Industrial Engineering Journal
ISSN: 0970-2555
Volume: 53, Issue4, April : 2024

Emotional Mood Enhancer Using Multi Layer Perceptron

[1]K Bilwa Shankaran, [2] Sreya Pothala, [3]N Balaji, [4]G Shanmukha Naidu
[5]S Venkata Lakshmi,[6]MKalyani.

[1,2,3,4] Student, Department of Computer science & Engineering with specialization of Data Science, Dadi Institute of Engineering & Technology, Vizag, Andra Pradesh, India.
[5]Assistant Professor, Department of computer Science &Engineering, DadiInstitute of Engineering & Technology, email: venkatalakshmisalapu@gmail.com
[6]Assistant Professor, Department of computer Science &Engineering, DadiInstitute of Engineering & Technology, email: mallakalyani1@gmail.com

ABSTRACT:

In today's fast-paced world, finding stable emotional connections has become increasingly challenging, contributing to the rise of mental health issues like schizophrenia, depression, anxiety, and loneliness. To address this issue stemming from our increasingly mechanical lifestyles, our project offers a solution by providing emotional support and companionship. We achieve this by analyzing voice tone and speech expressions to better understand and respond to emotions. At the heart of our project lies a vast collection of emotional data, including speech patterns and textual cues. This extensive dataset serves as the foundation for training a deep neural network. We utilize cutting-edge algorithms such as Multi Layer Perceptron (MLP) and transformer models to delve deeply into the complexities of human emotions. By leveraging pre-trained Transformer models, we can provide personalized interventions tailored to individual emotional needs.

Index Terms: Emotion recognition, personalized interventions, interactive sessions, Supervised learning Techniques, Pre-trained Transformer

INTRODUCTION

In an era dominated by the relentless pace of modern life, characterized by constant connectivity and demanding schedules, our project, the Emotion Mood Enhancer, emerges as a pioneering response to the profound challenges individuals face in maintaining emotional well-being. Recognizing the intersection of our digital existence with emotional health, our initiative seeks to bridge the gap between the demands of the fast-paced world and the fundamental human need for emotional connection.

At the core of our project lies a commitment to leveraging advanced deep learning techniques. It transcends conventional approaches by not only detecting emotions but also introducing a novel dimension— personalized response generation. In the face of stress and emotional turbulence, the Emotion Mood Enhancer aspires to provide tailored responses that uplift and comfort individuals, offering a digital companion in times of need. The belief that understanding and responding to human emotions is not just an aspiration but an indispensable necessity forms the guiding principle of our endeavor. In a society increasingly intertwined with technology, this project represents a crucial step towards harmonizing our digital interactions with our emotional states.

Embarking on this transformative journey with us means exploring the potential for the Emotion Mood Enhancer to revolutionize the way we engage with and enhance our emotional well-being. Beyond a mere technological innovation, it stands as a testament to our commitment to creating a future where technology serves as a supportive force in fostering brighter and more resilient lives. As we navigate the complexities of modern existence, join us in envisioning a world where the Emotion Mood Enhancer contributes to a more balanced, connected, and emotionally enriched human experience in the digital age. Together, let's embrace the possibilities of a future where technology and emotional intelligence converge to empower individuals on their journey toward emotional well-being.

BACKGROUND

A component of machine learning, which draws inspiration from the composition and operations of the human brain, is deep learning. In order to learn hierarchical data representations, multilayer neural networks, also known as deep neural networks, are used. In order to identify human emotions, deep learning in emotional recognition uses sophisticated algorithms to assess a variety of data sources, including text content, voice tones, and facial expressions. Convolutional Neural Networks (CNNs) analyze visual cues such as facial features, while Recurrent Neural Networks (RNNs) capture temporal patterns in sequential data like speech. Multi-modal fusion combines information for a holistic understanding. In response generation, Generative Models and Transformer Models create contextually relevant and diverse outputs, fostering more emotionally resonant interactions.

MLP: Neural networks that can process intricate patterns in data are known as multilayer perceptrons, or MLPs. A network of nodes that typically consists of an input layer, one or more hidden layers, and an output layer makes up an MLP. Because each component of a layer has weighted connections to every other layer, the network can identify and depict nonlinear relationships in the data. Each component of the output has an activation function applied to it, which displays nonlinearities crucial for capturing intricate patterns. Using backpropagation, which is the process of propagating parameters back through the network, the network modifies these weights throughout training in order to minimize the discrepancy between the predicted and actual output.MLP excels in many areas of machine learning, such as classification, regression, and pattern recognition, making it the foundation of artificial intelligence. MLP, with its versatility and ability to learn relationships in data, is still an important choice for solving many problems in many areas of the world.

