Developing an Empathetic Interaction Model for Elderly in Pandemics

Human-Centered Approach to Addressing Social Isolation

P Sathya Narayan
U.G Student,Computer Engineering
College of Engineering Chengannur,IHRD
Kerala, India
ORCID : 9587-0962
sathyanmbpk@gmail.com

Safad Ismail
M-Tech, Asst.Professor, Dept of Computer Engineering.
College of Engineering Chengannur,IHRD
Kerala, India
ORCID : 5048-1920
sanu9225thadathil@gmail.com

***Abstract*—This paper presents research on integrating artificial emotional intelligence in a social robot (Sugunan) and its effectiveness in engaging older adults. The empathic Sugunan utilises a multimodal emotion recognition algorithm and a multimodal emotion expression system, detecting user's emotional state and generating an appropriate response through an effective dialogue manager. To facilitate interaction with Sugunan, an android app is used as an interface, utilising the mobile phone's inbuilt camera and microphone for facial expression and speech sentiment detection. The captured data is processed in the cloud, enabling the model to be deployed remotely. The study aims to create an empathetic interaction model for senior citizens , particularly during pandemics like COVID-19 to prevent feelings of isolation.**

***Index Terms*—Deep Neural networks, ResNet,Artificial Intelligence,Sentiment Analysis,SAR**

I. Introduction

Embedding artificial intelligence (AI) with emotional intelligence in the form of empathy could revolutionize how we interact with machines. Empathy is the ability to understand and share another's feelings, and when incorporated into AI, it can facilitate new perspectives and more human-like interactions. This matter is especially pertinent given the worldwide circumstances in 2019, when the elderly experienced mounting isolation as a result of the COVID-19 outbreak.

To address this issue, we present recent research on the integration of artificial emotional intelligence in a social robot named Sugunan. Sugunan is a socially assistive robot designed to provide companionship for older adults with depression and dementia through conversation. In this work, we developed an empathic version of Sugunan which utilizes a multimodal emotion recognition algorithm and a multimodal emotion expression system. It uses different input modalities for emotions, such as facial expression and speech sentiment, to detect user's emotional states and generate an appropriate response.

To implement this system, we have utilized a modified version of the Deep Neural Network called Residual Neural Network. The dialogue processing task was accomplished using Program R. To facilitate interaction with Sugunan, we have deployed this model in the cloud and used an android app as an interface to talk to Sugunan. The app utilizes the mobile phone's inbuilt camera and microphone for facial expression and speech sentiment detection. The captured result is then sent to the cloud for processing and the appropriate response is generated.

The following scenario illustrates how a SAR can be developed and what tools are required to accomplish this. Let us consider Kunjappan, an elderly individual who lives alone. As Kunjappan usually does not have anyone to converse with, he often feels lonely, which is especially detrimental to an elderly person. Fortunately, Kunjappan is accompanied by our emotionally intelligent robot Sugunan. An excerpt from the conversation between Kunjappan and Sugunan is provided below.

Sugunan: "Kunjappan, what's your day looking like?" [*Preventative conversation starts by robot*]
Kunjappan: "I'm doing fine Sugunan." [*User responds, but looks sad*]
Sugunan: "Are you sure? But you're not smiling." [*The robot tries to make the user talk about his feelings*]
Kunjappan: "Maybe a joke would cheer me up." [*The user acknowledges that he is sad and asks for help.*]

As shown in this example, there are a number of components that can contribute to the development of a friendly robot. In a proactive manner, Sugunan inquires about Kunjappan's wellbeing. When a human-oriented robot initiates a conversation with a user, it is useful for the robot to be able to discern the length of time the user has been in the room. It is likely that a user has been alone for a considerable period of time. If the robot detects that the user has spent a significant amount of time in their room, then the user has probably not had many social interactions during that period. As part of the conversation, the robot should be capable of engaging in a spoken dialogue with the user [8]. It is displayed in the example above that the robot uses Sentiment Analysis (SA) and Facial Expression Recognition (FER) in order to detect a discrepancy between Kunjappan's response and his facial expression. Emotional intelligence requires a multimodal emotion perception system [9]. To improve Kunjappan's mood, the robot decides to tell a joke. This means that the robot needs multiple channels to express emotional information.

