Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 187–194 DOI: 10.15439/2024F3586 ISSN 2300-5963 ACSIS, Vol. 39

# A Machine Learning Approach for Anxiety and Depression Prediction Using PROMIS® Questionnaires

Arthur R. S. Vitória*, Murilo O. Guimarães*, Daniel Fazzioni*, Aldo A. Díaz-Salazar*,
Ana Laura S. A. Zara†, Iwens G. Sene Junior*, Renato F. Bulcão-Neto*
*Institute of Informatics, Federal University of Goiás, Brazil
†Institute of Tropical Pathology and Public Health, Federal University of Goiás, Brazil
arthurvitoria@discente.ufg.br
0000-0003-1746-9668

***Abstract*—A mental disorder is a clinically significant disturbance in an individual's cognition, emotional, or behavioral functioning. Mental disorders such as anxiety and depression can be accessed by psychiatrists using auxiliary tools such as the depression anxiety stress scale (DASS), patient reported outcome (PRO), patient reported outcome measures (PROMs) and patient reported outcomes measurement information system (PROMIS®). However, many individuals affected by the symptoms of mental disorders do not receive a proper diagnosis. In that context, this work proposes a machine learning approach to predict the score of anxiety and depression using PROMIS® questionnaires by performing a comparative study between supervised learning models to estimate the scores of anxiety and depression from individuals. Through the proposed model an average MAPE of 6.31%, $R^2$ of 0.76, and Spearman coefficient of 88.86 were achieved, outperforming widely used linear models such as support vector machines (SVM), random forest (RF), and gradient boosting (GB). In conclusion, the utilization of machine learning algorithms with PROMIS® questionnaires has shown promise as a methodology for assessing anxiety and depression scores from the participants' perspective, aligning with their perceptions of well-being.**

***Index Terms*—Anxiety, Depression, PROMIS®, Machine Learning, Mental Health**

## I. Introduction

THE CONDITIONS manifested through the symptoms associated with mental disorders have a far-reaching impact that permeates and affects the personal relationships, occupational pursuits, and general well-being of millions of people [1]. The World Health Organization (WHO) has recorded a significant global incidence of mental disorders, with 970 million individuals affected, with anxiety and depression being the most common. Moreover, a notable surge in these numbers were observed during the COVID-19 pandemic, with an increase of 26% and 28% for anxiety and depression, respectively [2].

As reported by the Brazilian Ministry of Health (MH), 18.6 million Brazilians are affected by anxiety, and mental disorders constitute a significant contributing factor of disabilities across the Americas. In Latin America, Brazil is the country with the highest prevalence of depression, reaching around 15.5% of the population [3]. Additionally, symptoms of these mental disorders between children and teenagers rose to 25.2% and 20.5% for depression and anxiety, respectively, during the COVID-19 pandemic. Underscoring the pressing need for comprehensive strategies aimed at enhancing the care and support for those affected by this reality.

Mental disorders can be assessed by psychiatrists using tools, such as DASS42 and DASS21 questionnaires, with a 42-item and 21-item, respectively. These self-administered instruments are designed to measure the magnitude of three negative emotional states: depression, anxiety and stress [4]. Mental disorders screening can also be carried out using patient reported outcome (PROs) and patient reported outcome measures (PROMs) with a validated accuracy. These measures rely on patient self-assessments covering aspects of well-being, including quality of life, symptom or symptom burden, experience of care, and mental health indicators like anxiety and depression [5], thus, providing insights from the patient's perspective and ideas of their own health [6].

The treatment for these mental disorders usually consists of medications and psychotherapy. However, a substantial challenge arises from the delayed diagnosis of these mental disorders, resulting in limited access to timely and proper interventions [7]. As the condition deteriorates, the individual's psychological capacity to seek treatment decreases, leading to an increase in the number of undiagnosed individuals [8]. Moreover, in primary care, only 50% of the patients with depression receive a diagnosis and only 15% receive a proper treatment [9]. In this way, emerging methodologies designed to enhance mental disorder screening are essential, as they can facilitate appropriate treatment and a decrease in the number of undiagnosed individuals.

Several studies have explored the application of statistical and machine learning models to predict therapy outcomes for a wide range of mental disorders [10]. Several studies primarily focus on making predictions regarding the long-term outcomes of patients with various conditions, either before diagnosis or during the course of treatment [11]. These conditions encompass schizophrenia [12], stress [13], [14], depression [15], [16], anxiety [17], [18], as well as other mental health disorders [19], [20]. However, since these mental disorders often manifest gradually, with symptoms becoming discernible in their early stages, it is crucial to explore predictive modeling to anticipate and address potential issues before they advance further.

In that context, this work proposes a machine learning approach to predict the score of anxiety and depression using PROMs questionnaires by performing a comparative study between supervised learning models to estimate the scores of anxiety and depression from individuals. Furthermore, we investigated the relative contribution of the input variables to improve the understanding and importance of each variable related to the individuals' perspective and ideas of their own health.

