Well-being Analysis Using Machine Learning

Dr. N. Mangathayaru
Professor, Department of
IT VNR VJIET
Hyderabad, India

Dr.B.Venkata SeshuKumari
Associate Professor
Department of IT
VNR VJIET
Hyderabad, India

M.Susmitha
Assistant Professor
Department of IT
VNR VJIET
Hyderabad, India

Abstract—Psychological well-being is one of the main contributors to an individual's prosperity, irrespective of age and education. Tragically, these issues have been increasing step by step, most commonly in the case of secondary school and understudies. There is a great deal of need for successful methods to hold this issue within proper limits, or else they may lead to many significant issues such as depression and suicides. To keep the problem in check, it is necessary that the confusion is detected at the beginning phase and the individual is treated appropriately. This paper centers around the psychological well-being issues that are usually looked at by understudies between the ages of 15 to 27 and pursuing their schooling between grade 10 and post-graduate studies. The information has been gathered straightforwardly from the understudies through a survey, and machine learning procedures have been utilized to pre-cycle and concentrate the useful data present. By utilizing the machine learning algorithms, the level of mental prosperity has been calculated from the review information, which gives an early identification of issues being looked at by the understudy. This model further provides a few precise ways that the individual can rehearse to defeat the problems being faced.

Keywords—Mental Health, Students, Education, Depression, Anxiety, Suicide

I. INTRODUCTION

Psychological wellness is a vital component that influences one's feelings, reasoning, and ability to think. Some of the most serious mental health conditions don't just appear overnight. Issues identified in this field are caused by numerous variables, which differ from one individual to another, considering the situations faced, and include past medical problems. Extra care and treatment can be given if aberrant mental states are identified early in the course of the illness. While threatening or upsetting stimuli can cause anxiety, depression, social dysfunction, and even suicidal thoughts, challenging stimuli can have positive effects like motivation and enhanced task performance.

Mental illnesses, which are also called mental health disorders, are a broad category of mental health conditions that affect your emotions, thoughts, and behaviour. They include psychological conditions such as eating disorders, compulsive behaviours, schizophrenia, depression, and anxiety disorders.

A mental illness can ruin your life and cause problems for routine tasks like relationships, job, and education. Most of the time, talk therapy along with medication helps manage symptoms. (psychoanalysis).

If you notice any signs or symptoms of a mental illness, speak with your healthcare provider or a mental health professional. Most mental illnesses don't go better on their own, and if they aren't treated, they can worsen over time and result in serious problems.

Many people occasionally experience problems with their mental health. However, a mental health issue turns into a mental disease when persistent symptoms put you under a lot of stress and impair your capacity to perform daily tasks.

For the students that are centred on this examination, these factors incorporate companion pressure, family issues, instructive and professional settlement, etc. Such issues hinder the mental strength of individuals early on, which prompts various chronic problems down the road. If the issues are diagnosed at the beginning phase, the likelihood of fixing them is higher. This paper focuses on the causes behind issues looked at by the students and provides a potential answer for them.

This paper focuses on the use of supervised machine learning algorithms to find a group of students who are in danger of suffering from mental health issues. These algorithms focus on building the model using a target or a class variable. This method allows the model to learn over time and provide the desired outputs for the given inputs. Some of the algorithms used in this project include Support Vector Machines, Decision Trees, and Logistic Regression.

Finding mental health through surveys can be a beneficial tool for individuals, researchers, and experts alike. To gain insights into a human being's general well-being and pinpoint potential areas of concern, surveys are made to gather information about people's opinions, emotions, and behaviours relating to mental health. The raw data that has been used for the purpose of training the desired model has been collected through a survey, by interacting with students of different age groups and pursuing different educational backgrounds. Questions were asked in a generalized manner, by taking mental health triggers into consideration.

II. LITERATURE SURVEY

2.1 Relevant Works: Health is described by the World Health Organization (WHO) as "a state of full mental, social, and physical health and not just being free of illness or impairment"[1]. The commonly utilized WHO-5 Happiness Index evaluates five elements of subjective well-being: good mood, vigor, general passions, confidence, and ability to deal with everyday challenges as given by] Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015) [2]. Studies show that mental health differs among various populations. According to the research given by Baxter A.J. et al. (2014) [3], the statistics given across 44 countries give major factors such as age, gender, economic status, and urban conditions responsible for the variance of mental health. The prevalence of anxiety disorders ranges from approximately 5% in African countries to 10% in European countries. According to reports given by Kessler R.C. et al. (2005) [4], the annual prevalence of anxiety disorders in the USA was 18%. These investigations illustrate the serious consequences that mental health illnesses carry as well as the requirement for efficient preventative and treatment methods.

The grouping of mental health conditions evolved over time, resulting in different classification schemes in existence. For instance, while the International Classification of Diseases (ICD) is used globally, the Diagnostic and Statistical Manual of Mental Disorders (DSM) is frequently used in the US. The reliability and validity of these criteria are, however, still up for debate by Insel T.R. and Cuthbert B.N. (2013) [5][6]. This underlines the necessity of continued research to enhance diagnostic precision and guarantee that patients receive the right care.

It has been determined that a variety of biological, psychological, and environmental factors have a role in the emergence of mental health illnesses. The development of posttraumatic stress disorder (PTSD), for instance, may be influenced by genetic, epigenetic, and neurobiological variables, according to Koenen K.C. et al. (2017) [7]. On the other hand, factors that positively affect well-being include income, education, and support, through the research given by Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L et al. (1999) [8]. Boehm, J. K., & Kubzansky, L. D. et al. (2012) [9] however state psychological resilience, optimism, and social connections as the major contributors. Demyttenaere et al. (2014) showed that lifetime prevalence rates for mood and anxiety disorders were 12.8% and 14.6%, respectively, throughout Europe [32]. A study by Prinz et al. (2018) [33] found that 43% of computer science students exhibited symptoms of depression, anxiety, and stress in the setting of technology and engineering.

Two significant obstacles to obtaining care for mental health disorders are stereotypes and lack of access to professionals in the field. Only 11% of college students with mental health issues sought professional help, according to a study by Eisenberg et al. (2010) [53]. Similar findings were made by Cheung et al. (2016) [54], who revealed that stigma greatly discouraged engineering students from seeking aid. According to Prinz et al. (2018) [33] , students in the fields of technology and engineering were unwilling to seek assistance due to worries about stigma and lack of anonymity.

Numerous interventions have been created to support the general public's mental health and well-being. In a study by Cuijpers et al. (2009) [34], it was discovered that cognitive behavioral therapy (CBT) delivered online was successful in easing the symptoms of anxiety and depression. In a similar vein, a study by Richards et al. (2016) [35] discovered that a mobile app-based intervention was successful in lowering university students' symptoms of anxiety and sadness. A study by Munn et al. (2018) [36] in the field of technology and engineering discovered that a peer support program was successful in enhancing mental health among students studying computer science.

