Article

Explainable Machine Learning in the Prediction of Depression

Christina Mimikou , Christos Kokkotis , Dimitrios Tsiptsios [*] , Konstantinos Tsamakis , Stella Savvidou , Lilian Modig , Foteini Christidi , Antonia Kaltsatou , Triantafyllos Doskas , Christoph Mueller , Aspasia Serdari , Kostas Anagnostopoulos , Gregory Tripsianis

Article

Explainable Machine Learning in the Prediction of Depression

Christina Mimikou [1], Christos Kokkotis [2], Dimitrios Tsiptsios [3,*], Konstantinos Tsamakis [4], Stella Savvidou [3], Lilian Modig[3], Foteini Christidi [5], Antonia Kaltsatou [6], Triantafyllos Doskas [7], Christoph Mueller [4], Aspasia Serdari [8], Kostas Anagnostopoulos [9] and Gregory Tripsianis [1]

1. Laboratory of Medical Statistics, School of Medicine, Democritus University of Thrace, 68100, Alexandroupolis, Greece
2. Department of Physical Education and Sport Science, Democritus University of Thrace, 69100 Komotini, Greece
3. 3rd Department of Neurology, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece
4. Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London SE5 8AB, UK
5. Department of Neurology, Democritus University of Thrace, 68100, Alexandroupolis, Greece
6. Physical Education and Sport Science, University of Thessaly, 42132, Trikala, Greece
7. Department of Neurology, Athens Naval Hospital, 11521 Athens, Greece
8. Department of Child and Adolescent Psychiatry, School of Medicine, Democritus University of Thrace, 68100, Alexandroupolis, Greece
9. Laboratory of Biochemistry, School of Medicine, Democritus University of Thrace, 68100, Alexandroupolis, Greece
* Correspondence: tsiptsios.dimitrios@yahoo.gr

Abstract: Background: Depression constitutes a major public health issue, being one of the leading causes of burden of disease worldwide. The risk of depression is determined by both genetic and environmental factors. While genetic factors cannot be altered, the identification of potentially reversible environmental factors is crucial in order to try and limit the prevalence of depression. **Aim:** A cross-sectional questionnaire-based study on a sample from the multicultural region of Thrace in northeast Greece was designed to assess the potential association of depression with several socio-demographic characteristics, lifestyle and health status. The study employed four machine learning (ML) methods to assess depression: Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, and Neural Networks (NNs). These models were compared to identify the best-performing approach. Additionally, a Genetic Algorithm (GA) was utilized for feature selection and SHAP (SHapley Additive exPlanations) for interpreting the contributions of each employed feature. **Results:** The XGBoost classifier demonstrated the highest performance on the test dataset to predict depression with excellent accuracy (97.83%), with NNs a second close (accuracy, 97.02%). The XGBoost classifier utilized the 15 most significant risk factors identified by the GA algorithm. Additionally, the SHAP analysis revealed that anxiety, education level, alcohol consumption and body mass index were the most influential predictors of depression. **Conclusions:** These findings provide valuable insights for the development of personalized public health interventions and clinical strategies, ultimately promoting improved mental well-being for individuals. Future research should expand datasets to enhance model accuracy, enabling early detection and personalized mental healthcare systems for better intervention.

Keywords: artificial intelligence; neural networks; logistic regression; support vector machine; XGBoost; interpretation

Introduction

Depression, a chronic mood disorder characterized by loss of interest and a persistent feeling of sadness [1], affects approximately 280 million people globally [2]. It is one of the leading causes of the global burden of disease [3], thus posing a challenging public health issue. Many studies have documented robust relationships between depression and hopelessness and subsequent suicidal thoughts and behaviors [4]. Apart from its debilitating impact on the sufferer, depression also affects their close environment, as caregivers of individuals with depression often endure emotional and physical challenges, increasing the risk of experiencing psychological issues themselves [5]. The pathogenesis of depression is associated with both genetic and environmental factors, with environmental features potentially having the greatest influence [6]. Due to the detrimental effects on people's health, early diagnosis of depression is essential.

Machine Learning (ML) is a powerful Artificial Intelligence (AI) tool used by researchers in the medical field, to predict, calculate and generate patterns for specific diagnoses. Over the past two decades, ML has been widely used to process statistical data to predict possible outcomes of complex biological systems [7]. The goal of ML is to detect underlying patterns within a sequence of observations by performing specific tasks to analyze data points collected by the physician's team, ultimately producing predictions or even enabling early diagnoses. ML is a combination of algorithms exploring how computer systems can learn rules from multiple examples without explicit programming [8]. ML is gaining prominence in the field of medicine, demonstrating impressive results in predicting survival and prognosis among patients [9]. ML algorithms can handle and analyse large datasets more efficiently than traditional methods, allowing for the extraction of meaningful insights and physical laws that might otherwise be missed [10]. Neural networks are vital components of ML algorithms which are modeled after the human brain. They function via pattern recognition, diagnosis, and prognosis in neurology. In a recent study, neural networks have been seen to achieve an 87% accuracy, suggesting that such models can effectively assist neurologists in diagnosing and understanding Multiple Sclerosis (MS) [11].

ML has not only been used in psychiatry but also in a vast number of specialties including surgery, nephrology and genomic medicine. In surgery, it has been used to analyse the surgeon's technical skill by detecting instrument motion, recognise patterns in video recordings, track eye movements, and cognitive function of the surgeon [12]. Another function for the use of ML, is the benefit to Chronic Kidney Disease (CKD). CKD is known to be a costly disease and thus, with the help of ML, physicians can proceed to reduce the costs and provide more care to a greater patient population. In primary care settings, these algorithms can help address the issue by triggering early nephrology referral and improving outcomes in kidney disease patients [13]. Another example is the use of these programs in the field of genomic medicine, where the scope of ML can sieve through complex genomic data to identify existing patterns associated with diseases such as cancer. Here, applying ML can help detect mutations in lesions or tumors. This integration allows for the identification of customized treatment recommendations, ultimately leading to enhanced patient outcomes [14].

Neurological disorders such as stroke, spinal cord injury and Parkinson's disease require accurate diagnosis and long-term neurorehabilitation, as they cause chronic disability. Diagnosis made by neuroimaging and physiological tools are important for accurately guiding the subsequent rehabilitation [15]. "Neuroscience and AI share a long history of collaboration" as Macpherson et al. [16] claim; AI and ML algorithms are able to sort through vast amounts of complicated data, such as neuroimaging sets while recognising specific patterns, valuable for prognosis and guidance in treatment [16]. Therefore, "these newer technologies can offer better rehabilitation outcomes and patient care through more personalized treatments based on (such) data" [15].

In regards to mental health disorders, there is currently no available FDA-approved AI application. However, considering the chronicity and the significant burden of psychiatric disorders, there is a significant need for the utilization of AI and ML algorithms, to assist especially in identifying individuals at risk [17]. Mental health illnesses can pose a challenge in terms of diagnosis

as their disease patterns are interchangeable and complex. In this case, AI and ML could potentially address the challenge through their capacity to analyze extensive patient data, "including medical records, genetic information and behavioral patterns" [19], thus enhancing diagnostic accuracy. Utilising AI in the field of mental health also has the potential to establish diagnoses more objectively and detect early stages of disease where signs are frequently overlooked [20].

Utilization of AI and ML algorithms, for depression specifically, provides meaningful insight into the disease, more effective drug regimens, and some predictive ability regarding patient outcomes [21]. Diagnosis of depression can be challenging, as it is highly heterogeneous, while it can also be underdiagnosed, since many individuals do not seek medical care to the perceived stigma [22]. In the case of depression, prevention is of utmost importance, even more so than the diagnosis on occasions, as preventative actions significantly limit prevalence [23]. AI and ML algorithms have the capacity to possibly predict the development of depression by simply identifying certain environmental factors which put an individual at greater risk [24].

