iPonder: A Novel Approach to Multimodal Mental Health Diagnosis and Therapy for Teens via Deep Transfer Learning and Computer Vision

Vivek Kogilathota
Computer Science and Engineering Rick Reedy High School
Frisco, United States
vkogilathota@gmail.com

2021 IEEE International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC) | 978-1-6654-4102-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIRC52191.2021.9545055

***Abstract*—Suicide is the second leading cause of death for teenagers, tarnishing the future potential of the world. More teenagers die from suicide than from cancer, heart disease, stroke, and many more diseases, combined. Although the greatest mental health improvement occurs from a conversation, over 50% of suicidal teenagers stay quiet about their depression due to shame. Thus, iPonder, a cross-platform accessible mobile application for teenagers to privately discuss, confront, and overcome their depression, was created. Using a four-level diagnosis approach based on user data, responses to psychological questionnaires, choice specification, and free-response answers, iPonder constructs a hyper-personalized user profile, which can then be used to pair teenagers anonymously online with similar profiles to discuss and vent about their issues. Questionnaires include the Patient Health Questionnaire-9, General Anxiety Disorder- 7, UCLA Loneliness Scale, and Myers-Briggs Personality Type Indicator test. Using Reddit's Self-Reported Mental Health Diagnosis text-corpus with over 200,000 data points, 11 artificial neural networks were constructed to predict user risk of various mental illnesses based on their free-text-responses, using transfer learning from a token-based text embedding neural network with over 48 million parameters trained on the English Google News 7B corpus with over 4 billion data points. Additionally, two other artificial neural networks were created, trained on the Distress Analysis Interview Corpus with tremendous labeled video data regarding facial landmark tracking and audio analysis for depression-diagnosed people, to predict the risk of depression from a user-submitted video. All neural networks were evaluated on their accuracy, precision, recall, and f1-score. Besting state-of-the-art metrics on similar tasks in the relevant literature, iPonder achieves over 95% accuracy, recall, precision, and f1-score per neural network, pointing to a new age of accessible, effective, and anonymous mental health therapy in teens.**

***Keywords*—mental health, transfer learning, deep learning, natural language processing, computer vision, multimodal screening**

I. Introduction

Suicide is the second leading cause of death for teenagers, and more teenagers and young adults die from suicide than from cancer, heart disease, AIDS, birth defects, stroke, pneumonia, influenza, and chronic lung disease, combined. Even with suicide rates in teenagers being horrifyingly high, over 50 percent of suicidal teens don't ask for help or reach out in any way. Whereas depressed adults have the facilities to independently seek therapy, depressed teens not only feel an obligation to deal with these serious issues by themselves to avoid burdening those around them but also keep quiet due to their own shame. Furthermore, society is very poor at identifying those with mental health issues, evident in the fact that four out of five teenagers who have committed suicide have shown clear warning signs. To further stress the urgency of the issue, approximately 3,703 teen suicides happen every single day and 1.9 million teens in the United States have been diagnosed with depression. Still, these findings do not change the fact that depression and mental illness are indeed best addressed through conversational and intimate mental health therapy. The best way to face mental health issues is to confront them in a guided manner. This dilemma here between social stigma and effective treatment poses perhaps the biggest threat to the future of our youth, a threat exacerbated by the COVID-19 pandemic.

II. Background Research and Datasets

A. Transfer Learning and Multimodality

To determine how exactly artificial intelligence could be used to accomplish the task posed in the research question, the researcher set out to discover modern technologies that would improve the accuracies of deep learning models to a major extent. Here, transfer learning was significant: a method by which one model can train on a substantial dataset, draw connections, and pass those connections for a second model to use on different data. This is highly efficient as it eliminates the need for some models to start from scratch on complex data. Additionally, several research papers pointed to objective methods to predict disorder tendencies in a non-invasive manner such as the analysis of facial structure, eye gaze over time, social media/conversational data sentiment, tone quality, and pitch [1–5]. To extract all this data to create the most holistic and accurate means of diagnosis, users would need to submit video and text data.

B. Datasets and Questionnaires

1) Reddit Corpus and Twitter Data

To accurately predict depressive tendencies in teens via text-based methods, the most important tool to use is conversational data [6] [7]. Twitter is known as the number one text database in the world with a comprehensive corpus of cultures, words, feelings, and more. Twitter also hosts an unbelievable amount

ISBN: 978-0-7381-4533-4

of tweets that show depressive tendencies. The most accurate text-based depression data on the internet is available via the Reddit Self-Reported Depression Diagnosis Corpus (RSDD) and the Self-Reported Mental Health Diagnosis Corpus (SMHD). These two corpora contain over 20 gigabytes of data and over 1.5 million text components. This corpus contains an abundance of negative text data to predict tendencies of the following mental health issues: Addiction, Alcoholism, Anxiety, ADHD, Autism, Bipolar Disorder, Borderline Personality Disorder, Depression, Loneliness, PTSD, Schizophrenia, and Suicide Watch. Additionally, it also contains an abundance of positive or neutral text data of the following categories: Fitness, Finance, Teaching, Parenting, and Law. A release form was signed to obtain this data.