2.2 Analysis of Existing System: There are many existing methodologies that have contributed to the analysis of mental health. Jet Li Chung and Jason Teo et al. (2021) [10] and Anu Priya, Shruthi Garg, and Neha Prerna Tigga et al. (2020) [11] recognize several Machine Learning techniques including logistic regression, decision trees, random forests, stacking, naïve Bayes, support vector machine, and K nearest neighbor classifiers. They utilized every one of the seventy-seven characteristics from the text archives to run these calculations. However, Anu Priya, Shruti Garg, and Neha Prerna Tigga [11] classified the best-model selection based on the f1 score due to the imbalance in the data. Though the accuracy of Naive Bayes was found to be the highest, the Random Forest technique proved to be the best model.

In the analysis given by Dr.J.Arokia Renjit, Adlin Sajeesha M.J, Sangavai V.D, Sree Devi D.S et al. (2022) [12], the dataset was collected, pre-handled, and mapped into graphs with the end goal of information quality. Then, various algorithms were utilized to foresee the result with high precision. The primary parts of this paper are data graphing and designing the classification model. The information purification, lost values, examining translation, model structure, and assessment were the most important phases in the analysis cycle.

Roger Garriga, Javier Más, Semhara braha, Jonnolan, Oliver Harrison, George Tadros, and Aleks Andar Matic et al. (2022) [13] explored the possibility to forecast any mental health crisis event and if such predictions can benefit medical care. They used machine learning approaches to analyze longitudinally collected EHR data. The result showed the possibility of forecasting mental health crises and obtained an AUROC of 0.797 for the general model.

Sofianita Mutalib, et. al. (2021 [14] focused on mental health problems in Malaysian students. Twenty Public Universities in Malaysia had 552,702 students enrolled as of the end of December 2018, and 119,345 of them had graduated, according to statistics. The Certificate, Diploma, bachelor's degree, master's degree, and Doctor of Philosophy are the five stages of higher education qualification. They begin by describing the mental health disorders and causes common among college students and classify the difficulties into three groups: stress, depression, and anxiety. Using the characteristics' score to label the individuals in the dataset, they demonstrated how the DASS-21 data may be used for modeling. Support Vector Machine for depression and Decision Tree for stress are the best models with the highest accuracy. The two models that provide precise results for anxiety with an accuracy range between 68% and 88% are linear regression and neural networks.

In the context of predicting and diagnosing mental health diseases, machine learning has shown promise in the field of mental health analysis. Machine learning was utilized in a 2013 study by De Choudhury et al. [15] check to predict depression using social media data, with a 70% accuracy rate. In a similar vein, Chekroud et al. (2016) [16] achieved an accuracy rate of 73% when using machine learning to predict the effectiveness of antidepressant drugs. In their 2014 study, Coppersmith [52] examined how social media data could potentially utilised for predicting depression. The study showed the promise of machine learning for early detection and prevention of depression by analyzing social media data from depressed participants and healthy controls.

III. DRAWBACKS OF EXISTING SYSTEM

- Machine Learning algorithms can investigate the information given, and don't have a similar degree of knowledge and setting as a psychologist. Subsequently, they might miss significant elements that could affect efficiency.

- Moreover, Machine Learning Algorithms are simply ready to examine the information that is accessible to them, which may not catch the full scope of elements that affect prosperity. This could bring about inadequate or incorrect outcomes.

- Numerous systems receive training using skewed data, which can lead to incorrect predictions. For instance, if a dataset contains a large proportion of white people, the algorithm might not work as effectively with data from other racial and ethnic groups.

- Some systems have a limited ability to forecast other mental health illnesses because they are primarily intended to predict certain ones, like depression or anxiety.

- Many current systems rely on population-level data rather than patient-specific information to generate predictions. For people who do not fit the normal profile of the population being analyzed, this may lead to erroneous forecasts.

- Instead of using data specific to each patient, many current systems produce predictions using data collected at the population level. Forecasts may be inaccurate for individuals who do not fit the typical profile of the population being studied.

- Concerns exist around the collection, storage, and usage of sensitive mental health data. As a result, people may be less willing to contribute their data, which could further lower prediction system accuracy.

IV. PROPOSED SYSTEM

From our study of mental well-being and the existing systems that have been made through this foundation, it has been observed that mental health is an essential aspect of our daily lives, which is often neglected by people, especially that of students in the ages of 15 to 27. In a world with a lot of technological advancement, we aim to provide an analysis of the mental health of students and identify the factors which cause it among students from study groups such as 10th grade, intermediate, Undergraduate, and Postgraduate, where the tendency to get mentally affected is more. Using supervised machine learning techniques, one can predict whether they are on the edge of being mentally affected and can contact a therapist for the same.

4.1 Objective of the proposed system:

1. To determine if a student reflects the symptoms of being affected by mental health problems.
2. To identify the main elements influencing students' mental health through analysis.
3. To bring about awareness among students regarding mental health and its care.

4.2 Problem Definition: Mental health is a commonly overlooked factor, especially with the busy lives that today's youth lead. This can lead to bigger consequences in the future if not taken care of. To find a solution to this problem, we made a quiz, which is connected to the trained machine-learning model at the back end. The quiz contains general questions about the respondents' daily habits and the intensity of the habit. Through the responses given, it is determined whether the student reflects the symptoms of displaying mental health problems or not and connect them to a therapist if they tend to show more vital signs.

4.3 System Design: The workflow of the system can be illustrated through the following diagram:

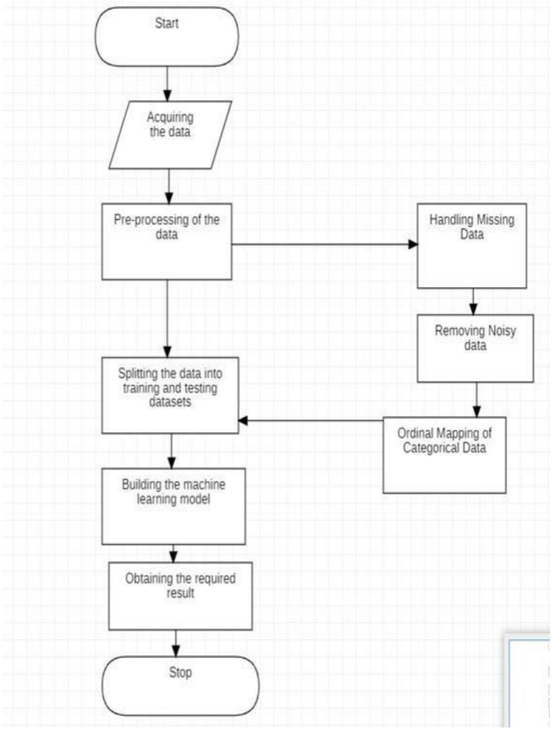


Fig 1. System Design

1. Dataset Collection: The data for training the mental health prediction model has been collected in the form of a survey, with the responses coming directly from the students. It consists of 20 attributes in text format, with the class labels checked. The feature names include Age Group, Gender, Education Pursuing, Family Size, Physical Health Issues, Mental Health Problems in the family, Previous Mental Health Problems, Sleeping Trouble, Anticipatory Anxiety, Relaxation issues, Suicidal thoughts, Factors that externally affect their mental health; Friends, Family, Education, Health, Strangers, Career, Lover, have they tried to overcome the problem.

