

DEVELOPMENT OF NOVEL SCREENING APP TO ASSIST DOCTORS IN DIAGNOSING MAJOR DEPRESSIVE DISORDER AMONG INDIAN STUDENTS**Dr Anandkumar K M^{1*}, Ajith S², Bharathkumar J³, Hafeeluddeen V B T⁴**¹ Professor, Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India^{2, 3, 4} Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, IndiaEmail: ¹ anandkumar.m@eec.srmrmp.edu.in, ² ajith.s.2002@gmail.com, ³ bharathkumarj2002@gmail.com, ⁴ hafeeluddeen@gmail.com**Abstract**

Major Depressive Disorder is a serious mental health issue that can have a significant impact on a person's quality of life, relationships, and overall wellbeing. In response to this growing concern, SOLWOE is a novel app designed to diagnose and treat depression in teenagers. SOLWOE offers two types of care to its users: guided care and self-care. Guided care allows users to book appointments with psychologists and psychiatrists, who can provide tailored treatment plans to help them manage their depression symptoms. Self-care, on the other hand, provides basic self-healing activities such as yoga and meditation to improve mental wellness. By offering both types of care, SOLWOE aims to provide comprehensive support to people struggling with depression. The app uses a regression algorithm to analyse and categorize the depression level of users from the questionnaires they fill. In addition to that the app asks the users to write their journal entries in order to keep track of their mental health being and records their statements using BERT algorithm. Additionally there is a web cam feature opted in this project where the application records the images of the user for a specified time frame and allows the images to run in a CNN Model to determine the emotions of the user. This simplifies the job of psychologists and psychiatrists, who can use the app's data to provide better care to their patients. Overall, SOLWOE is a promising tool for diagnosing and treating depression in people. Its combination of guided and self-care, along with its use of regression algorithms and deep learning model, offers a holistic approach to mental health care. By providing early intervention and support, SOLWOE has the potential to improve the lives of millions of teenagers worldwide.

Keywords: *Major Depressive Disorder, Teenagers, Depression Detection, PHQ-9, Regression, Self-Care, Guided Care, Bidirectional Encoder Representations from Transformers, Daily Journals, Convolutional Neural Network.*

INTRODUCTION

Depression is a major global health issue that has significant consequences for individuals and society as a whole. According to the World Health Organization, depression is the leading cause of disability worldwide, with more than 264 million people affected globally. Depression can have a significant impact on a person's mental health and well-being, as well as their ability to function in daily life. For teenagers, depression can be particularly concerning, as it can impact their ability to learn, socialize, and develop relationships. Unfortunately, existing methods of diagnosing and treating depression in teenagers often involve lengthy wait times and high costs. These methods can be challenging for both patients and healthcare providers, as they require significant effort to diagnose and treat patients effectively. To address these challenges, SOLWOE is a novel app designed to diagnose and treat depression in teenagers. SOLWOE provides a user-friendly and accessible platform for people to receive support for their depression symptoms. The app offers both guided care and self-care options to users, providing them with the flexibility to choose the type of care that best suits their needs. Guided care enables users to book appointments with psychologists and psychiatrists, while self-care offers basic self-healing activities such as yoga to improve mental wellness. By using a regression algorithm to analyse and categorize the depression level of users from the questionnaires they fill and by enabling the users to write and record journals to analyse their mental health well being by employing deep learning algorithm. Additionally there is a web cam feature opted in this project where the application records the images of the user for a specified time frame and allows the images to run in a CNN

Model to determine the emotions of the user. SOLWOE simplifies the job of psychologists and psychiatrists, allowing them to provide more effective care to their patients.

Fig 1, illustrates the gender differences in suicide in percentage [4]. Suicide rate of male is 66.20% and that of female is 33.80%. Suicide by males is greater than that of females.

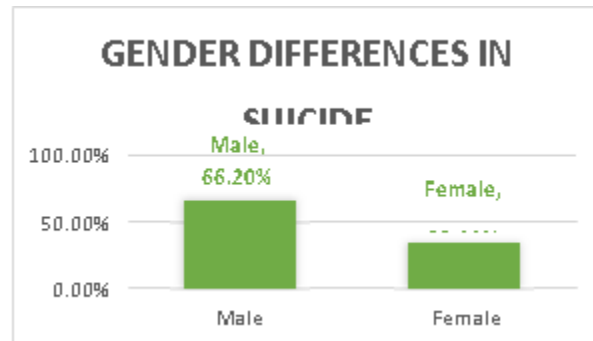


Fig. 1. Gender Differences in Suicide

Overall, SOLWOE represents a promising new approach to diagnosing and treating depression in teenagers. By providing early intervention and support, SOLWOE has the potential to improve the lives of millions of teenagers worldwide, and to reduce the global burden of depression.

This paper is divided into sections: section 2 for literature survey, section 3 for system model, section 4 describes the working of SOLWOE, section 5 about the results of the application and finally section 6 about the conclusion and future scope of SOLWOE

LITERATURE SURVEY

The papers referred for survey addresses the growing concern for mental health in today's fast-paced and electronically integrated world by discussing creative methods to monitor and recognize depression in individuals. Using commonplace digital gadgets, such smartwatches and smartphones with sensors, shows promise in detecting depressive symptoms early on. To identify such cases, a number of measures are associated, including digital screen time, travel habits, physical activity, sleep patterns, and levels of weariness. The study emphasizes the significance of quantifying positive and negative phrases by using the Patient Health Questionnaire-9 (PHQ-9) and therapeutic diaries with sentiment analysis to analyze emotional states.

The physical architecture of the system consists of a medical practitioner module that displays statistical data on a dashboard and a patient module that collects data and calculates diagnoses. The usage of user-provided text and a lack of integration of insights with clinical care are limitations that call for more study to be done in order to address ethical and privacy concerns as well as improve the system.

Machine learning algorithms are used to investigate suicide inclinations on social media sites such as Twitter. It is emphasized that a good classifier for textual emotion identification is logistic regression. Multimodal fusion is used to address identity deception on collaborative platforms. It combines verbal and nonverbal behavior analysis for precise identification.

