Saathi – An AI Companion

Ankita Ramesh Upadhyay
Department of Computer Engineering
Shree L.R. Tiwari College of Engineering
Mumbai, India
ankita.upadhyay@slrtce.in

Amey Atul Bhavsar
Department of Computer Engineering
Shree L.R. Tiwari College of Engineering
Mumbai, India
amey.bhavsar@slrtce.in

Deepanshu Kailashnath Yadav
Department of Computer Engineering
Shree L.R. Tiwari College of Engineering
Mumbai, India
deepanshu.yadav@slrtce.in

Pooruvi Virendra Singh
Department of Computer Engineering
Shree L.R. Tiwari College of Engineering
Mumbai, India
pooruvi.singh@slrtce.in

Neelam Phadnis
Department of Computer Engineering
Shree L.R. Tiwari College of Engineering
Mumbai, India
neelam.phadnis@slrtce.in

***Abstract*— The research paper presents a novel approach for developing a mental health chatbot (Saathi) to provide empathetic and human-like support. Using an open domain system, the model behind chatbot predicts the probabilities from message for signs of mental illness like depression, anxiety, and bipolar disorder & emotions. This paper demonstrates a generative approach to create outputs sent to the user keeping the recent history of signs of mental illness and emotions as inputs to the generative model. This results in a more appropriate output since it is dependent on the current state of the user. Saathi has the potential to improve the early detection and management of mental health disorders by providing users with a convenient and confidential way to receive help and support. Focused on adolescents, we aim to bridge the gap between people and mental health aid using this application.**

Keywords—Mental health, Chatbot, NLP, Depression, Anxiety, Bipolar Disorder, CNN, Emotion Detection, Neural Networks, Logistic Regression, Disease Detection, Machine Learning, Deep Learning.

I. INTRODUCTION

Mental health can be defined as a condition that enables individuals to manage life's stressors, develop their potential, learn and function successfully, and give back to their communities. More than only the absence of mental disorders, mental health is a crucial aspect of physical and mental well-being [1]. Our mental health includes all aspects of our wellbeing-emotional, psychological, and social. It influences our thoughts, emotions, and behaviour. It also adversely affects our decision-making, social relations, and reflex to stress. Every period of life, from childhood and adolescence to maturity, is vital for mental health [2]. Mental health issues may have an impact on your thinking, mood, and behavior over the course of your life. Biological variables, like genes or brain chemistry, are among the many factors that affect mental health, experiences with trauma or abuse during life, a history of mental illness in the family, etc. [2].

Conversational AI is a sub-domain of Artificial Intelligence that deals with speech based or text based AI agents that have the capability to simulate and automate conversational and verbal interactions. Conversational agents are agents that interact with users via written or spoken natural language [4]. They have the potential to provide a platform for individuals to access mental health support in a convenient and accessible way. The widespread use of mobile devices and the internet has made it possible to provide mental health services through digital platforms, such as chatbots.

II. RELATED WORK

A. *'A Survey on Conversational Agents/Chatbots Classification and Design Techniques.'[4]*

The use of natural language-based interfaces in human-computer interaction has become increasingly popular for personalization purposes. Chatbots, which are designed to converse with humans using natural language, can be categorized into task-oriented and non-task oriented. Task-oriented chatbots are designed to perform specific tasks based on user instructions, while non-task oriented chatbots have multiple purposes but are unable to perform any particular task.

Task-oriented chatbots can use simple algorithms based on pattern matching to process similar inputs, while non-task oriented chatbots require different algorithms and strategies to ensure the correct response. Generative-based chatbots are more challenging to build and operate than retrieval-based chatbots. This study's findings show that it has the potential to provide an alternative and engaging method of delivering CBT to the approximately 10 million college students in the United States who suffer from debilitating anxiety and depression.

B. *'Emotionally Intelligent Chatbots: A Systematic Literature Review'[5]*

The findings of the study show that most studies are based on an open-domain generative chatbot architecture. Researchers are primarily concerned with accurately detecting the user's emotion and generating emotionally appropriate responses.

Emotion and appropriate response are required for effective communication. Empathetic and emotionally intelligent agents are being developed to detect user sentiments and generate appropriate responses in order to accomplish this. Furthermore,

chatbots simulate human behaviour and emotions, which can lead to increased rapport, higher motivation, and better engagement.