II. SUGUNAN, AN EMOTIONALLY INTELLIGENT ROBOT

In order to recognize and understand human emotions, Sugunan uses a multimodal emotion recognition algorithm trained using AffectNet Dataset to recognize and understand facial expressions and speech sentiments.

AffectNet is a large-scale, multi-label, and continuously annotated dataset of facial expressions, valence, and arousal ratings, which was developed for training and evaluating algorithms for facial expression recognition and affect estimation. It contains more than one million images, with each image labeled with a set of 11 basic emotions, valence, and arousal values. The emotions labeled in the dataset include happiness, sadness, anger, fear, surprise, disgust, contempt, valence and neutral, with valence indicating the positivity or negativity of the emotion, and arousal representing the intensity of the emotion. This dataset was developed through a combination of crowdsourcing and manual annotation techniques, making it one of the largest and most comprehensive datasets on facial expressions available currently. There has been a significant use of the AffectNet dataset for developing and evaluating facial expression recognition algorithms, research in affective computing, psychology and neuroscience.

There are several models of emotions in the literature [7], where Russell's [6] and Ekman's [5] are the most common models used in HRI studies [2], [4]. For the purpose of measuring emotional facial expression, we utilize Russell's dimensional model. The use of a RGB camera on a mobile phone enables Sugunan capture 10 images per second. Every image is fed into a face detector which utilizes the Viola Jones algorithm [22]. We then crop the identified face and feed it into a deep neural network (DNN) running on the cloud for facial expression recognition (FER). Based on the results of the FER algorithm, three categories of emotion can be determined: Positive, Neutral and Negative.

As the user's facial expression may change multiple times during the conversation with the robot, our FER algorithm returns 10 estimated values per second for the user's facial expression. It is likely that the last frame before the user stops speaking is not the best candidate to represent their facial expression at this time. Consequently, a misclassification could occur. In certain circumstances, the output of the FER system may be inaccurate, such as when a user is blinking, yawning or covering their face. Using the data from the last 30 frames, we are able to avoid noise and create a more stable method of measuring emotional status (Fig. 2). Nonetheless, we assign higher weights to the more recent frames in order to make the algorithm more sensitive to recent changes in the subject's facial expression. For each new frame, the value (-1, 0, +1) is added to the list and the oldest frame is removed.