These technologies can be used in high-stress settings like schools and workplaces where the COVID-19 epidemic is present. The study utilizes machine learning algorithms to analyze datasets and improve employee well-being, with a focus on feature visualization, preprocessing, and data preparation procedures. Using multimodal signals (text, audio, and video) for automatic depression diagnosis is a novel approach. Grammar and settings pertinent to the emotion domain can be learned with the help of task-oriented text embeddings and the Distress Analysis Interview Corpus dataset. The article emphasizes the useful applications that might be used in workplaces to identify and support employees who are experiencing mental health problems. All things considered, these many approaches

present a thorough and changing scene at the nexus of technology, mental health monitoring, and diagnostic instruments.

SYSTEM MODEL

SOLWOE is an innovative application designed to diagnose depression among people and provide them with effective self-care and guided care options. The app utilizes the PHQ 9 questionnaire a widely used and validated tool, and in addition to that encourages the users to write down their journals periodically to record and analyse their mental health well being, to predict the level of depression experienced by the user. Based on the questionnaire results and journal entries made by the user, additionally there is a web cam feature opted in this project where the application records the images of the user for a specified time frame and allows the images to run in a CNN Model to determine the emotions of the user, the app offers two types of care: self-care and guided care. Self-care is offered to users with mild or minimal depression and includes a variety of activities such as yoga, exercise, and mindfulness techniques to help them manage and overcome their depression. Guided care is offered to users with high levels of depression and enables them to book an appointment with a qualified psychiatrist or psychologist for personalized care and treatment. Fig 2. Illustrates the architectural model of SOLWOE.

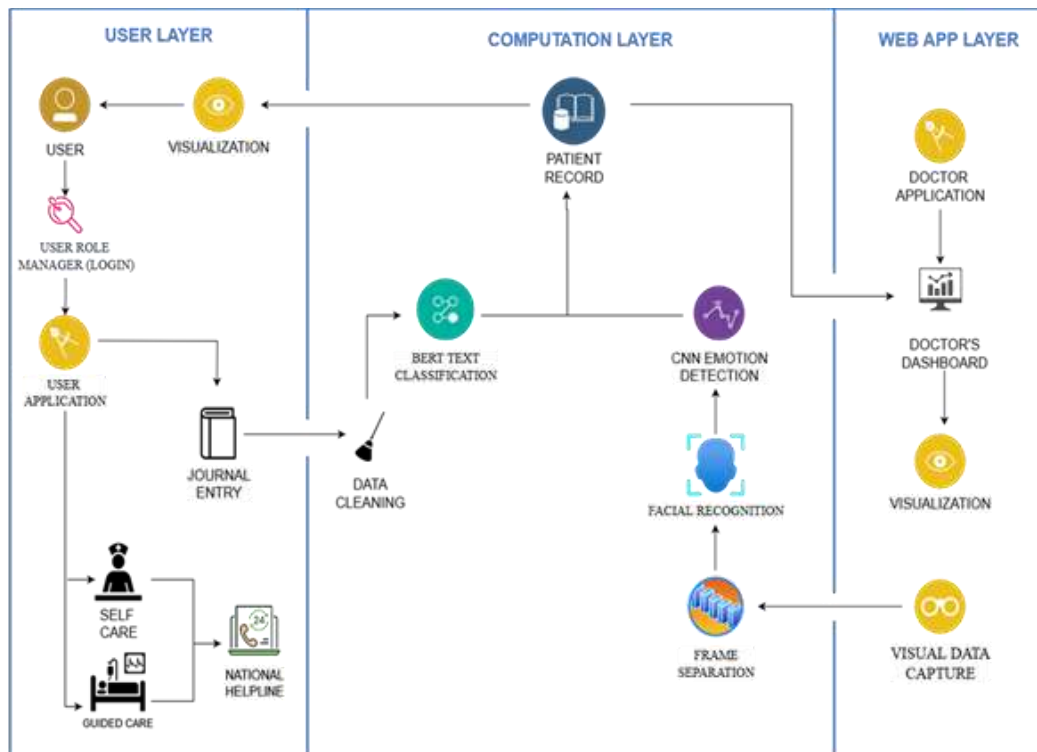


Fig. 2. Architectural Model of SOLWOE

SYSTEM DESIGN

SOLWOE is developed using Flutter, a powerful and popular open-source framework for building high-performance mobile applications. The app is also integrated with Google Firebase, a cloud-based platform that offers scalable and secure storage for user data. The use of Flutter and Firebase allows SOLWOE to deliver a seamless and responsive user experience while ensuring data security and reliability. The SOLWOE app uses the PHQ 9 dataset for training and testing its logistic regression model to diagnose depression in teenagers. This dataset contains 170,000 rows and 11 columns, with 9 columns representing the 9 questions in the PHQ 9 questionnaire, 1 column for the total score, and 1 column for the professional diagnosis. The dataset is used to train the regression model to predict the depression level of users based on their responses to the questionnaire. In

In addition to that PHQ 9 Questionnaire the journal entries of the user are categorized based on the BERT algorithm which uses a pre-trained dataset model which consists of more than 110,000 rows in which words are categorized into ‘positive’, ‘negative’ or ‘neutral’, Based on the wordings used by the user this model gauges the mental health well being of the user.

