Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 11s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Chatbot Emotion Recognition & Music Recommendation

Satya Vaishnavi Yanamandra [1], Dr. Y. Md. Riyazuddin[2]

[1]GITAM Deemed to be University, Rudraram, Telangana, India
[2]GITAM Deemed to be University, Rudraram, Telangana, India
Email ID: yanamandravaishnavi@gmail.com
Email ID: rymd@gitam.edu

Cite this paper as: Satya Vaishnavi Yanamandra, Dr. Y. Md. Riyazuddin, (2025) Chatbot Emotion Recognition & Music Recommendation. *Journal of Neonatal Surgery*, 14 (11s), 476-485.

ABSTRACT
Emotion popularity and music suggestion through chatbots make use of deep learning methodologies to improve human-computer interplay thru textual content, vocalization, and facial expressions. This look at gives a multimodal technique that carries Convolutional Neural Networks (CNN), long short-term memory (LSTM), Bidirectional LSTM (BiLSTM), and LSTM-GRU networks, alongside BERT embeddings, to identify emotions throughout various input modalities. The Kaggle Emotion Dataset is employed for training and contains tagged textual statistics that indicates diverse emotional states. Preprocessing methods embody text sanitization and numerical vectorization making use of BERT for characteristic extraction. The assessment of version overall performance is performed by way of accuracy, precision, recall, and F-score measurements. The BiLSTM version attains a most accuracy of 91.84%, illustrating its exceptional functionality in spotting sequential dependencies and contextual subtleties in textual input. A web-primarily based framework is established to permit real-time emotion recognition, allowing customers to have interaction with the chatbot through text enter, facial expressions captured with the aid of a webcam, or voice recordings. A customized music advice system provides tracks that correspond to the user's emotional state based at the diagnosed emotion. This approach improves human-computer connection by means of providing a sensible and responsive emotion-sensitive system.

Keywords: *Chatbot, Emotion Recognition, Music Recommendation, Deep Learning, Multimodal Approach, BERT Embeddings.*

1. INTRODUCTION

Emotion detection is vital for improving human-computer interaction, permitting structures to come across, analyze, and reply to human emotions proficiently. The amalgamation of artificial intelligence with chatbot generation has markedly more suitable person studies via facilitating tailored and dynamic interactions knowledgeable through perceived feelings [1]. This device emphasizes real-time emotion identity across three modalities—textual content, voice, and facial expressions—to signify music that corresponds with the person's emotional nation, as a result enhancing engagement and emotional well-being [2]. The device discerns feelings together with happy, unhappiness, marvel, and neutrality thru the analysis of user inputs, facilitating an enticing and human-like experience [3].

Deep learning methodologies are vital in enhancing the precision of emotion belief throughout numerous modalities. Textual inputs undergo sentiment analysis, using sophisticated natural language processing algorithms to parent emotional signs [4]. simultaneously, vocal tone and pitch are tested to ascertain emotional states, making use of speech emotion popularity fashions for greater precision [5]. Furthermore, facial expressions recorded via webcams are analyzed the use of convolutional neural networks (CNNs) to assess instant emotional variations. The combination of these modalities improves the chatbot's potential to become aware of and react suitably to human emotions [6].

Emotion-based totally song hints facilitate a dynamic and engaging courting through mood-specific song suggestions. The chatbot no longer best identifies emotions but also modifies its responses thus, facilitating an unbroken and sympathetic verbal exchange [7]. This technology improves user engagement and personalizes virtual interactions by way of integrating AI-driven emotion popularity with actual-time chatbot verbal exchange. An internet-based totally interface enables user interaction through text input, voice commands, or facial expressions, rendering the gadget flexible and handy.

The usage of AI in emotion recognition and track notion has been significantly investigated, showcasing its efficacy in improving human-pc interactions. Prior research has applied deep mastering methodologies, together with recurrent neural networks (RNNs) and transformers, to derive considerable functions from multimodal inputs, therefore enhancing recognition precision [2]. Moreover, emotion-conscious chatbot systems were established to beautify consumer happiness, indicating its applicability across numerous domains such as intellectual health, leisure, and customer service [5].

This system seeks to decorate personalized touch via incorporating powerful AI fashions, handing over a shrewd, empathic, and responsive chatbot experience. Utilising a multimodal approach complements accuracy and engagement, consequently confirming AI's role in developing emotionally responsive digital environments [3].

