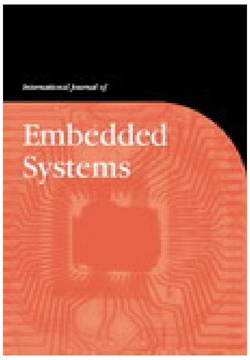
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Embedded Systems

ISSN online: 1741-1076 - ISSN print: 1741-1068
https://www.inderscience.com/ijes

Chatbot for mental health diagnosis using data augmentation techniques and deep learning

Neel Ghoshal, Vaibhav Bhartia, B.K. Tripathy, A. Tripathy

DOI: 10.1504/IJES.2024.10065115

Article History:

Received:	29 April 2023
Last revised:	28 July 2023
Accepted:	16 September 2023
Published online:	10 February 2025

Chatbot for mental health diagnosis using data augmentation techniques and deep learning

Neel Ghoshal and Vaibhav Bhartia

SCOPE,
Vellore Institute of Technology,
Vellore, 632014, Tamil Nadu, India
Email: neelghoshal02@gmail.com
Email: bhartiavaibhav@gmail.com

B.K. Tripathy*

SCORE,
Vellore Institute of Technology,
Vellore, 632014, Tamil Nadu, India
Email: tripathybk@rediffmail.com
*Corresponding author

A. Tripathy

Carnegie Mellon University,
5000 Forbes Ave, Pittsburgh, PA 15213, USA
Email: anurag7642@gmail.com

Abstract: Mental illness has become widespread among people world over. Although several chatbot based models have been designed, their efficient utilisation by people is not properly confirmed. In this paper a customised chatbot framework is proposed and developed using natural language understanding (NLU) mechanisms, which comprises a unique two-tier modular functionality of an empathetic conversational model with a simultaneous implementation of a classification model. The framework provides a holistic service to a user. The dataset is prepared manually to include the various mental health diseases and the appropriate responses provided by professionals. The model uses conversational therapeutic data and uses RASA as its structural framework, which is trained to perform efficient and vicarious dialogue with a user. The mental health-based categorical dataset undergoes various models such as logistic regression, decision tree random forest, naive Bayes and Google BERT leading to an accuracy of 91.08%.

Keywords: chatbot; mental health; natural language processing; NLP; deep learning; soft computing.

Reference to this paper should be made as follows: Ghoshal, N., Bhartia, V., Tripathy, B.K. and Tripathy, A. (2024) 'Chatbot for mental health diagnosis using data augmentation techniques and deep learning', *Int. J. Embedded Systems*, Vol. 17, Nos. 3/4, pp.171–182.

Biographical notes: Neel Ghoshal is currently pursuing his Bachelor's degree in Computer Science and Engineering from VIT, Vellore. He has multiple contributions accepted in renowned publications and journals including IEEE, Springer and Wiley. He has worked with 'Reliance Jio' as a Machine Learning Intern and is currently interested in tackling research oriented problems in machine learning and its applications.

Vaibhav Bhartia is currently doing his BTech in Computer Science and Engineering from VIT, Vellore. He has published over seven papers in journals and proceedings of international conferences sponsored by Springer, Elsevier and IEEE. He has presented papers in all types of international conferences. His current areas of interest include deep learning and computer vision.

B.K. Tripathy is a Professor at VIT, Vellore since 2007. He has supervised 52 research degrees, published 755 papers in international journals, proceedings of international/national conferences and chapters in edited research volumes. He has published two books, two monographs and edited ten research volumes. He is a fellow of IETE, IEI and a senior member of ACM, IEEE, IRSS and CSI. He is in the editorial board/reviewer of more than 100 international journals and has adjudicated PhD thesis of more than 30 universities. He has served as member of advisory committee/TPC of more than 190 international conferences.

A. Tripathy did his BTech in Computer Science and Engineering from VIT, Vellore and had served at 'Publicis Sapient' as an Associate Quality Assurance L2 in Bengaluru for three years. He then joined CMU in the USA for his Master's in ISM in May 2018. At present, he is working as a Senior Developer for Kijiji Canada. He has published ten papers in journals and proceedings of international conferences sponsored by Springer, ACM and IEEE. He has presented papers in all types of international conferences. His current areas of interest include machine learning and MIS.

1 Introduction

Nobody wants to talk about the axiomatic problem in the world that is mental health (World Health Organization, 2005). There is hardly any public discussion about how to prevent or cure mental illness in India (Reddy, 2019), despite the country being on the cusp of a crisis. Almost little action is being taken on the scale necessary to address the rising number of individuals with mental health difficulties.

There is a significant gap between the support that is readily accessible, easily affordable, and the treatment that should be offered. Even in wealthy countries, the ratio of psychiatrists, psychologists, psychiatric social workers, and mental health nurses to patients is one to 10,000 (Katschnig, 2010). The system's flaw ensures that the majority of those with mental health problems never get the assistance they require. Numerous digital interfaces are emerging as practical supplemental services to fill some of the needs of artificial intelligence (AI)-based solutions that offer help and frequently some form of companionship. These solutions were developed in close collaboration with healthcare professionals. Additionally, it may lower the price of psychiatric diagnosis and care. When it comes to psychiatric diseases, most people have experienced the stigma that pervades our society and frequently prevents proper treatment. Using spoken, written, and visual languages, chatbots are natural language processing (NLP)-based frameworks that communicate with human users (Adamopoulou and Lefteris, 2020). Chatbots, created expressly to communicate with persons with mental health issues, have the potential to be helpful resources. On multiple platforms, a chatbot can imitate a conversation using text, audio, and even video. While some chatbots use human interface, others are totally automated. To enable these bots to be compatible with the complexity of human communication and the capacity to recognise cultural subtleties, the AI frameworks need to be trained with a lot of data (Brandtzaeg and Følstad, 2017). Chatbots can offer company, assistance, and treatment, which significantly lessens the workload for therapists. It presents itself as a practical choice for those who struggle with accessibility and cost.

The structure of the paper is as follows, the upcoming sections will first focus upon the related works to our current proposal, then we shall discuss the proposed methods that we have used, depicting the pipeline flow of the entire data, in which the details of the workings of the conversational model as well as the classification model are discussed while also simultaneously explaining the response generation process. After this, the next section shall describe the experimental analysis of the results obtained in the previous sections, after which a discussion section is provided for the portrayal of the inferences of the results which is followed by the section on conclusions and future scope which allows for a compact and precise analysis of the specificities of results obtained and the scopes that may exist for future work upon the same.

2 Related work

Tech companies around the world are combining the power of AI with the portability of smartphones to create chatbots designed to assist patients with mental health issues. These conversational agents assist patients while maintaining a high level of privacy and anonymity.