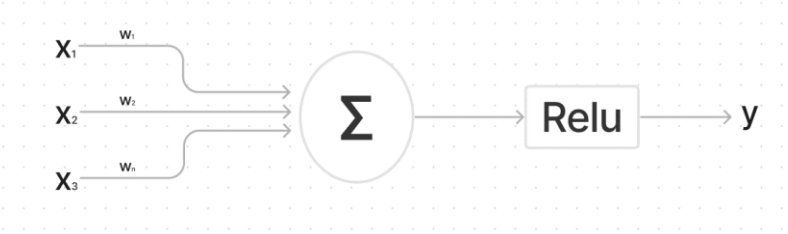

Fig 1: Perceptron neuron model

Data Set: The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) served as the foundation for our project's emotional speech analysis research. The dataset gave us a strong basis for training and testing our models, with 24 experienced actors providing a wide range of emotions, including happiness, sadness, anger, fear, surprise, and neutral states. (RAVDESS) There are 7356 files in all, weighing in at 24.8 GB. Two comparable lines from the literature on American idioms are acted out by 24 actors—12 men and 12 women. Peace, joy, joy, grief, wrath, fear, surprise, disdain, etc., can all be expressed in words or sentences. There are two thought intensity levels (strong and normal), and an intermediate expression is made for every expression. Three modal formats—audio-only (16 bit, 48kHz.wav), audio video (720p H.264, AAC 48kHz,.mp4), and video-only (480p H.264, 264, AAC 48kHz,.mp4)—are available to meet all needs. The inclusion of both monologue and dialogue modes, as well as singing samples, allowed us to explore various facets of emotional expression in audio. RAVDESS became an integral part of our project, enabling us to develop and fine-tune algorithms for emotion recognition and response generation. Its widespread use in the research community underscores its significance as a benchmark dataset for evaluating the efficacy of models in understanding and categorizing emotions in audio signals. Our reliance on RAVDESS exemplifies its role as a valuable resource, contributing to the success and depth of our work in the realm of emotional intelligence and technology.

LITERATURE SURVEY

Numerous research works have been done on the recognition of emotions from auditory information. In order to identify emotions in speech, one must first extract the relevant features from a corpus of emotionally charged speech that has been chosen or used. Then, emotions are categorized based on the qualities that have been retrieved. A successful characteristic extraction process is critical to the classification of emotions' performance. Numerous features, including spectral and prosodic characteristics as well as combinations of these features (e.g., the MFCC acoustic feature and the energy prosodic features in [30]) were recovered by the researchers during their investigation.

We need a mathematical model that describes emotions in order to be able to categorize them using computer algorithms. Psychologists have created a standard approach that is based on three measures that together form a three-dimensional space that encompasses all emotions. Pleasure, arousal, and domination are these metrics or dimensions [39, 40]. We can report the most pertinent feeling based on a vector that is in one of the established emotion areas, which is created by a combination of these features [31].
We can characterize nearly any feeling using pleasure, arousal, and dominance; but, implementing such a deterministic system for machine learning will be highly intricate. Consequently, we usually employ statistical models and cluster samples into one of the designated qualitative emotions, like happiness, sadness, rage, and so on, in machine learning investigations. In order to categorize and group any of the aforementioned emotions, we must first model them using features taken from the speech; typically, this involves taking several prosody, voice quality, and spectral featurecategories [32] .While certain emotions are better classified by any of these categories, others are harder to identify. Prosody characteristics that are typically concentrated on fundamental frequency (F0), speaking rate, duration, and intensity are unable to accurately distinguish between happy and angry emotions [32]. Features of voice quality typically have a greater influence when identifying the emotions of the same speaker. However, because they vary from speaker to speaker, it is challenging to employ them in a situation where there is no specific speaker [33]. A great deal of analysis has been done on spectral patterns to extract emotions from speech. Their initial advantage over prosody features is their ability to tell the difference between furious and cheerful with confidence. An emotion detection system would need to be speech content-aware because of the issue that the magnitude and shift of the formants for the same emotions varied across various vowels [34].

