Underdiagnosed Depression in Older Adults: Analysis of the National Health Survey and Other Aggregate Factors

Eduardo H. S. Paraíso[a], Ariane C. B. da Silva[b] and Cristiane N. Nobre[c]
Institute of Exact Sciences and Informatics, Pontifical Catholic University of Minas Gerais, Dom José Gaspar, Belo Horizonte, Brazil

Keywords: Older Adults, Depression, Underdiagnosis, Machine Learning, Risk Factors.

Abstract: According to the World Health Organization, the total number of people living with depression worldwide is more than 300 million, with depressive disorders ranked globally as the third leading cause of disability. Among older adults, depression is the most common mental illness. This study addressed the cultural stigma surrounding depression in older adults and investigated factors contributing to underdiagnosis and undertreatment. We used data from older adults participating in the National Health Survey (NHS). We applied machine learning algorithms to predict the disorder (Random Forest, Support Vector Machine, Logistic Regression, Gradient Boost, XGBoost, Decision Tree, and Multilayer Neural Network), carefully interpreting the result obtained. Through the interpretability of ML models, the study identified risk factors associated with depression, and using silhouette index and attribute comparison, we found evidence of potential individuals who, although undiagnosed, may be suffering or about to suffer from depression, requiring appropriate care and treatment. This study represents a significant advance in mitigating the impact of cultural stigma on mental health diagnoses in the older population in Brazil.

1 INTRODUCTION

Among older adults, depression is the most common mental illness (Alexopoulos, 2005). Depressive disorders are characterized by sadness, loss of interest or pleasure, guilt or low self-esteem, sleep or appetite disturbances, tiredness, and lack of concentration, leading to suicide in its most severe form. Prevalence rates vary by age, peaking in older adulthood, affecting 5% of adults and 5.7% of adults older than 60 (WHO, 2023). In Brazil, it affects over 11 million people (5,8% of the population), having one of the highest rates of adulthood depression in all of Latin America (Errazuriz et al., 2023). Since the share of people aged 60 or over jumped from 11.3% to 14.7% of the population between 2012 and 2021 in this country, going from 22.3 million to 31.2 million (IBGE, 2021), that peak of prevalence is very worrying.

Several initiatives have been implemented to address depression worldwide (Rhee et al., 2018) (DeSouza et al., 2021) (Maștaleru et al., 2022) (Errazuriz et al., 2023), but the cultural stigma about depression that occurs in many cultures, such as the social disapproval, discrimination, and prejudice that people with depression face, underestimating symptoms and consequences of mental disorders, as well as the lack of integration between mental health services and primary health care in the public health system contribute to the underdiagnosis and undertreatment of depression (Scazufca et al., 2020).

Therefore, our study aims to analyze the population of older adults participating in the last NHS carried out in Brazil. To do this, we use the attribute Q092 (“Has a doctor or mental health professional (such as a psychiatrist or psychologist) ever diagnosed you with depression?”) to classify the database and separate it into two groups: older people diagnosed with depression and older people without a diagnosis of depression. Based on this classification, we applied machine learning techniques to train different classifier algorithms like Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Gradient Boost, Decision Tree (DT), and Multilayer Neural Network (MLP).

Thus, our objective was to select the models that best identifies depression in the study age group (60 years or more) and seek to identify the main risk fac-

[a] https://orcid.org/0009-0004-9236-8300
[b] https://orcid.org/0000-0003-2477-4433
[c] https://orcid.org/0000-0001-8517-9852

Paraíso, E., B. da Silva, A. and Nobre, C.
Underdiagnosed Depression in Older Adults: Analysis of the National Health Survey and Other Aggregate Factors.
DOI: 10.5220/0012365000003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 91-102
ISBN: 978-989-758-688-0; ISSN: 2184-4305

tors associated with this disorder, as well as identify potential individuals who, although not diagnosed, may be suffering or about to suffer from depression. To achieve this, we used the Silhouette Index (Si) and visualization to evaluate instances erroneously classified as diagnosed (false positives). We analyzed their proximity to diagnosed individuals to verify the association of these undiagnosed people with depressive disorder requiring special attention. We aim to obtain information about the determinants, conditions, and health needs of this target audience, allowing us to identify risk factors for the prevalence of depression in older adults, capable of assisting in developing public policies and achieving greater effectiveness in health interventions.