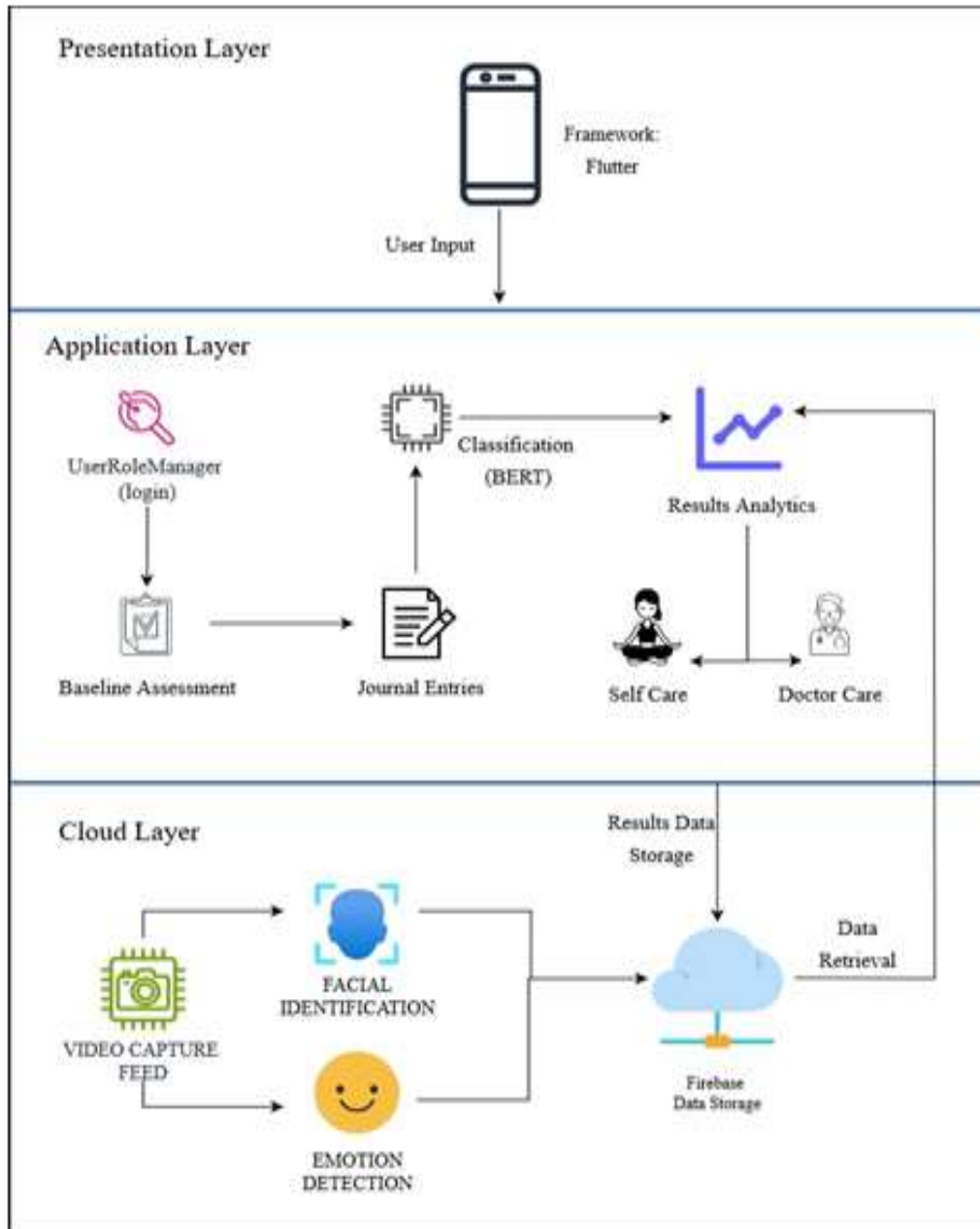


Fig. 3. Three-Tier Architecture of SOLWOE

Fig 3. represents the Three-Tier Architecture of SOLWOE, consisting of presentation, application, and data layers:

1. **Presentation Layer:** This layer is responsible for handling user interactions and displaying information to the user. In SOLWOE, the presentation layer is implemented using the Flutter framework for building cross-platform mobile applications.

2. **Application Layer:** This layer contains the business logic of the application and communicates with the presentation layer for user input and with the data layer for data storage and retrieval. The application layer in SOLWOE is responsible for implementing the PHQ 9 questionnaire, processing user responses, journal entries, and predicting depression levels using the logistic regression and deep learning model. Additionally, the webcam records the images of the user for a specified time frame and detects emotions of the user using CNN model. This layer also handles the logic for offering self-care and guided care features based on the predicted depression level.
3. **Data Layer:** This layer stores and retrieves data used by the application. In SOLWOE, the data layer is implemented using Google Firebase for secure data storage and retrieval.

Overall, the SOLWOE software architecture is designed to be modular, scalable, and easily maintainable. The three-tier architecture provides a clear separation of concerns between the presentation, application, and data layers, allowing for flexibility in development and ease of maintenance.

4.1 Diagnostic Module

The Diagnostic Module is responsible for assessing the user's symptoms and providing a diagnosis. The Assessment component of the Diagnostic Module uses a combination of user input, regression algorithm, records journals of the user using BERT model and records emotions of the user using CNN model, to generate a diagnosis. In diagnostic module, the functional architecture comprises of questionnaire and a database for read and write operations. The functional architecture of the diagnostic module can be seen in Fig 5., as a block diagram.

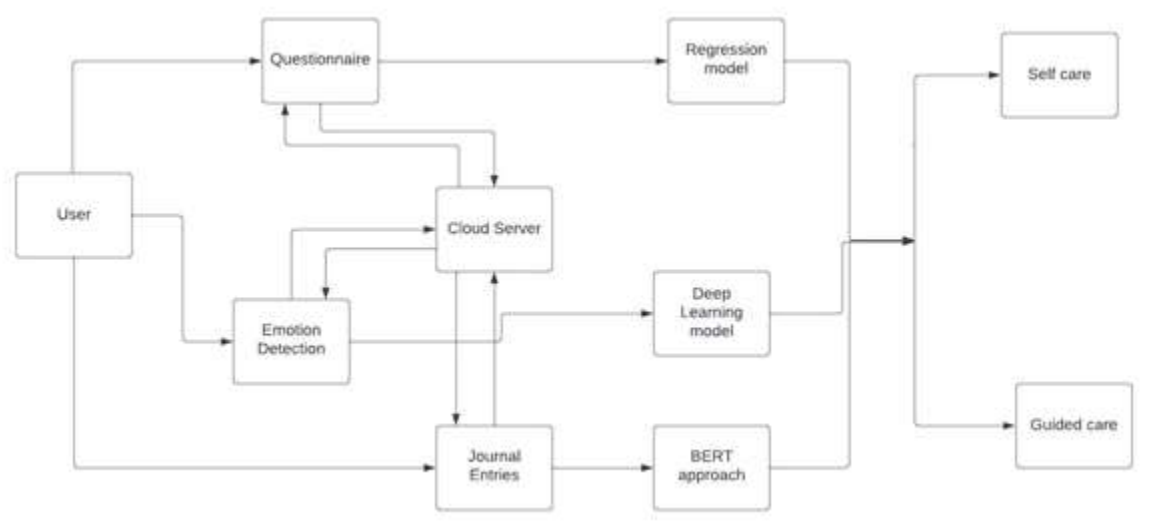


Fig. 4. Functional Architecture of Diagnostic Module

Fig 5., depicts the flow of the diagnostic module, where the user takes an assessment records assessment and records emotions of the user using webcam and the input from the above assessments are fed into the regression model, BERT model and CNN model respectively and a report is generated.