A flexible emotion identification system based on the examination of both visual and aural inputs was presented by Noroozi et al. Using Principal Component Analysis (PCA) in feature extraction, the researcher employed 88 features (Filter Bank Energies (FBEs) and Mel Frequency Cepstral Coefficients (MFCC)) in order to lower the dimension of previously derived features [2].Using the

Berlin Emotional Speech database, Bandela et al. employed the GMM classifier to detect five emotions by the merging of an acoustic feature, the MFCC, with a temporal energy operator (TEO) as a prosodic [1].

Natural language understanding (NLU) has advanced significantly over the past ten years thanks in large part to Deep Learning approaches. Proposed by Kim et al. [5] and Zheng et al. [6], Deep Belief Networks (DBN) for SER demonstrated a noteworthy enhancement over baseline models [3] [4] that do not utilize deep learning. This implies that high-order non-linear interactions are more suitable for emotion detection. In order to determine utterance level emotions, Han et al. [7] presented a DNN-Extreme Learning Machine (ELM), which combines a single hidden layer neural net with utter-ance-level features from segment-level probability distributions. However, accuracy improvements were not very significant. Fayek and associates.

The feed-forward structures that comprise one or more underlying hidden layers between inputs and outputs are the foundation of Deep Neural Networks (DNNs). For processing images and videos, feed-forward architectures like Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) yield effective results. Conversely, voice-based classification tasks like natural language processing (NLP) and speech recognition (SER) benefit greatly from the use of recurrent architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) [11].MLP is utilized for many applications, including as regression and classification, and is non-deterministic and universal in the sense that it can adequately estimate any continuous nonlinear function on a compact interval[8, 9]. The neurons in an MLP are organized in layers, starting with the input layer and going all the way up to the output layer through hidden layers. The network is feedforward and allows for connections between the two adjacent layers [10].

[12] put forth a SER system that records emotional content from different speaking modalities using both visual and audio modalities. Za-deh along with others. [13] presented a Tensor Fusion Network that is perfect for the erratic nature of online language since it learns both intra- and intramodality dynamics end-to-end. In order to obtain good accuracy, Ranganathan et al. [14] experimented with Convolutional Deep Belief Networks (CDBN), which learn significant multimodal characteristics of utterances.

The authors of [15] made the assumption that people might communicate their emotions in different ways, and that Speech Emotion Recognition (SER) ought to be trained using the speaker's identity as a basis. Therefore, one of the main contributions of [15] is that the authors have used a key-query-value attention called Self Speaker Attention (SSA) to condition emotion classification to speaker identity. This allows one to focus on the emotion-relevant portions of an utterance by computing both self and cross-attribute attention scores (the relationship between speaker identity and emotions). [15] employed a three-channel input 3-D Log-Mel spectrogram for feature parameters. Deltas and delta-deltas, which can be thought of as approximations of the first and second derivatives of the first channel, are the second and third channels, respectively, while the static of the Log-Mel spectrogram from 40 filter banks makes up the first channel. A 10-fold cross-validation method was used for assessments. In [15], there was no data augmentation.

While deep CNNs utilizing spectrograms are presented in [18], deep hierarchical models, data augmentation, and regularization-based DNNs for SER are proposed in [19]. utilizing a probabilistic CTC loss function, DNNs are trained for SER utilizing the acoustic features that are retrieved from the brief speech intervals in [20]. Compared to DNN-ELM [16], the bidirectional LSTM-based SER in [21] obtains higher accuracy after being trained on feature sequences. Even better results are obtained with the Deep CNN+LSTM-based SER in [22]. The SER accuracy is increased by the hybrid deep CNN + LSTM, but the overall computing complexity is increased. A SER based on auditory-visual modality (AVM) in [23] is capable of capturing emotional content from various speech modalities.
Intra- and inter-modality dynamics are learned by the Tensor Fusion Network (TFN)-based SER in [24]. Multimodal feature representations associated with expressions are learned by the convolutional deep belief network-based SER in [25]. Due to the loss of certain fundamental sequential information during the convolutional process, the single plain CNN model performs poorly when it comes to accurately identifying the speaker's emotional state. Consequently, the problem of speech loss of significant information can be resolved by using two CNN models operating in parallel. Two parallel CNN models are demonstrated in the work in [26], and SER makes use of them appropriately.

