*Journal of Knowledge Learning and Science Technology*
*ISSN: 2959-6386 (Online)*
2024, Vol. 4, No. 1, pp. 52–60
DOI: https://doi.org/10.60087/jklst.v4.n1.006

Research Article

# A Privacy-Focused, Adaptable Chatbot for Detecting and Preventing Depression

**Matviy Amchislavskiy**[1*]

[1]The Governor's Academy, Byfield, USA

## Abstract

According to the World Health Organization, approximately 280 million people have depression. However, mental health diagnostic and intervention tools remain largely inaccessible, unaffordable, and stigmatized. To address this global need, a novel, accessible, and nonintrusive diagnostic and assistive system was created through the development of two machine learning (ML) models and a web app. Psych2Go uses two ML models to nonintrusively detect depression (model 1) and emotion (model 2). Both achieve their respective goals by analyzing prosodic features in speech rather than the content. The first ML model achieves a depression detection accuracy of 75.54% and the second achieves an emotion detection accuracy of 77.60%. The assistive system is powered by the GPT-3.5 Turbo API. The API, using a custom prompt template, tailors responses and therapy techniques to the user-provided demographic information (name, gender, age) and the detected emotion from the second ML model. The prompt enables the GPT-3.5 Turbo API to apply cognitive behavioral therapy principles, identifying and addressing depression-related negative thoughts. Adhering to strict privacy standards, the chatbot eschews storage of personal conversations, focusing instead on session-specific data (the user's time of a session, depression score, emotion, name, age, and gender). Psych2Go was deemed successful in providing privacy-focused, personalized emotional support and depression and emotion detection. The chatbot's unprecedented privacy-focused approach and personalization allows for it to act as an aid for therapists to monitor progress and a support system for the user between therapy sessions.

## Keywords

Artificial intelligence, deep learning, depression, social functioning, speech, chatbot

[*]Corresponding author: Matviy Amchislavskiy

**Email addresses:** matvei.amchislavskiy@govsacademy.org

**Received:** 01-11-2024; **Accepted:** 01-12-2024; **Published:** 06-12-2024

# 1. Introduction

Mental health comprises people's psychological, emotional, and social welfare as a state of well-being in which individuals realize their abilities, cope with normal life stressors, work productively, and contribute to their communities (Galderisi et al., 2015). Conversely, mental illness refers to functional disorders characterized "by anxiety (neurosis) and/or disorganization (psychosis)" (Gove & Tudor, 1973). Various factors, such as genetics, environment, and life experiences, contribute to the development of mental illnesses (Inglis et al., 2016). Many mental illnesses impair one's ability to think, feel, and relate to others. While the scope of mental illness's prevalence varies across countries, at any given time, approximately one in five adults experience it (Gopalkrishnan, 2018). Notably, mental illness is not limited to a specific demographic, with mood and anxiety disorders (e.g., generalized anxiety disorder and major depressive disorder) as the most common types. However, most people with either a mood or anxiety disorder are untreated or misdiagnosed (Puyat et al., 2016).

Unlike physical ailments, where symptoms may be outwardly evident, the internal struggles associated with mood and anxiety disorders can often go unnoticed by the casual observer (Kasper, 2006). This invisibility fosters a lack of awareness, perpetuating various stereotypes associated with these disorders. For instance, some view those with depression as having weak personalities (Yokoya et al., 2018), while individuals with post-traumatic stress disorder may be labeled as crazy (Mittal et al., 2013). These misinformed perspectives lead to stigma, a mark of disgrace or discredit that sets a person apart from others.

Many individuals avoid seeking care for their mental health issues because of the social distance others may maintain from persons seeking psychiatric therapy or the self-stigma they experience due to earlier discrimination (e.g., internalized shame; Schulze, 2007). In the absence of treatment, these conditions may worsen, having devastating effects on a person's mental and physical health (Borenstein, 2020). Although seeking psychotherapeutic or medical aid for mood or anxiety disorders may make individuals uncomfortable, at least temporarily, it is a crucial step to preventing the condition from worsening and affecting daily life (National Alliance on Mental Illness California, n.d.). In the realm of treatments for mood and anxiety disorders, two primary methods have emerged as the most prevalent: psychotherapy and medication (Bandelow et al., 2022).

Psychotherapy is a broad range of treatments that can help individuals understand their illnesses, cope with their reality, and adjust to their challenges (National Institute of Mental Health, 2023). In comparison, medication involves the use of drugs to treat the symptoms and underlying causes of these disorders. Drugs are usually palliatives rather than cures (Parish et al., 2023). Among the various psychotherapeutic approaches, cognitive behavioral therapy (CBT) has been one that has received much attention.