2. Data Pre-Processing: The data collected from the respondents is raw and contains many ambiguities and inconsistencies. To prevent these factors from coming in the way of training an accurate model, the following steps have been followed to get proper data that the learning model can understand.

i. Missing Values: The collected data contains many missing values that are a hinderance to training the model and can result in various errors occurring. To avoid this, the dataset is checked for missing values, and are replaced using forward fill method.

ii. Feature Selection: There are 20 different features that have been collected from the students. Some of these features may not be useful for training the model and can affect the accuracies which we obtain in the output. So, only a few features have been selected using Pearson Coefficient.

iii. Noisy Data: In raw data, there might be some data that may not correspond to the feature, and thus will not contain any useful insights during training. Such data needs to be either replaced or removed. In this scenario, the data has been replaced using data smoothing.

3. Encoding Data: In this step, label encoding is done on the now cleaned data. This step is necessary as the data that is being dealt with is categorical data, and they need to be assigned a numerical label as the training model can understand only numerical values. Moreover, label encoding helps in identification of the categorical values.

4. Model Building: After the dataset is ready, it is split into training and testing data, and cross-validation is performed. Here, 70% of the data is taken as training data and 30% of the data has been taken as testing data. The six different models that were taken into consideration are Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest Classifier, and a Hybrid model of SVM and Decision Tree.

5. Model Training: The training data is now fit into the various supervised learning algorithms and the outputs are predicted. The accuracy metrics are derived based on the outcome of the predictions.

6. Mental Health Prediction: Now, to test the accuracy of the model, sample data is given and fed to the system. The result is given in the form of a message, which tells if the person is suffering or not.

7. Model Pickling: With the models trained and the outcomes predicted with the accuracy metrics, the models are now pickled as a .pkl file to incorporate them into a web interface, which works based on the learned data. The web interface is linked to the pickle file by using the Flask python library.

8. User Interface: The quiz is connected to a web interface for simplicity of use, with the use of little hardware and software requirements.

V. DATASET USED

The data used for training the machine learning model has been collected as a survey, which consists of 5000 different entries from students of different age groups and educational backgrounds.

The dataset consists of 17 different features, which include physical health issues, sleeping troubles, anxiety problems, suicidal thoughts, and relaxing problems.

Another aspect of the dataset is how different external factors such as family, friends, education, health, and so on tend to affect the person. Together, these different features have been used to determine if the student is at the risk of suffering from mental health issues.

From Fig.1 we can see a few entries of the data that has been collected from students for the purpose of this paper. Fig. 2 shows some of the external factors that students report, affecting them mentally.

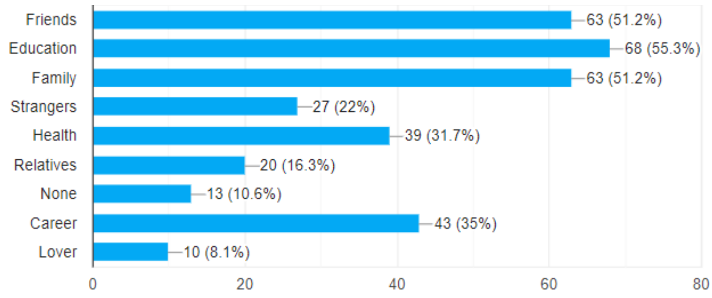


Fig. 2 Factors that affect mental health.

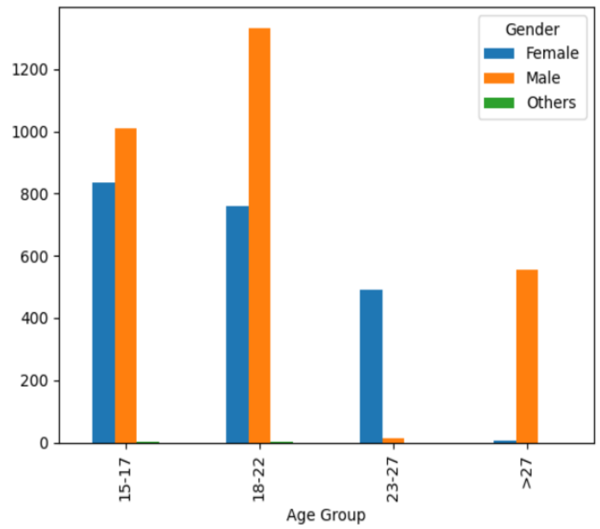


Fig. 3 Age Groups of Students surveyed.

Fig. 4 Shows that many of the respondents to the survey are between the ages of 18 to 22. To train the model in a better fashion, more responses have been collected from age groups of 23-27 and people of age greater than 27.

VI. METHODOLOGY

6.1 Algorithms used

i. **Support Vector Machine:** Support-vector machines, sometimes referred to as support-vector networks in machine learning, are supervised learning models that examine data for regression and classification along with corresponding learning methods. In an effort to increase accuracy, the subsequent actions have been taken:

Step 1: Split the data into training and testing parts.

Step 2: Choose the type of kernel to be used. In this scenario, a polynomial kernel has been used.

Step 3: Set the regularization parameter (C) to 1, to set the strength of regularization to a considerable amount.

ii. **Decision Tree:** This is a commonly utilized supervised machine learning method employed in data mining. A decision tree is a visual representation that people use to depict statistical probabilities or to map out the order of events, actions, or results. The following steps are implemented to build a decision

Step-1: The complete dataset has been taken as the root node of the decision tree.

Step-2: Using the attribute selection measure (ASM), the best attribute combinations to give accurate results have been found.

Step-3: The root node was divided into sub-trees that are more probable to give the best combination of attributes.

Step-4: The node that gave the better combination was taken into consideration.

Step-5: Repeatedly new sub-trees were created by following steps 1 to 3 till no further classification can be made.

iii. **Logistic regression:** For binary classification issues, where the objective is to predict one of two possible outcomes, logistic regression is a statistical machine learning approach that is utilized. In many different applications, including fraud detection, picture classification, and medical diagnosis. The procedure followed to implement logistic Regression for the system is as follows:

Step 1: Feature scaling was done using standard Scaler on the data to put all features in the same range, regardless of their relevance.
Step 2: To handle class imbalance, normalization is done on the data using class weights.

Step 3: Finally, the model is trained to predict the correct outputs.

iv. **K-Nearest Neighbors:** It is a non-parametric machine learning algorithm used for classification and regression problems. It is a form of instance-based learning where the algorithm predicts new data points based on the training data itself. For this project, KNN has been implemented in the following way:
Step 1: The data has been split into training and testing data.