This paper is organized as follows. In Section II the most relevant and related works are reported, describing the applied methodologies, results, and conclusions. In Section III, materials and methods used for the prediction of the scores of anxiety and depression are described, such as the proposed approach for training an MLP-based model and its evaluation. Section III-C describes and discusses the results of the relative importance of the input variables to the MLP model. Section V reports the results and Section VI describes the conclusions.

## II. Related Work

Literature search was conducted for articles that addressed the prediction of anxiety and depression levels using machine learning algorithms to provide context for current research and highlight advances and challenges in the prediction of mental health disorders. The databases used included PubMed, Scopus, Google Scholar, IEEE and IET. The keywords used were "anxiety", "depression", "machine learning", "deep learning", "questionnaires". As a result, 6 articles were selected as foundation for the current research.

In the research of [21], an attempt was made to determine five different levels of severity of anxiety, depression and stress. For the dataset, the DASS 21 questionnaire, which measures the level of anxiety, depression, and stress, was applied to 348 participants aged between 20 and 60. These questionnaires were used in five models: Decision Tree, Random Forests, Naïve Bayes, and $k$-Nearest Neighbor (KNN). Due to the unbalanced classes, the metric chosen was the F1-score. The model that showed the best F1-score for stress prediction was Random Forest with 71%. The best F1-score for depression was Naïve Bayes with 83%. For anxiety, none of the models performed well, reaching an average of 50%.

In the survey of [22], the DASS 42 questionnaire was used and filled in online by randomly chosen users between 2017 and 2019. Eight models were used: Naïve Bayes, Bayesian networks, $k$-nearest neighbors, multi-layer perceptron (MLP), radial basis function network (RBFN), random forest and J48. The results showed that the RBFN obtained the best accuracy in classifying the conditions of anxiety, depression and stress, with an average of 96% for each of the variables. A second test using the DASS 21 questionnaire showed that the MLP model achieved the best accuracy with 96% for stress, 93% for depression, and 98.8% for anxiety.

The study of [23] used a questionnaire with 55 questions. The answers of 604 participants were recorded. Depression was assessed using the Burns Depression Checklist (BDC). The study used various feature selection techniques to identify the most relevant ones. Six machine learning algorithms were applied to predict depression, with AdaBoost and the SelectKBest techniques achieving the highest accuracy of 92.56%.

In the study of [24], a group of people with autoimmune diseases were assessed, 637 participants completed a structured clinical interview for SDM-IV-TR axis disorders (SCID) and various PROMs. The models used include Logistic Regression (LR), Neural Networks (NN) and Random Forests, and were trained to predict anxiety and major depressive disorder (MDD) scores. As a result, the area under the curve (AUC) and Brier scores ranged from 0.87 to 0.91 and 0.07 (*i.e.*, no variation) for the MDD models and from 0.79 to 0.83 and 0.09 to 0.11 for the anxiety disorder models. In the LR and NN models, few PROMs items were needed to achieve an optimal performance.

The study of [25] aimed to develop an appropriate predictive model to diagnose anxiety and depression among elderly patients based on sociodemographic and health-related factors. Ten classifiers were evaluated using a dataset of 510 geriatric patients and tested using a 10-fold cross-validation method. The highest prediction accuracy of 89% was obtained with the random forest (RF) classifier. This RF model was tested with another dataset of 110 separate elderly patients for external validity. Its predictive accuracy was 91%, and the false positive rate was 10%, compared to the standard tool.

In [26], data collected from 935 university students in Bangladesh was used. The data included student demographic information, such as academic year, grade point average, and the results of two depression assessment scales: the Beck Depression Inventory (BDI-II) and the Anxiety, Depression and Stress Scales - Bangla Version (DASS 21-BV). In addition, the students answered a set of 16 questions related to the reasons for their depression. This data was used to train and test three machine learning algorithms: $k$-Nearest Neighbor (KNN), Random Forest (R), and Support Vector Machine (SVM). The RF algorithm showed the best performance, achieving an accuracy rate of 75% in identifying depression.

## III. Material and Methods

This section presents a comprehensive overview of the data collection process, the definition of the multi-layer perceptron model (MLP), and how we measure the relative contribution of its input variables.

### A. Dataset

A cohort study was carried out with primary data collection, approved by the Institutional Review Board (n. 5513411). The inclusion criteria were students, teachers or administrative technicians from a Brazilian public university, aged $\geq$ 18 years, attending one of the 22 monitored environments for at least two hours.

Data collection for this study was carried out using the Research Electronic Data Capture (REDCap) [27] platform, which is a highly versatile tool designed to meet the needs of research in the medical, public health and social sciences fields. Using REDCap, we developed customized electronic forms to collect relevant information.