C. 'Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial' [6]

This paper introduces WoeBot. The study's goal was to find out whether a completely automated conversational agent might give a self-help programme to college students who self-report having symptoms of anxiety and depression and whether it would be feasible, acceptable, and potentially effective. This study examined the effects of a text-based conversational agent (Woebot) on 70 participants' intake of self-help content on a university community social media site. The participants varied in age from 18 to 28. The results demonstrated that the Woebot group's depressed symptoms during the study period were dramatically reduced as compared to the information control group. An intent-to-treat univariate analysis of covariance revealed both groups considerably decreased anxiety as assessed by the GAD-7, and the study also revealed a significant group difference with regard to depression.

This study's findings show that it has the potential to provide an alternative and engaging method of delivering CBT to the approximately 10 million college students in the United States who suffer from debilitating anxiety and depression.

D. 'Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users' [3]

A pilot study was conducted to assess the effectiveness of a health app, Ada—Your Health Guide, in diagnosing mental illnesses. The study involved 20 case vignettes with various mental illnesses, and participants, including psychotherapists, psychology students, and laypeople, entered symptoms into the app to receive a prognosis. The study found that the app had poor diagnostic agreement for mental disorders in childhood and adolescence, moderate agreement for mental disorders in maturity, and good agreement for adult mental diseases among psychotherapists. The study concluded that the app has potential as a diagnostic screening tool for mental disorders in adults but needs improvement for mental diseases affecting children and adolescents.

The findings of the study suggest that user factors, such as the level of expertise and experience, affect the diagnostic accuracy of the app. The app took an average of 7 minutes and 34 questions to complete. The study recommends improving the app's accuracy for mental disorders affecting children and adolescents and highlights the potential of the app to support healthcare professionals, particularly diagnosticians, in their work. Overall, the pilot study provides preliminary indications of the app's diagnostic capabilities and highlights the need for further research and development to improve the app's accuracy and effectiveness.

E. 'Increased trait-like impulsivity and course of illness in bipolar disorder'[7]

This research investigated the association between demographic and illness-course factors of bipolar disorder and trait-like impulsivity measured by the Barratt Impulsiveness Scale. The study involved 114 bipolar patients and 71 healthy control participants. The results showed that bipolar patients had higher scores on each BIS-11 subscale compared to the control group. The findings suggest that high trait impulsivity is a significant contributing factor to bipolar disorder and is linked to a more severe course of the disease.

In summary, the study indicates that bipolar disorder patients exhibit higher trait impulsivity than healthy controls, independent of symptomatology, age, education, gender, mental symptoms, and illness-related traits. The research underscores the importance of trait-like impulsivity as a significant factor in bipolar disorder and its potential role in predicting disease severity. The findings provide insights into the development of targeted interventions to address trait impulsivity in bipolar disorder patients.

F. 'Depression and anxiety' [8]

This paper studies depression and anxiety disorder and outlines a flowchart for treating the same. It discusses psychological treatments for depressive and anxiety disorders and pharmacotherapy for unipolar depression and anxiety disorders. There is a lack of data about treatment for anxiety and depression comorbidity, leading to more than a third of patients not seeking treatment and almost half being offered treatments that may not be beneficial. Public awareness and professional education are needed to promote better mental health outcomes.

The comparison of the currently available chatbots for therapy and mental health apps/websites is shown in Table 1. The many applications examined were Woebot, Wysa, Healthily, Amaha and RiseUp.

The study suggests need for a system with unique features:

1. Existing systems do not use interactive or open-ended questions for sentiment analysis.
2. Chatbots tend to follow a clinical flow, resembling that of therapists, which may not be as effective in encouraging users to open up.
3. Proposed system would adopt a friendly conversational style.
4. Study aims to develop a completely free system to assist users regardless of their financial situation.

TABLE I. PROS AND CONS OF EXISTING SYSTEMS

App	*Rating*	*Advantages*	*Disadvantages*
Weobot	4.8	Popular, Interactive, Mood tracker	No custom responses
Wysa	4.5	Provides therapy sessions with professionals. Free registration, Accessible	Meditation session not useful, lack of interaction.