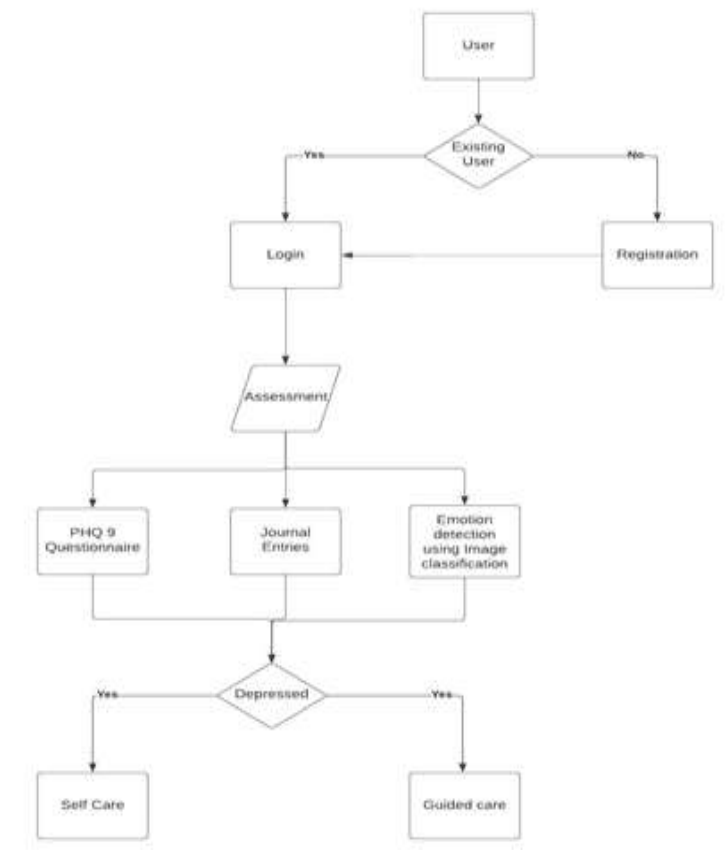


Fig. 5. Flowchart of the Diagnostic Module

4.2 Regression Model

a. Splitting dataset into train and test data

The PHQ-9 dataset has 170,000 rows, were 65 % of the data is used for training the model and 35 % of the data is used for testing/validating the model. Fig 6, shows the code for loading the data from a csv file and splitting the dataset into train and test data sets.

```

# Load the data from a CSV file
data = pd.read_csv("phq9.csv")

# Split the data into features (X) and labels (y)
X = data.iloc[:, :-2].values
y = data.iloc[:, -1].values

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.35, random_state=42)
  
```

Fig. 6. Splitting dataset into train and test data

b. Train, Test and Evaluate the Regression Model

The logistic regression model is initialized by assigning random values to the coefficients and trained using the training data to optimize the coefficients using gradient descent. The performance of the model is evaluated using

the testing data. The accuracy of the model was found to be 98%. Fig 7, shows the code for creating a logistic regression model, train the model with train data sets and test the model with test data sets. Finally, the model's performance is measure using parameters such as accuracy, precision and recall.

```
# Create a logistic regression model
model = LogisticRegression(penalty='l2',C=0.1,max_iter=1000)

# Train the model
model.fit(X_train, y_train)

# Test the model
y_pred = model.predict(X_test)

# Evaluate the model's performance
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred,average=None)
recall = recall_score(y_test, y_pred,average=None)

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
```

Fig. 7. Train, Test and Evaluate the Regression Model

c. Visualization of the Regression Model

The feature importance and confusion matrix are used for visualization, which helps to understand the significance of the independent variables and assess the model's performance.

```
# Visualize the logistic regression coefficients
plt.figure(figsize=(10, 5))
plt.bar(range(len(model.coef_[0])), model.coef_[0])
plt.xticks(range(len(model.coef_[0])), data.columns[:-2],
           rotation=45, ha='right')
plt.xlabel("Feature")
plt.ylabel("Coefficient")
plt.title("Logistic Regression Coefficients")
plt.show()

# Create confusion matrix
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, cmap='Blues',fmt='.0f')
plt.title('Confusion Matrix')
plt.xlabel('Predicted Labels')
plt.ylabel('True Labels')
plt.show()
```

Fig. 8. Code Snippet for Visualization of the Regression Model

Fig 8, shows the code for create a visual for the logistic regression feature importance by feeding the 9 answers from PHQ-9 as features. It also shows the code for creating a confusion matrix which in helps in identifying the true positives, true negatives, false positives and false negatives.

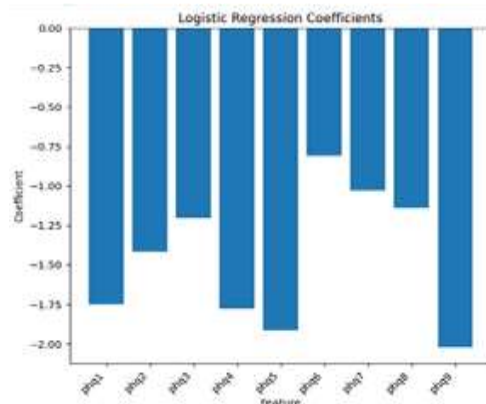


Fig. 9. Logistic Regression Coefficients

Fig 9, represents the feature importance of the logistic regression model. The feature importance in logistic regression can be determined by comparing the magnitude of the coefficients for each feature. Features with larger coefficients are considered to be more important in predicting the outcome variable than those with smaller coefficients.