One may almost characterize all feelings using dominance, pleasure, and excitement; nevertheless, it is highly difficult and complex to create such a deterministic system using DL. As a result, in DL, emotions like happiness, rage, and melancholy are qualitatively categorized using statistical models and data clustering. Speech features must be retrieved in order to classify and cluster emotions. Typically, this involves using spectral features, voice quality, and prosody of various kinds [27]. Prosody characteristics, which typically consist of the fundamental frequency (F0), intensity, and speaking rate, are unable to reliably distinguish between the emotions of anger and happiness. The characteristics of a speaker's vocal quality are typically the ones that work best at expressing their emotions. It is challenging to apply these properties in speaker-independent situations, nevertheless, because they differ between speakers [28]. However, spectral patterns are frequently employed to extract emotional information from speech. These characteristics can confidently discriminate between happy and rage. The speech emotion identification system is more difficult since the shifts and magnitudes of the formant frequencies for the same emotions vary across vowels[29].

METHODOLOGY

The Emotional Mood Enhancer is a two-phase project designed to create a sophisticated system that not only accurately recognizes users' emotional states through speech but also responds to them with personalized and empathetic interactions. By combining advanced deep learning techniques with real-time emotional analysis, the system aims to enhance users' emotional well-being and provide a supportive virtual environment.

4.1 Phase 1: Emotion Recognition through Speech, the main objective of this phase is to develop a robust emotion recognition system focusing on speech analysis to accurately identify and understand users' emotional states. The approach for this phase is

I. **Feature Extraction:** Implement state-of-the-art deep learning techniques to extract crucial features from speech, including pitch, tone, and rhythm. This involves preprocessing raw audio data and transforming it into meaningful representations for subsequent analysis.

II. **Model Training:** Using a synergistic combination of Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs), our training methodology makes use of a variety of datasets, such as the extensive RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song). This deliberate combination of CNNs and MLP allows the model to go through various layers of the perceptron in addition to facilitating the investigation of complex patterns in the dataset. This meticulous strategy ensures that our model is immersed in a rich tapestry of emotional expressions and speech patterns. By exposing the model to a wide spectrum of emotional nuances, our training methodology guarantees a robust understanding and extraction of features across various layers of the perceptron architecture, thus enhancing the model's capability to discern and interpret diverse emotional cues effectively.

III. **Validation and Testing:** Rigorously validate the model's accuracy by assessing its performance across various speech patterns and emotions. Thorough testing is conducted to ensure reliability and generalization to real world scenarios.

4.2 Phase 2: Personalized Emotional Responses, the main objective of this phase is to develop a responsive system capable of crafting personalized and empathetic responses based on the detected emotional states. The approach for this phase is

I. **Emotional Profiling:** aiming to establish a robust baseline for personalized emotional analysis by delving into the intricate aspects of vocal tones and nuances. This multifaceted process involves a meticulous examination of the user's speech patterns to create a comprehensive understanding of their emotional landscape. our system endeavors to capture the nuances that reflect the user's emotional states. This nuanced emotional profiling forms the foundation for tailoring the system's responses in a highly individualized and adaptive manner, ensuring that our technology can engage with users on a more profound and empathetic level.

II. **Deep Learning Response Generation:** Utilize transformer models, known for their contextual understanding and language generation capabilities, to craft context-aware and emotionally resonant responses. This ensures that the system's responses are not only relevant but also emotionally appropriate.

III. **Diverse Intervention Strategies:** Implement a variety of interventions based on the user's emotional state. This includes words of encouragement, guided relaxation techniques, and even customizable messages delivered in familiar voices to enhance the overall user experience.

IV. **User Feedback Loop:** Establish a continuous feedback loop by incorporating user feedback. This iterative process allows the system to learn and adapt to individual preferences and further refines the emotional response generation model over time.