2) *Distress Analysis Interview Corpus*

The data described above deals with text data. However, video, audio, and survey response data are also necessary to make the diagnostic process multimodal for maximum user-specific data collection. This secondary data is present in the Distress Analysis Interview Corpus-Wizard of Oz Corpus which was created by the Institute of Creative Technology at the University of Southern California [8]. This database is a culmination of a series of interviews carried out at the university on patients with mental health issues. There are 189 40-minute recorded interviews in the corpus with 189 distinct patients, which were scanned for features such a tone, eye movement, posture, and facial expression, and asked to take the PHQ-9 - a reputable survey to predict depressive tendencies. It was about 75 gigabytes. The scanned features were in CSV files and pre-formatted. This data consists of a primarily important feature that uses a 68-point facial landmarking system that plotted points on the face of participants and tracked them over time. This time-series data was essential to the development of a computer vision algorithm that predicted depression tendency based on a face-cam video.

3) *Diagnostic Tests*

The PHQ-9 (Patient Health Questionnaire-9) is a famous behavioral science nine-question exam that evaluates patients for depressive tendencies. It is often the first thing a patient receives upon walking into a general physician checkup. The GAD-7 (General Anxiety Disorder-7) is analogous to the PHQ-9 as a seven-question exam to predict tendencies of anxiety. The UCLA Loneliness Test is a 20-question exam to predict tendencies of loneliness. Lastly, the Myers-Briggs Personality Type Indicator Test is a 50-question exam to classify the user into one of 16 distinct personality types. Each of these exams provides a Level 2 diagnostic measure that collects user-specific data [9]. The download for each of these tests is simple as many copies of them are available on the internet.

III. Hypothesis and Contribution

The methods proposed for this project are for both science and engineering combined. In this project, the answer to the following question is being sought: 1) How can artificial intelligence be used to accurately predict depressive tendencies in teens non-invasively and to provide private and effective mental health therapy for teens?, 2) The purpose of finding the answer to this research question has been extensively identified in the rationale for this project, and 3) Where mental health therapy and diagnosis are needed but so is the desire to avoid social stigma, artificial intelligence is the best resource, especially so as it can pick up aspects of behavior that humans can't quantitatively identify? Additionally, the contribution of this project is to the field of behavioral science, specifically mental health. Professionals and researchers agree that in-person therapy is the best means to confront mental health issues. However, they also recognize the social stigma of the matter and the tendency for teens to reject help in this manner [10,11]. Finding the answer to the research questions above would revolutionize the way for mental health issues of teens who are not ready to visit therapists. If the findings are significant, it prompts a new avenue for professional behavioral scientists.

The following testable hypothesis to answer the research questions was formulated. The researcher hypothesized that the use of text-and audio-based, so image-based deep learning neural networks predicts depressive tendencies in teens with accuracy, precision, and recall of over 95% and collects enough user-specific data to accurately and anonymously pair users online based on their diagnostic data in a way that satisfies the user.

IV. Methods

A. *Dataset Formatting*

1) *Twitter Data*

The first step to finding data to train the text-based models was web-scraping to collect tweets under the hashtag of depression through a negative text-sentiment test. Using Python functionalities, a program collects the text component 60,000 tweets under a hashtag (in this case "#depressed") and runs each through a sentiment analysis via the Python "sentiment analysis" module. If the tweet has a sentiment score of under 0.3 (a negative tweet), the program saved it into a CSV file and labeled it with the value 1. If the tweet did not meet this criteria, the program simply ignored it. The program kept scraping and processing tweets until there were exactly 60,000 in the file. However, for every deep learning supervised learning classification model, both positive and negative data are required. Thus, the process described above was also repeated for other unrelated positive hashtags. This time, the tweets were added to the same CSV file and labeled with a value of 0 if they had a sentiment score of above 0.2 (indicating neutral or positive). After this process, there was one CSV file with 120,000 data points where 60,000 of them indicate depressive tendencies and the rest do not. Lastly, a program was split the CSV file into two different ones: a test file and a train file. Each file had 30,000 positive and negative tweets.