AI and ML in the context of depression, could potentially be used to identify even minor signs, suggesting the presence of the disease based on behavioral and linguistic patterns. For instance, the patient's vocal tone and pattern could point the algorithm towards a direction ranging from major depressive disorder to mild anxiety. Additionally, AI algorithms show promise in the ability to analyse specific brain areas, such as the amygdala, anterior cingulate cortex and prefrontal cortex, that have been linked with anxiety and depression based on neuroimaging data [25]

The aim of our study is to explore the association between depression and certain environmental factors, such as demographic characteristics, socioeconomics, general health and habits using four machine learning methods. Identifying which factors show a positive association and which are protective, would allow for the creation of an algorithm that could predict and accurately diagnose depression, leading to earlier diagnosis and therefore prevention of worse outcomes, as well as adequate adaptation of therapy and treatment, thus limiting depression prevalence.

Materials and Methods

Study Sample and Research Design

The population of this cross-sectional study comprises 1227 participants, 657 females (53.5%) and 570 males (46.5%), with an average age of 49.94 ± 14.87 years old (ranging from 19 to 76; median age 50 years). The specimen selection was based on a system of stratified sampling of two stages on adult individuals (ages ≥18 years) living in the region of Thrace, the Northeastern prefecture of Greece, which is characterized by cultural diversity with various national, ethnolinguistic and religious groups; it was conducted from September 2016 until June 2022. The research design of this study is reported in Serdari et al. [26]. The overall response rate was 72.2%, which is fairly good for Greek standards (compared to 44.5% and 72% in the studies of Paparrigopoulos et al. [27] and Touloumi et al. [28], respectively). The sampling scheme ensured that the sample was randomly selected and representative of the general population of Thrace; specifically, 42.7% of the final sample were from urban areas and 57.3% from rural areas, while 65.8% were Greek Christians, 29.2% were Greek Muslims and 5.1% were Greek expatriates. The study excluded individuals under 18 years of age, pregnant women, night shift workers, people living in institutions for chronic illnesses, people living in retirement homes and correctional facilities due to their special characteristics in terms of habits and daily lives.

Ethics

All the procedures included in the study were carried out according to the ethics standards of the Democritus University Ethics Committee, who approved the realization of the study according to the standards of the Declaration of Helsinki (1964) and its subsequent amendments. Finally, all the participants in the study granted their consent.

Covariates

A structured questionnaire was used to collect: a) formal sociodemographic characteristics (gender, age, place of residence, education level, presence of child <6 years old, marital, cultural, financial and employment status), b) lifestyle and dietary habits (smoking, alcohol consumption, daily consumption of coffee, adherence to choice of Mediterranean diet [29], physical activity, midday sleep, duration of sleep), c) characteristics related to health (subjective general health status, body mass index [30], chronic disease morbidity, number of chronic diseases illnesses, anxiety [31], depression [32,33], family history of depression, traumatic events in the life of the participants, presence of insomnia or somnolence and sleep quality [34–36] (Appendix).

Assessment of Depression

Depression symptoms were assessed using the Greek version of the Beck Depression Inventory (BDI) [32,33], a widely used questionnaire that measures characteristic attitudes and symptoms of depression. It consists of 21 self-reporting Likert scale items, which are rated by respondents according to how each item applied to them during the past two weeks, using a 4-point scale ran ging from 0 (i.e. I do not feel sad) to 3 (i.e. I am so sad and unhappy that I can't stand it). Items are summed to create a total score, with higher scores indicating higher levels of depression. A total score of 13 was considered as a screening cut-off point for significant depression due to the high sensitivity [37].

Problem Definition

The participants were classified in a binary manner of "with depression" or "without depression". Almost thirty percent of the entire cohort (29%; 352 participants; Class 1) presented with depression disorders, while the rest of them had no depression disorders (29%; 352 participants; Class 0). The employed dataset consists of 27 variables at baseline with the target/dependent variable being the existence or non-existence of depression. Figure 1 presents the percentages of each class.
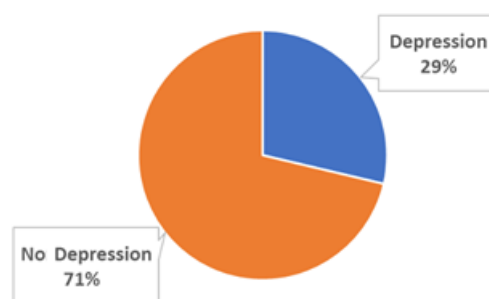

Figure 1. Grouping of the employed participants; *No depression: Class 0 (*n=875 participants) and *Depression: Class 1 (*n=352 participants).

Machine Learning Workflow

To handle missing data in the dataset, the mode imputation strategy was used, which involves replacing missing values with the most frequently occurring value in the dataset. The study employed the Genetic Algorithm (GA) as feature selection method to identify the optimal subset of features for improving the performance of the classifier. Four classifiers such as Logistic Regression (LR), Support Vector Machines (SVMs), XGBoost and Neural Networks (NNs) were used in the learning process, and a 70%/30% training/testing validation strategy was employed. Internal 10-fold cross-validation was used during the training phase to tune the hyperparameters after the undersampling step in the internal phase. The validation metrics included accuracy, recall, precision, f1-score, and specificity. The SHapley Additive exPlanations (SHAP) model assigns feature importance values using the concept of Shapley values from cooperative game theory and is a

powerful tool for understanding the decision-making process of a ML model. All code for the development, training, and evaluation of the ML models was written in Python, utilizing the Scikit-learn library (https://scikit-learn.org/, accessed on 30 March 2025) as the primary framework for implementing ML algorithms and techniques.

Statistical Analysis

Chi-squared analysis was used to evaluate whether the distribution of categorical variables, including subjects' demographic characteristics, lifestyle habits, and health-related factors, differs significantly between individuals with depression and those without. The analysis revealed significant associations, indicating that variations in these factors are linked to differences in the prevalence of depression.

Results

In this section, the epidemiological profile and depression prevalence among subjects, the description of the 15 most significant risk factors, the testing results of the ML classifiers that were trained using the aforementioned risk factors, and the interpretation of the best ML model output are presented.

Epidemiological Profile and Depression Prevalence among Subjects

The association of demographic characteristics with the prevalence of depression (Table 1) revealed that, while gender was not significantly associated with depression (p = 0.145), age, marital status, cultural status, place of residence, education level, unemployment, and financial status showed significant differences in depression prevalence (all p < 0.001). In particular, older individuals, divorced subjects, those residing in rural areas, and participants with lower education or poorer financial conditions were more likely to experience depression. The absence of a child under six years old also showed a significant association (p = 0.029) with a higher prevalence of depression.

Table 1. Prevalence of depression in relation to subjects' demographic characteristics.