A confusion matrix, Fig 10, is a table that summarizes the performance of the logistic regression model. The table is often used to evaluate the accuracy of a classifier by comparing predicted labels to actual labels. The predicted labels and actual labels are compared to determine the number of true positives, true negatives, false positives, and false negatives.

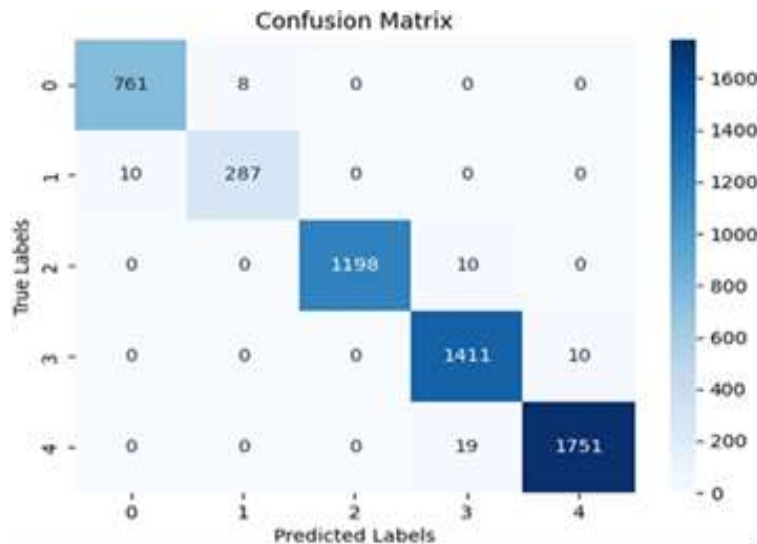


Fig. 10. Confusion Matrix

d. Logistic Regression Formula

The formula used by the ‘LogisticRegression’ class in scikit-learn is:

$$y = \frac{1}{(1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)})} \quad (1)$$

where:

- y is the predicted output
- b₀ is the intercept term (bias)
- b₁ to b_n are the coefficients (weights) for the input features x₁ to x_n
- e is the base of the natural logarithm (approximately 2.71828)
- x₁ to x_n are the input features (independent variables)

In this model, the number of features, that is, n ranges from 1 to 9 for each question of the PHQ 9.

To avoid overfitting of the model to the training data. We go for regularization of coefficients in logistic regression. Regularization helps to avoid this by adding a penalty term to the cost function that the model tries to minimize during training. This penalty term discourages the model from assigning too much importance to any one feature or coefficient, which can help to improve generalization performance on new data. The mathematical formula for logistic regression with L2 regularization is written as:

$$P(y = 1|x) = \frac{1}{(1+e^{-(b_0+b_1x_1+b_2x_2+\dots+b_nx_n)})} \quad (2)$$

where:

- $P(y=1|x)$ is the probability of the target variable (y) being 1 given the input features(x).
- The penalty parameter determines the type of regularization used. In this case, penalty=L2.

4.3 BERT model

This code installs the 'transformers' library and imports components for natural language processing. It utilizes the BERT-based 'bert-base-uncased' model for sequence classification. After tokenizing an example sentence, the code passes it through the model, applies softmax to obtain class probabilities, and prints the distribution. The code demonstrates basic usage of Hugging Face's Transformers library for text classification tasks.

```
!pip install transformers

from transformers import AutoTokenizer
from transformers import AutoModelForSequenceClassification
from scipy.special import softmax
```

This code utilizes the pre-trained sentiment analysis model 'cardiffnlp/twitter-roberta-base-sentiment' from Hugging Face's Transformers library. It first loads the corresponding tokenizer and then instantiates the sequence classification model. The model is specifically designed for sentiment analysis, allowing users to tokenize text and predict sentiment scores for various sentences.

```
MODEL = f"cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)
```

This function, named 'analysis', takes an input text and performs sentiment analysis using the previously defined tokenizer and model. It first encodes the input using the tokenizer, then feeds it to the model to obtain raw scores. The softmax function is applied to convert these scores into probabilities. The final result is a dictionary containing the sentiment scores categorized as negative, neutral, and positive. The function prints this dictionary, providing a user-friendly representation of sentiment probabilities for the given input text.

```
# Run MODEL
def analysis(input):
    encoded_text = tokenizer(input, return_tensors='pt')
    output = model(**encoded_text)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    scores_dict = {
        'NEGATIVE_SCORE' : scores[0],
        'NEUTRAL_SCORE' : scores[1],
        'POSITIVE_SCORE' : scores[2]
    }
    print(scores_dict)
```

The provided input text, "I don't know where I am going," is subjected to sentiment analysis using the 'analysis' function. This function tokenizes the input using the specified tokenizer and passes it through the pre-trained

sentiment analysis model. The resulting sentiment scores are then converted into probabilities using softmax. The function prints a dictionary with scores categorized as negative, neutral, and positive sentiments. This gives insight into the sentiment expressed in the input text, offering a breakdown of the likelihood of it being perceived as negative, neutral, or positive.

```
input = "I dont know where i am going"
analysis(input)

{'NEGATIVE_SCORE': 0.70395553, 'NEUTRAL_SCORE': 0.27261803, 'POSITIVE_SCORE': 0.023426468}
```

4.4 Treatment Module

The Treatment Module offers two types of care: Self-Care and Guided Care. Self-Care provides users with information and resources to manage their symptoms independently. Guided Care provides users with personalized care plans, including medication reminders, exercise routines, and follow-up appointments. In treatment module, the functional architecture comprises of self-care and guided care. The functional architecture of the treatment module can be seen in Fig 11., as a block diagram.