At the beginning, participants filled in the PROMIS®Global-10 (or PROMIS®10) V2 Questionnaire [28] to identify the physical and mental health scores. After this initial phase, participants were instructed to take part in the follow-up process, supervised by recruiters. As some studies report, individuals mental health can be influenced by the quality of the air and the indoor environment in which they spend time [29], [30]. In that context, during this phase, after at least 30 minutes of stay in the monitored environment, the participants were invited to answer three sets of questions once a day:

- Perception of indoor air quality, temperature, and humidity [31].
- PROMIS® Anxiety Short Form: 8 questions related to the participants' feelings in order to calculate an anxiety score.
- PROMIS® Depression Short Form: 8 questions designed to calculate a depression score.

In the follow-up period, participants had the opportunity to answer the questionnaires up to 15 times. This made it possible to obtain a comprehensive and detailed longitudinal view of the trends and variations over time of the input variables under study.

The dataset has the following characteristics:

- Total variables: 80
- Date of first follow-up: 05/15/23
- Date of last follow-up: 08/14/23
- Total participants: 249
- Total tracking records: 1924

### B. *Artificial Neural Networks*

An Artificial Neural Network (ANN) consists of fully connected artificial neurons organized into distinct layers: input, output, and hidden layers. The network's input is derived from samples, which provide the information for the ANN to assimilate and *learn* during its training. The most popular structure among ANNs is the feedforward [32]. In this work, we use a MLP, wherein information flows exclusively from the network's input layer to the output layer through a series of interconnected neurons.

Throughout the forward process during training an MLP model characterized by a predefined set of hidden layers $(h_1, h_2, ..., h_N)$, each layer $h_i$ containing $n_i$ neurons, the sequence of calculations for each layer can be described as follows:

$$h_i^j = f\left(\sum_{k=1}^{n_{i-1}} w_{k,j}^{i-1} h_{i-1}^k\right) \tag{1}$$

where $i = 2, ..., N$, since the first layer receives the input data directly, and $j = 1, 2, ..., n_i$. The weights $w_{k,j}^i$ represents the connection weights between the neuron $k$ in the hidden layer $i$ and the neuron $j$ in the next hidden layer, the number of neurons in the $i$th hidden layer is denoted by $n_i$. After the forward pass, a cost function computes the prediction-reference difference, leading to error calculations. This error drives the back-propagation algorithm, adjusting MLP weights and biases. The process continues until it meets the predefined stopping criterion, which, in this study, was determined only by the number of epochs.

### C. *Feature Importance*

Efforts to improve the interpretability of neural network models have led to the development of methods designed to estimate the relative contributions of input variables [33]. An ANN model can be represented as three set of layers, input, hidden, and output, and is often represented by weight matrices. However, the representation by weight matrices alone does not inherently convey valuable knowledge. In that context, these approaches often rely on the utilization of weight matrices within the neural network model to estimate the contributions of input variables [34].

The connection weight algorithm [35] measures the contributions of input variables by multiplying the raw connection weights from input neurons to hidden neurons with the connection weights from hidden neurons to output neurons and then summing these products across all the input neurons. The relative importance of a variable can be measured as

$$R_i = \sum_{j=1}^{h} w_{ij} w_{jo} \tag{2}$$

where $R_i$ is the relative importance of the input variable $i$, $h$ is the total number of hidden nodes in a layer, $w_{ij}$ is the weight of the connection between input node $i$ and hidden node $j$, and $w_{jo}$ is the weight of the connection between the hidden node $j$ and the output node $o$.

## IV. Experiments

The follow-up data of the participants contained a longitudinal view of the trends and variations over time of the input variables under study and were used in order to predict the scores of anxiety and depression. An overall structure of the experiments is presented in Figure 1, with a total of 249 participants and 74 input variables associated to each one. As part of the data preparation process, the collected responses were preprocessed carefully to enable model training. Additionally, the raw data is normalized between -1 and 1, allowing a adjust to a common magnitude scale, providing a more effective weight adjustment during the training time [36]. All the models were optimized using a grid search approach to identify hyperparameters that best matched the specific data settings with the aim of mitigating the risk of overfitting.

The participant data was partitioned into training, validation, and test sets using a 5-fold cross-validation approach. A representation of such process is depicted in Figure 2, each fold splits $N$ participants into 5 sets for training, validation and