2) *Reddit Data*

Once Reddit approved the request, the data was downloaded as a compressed file and pre-formatted. For each category listed above, the data was put into a single CSV file. For each negative/mental health issue category, the data were separated into two CSV files: a train file and a test file each with half the data as the original. Then, each created CSV file was populated with an equivalent amount of positive data as negative data by reading text from the positive CSV files. Once all was finished in this step, there were 24 CSV files each with an equal amount of positive and negative data points.

B. Four Level Diagnostic Process

1) Create the Rule-Based Programs for Diagnostic Levels 1, 2, and 3

For a quick recap, Level 1 involves the collection of user demographic data including age and sex. Level 2 involves the recording of responses to major behavioral science exams, and Level 3 involves the recording of answers to a series of multiple-choice questions used for specification of the patient issue. As each of these three levels involves a similar rule-based pattern to receive and record answers, all three levels are built using a similar programming mechanism. To store user data, a JSON file was created so that all specific aspects to the user including all responses to the first three levels were stored. For each level, one function or a group of functions in Python was created as one dictionary parameter. This dictionary was sent from the frontend and parsed to store relevant information in the User JavaScript Object. For the function for Level 1, it took in a dictionary that had keys of "age" and "sex" and stored them into the User JavaScript Object. Level 2 had 4 distinct functions (one per test) which each accepted a dictionary parameter. However, each test has its way of calculating the composite score via various point values per answer chosen. Thus, a dictionary was created mapping in each function that maps an answer to a test to a specific number of points. The parameter dictionary was simply a mapping of all the user answers. Therefore, the program parsed through the parameter and returned a specific point value for each function, which then was be stored in the main User JavaScript Object. Lastly, for Level 3, there was one function that stored all user responses to the multiple-choice section in the same way Level 2 did. Additionally, the free-response answers to Level 4 were also stored in the User JavaScript Object via the Level 3 functions but were manipulated in the following steps.

2) Train the Text-Based Artificial Neural Network via Transfer Learning

As done previously, the Twitter datasets were split into one train CSV file and one test CSV file. To create something to predict whether a new unseen inputted text shows depressive tendencies, an artificial neural network was created whose metrics confirmed or negated the hypothesis. An artificial neural network (ANN) is a feature of modern technology and artificial intelligence as it uses a series of nodes and connected layers to take in a series of inputs to predict a classification based on distributed node weights. ANNs are designed to mimic the way the human brain makes predictions. However, an even more modern approach to generate high accuracies via ANNs is known as transfer learning.

TABLE I. TRIAL 1: POSITIVE TEXT SENTIMENT DATA

	Accuracy	Precision	Recall
Anxiety	98.33%	98.12%	98.50%
Alcoholism	98.11%	97.90%	98.33%
Addiction	98.26%	98.14%	98.42%
ADHD	97.11%	97.01%	97.32%
Autism	96.40%	96.15%	96.59%
Bipolar	98.67%	98.50%	99.01%
BPD	97.94%	97.92%	98.23%
Depression	99.28%	98.99%	99.30%
Loneliness	98.68%	98.40%	98.73%
PTSD	99.64%	99.10%	99.57%
Suicide	99.10%	98.90%	99.29%
Landmarking	96.25%	96.12%	96.24%

Transfer learning is a technique that uses node weights from similar models that have extremely high accuracies and copies them to new ANNs. This allows new networks that are trained on different data to reap the benefits of previous accuracies. Firstly, TensorFlowHub presented the token-based text embedding ANN trained on over 4 billion tokens from the Google News 7B Corpus. This embedding was inserted as a layer in the new ANN trained via the Python Keras module [12]. To evaluate the model, the following metrics were used: Binary Accuracy (out of all predictions that involve either 1 or 0, how many were correct), Precision (out of all data points predicted positive), Recall (out of all data points positive), Root Mean Squared Error, Area Under Curve, Receiving Operator Characteristics, False Positives, False Negatives, True Negatives, and True Positives. The model was trained with the train data and a validation data size of 20% on 5 epochs.

V. RESULTS AND ANALYSIS

After taking all the trials across various sentiment types, the data confirmed that the accuracy, precision, and recall for every single model were above 95% thus confirming the hypothesis. Additionally, for each model, the recall values were higher, which proves that the models predict disorders solely based on the content of the inputs rather than extraneous features of the text.