	Number (%)	Depression		p value
		Frequency	Proportion (%)	
Gender				0.145
Males	570 (46.5)	152	26.7	
Females	657 (53.5)	200	30.4	
Age (years)				<0.001
≤40	341 (27.8)	42	12.3	
41 – 60	571 (46.5)	164	28.7	
>60	315 (25.7)	146	46.3	
Marital status				<0.001
Married	825 (67.2)	257	31.2	
Single	252 (20.5)	41	16.3	
Divorced	102 (8.3)	42	41.2	
Widowed	48 (3.9)	12	25.0	
Cultural status				<0.001
Greek Christians	807 (65.7)	194	24.0	
Greek Muslims	358 (29.2)	132	36.9	
Expatriated Greeks	62 (5.1)	26	41.9	

Place of residence				<0.001
Urban	524 (42.7)	88	16.8	
Rural	703 (57.3)	264	37.6	
Education level				<0.001
Low	406 (33.1)	211	52.0	
Medium	431 (35.1)	98	22.7	
High	390 (31.8)	43	11.0	
Presence of child <6 years				0.029
No	1128 (91.9)	333	29.5	
Yes	99 (8.1)	19	19.2	
Unemployment				<0.001
No	1121 (91.4)	303	27.0	
Yes	106 (8.6)	49	46.2	
Financial status				<0.001
Low	614 (50.0)	213	34.7	
Medium	258 (21.0)	33	12.8	
High	180 (14.7)	29	16.1	

The association of lifestyle habits with the prevalence of depression (Table 2) revealed that depression was statistically significantly associated with alcohol consumption, coffee consumption, physical activity, and sleep duration (all $p < 0.001$). Subjects consuming more than four cups of coffee daily or those reporting short sleep duration had substantially higher depression rates, whereas higher levels of physical activity and lower or moderate alcohol consumption were linked to lower depression prevalence. In contrast, smoking status ($p = 0.242$), adherence to the Mediterranean diet ($p = 0.080$), and midday sleep ($p = 0.101$) did not show any statistically significant association with depression.

Table 2. Prevalence of depression in relation to subjects' lifestyle habits.

	Number (%)	Depression		p value
		Frequency	Proportion (%)	
Smoking status				0.242
Never/ex-smoker	808 (65.9)	223	27.6	
Current smoker	419 (34.1)	129	30.8	
Alcohol consumption				<0.001
None	621 (50.6)	212	34.1	
1 – 3 glasses/week	316 (25.8)	69	21.8	
4 – 6 glasses/week	215 (17.5)	42	19.5	
>6 glasses/week	75 (6.1)	29	38.7	
Coffee consumption				<0.001
None	113 (9.2)	33	29.2	
1 – 2 cups/day	723 (58.9)	179	24.8	
3 – 4 cups/day	322 (26.2)	99	30.7	
> 4 cups/day	69 (5.6)	41	59.4	
Adherence to MED diet				0.080
Low	968 (78.9)	289	29.9	

	Number (%)	Frequency	Proportion (%)	p value
High	259 (21.1)	63	24.3	
Physical activity				<0.001
Low	1031 (84.0)	321	31.1	
High	196 (16.0)	31	15.8	
Midday sleep				0.101
No	520 (42.4)	162	31.2	
Yes	707 (57.6)	190	26.9	
Sleep duration				<0.001
Short	273 (22.2)	130	47.6	
Normal	780 (63.6)	176	22.6	
Long	174 (14.2)	46	26.4	

Health-related factors were strongly associated with the prevalence depression (Table 3). Individuals with poor subjective health, chronic illnesses (especially those with multiple conditions), a positive family history of depression, exposure to traumatic life events, and anxiety symptoms were significantly more likely to be depressed (all $p < 0.001$). Additionally, the presence of insomnia ($p = 0.042$) and poor sleep quality ($p = 0.008$) were associated with higher depression rates, while BMI status ($p = 0.103$) and excessive daytime sleepiness ($p = 0.704$) did not demonstrate any statistically significant association with depression.

Table 3. Prevalence of depression in relation to subjects' health related characteristics.

	Number (%)	Depression		p value
		Frequency	Proportion (%)	
BMI status				0.103
Normal	415 (33.8)	113	27.2	
Overweight	352 (28.7)	91	25.9	
Obese	460 (37.5)	148	32.2	
Subjective health status				<0.001
Good	941 (76.7)	168	17.9	
Bad	286 (23.3)	184	64.3	
Morbidity of chronic illness				<0.001
No	534 (43.5)	94	17.6	
Yes	693 (56.5)	258	37.2	
Number of chronic diseases				<0.001
None	534 (43.5)	94	17.6	
One	360 (29.3)	97	26.9	
Two	208 (17.0)	87	41.8	
More than two	125 (10.2)	74	59.2	
Family history of depression				<0.001
No	812 (66.2)	199	24.5	
Yes	415 (33.8)	153	36.9	
Traumatic events in life				<0.001
No	716 (58.4)	155	21.6	
Yes	511 (41.6)	197	38.6	
Anxiety symptoms				<0.001

No	813 (66.3)	119	14.6	
Yes	414 (33.7)	233	56.3	
Excessive daytime sleepiness				0.704
No	1120 (91.3)	323	28.8	
Yes	107 (8.7)	29	27.1	
Presence of insomnia				0.042
No	1015 (82.7)	279	27.5	
Yes	212 (17.3)	73	34.4	
Sleep quality				0.008
Good	765 (62.3)	199	26.0	
Bad	462 (37.7)	153	33.1	

Feature Selection

Table 4 shows the most significant 15 risk factors with the highest level of significance identified using a genetic algorithm as a feature selection technique for predicting depression in a binary classification problem.

Table 4. Ranking of most informative risk factors in depression diagnosis.

Risk Factor	Description	Type of variable
Gender	Gender (male/female)	Categorical
Marital status	Marital status (single/married/divorced/widowed)	Categorical
Residence	Area of residence (urban/rural)	Categorical
Education	Education level (low/medium/high)	Categorical
Unemployment	Unemployment (no/yes)	Categorical
Income	Income (low/medium/high)	Categorical
Chronic diseases	Chronic diseases (no/yes)	Categorical
BMI	Body mass index (normal/overweight/obese)	Categorical
Alcohol	Alcohol consumption/week (none/1-3 glasses/4-6 glasses/>6 glasses)	Categorical
Coffee	Coffee consumption/day (none/1-2 glasses/3-4 glasses/>4 glasses)	Categorical
Med diet	Adherence to Mediterranean diet (no/yes)	Categorical
Child <6 years	Presence of a child younger than 6 years of age (no/yes)	Categorical
Sleep duration	Sleep duration (short/normal/long)	Categorical
Sleepiness	Excessive daytime sleepiness (no/yes)	Categorical
Anxiety	Anxiety (no/yes)	Categorical

Testing Performance

Table 5 summarizes the testing performance metrics of comparative analysis between the employed ML classifiers in this binary task. The XGBoost classifier achieved the best testing performance scores with the 15 most significant risk factors as they were selected from the GA algorithm. Specifically, 97.83% accuracy, 97.85% f1-score, 97.94% precision, 97.83% sensitivity, and

97.44% specificity were achieved by XGBoost. On the other hand, the lowest performance metrics were achieved by the LR classifier. In particular, LR achieved 79.95% accuracy, 79.04% F1-score, 78.82% precision, 79.95% sensitivity, and 90.84% specificity.

Table 5. Metrics of testing performance for the employed classifiers.