Step 2: The nearest data points have been determined by choosing the K value as 3.

Step 3: Implement the following steps for all the data points:
• Calculate the distance between the testing data point and each of the training data points with the help of distance measuring metrics. Here, Manhattan distance has been used.
• Sort the distances calculated in increasing order, and choose the top K rows accordingly.
• Class is assigned according to the highest frequency.

v. **Random Forest Classifier:** This ensemble learning method aids in solving regression and classification issues. In this instance, it is an extremely helpful algorithm for managing big datasets. To create a more accurate and dependable model, it combines different decision trees. To apply the random forest classifier, these steps have been taken:

Step 1: From a given training set or data, choose random samples.

Step 2: A decision tree will be built for every data sample.

Step 3: The decision tree that is obtained will be averaged to conduct voting.

Step 4: The outcome that receives the most votes is chosen as the winning prediction.

vi. **Hybrid Classifier:** A hybrid classifier is created when the predictions of two different learning models are taken into consideration. We have implemented the hybrid classifier using the following steps:

Step 1: A model of the Support Vector Machine is built, using a kernel of choice.

Step 2: The decision tree model is created by selecting the attribute combination that gives the best results.

Step 3: The prediction values of both models were analyzed, and those data points that gave the same outputs were taken into consideration.

Step 4: Now, sample test data is taken to check the accuracy of the newly built algorithm.

6.2 Libraries used:

i. **NumPy:** Popular Python module NumPy is used for numerical computation in machine learning. It offers effective data structures for displaying arrays and matrices as well as a large selection of mathematical procedures for using these arrays. Because they enable effective and quick computations on huge datasets, NumPy arrays are particularly helpful in machine learning. Additionally, they offer a uniform user interface for handling data, regardless of its origin or format.

ii. **Pandas:** Pandas is a tool for data preparation, analysis, and manipulation in machine learning. It supports a large variety of data formats and activities and offers a high-level interface for working with tabular data in databases and spreadsheets. The Data frame, which is simply a two-dimensional table with rows and columns, is the primary data structure used by Pandas. Several data types, including CSV and Excel files, SQL databases, and other data formats, can be used to construct Data Frames. Pandas offers several functions for cleaning, altering, and manipulating the data once it has been loaded, including handling missing data, merging, and connecting data from different sources, and selecting subsets of data.

iii. **Matplotlib:** A well-liked Python package for data visualization in machine learning is called Matplotlib. It offers a wide number of tools for building several kinds of plots and graphs, including histograms, line plots, scatter plots, and bar charts, among others. With Matplotlib, you can easily change the colors, labels, and axes scales of plots to suit your preferences. It may also be used to visualize data from many sources, including NumPy arrays, Pandas Data Frames, and SQL databases, and it supports a broad variety of data formats.

iv. **Sci-Kit Learn:** It is used for data analysis, modeling, and prediction. Sci-Kit Learn offers a variety of algorithms and functions for classification, regression, clustering, and dimensionality reduction, among other machine-learning applications. Scikit-learn is a crucial tool for data scientists and machine learning practitioners because it offers a user-friendly interface for data preparation, feature engineering, and model validation. Additionally, it has many tools for transforming data, such as methods for handling missing data and tools for normalizing, scaling, and encoding data.

v. **Pickle:** In machine learning, trained models are frequently saved to files so they may be loaded later and applied to new data to create predictions. Pickle makes it simple to store learned machine-learning models in files that can subsequently be loaded to make predictions. When you have spent a lot of effort training a complex model and want to be able to reuse it without having to retrain it, this is quite helpful.

vi. **Flask:** Flask is a compact and adaptable framework that offers tools and utilities to developers so they may make web applications quickly and efficiently. When building online applications, Flask offers a straightforward and user-friendly interface with integrated support for managing HTTP requests and answers, routing, and URL management. Additionally, it supports templating and offers a straightforward method for creating HTML pages dynamically from user input.

6.3 Comparison of the Algorithms: To figure out which algorithm is the best; we have used the following metrics:
Accuracy Score:

	Algorithms	Accuracy
0	Logisitic Regression	0.974700
1	SVM	0.989348
2	Random Forest	0.986684
3	KNN	0.986019
4	Decision Tree	0.984021
5	Hybrid Classifier	0.988682

From the below graph, it is shown that the Support Vector Machine (SVM) has the highest accuracy, followed by the Hybrid Classifier. Logistic Regression (LR) gives the lowest accuracy among all the algorithms.

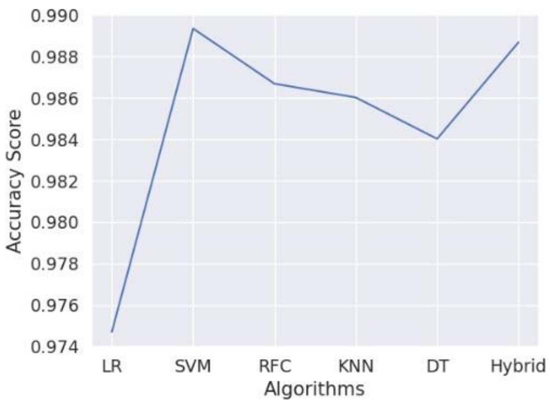


Fig. 4 Accuracy Score Comparision

Precision Score:

	Algorithms	Precision
0	Logisitic Regression	0.972934
1	SVM	0.988604
2	Random Forest	0.981638
3	KNN	0.980254
4	Decision Tree	0.980170
5	Hybrid Classifier	0.991392

The graph shows that the Hybrid Classifier obtains the highest precision score of 0.991, which is followed by the Support Vector Machine (SVM). The Random Forest Classifier (RFC) and K-Nearest Neighbors almost have the same scores. Logistic Regression (LR) has the lowest precision score of 0.972.

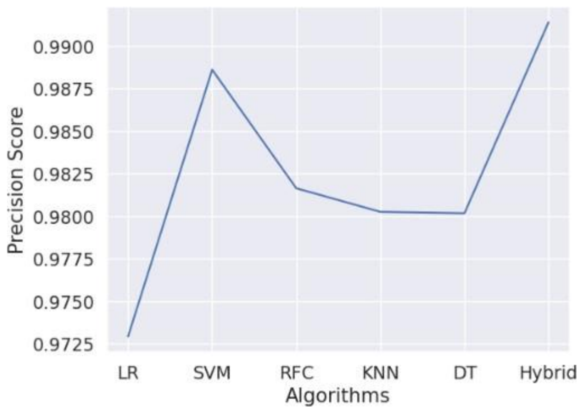


Fig. 5 Precision Score Comparison

Recall Score:

	Algorithms	Recall
0	Logisitic Regression	0.972934
1	SVM	0.988604
2	Random Forest	0.990028
3	KNN	0.990028
4	Decision Tree	0.985755
5	Hybrid Classifier	0.984330

After obtaining the results, it is observed that the Random Forest Classifier (RFC) and the K-Nearest Neighbors (KNN) classifiers get the highest recall scores, and the Logistic Regression algorithm obtains the lowest recall score.