CBT is a structured, goal-oriented therapy that aims to identify and change negative thought patterns and behaviors that contribute to and perpetuate mood and anxiety disorders. The primary goal of CBT is to equip individuals with practical skills to manage their disorder and reduce the risk of relapse. This form of therapy is grounded in the belief that people's thoughts, feelings, and behaviors are interconnected, so by identifying and addressing maladaptive thought patterns, one can bring about positive changes to their emotional well-being and actions (Parish et al., 2023).

Although CBT has been shown to be effective, its accessibility is limited since some individuals are embarrassed to enter a therapist's office, admit they need treatment, or pay for its cost. In response to these limitations, in 1966, the first chatbot psychotherapist, ELIZA, was created using pattern matching (Bendig et al., 2022) and template-based answers to implement a therapy approach called Socratic questioning. Today, Woebot (Woebot Health, Inc.), Youper (Youper, Inc.), Wysa (Wysa, Ltd), Replika (Luka, Inc.), Unmind (Unmind, Inc.), and Shim (Shim, Inc.) are a few examples of the health-focused chatbots publicly available as mobile app features (Xu et al., 2021). Woebot, based on CBT, was studied and shown to be effective in reducing depressed symptoms and increasing receptivity in patients compared to conventional treatments (Fitzpatrick et al., 2017). Consistent with the findings from Shim, who used the same therapeutic approach, a study found that the intervention was extremely engaging, increased well-being, and decreased stress (Ly et al., 2017). The user's motivation increased while tension decreased when a chatbot was created using the structured association-approach counseling method.

Similarly, a graph-based chatbot was suggested to utilize sentiment analysis to determine users' emotional states and engage in conversations to alleviate their distress (Xu et al., 2021). Moreover, cognitive and behavioral therapies provided by Vivobot (HopeLab, Inc.) aim to disseminate knowledge of positive psychology and foster flourishing (Xu et al., 2021). Young individuals undergoing cancer treatment found this chatbot helpful for emotional support and engagement.

However, chatbots are not where the ability of machine learning (ML) in mental health ends. By examining data for patterns that possibly indicate a mental health disorder, ML is used to make diagnoses (Shatte et al., 2019). Patient charts, brain scans, and social media postings are examples of where this information may be gathered and analyzed. Both

supervised learning algorithms (such as the convolution neural network [CNN] used in this study), requiring labeled data for training, and unsupervised learning algorithms, which discover hidden patterns in data without labels, have been utilized for this purpose (Osisanwo et al., 2017). Once a model has been trained on the acquired information, it can estimate the probability that a person has a specific mental health condition based on data. Researchers working in ML produce this forecast by feeding fresh data into their models and interpreting the results to draw conclusions. Nevertheless, the traditional method of diagnosis via questionnaires (such as the PHQ-9) is limited to the disorders they cover, while the invasive and vague nature of the questionnaire can make people uncomfortable.

The primary aim of my research was to develop a novel CBT chatbot utilizing the Generative Pre-trained Transformer (GPT)-3.5 Turbo Application programming interface (API) that is both privacy-focused and personalized. This chatbot was designed to discern emotion and depression indicators from prosodic features in spoken language. My approach was bifurcated into two primary components: 1) the detection of emotion and depression from audio and 2) the creation of an interactive CBT chatbot interface.

In the first phase, I created two ML model. One detected depression and the second emotion based on prosodic features in speech. The audio, based on which the two ML models were trained, was sources from two databases: the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Distress Analysis Interview Corpus (DAIC). Leveraging the capabilities of the Librosa Python module, I extracted Mel-frequency cepstral coefficients (MFCCs) from these samples, forming the bedrock of my 2D CNN. This model underwent extensive training, encompassing several layers, which equipped it to better detect emotion and depression.

The second phase witnessed the integration of these ML models with a web application developed in ReactJS. This integration facilitated interactions between the models and the application via the Hugging Face inference endpoint, ensuring that the detected emotional state and depression level, augmented with user-specific data such as gender, age, and name, were effectively employed to tailor the chatbot's interaction with the user. This level of customization enabled my chatbot, built from OpenAI's GPT-3.5 Turbo API with a custom prompt template, to conduct adaptive CBT sessions personalized for each user.

My research focused on developing a CBT chatbot to offer accessible, privacy-centric, and tailored mental health support. The study's objectives were: 1) to surpass the detection accuracy rate of primary medical providers' initial diagnoses, currently around 50% (Al-Huthail, 2008); 2a) to provide highly personalized emotional support sessions, grounded in CBT principles, by considering the user's emotion and demographic data (name, age, and gender) to select the most appropriate CBT technique for each individual; and 2b) to nonintrusively detect depression and emotion by identifying prosodic features in speech, avoiding the analysis of recording contents or text chats.