While some chatbots, like Ellie (Kim et al., 2022), can detect minor changes in our facial ex-pressions, speech speeds, or length of pauses and create a diagnosis accordingly, Woebot is a completely automated conversational robot that treats depression and anxiety using a digital version of cognitive behaviour therapy (CBT) (Fitzpatrick et al., 2017). The patient is given the option to meet with a real therapist if a serious issue is found, and pertinent hotline numbers are offered (Darcy et al., 2022). Analytical methods using social media-based data have also been carried out for natural language and categorical inferences (Lijo and Seetha, 2021). Advancements have also been made concerning social media data usage towards patient health analysis (Panda et al., 2023).

A chatbot that can speak in the voice of the departed person can help a person who is grieving after the loss of a close relative (Tracey et al., 2021). Messages are delivered over time that aid the recipient in erasing the trauma brought on by an unexpected loss and the lack of closure. In a fantastic advertisement for the Bixby Voice Assistant (Nobles et al., 2020), Samsung's #VoiceForever, this idea has been thoroughly explored. The commercial features a mother's voice working in tandem with a voice assistant to assist a young girl in coming to terms with the death of her mother (Narmadha et al., 2019).

The cognitive-behavioural therapy model (CBT) serves as the foundation for chat-bots that address mental health (Fenn and Majella, 2013). A step-by-step software manual or chatbot that employs CBT can help people examine and alter their mental patterns by using structured activities (Sulaiman et al., 2022). The management of mental health issues can benefit from prompt chatbot interventions with patients (Gabrielli et al., 2020). This is mostly accomplished by encouraging patients to transform their negative ideas into positive ones by utilising clinical skills and natural language processing. For the end-user, this serves to provide a cathartic and healing experience. The Silicon Valley startup X2AI created a chatbot named Karim that speaks Arabic and assists Syrian refugees with their mental health difficulties (Sekkat et al., 2021).

In Cameron et al. (2017), the authors have created a chatbot interface, wherein the chatbot initiates a conversation by asking how the user is feeling, and the user can respond using any one of the 'emojis' provided. Then, the user can pick from a set of issues that they may be facing and according to that, tips are given as a response. An Open AI-based architecture has recently been proposed by Ivanovic et al. (2023) who have demonstrated methodologies concerning medical data usage and management.

We have used an artificial neural network (ANN) system for the conversational chatbot using natural language understanding as well, to generate and carry forward textual conversations according to emotion and context of user, and parallel to that a systemic classification model runs on the user's responses to detect what problems that the user is facing and generate appropriate responses for that. In Grové (2021), the author has created a conversational chatbot for user interaction, wherein intent of the user is appropriated using machine learning, and appropriate responses are given out according to this. A risk categorisation is done according to level of risk the youth may be in, and a trigger to alarm system is present for high risk situations. Some more application developments using chatbots are found in Verma et al. (2022), Srividya et al. (2020, 2021) and Bhattacharya et al. (2022).

A categorisation model to detect the specific mental health problem the user may be facing is not present. We have developed a cohesive mental health and empathetic conversational chatbot, along with this we also categorise the specific and medical term of mental health issue that the user is facing using a classification model and give appropriate responses according to this as well.

3 Proposed method

The proposed method emphasises on mainly three parts: namely:

a getting user input as input data through a conversational chatbot model

b classification of various mental health diseases based on the previous input using state-of-the-art classification models

c generating appropriate responses for the user based on the output of the previous models as shown in Figure 1.

In general, chatbots either directly classify the diseases without conversing with the user or just has normal conversation without helping the user with its issues. The idea behind the proposed work is creating a well-balanced chatbot which gives the user solutions to its mental health issues while conversing with the user to give them a platform the user can only get while conversing with a professional. Our proposed framework is extremely unique due to its cohesion of two vital functionalities concerning both the conversational needs of the users and the corresponding knowledge of their mental health diagnosis. The method proposed has extensive applications in various industries, for example, the clinical industry can incorporate the conversational model along with its clinical procedures for a holistic and incorporative approach. The chatbot can assist in specific screening and assessment needs, insight generation via empathetic conversation, deliverance of educational and self-care requirements and progress tracking. Simultaneously, the corresponding classification model can be used by the clinical industry for preliminary analysis tasks, general progress tracking reasons, and predictive insight generation. Apart from the clinical industry, other industries can also incorporate this holistic model for employee-based use, allowing for diagnostic and preventative measures. Considering the lack of data in this domain, the dataset was scraped and manually extracted from Reddit as well as various surveys and websites present on the internet between patients and therapists. Mental health being a sensitive topic, the responses generated were thoroughly surveyed and validated by professionals to maintain a user-friendly environment while conversing with the chatbot. The architecture of the proposed framework is shown in Figure 2. The proposed method is divided into three components as:

- conversational model
- classification model
- response generation.

3.1 Conversational model

A specific and empathetic chatbot is created using mental health and therapist related data, which allows seamless and on topic conversations with user entailed, the chatbot allows a healthy and beneficial conversation with a potential patient while simultaneously gathering the user data for further processing and classification. The dataset used for this is comprised of genuine therapist and patient conversational data, which allows for a few key advantages for the model, including, embedded professional insights, clinical understanding, empathy and best professional practices. The parts of the framework are as follows:

- dataset description
- model.

Figure 1 General overview of the proposed method

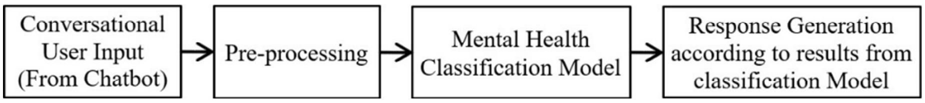


Figure 2 Architecture of the proposed framework

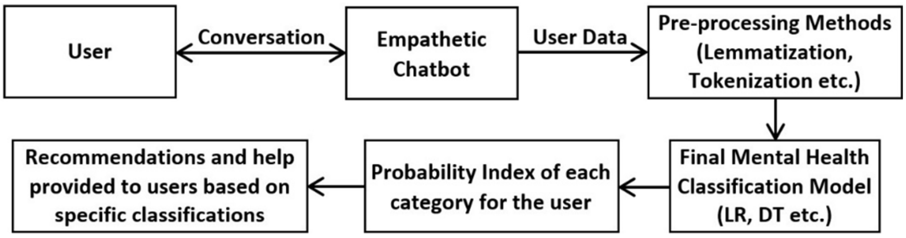
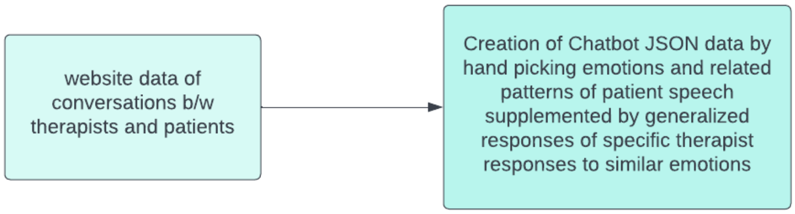