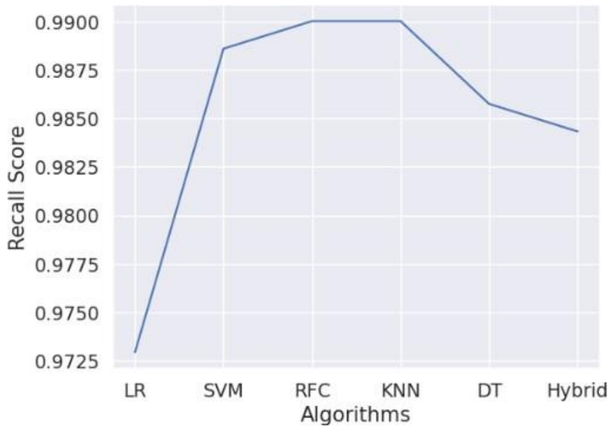


Fig. 6 Recall Score Comparision

VII. RESULTS

The accuracy metrics of the predicted models were trained and tested based on several factors in the dataset such as age, gender, education pursuing, sleeping troubles, anxiety and relaxation issues, external stimulus, and so on. The machine learning parameters include the size of the training model, kernel type for SVM model, number of estimators and the maximum depth of the tree for the Random Forest Classifier model, bias values for the Logistic Regression model, and the 'k' value for the K-Neighbors classifier. The metrics show that the Hybrid Classifier is the model with the highest accuracy of 98.8 percent.

	Accuracy	Precision	Recall
Logisitic Regression	0.974700	0.972934	0.972934
Support Vector Machine	0.989348	0.988604	0.988604
Random Forest	0.986684	0.981638	0.990028
KNN	0.986019	0.980254	0.990028
Decision Tree	0.984021	0.980170	0.985755
Hybrid	0.988682	0.991392	0.984330

Fig. 7 Accuracy Metrics of the learning model

Confusion Matrix: A confusion matrix gives us the summary of all the predictions in the form of a matrix. The key to the confusion matrix is all the values that have been sorted correctly and incorrectly. The following are the obtained confusion matrix results for the learning models.