## 2. Materials and Methods

The primary aim of this section was to develop a ML framework for detecting emotion and depression based on prosody in speech. This study used the RAVDESS database for emotion detection, which contains audio files of different emotions, including anger, calmness, fear, happiness, and sadness. For detecting depression, I utilized the DAIC database, in which there were interviews with participants exhibiting varying degrees of depression.

I used the audio files from the RAVDESS database, consisting of audio-visual recordings of performers displaying a range of emotional expressions. Only five emotions (i.e., anger, calmness, fear, happiness, and sadness) were chosen for investigation to reduce any confusion in emotion categorization. The acquired WAV dataset was split 80/20 between training and testing sets for each emotion. Figure 1 shows the distribution of target emotions in the RAVEDESS database.

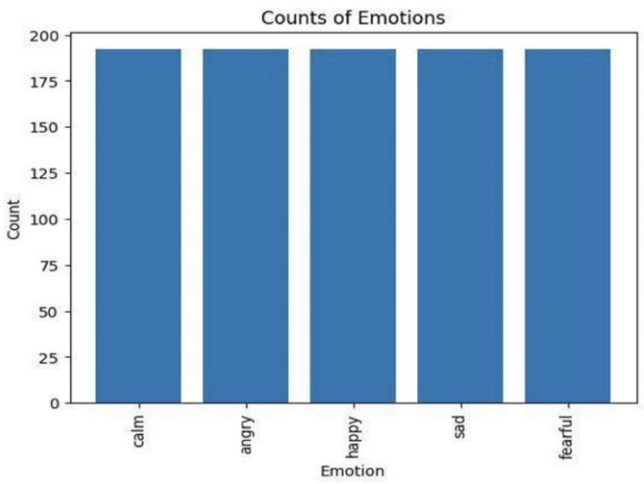


***Figure 1.*** *Distribution of Target Emotions*

The DAIC database, an assembly of audio-visual recordings featuring interactions with a virtual interviewer, was used to detect depression through speech and language patterns. I specifically utilized the audio component of the database, acquired in WAV file format. To refine the database for analysis, several preprocessing steps were executed: firstly, noise reduction techniques were employed to eliminate background

disturbances such as hissing, humming, and buzzing, thereby enhancing audio clarity. Secondly, speech diarization was conducted to isolate and remove the virtual interviewer's voice, which was deemed irrelevant noise in the context of this study. Thirdly, the speech segments pertaining to individual speakers were amalgamated into singular, continuous recordings. The database's composition, as depicted in Figure 2, revealed a skewed distribution of targets. Both databases were subsequently divided into training and testing sets, adhering to an 80/20 split ratio. The subsequent phase involved feature extraction from these preprocessed databases.

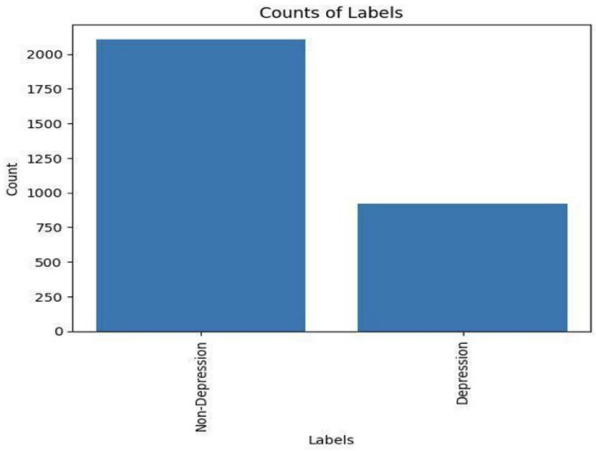


***Figure 2.*** *Target Distribution of Depression Audio*

The Librosa Python module was used to extract MFCCs from both databases. MFCCs are widely used in speech recognition and emotion detection tasks, as they provide a compact representation of the audio signal's spectral characteristics. The extracted MFCCs are used as feature vectors, with a shape of (30, 2584), which are then reshaped into (30, 2584, 1) to serve as input for the 2D CNN model. Similarly, for the DAIC database, MFCCs are extracted using the same Librosa library. The feature vectors in this case have a shape of (30, 216), and they are also reshaped into (30, 216, 1) to be used as input for the 2D CNN model.