Figure 3 Chatbot dataset (see online version for colours)



3.1.1 Dataset description

For the dataset, we have scraped authentic conversation data between therapists and patients publicly available websites. The data is then categorised manually to tag specific emotions and their specific contexts to allow for proper response generation and efficient training. The responses encoded to be given out for each emotion tag is curated to be a general and appropriate response, gathered from authentic therapist responses. The dataset contains patterns, specific emotion tags specific to them and appropriate responses for them, encoded in JSON format. The tags contained are: 'greetings', 'goodbye', 'self-esteem-A', 'self-esteem-B', 'relationship-a', 'relationship-b', 'angermanagement-a', 'angermanagement-b', 'domesticviolence', 'griefandloss', 'substanceabuse-a', 'substanceabuse-b', 'familyconflict'. The conversation dataset diagram is shown in Figure 3. The dataset can be viewed at this link: https://www.kaggle.com/datasets/neelghoshal/therapist-patient-conversation-dataset. As mentioned, all of the above subcategories refer to a certain situation or emotion either felt by the patient or by someone close to them. Some possible deficiencies in the dataset would be limited availability of data due to restrictions, lack of standardisation, limited labelling and validation and limited diversity.

This dataset has been further augmented and reformatted to fit in a '.yml' file format to be used as training data in the model described below. The augmentation of the dataset has been done using text data augmentation techniques, which in essence to generate text-based training data by artificially augmenting it various embedded methodologies and frameworks like character level augmentation, word level augmentation, phase level augmentation and document level augmentation.

The Python library used for the chatbot dataset extension is 'nlpaug'. The implementation of the same encapsulates the use of techniques like word replacements, word embeddings, back translation and text generation. All of these methods coalesce together to allow the creation of synthetic augmented data which can be used to train the machine learning model in a much more stable and accurate environment.

The specific techniques we have used and the components used to produce the results are as follows:

1 Character augmentation techniques: these techniques work with NLP augmentation at the character level of a given text sample. This is done via the following methods,

   a OCR augmentation: this augmentation technique encapsulates the methodology of simulating OCR error by using random values. OCR here refers to 'optical character recognition'. The random values selected correspond to specific errors common when using OCR technology.

   b Keyboard augmentation: this method works with introducing random values in a given text sample which correspond to the context of errors which may be caused due to a keyboard interface.

   c Random augmenter: this technique follows through its task by either substituting, swapping or deleting values present in the text sample.

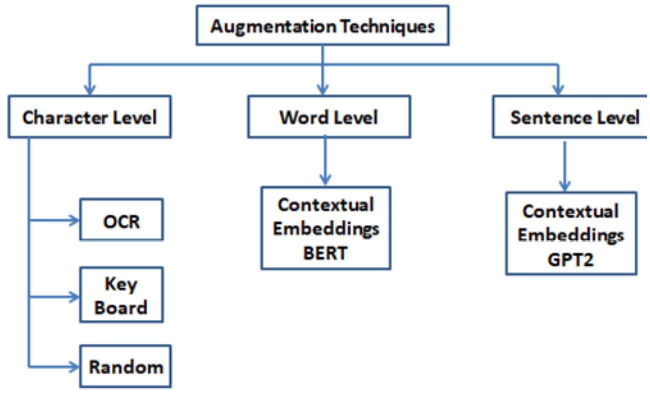
2 Word augmentation techniques: these techniques work with NLP augmentation at the word level of a given text sample. This is done via the following method,

   a Contextual word embeddings: this method works with inserting and substituting words using Google BERT as a model path.

3 Sentence augmentation techniques: this method works with NLP augmentation at the sentence level of a given text sample. This is done via the following method:

   a Contextual word embeddings: here, we have used GPT2 as a model path to generate augmented textual data, which enables for the creation of newer sentences.

Figure 4 Augmentation techniques used (see online version for colours)

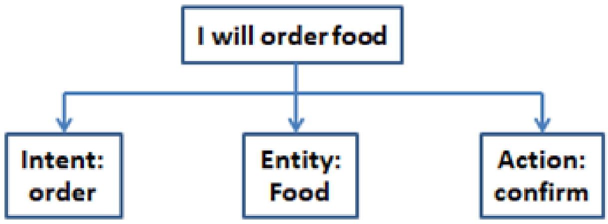


3.1.2 Model

For training and creation of our conversational chatbot model, we have used an ANN classification structure. An ANN is a collection of layers each having a specific number of nodes, each ascertaining to a mathematical function/formula allowing for cohesive training for creating models. The model consists of an input layer, 1 hidden layer, and an output layer along with dropout regularisation for the layers. We have also used regularisation to tackle the problem of overfitting and underfitting while training the data. For regularisation, we have used dropout regularisation which allows for random disablement of some nodes in the respective layers allowing for a more distributed and proper training mechanism. Dropout allows for more efficient training process due to the removal of noise carried on in the layers while training. All the above methodology has been fitted into the conversational model framework.

The chatbot model has been further and finally equipped with natural language understanding methodology via implementing the open-source software 'Rasa'. So, instead of directly feeding the training data into an ANN architecture, using NLU, the training data will be categorised into a set of subdivisions which would include categorisation of words into either an intent, entity or action. Intent can be considered as an understanding of a verb in the English context, as it generally refers to an action word in the sentence. Entity generally refers to an understanding of a noun in the English context. Action refers to the embedded response architecture of the model with respect to a given sentence, i.e., the final model refers to the action subdivision of a sentence to then carry forth towards an action like confirmation, agreement, specific response etc. According to these categorisations, the model is then fed into a neural network.

Figure 5 Word categorisation (see online version for colours)



The model allows the capability of taking the input in a more structured and formatted version to fed into the model with a custom pipeline build in with all mentioned and implemented parameters for proper results. The conversational model pipeline uses something called as 'memorisation policy' which refers to the ability of the model to remember 'story' structures mentioned. A 'story' refers to a custom conversational format that is frequently followed and falls under the category of conversational pre-suppositions that are provided to the model for it to be able to give better results while conversing with the user.