This article is organized following this structure: in Section 2, the theoretical foundation is presented, which brings the main concepts related to work. The works related to the theme are covered in Section 3. Section 4 presents the methodology used, with a detailed description of the database and the preprocessing steps. Section 5 gives the discussions regarding the results found. Finally, in Section 6, the final considerations are exposed.

2 BACKGROUND

2.1 Depression in Older Adults

The Diagnostic and Statistical Manual of Mental Disorders V (DSM-V) identifies a general set of depressive symptoms, such as depressed mood, loss of interest and pleasure, weight loss or gain, fatigue, insomnia or hypersomnia, psychomotor agitation or retardation, decreased concentration, thoughts of death or suicide, and worthlessness. However, according to Devita et al. (2022), some symptoms are more prominent in older adults, such as anhedonia, lack of appetite, and insomnia, whilst others are less frequent, such as depressed mood and agitation. In older adults, in addition to the symptoms mentioned, depression may also be related to difficulties inherent to the aging of the body, such as frailty (Maștaleru et al., 2022), increased need for help to carry out daily tasks, both inside and outside the home, cognitive impairment (Invernizzi et al., 2022), mobility, vision, and subjective memory, significantly increasing the risk of depression (Weyerer et al., 2013).

2.2 Machine Learning Algorithms

Machine Learning techniques have proven to be very useful in classification problems and identifying relevant characteristics to distinguish classes, with the interpretability of the methods used being one of the most significant issues, especially in the health area. A brief presentation of the machine learning techniques covered is presented below.

Logistic Regression (Cramer, 2004) is a linear model classifier used for binary and multiclass classification. The model is built from a logistic function that transforms the independent variable into a probability between 0 and 1. The logistic function is an “S” shaped curve approaching 0 and 1 at the extremes. The model is trained to find the coefficients that best fit the training data, minimizing the loss function.

A *Decision Tree* is a technique that uses a graph or model of decisions and their possible consequences, including the usefulness of each attribute in decision-making. Its sheets show the classification obtained. Thus, the decision tree creates nodes (database attributes) that are linked through a hierarchy, where the most valuable node is the root node, and the results presented are the leaf nodes (Salzberg, 1994).

Random Forest (Breiman, 2001) is considered an ensemble method as it builds a set of classifiers and presents the weighted vote of their predictions as a result. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector independently sampled and with the same distribution for all trees in the forest.

XGBoost (Chen et al., 2015) is a decision tree-based machine learning algorithm that uses a *Gradient Boosting* framework, applying the principle of boosting weak learners through gradient descent architecture (iterative optimization method).

Support Vector Machine (Vapnik, 1997) is an algorithm that uses a non-probabilistic binary classifier, applied to classification, regression, and outlier detection problems, whose objective is to find a hyperplane in an N-dimensional space (N – the number of features) that distinctly classifies the data points.

A *Multi-layer Neural Network*, also known as a Multi-layer Perceptron (MLP), is a type of artificial neural network (Kubat, 1999) that contains more than one layer of artificial neurons or nodes. It’s a supervised learning algorithm that learns a function by training on a dataset. MLP’s are capable of handling both linearly separable and non-linearly separable data.

2.3 Group Assessment Metrics - Silhouette Index

The Silhouette Index (Si) is a technique widely used to evaluate the quality of clusters. Although the clustering problem is unsupervised (without a classification attribute), the desired clustering result may be helpful in some scenarios.