The model architecture used for both emotion and depression detection is based on a 2D CNN implemented using TensorFlow. It consists of several layers. The first layer is a 2D convolutional layer with 256 filters and a kernel size of 5, using the ReLU (rectified linear unit) activation function. This layer is applied to the input feature vector to learn various features and patterns from the input data. Then, there is another 2D convolutional layer with 128 filters and a kernel size of 5, also using the ReLU activation function. This layer further refines the features learned by the previous layer. A dropout layer with a dropout rate of 0.3 follows, which helps in reducing overfitting by randomly setting input units to 0 during training. Thereafter, a max-pooling layer that downsamples the input feature map to reduce spatial dimensions and computational complexity is also added. Then a third 2D convolutional layer with 128 filters and a kernel size of 5, using the ReLU activation function, followed by another 2D convolutional layer with 64 filters and a kernel size of 5, also using the ReLU activation function is integrated. Yet another dropout layer with a dropout rate of 0.3 is included to further prevent overfitting. A flatten layer that reshapes the 3D output tensor into a 1D tensor, allowing it to be used as input for the subsequent dense layer was also incorporated. Finally, a dense output layer with five nodes and a softmax activation function is applied to the emotion detection model, while for the depression detection model, a dense output layer with one node and a sigmoid activation function is used. The total number of parameters for the emotion detection model is 2,682,245, while the total number of parameters for the depression detection model is 2,680,897. These parameters are learned during the training process to optimize the model's performance in recognizing emotions or detecting depression from the input audio features.

## 2.1. Hugging Face

The solution deployment used a custom Hugging Face inference endpoint for the ML models, complemented by a web server for the interface. Hence, a Supabase PostgreSQL database was created by accessing the database section of the project to retrieve the PostgreSQL connection string, which was essential for linking the database to the project. Simultaneously, a new model repository was created under the "Models" section within the Hugging Face interface.

The next step pushed the ML models and the inference code to this repository, which begins with cloning the repository to a local machine using Git. The ML models and a handler.py file containing the inference code were added to this repository. This included a requirements.txt file, which was also imperative to list all dependencies needed to run the inference code. This comprehensive setup ensured seamless integration of the ML models with the web interface, providing an efficient deployment.

The subsequent phase involved creating a custom inference API and setting up the front-end repository. The custom inference API was created by navigating to the solutions tab on the Hugging Face platform and then to the "Inference API" tab. Clicking it produced an option to "Create Inference API," which led to a setup to choose the specific branch and the Python runtime version that best suited the project.

Subsequently, attention was shifted toward cloning the front-end repository, which involved laying out the structure for the front-end part of the project with various dependencies to be installed to function correctly. The first dependency was

Langchain, installed using the npm command. Langchain was essential for integrating language processing capabilities into the front end. The second dependency was the Supabase client. This client enabled the front end to interact with the Supabase database, allowing for operations like data retrieval.

Finally, setting up database integration involved configuring environment variables in the project, including "NEXT_PUBLIC_SUPABASE_URL" and "NEXT_PUBLIC_SUPABASE_ ANON_KEY," which were set to the specific Supabase URL and anonymous key, respectively. These environment variables allowed the front-end application to communicate securely and effectively with the Supabase database, ensuring it had the necessary permissions and pathways to access and manipulate data as required. Through these steps, the front end was effectively prepared to work with the back end, creating a seamless integration between the ML models and the user interface.

The next steps involved database configuration and front-end development for a web application. Firstly, within the Supabase dashboard, the SQL editor was used to create a specific table to store user interactions with columns for "Name," "Age," "Gender," "Depression," and "Timestamp." The development then shifted toward creating the homepage of the application.

The homepage's primary functionality revolves around user interaction through a chat interface. The key tasks for this page include displaying a prompt that asks the user to talk about their day and asking the user to read the prompt. Once the user reads the prompt, the application records the response by providing a sample code structure that outlines the process of capturing audio input from the user using JavaScript's "navigator.mediaDevices.getUserMedia" method.

The code snippet checks if the user's browser supports the "getUserMedia" API, which is necessary for accessing the device's audio input capabilities. If supported, the code proceeds to request access to the audio input. The "getUserMedia" method is configured to capture only audio data ("{ audio: true }"). A stream is obtained upon successful access to record the user's voice after the prompt is read. In case of an error or if the browser does not support "getUserMedia," appropriate messages are logged to the console.

The next phase involves the interaction between the recorded audio and the Hugging Face API. The recorded audio is sent to the Hugging Face API. The API processes this audio input using the two ML models created prior.

This interaction between the application and the Hugging Face API is executed through a specific request-response mechanism. The request sent to the Hugging Face API is structured in a JSON format. It includes a key named "inputs," within which the "audio_data" is embedded. This "audio_data" is the recorded audio from the user, encoded in base64 format. Base64 encoding is a standard method to convert binary data (in this case, audio) into a string format, making it suitable for transmission over the network. Upon receiving this request, the Hugging Face API processes the audio data. It utilizes ML capabilities to analyze the audio input, identifying the emotion and a depression score based on prosody in the user's speech. The response from the Hugging Face API is also in a JSON format, comprising two key pieces of information: "emotion" and "depression_score." The "emotion" field contains a string representing the detected emotional state from the audio, while the "depression_score" field conveys a string that quantifies the score of depression.