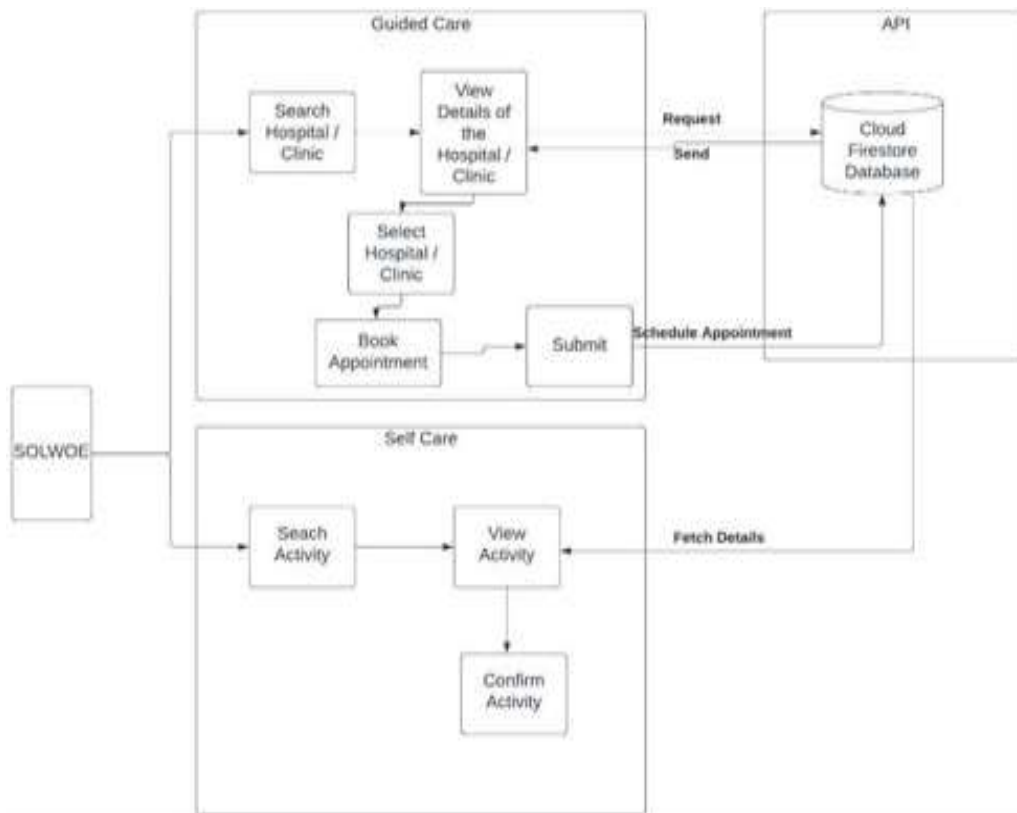


Fig. 11. Functional Architecture of Treatment Module

Fig 12, depicts the flow of self-care feature in the treatment module, where the user chooses any one of the activities listed and performs the activity. Fig 10, also depicts the flow of guided care feature in the treatment module, where the user chooses any one of the specialists listed and seek their help.

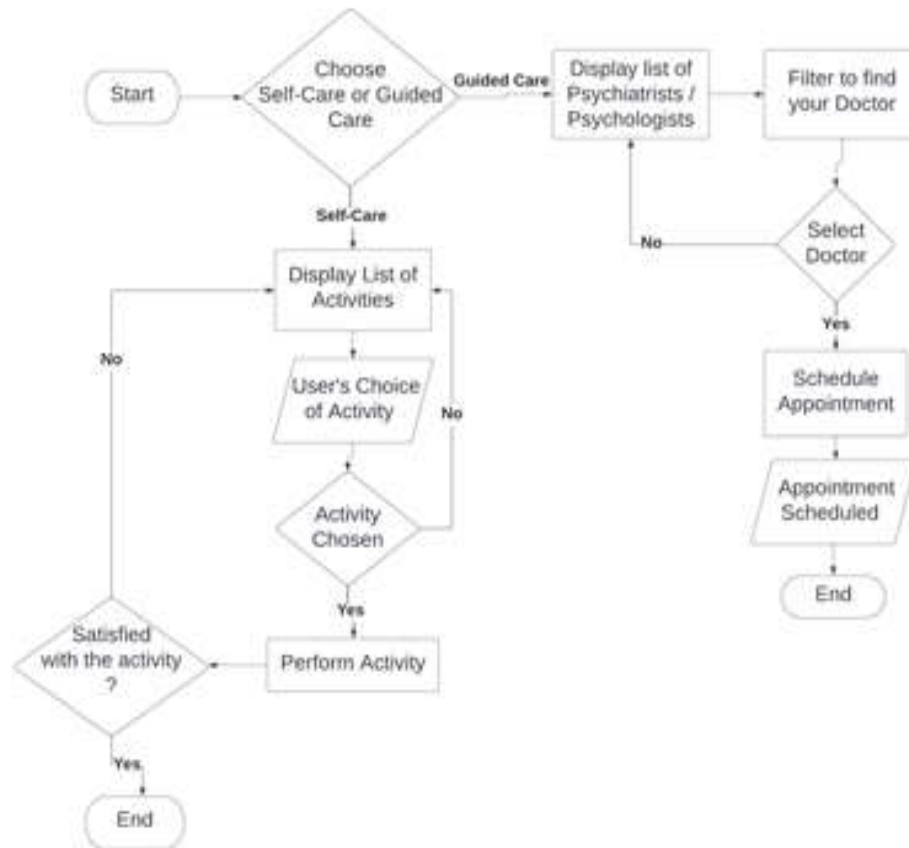


Fig. 12. Flowchart of the Treatment Module

RESULTS

The SOLWOE app is a promising new approach to diagnosing and treating depression in teenagers. The app can analyse and categorize the depression level of users, enabling healthcare providers to develop more tailored treatment plans that are specific to individual needs. The app's guided care option also provides users with access to professional care and support from psychologists, allowing users to book appointments with psychologists and receive personalized care and support.

In addition to its guided care option, SOLWOE offers a self-care option that can help users manage their condition more effectively. The app provides basic self-healing activities like yoga to improve mental wellness. This option can be used as a supplement to guided care or as a standalone option for users who prefer to manage their condition on their own. By providing users with access to self-care options, SOLWOE enables users to take an active role in their own treatment and recovery.