The Silhouette Index (Si) is an evaluation measure that verifies the cohesion and separation of clusters and is based on the difference between the average distance of points belonging to the closest cluster to points in a group . For each point in the database, xi, the value of the Si, is calculated according to the Equation 1 (Rousseeuw, 1987).

$$S_i = \begin{cases} \dfrac{\mu_{out}^{min}(x_i) - \mu_{in}(x_i)}{max\{\mu_{out}^{min}, \mu_{in}(x_i)} \end{cases} \quad (1)$$

Where $\mu_{in}(x_i)$ is the average distance from x_i to the instances in its cluster, and μ_{out}^{min} is the average distance from x_i to instances in the nearest cluster.

Thus, this calculation is made for each instance in the database. The S_i value of an instance lies in the range $[-1,1]$. A value close to 1 indicates that x_i is more relative to points in its cluster and distant from neighboring clusters. A value close to 0 indicates that x_i is close to the border of two clusters. Finally, a value near -1 indicates that x_i is close to points that belong to another cluster.

In this work, we used the silhouette metric to check how close the instances erroneously classified as diagnosed are to the truly diagnosed ones.

3 RELATED WORKS

For some time now, researchers have been conducting studies to understand better the symptoms and risk factors associated with depression and the most effective approaches to alleviating its effects. In this scenario, machine learning techniques have shown promise.

The study of Kim et al. (2019) aimed to develop a machine learning algorithm to predict the classification of depression clusters among older people living alone. The authors focused on utilizing depression-related data from 47 older adults collected through research and data from wearable devices (Actiwatch). Through the proposed approach, the authors presented a method to identify underdiagnosed subgroups and monitor daily progression toward treatment or therapeutic intervention in the community setting.

Xu et al. (2019) developed a deep learning model for individualized prediction of the onset of depressive disorder among older adults in the United States. The model can provide information about disease progression and suggest timely targeted interventions. It also possible to extract information from longitudinal bases and temporal patterns of risk factors associated with the onset of depressive disorder. It can predict whether the person will become depressed in two years, but it depends on longitudinal studies rarely conducted in low- and middle-income countries.

The work of Zhang et al. (2021) is similar to ours. The authors sought to establish a risk factor-based depression assessment model using machine learning. Five ML models were used to assess depression among middle-aged and older adults in a American longitudinal database. The authors were able to have the learning models identify the occurrence of depression in the data set through social demographics, lifestyle, laboratory data, and other data from middle-aged and older people. However, In addition to identifying risk factors, we intend to identify individuals who are already affected by the disorder but without a confirmed diagnosis.

DeSouza et al. (2021) advocated using Natural Language Processing (NLP) in evaluating acoustic and linguistic aspects of human language derived from text and speech, integrated with machine learning approaches to classify depression and its severity. The goal was to help combat underdiagnosis or misdiagnosis due to subjective reporting of symptoms and the distinct cognitive, psychomotor, and somatic characteristics of depressive disorder.

Devita et al. (2022) performed a systematic review (LSR) analyzing elderly-population-depression and its primary clinical characterizations to propose guidelines from "everyday" clinical activity around the world. Their findings are significant, but because it does not focus on specific regions and due to the large amount of work carried out in higher-income countries, their conclusions could be different in low- and middle-income countries.

Errazuriz et al. (2023), in turn, performed a systematic review (LSR) in Latin America countries. They analyzed adult-population-based studies that reported primary data on the prevalence of depressive disorder in Latin America from 1990 to 2023. They found a high prevalence of depressive disorders in Latin America and its association with inequality, development, and crime indicators. However, they analyzed the adult population yet to be affected by the cognitive and physical limitations that can impact the lifestyle and independence of older people.

Although it is now possible to find studies investigating depression in older people, they are still very scarce in low- and middle-income countries, especially studies regarding underdiagnosis or incorrect diagnosis. With our study, we intend to carry out a more careful analysis of a nationwide public database, identifying individuals who could be at risk of developing depression or even already suffering from this disorder.