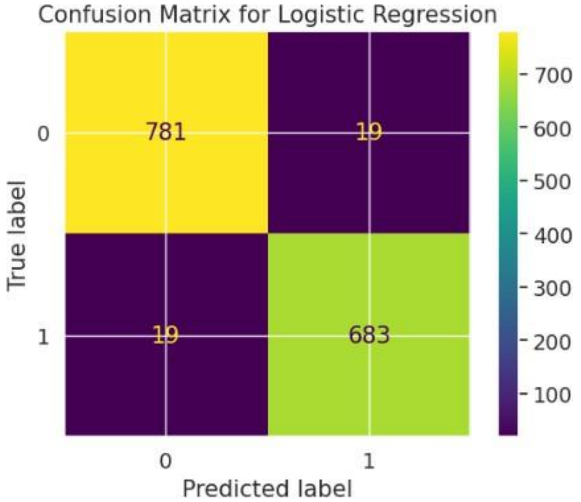


Fig. 8.1 Confusion Matrix for Logistic Regression

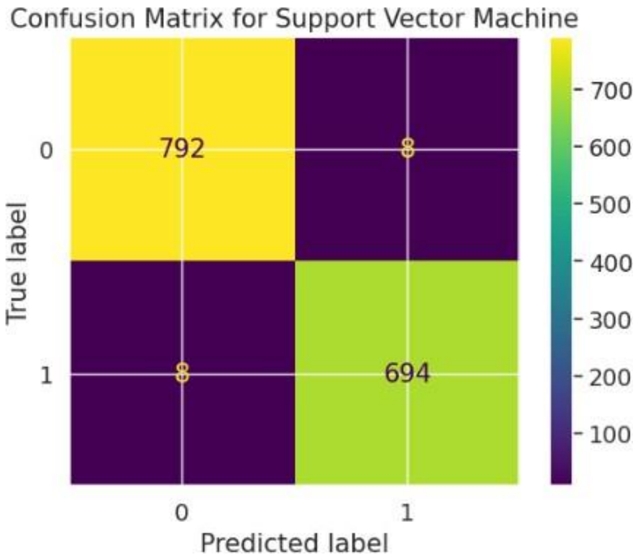


Fig. 8.2 Confusion Matrix for Support Vector Machine

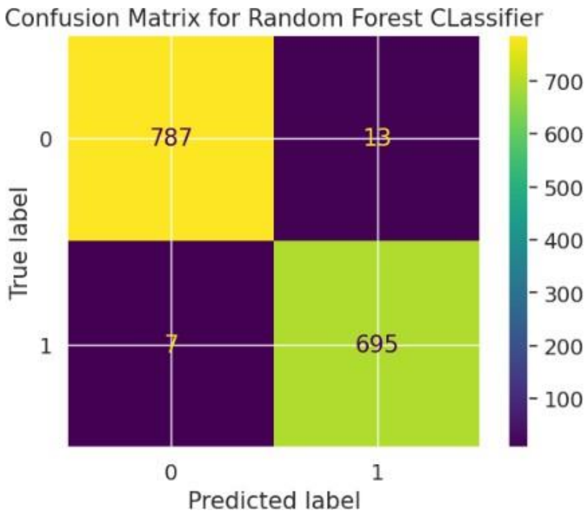


Fig. 8.3 Confusion Matrix for Random Forest Classifier

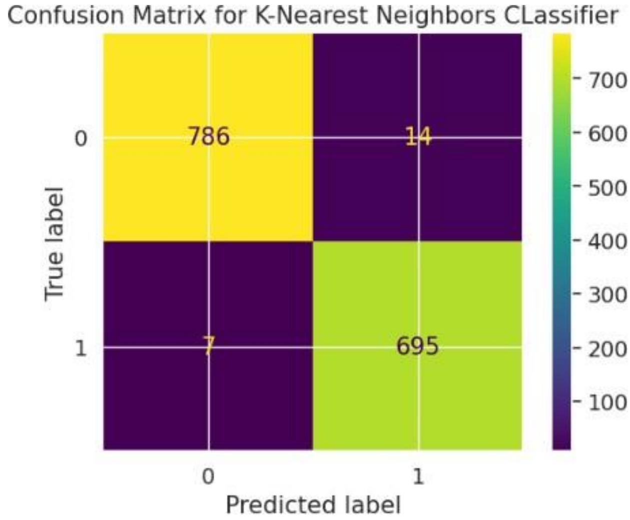


Fig. 8.4 Confusion Matrix for K Nearest Neighbors Classifer

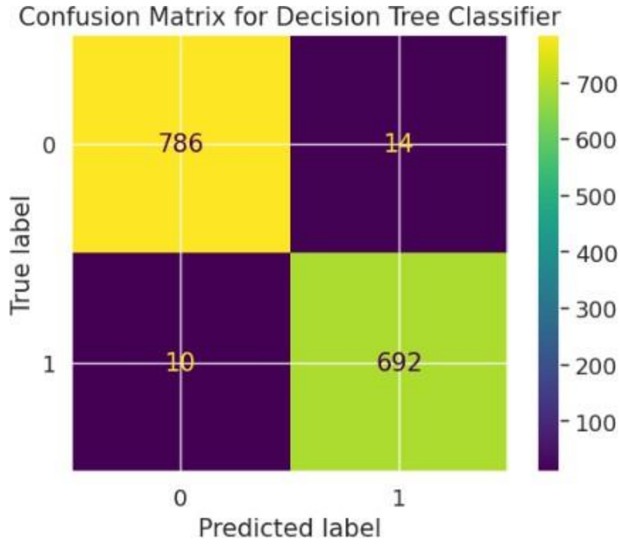


Fig. 8.5 Confusion Matrix for Decision Tree Classifier

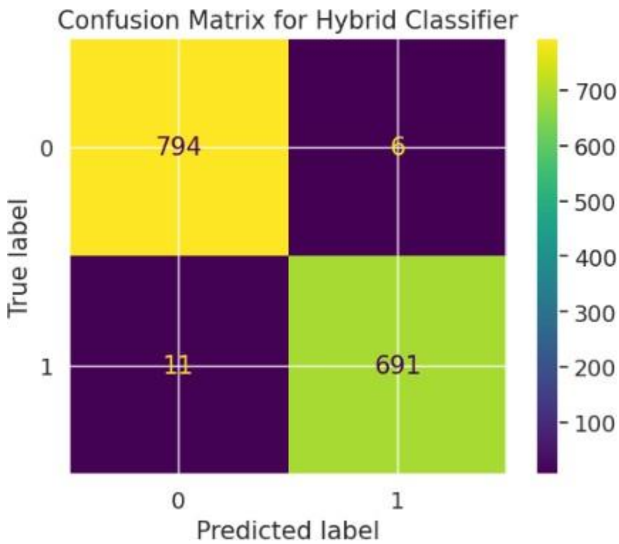


Fig. 8.6 Confusion Matrix for Hybrid Classifier

The top left part and the bottom right part of the confusion matrix give the number of data points that were correctly classified by the machine learning model, while the top right and the bottom left part give the number of data points that were incorrectly classified. From the given figures, we observe that the Support Vector Machine has the highest number of correctly classified points, making it the most accurate algorithm to be used.

Reasons for Mental health: The web interface explores the information on why mental health gets effected and the different factors that affect it externally. It also provides a way to contact therapists if help is needed.

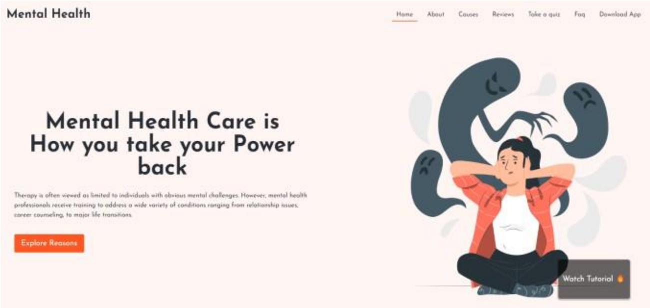
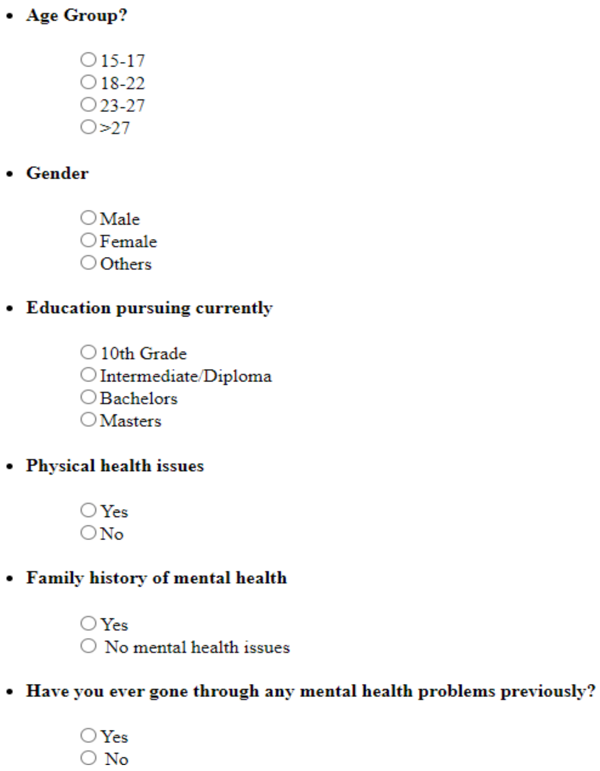


Fig. 9.1 Website for Mental health

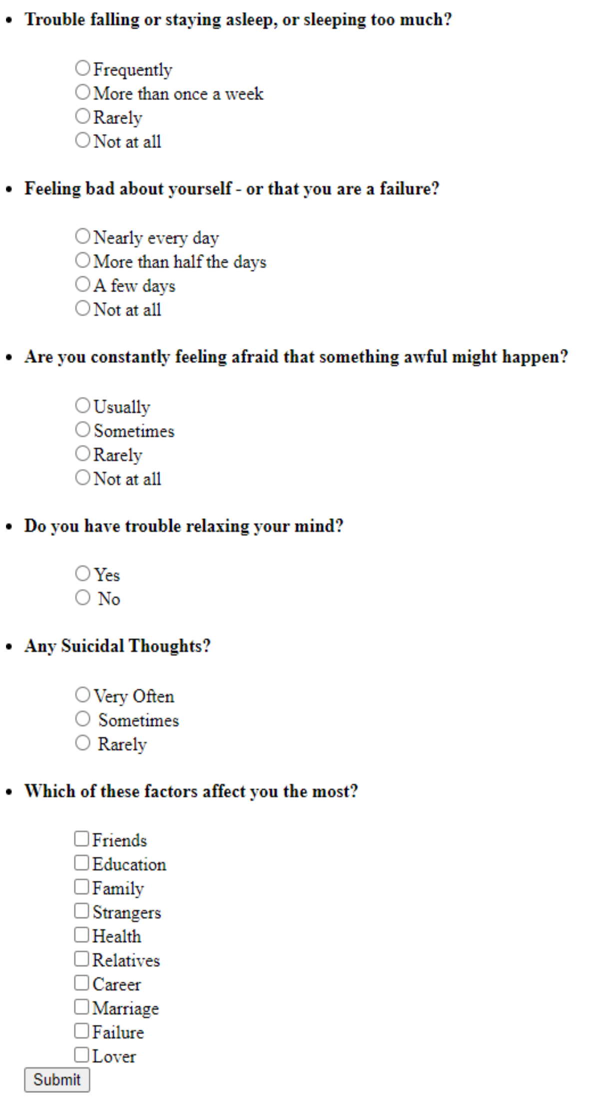
Fig. 9.2 Contacting Doctors

Mental Health quiz: This quiz will be taken by the students to give an analysis of the status of their mental health, and if they need to contact a therapist or not, based on the responses given.
The output is given either as ‘mentally depressed’ or ‘not mentally depressed’ accordingly.