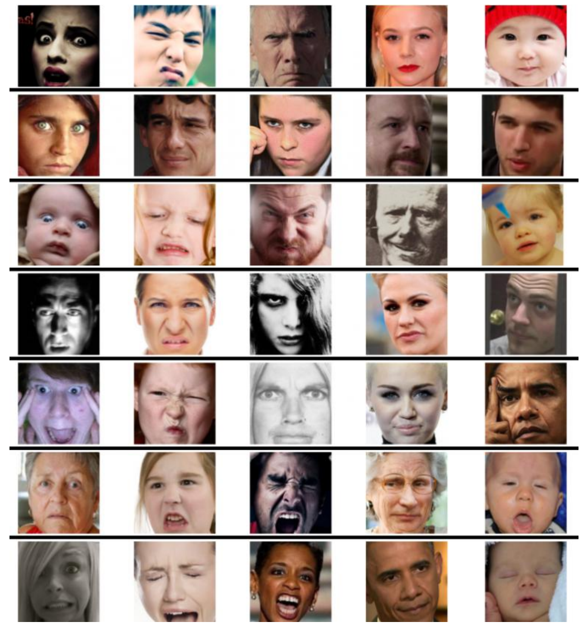

Fig. 1: AffectNet Dataset

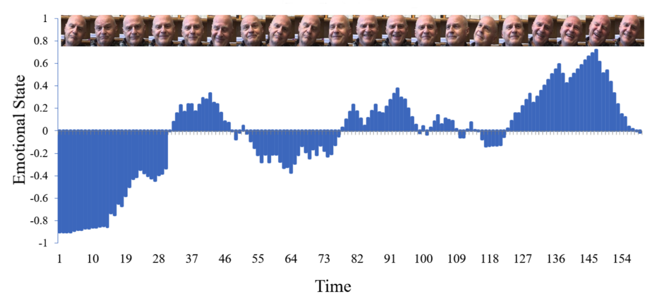

Fig. 2: In this chart [1], the horizontal axis represents time and the vertical axis represents emotional state, with a range of -1 (Negative) to +1 (Positive)

By deploying the model in the cloud, data collected from the mobile app interface, which uses the inbuilt camera and microphone on the mobile device to capture the user's facial expressions and speech sentiment, can be processed efficiently. The captured result is then sent to the cloud for processing, and the response is generated using an effective dialogue manager.

Known as template-based dialogue systems, Program-R is an artificial intelligence-based dialogue system that combines visual and textual information and responds to users accordingly. In contrast to most dialogue systems, Program-R is an active agent, meaning it initiates the conversation and engages the user in a controlled conversation. XML-based AIML [18] is used by different chatbots such as Alice [19] to organize the dialogues. A dialogue system using AIML uses Regex matching to determine the best response (responses are stored in AIML's template tag) for every user input utterance. Fig. 7 depicts the architecture of the proposed system.

As part of the proposed Socially Assistive Robot (SAR) architecture, two major modules are presented: the Cloud module and the Android module. The Cloud module comprises the Dialogue Processing module and Advanced Face Recognition module, which is responsible for natural language processing and response generation. On the other hand, the Android module includes the Basic level Face Recognition module. This utilizes the smartphone's built-in camera and microphone to capture the user's facial expressions and speech sentiments.

Once the data is processed, the Dialogue Processing module generates appropriate responses based on the user's emotional state as shown in Fig. 4.

Preprocessing involves the pretreatment of the raw text so that unnecessary punctuation is removed, normalized, and it will be segmented by sentence type. In order to arrive at a score, the sentiment of the text is combined with the output from the Facial Expression Recognition module (Emotional State) making use of the Sentiment Analysis module. By making use of the Brain's Question Handler, we can ensure that the context, sentiment and information related to the conversation will be factored by using the Context Manager and thus we are able to ensure that we are taking into account the conversation context. As a result of the information provided by the Context Manager, as well as the value calculated based on the Emotional State and Sentiment, the Question Handler creates a response based on the information provided. Upon completion of the process, the selected answer is sent to the Answer Handler and Postprocessing module for processing in a further step which are then sent back to the Android module.