2. RELATED WORK

Numerous academics have investigated the amalgamation of emotion identification and music recommendation structures via synthetic intelligence and deep learning methodologies. Hazmoune and Bougamouza [8] carried out an in depth evaluation of transformer-based multimodal emotion recognition, categorizing current techniques into taxonomies and evaluating the leading methodology. Their research highlighted the efficacy of transformers in capturing contextual subtleties throughout numerous modalities, along with text, audio, and facial expressions, to improve emotion identity. Mathew, Chooramun, and Sharif [9] brought an emotion-driven chatbot music advice system that employs person sentiment evaluation to tailor tune hints. Users could interface with NLP-supported deep-learning technology to enter text which would result in suggested songs. Yuga showed effective accuracy in tracking user emotions while it simultaneously increased overall human-machine interaction engagement.

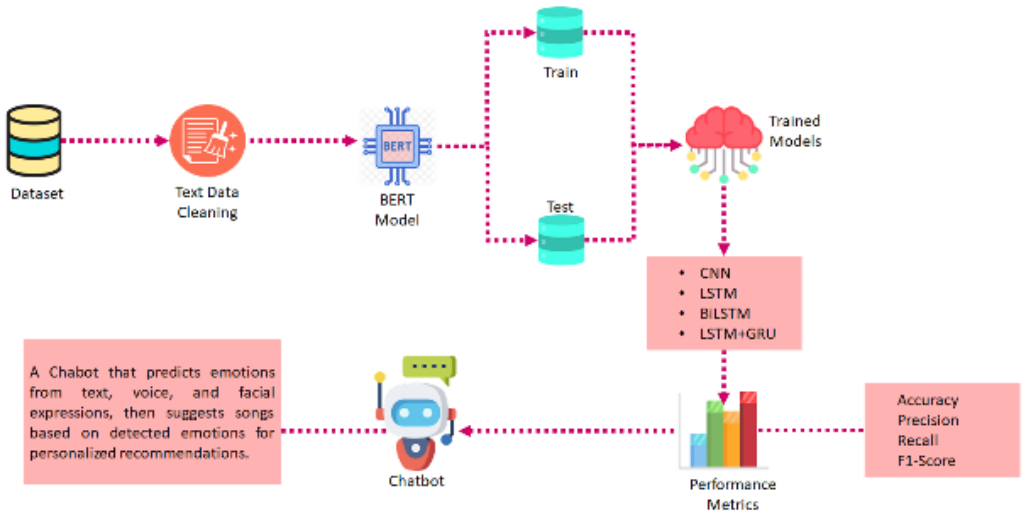
Singhal, Sahu, Jaiswal, Ahmad, and Gaur [10] evolved a track recommendation system employing convolutional neural networks (CNNs) to categorize and suggest songs based totally on consumer sentiment and historical listening patterns. Their method utilized feature extraction from audio alerts and emotion-labeled datasets to improve recommendation accuracy. P. J. R, M. M, R. k, and P. k. G [11] created Chatplayer, a chatbot-based music advice system that makes use of the Spotify API for instantaneous music suggestions. Their research targeting improving suggestion performance through the analysis of textual and speech-derived emotional inputs. Anusha and Srinivasan V. [12] brought a chatbot tune recommendation machine that incorporates sentiment analysis and advice algorithms to tailor user experiences. Their research illustrated the efficacy of AI-driven methods in improving user engagement through emotion-sensitive music selections.

Chaudhary et al. His approach incorporates the spirit analysis from NLP-Push together with knowledge categorization systems to enhance the accuracy of emotional prestige and milk selection. Research stressed that the unit accountability must be supported by numerous decorating methods. The paper published by Ijraset [14] described a track suggestion platform with chatbots as it demonstrated the necessity of real-time evaluation for adapting tracks. He studied various AI applications for extracting functions while developing different types as well as enhancing Council accuracy levels. Suvarna Bahir et al. Among his generation there was wide promotion toward making customers satisfied and adapting to new information.

The research demonstrates how the recommendation system for emotional hair mats develops while user involvement and suggestions indicate how AI and deep learning and NLP enhance user precision.