Fig. 9.3 Questions based on personal factors.



- **Age Group?**
  - ○ 15-17
  - ○ 18-22
  - ○ 23-27
  - ○ >27
- **Gender**
  - ○ Male
  - ○ Female
  - ○ Others
- **Education pursuing currently**
  - ○ 10th Grade
  - ○ Intermediate/Diploma
  - ○ Bachelors
  - ○ Masters
- **Physical health issues**
  - ○ Yes
  - ○ No
- **Family history of mental health**
  - ○ Yes
  - ○ No mental health issues
- **Have you ever gone through any mental health problems previously?**
  - ○ Yes
  - ○ No



- **Trouble falling or staying asleep, or sleeping too much?**

  ○ Frequently
  ○ More than once a week
  ○ Rarely
  ○ Not at all

- **Feeling bad about yourself - or that you are a failure?**

  ○ Nearly every day
  ○ More than half the days
  ○ A few days
  ○ Not at all

- **Are you constantly feeling afraid that something awful might happen?**

  ○ Usually
  ○ Sometimes
  ○ Rarely
  ○ Not at all

- **Do you have trouble relaxing your mind?**

  ○ Yes
  ○ No

- **Any Suicidal Thoughts?**

  ○ Very Often
  ○ Sometimes
  ○ Rarely

- **Which of these factors affect you the most?**

  ☐ Friends
  ☐ Education
  ☐ Family
  ☐ Strangers
  ☐ Health
  ☐ Relatives
  ☐ Career
  ☐ Marriage
  ☐ Failure
  ☐ Lover

Submit

Fig. 9.4 Questions based on behavioral and external factors

VIII. CONCLUSION

The study of well-being is a complicated and varied topic that necessitates considering a wide variety of variables and perspectives, it can be inferred from a thorough investigation utilizing machine learning. The analysis's findings revealed that while stress, loneliness, and sedentary behavior are all negatively correlated with well-being, other elements, such social support, physical activity, and mindfulness practices, are positively associated with it.
Also, the investigation has shown how machine learning approaches can be used to analyze and forecast well-being results, providing fresh information in this area. Researchers can build more precise and individualized interventions by utilizing artificial intelligence to better grasp the intricate correlations between diverse well-being markers.

IX. FUTURE SCOPE

Customized well-being recommendations: By using these developed project algorithms by identifying individuals' day to day behaviors, activities, and habits, we can work on providing personalized recommendations to know their mental health status and by then can improve their well-being health accordingly. For example, the system could suggest changes in diet or exercise based on a person's daily routine and lifestyle.

Emotional well-being examination: AI calculations can be prepared to break down designs in discourse, virtual entertainment action, or different information sources to distinguish indications of emotional well-being issues. The framework could give constant criticism to people and medical care experts on changes in prosperity pointers, which could help forestall and treat psychological well-being issues.

REFERENCES

[1] World Health Organization. (1948). Constitution Of The World Health Organization. Retrieved From Https://Www.Who.Int/About/Who-We-Are/Constitution

[2] Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-being Index: A systematic review of the literature. Psychotherapy and Psychosomatics,84(3),167–176.

[3] Baxter, A.J., et al. (2014). "Global prevalence of anxiety disorders: a systematic review and meta-regression." Psychological Medicine, 44(11), pp. 2363-2374.

[4] Kessler, R.C., et al. (2005). "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication." Archives of General Psychiatry, 62(6), pp. 593-602.

[5] Insel, T.R. (2013). "Psychiatry in crisis." JAMA, 369(24), pp. 2289-2290.

[6] Cuthbert, B.N., & Insel, T.R. (2013). "Toward the future of psychiatric diagnosis: the seven pillars of RDoC." BMC Medicine, 11(1), p. 126.

[7] Koenen, K.C., et al. (2017). The hidden wounds of war: insights from neuroimaging studies of post-traumatic stress disorder.

[8] Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. Psychological Bulletin, 125(2), 276–302.

[9] Boehm, J. K., & Kubzansky, L. D. (2012). The heart's content: The association between positive psychological well-being and cardiovascular health. Psychological Bulletin, 138(4), 655–691.

[10] Well-being analysis using machine learning Taxonomy, Applications, and Challenges Jet Li Chung and Jason Teo (2021) Faculty of Computing and Informatics, University Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.

[11] Estimating Anxiety, Depression and Stress in modern Life using Machine Learning Algorithms; Anu Priya, Shruti Garg, Neha Prerna Tigga (2020) Procedia Computer Science 167(6):1258-1267.

[12] Dr.J.Arokia Renjit, Adlin Sajeesha M.J, Sangavai V.D, Sree Devi D.S (2022). Prediction of Mental Health using Machine Learning Journal of Emerging Technologies and Innovative Research, Volume 9, Issue 5.

[13 Roger Garriga, Javier Más, Samhara varaha, Jonnolan, Oliver Harrison, George Tadros, Aleks Andar Matic (2022) Machine Learning Model to predict Mental Health Crises from Electronic Health Records. Nat Med. 28(6): 1240-1248

[14] Sofianita Mutalib, et. al. (2021) Mental Health Prediction Models Using Machine Learning in Higher Education Institution. Turkish Journal of Computer and Mathematics Education (TURCOMAT) Vol. 12 No. 5

[15] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In Seventh International AAAI Conference on Weblogs and Social Media.

[16] Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., & Cannon, T. D. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry, 3(3), 243-250.

[17] Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., & Picard, R. (2015). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: A pilot study. In International Conference on Wireless Mobile Communication and Healthcare.

[18] Wang, R., Wang, W., Aung, M. S., Abdullah, S., Huang, Z., & Choudhury, T. (2018). Tracking depression dynamics in college students using mobile phone and wearable sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(4), 1-24.

[19] Kauer, S. D., Reid, S. C., Crooke, A. H., Khor, A., Hearps, S. J., Jorm, A. F., & Patton, G. (2014). Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. Journal of medical Internet research, 16(7), e173.

[20] Wahle, F., Kowatsch, T., Fleisch, E., & Rufer, M. (2016). Mobile sensing and support for people with depression: A pilot trial in the wild. JMIR mHealth and uHealth, 4(3), e111.