One of the key benefits of SOLWOE's self-care option is its low cost and accessibility. Many individuals may not have access to professional care or may not be able to afford it. By providing a self-care option, SOLWOE enables users to manage their condition effectively without incurring high costs.

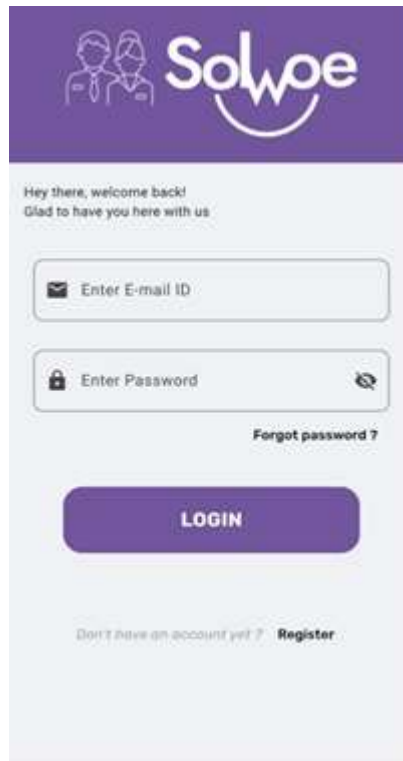


Fig. 13. Login Page for New/Existing Users

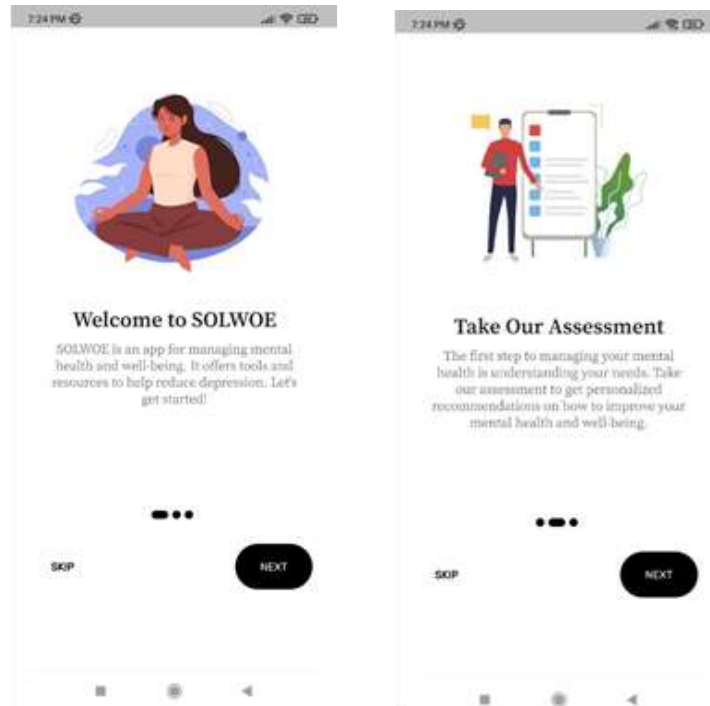


Fig. 14 (a), (b). Walkthrough of SOLWOE



Fig. 14 (c). Walkthrough of SOLWOE

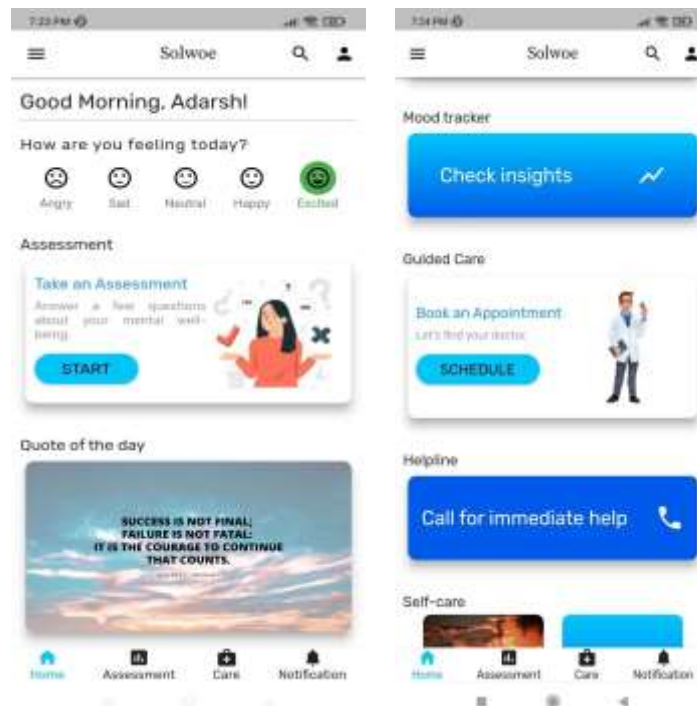


Fig. 15 (a), (b). Home Screen of SOLWOE

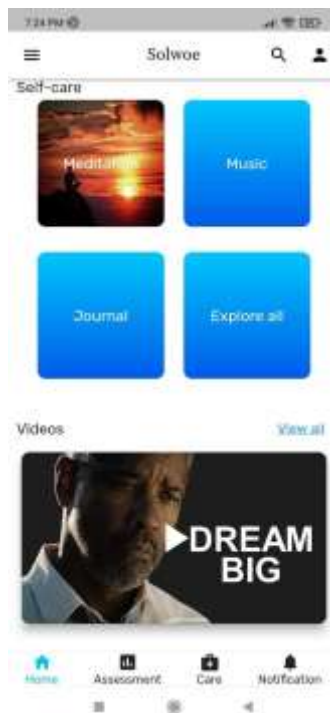


Fig. 15 (c). Home Screen of SOLWOE

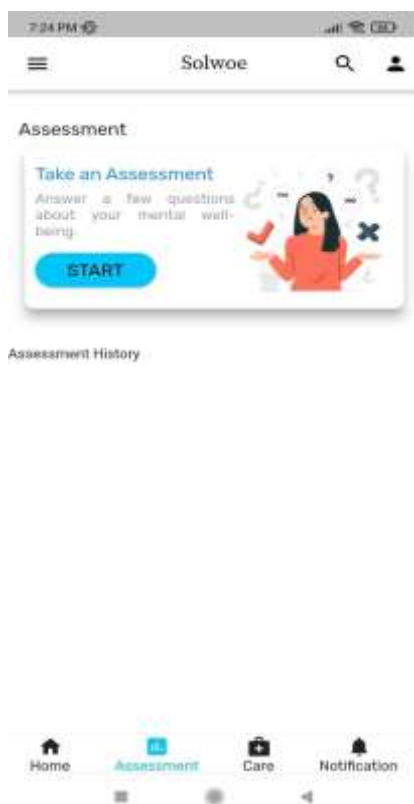


Fig. 16. Assessment Screen of SOLWOE

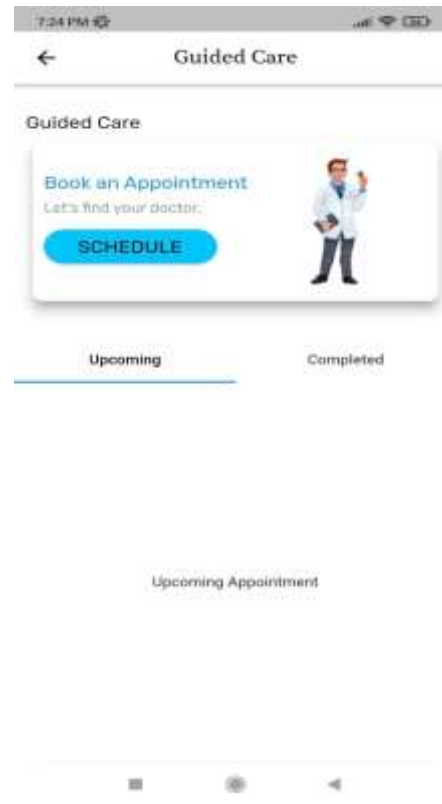


Fig. 17. Guided Care Screen of SOLWOE

CONCLUSION AND FUTURE SCOPE

In conclusion, SOLWOE has the potential to significantly improve the diagnosis and treatment of depression in teenagers. By providing early intervention and support, SOLWOE can improve the lives of millions of teenagers worldwide and reduce the global burden of depression. The app's guided care option provides users with access to professional care and support from psychologists, while the self-care option offers a low-cost and accessible alternative for users who prefer to manage their condition on their own. Overall, SOLWOE is a promising new tool for improving mental health and wellbeing in teenagers.

Moving forward, SOLWOE can be further enhanced with the integration of artificial intelligence and machine learning algorithms. By analyzing user behavior and patterns, these technologies can help predict depression levels with even greater accuracy, and identify early warning signs of depression. This would enable healthcare providers to intervene early and provide more effective treatment, potentially preventing the onset of severe depression. Additionally, SOLWOE can be expanded to include other mental health conditions, such as anxiety or bipolar disorder, providing a comprehensive platform for mental health care. Furthermore, the app can be translated into multiple languages, making it accessible to individuals worldwide, regardless of their language or location. Finally, incorporating user feedback and preferences can help improve the user experience and ensure that the app remains relevant and effective in addressing the evolving needs of its users.

REFERENCES

- [1] Amanat et al., “Deep Learning for Depression Detection from Textual Data,” *Electronics (Switzerland)*, vol. 11, no. 5, Mar. 2022, doi: 10.3390/electronics11050676.