App	*Rating*	*Advantages*	*Disadvantages*
Healtily	3.6	Interactive	Not mental health specific, inaccurate.
Amaha	4.4	Journals, Assessments, Mood tracking, Activities & courses, Book session with therapists.	Chatbot is Repetitive, No custom responses.
RiseUp	4.2	Includes Personality test, Emotion Tracker. Depression & Anxiety tests.	No custom Responses.

III. PROPOSED SYSTEM

In this research paper, we propose a mental health chatbot - Saathi, that can detect signs of depression, anxiety, and bipolar disorder using CNN (Convolutional Neural Network) and emotions of anger, disgust, fear, joy, neutral, sadness, shame, and surprise using logistic regression. The chatbot will have language generation and understanding capabilities through the implementation of the RASA framework. The chatbot will provide a platform for users to connect with professional help and support, thus improving access to mental health services. The proposed chatbot will be trained on a large dataset of text data from individuals with mental health disorders, as well as data from individuals without mental health disorders.

The paper conducted a survey to collect data on users' experiences and perceptions of mental health chatbots. The survey aimed to examine users' attitudes towards using mental health chatbots, their satisfaction with the chatbot's responses, and the effectiveness of the chatbot in addressing their mental health concerns. Majority of the respondents were from India. 6% of the respondents were willing for chatbot only therapy while 29.6% students agree for both chatbot & in-person therapy. 30.2% respondents agreed to use a mental health chatbot with a platform to provide professional help from a certified professional. The results of the survey suggest that mental health chatbots can be an effective tool in providing mental health support and guidance, particularly for individuals who may not have access to traditional mental health services.

The chatbot will use deep learning techniques to analyse the text data and identify patterns that are indicative of mental health disorders and emotions. Saathi will also be able to detect emotions expressed in the text and respond in an appropriate and empathetic manner. Saathi will also provide a platform for individuals to access mental health support in a confidential and convenient way. The chatbot will also provide a platform for users to connect with professional help if needed. Saathi will be accessible via mobile devices and web browsers, making it easy for individuals to access the chatbot at any time and from any location.

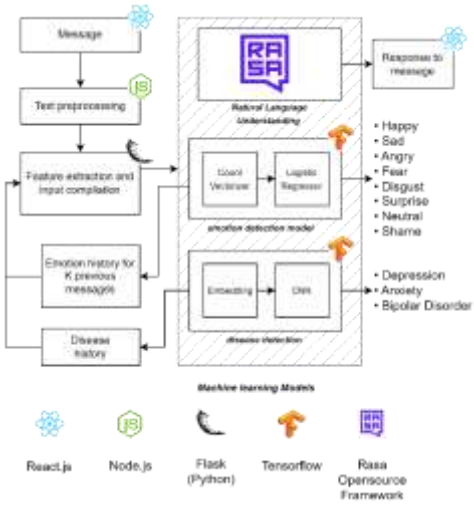

Fig. 1. Architecture of Saathi

Our proposed architecture consists of 5 components, as illustrated in Fig: 1.

A. *Application*

The user interface for Saathi is a divided into three components:

1) *Chat experience:* The chatting functionality allows users to communicate freely without any hesitation or judgment with chatbot.

2) *Dashboard:* Application's dashboard offers a visual representations of the user's mental state such as graphs which displays the user's emotional trends and the severity of the detected symptoms gathered during the chat sessions.

3) *Consultation:* On Saathi, the user can book an appointment with a psychiatrist by selecting a preferred date and time for their appointment along with a list of available mental health counsellors.

B. *Natural Language Understanding*

To facilitate NLP, RASA Open Source framework can be utilized. RASA has two primary components that provide its core functionality:

1) *Intent Generation:* Rasa uses the concept of intents to describe how user messages should be categorized. Rasa NLU will classify the user messages into one or also multiple user intents. The two components between which you can choose are:

a) Pretrained Embeddings (Intent_classifier_sklearn): This classifier loads pretrained language models from the spaCy library and then utilises those models to represent each word in the user message as a word embedding. In order to get the appropriate parameters for the support vector classifier, which categorises the averaged embeddings into the various intents, Rasa NLU averages all word embeddings inside a message.

b) Supervised Embeddings - (Intent_classifier_tensorflow_embedding): This method learns

word embeddings from scratch rather than utilising pre-trained embeddings and then training a classifier on top of them.

2) Entity Extraction: Rasa NLU, an open-source framework, emphasises complete customizability in particular. As a consequence, Rasa NLU offers you a variety of entity recognition components that may be customised to meet your needs.