ARCHITECTURAL DESIGN

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 53, Issue4, April : 2024

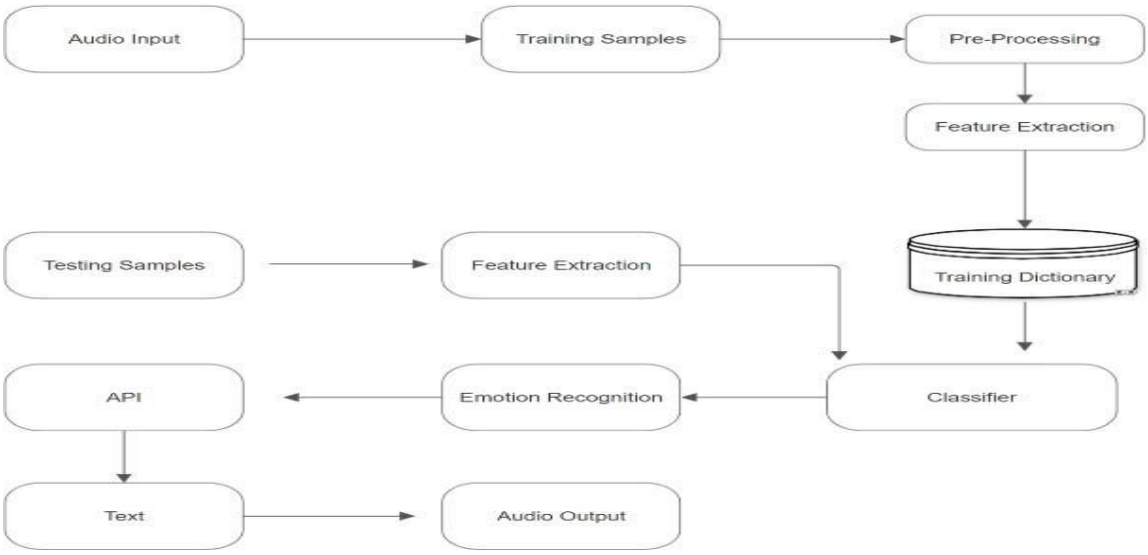

Fig 2: Architectural design

We have developed a web interface that interacts with users through speech and provides responses tailored to their emotions, aiming to improve their well-being.

1. Users provide audio samples as input to the system.
2. The system processes this audio speech to create a dataset for training.
3. The audio data undergoes preprocessing to remove noise and unwanted segments, ensuring the quality of the dataset.
4. Relevant features such as pitch, tempo, and intensity are extracted from pre-processed audio data. These features serve as input for the mood classification model.
5. A classifier, such as an ML model, is trained using extracted features and corresponding mood labels. Once trained, audio will enhance our mood.
6. The system sends the extracted features and mood recognition results to an external API for further processing.
7. The API processes the input and generates text-based output, which is then returned to the system for display.

RESULTS

Our interface is designed to accommodate both account holders and non-account holders. Account holders enjoy certain benefits, suchas access to their conversation history. Moving to the interaction page, users can engage with the system through either text chat or speech audio by simply clicking on the record button. Notably, the system's capability to recognize and respond to emotions is exclusive to interactions involving audio. This feature adds a layer of depth and personalization to the user experience.