- [2] K. A. L. Govindasamy and N. Palanichamy, “Depression detection using machine learning techniques on twitter data,” in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, May 2021, pp. 960–966. doi: 10.1109/ICICCS51141.2021.9432203.
- [3] Lekkas, R. J. Klein, and N. C. Jacobson, “Predicting acute suicidal ideation on Instagram using ensemble machine learning models,” *Internet Interventions*, vol. 25, Sep. 2021, doi: 10.1016/j.invent.2021.100424.
- [4] P. K. Swain, M. R. Tripathy, S. Priyadarshini, and S. K. Acharya, “Forecasting suicide rates in India: An empirical exposition,” *PLoS One*, vol. 16, no. 7 July, Jul. 2021, doi: 10.1371/journal.pone.0255342.
- [5] Priya, S. Garg, and N. P. Tigga, “Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms,” in *Procedia Computer Science*, 2020, vol. 167, pp. 1258–1267. doi: 10.1016/j.procs.2020.03.442.
- [6] R. Primartha, B. Adhi Tama, A. Arliansyah, and K. Januar Miraswan, “Decision tree combined with pso-based feature selection for sentiment analysis,” in *Journal of Physics: Conference Series*, Apr. 2019, vol. 1196, no. 1. doi: 10.1088/1742-6596/1196/1/012018.
- [7] R. S. Skaik and D. Inkpen, “Predicting Depression in Canada by Automatic Filling of Beck’s Depression Inventory Questionnaire,” *IEEE Access*, vol. 10, pp. 102033–102047, 2022, doi: 10.1109/ACCESS.2022.3208470.
- [8] S. G. Burdisso, M. Errecalde, and M. Montes-y-Gómez, “A text classification framework for simple and effective early depression detection over social media streams,” *Expert Syst Appl*, vol. 133, pp. 182–197, Nov. 2019, doi: 10.1016/j.eswa.2019.05.023.
- [9] Sau and I. Bhakta, “Screening of anxiety and depression among the seafarers using machine learning technology,” *Inform Med Unlocked*, vol. 16, Jan. 2019, doi: 10.1016/j.imu.2018.12.004.
- [10] Shanghai Dianji University. School of Electronics & Information, M. IEEE Systems, and Institute of Electrical and Electronics Engineers, 2016 3rd International Conference on Systems and Informatics: 19-21 November 2016: Shanghai Jingyue International Aviation Hotel, A big data application to predict depression in the university based on the reading habits, Shanghai, China: ICSAI 2016.
- [11] Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella, and S. V. Alluri, “Real-time Acoustic based Depression Detection using Machine Learning Techniques,” in *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, Feb. 2020. doi: 10.1109/ic-ETITE47903.2020.394.