3) Stories: Stories are a type of training data used to train your assistant's dialogue management model. Stories can be used to train models that are able to generalize to unseen conversation paths.

4) Actions: The dialogue management model for your helper is trained using stories as a sort of training data. Models that can generalise to unheard conversation pathways may be trained using stories.

C. Emotion Detection

Emotion detection Model: The dataset used for emotion detection was extracted from twitter. The Dataset has been trained for the following emotions:- joy, sadness, fear, anger, surprise, neutral, disgust, and shame. NeatText library is a tool for tidying up text data (neattext.functions) was used to perform data cleaning & preprocessing tasks.The dataset was split into training and testing data with the ratio of 70:30. A pipeline was built using Logistic Regression and CountVectorizer which is a method of converting a phrase to individual words.

This results in an encoded vector that contains the total number of unique words in the text, and the number of occurrences of each word in the sentence. The emotion detection model gives a probability score for each emotion label and outputs the emotion with maximum score.

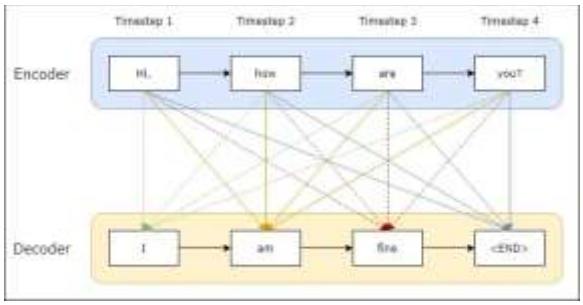

Fig. 2. Seq2seq Architecture diagram

In our study, we obtained an official copy of the dataset provided by the founders of CounselChat, where therapists respond to questions posed by clients, and users can like responses that they find most helpful. [9]

IV. Results

In this paper, we have designed Saathi - An AI Chat Companion that can detect early symptoms of Depression with an accuracy 86.76% against the training duration of 75.0 minutes, anxiety with an accuracy of 89.79% against the training duration of 61.0 minutes and Bipolar Disorder with an accuracy of 95.51% against the training duration of 77.0 minutes.

D. Illness Detection

The First Phase of Disease Detection was Data Pre-processing, Real world data was extracted from Reddit Mental health Forum Pages using web scraping. In the Illness Detection Model, our main aim was to cover Mental illnesses like 'Depression', 'Anxiety' and 'Bipolar Disorder'. Data Cleaning measures like removal of duplicates, spaces, emojis, special characters, URLs were performed along with custom tokenization, wordnet lemmatization. The Machine Learning models were individually trained against labels (0,1). Feature Extraction was performed for unique tokens. The Training and Testing data were divided in the ratio of 80:20.

Word embeddings of size 127 were passed to a CNN layer with Relu activation function, followed by 64 node CNN layers, the output from CNN layers were down sampled by using MaxPooling1D, and then flattened into a single vector using Flatten(), this vector is passed to 2 Dense layers with 250 and 3 nodes each and uses Relu activation function, the final 3 nodes give the output for each mental disorder.

E. Natural Language Generation

The seq2seq model, also called the encoder-decoder model, uses Long Short-Term Memory- LSTM for text generation from the training corpus. It predicts a word given in the user input and then each of the next words is predicted using the probability of likelihood of that word to occur. Using a dataset with question-answer pairs or message-reply pairs, we train the model to generate an output for given input.

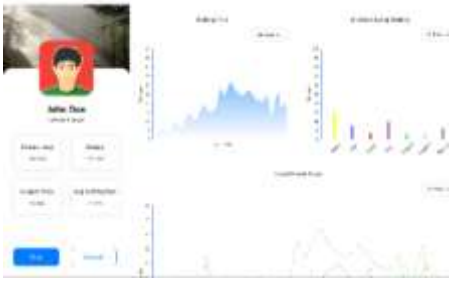

Fig. 3. Saathi – User's Dashboard

Fig. 4. Sample Chat

Saathi can also do opinion mining on the ongoing conversation and detect the emotion like anger, joy, neutral, shame, disgust, fear, sadness, surprise with an accuracy of 62% and respond accordingly. The integration of the CNN and logistic regression models with RASA results in a chatbot that can understand natural language inputs and generate appropriate responses in a way that mimics human conversation. Currently, the Saathi system can only evaluate the last text sent and will send a score for emotion and disease detection and the average of all the scores from the previous conversation will be sent as an output in the Analysis Report of the user.