3. MATERIALS AND METHODS

The proposed system serves as a sophisticated approach for smart emotion detection and music suggestions through combined processing of text data with speech and facial expressions using deep learning methods and many different computational models [1] [2]. A system framework combines CNN and LSTM alongside Bilstm as well as a hybrid model of LSTM-Gru that utilizes improved functionality through Bert extraction [3]. The text input passes through Burt for preaching and vectorization operations before deep learning analyzes speech patterns [4]. The conviction network appointed by facial expression identity extracts symptoms from live webcam pronoun [6]. Through their Internet-based user interface users can report using live videos or audio transmissions or live emotional detection. People receive personalized playlists through technological detection of felt emotions based upon their usage patterns and interactive habits [2] [3].



"Fig.1 Proposed Architecture"

The system structure shown in the image (Fig.1) demonstrates that the chatbot guesses emotions from text, speech, and facial expressions, then recommending songs based at the diagnosed feelings. The dataset's textual statistics is sanitized and input into a BERT model. Emotion recognition employs trained models such as CNN, LSTM, BILSTM, and LSTM+GRU. The chatbot engages with human beings, and its overall performance is assessed using metrics consisting of accuracy, precision, recall, and F1-score. This offers tailored tune tips depending on user emotions.

i) Dataset Collection:

The data series Emotion comprises a compilation of English Twitter posts categorised into six classes: anger, worry, joy, love, sadness, and wonder. The dataset has 16,000 times, presenting attributes: one for textual content statistics and the other for emotion labeling. The dataset, obtained from Kaggle, has been preprocessed in step with the CARER method. It affords a basis for training deep learning models in emotion identity and music recommendation.

	text	label
0	i didnt feel humiliated	0
1	i can go from feeling so hopeless to so damned...	0
2	im grabbing a minute to post i feel greedy wrong	3
3	i am ever feeling nostalgic about the fireplac...	2
4	i am feeling grouchy	3
...	...	...
15995	i just had a very brief time in the beanbag an...	0
15996	i am now turning and i feel pathetic that i am...	0
15997	i feel strong and good overall	1
15998	i feel like this was such a rude comment and i...	3
15999	i know a lot but i feel so stupid because i ca...	0

16000 rows × 2 columns

"Fig. 2 Dataset Collection"

ii) Pre-Processing:

The preprocessing phase ensures superior input for emotion recognition. The manner entails the purification of textual content facts via eliminating superfluous symbols, stopwords, and punctuation. Finally, the BERT model is utilized for feature extraction, reworking text into good sized embeddings for classification based on deep learning.

a) Text Data Cleaning: Text data's adaption process calls for purification procedures to attain versioning's greater precision.

The book becomes lowercased throughout this process, and a struggle with Whitespace starts until all special characters, stop words, and punctuation are eliminated. The book serves as a factor causing significant fragmentations. Lemmatization preserves natural sentence structures, therefore enhancing the extended characteristic extraction techniques. The Bert version has enhancements in preparation that let archaic entrance detection for different emotional classes.

b) BERT Model: Bert (two-way coder representation from transformer) represents an advanced teaching model which research organizations developed for natural language treatment purposes. Bert analyzes sentences through two directions to obtain understanding of the material. Burt completes preschool education utilizing adequate datasets before achieving exceptional capabilities for certain emotional recognition duties. The system handles both symbol and hygiene material processing before turning the information into classification useful data. The detection of emotions shows increased accuracy through this model because it handles extensive linguistic microfies.

iii) Algorithms:

Convolutional Neural Network (CNN)

The CNN statement functions as a deep learning structure specifically designed for text operations through sequence pattern detection. The approach enhances emotion and emotion analysis through regular filters that allow hierarchical operations to simplify accuracy enhancement methods [10] [13].

Long Short-Term Memory (LSTM)

LSTM, a size of the recurrent nerve network (RNN), effectively captures long -distance addiction in sequential text. It reduces the problems with the lack of shield, caused by the use of memory cells and gates, which is effective for real -time popularity by placing the necessary relevant facts [1] [3].

Bidirectional Long Short-Term Memory (BiLSTM)

Examining text from both ways helps car stm improve LSTM performance by leading to mutual addiction across past and present settings. Higher degrees of accuracy in both real-time emotional evaluation and Chatbot exchanges help the system to improve emotional awareness. [4] [5].

LSTM+GRU

The LSTM+GRU model integrates LSTM's long-term memory features with GRU's efficiency, enhancing each precision and computational performance. This hybrid methodology is efficacious for emotion identity, as it equilibrates complexity even as maintaining critical contextual linkages in textual data [2] [8].