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 53, Issue4, April : 2024

Fig 3: welcome Page

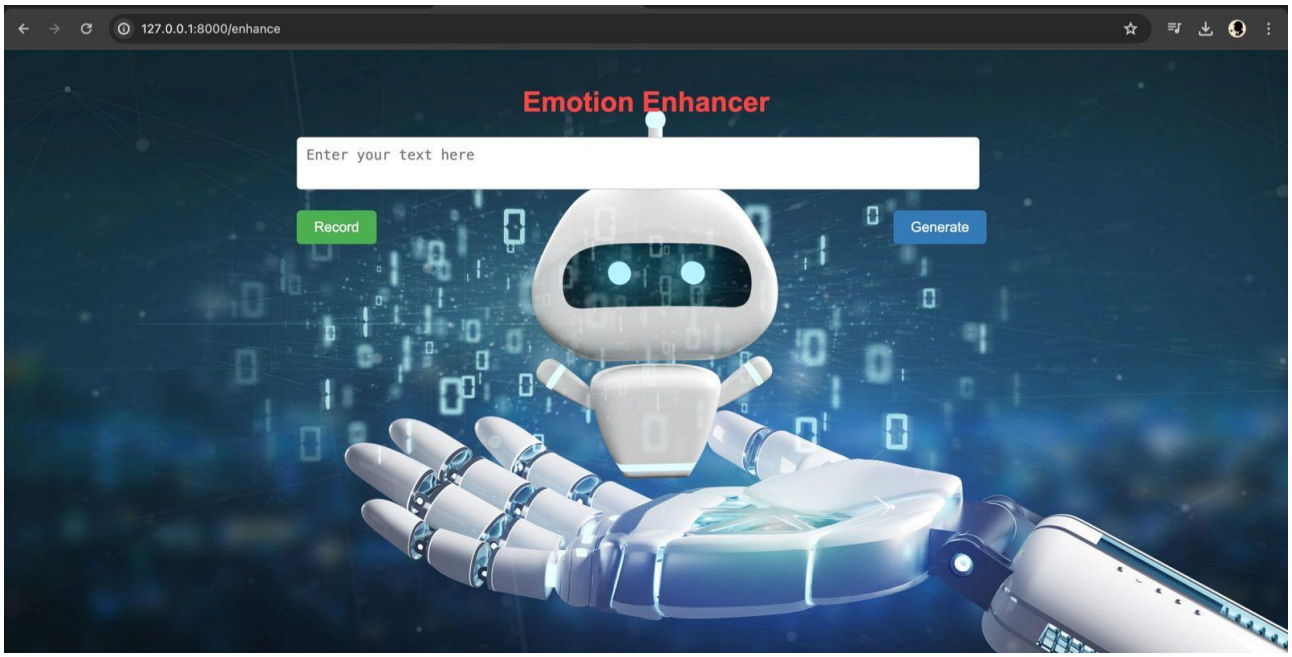

Fig 4: Interactive page

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 53, Issue4, April : 2024

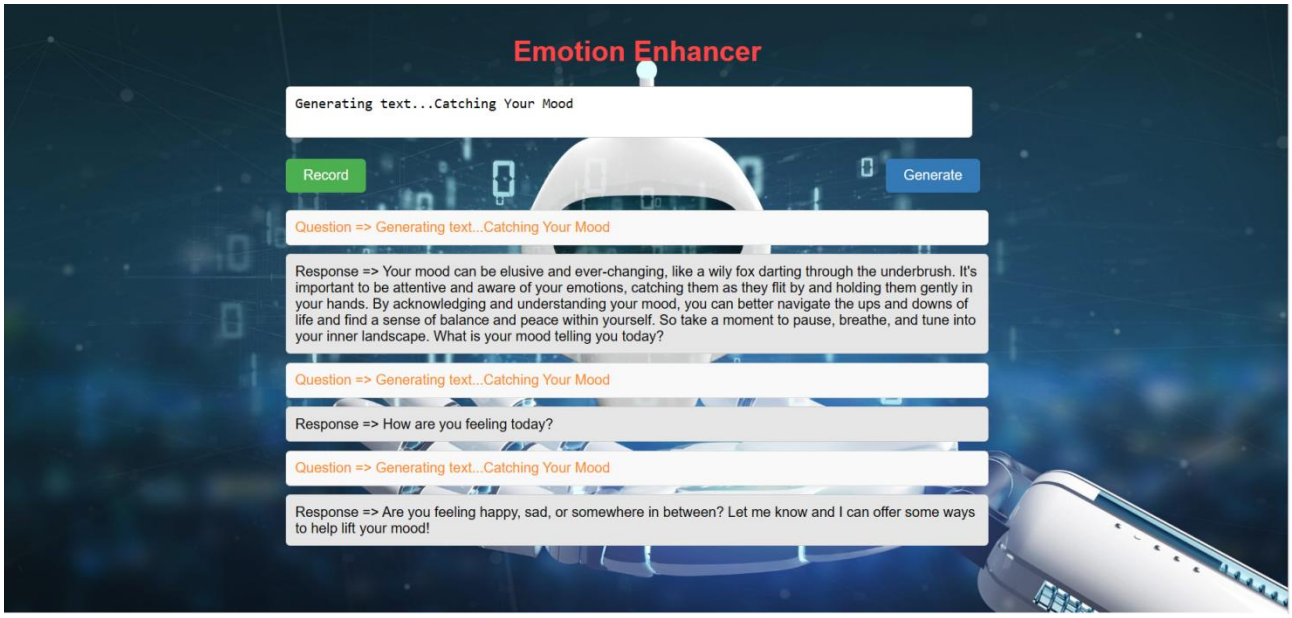

Fig 5: Response Generation based on the mood

CONCLUSION

We worked on a project called the Emotion Mood Enhancer. Our goal was to make a cool system that uses Deep Learning technology to make people feel better emotionally. First, we talked about what we wanted to achieve with the project, which was to help people feel happier in today's world. We used smart computer techniques to figure out how people are feeling and to give them helpful responses right away. Throughout our work, we outlined the objectives of the project, which were centered around addressing the increasing need for personalized emotional support systems in today's society. Through extensive testing and evaluation, we have demonstrated the effectiveness and reliability of the Emotion Mood Enhancer system in accurately recognizing responses in real-time. Our efforts involved a thorough examination of existing emotion recognition systems, where we identified shortcomings and proposed innovative solutions to overcome these challenges. Through extensive testing and evaluation, we have demonstrated the effectiveness and reliability of the Emotion Mood Enhancer system in accurately recognizing emotions and providing meaningful support to users.