4 MATERIALS AND METHODS

4.1 Database

The NHS is a household-base survey with national scope, carried out by the Ministry of Health in partnership with the Brazilian Institute of Geography and Statistics in 2013 and 2019. This study used the NHS 2019[1], which collected information in 108 thousand households about the performance of the national health system, as well as the population's health conditions, surveillance of chronic non-communicable diseases, and associated risk factors. Table 1 shows the NHS 2019 structure.

Table 1: NHS 2019 questionnaire structure.

Questionnaire	Description
Identification and Control	General household information
Household	Detailed Household Information
	Home visits by the Health Team
	and Endemic Disease Agents
All Residents Questionnaire	General characteristics of residents
	Features of education of residents
	Work characteristics (14 and over)
	Income from other sources
	Disabilities (age 2 and over)
	Health Plan Coverage
	Use of health services
	Health of individuals (60 and over
	Children under two years
Selected Resident Questionnaire	Features of work and social support
(People aged 15 or over)	Perception of health status
	Accidents
	Lifestyles
	Chronic diseases
	Women's Health (age 15 and over)
	Prenatal Care (age 15 and over)
	Oral Health
	Paternity, partner prenatal care
	Violence (age 18 and over)
	Communicable diseases
	Sexual activity (age 18 and over)
	Medical care (age 18 and over)
	Anthropometry

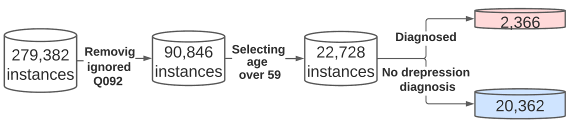

Figure 1: Instance selection on NHS database.

The original database has 279382 instances and 1087 attributes organized into 29 modules. Figure 1 outlines the selection of instances carried out in the database. Since this work aims to analyze older people who have ever been diagnosed with depression, the attribute Q092, *"Has a doctor or mental health professional (such as a psychiatrist or psychologist) ever diagnosed you with depression?"* became the classification attribute. This way, all instances that had this null attribute were eliminated. We selected only individuals aged over 59 years remaining 22,728 instances, with 20,362 that were never diagnosed with depression and 2,366 that had already been diagnosed, as represented in Figure 1.

Among the 1087 attributes in NHS 2019, a total of 100 were considered most relevant in this work context. We selected attributes related to risk factors known in the literature and the others were selected in descending order of their information gain. Table 2 shows some selected attributes and their relationship with depression in older adults.

4.2 Methodology

Figure 2 depicts the schematic representation of the adopted methodology. The methodology consisted of five general stages: 0) Data preparation; 1) Training of the ML algorithms; 2) Test of the trained classifiers; 3) Evaluation of the obtained results; and 4) Interpreting the obtained ML models. To execute all these stages, we employed the Python v3.10 libraries: Pandas v.1.5.3, Scikit-learn v.1.2.2, Matplotlib v.3.7,1, Scipy v.1.11.3, and Numpy v.1.23.5, prevalent within the data science and machine learning community.

During Stage 0, we executed a series of preprocessing steps to mitigate noise within the database. These steps encompassed the handling of duplicates, outliers, and null values. The procedure adhered to the following sequential process:

1. We investigated the dataset and removed attributes that showed incomplete information.
2. Outliers were identified and removed using the interquartile range technique, excluding observations outside this range.
3. We segmented the database based on its distinct labels for duplicate removal and eliminated duplicates among the instances.
4. Due to the significant imbalance towards the minority class, the majority class underwent undersampling using the k-means method for training and cross-validation (Kohavi et al., 1995). The number of clusters was determined using the elbow method, utilizing the metric of Within-Cluster-Sum-of-Squares (WCSS). We took random samples from each defined cluster until a balance with the minority class was achieved.
5. Data were normalized by standardizing the features by removing the mean and scaling to unit

[1] NHS 2019 database can be downloaded at https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html?caminho=PNS/2019/Microdados/Dados

Table 2: Some selected attributes and their relation with older adults depression.