Subsequently, interactive user engagement and data handling are established, leveraging the capabilities of Hugging Face, Langchain, and Supabase. After receiving the text response from the Hugging Face API, which includes the user's emotional state and depression score, the application displays a form to the user to collect the user's name, age, and gender. Next, the process involves creating a Langchain ChatPromptTemplate, which is achieved using the "fromTemplate" method of "SystemMessage." The template generates a chat prompt, as shown in Figure 3, incorporating placeholders for the data received from the user and the Hugging Face API.

A detailed prompt for using CBT techniques with the user was created through trial-and-error. This prompt is designed to guide users through simulated therapy sessions, starting with assessing the most suitable CBT technique for the users' problems, followed by step-by-step navigation through the therapy process. The prompt is personalized with placeholders for users' names, genders, ages, and detected emotions. The aim is to initiate personalized conversations with the users and assist them in finding resolutions to their issues, adapting the approach if necessary. Now, the application provides a responsive chat interface and serves as a virtual therapist, assisting users through a therapeutic session tailored to their emotional states and personal information. This approach enhances the application's utility in providing mental health support and personal development guidance.

Prompt : I'll ask you to practice a simulation of the Cognitive-Behavioral Therapy techniques (cognitive restructuring, behavioural activation, exposure therapy, problem-solving skills training, relaxation and stress management, assertiveness training, and thought records and journaling) with me. First, you need to assess which technique will suit better to my problem. Then tell me how the process will go on. Then start navigating me through it step by step. Navigate me through this process as though you are my psychologist. Start with the first step, then stop and wait for my response until you go on; your response has to be feedback on my reaction. Don't list all of the steps from the start. My name is [your_name], my gender is [your_gender], and my age is [your_age]. My main issue is [detected_emotion]. Initiate my inner conversation and help me to come up with some resolution. If one technique does not work out, proceed with the next one.

***Figure 3.*** *Prompt Template*

Then, a MemoryPlaceholder and a HumanMessage complete the ChatPromptTemplate. The methodology then called for creating a Supabase client to upload data from each session including the user's name, age, gender, emotion detected, depression index (0-1), and timestamp of the session to the previously set up Supabase database. This step is crucial for maintaining a record of a user's emotional assessments. Furthermore, an LLMChain was created using ChatOpenAI as the language learning model (LLM) with GPT-3.5 Turbo as the chosen model. The prompt created earlier was used as the input for this LLMChain. The LLMChain's memory type was set to "ConversationSummaryMemory" to manage the conversation context effectively, which involved creating a new memory instance using the "ConversationSummaryMemory" and "LLMChain" from Langchain.

Next, I implemented user interaction and integrated the ChatOpenAI model into the web application. After the initial setup, the next step was to invoke the "run" method on the LLMChain with an empty input. This action prompted the chain to return data from the OpenAI API, serving as a starting point for the conversational interaction while allowing the application to initialize the chat session and prepare for user input.

Subsequently, the application was configured to enable text chat input. Hence, the "run" method is called whenever users enter messages. This functionality is critical for creating an interactive chat experience where the application can process and respond to user inputs in real time. Modifications were made to the "pages/home/index.tsx" file of the application to implement these features. Moreover, setting the OpenAI API key in the ".env" file was imperative for the application to authenticate and interact with the OpenAI services.

# 3. Results

## 3.1. Emotion Detection

The experiment for emotion detection was performed by using the RAVDESS database, which consists of audio samples for five different emotions - anger, calmness, fear, happiness, and sadness. The model was trained on the train split (80%) and evaluated on the test split (20%). The training process involved 100 epochs, using the Adam optimizer with a learning rate of 2e-6 and a batch size of 16. The loss function employed was sparse categorical cross-entropy.

Model checkpointing was employed to retrieve the best model weights throughout the training process. The test loss and test accuracy recorded after the evaluation were 0.6908 and 0.7760, respectively. The model's loss and accuracy plots during training and evaluation are presented in Figures 4 (loss) and 5 (accuracy).