The model employs characteristics and functionalities like:

- Dual intent and entity transformer (DIET) has the functionality that allows the model to employ a conditional random field tagging substructure over the framework of a transformer overlap. DIET is unique and different from other language models as it has a modular architecture, parallels large-scale pre-trained language models and improves upon current best architectures.
- Whitespace tokeniser, which basically allows us to obtain the tokenised words. Tokenisation of a sentence or textual collection refers to the categorisation of the entirety of the textual content into words, terms or some other meaningful elements.
- Transformer embedding dialogue (TED) is a policy, is an architecture for predictive analysis and entity recognition. It uses a conditional random field (CRF) tagging layer and a sequence transformer encoder and single semantic vector space methodologies.
- Rule policy, which handles conversation parts that follow a fixed behaviour, which in essence refers to the behaviour of the machine learning model to allow for proper responses according to a sequential rule code set already defined before.

3.2 Classification model

This subsection refers to the model to be employed at the end of a conversation session between the user and the conversational chatbot model. The specific task of this model is for the classification of various mental health diseases such as depression, stress, anxiety, bipolar disorder, and personality disorder is done using state-of-the-art models such as logistic regression (LR), decision tree (DT), random forest (RF), multinomial naïve Bayes (NB) and Google BERT. The dataset is extracted from Reddit and is preprocessed using various preprocessing techniques. This model provides a rich, contextual and comprehensive understanding of the user input, it is possible to do this due to the distinctive nature of its natural, unrestrained and germane dataset, due to which the model is bolstered in its functionality to provide outputs and categorisations while taking into consideration important aspects including real world context, unstructured and tacit language, and diversity factors. The stages of the framework are:

- dataset description
- data cleaning and exploration
- models
- continuous automated model retraining

3.2.1 Dataset description

The data is collected by scraping the Reddit website using Reddit API. The dataset is scraped using the API-based service to gain access to a page on the website, which is commonly referred to as a 'Subreddit', through which an ordering mechanism is followed for each and every post present on that specific 'Subreddit', which is then requested from the servers and directly appended to the cumulative dataset collection. The cumulative dataset contains information regarding 5 diseases such as depression, stress, anxiety, bipolar disorder, and personality disorder. It mainly has 3 columns namely, title, post text and target value. It contains 1818 documents (rows) for each said disease with the three features(columns) mentioned. The data was collected from the filtered subsection of the disease subreddit which are the most viable according to the users (hot). The dataset was split into training and testing in the ratio 4:1. Some possible deficiencies in this dataset would be the presence of anonymous and misleading information as data points, presence of spam, irrelevant information or noise in the dataset, limited labelling measures, class imbalance etc. The classification model dataset is shown in Figure 6.

3.2.2 Data cleaning and exploration

The dataset being directly scraped and extracted from Reddit needs to undergo various preprocessing techniques. The data is thoroughly cleaned to remove any kind of noise in the dataset. Duplicity is avoided to remove any noise which could hurt the classification models. Null values occurred in the text feature; this occurs when a user on Reddit decides to use only the title feature. Since the title feature are an important part of the dataset, the null values were not dropped instead were filled with unique text. Furthermore, the data was visualised into a word cloud with the top 100 words in all five subreddits to give a better understanding of the dataset (Figure 7).

3.2.3 Models

The classification of the diseases was done using various machine learning classifiers. Count vectoriser was used for feature engineering to convert the collection of text document, i.e., the rows of the dataset into a matrix containing token counts. Various other feature engineering techniques were used such as stop word removal, stripping accents along with a n-gram model with a fixed threshold frequency for the document. The subsequent classification models that were used are:

3.2.3.1 Logistic regression

Logistic regression is used for the prediction of a binary event occurring. It is a supervised machine learning algorithm which predicts the output of a categorical dependent variable, giving probabilistic values as output.

Two LR models were performed with different parameters.

1 The first model had class weight as balanced, warm state as true, and solver as bilinear. It had 'C' value set as 1 with penalty set as 'l1'. The random state of the model was set to 42. Simultaneously the verbose value was set as 1 and 'n' jobs value was set as -1.

2 The second model had 'C' value set as 2.5 with penalty set as 'l2'. The class weight was set as balanced and the warm set was set as true. The random state here was 42 and the solver used here was lbfgs. Here also, the verbose was set as 1 and the 'n' jobs was set as –1.

The first model showed better results in comparison to the second model.

Figure 6 Classification dataset (see online version for colours)

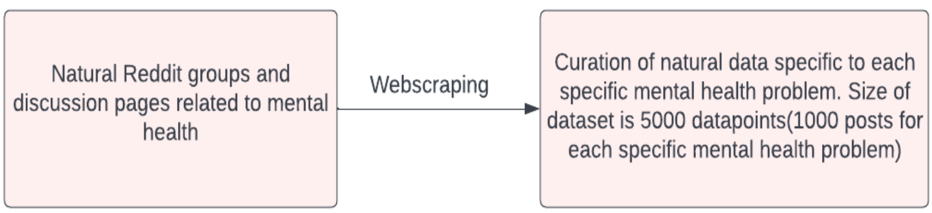
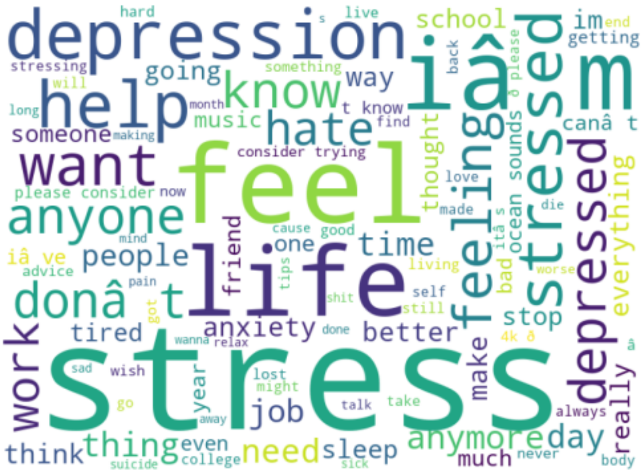


Figure 7 Sample word cloud (see online version for colours)



3.2.3.2 *Decision tree*

Decision tree is a classifier that enables you to make decisions according to some process. This algorithm is a supervised machine learning algorithm technique which can be used for both classification and regression. Here internal nodes represent the decision rules and each leaf node represents the outcome.

Like LR, two DT models were performed.