Feature	Justificative	Attributes
	Characteristics of the individuals	
Region	Understanding the regional context is crucial for implementing appropriate intervention and support strategies (Gonyea et al., 2017).	V0001
Age	Neurobiological changes, chronic illnesses, the loss of loved ones, and social isolation heighten the vulnerability (Diniz et al., 2014).	C008
Race	Social disparities and access to resources can exacerbate the situation (Foong et al., 2021).	C009
Gender	Older women exhibit a significantly higher prevalence of depression than men in the same age group (Lytle et al., 2018).	C006
Sexual orientation	Older LGBTQ+ persons are more likely to face social isolation and contend with prejudice, factors that can contribute to depression (Lytle et al., 2018).	Y008
Marital status	Individuals who live alone, especially those who are widowed, are more predisposed to developing the condition (Pan et al., 2022).	C011, C01001
Body Mass Index (BMI)	A significant BMI imbalance may be associated with depressive symptoms due to malnutrition, physical frailty, or obesity (Chang and Yen, 2012).	P00104, P00404
	Socioeconomic Conditions	
Income	Income plays a crucial role in the context of depression, directly impacting access to mental health resources (Gonyea et al., 2017).	VDF003
Health insur.	Health insurance can represent access to mental health services, specialized therapies and treatments (Li et al., 2023).	I00101, I00102
Job	Abrupt retirement or professional devaluation can lead to a loss of purpose, social isolation, and stress (Fernández-Niño et al., 2018).	E017, E019 VDE001, M05010
Education	Education plays a crucial role in understanding depression, directly influencing the ability to access mental health (Belo et al., 2020).	D001, VDD004A
	Lifestyle	
Food intake	Nutrients play a direct role in brain function and the regulation of neurotransmitters. Balanced diets provide the essential elements for mental health, while nutritional deficiencies can increase the risk and severity of depression (Rogozinski and Zisberg, 2020).	P006, P013, P015 P018, N015, P1101 P02501, P02602
Alcohol intake	Excessive alcohol can exacerbate feelings of sadness, hopelessness and social isolation. (Quittschalle et al., 2021).	P02801
Tobacco	The nicotine in tobacco can affect the chemical balance in the brain, exacerbating depressive symptoms (Quittschalle et al., 2021).	P050
Sleep problem	Sleep deprivation impairs emotional regulation, hampering the ability to cope with stress and negative emotions (Pigeon et al., 2008).	N010, N011
Focus problem	Difficulty in maintaining focus and attention can negatively impact daily functionality and quality of life (Devita et al., 2022).	N012N013
Daily energy	Persistent fatigue and lack of motivation, are significant indicators of depression. (Lee and Holtzer, 2021).	N015
Depressive feelings	Depressive feelings are vital indicators in characterizing depression, as they reflect the subjective experience of the condition (Devita et al., 2022).	N016, N017, N018
Screen time	Excessive screen time can indicate association with social isolation, sedentary behavior, and sleep disturbances (Wang et al., 2022).	P04501, P04502
	Health Conditions	
Impairments	Physical or cognitive impairments can be a significant factor in characterizing depression, as they can increase emotional and social vulnerability, limit participation in activities, and cause frustration and isolation—all (Devita et al., 2022).	G033, G048 G059, G072
Mental health	Mental health directly impacts emotional well-being, the ability to face challenges and quality of life. (Devita et al., 2022).	Q11006, Q092
Polypharmacy	Using multiple medications can indicate physical comorbidities, drug interactions, and complexity in clinical management (Palapinyo et al., 2021).	K04302
Need for help	Difficulties in performing simple tasks (dressing, eating, personal hygiene) significantly impacting quality of life (Devita et al., 2022).	K01901, K03401
Chronic diseases	The physical and emotional burden of these conditions can increase the risk of developing depressive symptoms due to their impact on quality of life, functional limitations, and the need to adapt to new lifestyles, leading to stress and anxiety (Papageorgiou et al., 2022).	Q00201, Q03001 Q060, Q06306
	Violence	
Insults or threats	They can lead to increased stress, helplessness, and social isolation. It is resulting in a decline in mental health and emotional well-being. (Mendes et al., 2021)	V00201, V00202 V00203
Intimidation or aggression	These experiences can cause psychological trauma, leading to helplessness, social isolation, and low self-esteem. (Mendes et al., 2021)	V00204, V00205 V01402, V01403 V01404, V01405
Sexual assault	A traumatic experience can trigger helplessness, shame, and guilt. This emotional burden can persist over time, contributing to the development or worsening of depression (Mendes et al., 2021).	V02701, V02702 V02801, V02802