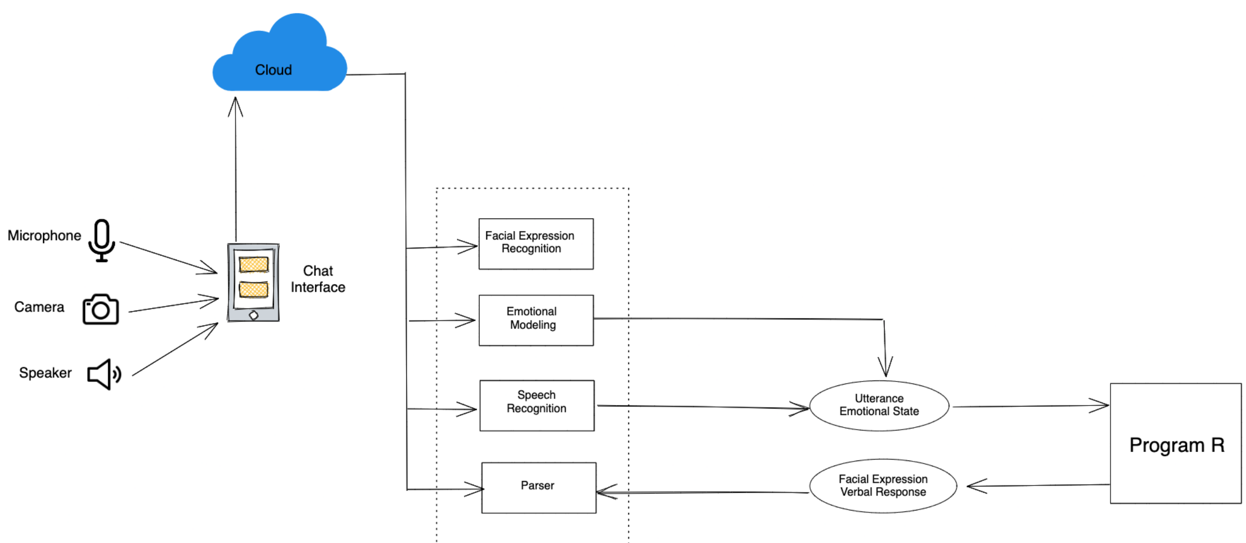

Fig. 3: Architecture of the proposed system

Face Recognition module is used to capture the user's image and audio data, which is then sent to a Cloud module for further processing. For the purpose of recognizing facial expressions, a modified version of the Deep Neural Network called the Residual Neural Network (ResNet50) [17] is applied to the data. Speech sentiment analysis is carried out with the aid of a Multimodal Emotion Recognition algorithm. This algorithm integrates audio features such as pitch and intensity with linguistic features in order to detect emotions within the speech of the user.

The Android module receives the response from the Cloud module and displays it on the mobile interface, providing an interactive platform for the user to communicate with the SAR.

The proposed architecture (Fig 3) leverages cloud technology to enable efficient processing of data and generation of responses, while the Android module provides a user-friendly interface for the older adults to interact with the SAR. The integration of facial expression recognition and speech sentiment analysis allows the SAR to detect the user's emotional state and respond accordingly, providing empathetic and engaging companionship to older adults with depression

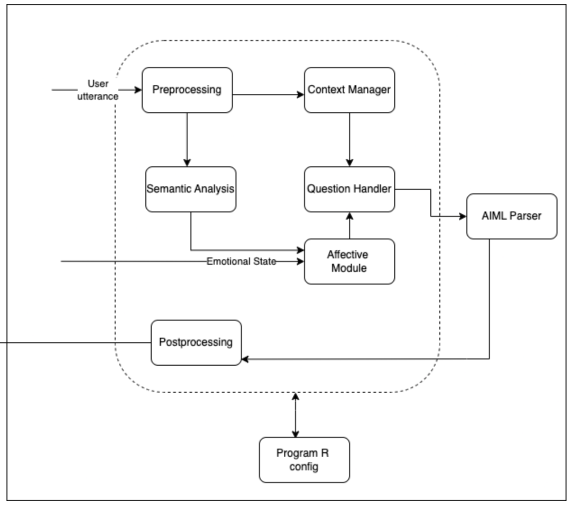

Fig. 4: Framework of Program R

and dementia.

III. APP INTERFACE

In terms of the mobile interface, the first interaction that a user has with it is the login page that appears when they first launch the app. Authorized users must be authenticated

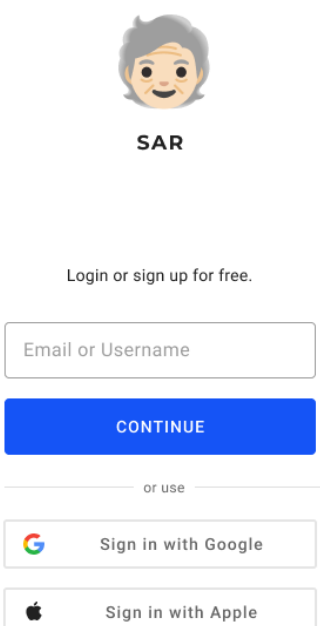

Fig. 5: Login Interface of Sugunan

in order to access the system. A form will appear once the user reaches this page asking for his or her user name and password.