1. For the first parameter setting, the criterion for it is set as 'gini', and the max depth of the model is categorised at 4, 24 and 54. The minimum samples split is set as 5, 7, 11 and 14 whereas the max features are set as 'none', 'log2', 'auto', .40, .50 and .70. The random state for the model is set as 42 for the simulation.
2. For the second parameter setting, the criterion used for simulation is entropy, and the max depth is set as 4, 24 and 54. The minimum sample splits are categorised by the numerics 5, 7, 11 and 14 similar to the above setting.

Both the models showed almost identical results.

3.2.3.3 *Random forest*

Random forest is a collection of decision trees known as 'forest' which used for both classification and regression. It uses a supervised learning algorithm for its processes. The model collects the result from each tree and then calculates the prediction's majority votes for it to choose it is necessities.

1. The first parameter setting has n_estimators set as 15, 24 and 30. The criterion for the model is set as gini. The max_depth is set as the values of 'none', 5, 13 and 21. The bootstrap is set as true and the minimum samples split are 5, 7, 15 and 25. The maximum features of the model is set as 'none', 'log2', 'auto', .10, .25 and .50. The verbose is set as 1 and the n_jobs as set as the value of –1.
2. The first parameter setting has n_estimators set as 15, 24 and 30. The criterion for the model is set as entropy. The max_depth is set as the values of 'none', 5, 13 and 21. The bootstrap is set as true and the minimum samples split are 5, 7, 15 and 25. The maximum features of the model is set as 'none', 'log2', 'auto', .10, .25 and .50. The verbose is set as 1 and the n_jobs as set as the value of –1.

The second model showed better results than the first model.

3.2.3.4 *Multinomial naïve Bayes*

Multinomial naïve Bayes (NB) is a classifier used for classification of datasets with discrete features. The two variations of the models had fit prior hyperparameter as the distinction. Baye's theorem is used to predict tags of the text for predictive outcomes.

1. The first parameter setting employs parameters depicting a fit_prior value of true and an alpha value of 0, 0.5 and 1.
2. The second parameter setting employs a fit_prior value of false and the alpha values.

The second model showed comprehensively better results than the first model.

3.2.3.5 *Google BERT classification model*

BERT or bidirectional encoder representations from transformers is a transformer-based pre-trained machine learning model used for NLP-based predictive methodologies. BERT uses a methodology of masked language model which enforces bidirectional learning from text by hiding words in a sentence. It is based on transformer architecture and also employees the use of functionalities like next sentence prediction.

Continuous automated model retraining, we have used the concept of automatic model training for the classification model subsection, in which, we have deployed models on the internet as a web service page, upon which we have used the methodology of using code-based functionality to automatically carry out the following steps:

a. Data is fetched from all the aforementioned pages on the Reddit website, which is accumulated via using the Reddit-API to gather the respective post data in each subreddit, which is appended to the cumulative dataset for the conversational model.
b. The data is periodically cleaned and techniques like, segmentation, tokenisation, lemmatisation etc. are applied on the dataset.
c. The data is fed into the present models and the models therefore get retrained periodically over time.

The current periodic time variable is set as 30 days worth of time, before the entirety of the above mentioned steps is computationally re-performed. The code is currently deployed on the website ‘Render’ for automatic retraining over the allotted periodic time slot.

The need of this specific methodology is due to certain factors like the staling of machine learning models with time, referring to the sensitive real world changes is conversation patterns and data leading to the overall decay of the model. Another requirement which is satisfied by continuous retraining is concept drift, which is the terminology defining the phenomenon of propertied of predictive values change over time leading to the change of the concept which the model will be severely unprepared for.

Figure 8 Automated continuous retraining (see online version for colours)

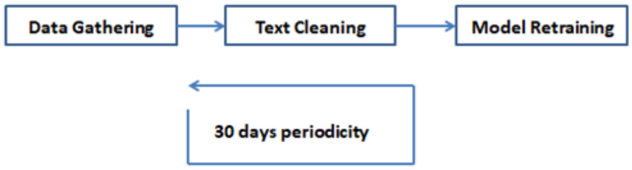


Figure 9 Chatbot model responses (see online version for colours)

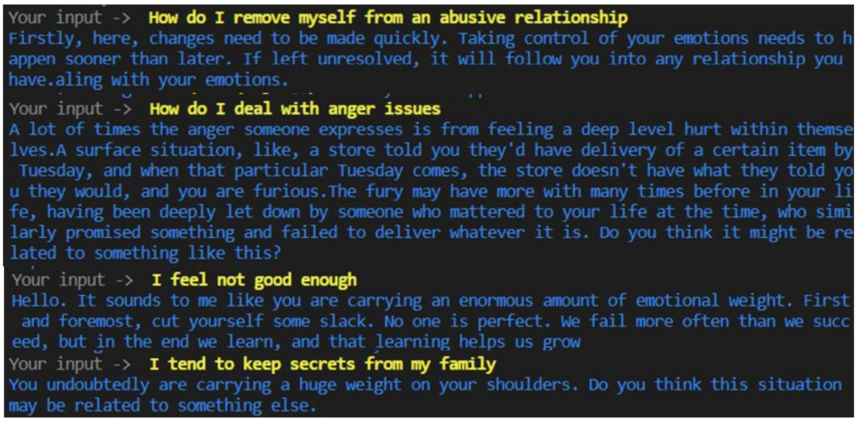
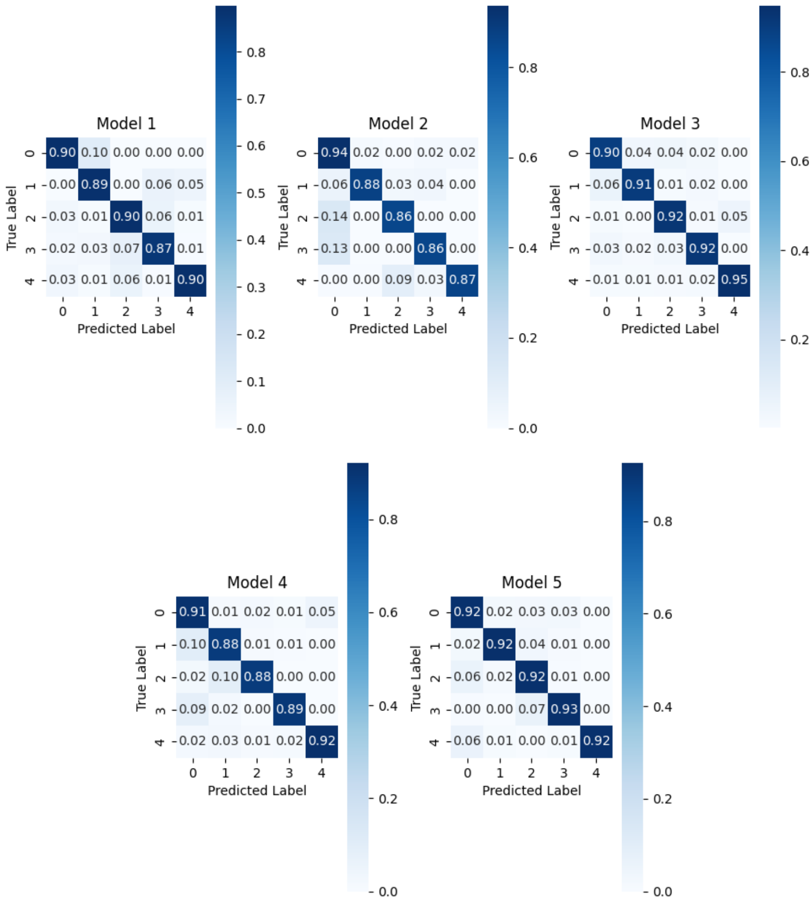