In binary classification situations, this indicates that the models are strict. It means that if a person seems as they are on the fence of having or not having a disorder, the model weighs them towards having it. This is important as it ensures that no one with mental health issues is left out even if some users without mental issues are classified. For each model, the user ensures a high chance of being diagnosed correctly, which is a testament to the effectiveness of artificial intelligence. Once the substantially high model accuracies are confirmed, the researcher begins drawing relationships between what factors change the tendency probability. From the two graphs plotting

ISBN: 978-0-7381-4533-4

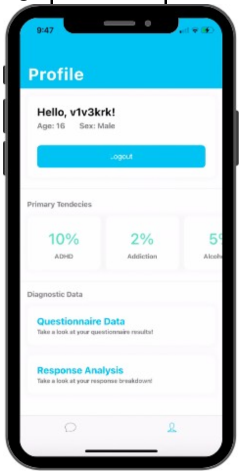
the percentage change of depressive tendencies versus an essential characteristic of the body of text, two important features are observed. In the way the model was trained, it negatively associated the positivity in the sentiment of the data with the chance of depression with an R-squared value of 0.74. This means that the models have intelligently figured out that more depressive texts showed more negative sentiment in their word choice. An additional characteristic of the model is that the length of the text has no effect on the probability of depressive tendencies with an R-squared value of 0.002. This is an important aspect to realize as it indicates that the models never associated text length with disorder chance. This ability is essential for private objective screening as it is not biased by irrelevant features of the text and predicts disorder tendencies independently of how much the user said [13].

TABLE II. TRIAL 2: NEGATIVE TEXT SENTIMENT DATA

	Accuracy	Precision	Recall
Anxiety	96.74%	96.54%	97.23%
Alcoholism	97.23%	97.10%	97.56%
Addiction	97.90%	97.75%	98.12%
ADHD	97.24%	97.13%	97.54%
Autism	96.90%	96.85%	97.05%
Bipolar	95.90%	95.50%	96.03%
BPD	97.80%	96.99%	97.89%
Depression	97.50%	97.12%	97.75%
Loneliness	98.68%	98.36%	98.82%
PTSD	97.90%	97.39%	98.19%
Suicide	98.76%	98.46%	99.02%
Landmarking	94.59%	94.57%	94.97%

TABLE III. TRIAL 3: NEUTRAL TEXT SENTIMENT DATA

	Accuracy	Precision	Recall
Anxiety	96.31%	96.11%	97.14%
Alcoholism	97.76%	97.42%	97.92%
Addiction	97.31%	97.12%	98.03%
ADHD	97.05%	96.86%	97.28%
Autism	96.17%	96.10%	97.05%
Bipolar	95.84%	95.77%	95.90%
BPD	97.36%	97.31%	97.51%
Depression	96.65%	96.34%	96.80%
Loneliness	98.03%	97.88%	98.43%
PTSD	96.47%	96.31%	96.82%
Suicide	98.43%	98.30%	98.44%
Landmarking	95.65%	95.41%	95.70%



Hello, v1v3krk!

Age: 16 Sex: Male

Take a look at your questionnaire results!

Take a look at your response breakdown!

Fig. 1. Home Page image of iPonder application.

VI. MOBILE APPLICATION CREATION

A. Create Application Programming Interface

Although all testing was finished, a finished product was created as the main goal was to distribute the technology to the public at some point. The finished product was a mobile application. In order to allow the mobile application to access all the Python code, an Application Programming Interface (API) via Flask was created. Flask is a Python web framework but is a popular tool that creates APIs. To create the API, only one new file was created: app.py. This file imported all the Python Flask modules and functionalities as well as the functions created previously. Additionally, it contained a series of API endpoints. API endpoints are functions that are accessed via a URL and passed data. Each function operated on a “POST” request (receives data from the frontend, processes it, and returns something back). There were 5 API endpoints: Level 1, Level 2, Level 3, Level 4, and Chat. Each endpoint received user information from the frontend as substantially detailed previously in the procedure and respond after computation.

B. Create Frontend Design

To ensure the highest amount of accessibility for users, the mobile application was created using the Flutter App Framework and the Dart Programming Language created by Google. Flutter allows sleek and modern-looking apps that demonstrate high levels of professionalism. To ensure a calm environment for the user, color psychology was researched: a branch of behavioral science. Based on numerous color psychology articles, a light blue and white theme application with modern and abstract images. Firstly, Level 1, Level 2, Level 3, and Level 4 of the mobile application were created, which are simply quiz-user interfaces. Flutter has many built-in functionalities for text boxes and multiple choice pages. Once the diagnostic portion of the frontend is complete, the chat page for anonymous pairing and chatting was created.

C. Connecting API to the Frontend

As of this step, the backend has been created and the frontend has been created, but they are not connected yet. Thus, the app does not have functionality. To connect this, Flutter’s Future object types were used as well as in-built fetch methods to call the API endpoints via POST requests and pass frontend data along with it. This process is fairly simple. With the connection of the API to the Frontend, the mobile application as well as the project is complete.