Classifier	Accuracy (%)	f1-score (%)	Precision (%)	Sensitivity (Recall) (%)	Specificity (%)	Hyperparameters
LR	79.95	79.04	78.82	79.95	90.48	C: 1, penalty: l2
SVM	95.66	95.64	95.63	95.66	97.80	C: 10, kernel: rbf
XGBoost	**97.83**	**97.85**	**97.94**	**97.83**	**97.44**	**gamma: 0, max_depth: 7, min_child_weight: 1**
NN	97.02	97.03	97.06	97.02	97.44	activation: tanh, alpha: 0.0001, hidden_layer_sizes: (10, 20, 50), learning_rate: constant, solver: adam

Additionally, Figure 2 depicts the normalized confusion matrix and the receiver operating characteristics (0.98) for our best ML classifier. Specifically, the XGBoost classifier achieved 0.99 sensitivity and 0.97 specificity in this binary task.
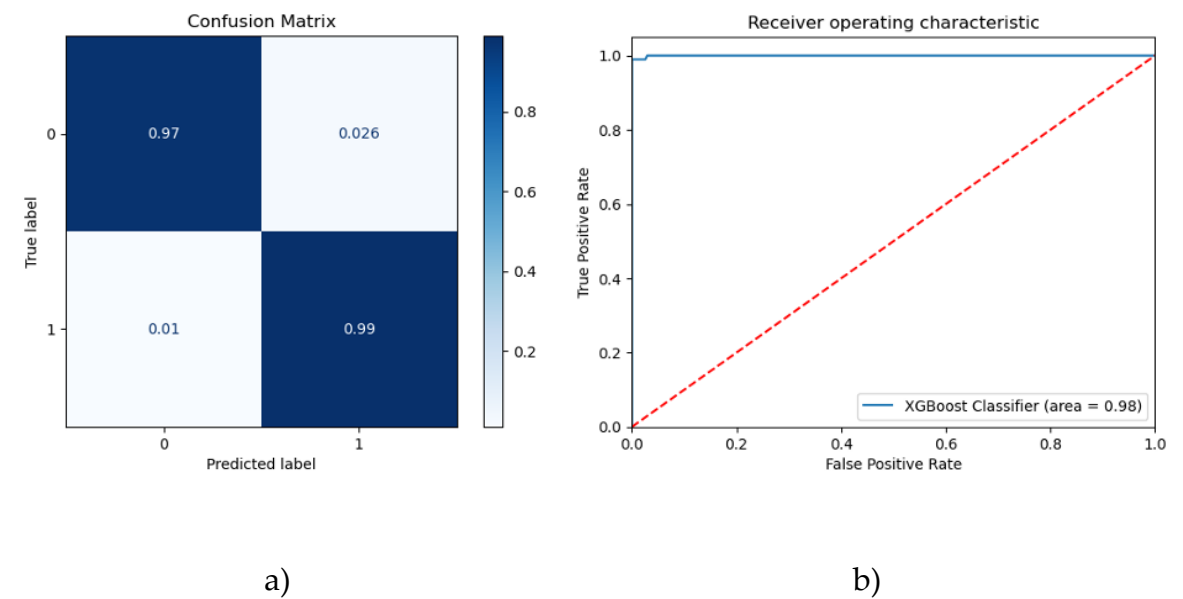

a)   b)

Figure 2. For the best ML classifier (XGBoost) a) confusion matrix and b) receiver operating characteristics are presented.

Explainability

In Figure 3 the effects of the 15 most significant risk factors on the output of the top-performing ML model (XGBoost) is illustrated. Figure 3a shows the mean absolute value of the SHAP values, which is an indicator of the SHAP global feature importance. Notably, the risk factor anxiety, education, alcohol, BMI, and coffee had the greatest impact on the prediction output and were considered the most important features. Figure 3b displays the effect of each feature on the output of the final model (XGBoost) applied to the depression dataset. The features are sorted based on the sum of their SHAP value magnitudes across all samples, SHAP values are based on game theory and assign an importance value to each feature in a model. Features with positive SHAP values positively impact the prediction, while those with negative values have a negative impact. The magnitude is a measure of how strong is the effect [38].

The color of each feature represents its value (blue for low, red for high). This analysis reveals that high levels of anxiety among the participants lead to an increase in their predicted depression status. Moreover, high consumption of coffee, chronic diseases, unemployment, med diet and sleepiness have a positive impact on the development of depression. On the contrary, higher education level, excessive drinking versus moderate drinking, higher BMI, females, high income, residence in the country and long sleep durations are negatively correlated with the existence of depression.
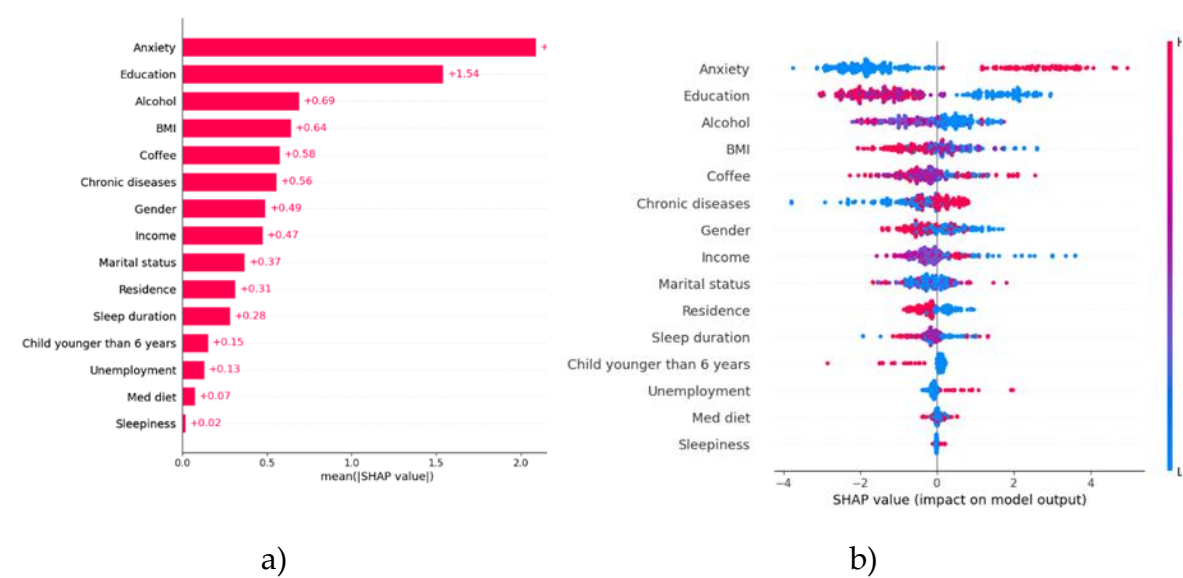

a)   b)

Figure 3. Risk factors on XGBoost ML classifier output for the diagnosis of depression. This figure presents (a) the SHAP feature importance and (b) the SHAP summary plot for the XGBoost trained on the risk factors selected by the GA.

Discussion

This study investigated the association between depression and multiple environmental factors, including sleep patterns, BMI, and diet. Data were collected through random phone number sampling, achieving a response rate of 72%. Participants completed a one-hour interview with healthcare professionals via phone call from their homes. The collected data were analyzed using multiple ML algorithms, including LR, NNs, SVMs, and XGBoost, with XGBoost demonstrating the highest reliability and accuracy. SHAP analysis identified several environmental factors with either positive or negative impacts on depression development. In this discussion, we compare our findings to previous studies to better understand the factors influencing the prevalence and diagnosis of depression.

The prevalence of depression in the present study was high (28.7%), aligning with Kokaliari [39], who reported a 22.5% prevalence of moderate to severe depression within the Greek population.

Similarly, Papadopoulos et al. [40] identified a high prevalence among individuals over 60 years of age living in rural Greece. Our study utilized the Greek version of the Beck Depression Inventory, which, while more effective as a screening tool than a diagnostic one, reliably identifies individuals at high risk or already experiencing depression [41].

Increased depression prevalence was observed among minority groups, with rates of 36.9% among Greek Muslims and 41.9% among Greek expatriates, compared to 24% among indigenous Greeks. This supports the hypothesis that minority status is associated with a higher risk of depression, consistent with findings by Bailey et al. [42], who identified exclusion, lower socioeconomic status, and limited access to psychiatric care as key factors. Furthermore, belonging to a minority group often reduces the likelihood of seeking mental health support [42], despite evidence that any form of social identity can confer protection against mental illness [43].