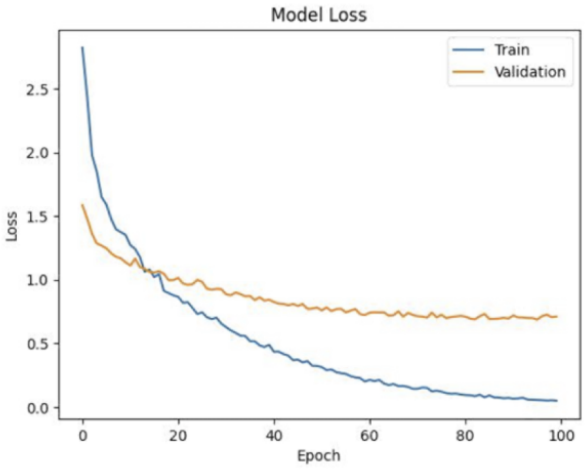


***Figure 4.*** *Emotion Model Loss Plot*

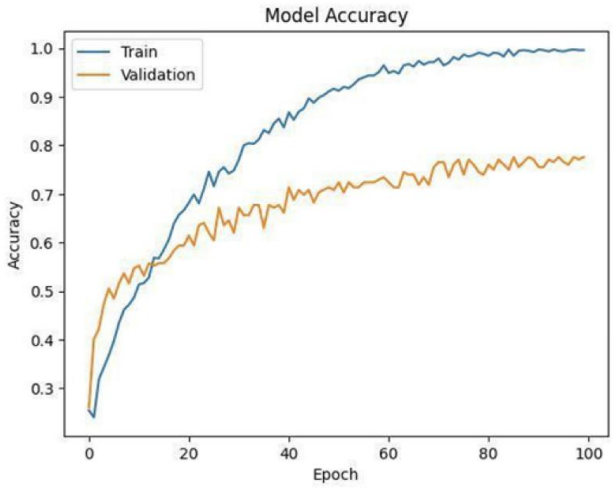


*Figure 5. Emotion Model Accuracy Plot*

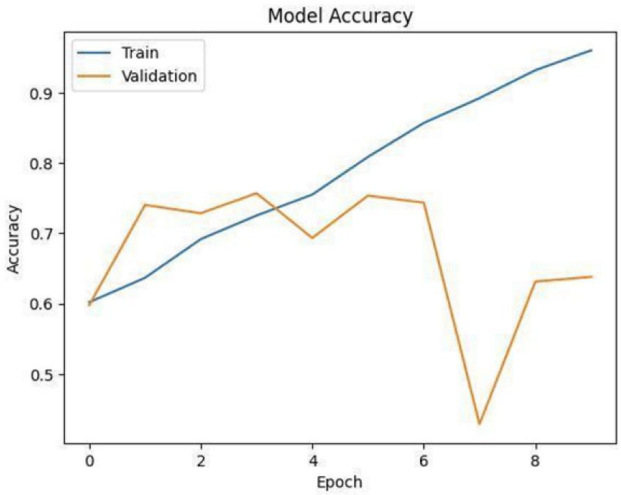


*Figure 7. Depression Model Accuracy Plot*

## 3.2. Depression Detection

The depression detection experiment was conducted using the DAIC database. Similar to the emotion detection experiment, the model was trained on the train split (80%) and evaluated on the test split (20%). The training was carried out for 10 epochs, using the Adam optimizer with a learning rate of 5e-6 and a batch size of 8. The loss function used for this experiment was binary cross-entropy.

Model checkpointing was applied to obtain the best model weights during training. The test loss and accuracy recorded after the evaluation were 0.5356 and 0.7554, respectively. The model's loss and accuracy plots during training and evaluation are presented in Figures 6 (loss) and 7 (accuracy).

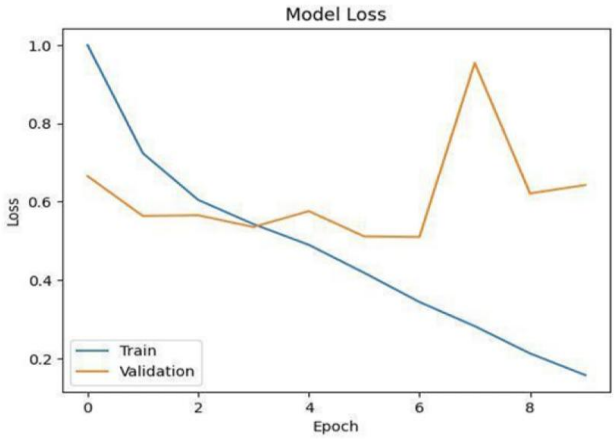


*Figure 6. Depression Model Loss Plot*

## 4. Discussion

This study affirms objectives 1, 2a, and 2b, underscoring the success of the project in developing a privacy-focused chatbot that provides personalized emotional support using CBT principles and detects depression and emotion.