4. RESULTS & DISCUSSION

Precision: Precision assesses the proportion of accurately classified cases among those identified as positive. Consequently, the formula for calculating precision is expressed as:

$$"Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (1)"$$

Recall: recall is a metric in machine learning that assesses a version's capacity to understand all pertinent times of a particular class. it is the ratio of accurately expected superb observations to the overall actual positives, offering insights into a model's efficacy in identifying occurrences of a specific class.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} (2)$$

mAP: mean average Precision (MAP) is a statistic for assessing ranking quality. It evaluates the quantity of pertinent tips and their placement within the listing. MAP at k is decided as the arithmetic suggest of the average Precision (AP) at k for all users or queries.

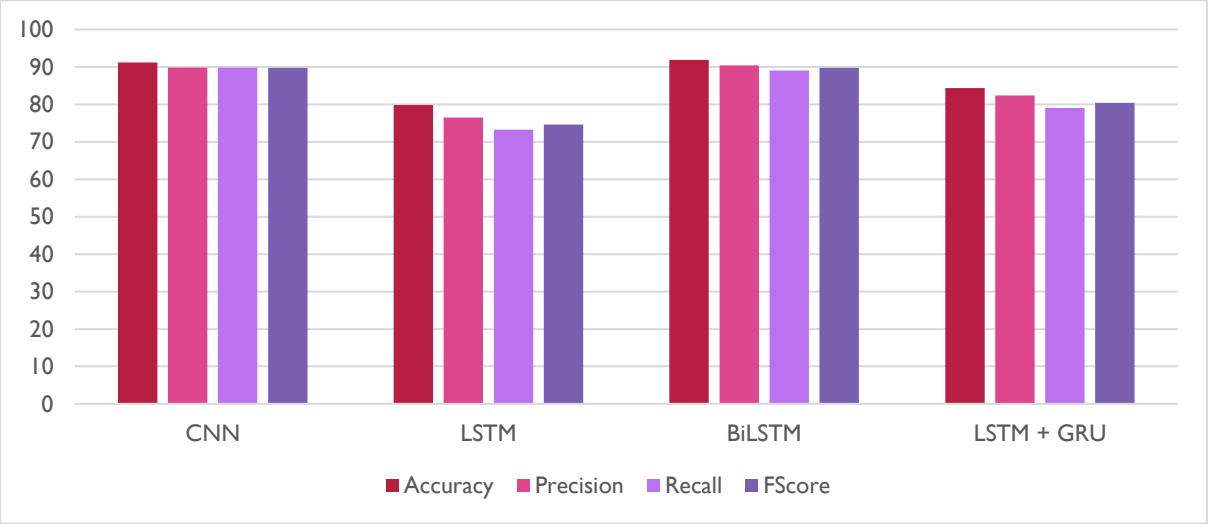
$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k (3)$$

Table (1) assess the performance metrics—Accuracy, Precision, recall, F-rating—for each algorithm. The BiLSTM routinely surpasses all other algorithms across all metrics. The tables provide a comparative exam of the metrics for the alternative methods.

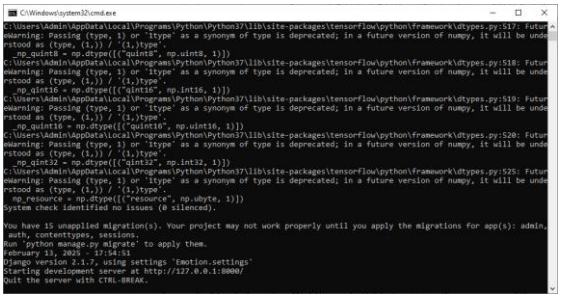
"Table.1 Performance Evaluation Table"

Algorithm Name	Accuracy	Precision	Recall	FScore
CNN	91.18750	89.820577	89.826949	89.751423
LSTM	79.78125	76.482030	73.226682	74.586106
BiLSTM	91.84375	90.427669	89.087627	89.718936
LSTM + GRU	84.31250	82.401636	79.011299	80.369628

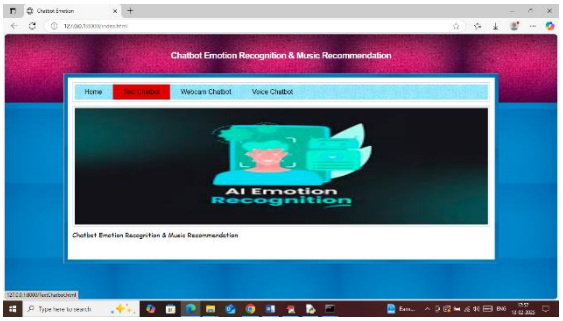
"Graph.1 Comparison Graphs"