ACKNOWLEDGEMENT

I would like to express our sincere gratitude to professor Mrs S.Venkata Lakshmi for her invaluable guidance and her expertise and insightful feedback have greatly contributed to the success of this work.

REFERENCES

[1]Bandela, S. R. and Kumar, T. K. (2017) "Stressed speech emotion recognition using feature fusion of teager energy operator and MFCC," in 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-5.
[2]Faroozi, F., Anbarjafari, G., Marjanovic, M., Njegus, A., and Escalera, S. Emotion identification in audio-visual footage. Affective Computing Transactions, IEEE, 2017.
[3] Jin, Q., Li, C., Chen, S., Wu, H.: Using acoustic and lexical characteristics for speech emotion recognition. 2015, pp. 4749–4753, IEEE International Conference on Acoustics, Speech, and Signal Processing.
[4] Wu, C. H., Liang, W. B.: Acoustic-Prosodic Information and Semantic Labels for Multiple Classifiers in the EmotionRecognition of Affective Speech. published in: IEEE Transactions on Affective Computing 2.1 (2011), pages 10–21.
[5]. Provost, E. M., Lee, H., and Kim, Y.: Robust feature generation in audiovisual emotion recognition using deep learning. (2013), pp. 3687–3691, IEEE International Conference on Acoustics, Speech, and Signal Processing.
[6] Zheng, W. L., Zhu, J., Peng, Y.: Deep belief networks for EEG-based emotion classification. 1-6 (2014) in IEEE International Conference on Multimedia & Expo.
[7] Han, K., Yu, D., Tashev, I.: Deep neural network and extreme learning machine for speech emotion recognition. In: 2014 INTERSPEECH.
[8] "Multilayer perceptron, fuzzy sets, classification," S. K. Pal and S. Mitra, 1992.
[9] "Learning representations by back-propagating errors," D. E. Rumelhart, G. E. Hinton, and R. J. Williams, nature, vol. 323, no. 6088, 1986, pp. 533-536.
[10] A comparison research using support vector machine and artificial neural network (MLP, RBF) models for river flow prediction was conducted by M. A. Ghorbani, H. A. Zadeh, M. Isazadeh, and O. Terzi in Environmental Earth Sciences, vol. 75, no. 6, p. 476, 2016.