The project uses two main databases: RAVDESS for emotion detection and DAIC for depression detection. These databases were preprocessed with a number of techniques, including noise reduction, speech diarization, and conversion into continuous recordings. Then the preprocessed datasets were split into training and testing sets. Feature extraction was a critical step, employing the Librosa Python module to extract MFCCs from both databases. These features were used as input for a 2D CNN model implemented using TensorFlow. The CNN model architecture consisted of several layers, including convolutional layers with ReLU activation, dropout layers, max-pooling layers, and a dense output layer. The first model was trained to detect five different emotions and the second to detect depression. Deployment involved using a Hugging Face inference endpoint and a web server interface. A Supabase PostgreSQL database was set up for data management. The front-end included capturing audio input and sending it to the Hugging Face API for processing along with a chat interface for user interaction. The API analyzed the audio to detect emotion and depression. Interactive user engagement was facilitated by integrating Hugging Face, Langchain, and Supabase. A chat prompt template for personalization of the therapy sessions was created, leveraging a Langchain ChatPromptTemplate with the GPT-3.5 Turbo API. This approach followed the prior literature on detecting emotions and depression using prosody in speech.

For example, Afshan et al. (2018) focused on the effectiveness

of voice quality features in detecting depression. They used voice quality features (F0, F1, F2, F3, H1-H2, H2-H4, H4-H2k, A1, A2, A3, and CPP) inspired by a psychoacoustic model of voice quality with MFCCs. The model was validated with an 80/20 split of each database into training and testing sets. The validation phase was crucial in fine-tuning the model parameters, ensuring an optimal balance between sensitivity and specificity in detecting depression.

My ML model for depression detection achieved a peak diagnostic accuracy of 75.54% (see Figure 7), along with a loss rate of 53.56% (see Figure 6). This performance surpasses the initial diagnostic accuracy rate of primary medical providers, which is approximately 50%. Furthermore, my ML model for emotion detection achieved a peak accuracy of 77.60% (see Figure 5), along with a loss rate of 69.08% (see Figure 4). The support was personalized due to the use of a custom prompt template (see Figure 3), enabling the GPT-3.5 Turbo API to use the emotion detected by the ML model along with the users' self-reported demographics to offer highly personalized emotional support based on CBT principles. Importantly, the chatbot preserved privacy by relying on prosodic speech patterns instead of content analysis to detect emotion and depression. It did not store actual conversation details but retained only demographic data, the detected emotion, a depression index (0 to 1), and a timestamp for each session. Hence, Psych2Go was deemed successful in diagnosing depression with a greater accuracy than a primary medical provider (objective 1), providing highly customizable emotional support based on CBT principles by considering the user's emotion and demographic information (objective 2a), and detecting depression and emotion nonintrusively using prosody in speech (objective 2b).

The study's results align with previous research. For instance, Afshan et al. (2018) reported a 77% accuracy in depression detection using 10-second voice samples. Additionally, Hema and Marquez (2023) achieved a 78% peak accuracy in emotion detection based on pitch and loudness, highlighting the feasibility of using prosody in speech for both depression and emotion detection. Psych2Go, however, is unique in its characteristics as a chatbot. Unlike existing chatbots like Woebot that directly inquire about emotions and do not detect depression, my chatbot uniquely offers tailored techniques based on user demographics and detected emotion.

A limitation of this study is the distribution of data. Figure 2 shows far more non-depressed than depressed audio samples, leading to lower accuracy in depression detection. Furthermore, a limitation of the chatbot is that while it can constantly adjust to users' needs, it does not build on prior conversations to preserve maximum privacy for the user.

Future parts of this study could increase the amount of training data and improve the model's generalization capabilities. Moreover, I could apply data augmentation techniques such as pitch shifting, time stretching, and adding background noise to the audio files to investigate using pre-trained models like OpenL3 and VGGish on the audio files to help improve feature extraction and decrease training time.

## 5. Conclusions

The significance of my research lies in its innovative approach to mental health support through the use of prosodic speech analysis. By focusing on the nuances of a user's speech patterns, the ML models are able to detect emotion and depression. Beyond this, the custom prompt integrates a user's demographic information to tailor their therapeutic experience. This dual focus allows for the creation of a highly personalized CBT framework, adaptable to each individual's emotion and unique demographic profile. Moreover, the nonintrusive nature of this technology ensures that it can be seamlessly integrated into daily life, providing users with accessible mental health support without the stigma or inconvenience often associated with traditional therapy methods. In practical application, this research has the potential to function as a supplementary tool for professional therapists. It could be used to monitor client progress and act as a support system for the user between therapy sessions. This represents a significant step forward in bridging the gap between technology and mental health, offering a novel, efficient, and sensitive approach to therapy that respects and responds to the individual needs of each user.

## Author Contributions

Matviy Amchislavskiy: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review and editing.

## Funding

This work is not supported by any external funding.

## Data Availability Statement

The data supporting the outcome of this research work has been reported in this manuscript.

## Conflicts of Interest

The author declares no conflicts of interest.