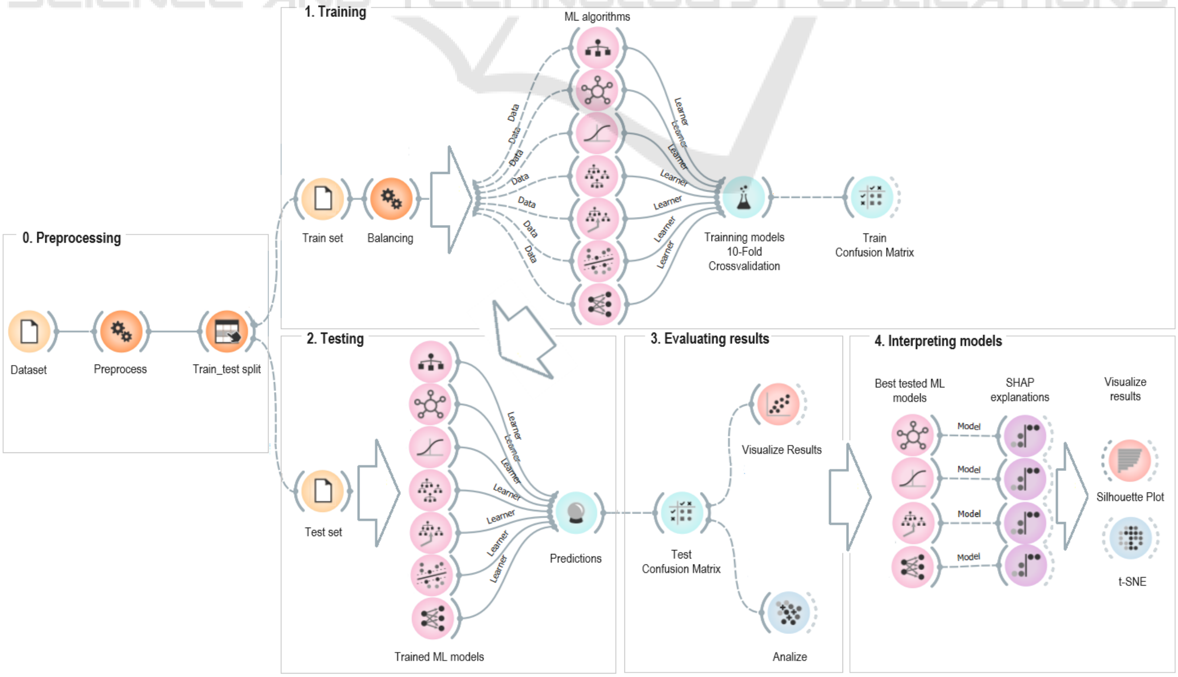

Figure 2: The proceedings of preparation, training, testing, and evaluation of the results.

variance. Each value is reduced from the mean of the samples of that attribute, and the result is divided by the standard deviation of the attribute.

Table 3: Number of instances per class for training/validation and testing.

Class	Original set	Filtered by age	Unbalan. train set	Balanced train set	Test set
Diag.	8,332	2,366	2,011	2,011	355
No diag.	82,512	20,362	17,307	2,011	3,055
Total	**90,844**	**22,728**	**19,318**	**4,022**	**3,410**

At the end of the preprocessing stage, the dataset was divided in an 85% to 15% ratio for training and testing, respectively. The division was stratified based on the labels in the central database, guaranteeing the exact original proportionality between classes.