Figure 10 Heat maps for LR, DT, RF, NB and BERT in models 1, 2, 3, 4 and 5 respectively (see online version for colours)



3.3 Response generation

While conversing with the user, the conversational model gives out specific, friendly and knowledgeable responses. These responses are specific to the emotion and context the user is speaking in. The response dataset is curated from and vetted by therapists dealing with these specific problems and allows for a friendly and medical environment during the conversation. When the conversation reaches its end, the classification model would be applied upon the users input and according to the specific mental health problems the user may be facing, and correctly classify the specific mental health medical diagnosis of the user, enabling for the culmination of the conversational model and the classification model simultaneously working for the desired results. The accuracy levels obtained by all of the classification models used are mentioned in Table 1.

Table 1 Performance metrics obtained for the classification models used

Model	*Train accuracy*	*Test accuracy*	*F1 score (avg)*
LR	96.2121	88.6486	0.89
DT	85.1731	86.1621	0.88
RF	93.3982	89.4054	0.91
NB	88.3116	88.1081	0.89
BERT	91.3	89.6	0.92

4 Case study

We shall take into consideration a case study pertaining our proposed framework. Our patient, John, is a 32-year-old male who works as a software engineer. John has consequent issues with persistent sadness, low motivation and disinterest. He also reports fatigue and insomnia. Due to these conditions, his work and personal life have been impacted. John's symptoms have come into existence due to work deadlines and other pressures.

John was provided with the empathetic chatbot for open and inclusive conversation sessions. Due to its professional, clinical and dialogue-based nature, John was able to converse with the chatbot about important points like his current mental state, reasons for it, any previous incidents etc. while simultaneously providing him with empathetic and therapeutic responses. After the required session, the classification model comes into play, and John receives a clinical classification about his current mental health condition, being categorised as depression. After this point, John can either consult a doctor or self-assess his condition to better analyse and deal with the same. He now knows about the steps he needs to take to resolve his issue and the possible reasons leading up to his condition.

5 Experimental analysis

This part of the paper deals with the overall performance of the different classifiers in the terms of metrics. The results obtained from each model has been determined systemically using GridSearchCV. Since, two models were taken into consideration for each classifier withholding different parameters, the model with the better results are taken for visualisation.

The accuracy obtained for the models performed are presented in Table 1. The confusion matrixes obtained are presented in Figure 6. The comparison among various classifiers based on accuracy is presented in Figure 7. The BERT classifier has performed best here with an accuracy of 91.08%.

The receiver operating characteristic (ROC) curve, which defines the classification model at all classification thresholds, it shows the parameters of true positive rate and false positive rate. The ROC curve for the BERT classifier is shown visually in Figure 12. The precision recall curves of the base models are given in Figures 13 to 15.

Figure 11 Comparison graph of base classification models based on test accuracy (see online version for colours)

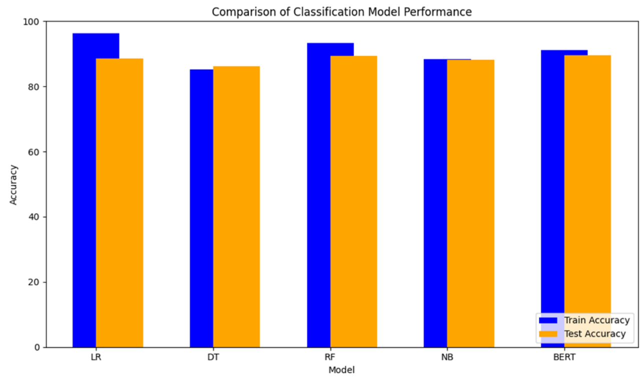


Figure 12 ROC curve for BERT (see online version for colours)

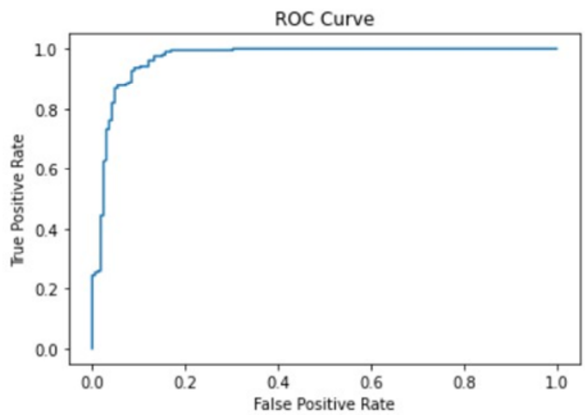


Figure 13 Precision-recall curves for naïve Bayes (both parameters) (see online version for colours)

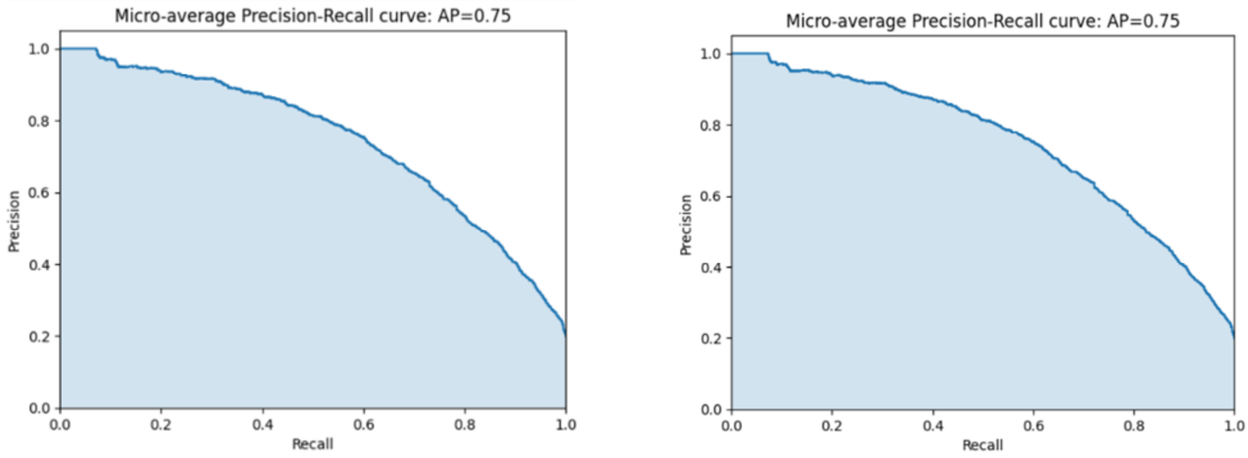


Figure 14 Precision-recall curves for decision tree (both parameters) (see online version for colours)