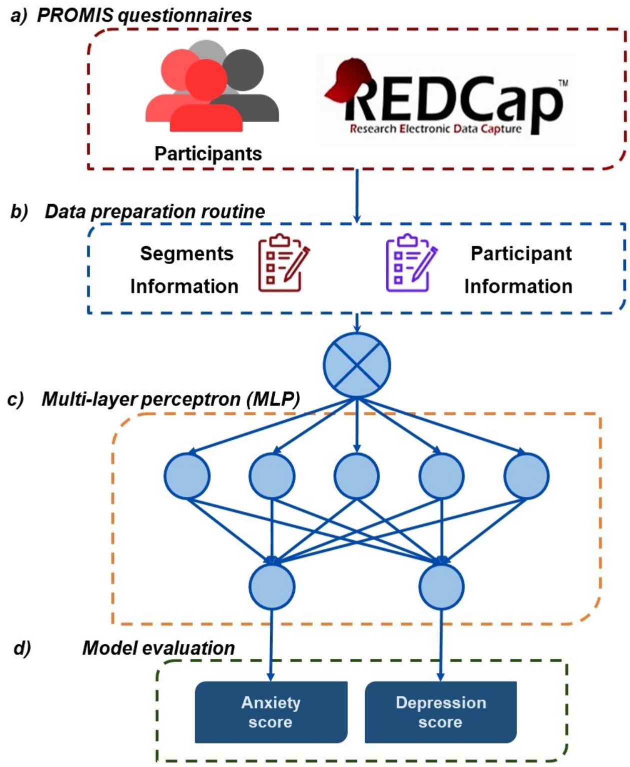


Fig. 1. Overall structure of the experiments. a) During the follow-up process, the input variables of the participants were collected using PROMIS® questionnaires along with questions related to their perception of indoor air quality; b) To prepare the data for the model training, we applied preprocessing routines that included various steps, such as data cleaning, feature engineering, and data normalization; c) An overall representation of the model structure, where each input variable is used as a feature, along with all participants information to train and test the model; d) A model evaluation routine was applied to systematically assess and analyze the outcomes achieved during training and testing phases.

testing. To prevent data leakage, all the follow-up information related to the participants was used exclusively within one set for each fold.

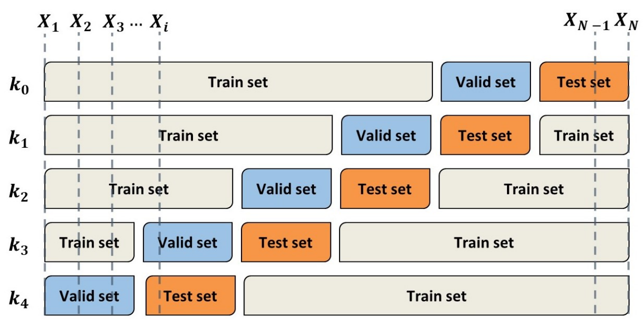


Fig. 2. The 5-fold cross-validation strategy.

The search space of hyperparameters in the proposed MLP model was tested extensively with the focus of determining the optimal number of neurons within each layer of the model, encompassing a wide array of combinations. The Adam [37] algorithm was used as an optimizer for the training. The model's performance was assessed across multiple randomization seeds to measure their impact in different initialization. This assessment plays an important role in guiding the learning process away from undesired local minima, a critical aspect that needs careful consideration.

TABLE I
SEARCH SPACE OF MLP HYPERPARAMETERS.

| Hyperparameter | Search Space |
|---|---|
| Hidden Layers | $1, 2, 3$ |
| Number of Neurons | $16, 32, 64, 128, 256, 512$ |
| Batch Size | $128$ |
| Epochs | $500$ |
| Seeds | $73, 42, 10, 3407, 103$ |
| Learning rate | $1e{-4}$ |

Four other models were chosen as a baseline for comparison to the the MLP model: a model that combines multiple decision trees to make predictions (Random Forest); a supervised learning model that seeks to find a hyperplane for classification or regression (SVM); decision tree (DT), which makes decisions in a tree-like structure by splitting data based on rules; and gradient boosting (GB), an ensemble model that combines multiple DTs sequentially.

Each model was initialized and trained with the training set corresponding to the 5-fold cross-validation iteration. The training data included the input variables presented in the PROMIS® questionnaires, the perception of indoor air quality and, the corresponding output scores of anxiety and depression used as supervised labels. After training, each model was tested using the test set from the corresponding cross-validation iteration. This allowed the assessment of the model's performance on independent data.

Each model was assessed in terms of predictive capability using diverse criteria to enable the evaluation of various aspects of their generalization capabilities. These criteria include the mean absolute percentage error (MAPE), represented in Equation 3, where $n$ represents the amount of data, $\hat{y}_i$ is the predicted score and $y_i$ is the measured score, addressing the performance of the models based on the average percentage difference between predicted and measured values.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} * 100 \tag{3}$$

The coefficient of determination $R^2$ described in Equation 4 expresses the capabilities of the model to accommodate the variance in the data. It shows the proportion of the variability in the dependent variable $y$ that is explained by the independent variables, where $n$ is the number of observations, $y_i$ is the measured value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of measured values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

The Spearman Correlation Coefficient represented in Equation 5 allows a quantification of the degree and direction of the monotonic relationship between predicted and measured values, where $n$ is the number of observation pairs and $d_i$ represents the difference between ranked values of the two variables. It offers a non-parametric measure of the association of predicted and measured values. It varies from $-1$ (perfectly decreasing correlation) to 1 (perfectly increasing correlation), with 0 indicating no correlation.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{5}$$

## V. Results and Discussion

The dataset was splited using the $k$-fold cross-validation, with $k = 5$ to divide the data in distinct folds, with each fold alternately serving as training, validation and test set. This allowed the models to be repeatedly trained and evaluated on different data combinations. In each iteration of $k$-fold, the models were trained on the training data, specific to that iteration, and their performance was assessed using the corresponding test data. The process was iterated on all possible fold combinations, resulting in a comprehensive evaluation of the models' performance across different scenarios. The results were evaluated in terms of MAPE, $R^2$, and the Spearman Correlation Coefficient.