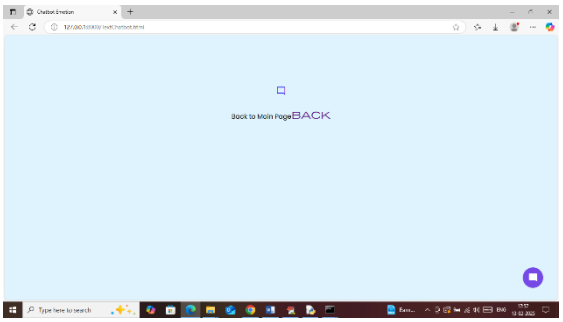
Accuracy is depicted in blue, precision in orange, recollect in inexperienced, and F-score in sky blue in Graph (1). Compared to the opposite models, the BiLSTM demonstrates greater performance across all criteria, reaching the very best values. The graphs above visually constitute those findings.



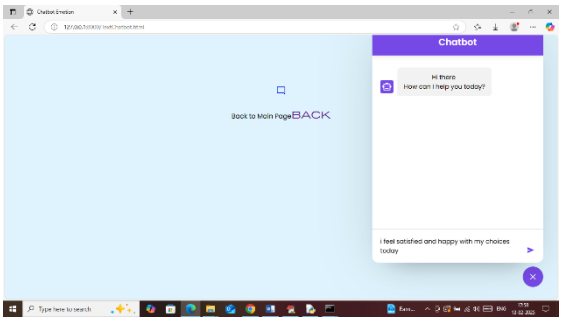
In the previous screen, the Python server has been initiated. Subsequently, open a browser and enter the URL "http://127.0.0.1:8000/index.html", then click on the input key to show the subsequent web page.



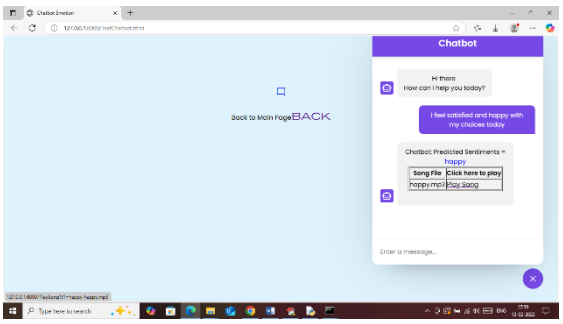
Click on on the 'text Chatbot' link at the top display to get admission to the following web page.



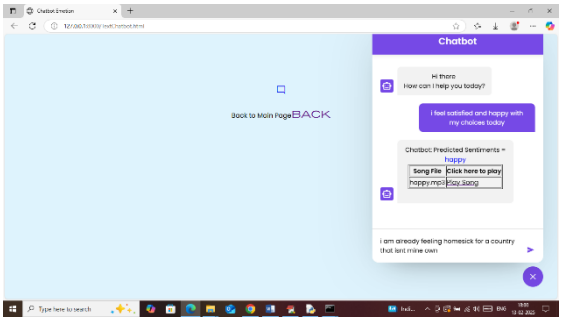
inside the aforementioned display, click on at the 'Blue' icon located at the bottom right to initiate the Chabot and get admission to the subsequent page.



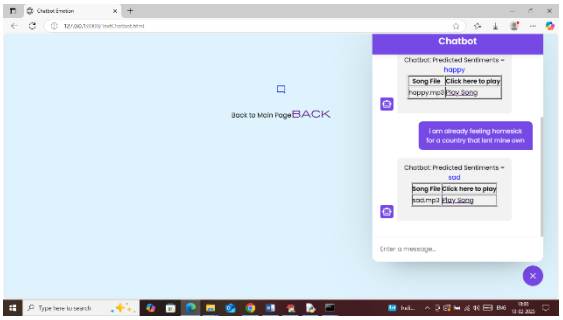
Within the above screen, I input a message and sooner or later hit the arrow icon to gain the prediction result below.