## References


[1] Afshan, A., Guo, J., Park, S. J., Ravi, V., Flint, J., & Alwan, A. (2018). Effectiveness of voice quality features in detecting depression. *Proceedings of Interspeech 2018*, 1676–1680. https://doi.org/10.21437/interspeech.2018-1399

[2] Al-Huthail, Y. R. (2008). Accuracy of referring psychiatric diagnosis. *International Journal of Health Sciences*, *2*(1), 35–38. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3068718/#:~:text=The%2520accuracy%2520of%2520psychiatric%2520diagnosis

[3] Bandelow, B., Michaelis, S., & Wedekind, D. (2022). Treatment of anxiety disorders. *Generalized Anxiety Disorders*, *19*(2), 93–107. https://doi.org/10.31887/dcns.2017.19.2/bbandelow

[4] Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2022). The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health – A scoping review. *Verhaltenstherapie*, *32*(S1), 64–76. https://doi.org/10.1159/000501812

[5] Borenstein, J. (2020, August). *Stigma, prejudice and discrimination against people with mental illness*. American Psychiatric Association. https://www.psychiatry.org/patients-families/stigma-and-discrimination

[6] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (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*, *4*(2), Article e19. https://doi.org/10.2196/mental.7785

[7] Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2015). Toward a new definition of mental health. *World Psychiatry*, *14*(2), 231–233. https://doi.org/10.1002/wps.20231

[8] Gopalkrishnan, N. (2018). Cultural diversity and mental health: Considerations for policy and practice. *Frontiers in Public Health*, *6*, Article 179. https://doi.org/10.3389/fpubh.2018.00179

[9] Gove, W. R., & Tudor, J. F. (1973). Adult sex roles and mental illness. *American Journal of Sociology*, *78*(4), 812–835. https://doi.org/10.1086/225404

[10] Hema, C., & Marquez, F. (2023). Emotional speech recognition using CNN and deep learning techniques. *Applied Acoustics*, *211*, Article 109492. https://doi.org/10.1016/j.apacoust.2023.109492

[11] Inglis, A., Morris, E., & Austin, J. (2016). Prenatal genetic counselling for psychiatric disorders. *Prenatal Diagnosis*, *37*(1), 6–13. https://doi.org/10.1002/pd.4878

[12] Kasper, S. (2006). Anxiety disorders: under-diagnosed and insufficiently treated. *International Journal of Psychiatry in Clinical Practice*, *10*(S1), 3–9. https://doi.org/10.1080/13651500600552297

[13] Ly, K. H., Ly, A.-M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, *10*, 39–46. https://doi.org/10.1016/j.invent.2017.10.002

[14] Mittal, D., Drummond, K. L., Blevins, D., Curran, G., Corrigan, P., & Sullivan, G. (2013). Stigma associated with PTSD: Perceptions of treatment seeking combat veterans. *Psychiatric Rehabilitation Journal*, *36*(2), 86–92. https://doi.org/10.1037/h0094976

[15] National Alliance on Mental Illness California. (n.d.). *About mental illness*. https://namica.org/what-is-mental-illness/

[16] National Institute of Mental Health. (2023, January). *Psychotherapies*. https://www.nimh.nih.gov/health/topics/psychotherapies#:~

[17] Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138. https://doi.org/10.14445/22312803/ijctt-v48p126

[18] Parish, A. L., Gillis, B., & Anthamatten, A. (2023). Pharmacotherapy for depression and anxiety in the primary care setting. *Journal for Nurse Practitioners*, *19*(4), Article 104556. https://doi.org/10.1016/j.nurpra.2023.104556

[19] Puyat, J. H., Kazanjian, A., Goldner, E. M., & Wong, H. (2016). How often do individuals with major depression receive minimally adequate treatment? A population-based, data linkage study. *Canadian Journal of Psychiatry*, *61*(7), 394–404. https://doi.org/10.1177/0706743716640288

[20] Schulze, B. (2007). Stigma and mental health professionals: A review of the evidence on an intricate relationship. *International Review of Psychiatry*, *19*(2), 137–155. https://doi.org/10.1080/09540260701278929

[21] Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448. https://doi.org/10.1017/s0033291719000151

[22] Xu, L., Sanders, L., Li, K., & Chow, J. (2021). Chatbot for healthcare and oncology applications using artificial intelligence and machine learning: Systematic review. *JMIR Cancer*, *7*(4), Article e27850. https://doi.org/10.2196/27850

[23] Yokoya, S., Maeno, T., Sakamoto, N., Goto, R., & Maeno, T. (2018). A brief survey of public knowledge and stigma towards depression. *Journal of Clinical Medicine Research*, *10*(3), 202–209. https://doi.org/10.14740/jocmr3282