VII. DISCUSSION

A. Impact

The impact of the findings of the iPonder project is far-reaching. With modern technology in transfer learning, it reached accuracies with models that have never been seen before, which allows the integration of artificial intelligence into the field of behavioral science. Additionally, artificial intelligence seems to be the only viable and effective solution

ISBN: 978-0-7381-4533-4

to the problems discussed initially. Teens with mental health disorders have a way to deal with their problems early on while avoiding the social stigma of visiting a therapist or talking through it with anyone else. Like anything in life, one needs to be eased into the optimal solution. With faulty mental states, it is almost impossible to take initiative to organize a therapist, so private therapy helps ease one into the standard solution. Additionally, the cost of this project was under $10, which means there is a substantial market for those who cannot afford or spend on mental health therapy. Ultimately, with iPonder, the community strives to rectify the ongoing mental health crisis in the youth in order to ensure that they have the right mental state to change the world on their own. Mental health is the most important health.

B. *Future Steps*

As demonstrated, the mobile application portion of the project was an example of the potential of a personalized user object of data based on diagnostics. The user object storage model initiates an enormous amount of personalized mental health therapy initiatives. For example, in the future, the researcher plans to integrate the user-object model from diagnostics to integrate a physical fitness and nutrition portion to the mobile application for physical and mental health. Additionally, further increasing the accuracy of the models by integrating ensemble learning may be a fruitful endeavor. In this case, for storage optimization, rather than training 13 separate models, 3 were trained with ensemble learning to differentiate between tendencies of distinct disorders with higher predictive power.

VIII. Conclusions

In conclusion, the iPonder project has proved the hypothesis correct by proving that the use of text-, audio-, and image-based deep learning neural networks predicted mental health disorder tendencies in teens with accuracy, precision, and recall of over 95% and collected enough user-specific data to create a personalized therapy plan that satisfied the user. The model does not exhibit any real structural weak points after analyzing its ability to conclude the high association between sentiment and disorder tendencies and the non-association between text length and disorder tendencies. Additionally, high accuracies using modern technology such as transfer learning show how artificial intelligence plays an important role in the field of behavioral science and mental health. The use of computer vision, natural language processing, deep learning, transfer learning, API creation, and mobile application creation in one behavioral science project also demonstrates the high degree of interconnectedness of the two fields.

References

[1] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, and M. Breakspear, "Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Transactions on Affective Computing*, vol. 9, pp. 478–490, 10 2018.

[2] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," IEEE Xplore, p. 858–862, 12 2017. [Online]. ttps://ieeexplore.ieee.org/document/8389299

[3] S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx, and M. Breakspear, "Head pose and movement analysis as an indicator of depression," IEEE Xplore, p. 283–288, 09 2013. [Online]. Available: ttps://ieeexplore.ieee.org/abstract/document/6681444

[4] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, "Eye movement analysis for depression detection," IEEE Xplore, p. 4220–4224, 09 2013. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6738869

[5] A. P. Shah, V. Vaibhav, V. Sharma, M. Al Ismail, J. Girard, and L.-P. Morency, "Multimodal behavioral markers exploring suicidal intent in social media videos," *2019 International Conference on Multimodal Interaction*, 10 2019.

[6] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T. Padilla, F. Zhou, and F. D. l. Torre, "Detecting depression from facial actions and vocal prosody," IEEE Xplore, p. 1–7, 09 2009. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5349358

[7] Y. Yang, C. Fairbairn, and J. F. Cohn, "Detecting depression severity from vocal prosody," *IEEE Transactions on Affective Computing*, vol. 4, pp. 142–150, 04 2013.

[8] J. R. Williamson, E. Godoy, M. Cha, A. Schwarzentruber, P. Khorrami, Y. Gwon, H.-T. Kung, C. Dagli, and T. F. Quatieri, "Detecting depression using vocal, facial and semantic communication cues," *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge - AVEC '16*, 2016.

[9] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, p. 588–601, 03 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8580405

[10] S. G. Burdisso, M. Errecalde, and M. Montes-y G o´mez, "A text classi- fication framework for simple and effective early depression detection over social media streams," *Expert Systems with Applications*, vol. 133, pp. 182–197, 11 2019.

[11] C. Karmen, R. C. Hsiung, and T. Wetter, "Screening internet forum participants for depression symptoms by assembling and enhancing multiple nlp methods," *Computer Methods and Programs in Biomedicine*, vol. 120, p. 27–36, 06 2015. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260715000620

[12] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 12 2017.

[13] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health Information Science and Systems*, vol. 6, 08 2018.

ISBN: 978-0-7381-4533-4