[11]The article "Deep learning in neural networks: An overview" was published in January 2015 in Neural Netw., vol. 61, pp. 85-117.
[12] Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B., Zafeiriou, S.: Deep neural networks for end-to-end multi-modal emotion recognition. published in 2017 in the IEEE Journal of Selected Topics in Signal Processing.
[13]Tensor fusion network for multimodal sentiment analysis, Zadeh, A., Chen, M., Poria, S., Cambria, E., Morency, L.P. [13]. In: EMNLP (2017).
[14] Ranganathan, H., Chakraborty, S., & Panchanathan, S.: Deep Learning Arc for Multimodal Emotion Recognition.
[15]Clément Le M., Obin N., Roebel A. Speaker Attentive Speech Emotion Recognition; Proceedings of the International Speech Communication Association (INTERSPEECH); Brno, Czechia. 30 August–3 September 2021. [Google Scholar]
[16] Guan, H.; Guo, L.; Wang, L.; Dang, J.; Liu, Z. investigation of supplementary characteristics for kernel extreme learning machine-based voice emotion identification. 2019, IEEE Access 7, 75798–75809. [Scholar Google] [Cross Reference]
[17] Han, K.; Tashev, I.; Yu, D. Using an extreme learning machine and deep neural network, speech emotion recognition is achieved. In The Interspeech Proceedings, Singapore, September 14–18, 2014. [Scholar Google]
[18]Panda, A.; Kopparapu, S.K.; Tiwari, U.; Soni, M.; Chakraborty, R. Speech emotion recognition in noisy environments utilizing generative noise model with multi-conditioning and data augmentation. 7194–7198 in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2020—2020, held in Barcelona, Spain, May 4–8, 2014. [Scholar Google]
[19] Speech emotion identification from spectrograms using a deep convolutional neural network, Badshah, A.M.; Ahmad, J.; Rahim, N.; Baik, S.W. In the proceedings of the 13–15 February 2017 Busan, Republic of Korea International Conference on Platform Technology and Service (PlatCon); pp. 1–5. [Scholar Google]
[20] Yang, X., and Dong, Y. Affect-salient event sequence modelling for continuous speech emotion recognition. 2021; Neurocomputing 458, 246–258. [Scholar Google] [Cross Reference]
[21] Huang, G.; Chen, Q. A new dual attention-based BLSTM for speech emotion recognition using hybrid characteristics. Artif. Intell. Eng. Appl. 2021, 102, 104277. [Scholar Google] [Cross Reference]
[22] \engür, A.; Atila, O. A 3D CNN-LSTM model with attention guidance is used to accurately identify emotions in speech. 2021; Appl. Acoust., 182, 108260. [Scholar Google] [Cross Reference]
[23] Kreifelts, B.; Wildgruber, D.; Lambert, L. Impact of sensory modality and emotional category on gender differences in emotion detection. 2014, Cogn. Emot. 28, 452–469. [Scholar Google] [Cross Reference]
[24] Ishi, C.T.; Ishiguro, H.; Fu, C.; Liu, C. Emotion recognition approach with many modalities and GAT-based multi-head inter-modality attention. 2020, 20, 4894; sensors. [Scholar Google] [Cross Reference]
[25] Wang, Z.; Diao, G.; Chen, L.; Liu, D. Deep belief network-based multimodal emotion identification in speech expression. J. Grid Comput. 19, 22 (2021). [Scholar Google] [Cross Reference]
[26] Researchers Zhao, Li, Zhang, Zhao, Cummins, Wang, H., Tao, J., and Schuller, B.W. combined a self-attention dilated residual network and a parallel 2d convolutional neural network for ctc-based discrete speech emotion identification. 2021; Neural Network, 141, 52–60. [Scholar Google] [Cross Reference]
[27] Yegnanarayana, B.; Kadiri, S.R.; Gangamohan, P. Review of emotional speech analysis. Believable Behaving Systems, Towar. Robot. Soc., 2016, 1, 205–238. [Scholar Google]
[28] Gobl, C.; Chasaide, A.N. Voice quality's function in conveying attitude, emotion, and mood. Speech Communication 40 (2003) 189–212. [Scholar Google] [Cross Reference]
[29]Blasenko, B., Prylipko, D.; Böck, R.; Siegert, I.; Wendemuth, A.; Philippou-Hübner, D. The examination of vowels formants makes it simple to identify high-arousal moods. Selected papers from the 2011 IEEE International Conference on Multimedia and Expo, held July 11–15, 2011, in Barcelona, Spain, pp. 1–6. [Scholar Google]
[30]"Speech emotion recognition using kernel sparse representation-based classifier," P. Sharma, V. Abrol, A. Sachdev, and A. D. Dileep, in 2016 24th European Signal Processing Conference (EUSIPCO), pp. 374-377. [31] M. Albert. Pleasure-arousal-dominance: An all-encompassing framework for characterizing and quantifying individual variances in temperament. Curr. Psychol., 16, 261–292, 1996. [Scholar Google]
[32] Paidi, G.; Yegnanarayana, B.; Kadiri, S.R. Analysis of Emotional Speech: A Survey. Robotics, Society, Belief, Behavior, Syst. 2016, I, 205–238. [Scholar Google]
[33] Gobl, C., and Chasaide, A.N. Voice quality's function in conveying attitude, mood, and emotion. Speech Communication 40 (2003) 189–212. [Scholar Google] [Cross Reference]
[34] Philippou-Hübner, D.; Wendemuth, A.; Prylipko, D.; Vlasenko, B. Analysis of vowels formants makes it simple to identify high-arousal and impulsive emotions. In Proceedings of the International Speech Communication Association's Twelfth Annual Conference, August 27–31, 2011, Florence, Italy; pp. 1577–1580. [Scholar Google]