Higher income and financial stability were associated with a decreased risk of depression; however, consistent with previous studies, a U-shaped relationship was observed. Depression was more prevalent at very low and very high-income levels, while mid- to high-income levels were protective [44,45]. These findings echo those of Stylianidis and Souliotis [46], who reported a significant impact of unemployment and financial hardship on depression and suicidality during the Greek economic crisis.

Among all factors, educational attainment emerged as the strongest protective predictor against depression, supporting the findings of Biswas et al. [48]. Nevertheless, when coupled with unemployment, particularly during adolescence, the protective effect of education diminished. Unemployed adolescents with higher education levels showed increased anxiety and depression symptoms, driven by societal and familial pressures. Thus, the interplay between education and other socioeconomic factors should be considered when evaluating depression risk. Including vocational and skills-based courses in curricula could enhance future employment prospects [48].

Anxiety was the most significant risk factor for depression in our study, in line with existing research showing that approximately 85% of depression cases are comorbid with anxiety disorders [49–51]. Generalized Anxiety Disorder, in particular, frequently precedes depression [52]. Avoidant behaviors driven by anxiety can evolve into depression [53]. Treatments such as Cognitive Behavioral Therapy (CBT) and antidepressants benefit both conditions [49], and neuroimaging studies suggest shared brain alterations in emotion-processing circuits [54]. The STAR*D study further highlighted that comorbid anxiety-depression leads to more severe depressive episodes and increased suicide risk [55].

Interestingly, our findings diverged from the widely reported trend of higher depression rates among females, as we found a lower prevalence among women. Although epidemiological studies commonly show a 2:1 female-to-male ratio for major depression [56], differences in symptom presentation—internalizing symptoms in men versus externalizing in women [57]—and sensitivity to interpersonal versus extrinsic factors [58] could explain this discrepancy in our sample.

Contrary to expectations, heavy drinking was negatively associated with depression risk. Depression prevalence decreased with higher alcohol consumption and increased among moderate or non-drinkers. Although alcohol dependence has been linked to depression [59], some studies suggest moderate drinking may improve mood and cognitive function [60]. This complexity highlights the need for more nuanced evaluations.

Similarly, a higher BMI was negatively associated with depression risk in our study, whereas prior research, such as that by Kraus et al. [61], linked obesity with treatment-resistant depression and worse clinical outcomes. Badillo et al. [62] found obesity to be especially detrimental for men, largely mediated by poor sleep quality. Our findings align more closely with Cui et al. [63], who described a U-shaped relationship between BMI and mental health, suggesting that maintaining a healthy weight offers the best protection.

In terms of sleep, our findings revealed that both short and prolonged sleep durations were associated with depression, reflecting Zhai et al.'s meta-analysis [66]. Although some previous studies did not find a link between longer sleep duration and depression [64,65], our data, consistent

with Badillo et al. [62], suggest that sleep disturbances, potentially driven by inflammation, biochemical, or genetic mechanisms, play a key role in depression development.

Caffeine consumption also emerged as a risk factor for depression, likely through its negative effects on sleep and anxiety. However, Narita et al. [67] found that black coffee, without additives, might have protective effects due to lower inflammation and maintained brain-derived neurotrophic factor (BDNF) levels.

As expected, depression was more common among individuals with chronic diseases such as diabetes, arthritis, and asthma [68], consistent with Herrera et al. [69]. However, effective self-regulation and disease management appeared to mitigate the psychological burden for some patients.

In contrast to most studies [70–72], adherence to a Mediterranean diet (MD) was unexpectedly associated with a higher depression risk. Although traditionally protective, issues with low adherence or misreporting might explain this contradiction, as noted by Radkhah et al. [73]. Sánchez-Villegas et al. [70] demonstrated that while B vitamins showed a protective effect, Omega-3 fatty acids did not have a significant impact.

Living in rural areas was generally protective against depression, consistent with findings by Pérès et al. [74], who cited stronger social support during the COVID-19 lockdown. However, Nam et al. [75] identified farm workers as an exception due to unique occupational stressors.

In terms of model performance, XGBoost and Neural Networks (NNs) outperformed other ML models for predicting depression-associated factors. These findings align with those of Qasrawi et al. [81], who suggested that ML models can help healthcare professionals implement preventive interventions. XGBoost was particularly noted for its superior modeling capabilities over LR, SVM, and Decision Trees, as supported by Sharma and Verbeke [77] and Kessler et al. [78]. The consistent advantage of ML methods underlines the importance of using sophisticated algorithms, especially as the number of predictive factors increases. However, challenges remain. Richter et al. [80] noted inconsistencies in ML performance across different datasets and methods, suggesting a need for greater standardization.

Limitations

Despite the valuable insights gained from this study, several limitations must be acknowledged. First, the cross-sectional design prevents the establishment of causal relationships between environmental factors and depression. Second, self-reported data collected via phone interviews may introduce recall bias or social desirability bias, potentially affecting the accuracy of responses. Third, although random sampling was employed, selection bias cannot be fully excluded, particularly given the 28% non-response rate. Additionally, while the Greek Beck Depression Inventory is a validated screening tool, it is not a definitive diagnostic instrument, which may influence the estimated prevalence rates. Finally, although ML models such as XGBoost and NNs demonstrated strong predictive ability, model performance could vary with different datasets or demographic contexts, and external validation with independent samples is necessary to confirm generalizability.

Future Directions

Most predictive studies for depression to date have relied on small sample sizes, particularly when assessing treatment responses. Although small samples are useful for model development, larger datasets are essential for creating more powerful, generalizable models. As datasets grow, validation methods such as higher k-fold cross-validation will enable more robust model testing and better generalization. Moreover, feature reduction techniques will yield more meaningful results when applied to larger samples. Finally, selecting algorithms specifically designed for large datasets will enhance performance and predictive reliability.

Conclusions

In summary, depression is a pathological illness that can affect individuals of any age and gender. It is also more frequently observed in individuals with comorbid physical illnesses. ML approaches have shown significant promise in aiding the diagnosis of various mental health conditions, including schizophrenia, depression, bipolar disorder, autism spectrum disorders, and post-traumatic stress disorder. To detect such conditions, data derived from patients' social profiles, general clinical health status, and sensory mobile applications can be analyzed. In the present study, we examined contemporary research on the diagnosis of depression using ML-based approaches. Our aim was to provide information on the fundamental concepts of ML algorithms employed in mental health, particularly depression, and to explore their practical application. The results indicate that XGBoost outperforms traditional projection methods, demonstrating superior adaptability in predicting depression. Importantly, XGBoost's benefits extend beyond diagnosis, offering potential for predicting the future development of the disorder. A key advantage of this method is its applicability to individualized analysis.

Future studies could focus on expanding the dataset size to enhance training and validation processes, thereby improving the model's performance and reliability for clinical applications. As depression is a leading cause of impaired quality of life and remains challenging to predict, the application of advanced ML models like XGBoost offers a promising new direction in the therapeutic management of the disorder. The identified risk factors could contribute to the development of intelligent mental healthcare systems capable of detecting early signs of depressive symptoms, including within workplace environments.

References

1. Chand, S.P.; Arif. H. Depression. In *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.

2. Sousa, R.D.; Henriques, A.R.; De Almeida, J.C.; Canhão, H.; Rodrigues, A.M. Unraveling Depressive symptomatology and risk factors in a changing world. Int. J. Environ. Res. Public Health 2023, 20, 6575. https://doi.org/10.3390/ijerph20166575

3. Mathers, C.D.; Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Med. 2006, 3, e442, https://doi.org/10.1371/journal.pmed.0030442.