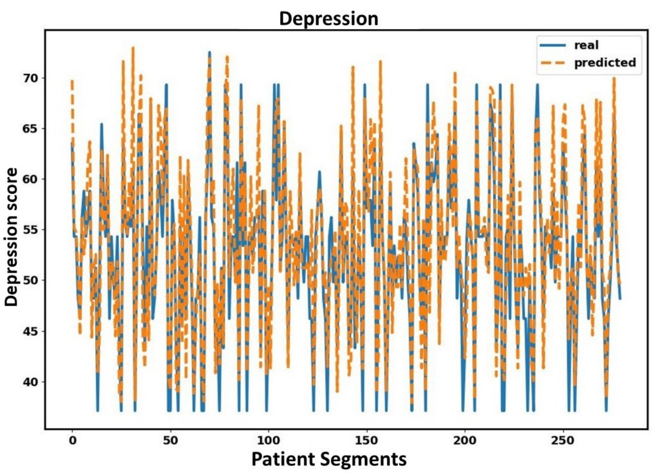


Fig. 3. Predictions for depression scores.

The difference in performance among the models is rooted in their intrinsic characteristics. Ensemble models, such as RF and GB, stand out for their capacity of diversification, variance reduction, and resist overfitting. They combine multiple decision trees, capturing different aspects of the data and aggregate predictions to provide robust results. In contrast, individual models like DT are simpler, with hierarchical rule structures that can limit their ability to capture complex relationships. They are prone to overfitting as its depths grows, especially in high-dimensional problems or data with intricate relationships.

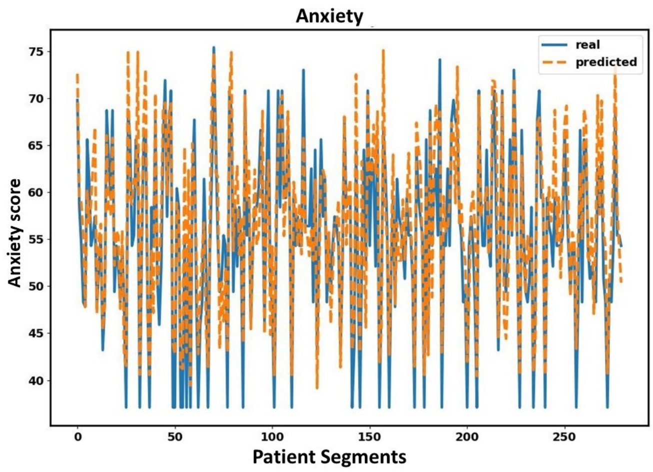


Fig. 4. Predictions for anxiety scores.

Table II provides a summary of the predictive capabilities of anxiety and depression for different models. The results show a better performance of the RF and GB regressors among five selected models in predicting anxiety and depression. These two models exhibited relatively lower MAPE values and higher $R^2$ scores, demonstrating strong ability to explain the variance in the data. In contrast, the DT model demonstrated an acceptable performance in predicting anxiety, but encountered challenges in explaining the variation in depression data, as evidenced by its lower $R^2$ score. The RF model displayed a robust performance, yielding a relatively low MAPE of $5.51\%$ for anxiety and $9.16\%$ for depression. Moreover, it achieved a high $R^2$ score, indicating its strong capacity to elucidate the variance within the data. The SVM model delivered reasonable results with a MAPE of $8.10\%$ for anxiety and $9.10\%$ for depression, although a slightly lower $R^2$ value when compared to the RF. The DT grappled with predicting depression, which is evident in its higher MAPE of $10.37\%$ and a lower $R^2$ of $0.30$. In contrast, the GB model achieved a MAPE of $7.05\%$ for anxiety and $9.81\%$ for depression, while maintaining fairly low $R^2$ values.

Figure 3 and Figure 4 show the predictions of the best trained model using $k$-fold and a grid-search optimization for anxiety and depression scores. A visual inspection in these results shows that the proposed MLP model performed well in both scenarios, where the only significant discrepancies that can be point out are in segments that reach the minimum score due to the lack of information provided by the participant. Using the connection weights algorithm proposed by [35], the relative contribution of each input variable was accessed and showed in Figure 5. It is evident that throughout the training process, the model consistently exhibited a tendency to prioritize similar variables for both anxiety and depression. Moreover, these input are identified in Table III and Table IV, where they are sorted in a descending order, where the first feature is the most relevant. It is worth mention that despite seventy four input variables, the model learned to give

TABLE II
RESULTS OF MAPE, $R^2$, AND CORRELATION COEFFICIENT OF SPEARMAN FOR ANXIETY AND DEPRESSION.

