiers. Although these methods proved to be viable, they could not work well in the real world because of the difference in pose, illumination, and expression strength [10],[17]. As deep learning has developed, the Convolutional Neural Networks (CNNs) are now the popular method of recognizing facial emotions [4],[5]. It has been demonstrated by several studies that CNN-based models can learn discriminative markers of the face automatically and have a higher level of accuracy on standard datasets like FER-2013 and Affect Net [14]. Such models are very effective compared to traditional techniques, but most of them are computationally expensive and reviewed in offline or desktop setup and hence cannot be accomplished in real time and mobile based networks. The research in conversational AI has primarily been directed to enhance language comprehension and generation response on transformer based models[7]. Although these models produce fluent and contextually relevant responses, the majority of conversational agents are emotionally neutral or they only use text based sentiment analysis. Text based sentiment methods are able to capture the linguistic polarity but are often inaccurate of the real emotional state of the user especially where non verbal feedback exists in opposition to textual information. The current solutions are usually inefficient in their way of deployment, emotional flexibility, or privacy-related concerns of the user. This is the reason why AffectAI is created as CNN based facial emotion recognition is used with an adaptive conversational model, which is optimized to achieve real-time on-device interaction.

III. IMPLEMENTATION

AffectAI implementation is interested in development of a feasible real time emotion conscious conversational system which integrates facial emotion recognition with the mobile based implementation. The system implementation is grounded on a deep learning architecture and tuned to be used on Android with minimal latencies, privacy and feasibility in real world environments. This overall implementation plan will entail the preparation of datasets, model development, optimization of model and its integration with Android applications.

A. Data collection and preparation

The dataset was collected and preprocessed using different methods to obtain the results reported in the article.

The primary dataset charges the FER-2013 including face pictures in the grey scale and the labels of seven simple categories of feelings, one of which is the feeling of anger, another is the feeling of disgust, the third is the feeling of fear, the fourth is the feeling of happiness, the fifth is the feeling of sadness and the sixth one is the feeling of surprise and neutral. To enhance the question of generalization and cope with the diversity of real world faces, additional samples of facial expression were examined making available bigger scale data sets such as Affect Net [3], where images were captured via the web under the entire range of luminance and pose and expression dimensions. The face detection process is accomplished with the help of openCV version of Haar cascade and face detectors (made using deep learning)[16] in order to get the appropriate facial area forwarded to an emotion classifier. They also employ the technique of data augmentation including horizontal flipping, rotation and brightness changes as a way of mitigating over-fitting as well as improving the model with respect to real-time changes in the environment.

B. Model Design using CNN

A Convolutional Neural Network (CNN) is created in a manner that is trained automatically to discriminate the discriminative features of the face in the input image. The Convolutional layers that follow the model network are multiple followed by ReLU functions of activation and max-pooling layers to extract spatial hierarchies of face features.

These pulled features are made susceptible to completely interconnected dense-layers followed by a Softmax output layer that forecasts the likelihood distribution of emotions. The model will be trained with the help of categorical cross-entropy loss and Adam optimizer. To find out the effectiveness of the classification of various emotions the monitoring of the performance measures of accuracy, precision and confusion matrix in the course of training is conducted.

C. Training and Evaluation on the Model.

The training of the model is done under the Python system using TensorFlow. It is divided into training, validation and testing to estimate the performance without bias level of the evaluation. It has been demonstrated that the trained CNN converges steadily after a number of epochs and the data augmentation and regularization strategies have been supplied to demonstrate improvements.

Measurement will be done on general classification measures like accuracy and precision. Confusion matrix analysis has shown that those emotions that are readily perceived such as happy and surprise are more likely to be recognized as compared to the subtly known emotions such as fear and sadness that show moderate overlap that also matches that found in the past facial emotion recognition research works.

D. Optimization of the Model, TFL Conversion.

To mimic real-time on-the-fly inferences on the mobile based processor, trained CNN model is transferred to the format of the TensorFlow Lite. This is what makes AffectAI be able to directly identify emotions on the device therefore enhancing response rates and preventing the loss of user information.

E. Android App. Integration

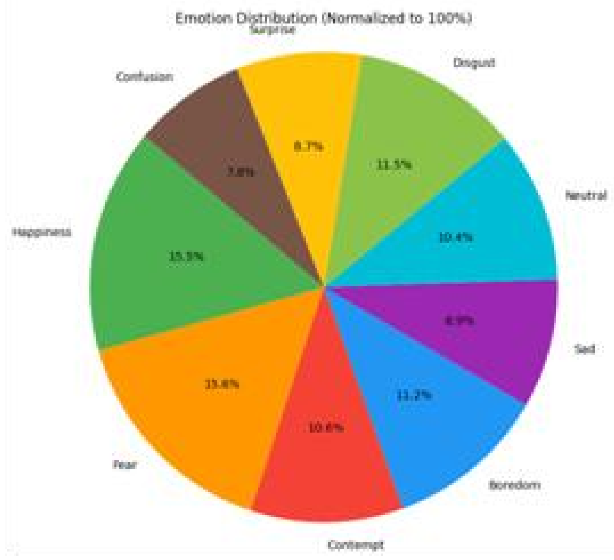
The final model which is built in TensorFlow is integrated with the AffectAI Android application. CameraX API to facilitate access to camera captures faces live using the front facing camera. These prediction results are furnished to the user as the type of emotion (such as Sadness) and their strength.