4. Ribeiro, J.D.; Huang, X.; Fox, K.R.; Franklin, J.C. Depression and hopelessness as risk factors for suicide ideation, attempts and death: meta-analysis of longitudinal studies. *Br. J. Psychiatry* **2018**, *212*, 279-286. doi:10.1192/bjp.2018.27

5. Sobieraj, M.; Williams, J.; Marley, J.; Ryan, P. The impact of depression on the physical health of family members. *Br. J. Gen. Pract*. **1998**, *48*, 1653–1655.

6. Nabeshima, T.; Kim, H. Involvement of Genetic and Environmental Factors in the Onset of Depression. *Exp. Neurobiol.* **2013**, *22*, 235–243. https://doi.org/10.5607/en.2013.22.4.235

7. Choi, R.Y.; Coyner, A.S.;Kalpathy-Cramer, J.; Chiang, M.F.; Campbell, J.P. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl. Vis. Sci. Technol.* **2020**, *9*, 14. doi:10.1167/tvst.9.2.14

8. Schlick, T.; Wie, G-W. Machine learning tools advance biophysics. *Biophys. J.* **2024**, 123, E1-E3. doi:10.1016/j.bpj.2024.07.036

9. Bi, Q.; Goodman, K.E.; Kaminsky, J.; Lessler, J. What is machine learning? A primer for the epidemiologist." *Am. J. Epidemiol.* **2019**, *188*, 2222-2239. https://doi.org/10.1093/aje/kwz189

10. Swanson, K.; Wu, E.; Zhang, A,; Alizadeh, A.A.; Zou, J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment." *Cell* **2023**, 186, 1772-1791. doi:10.1016/j.cell.2023.01.035

11. Ata, N.; Zahoor, I.; Hoda, N.; Adnan, S.M.; Vijayakumar, S.; Louis, F.; Poisson, L.; Rattan, R.; Kumar, N.; Cerghet, M.; Giri, S. Artificial neural network-based prediction of multiple sclerosis using blood-based metabolomics data. *Mult. Scler. Relat. Disord.* **2024**, *92*, 105942. https://doi.org/10.1016/j.msard.2024.105942

12. Egert, M.; Steward, J.E.; Sundaram, C.P. Machine Learning and Artificial Intelligence in Surgical Fields. *Indian J. Surg. Oncol.* **2020**, *11*, 573-577. doi:10.1007/s13193-020-01166-8

13. Singh, P.; Goyal, L.; Mallick, D.C.; Surani, S.R.; Kaushik, N.; Chandramohan, D.; Simhadri, P.K. Artificial Intelligence in Nephrology: Clinical Applications and Challenges. *Kidney Medicine* **2024**, 7, 100927. doi:10.1016/j.xkme.2024.100927

14. Chafai, N.; Bonizzi, L.; Botti, S.; Badaoui, B. Emerging applications of machine learning in genomic medicine and healthcare. *Crit. Rev. Clin. Lab. Sci.* **2024**, *61*, 140-163. doi:10.1080/10408363.2023.2259466

15. Calderone, A.; Latella, D.; Bonanno, M.; Quartarone, A.; Mojdehdehbaher, S.; Celesti, A.; Calabrò, R.S. Towards Transforming Neurorehabilitation: The Impact of Artificial intelligence on diagnosis and treatment of Neurological Disorders. *Biomedicines* **2024**, *12*, 2415. https://doi.org/10.3390/biomedicines12102415

16. Macpherson, T.; Churchland, A.; Sejnowski, T.; DiCarlo, J.; Kamitani, Y.; Takahashi, H.; Hikida, T. Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Networks*, **2021**, *144*, 603–613. https://doi.org/10.1016/j.neunet.2021.09.018

17. Lee, E.E.; Torous, J.; De Choudhury, M.; Depp, C.A.; Graham, S.A.; Kim, H.; Paulus, M.P.; Krystal, J.H.; Jeste, D.V. Artificial intelligence for Mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol. Psychiatry: Cogn. Neurosci. Neuroimaging* **2021**, *6*, 856–864. https://doi.org/10.1016/j.bpsc.2021.02.001

18. Wickelgren, I. Why do so many mental illnesses overlap? *Scientific American*. Available online: https://www.scientificamerican.com/article/why-do-so-many-mental-illnesses-overlap/ (accessed on 23 April 2025 Year)

19. Levkovich, I. Is artificial intelligence the next Co-Pilot for primary care in diagnosing and recommending treatments for depression? *Med. Sci.*, **2025**, *13*, 8. https://doi.org/10.3390/medsci13010008

20. Mansoor, M.A.; Ansari, K.H. Early Detection of Mental Health Crises through Artifical-Intelligence-Powered Social Media Analysis: A Prospective Observational Study. *J. Pers. Med.* **2024**, *14*, 958. doi:10.3390/jpm14090958

21. Park, Y.; Park, S.; Lee, M. Effectiveness of artificial intelligence in detecting and managing depressive disorders: Systematic review. *J. Affect. Disord.* **2024**, *361*, 445–456. https://doi.org/10.1016/j.jad.2024.06.035

22. Liu, X.; Jiang, K. Why is Diagnosing MDD Challenging? *Shanghai Arch Psychiatry* **2016**, *28*, 343-345. doi:10.11919/j.issn.1002-0829.216073

23. Cuijpers, P.; Beekman, A.T.; Reynolds 3rd, C.F. Preventing depression: a global priority. JAMA **2012**, *307*, 1033-1034. doi:10.1001/jama.2012.271

24. López Steinmetz, L.C.; Sison, M.; Zhumagambetov, R.; Godoy, J.C.; Haufe, S. Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine. *Front. Psychiatry* **2024**, *15*, 1376784. doi:10.3389/fpsyt.2024.1376784

25. Zafar, F.; Fakhare Alam, L.; Vivas, R.R.; Wang, J.; Whei, S.J.; Mehmood, S.; Sadeghzadegan, A.; Lakkimsetti, M.; Nazir, Z. The Role of Artificial Intelligence in Identifying Depression and Anxiety: A Comprehensive Literature Review." *Cureus* **2024**, *16*, e56472. doi:10.7759/cureus.56472

26. Serdari, A.; Manolis, A.; Tsiptsios, D.; Vorvolakos, T.; Terzoudi, A.; Nena, E.; Tsamakis, K.; Steiropoulos, P.; Tripsianis, G. Insight into the relationship between sleep characteristics and anxiety: a cross-sectional study in indigenous and minority populations in northeastern Greece. *Psychiatry Res*. **2020**, *292*, 113361. https://doi.org/10.1016/j.psychres.2020.113361

27. Paparrigopoulos, T.; Tzavara, C.; Theleritis, C. Insomnia and its correlates in a representative sample of the Greek population. *BMC Public Health* **2010**, *10*, 531. https://doi.org/10.1186/1471-2458-10-531.

28. Touloumi, G.; Karakatsani, A.; Karakosta, A.; et al. National Survey of Morbidity and Risk Factors (EMENO): protocol for a Health Examination Survey representative of the Adult Greek Population. *JMIR Res Protoc*. **2019**, *8*, e10997. https://doi.org/10.2196/10997

29. Panagiotakos, D.B.; Pitsavos, C.; Arvaniti, F.; Stefanadis, C. Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore. *Prev. Med.* **2006**, *44*, 335–340. https://doi.org/10.1016/j.ypmed.2006.12.009

30. World Health Organization, (2000). The world heart: health systems: improving performance.

31. Samakouri, M.; Bouhos, G.; Kadoglou, M.; Giantzelidou, A.; Tsolaki, K.; Livaditis, M. Standardization of the Greek version of Zung's Self-rating Anxiety Scale (SAS). *Psychiatriki* **2012**, *23*, 212–220

32. Beck A.T.; Ward C.H.; Mendelson, M.; Mock, J.; Erbaugh, J. An inventory for measuring depression. *Arch Gen Psychiatry* **1961**, *4*, 561-571.