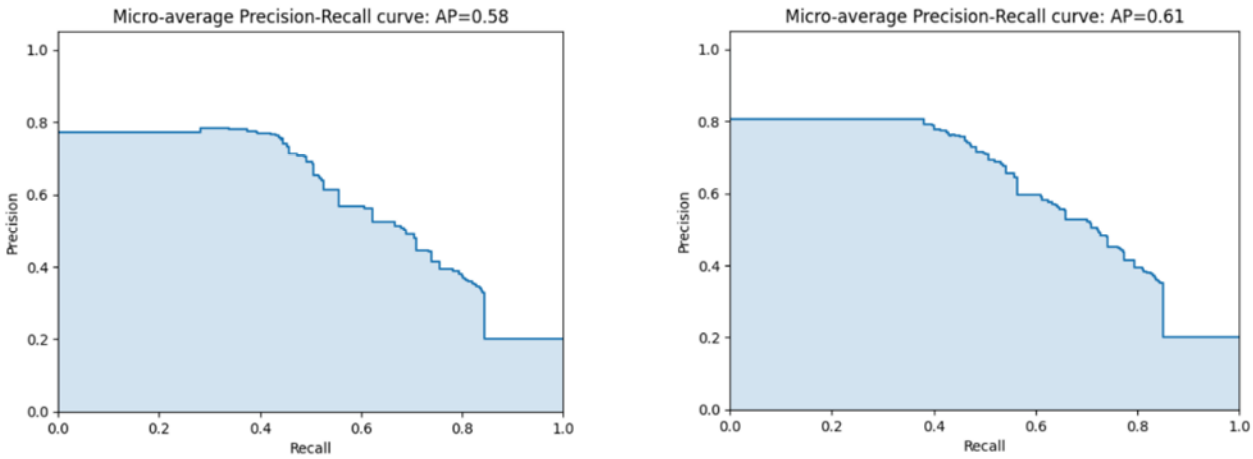
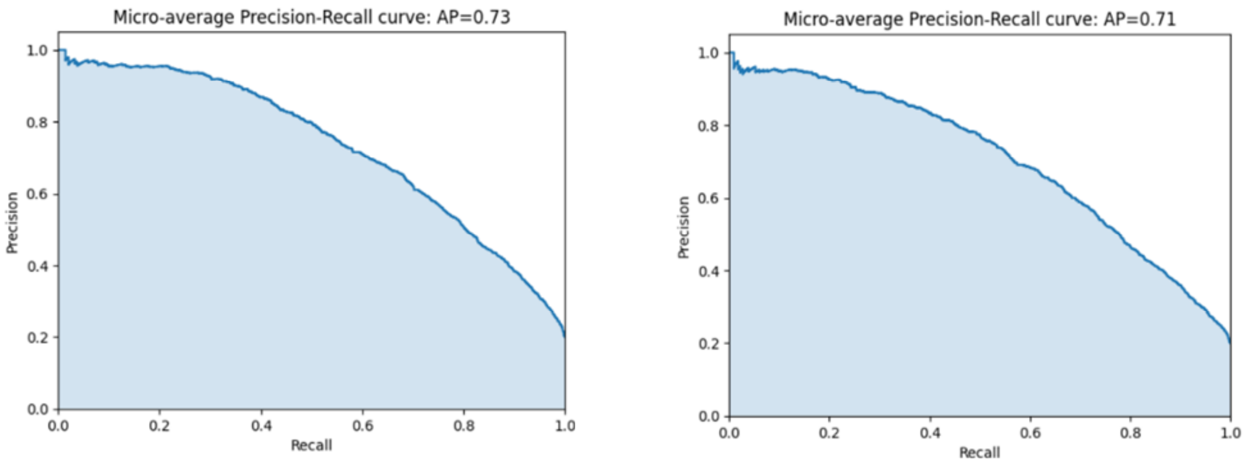


Figure 15 Precision-recall curves for logistic regression (both parameters) (see online version for colours)



6 Discussion

The conversational chatbot provides for an empathetic and dynamic conversation environment allowing for seamless exchange of conversation between the entities. It asks and answers appropriate and leading questions to and from the user and allows for proper data gathering for diagnosis, while simultaneously conversing with the user in a safe environment. Among the classification models the BERT classifier has performed the best with the accuracy of 91.08%. This performance of BERT can be attributed to its inherent mechanism of reading textual data in both left and right directions while simultaneously accounting for all the words in the textual sequence to come up with a better and deeper understanding of the context. Random forest works well here due to its inherent mechanisms for working with missing datapoints in the dataset and for the variance and lack of standardisation of the data used in the model. Logistic regression works well on the data because it assumes that only highly meaningful features are included, and that data is linearly separable. Naïve Bayes works properly because it assumes independence of all variables and classes, and normally requires relatively lesser training data for better outputs. As the data is independent and highly varied, decision tree model does not work properly here as all terms are assumed to interact and independency of variables is not tackled in the mechanism. All the models performed have approximately the same accuracy with slight deviations due to the above-mentioned reasons. The application for this comprehensive model is manyfold and can be used in multiple areas such as an application software, for research surveys and studies, as a medical tool for therapists and professionals etc. Overall, this entire model can be incorporated into multiple scenarios and products allowing for and providing an independent, free and open-source platform for people dealing with mental health issues in and around any required organisations or individuals.