V. CONCLUSION

In conclusion, this research aimed to develop a mental health chatbot that can detect mental illnesses like depression, anxiety, and bipolar disorder using a combination of machine learning techniques including CNN, logistic regression, and RASA platform. Furthermore, the chatbot was able to detect various emotions such as anger, disgust, fear, joy, neutral, sadness, shame, and surprise. The chatbot also provided a platform for users to connect with a professional for professional help and to share the Disease detection & Emotion detection Analysis with the Mental Healthcare Professionals. Saathi can be also used by Healthcare Organizations to monitor their patients better and have a deeper understanding of their emotional well-being & Mental Health. The results of this research demonstrate the potential of using advanced machine learning techniques to develop a highly accurate and efficient mental health chatbot. However, it is important to note that the chatbot is not a substitute for professional help and should only be used as a tool to support the diagnosis and treatment of mental illnesses. Saathi can be used by anyone irrespective of Mental illness or not, as it is trained on a wide dataset from individuals with or without any mental health disorders, and is accessible via mobile devices and web browsers, providing users with the flexibility to access the chatbot at any time and from any location. Further development can be done on the platform by upscaling the application, improving the Model Performances, making it a self-sustainable model, considering last n text messages for evaluation, Improving Analysis by using more efficient algorithms, and improving UI / UX.

References

[1] World Health Organization, “Mental Health,” WHO, 2022.

[2] U.S. Department of Health & Human Services, “What Is Mental Health?,” Mentalhealth.gov, Feb. 28, 2022. https://www.mentalhealth.gov/basics/what-is-mental-health

[3] Jungmann S, Klan T, Kuhn S, Jungmann F Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users, JMIR Form Res 2019;3(4):e13863, URL: https://formative.jmir.org/2019/4/e13863, DOI: 10.2196/13863.

[4] Hussain, S., Ameri Sianaki, O., Ababneh, N. (2019). A Survey on Conversational Agents/Chatbots Classification and Design Techniques. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds) Web, Artificial Intelligence and Network Applications. WAINA 2019. Advances in Intelligent Systems and Computing, vol 927. Springer, Cham. https://doi.org/10.1007/978-3-030-15035-8_93.

[5] Bilquise, Ghazala & Ibrahim, Samar & Shaalan, Khaled. (2022). Emotionally Intelligent Chatbots: A Systematic Literature Review. Human Behavior and Emerging Technologies. 2022. 1-23. 10.1155/2022/9601630.

[6] Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health. 2017 Jun 6;4(2):e19. doi: 10.2196/mental.7785. PMID: 28588005; PMCID: PMC5478797.

[7] Swann, A.C., Lijffijt, M., Lane, S.D., Steinberg, J.L. and Moeller, F.G. (2009), Increased trait-like impulsivity and course of illness in bipolar disorder. Bipolar Disorders, 11: 280-288. https://doi.org/10.1111/j.1399-5618.2009.00678.x

[8] Tiller, John W G. “Depression and anxiety.” The Medical journal of Australia vol. 199,S6 (2013): S28-31. doi:10.5694/mja12.10628

[9] https://github.com/nbertagnolli/counsel-chat

[10] Dongkeon Lee, Kyo-Joong Oh, and Ho-Jin Choi, “The chatbot feels you - a counselling service using emotional response generation,” 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Feb. 2017, doi: 10.1109/bigcomp.2017.7881752

[11] P. Rathnayaka, N. Mills, D. Burnett, D. De Silva, D. Alahakoon, and R. Gray, “A Mental Health Chatbot with Cognitive Skills for Personalised Behavioural Activation and Remote Health Monitoring,” Sensors, vol. 22, no. 10, p. 3653, May 2022, doi: 10.3390/s22103653.

[12] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O’Neill, Cherie Armour, Michael McTear, “Towards a chatbot for digital counselling”, Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017), July 2017, pp. 1-7

[13] Osorio, Mauricio, Claudia Zepeda Cortés and José Luis Carballido. “MyUBot: Towards an Artificial Intelligence Agent System Chat-bot for Well-being and Mental Health.” AAI4H@ECAI (2020).