33. Jemos, J. Beck Depression Inventory; validation in a Greek sample. Athens University Medical School, 1984.

34. Soldatos, C.R.; Dikeos, D.G.; Paparrigopoulos, T.J. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. *J. Psychosom. Res.* **2000**, *48*, 555–560. https://doi.org/10.1016/s0022-3999(00)00095-7

35. Tsara, V.E.; Serasli, E.V.; Amfilochiou, A.N., Constantinidis, Th.; Christaki, P. Greek version of the Epworth Sleepiness Scale. *Sleep Breath* **2004**, *8*, 91–95. https://doi.org/10.1007/s11325-004-0091-6

36. Kotronoulas, G.C.; Papadopoulou, C.N.; Papapetrou, A.N., Patiraki E. Psychometric evaluation and feasibility of the Greek Pittsburgh sleep quality index (GR-PSQI) in patients with cancer receiving chemotherapy. *Support Care Cancer* **2011**, *19*, 1831–1840. https:// doi:10.1007/s00520-010-1025-4

37. Lasa, L.; Ayuso-Mateos, J.L.; Vázquez-Barquero, J.L.; Díez-Manrique, F.J.; Dowrick, C.F. The use of the Beck Depression Inventory to screen for depression in the general population: a preliminary analysis. *J. Affect. Disord*. **2000**, *57*, 261–265.

38. Trevisan, V. Using SHAP values to explain how your machine learning model works. *Medium*. Available online: https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137 (accessed on 23 April 2025 Year)

39. Kokaliari, E. Quality of life, anxiety, depression, and stress among adults in Greece following the global financial crisis. *Int. Soc. Work* **2016**, *61*, 410–424. https://doi.org/10.1177/0020872816651701

40. Papadopoulos, F.C.; Petridou, E.; Argyropoulou, S.; Kontaxakis, V.; Dessypris, N.; Anastasiou, A.; Katsiardani, K.P.; Trichopoulos, D.; Lyketsos, C. Prevalence and correlates of depression in late life: a population-based study from a rural Greek town. *Int. J. Geriatr. Psychiatry* **2005**, *20*, 350–357. https://doi.org/10.1002/gps.1288

41. Edelstein, B.A.; Drozdick, L.W.; Ciliberti, C.M. Assessment of Depression and Bereavement in Older Adults. In *Handbook of Assessment in Clinical Gerontology*, 2nd ed.; Lichtenberg, P.A., Eds. 2003; pp. 3-43. www.sciencedirect.com/science/article/abs/pii/ B9780123749611100016.

42. Bailey, R.K.; Mokonogho, J.; Kumar, A. Racial and ethnic differences in depression: current perspectives. *Neuropsychiatr. Dis. Treat.* **2019**, *15*, 603-609. doi:10.2147/NDT.S128584

43. Brance, K.; Chatzimpyros, V.; Bentall, R.P. Increased social identification is linked with lower depressive and anxiety symptoms among ethnic minorities and migrants: A systematic review and meta-analysis. *Clin. Psychol. Rev*. **2023**, *99*, 102216. doi: 10.1016/j.cpr.2022.102216

44. Li, C.; Ning, G.; Wang, L.; Chen, F. More income, less depression? Revisiting the nonlinear and heterogeneous relationship between income and mental health. *Front. Psychol.* **2022**, *13*, 1016286. https://doi.org/10.3389/fpsyg.2022.1016286

45. Parra-Mujica, F.; Johnson, E.; Reed, H.; Cookson, R.; Johnson, M. Understanding the relationship between income and mental health among 16- to 24-year-olds: Analysis of 10 waves (2009-2020) of Understanding Society to enable modelling of income interventions. *PloS one* **2023**, *18*, e0279845. doi:10.1371/journal.pone.0279845

46. Stylianidis, S.; Souliotis K. The impact of the long-lasting socioeconomic crisis in Greece." *BJPsych Int.* **2019**, *16*, 16-18. doi:10.1192/bji.2017.31

47. Bjelland, I.; Krokstad, S.; Mykletun, A.; Dahl, A.A.; Tell, G.S.; Tambs, K. Does a higher educational level protect against anxiety and depression? The HUNT study. *Soc. Sci. Med.* **2008**, *66*, 1334-1345. doi:10.1016/j.socscimed.2007.12.019

48. Biswas, M.M., Das, K.C., Sheikh, I. Psychological implications of unemployment among higher educated migrant youth in Kolkata City, India. *Sci. Rep.* **2024**, *14*, 10171. https://doi.org/10.1038/s41598-024-60958-y

49. Wittchen, H.U.; Kessler, R.C.; Pfister, H.; Höfler, M.; Lieb, R. Why do people with anxiety disorders become depressed? A prospective-longitudinal community study. *Acta Psychiatr. Scand.* **2000**, *102*, 14-23. https://doi.org/10.1111/j.0065-1591.2000.acp29-03.x

50. Tiller, J.W.G. Depression and anxiety. *Med. J. Aust.* **2013**, *199*, S28-S31. https://doi.org/10.5694/mja12.10628

51. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 2021, *398*, 1700-1712. doi: 10.1016/S0140-6736(21)02143-7

52. Horn, P.J.; Wuyek, L.A. Anxiety disorders as a risk factor for subsequent depression. *Int. J. Psychiatry Clin. Pract.* **2010**, *14*, 244–247. doi.org/10.3109/13651501.2010.487979

53. Jacobson, N.C.; Newman, M.G. Avoidance mediates the relationship between anxiety and depression over a decade later. *J. Anxiety Disord.* **2014**, *28*, 437-45. doi:10.1016/j.janxdis.2014.03.007

54. McTeague, L.M.; Rosenberg, B.M.; Lopez, J.W.; Carreon, D.M.; Huemer, J.; Jiang, Y.; Chick, C.F.; Eickhoff, S.B.; Etkin, A. Identification of Common Neural Circuit Disruptions in Emotional Processing Across Psychiatric Disorders. *Am. J. Psychiatry* **2020**, *177*, 411–421. doi:10.1176/appi.ajp.2019.18111271

55. Fava, M.; Alpert, J.E.; Carmin, C.N.; Wisniewski, S.R.; Trivedi, M.H.; Biggs, M.M.; Shores-Wilson, K.; Morgan, D.; Schwartz, T.; Balasubramani, G.K.; Rush, A.J. Clinical correlates and symptom patterns of anxious depression among patients with major depressive disorder in STAR*D. *Psychol Med.* **2004**, *34*, 1299-1308. doi:10.1017/s0033291704002612

56. Salk, R.H.; Hyde, J.S.; Abramson, L.Y. Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychol. Bull.* **2017**, *143*, 783-822. doi:10.1037/bul0000102

57. Bartels, M.; Cacioppo, J.T.; van Beijsterveldt, T.C.; Boomsma, D.I. Exploring the association between well-being and psychopathology in adolescents. *Behav. Genet.* **2013**, *43*, 177-190. doi: 10.1007/s10519-013-9589-7

58. Kendler, K.S.; Gardner, C.O. Sex differences in the pathways to major depression: a study of opposite-sex twin pairs. *Am. J. Psychiatry* **2014**, *171*, 426-35. doi:10.1176/appi.ajp.2013.13101375

59. Kuria, M.W.; Ndetei, D.M.; Obot, I.S.; Khasakhala, L.I.; Bagaka, B.M.; Mbugua, M.N.; Kamau, J. The Association between Alcohol Dependence and Depression before and after Treatment for Alcohol Dependence. *ISRN psychiatry* **2012**, *2012*, 482802. doi:10.5402/2012/482802