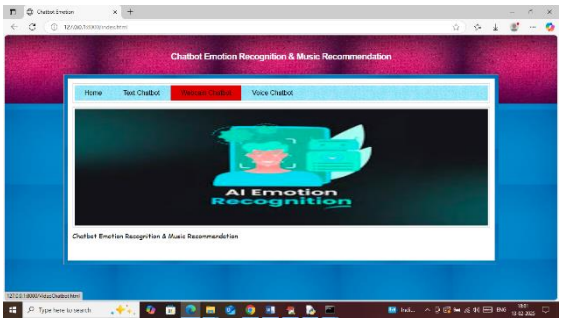
The text displayed above is anticipated as 'happy,' accompanied with the aid of tune guidelines, and you may click on at the 'Play music' link to initiate playback. Currently evaluating an additional sentence.



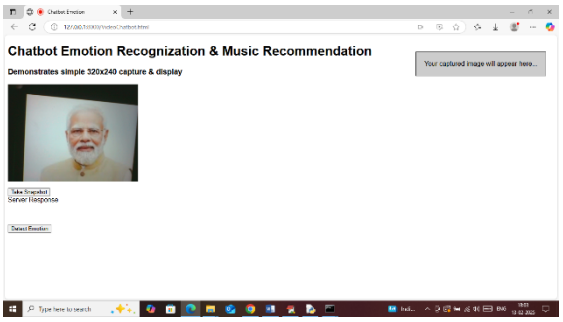
The upper display screen offers a further sentence, followed by the Chatbot's reaction below.



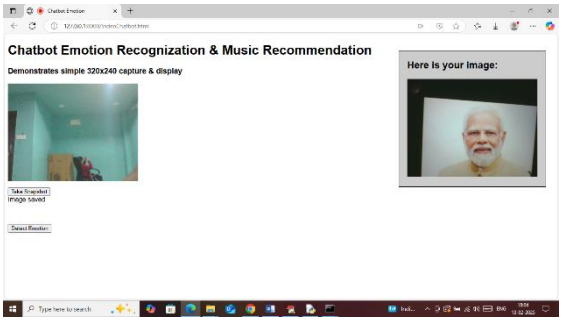
The anticipated emotion displayed above is 'sad'; likewise, enter any textual content to get hold of the corresponding emotion and a suggested track list. Please click on the 'back' hyperlink to return to the main web page.



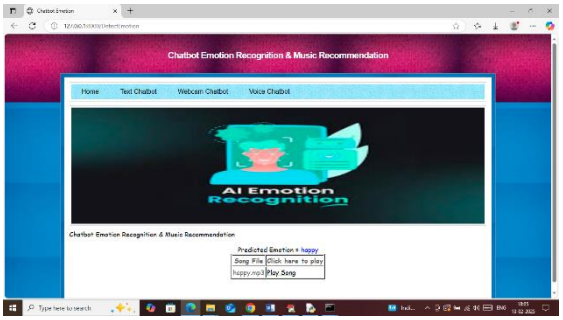
Click on the 'Webcam Chatbot' hyperlink inside the upper display screen to get admission to the subsequent page.



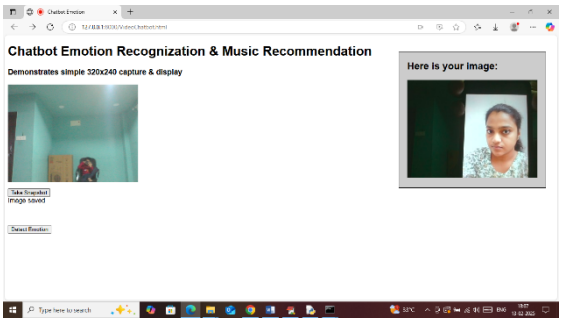
within the camera interface above, you may show your face after which click the 'Take snapshot' button to seize it, accompanied by way of clicking the 'hit upon Emotion' button to reap a list of songs.

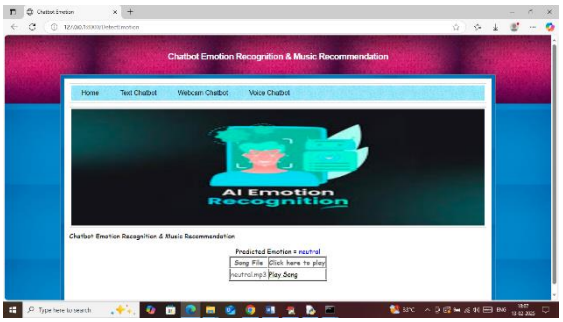


At the right side of the display screen, a detected face is seen; now click on the 'locate Emotion' link to get admission to the subsequent page.