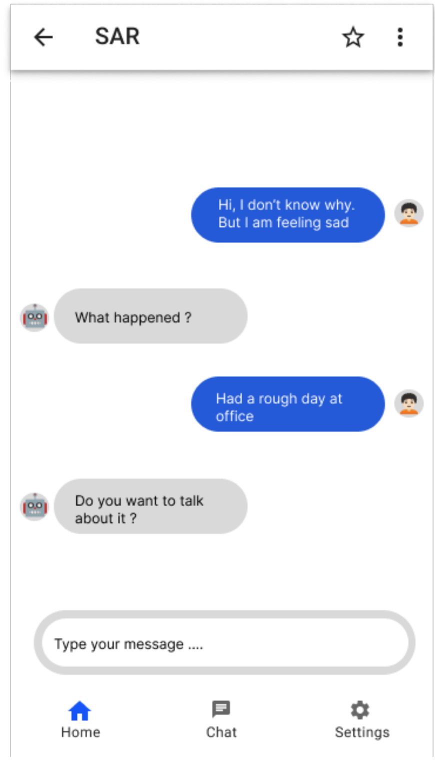

Fig. 6: A Chat Interface with Sugunan.

The second page of the mobile interface is the chat page, where the user can interact with the SAR Sugunan. The chat interface is designed to be user-friendly and intuitive, with a clean and simple design.The interface consists of a chat window where the user can type in their messages and view the responses from Sugunan. The chat window is designed to look like a typical messaging app with a white background and blue bubbles for the user's messages and gray bubbles for Sugunan's responses. The chat window also includes an avatar of Sugunan, which appears on the left-hand side of the screen. The avatar is designed to look friendly and approachable, with a smiling face and expressive eyes. The user's own avatar appears on the right-hand side of the screen, along with their messages.

Overall, the interface is designed to be easy to use, with emphasis on the dialogue between the user and Sugunan, and is generally designed to be simple and easy to use. The design of the SAR is intended to create a friendly and approachable atmosphere that will assist the user in feeling more comfortable as he or she interacts with it.

The above conversation depicts a sample interaction between a user and the socially assistive robot (Sugunan) using the mobile interface.

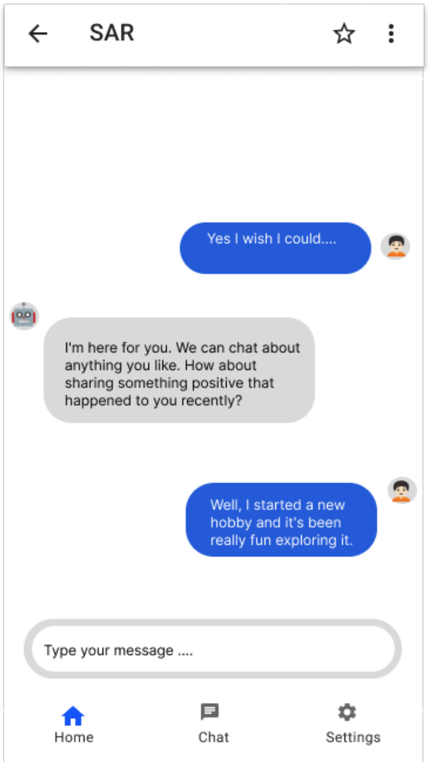

Fig. 7: Illustrates Sugunan's understanding level towards the context of the text

Fig. 8: Illustrates Sugunan handling human conversation

IV. CONCLUSION

The presented SAR with an empathetic interaction model has shown promising results in engaging older adults and providing them with companionship through conversation. The integration of emotional intelligence in the form of empathy has opened up newer perspectives in the field of human-robot interaction.