[21] Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. Journal of medical Internet research, 17(7), e175.

[22] Banaee, H., Ahmed, M. U., Loutfi, A., & Muhammad, G. (2013). A review on wearable sensors and systems with application in rehabilitation. Journal of neuroengineering and rehabilitation, 10(1), 1-20.

[23] Williams, A. C. D. C., & Craig, K. D. (2016). Updating the definition of pain. Pain, 157(11), 2420-2423.

[24] Torous, J., Roberts, L. W., & Needed, J. T. (2017). The ethical use of mobile health technology in clinical psychiatry. The Journal of nervous and mental disease, 205(1), 4-8.

[25] Lovejoy, C. A., Buch, V. & Maruthappu, M. Technology and mental health: the role of artificial intelligence. Eur. Psychiatry **55**, 1–3 (2019).

[26] Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. Annu. Rev. Clin. Psychol. **14**, 91–118 (2018).

[27] Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. Deep learning for healthcare: review, opportunities, and challenges. Brief. Bioinformatics **19**, 1236–1246 (2017).

[28] Vieira, S., Pinaya, W. H. & Mechelli, A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci. Biobehav. Rev. **74**, 58–75 (2017).

[29] Shatte, A. B., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: a scoping review of methods and applications. Psychol. Med. **49**, 1426–1448 (2019).

[30] Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. **151**, 264–269 (2009).

[31] Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods **16**, 111–116 (2019).

[32] J. Alonso, M. C. Angermeyer, S. Bernert, R. Bruffaerts, T. S. Brugha, H. Bryson, G. de Girolamo, R. de Graaf, K. Demyttenaere, I. Gasquet, J. M. Haro, S. J. Katz, R. C. Kessler, V. Kovess, J. P. Lépine, J. Ormel, G. Polidori, L. J. Russo, G. Vilagut, J. Almansa, S. Arbabzadeh-Bouchez, J. Autonell, M. Bernal, M. A. Buist-Bouwman, M. Codony, A. Domingo-Salvany, M. Ferrer, S. S. Joo, M. Martínez-Alonso, H. Matschinger, F. Mazzi, Z. Morgan, P. Morosini, C. Palacín, B. Romera, N. Taub, W. A. M. Vollebergh. (2014) Disability and Quality of life Imapct on Mental Disorders in Europe. 10.3390/healthcare11060921, 11, 6, (921),

[33] Veit Rothhammer, Davis M. Borucki, Emily C. Tjon, Maisa C. Takenaka, Chun-Cheih Chao, Alberto Ardura-Fabregat, Kalil Alves de Lima, Cristina Gutiérrez-Vázquez. (2018) Microglial control of astrocytes in response to microbial metabolites. Nature 557, 724-728.

[34] Pim Cuijpers, Annemieke van Straten, Lisanne Warmerdam, Gerhard Andersson et al. (2009). Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depress Anxiety 2009;26(3):279-88.

[35] David A Richards, David Ekers, Dean McMillan, Rod S Taylor, Sarah Byford, Fiona C Warren, Barbara Barrett, Paul A Farrand, Simon Gilbody, Willem Kuyken, Heather O'Mahen, Ed R Watkins, Kim A Wright, Steven D Hollon, Nigel Reed, Shelley Rhodes, Emily Fletcher, Katie Finning et al. (2016) Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial Lancet 2016 Aug 27;388(10047):871-80

[36] Zachary Munn, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur & Edoardo Aromataris et al. (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach BMC Medical Research Methodology volume 18, Article number: 143.

[37] American Psychological Association. Stress in America: Coping with Change, Part 1. Washington, DC: American Psychological Association; 2017.

[38] Merikangas KR, Ames M, Cui L, Ustun TB, Von Korff M, Kessler RC. The impact of comorbidity of mental and physical conditions on role disability in the US adult household population.Arch Gen Psychiatry. 2007;64(10):1180–1188.

[39] Scott KM, Lim C, Al-Hamzawi A, et al. Association of mental disorders with subsequent chronic physical conditions: work mental health surveys from 17 countries. JAMA Psychiatry. 2016;73(2):150–158.

[40] Glassman AH. Depression and cardiovascular comorbidity. Dialogues Clin Neurosci. 2007;9(1):9–17.

[41] Luppino FS, de Wit LM, Bouvy PF, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry. 2010;67(3):220–229.

[42] Melek SP, Norris DT, Paulus J, Matthews K, Weaver A, Davenport S. Potential Economic Impact of Integrated Medical-Behavioral Healthcare: Updated Projections for 2017. Milliman Research Report. Seattle, WA: Milliman, Inc.; 2018.

[43] US Department of Labor, Bureau of Labor Statistics. Databases, Tables & Calculators by Subject website. Labor force statistics from the Current Population Survey. Accessed July 3, 2018.

[44] Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity? J Occup Environ Med. 2008;50(4):401–410.

[45] Dewa CS, Thompson AH, Jacobs P. The association of treatment of depressive episodes and work productivity. Can J Psychiatry. 2011;56(12):743–750.

[46] Goetzel RZ, Anderson DR, Whitmer RW, et al; Health Enhancement Research Organization (HERO) Research Committee. The relationship between modifiable health risks and health care expenditures: an analysis of the multi-employer HERO health risk and cost database. J Occup Environ Med. 1998;40(10):843–854.

[47] Goetzel RZ, Pei X, Tabrizi MJ, et al. Ten modifiable health risk factors are linked to more than one-fifth of employer-employee health care spending. Health Aff. 2012;31(11):2474–2484.

[48] AACC (American Association of Community Colleges). 2019. Community college enrollment crisis? Historical trends in community college enrollment. Washington, DC.

[49] National Academies of Sciences, Engineering, and Medicine. 2021. Mental Health, Substance Use, and Wellbeing in Higher Education: Supporting the Whole Student. Washington, DC: The National Academies Press.

[50] ACCA (American College Counseling Association). 2020. Professional advocacy and public awareness (PAPA) paper: Outsourcing of counseling services. Alexandria, VA.

[51] ACHA. 2014. American College Health Association-National College Health Assessment II: Reference group data report, fall 2013. Hanover, MD.

[52] Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying Mental Health Signals in Twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 51–60, Baltimore, Maryland, USA. Association for Computational Linguistics.

[53] Nancy Eisenberg, Tracy L. Spinrad, and Natalie D. Eggum (2010) Emotion-Related Self-Regulation and Its Relation to Children’s Maladjustment. Annu Rev Clin Psychol. 2010 Apr 27; 6: 495–525.

[54] Teris Cheung, Siu Yi Wong, Kit Yi Wong, Lap Yan Law, Karen Ng, Man Tik Tong, Ka Yu Wong, Man Ying Ng, Paul S F Yip (2016) Depression, Anxiety and Symptoms of Stress among Baccalaureate Nursing Students in Hong Kong: A Cross-Sectional Study. Int J Environ Res Public Health 3;13(8):779.