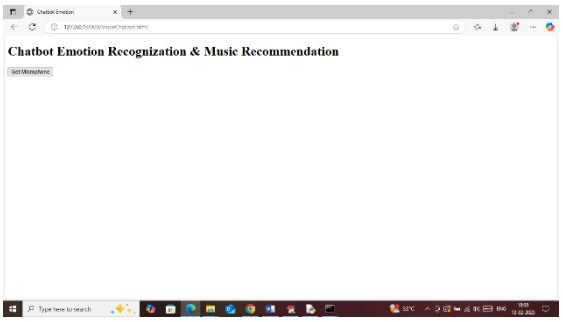


The top screen shows the detection of the 'satisfied' emotion from a face. in addition, you could test with many faces using a stay digital camera, as demonstrated in the display beneath, which checks every other live photograph.

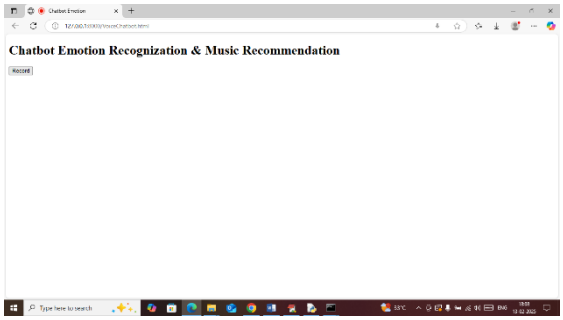




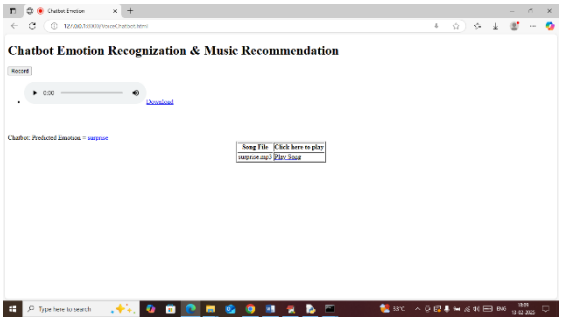
The display screen indicates a impartial facial features; now click on at the 'Voice Chatbot' option to document your voice and thereafter propose songs.



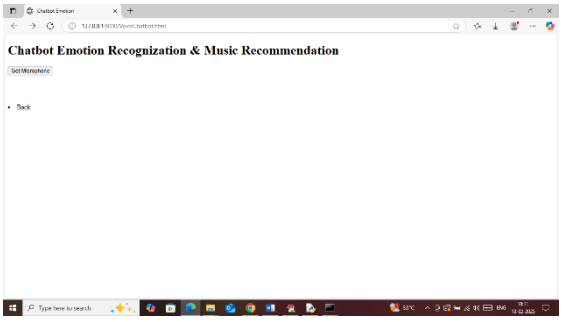
Click on the 'Get Microphone' hyperlink at the above screen to connect the application to the microphone, and then the following page will appear.



on the display screen above, click the 'record' button to start speaking, and once finished, click the 'stop' button to access the subsequent web page.



The vocal emotion depicted inside the recorded display screen is diagnosed as 'marvel'; in addition, you could test with another stay voices.



At any time on the above screen, click the 'back' button to return to the main page.

5. CONCLUSION

In summary, the suggested emotion identification and music recommendation gadget proficiently amalgamates deep getting to know methodologies to evaluate emotions derived from text, vocal intonations, and facial expressions. The gadget achieves first rate enhancements in emotion class accuracy by using utilising BERT embeddings for feature extraction and implementing deep gaining knowledge of architectures which includes CNN, LSTM, BiLSTM, and LSTM-GRU. Of the assessed models, the BiLSTM community attains the first-class accuracy of 91.84%, underscoring its efficacy in figuring out contextual dependencies within sequential statistics. The amalgamation of textual content, vocal intonation, and facial features evaluation right into a cohesive framework improves the machine's capacity to correctly become aware of feelings. An internet-primarily based interface permits actual-time user engagement, making an allowance for seamless textual content input, voice recording, and facial recognition for emotion detection. The gadget delivers customized song recommendations based totally on the user's emotional condition. This multimodal approach improves human-computer interplay by way of developing a more intuitive and emotionally shrewd chatbot system, for this reason enhancing user engagement and enjoy.