60. Baum-Baicker, C. The Psychological Benefits of Moderate Alcohol Consumption: A Review of the Literature. *Drug Alcohol. Depend.* **1985**, *15*, 305–322, www.ncbi.nlm.nih.gov/pubmed/4053968

61. Kraus, C.; Kautzky, A.; Watzal, V.; Gramser, A.; Kadriu, B.; Deng, Z.; Bartova, L.; Zarate, C.A.; Lanzenberger, R.; Souery, D.; Montgomery, S.; Mendlewicz, J.; Zohar, J.; Fanelli, G.; Serretti, A.; Kasper, S. Body mass index and clinical outcomes in individuals with major depressive disorder: Findings from the GSRD European Multicenter Database. *J. Affect. Disord.* **2023**, *335*, 349–357. https://doi.org/10.1016/j.jad.2023.05.042

62. Badillo, N.; Khatib, M.; Kahar, P.; Khanna, D. Correlation Between Body Mass Index and Depression/Depression-Like Symptoms Among Different Genders and Races. *Cureus* **2022**, *14*, e21841. doi:10.7759/cureus.21841

63. Cui, H.; Xiong, Y.; Wang, C.; Ye, J.; Zhao, W. The relationship between BMI and depression: a cross-sectional study. *Front. Psychiatry* **2024**, *15*, 1410782. https://doi.org/10.3389/fpsyt.2024.1410782

64. Vorvolakos, T.; Leontidou, E.; Tsiptsios, D.; Mueller C.; Serdari A.; Terzoudi, A.; Nena, E.; Tsamakis, K.; Constantinidis, T.C.; Tripsianis G. The Association between Sleep Pathology and Depression: A Cross-Sectional Study among Adults in Greece. *Psychiatry Res.* 2020, 294, 113502. https://doi.org/10.1016/j.psychres.2020.113502

65. Gehrman, P.; Seelig, A.D.; Jacobson, I.G.; Boyko, E.J.; Hooper, T.I.; Gackstetter, G.D.; Ulmer, C.S.; Smith, T.C.; Millennium Cohort Study Team. Predeployment Sleep Duration and Insomnia Symptoms as Risk Factors for New-Onset Mental Health Disorders Following Military Deployment. *Sleep* **2013**, *36*, 1009–1018, https://doi.org/10.5665/sleep.2798

66. Zhai, L.; Zhang, L.; Zhang, D. Sleep duration and depression among adults: a meta-analysis of prospective studies. *Depression and Anxiety* **2015**, *32*, 664–670. https://doi.org/10.1002/da.22386

67. Narita, Z.; Hidese, S.; Kanehara, R. et al. Association of Sugary Drinks, Carbonated Beverages, Vegetable and Fruit Juices, Sweetened and Black Coffee, and Green Tea with Subsequent Depression: A Five-Year Cohort Study." *Clin. Nutr.* **2024**, *43*, 1395–1404. https://doi.org/10.1016/j.clnu.2024.04.017

68. Lotfaliany, M.; Bowe, S.J.; Kowal, P.; Orellana, L.; Berk, M.; Mohebbi, M. Depression and Chronic Diseases: Co-Occurrence and Communality of Risk Factors. *J. Affect. Disord.* 2018, 241, 461–468. https://doi.org/10.1016/j.jad.2018.08.011

69. Herrera, P.A.; Campos-Romero, S.; Szabo, W.; Martínez, P.; Guajardo, V.; Rojas, G. Understanding the Relationship between Depression and Chronic Diseases Such as Diabetes and Hypertension: A Grounded Theory Study. *Int. J. Environ. Res. Public Health* **2021**, 18, 12130. doi:10.3390/ijerph182212130

70. Sánchez-Villegas, A.; Henríquez, P.; Bes-Rastrollo, M.; Doreste, J. Mediterranean diet and depression. *Public Health Nutr.* **2006**, *9*, 1104–1109. https://doi.org/10.1017/s1368980007668578

71. Mamalaki, E.; Ntanasi, E.; Hatzimanolis, A.; Basta, M.; Kosmidis, M.H.; Dardiotis, E.; Hadjigeorgiou, G.M.; Sakka, P.; Scarmeas, N.; Yannakoulia, M. The Association of Adherence to the Mediterranean Diet with

Depression in Older Adults Longitudinally Taking into Account Cognitive Status: Results from the HELIAD Study. *Nutrients*, **2023**, *15*, 359. https://doi.org/10.3390/nu15020359

72. Yin, W.; Löf, M.; Chen, R.; Hultman, C.M.; Fang, F.; Sandin, S. Mediterranean Diet and Depression: A Population-Based Cohort Study. *Int. J. Behav. Nutr. Phys. Act.* **2021**, *18*, 153. https://doi.org/10.1186/s12966-021-01227-3.

73. Radkhah, N.; Rasouli, A.; Majnouni, A.; Eskandari, E.; Parastouei, K. The effect of Mediterranean diet instructions on depression, anxiety, stress, and anthropometric indices: A randomized, double-blind, controlled clinical trial. *Prev. Med. Rep.* **2023**, *36*, 102469. doi:10.1016/j.pmedr.2023.102469

74. Pérès, K.; Ouvrard, C.; Koleck, M.; Rascle, N.; Dartigues, J.F.; Bergua, V.; Amieva, H. Living in rural area: A protective factor for a negative experience of the lockdown and the COVID-19 crisis in the oldest old population?. *Int. J. Geriatr. Psychiatry* **2021**, *36*, 1950-1958. doi:10.1002/gps.5609

75. Nam, S.M.; Peterson, T.A.; Seo, K.Y.; Han, H.W.; Kang, J.I. Discovery of Depression-Associated Factors from a Nationwide Population-Based Survey: Epidemiological Study Using Machine Learning and Network Analysis. *J. Med. Internet Res.* **2021**, *23*, e27344. https://doi.org/10.2196/27344

76. Naveen, K.S.; Goel, A.; Dwivedi, S.; Hassan, M. Adding life to years: Role of gender and social and family engagement in geriatric depression in rural areas of Northern India. *J. Family Med. Prim. Care* **2020**, *9*, 721. https://doi.org/10.4103/jfmpc.jfmpc_1019_19

77. Sharma, A.; Verbeke, W.J.M.I. Improving Diagnosis of Depression with XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Front. Big Data* **2020**, *3*, 15. https://doi.org/10.3389/fdata.2020.00015

78. Kessler, R.C.; van Loo, H.; Wardenaar, K.J.; Bossarte, R.M.; Brenner, L.A.; Cai, T.; Ebert, D.D.; Hwang, I.; Li, J.; de Jonge, P.; Nierenberg, A.A.; Petukhova, M.V.; Rosellini, A.J.; Sampson, N.A.; Schoevers, R.A.; Wilcox, M.A.; Zaslavsky, A. M. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. psychiatry* **2016**, *21*, 1366-1371. doi:10.1038/mp.2015.198

79. Nahm, F.S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol.* **2022**, *75*, 25-36. doi:10.4097/kja.21209

80. Richter, T.; Fishbain, B.; Richter-Levin, G.; Okon-Singer, H. Machine Learning-Based Behavioral Diagnostic Tools for Depression: Advances, Challenges, and Future Directions. *J. Pers. Med.* **2021**, *11*, 957. doi:10.3390/jpm11100957

81. Qasrawi, R.; Vicuna Polo, S.P.; Abu Al-Halawa, D.; Hallaq, S.; Abdeen, Z. Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis. *JMIR Form Res*. **2022**, *6*, e32736. doi: 10.2196/32736