| Models** | Anxiety | | | Depression | | |
|---|---|---|---|---|---|---|
| | **MAPE** | $R^2$ | **Spearman** | **MAPE** | $R^2$ | **Spearman** |
| RF | **5.51**% ± 0.02 | **0.79** ± 0.16 | **0.86** ± 0.10 | 9.16% ± 0.02 | 0.54 ± 0.02 | 0.82 ± 0.08 |
| SVM | 8.10% ± 0.01 | 0.66 ± 0.08 | 0.84 ± 0.05 | 9.10% ± 0.01 | 0.60 ± 0.06 | 0.82 ± 0.05 |
| DT | 5.67% ± 0.03 | 0.62 ± 0.31 | 0.80 ± 0.15 | 10.37% ± 0.03 | 0.30 ± 0.06 | 0.75 ± 0.13 |
| GB | 7.05% ± 0.01 | 0.76 ± 0.05 | 0.86 ± 0.03 | 9.81% ± 0.01 | 0.53 ± 0.06 | 0.83 ± 0.13 |
| MLP * | 6.98% ± 0.02 | 0.72 ± 0.03 | 0.85 ± 0.01 | **5.64**% ± 0.04 | **0.80** ± 0.02 | **0.91** ± 0.02 |

** Best results for optimization and cross validation processes.

TABLE III
INPUT RELATIVE CONTRIBUTION FOR ANXIETY SCORES.

| **Feature Importance – Anxiety Prediction** |
|---|
| How often have you been bothered by emotional problems, such as feeling anxious, depressed or irritable? |
| General clinical symptoms (weakness, tiredness, nausea, others). |
| Neurological symptoms (headache, migraine, dizziness, others). |
| Respiratory symptoms (sneezing, stuffy nose, runny nose, difficulty breathing, dry throat or sore throat, others). |
| In general, how would you rate your mental health, including your mood and your ability to think? |
| Dermatological symptoms (burning skin, redness, allergies, etc.) |
| In general, rate how well you manage to carry out your frequent social activities and functions (including activities at home, at work and in the community, and responsibilities as a parent, child, spouse, employee, friend, etc.). |
| Do you carry out any professional activities? |
| Do you have any of these chronic diseases? (choice = sinusitis) |
| How well can you perform daily physical activities such as walking, climbing stairs, carrying groceries from the supermarket or moving a chair? |

TABLE IV
INPUT RELATIVE CONTRIBUTION FOR DEPRESSION SCORES.

| **Feature Importance – Depression Prediction** |
|---|
| How often have you been bothered by emotional problems, such as feeling anxious, depressed, or angry? |
| General clinical symptoms (weakness, tiredness, nausea, others). |
| Neurological symptoms (headache, migraine, dizziness, others). |
| In general, how would you rate your mental health, including your mood and your ability to think? |
| Respiratory symptoms (sneezing, stuffy nose, runny nose, difficulty breathing, dry throat or sore throat, others). |
| Dermatological symptoms (burning skin, redness, allergies, etc.). |
| In general, rate how well you are able to carry out your frequent social activities and functions (including activities at home, at work and in the community, and responsibilities as a parent, child, spouse, employee, friend, etc.). |
| Do you have any of these chronic diseases? (choice = sinusitis) |
| On average, how would you rate your tiredness? |
| Do you carry out any professional activities? |