In the future, this model can be expanded to include a wider range of emotions and become more adaptable to individual user preferences. Additionally, the system can be improved to provide more individualized care and therapy to older adults, especially those who feel isolated as a result of global pandemics such as COVID-19.

It is also possible to extend this system in order to provide companionship and therapy for people of all ages who feel lonely or isolated for a variety of reasons. With the advancement of technology, SARs are becoming more widely used as therapeutic tools. In addition to addressing the mental health concerns of individuals who cannot access traditional therapy for a variety of reasons, the development of such systems can help to address other mental health issues.

It can be concluded that the presented SAR that has an empathetic interaction model, when extended to all people, will be able to become an effective and scalable approach to addressing the mental health concerns of individuals who feel lonely or isolated in the world.

REFERENCES

[1] Abdollahi, H., Mahoor, M., Zandie, R., Sewierski, J. and Qualls, S., 2022. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. IEEE Transactions on Affective Computing.

[2] M.Szabo ova ,M.Sarnovsky',V.MaslejKres˘n˘a kova ,andK.Ma- chova , "Emotion analysis in human–robot interaction," Electronics, vol. 9, no. 11, p. 1761, 2020.

[3] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1–1, 2017.

[4] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinc˘a k, and P. Dario, "Emotion modelling for social robotics applications: a review," Journal of Bionic Engineering, vol. 15, no. 2, pp. 185–203, 2018.

[5] P. Ekman and W. Friesen, The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, 1978.

[6] J. Russell, "A circumplex model of affect," Journal of personality and social psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[7] D. Sander, "Models of emotion: The affective neuroscience ap- proach." The Cambridge Handbook of Human Affective Neuroscience, 2013.

[8] R.Nielsen,R.Voyles,D.Bolan os,M.Mahoor,W.Pace,K.Siek,and W. Ward, "A platform for human-robot dialog systems research," in AAAI Fall Symposium: Dialog with Robots, 2010, pp. 161–162.

[9] R. W. Picard, Affective computing. MIT press, 2000.

[10] Papadopoulos, Fotios, et al. "Designing an empathic social robot for older adults: An exploratory study." International Journal of Human-Computer Studies 128 (2019): 68-81.

[11] Hong, Kai, et al. "Empathetic conversational system for elderly care." 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2018.

[12] Lee, Juheon, et al. ”Emotional conversation generation with artificial intelligence for a social robot.” International Journal of Humanoid Robotics 15.06 (2018): 1850051.
[13] Wang, Fei, and Tao Zhang. ”A Survey on Empathetic Social Robots.” Robotics 7.3 (2018): 47.
[14] Wang, Fei, et al. ”Empathetic service robots for depression management: A review.” Journal of Healthcare Engineering 2018 (2018): 1-18.
[15] D. Banerjee, M. Rai et al., “Social isolation in covid-19: The impact of loneliness,” International Journal of Social Psychiatry, vol. 66, no. 6, pp. 525–527, 2020.
[16] A. Adams, J. Beer, X. Wu, J. Komsky, and J. Zamer, “Social Activities in Community Settings: Impact of COVID- 19 and Technology Solutions,” Innovation in Aging, vol. 4, no. Suppl 1, p. 957, Dec. 2020. [Online]. Available: https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC7741293/
[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[18] R. Wallace, “The elements of aiml style,” Alice AI Foundation, vol. 139, 2003.
[19] R. S. Wallace, “The anatomy of alice,” in Parsing the Turing Test. Springer, 2009, pp. 181–210.
[20] Liu, Q., Gu, J., Yang, J., Li, Y., Sha, D., Xu, M., Shams, I., Yu, M. and Yang, C., 2021. Cloud, Edge, and Mobile Computing for Smart Cities. Urban Informatics, pp.757-795.
[21] Uddin MZ, Soylu A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. Sci Rep. 2021 Aug 12;11(1):16455. doi: 10.1038/s41598-021-95947-y. PMID: 34385552; PMCID: PMC8361103.
[22] P. Viola and M. J. Jones, “Robust real-time face detection,” International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.