Destiny endeavors will concentrate on augmenting the machine's resilience with the use of transformer-primarily based models which include RoBERTa and GPT to decorate textual content emotion reputation. In addition improvements will encompass actual-time emotional adaption through continuous gaining knowledge of, augmentation of the dataset for extra variety, and refinement of voice and face emotion reputation through sophisticated audio-visual fusion methodologies. The device will additionally inspect reinforcement learning for tailored song pointers. Additionally, implementation on cellular platforms can be evaluated to improve accessibility and consumer experience across various devices.

REFERENCES

[1] Jindal, N., Pandey, A., & Sharma, A. (2024). Music Recommendation using Chatbot.

[2] Sharma, A., Vishwakarma, S., & Mathew, L. T. (2024, May). Feel good ai: Voice-enabled emotion-based music recommendation system. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-6). IEEE.

[3] Mohan, G. B., Kumar, R. P., Korrayi, S., Harshitha, M., Chaithanya, B. S. S., Saiteja, K., & Rohan, G. V. (2024, February). Emotion-based music recommendation system-a deep learning approach. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-7). IEEE.

[4] Hamad, O., Hamdi, A., & Shaban, K. (2024). Asem: Enhancing empathy in chatbot through attention-based sentiment and emotion modeling. *arXiv preprint arXiv:2402.16194*.

[5] Zhang, J., Chen, Q., Lu, J., Wang, X., Liu, L., & Feng, Y. (2024). Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism and boundary conditions. *Tourism Management*, *100*, 104835.

[6] Kadyrgali, E., Yerkin, A., Torekhan, Y., & Shamoi, P. (2024, April). Group movie selection using multi-channel emotion recognition. In *2024 IEEE AITU: Digital Generation* (pp. 85-91). IEEE.

[7] Madhan, S., Sridharan, S., Deivasigamani, S., Rajesh, R., & Surendran, R. (2024, September). Facial Expression Analysis using K-Nearest Neighbor Classification Method: Enhancing Emotion Detection and Stress Monitoring in an Interactive Music Player. In *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1316-1322). IEEE.

[8] Hazmoune, S., & Bougamouza, F. (2024). Using transformers for multimodal emotion recognition: Taxonomies and state of the art review. *Engineering Applications of Artificial Intelligence*, *133*, 108339.

[9] N. Mathew, N. Chooramun, and S. Sharif, "Implementing a Chatbot Music Recommender System Based on User Emotion" in Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT 2023), University of Bahrain, Bahrain, November 20-21, 2023. DOI: [https://repository.uel.ac.uk/item/8wq01] IEEE.

[10] V. Singhal, A. Sahu, A. Jaiswal, F. Ahmad, and R. Gaur, "Song Recommender System by Convolutional Neural Network" in Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022, March 13, 2023. Available at SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4386515].

[11] P. J. R, M. M, R. K, and Mr. P. K. G, "Chatplayer - Chatbot Song Recommendation System Using Spotify API," International Journal of Progressive Research in Engineering Management and Science (IJPREMS), vol. 03, issue 05, pp. 1267-1271, May 2023. DOI: [https://www.ijprems.com/uploadedfiles/paper//issue_5_may_2023/31479/final/fin_ijprems1685 368519.pdf].

[12] Anusha, and Dr. Srinivasan V. "Chatbot Song Recommendation System" IJERT, vol. 11, no. 06, NCRTCA - 2023, Dec. 2023, pp. 1267-1271, DOI: [https://www.ijert.org/chatbot-song-recommendation-system].

[13] S. Chaudhari et al., "Chatbot with Song Recommendation based on Emotion" Int. J. Res. Pub. Rev., vol. 3, no. 5, pp. 2716-2719, May 2022. DOI: [https://ijrpr.com/uploads/V3ISSUE5/IJRPR4187.pdf].

[14] [18] IJRASET Publication, "Music Recommender System Using ChatBot," Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET), vol. 9, no. XII, pp. 2315-2323, 2021. DOI: [https://www.academia.edu/68135212/Music_Recommender_System_Using_ChatBot].

[15] P. Suvarna Bahir et al., "Chat Bot Song Recommender System," International Research Journal of Modernization in Engineering Technology and Science, vol. 04, no. 04, pp. 2121, April 2022. Available: [https://www.irjmets.com/uploadedfiles/paper/issue_4_april_2022/21895/final/fin_irjmets16514 04336.pdf]