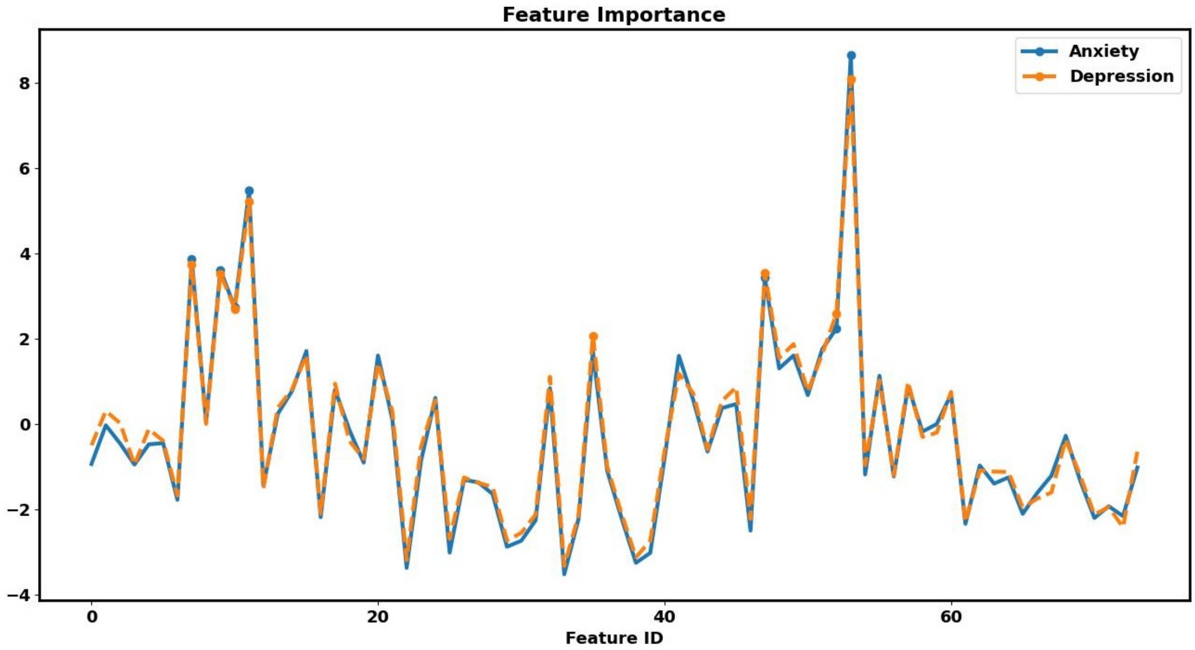


Fig. 5. Relative importance of the input variables for predict the anxiety and depression scores.

more importance around questions related to the general well-being of participants, as well as questions about symptoms they presented during the follow-up process, including general clinical, respiratory, and neurological, and dermatological in case of anxiety. Furthermore, some questions concerning participants' views on physical health, including aspects like fatigue, engagement in social activities, and functions that encompasses personal activities can be observed as more important.

## VI. Conclusion

This work investigates the potential of self-assessments covering aspects of well-being, mental health indicators, and the perception of indoor air quality, collected through PROMIS® questionnaires to measure scores of anxiety and depression for several participants. The dataset encompasses data from $219$ participants who volunteered during a follow-up period. Comprehensive information regarding well-being status and perception of the surroundings were periodically collected through questionnaires in predefined environments, with a total of $1924$ tracking records. However, despite a good amount of data from the participants, there was a lack of more representative scores for severe depression and anxiety to achieve a model with a good generalization.

A comparative study of supervised learning methods was conducted to measure scores of anxiety and depression for all the questionnaires under evaluation. Although simpler models demonstrated exceptional performance for anxiety scores, achieving $5.51\%$ of MAPE and $0.79$ of $R^2$ for the RF model, the prediction performance for depression scores was not as impressive, with the lowest MAPE of $9.16\%$ and a highest $R^2$ of $0.54$. The proposed MLP model outperformed the baseline, achieving an average MAPE for both anxiety and depression of $6.31\%$, an $R^2$ of $0.76$, and a Spearman coefficient of $88.86$. This suggests that machine learning models with the ability of capture more complex patterns in data, such as MLPs, might be better suited for addressing the prediction of anxiety and depression scores. Through the connection weight algorithm it was possible to determine the relative importance of each input feature. Additionally, future work could include other mental disorders, such as burnout [38].

## References

[1] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020.

[2] "Mental health and covid-19: Early evidence of the pandemic's impact," 2022.

[3] MH, "Depressão," https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/depressao, 2023, last acessed 05 Oct 2023.

[4] L. Parkitny and J. McAuley, "The depression anxiety stress scale (dass)," *Journal of physiotherapy*, vol. 56, no. 3, p. 204, 2010.

[5] T. Weldring and S. M. Smith, "Article commentary: patient-reported outcomes (pros) and patient-reported outcome measures (proms)," *Health services insights*, vol. 6, pp. HSI–S11 093, 2013.

[6] L. G. Tennenhouse, R. A. Marrie, C. N. Bernstein, L. M. Lix *et al.*, "Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease," *Journal of psychosomatic research*, vol. 134, p. 110126, 2020.

[7] R. Razavi, A. Gharipour, and M. Gharipour, "Depression screening using mobile phone usage metadata: a machine learning approach," *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 522–530, 2020.

[8] A. L. B. Melo and A. L. F. Alves, "Aplicação de técnicas de aprendizagem de máquina para diagnóstico de depressão, ansiedade e estresse," in *Anais do IX Encontro Nacional de Computação dos Institutos Federais*. SBC, 2022, pp. 13–16.

[9] A. J. Mitchell, A. Vaze, and S. Rao, "Clinical diagnosis of depression in primary care: a meta-analysis," *The Lancet*, vol. 374, no. 9690, pp. 609–619, 2009.

[10] S. Hornstein, V. Forman-Hoffman, A. Nazander, K. Ranta, and K. Hilbert, "Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach," *Digital Health*, vol. 7, p. 20552076211060659, 2021.

[11] A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological medicine*, vol. 49, no. 9, pp. 1426–1448, 2019.

[12] N. C. Hettige, T. B. Nguyen, C. Yuan, T. Rajakulendran, J. Baddour, N. Bhagwat, A. Bani-Fatemi, A. N. Voineskos, M. M. Chakravarty, and V. De Luca, "Classification of suicide attempters in schizophrenia using sociocultural and clinical features: A machine learning approach," *General hospital psychiatry*, vol. 47, pp. 20–28, 2017.