IV. RESULT AND ANALYSIS

The current section is used to demonstrate the proposed AFFECTAI facial emotion recognition system performance evaluation. The evaluation is done on real time deployment output, confidence behavior, and the temporal trend.

A. Categorizing Behavior and Model Analysis.

The trained Convolutional neural network (CNN) model was tested on several emotions, namely: happiness, fear, disgust, sadness, neutral, surprise, contempt, boredom and confusion. The confusion matrix, which is produced during the testing, demonstrates that there are major correct classifications on the diagonal axis and means that the model does manage to differentiate between several major emotional states[9].



[Fig. 1 Various emotions]

The ability to recognize emotional involvement is stronger in the strong expressive emotion like happiness and fear. There is moderate inter-class overlap in the subtle emotions of neutral, sadness and confusion. This is a pattern in line with the consistency of facial expressions with subtle emotions demonstrating similar muscle movements with lower level of variations.

The classification behavior generally suggests good behavior on feature learning and relatively wide functionality to separate the classes in the trained model.

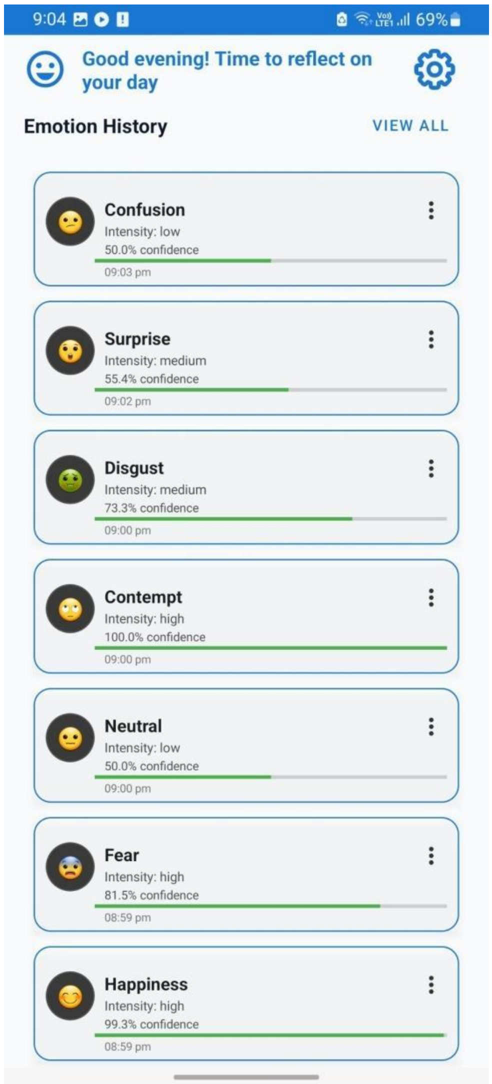
B. Real-Time Detection Performance of Emotion.

It introduced the CNN model to AFFECTAI mobile application to confirm the actual performance. The system makes predictions on predominant emotions and the level of confidence and intensity scores.

According to the outputs obtained:

- Happiness was also characterized by high levels of confidence of about 98-99% which is high intensity.
- Similar performance was exhibited by fear whereby the values of confidence were very close to 99 percent.
- There was a mixed result in Contempt where the levels of confidence ranged between moderate (~67) and very high (100).
- Disgust and Boredom were also identified with moderate levels of confidence comprising about 70-73.
- Confidence produced by Neutral, Sadness, and Surprise with all values within the range of 50 to 66 usually.
- There was a relatively low likelihood of confusion at 50%.

[Fig. 2 Real-time Emotion detection system performance]



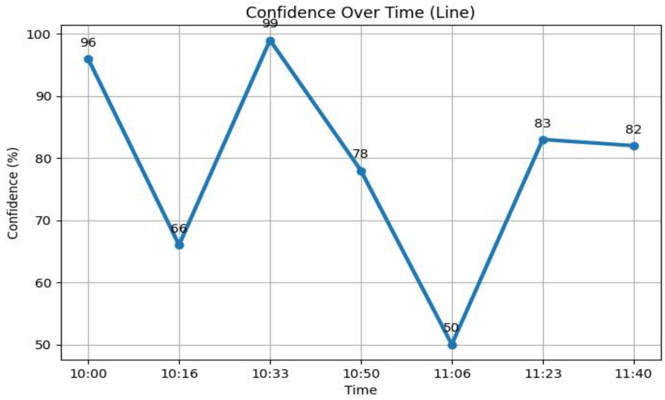
[Fig. 3 Real-time Emotion detection history]

According to these findings, the system works most reliably with strong and visually distinct emotional parses and less expressive or subtle states cause more complexity in classification.

C. Overtime Confidence Analysis.

In order to test the time stability, prediction confidence values were observed through the continuous time interval. The confidence-over-time graph proves that the model behaves extremely well when utilized in the real time.