7 Conclusions and future scope

In conclusion, we have created a multifunctioning and cohesive system for detection, prevention and support related to available mental health diseases under study. We have created a chatbot for conversing appropriately with the user in a very specific and medical manner, and enabled classification of the user's input data on whether they may be affected by a mental health disease or not, while simultaneously also providing the classification report of the specific mental health disease that the individual may be suffering with. The system works with two incorporated models which correspondingly work simultaneously to perform and provide the necessary outputs. Regarding the future of the paper, the following recommendations can be made:

- a The chatbot model can be trained using NLU mechanisms to generate automatic responses instead of generalised ones, the model can be added with more conversation data to enable longer conversations.
- b Time varying analysis can also be used when conversing with and obtaining data from the user. The user data can be analysed across time periods and time variance can be taken as a factor while training and creating the classification model.
- c Training data can be scraped in a corresponding manner.
- d The model can be deployed on cloud platforms and multiple virtual servers can be used to provide better efficiency in real time computational use of the software.
- e More Reddit features, like upvotes, downvotes, shares, awards etc. can be also taken into consideration while training the data.
- f More datapoints can be obtained by identifying and collaborating with institutions which work with corresponding issues.

References

Adamopoulou, E. and Lefteris, M. (2020) 'Chatbots: history, technology, and applications', *Machine Learning with Applications*, Vol. 2, No. 1, p.100006.

Bhattacharya, S., Kushwaha, A., Banu, S.K. andq Tripathy, B.K. (2022) 'Customizing backend logic using a chatbot', in Pandit, M., Gaur, M.K., Rana, P.S. and Tiwari, A. (Eds.): *Artificial Intelligence and Sustainable Computing. Algorithms for Intelligent Systems*, Springer, Singapore, https://doi.org/10.1007/978-981-19-1653-3_43.

Brandtzaeg, P.B. and Følstad, A. (2017) 'Why people use chatbots', *International Conference on Internet Science*, pp.377–392, Springer, Cham.

Cameron, G., Cameron, D., Megaw, G. et al. (2017) 'Towards a chatbot for digital counselling', in *Proceedings of the 31st international BCS human computer interaction conference (HCI 2017). BCS Learning & Development*, DOI: 10.14236/ewic/HCI2017.24.

Darcy, A., Aaron, B., Emil, C., Jade, D., Kim, G., Timothy, Y.M., Paul, W. and Athena, R. (2022) 'Anatomy of a Woebot® (WB001): agent guided CBT for women with postpartum depression', *Expert Review of Medical Devices*, Vol. 19, No. 4, pp.287–301.

Fenn, K. and Majella, B. (2013) 'The key principles of cognitive behavioural therapy', *InnovAiT*, Vol. 6, No. 9, pp.579–585.

Fitzpatrick, K.K., Alison, D. and Molly, V. (2017) '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 Mental Health*, Vol. 4, No. 2, p.e7785 (2017).

Gabrielli, S., Silvia, R., Sara, C. and Valeria, D. (2020) 'A chatbot-based coaching intervention for adolescents to promote life skills: pilot study', *JMIR Human Factors*, Vol. 7, No. 1, p.e16762.

Grové, C. (2021) 'Co-developing a mental health and wellbeing chatbot with and for young people', *Front. Psychiatry*, Vol. 11, p.606041. DOI: 10.3389/fpsyt.2020.606041.

Ivanovic, M., Autexier, S., Kokkonidis, M. and Rust, J. (2023) 'Quality medical data management within an open AI architecture – cancer patients case', *Connection Science*, Vol. 35, No. 1, DOI: 10.1080/09540091.2023.2194581.

Katschnig, H. (2010) 'Are psychiatrists an endangered species? Observations on internal and external challenges to the profession', *World Psychiatry*, Vol. 9, No. 1, p.21.

Kim, H., Hyejin Y., Dongkwang, S. and Jang, H.L. (2022) 'Design principles and architecture of a second language learning chatbot', *Language Learning & Technology*, Vol. 26, No. 1, pp.1–18.

Lijo, V.P. and Seetha, H. (2021) 'Tweets sentiment analysis using multi-lexicon features and SMO', *International Journal of Embedded Systems*, Vol. 14, No. 5, pp.476–485.

Narmadha, V., Ajay Krishnan, J.U., Kumar, R.P. and Kumar, R.R. (2019) 'Telepathic virtual assistant', *2019 3rd International Conference ICCCT*, IEEE, pp.321–325.

Nobles, A.L., Eric, C.L., Theodore, L.C., Zhu, S.H., Steffanie, A.S. and John, W.A. (2020) 'Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants', *NPJ Digital Medicine*, Vol. 3, No. 1, pp.1–3.

Panda, B., Panigrahi, C.R. and Pati, B. (2023) 'Patient reviews analysis using machine learning', *International Journal of Computational Science and Engineering (IJCSE)*, Vol. 26, No. 2, pp.111–117.

Reddy, V. (2019) 'Mental health issues and challenges in India: a review', *International Journal of Social Sciences Management and Entrepreneurship (IJSSME)*, Vol. 3, No. 2.

Sekkat, K., Nassim, E.H., Siham, E.A. and Asmaa, C. (2021) 'Designing a chatbot for young diabetics: unprecedented experiment at the endocrinology service of the Casablanca University Hospital', *Endocrine Abstracts*, Vol. 73, Bioscientifica, DOI: 10.1530/endoabs.73.AEP418.

Srividya, V., Tripathy, B.K., Akhtar, N. and Katiyar, A. (2020) 'AgentG: a user friendly and engaging bot to chat for e-commerce lovers', *Computer Reviews Journal*, Vol. 7, No. 1, pp.7–19.

Srividya, V., Tripathy, B.K., Akhtar, N. and Katiyar, A. (2021) 'AgentG: an engaging bot to chat for e-commerce lovers', in Tripathy, A., Sarkar, M., Sahoo, J., Li, KC. and Chinara, S. (Eds.) *Advances in Distributed Computing and Machine Learning. LNNS*, Vol. 127, Springer, Singapore, DOI: 10.1007/978-981-15-4218-3_27.

Sulaiman, S., Marzita, M., Rohaizah, A.W., Nur, A. and Alisa, N.A. (2022) 'Anxiety assistance mobile apps chatbot using cognitive behavioural therapy', *International Journal of Artificial Intelligence*, Vol. 9, No. 1, pp.17–23.

Tracey, P., Mo, S. and Chris, H. (2021) 'Applying NLP to build a cold reading chatbot', *2021 International Symposium on Electrical, Electronics and Information Engineering*, pp.77–80.

Verma, S., Singh, M., Tiwari, I. and Tripathy, B.K. (2022) 'An approach to medical diagnosis using smart chatbot', in Das, A.K. et al. (Eds.): *Proceedings of CIPR*, LNNS 480, pp.1–14.

World Health Organization (2005) *Mental Health Atlas 2005*, World Health Organization, Department of Mental Health, Substance Abuse, Mental Health Evidence and Research Team.