[13] J. L. Hagad, K. Moriyama, K. Fukui, and M. Numao, "Modeling work stress using heart rate and stress coping profiles," in *Principles and Practice of Multi-Agent Systems: International Workshops: IWEC 2014, Gold Coast, QLD, Australia, December 1-5, 2014, and CMNA XV and IWEC 2015, Bertinoro, Italy, October 26, 2015, Revised Selected Papers 5*. Springer, 2016, pp. 108–118.

[14] Y. Nakashima, J. Kim, S. Flutura, A. Seiderer, and E. André, "Stress recognition in daily work," in *Pervasive Computing Paradigms for Mental Health: 5th International Conference, MindCare 2015, Milan, Italy, September 24-25, 2015, Revised Selected Papers 5*. Springer, 2016, pp. 23–33.

[15] T. Hajek, C. Cooke, M. Kopecek, T. Novak, C. Hoschl, and M. Alda, "Using structural mri to identify individuals at genetic risk for bipolar disorders: a 2-cohort, machine learning study," *Journal of Psychiatry and Neuroscience*, vol. 40, no. 5, pp. 316–324, 2015.

[16] Y. Hou, J. Xu, Y. Huang, and X. Ma, "A big data application to predict depression in the university based on the reading habits," in *2016 3rd International Conference on Systems and Informatics (ICSAI)*. IEEE, 2016, pp. 1085–1089.

[17] F. Liu, W. Guo, J.-P. Fouche, Y. Wang, W. Wang, J. Ding, L. Zeng, C. Qiu, Q. Gong, W. Zhang *et al.*, "Multivariate classification of social anxiety disorder using whole brain functional connectivity," *Brain Structure and Function*, vol. 220, pp. 101–115, 2015.

[18] T. Tran and R. Kavuluru, "Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks," *Journal of biomedical informatics*, vol. 75, pp. S138–S148, 2017.

[19] M. Khondoker, R. Dobson, C. Skirrow, A. Simmons, and D. Stahl, "A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies," *Statistical methods in medical research*, vol. 25, no. 5, pp. 1804–1823, 2016.

[20] W. J. Bosl, T. Loddenkemper, and C. A. Nelson, "Nonlinear eeg biomarker profiles for autism and absence epilepsy," *Neuropsychiatric Electrophysiology*, vol. 3, no. 1, pp. 1–22, 2017.

[21] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020.

[22] P. Kumar, S. Garg, and A. Garg, "Assessment of anxiety, depression and stress using machine learning models," *Procedia Computer Science*, vol. 171, pp. 1989–1998, 2020, third International Conference on Computing and Network Communications (CoCoNet'19).

[23] M. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, and M. S. Uddin, "An in-depth analysis of machine learning approaches to predict depression," *Current Research in Behavioral Sciences*, vol. 2, p. 100044, 2021.

[24] L. G. Tennenhouse, R. A. Marrie, C. N. Bernstein, and L. M. Lix, "Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease," *Journal of Psychosomatic Research*, vol. 134, p. 110126, 2020.

[25] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, 2017.

[26] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, "Predicting depression in bangladeshi undergraduates using machine learning," in *2019 IEEE Region 10 Symposium (TENSYMP)*, 2019, pp. 789–794.

[27] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (redcap)—a metadata-driven methodology and workflow process for providing translational research informatics support," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 377–381, 2009.

[28] "Promis® instrument development and validation scientific standards version 2.0," 2013, revised May 2013.

[29] E. Finell and J. Nätti, "The combined effect of poor perceived indoor environmental quality and psychosocial stressors on long-term sickness absence in the workplace: A follow-up study," *International journal of environmental research and public health*, vol. 16, no. 24, p. 4997, 2019.

[30] T. Tuuminen, M. Andersson, S. Hyvönen, J. Lohi, and K. Vaali, "Indoor air nontoxicity should be proven with special techniques prior claiming that it may cause a variety of mental disorders." *International Journal of Hygiene and Environmental Health*, vol. 229, pp. 113 545–113 545, 2020.

[31] "Nbr 16401-3: instalações de ar-condicionado, sistemas centrais e unitários: parte 3: qualidade do ar interior," 2008.

[32] Y.-S. Park and S. Lek, "Artificial neural networks: Multilayer perceptron for ecological modeling," in *Developments in environmental modelling*. Elsevier, 2016, vol. 28, pp. 123–140.

[33] S. S. Haykin *et al.*, "Neural networks and learning machines." 2009.

[34] N. L. Costa, M. D. Lima, and R. Barbosa, "Evaluation of feature selection methods based on artificial neural network weights," *Expert Systems with Applications*, vol. 168, p. 114312, 2021.

[35] J. D. Olden and D. A. Jackson, "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks," *Ecological modelling*, vol. 154, no. 1-2, pp. 135–150, 2002.

[36] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[38] M. Grządzielewska, "Using machine learning in burnout prediction: A survey," *Child and Adolescent Social Work Journal*, vol. 38, no. 2, pp. 175–180, 2021.