The peaks of high confidence are associated with vivid emotional expressions whereas minor declines are witnessed during the switching of emotional states. Notably, there were no drastic changes and uncertain forecasts. This implies that CNN model can be called temporally coherent and the results should not be erratic during live inference.



[Fig. 4 Confidence over time(Line chart)]

The stability of the system at the dynamic conditions is supported by its temporal stability.

D. Confidence Distribution analysis.

The visualization analysis of the predictive confidence by the histogram and density curve show the most common range of distribution is the medium to the high range (around 60%-100%). The predictions are only realized in a little percentage of the 55%.

Such a distribution pattern implies that the trained model is capable of providing predictions that are trusted in most real-time situations. Higher percentage of greater confidence levels shows good extraction of features and classification power of the CNN architecture.

V. CONCLUSION

The aim of this paper was to design and test a practical, real-time facial emotion-recognition system that could be operated successfully inside a mobile application. The development of AFFECTAI shows that a well-balanced convolution neural network can prove to be successful in identifying various expressions of emotion while maintaining the same level of confidence over time. The outcome of the approach seems to show good results in detecting well expressed emotions, and at the same time, comparatively subtle expressions show areas for further refinement. This system goes beyond point prediction by combining classification, temporal tracking and visual analytics and tracks emotional trends over time.

The contribution of this work is to reduce the gap between the theoretical models of emotion recognition and the real-world implementation. Unlike many of the existing systems that focus on accuracy metrics, this framework has put more priority on usability, interpretability, and ongoing monitoring. Future research might be able to improve performance through multimodal integration, dataset enlargement, and minimum model optimization for better scalability. Human emotions will continue to be crucial to understanding and recognizing, as intelligent systems continue to evolve. This work is to build digital systems that are more intuitive and responsive to the emotions of humans.

FUTURE WORK

The current implementation of AFFECTAI suggests that there is a foundation for real-time emotion recognition, but there are some improvements that can make its performance and applicability even stronger. Improving the accuracy of detection will test multimodal emotion fusion, using facial curves, speech tone, physiological and user behavioral data. Better understanding of contextual awareness & depth conversation, System can be extended with transformer based language models (BERT and GPT variants).

Further research work will focus on privacy-preserving personalization, i.e., federated learning and adapt on-device training, to ensure data security in the face of improved performance at an individual level. Being able to integrate into healthcare, education, mental well-being platforms gives the opportunity of long term emotional trend analysis. As well as extending Multilinguistic and cross-cultural capabilities.

REFERENCES

[1] V. S. Manjula and L. D. S. S. Baboo, “Face detection identification and tracking by PRDIT algorithm using image database for crime investigation,” *Int. J. Comput. Appl.*, vol. 38, no. 10, pp. 40–46, Jan. 2012.

[2] W. Wang, Y. Jie, J. Xiao, L. Sheng, and D. Zhou, “Face recognition based on deep learning,” in *Proc. Int. Conf. Hum. Centered Comput.*, 2014

[3] Mollahosseini, Ali, Behzad Hasani, and Mohammad H. Mahoor. "Affectnet: A database for facial expression, valence, and arousal computing in the wild." *IEEE transactions on affective computing* 10.1 (2017).

[4] Li, Y., Zeng, J., Shan, S., & Chen, X. (2018). Occlusion aware facial expression recognition using CNN with attention mechanism. *IEEE transactions on image processing*, *28*(5).

[5] Barsoum, E., Zhang, C., Ferrer, C. C., & Zhang, Z. (2016, October). Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM international conference on multimodal interaction*.

[6] Tian, Chunwei, Jingyuan Xie, Lingjun Li, Wangmeng Zuo, Yanning Zhang, and David Zhang. "A Perception CNN for Facial Expression Recognition." *IEEE Transactions on Image Processing* 34 (2025).

[7] Xue, F., Wang, Q., & Guo, G. (2021). Transfer: Learning relation-aware facial expression representations with transformers. In *Proceedings of the IEEE/CVF International conference on computer vision*.

[8] Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., & Mihalcea, R. (2019, July). Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th annual meeting of the association for computational linguistics*.

[9] Barrett, Lisa Feldman, et al. "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements." *Psychological science in the public interest* 20.1 (2019).

[10] Li, Lixiang, et al. "A review of face recognition technology." *IEEE access* 8 (2020): 139110-139120.

[11] Chowdhary, Chiranji Lal, ed. Multidisciplinary Applications of Deep Learning-Based Artificial Emotional Intelligence. IGI Global, 2022.

[12] Følstad, Asbjørn, et al. "Future directions for chatbot research: an interdisciplinary research agenda." *Computing* 103.12 (2021).

[13] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." *IEEE signal processing letters* 23.10 (2016).

[14] Uniyal, S., & Agarwal, R. (2024, December). Analyzing facial emotion patterns in AffectNet with deep neural networks. In *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*.

[15] Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017).

[16] Jaiswal, A., Raju, A. K., & Deb, S. (2020, June). Facial emotion detection using deep learning. In *2020 international conference for emerging technology (INCET*).

[17] Verma, Amit, et al. "Face recognition: a review and analysis." *Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021* (2022).