In stage 1, we trained seven distinct classification algorithms using their default configurations. We performed an initial 10-fold cross-validation, ensuring data stratification. We applied the previously described balancing technique during each fold of cross-validation. The resulting metric averages for each classification algorithm allowed us to compare how each model would perform with the proposed balancing method. In the next stage, stage 2, we balanced the training set and used the resulting set for all models and the previously segmented test set.

After training and testing the models, in stage 3, we identify the four most effective ones based on recall and precision metrics using visualization and data analysis techniques.

Finally, in stage 4, we selected the intersection of these four models for the instances correctly classified, both positive and negative, including the cases of false positives and false negatives. We calculated the Silhouette Score for each instance in the test dataset to assess the clustering quality. Then, we examined the scores of correctly classified instances, false positives, and false negatives concerning this metric. This analysis allowed us to obtain an average silhouette value for the new clusters.

4.3 ML Algorithms

As for the ML algorithms, after the preprocessing stage, the training set was used for the training of seven classification algorithms: **DT** (criterion: entropy, max_depth: 4, max_features: None, min_samples_leaf: 19, min_samples_split: 14); **LR** (C: 3.937, max_iter: 100, penalty: l1, solver: saga); **RF** (bootstrap: False, max_depth: 25, max_features: log2, min_samples_leaf: 1, min_samples_split: 30, n_estimators: 252); **XGBoost** (colsample_bytree: 0.540, gamma: 0, learning_rate: 0.04, max_depth: 8, min_child_weight: 3, n_estimators: 419, reg_alpha: 0.01, reg_lambda: 1, subsample: 0.631); **GradientBoost** (learning_rate: 0.1, max_depth: 17, n_estimators: 498, subsample: 0.9); **SVM** (C: 0.502, class_weight: balanced, degree: 2, gamma: auto, kernel: rbf); **MLP** (activation: relu, alpha: 0.4, hidden_layer_sizes: 199, learning_rate: invscaling, max_iter: 468, solver: sgd).

All of them were built using the *Scikit-learn* library package version 1.0.2 (Pedregosa et al., 2011). The experiments were performed on Windows 11 operating system using an Intel(R) Core(TM) i7 processor, 2.60GHz, 32GB of RAM and the Jupyter v2022.1.3 tool.

4.4 Model Quality Assessment Metrics

Precision[2], Recall[3], and F-measure[4] metrics were used to assess the quality of the ML models. Precision is the rate of instances correctly classified as belonging to the class in question out of all those classified in the class. *Recall* refers to the percentage of class instances that were correctly predicted to belong to the class. The F-measure is a harmonic mean between Precision and *Recall*. The training of the ML models was carried out through a stratified 10-fold cross-validation method (Kohavi et al., 1995), in which the cross-validation procedure is repeated ten times and the mean value represents the test result.

5 RESULTS AND DISCUSSIONS

After training the ML algorithms, they were submitted to test set prediction with instances not yet seen. Table 4 presents the Recall, Precision, F1-score, and AUC metrics obtained in this phase.

All ML algorithms achieved more than 70 in the Recall metric, indicating that they were relatively efficient in perceiving diagnosed individuals' profiles based on the demographic, socioeconomic, lifestyle, and health condition attributes provided as input. Concerning older adults depression, Recall, also known as sensibility or detection rate, is one of the most critical metrics. Not recognizing a sick individual and letting the disease worsen without any treatment is more harmful and has a more significant impact on the chances of alleviating the disorder than eventually identifying a healthy individual as sick, as

[2] $Precision = \frac{VP}{VP+FP}$

[3] $Recall = \frac{VP}{VP+FN}$

[4] $F-Measure = \frac{2 \times Recall \times Precision}{Recall+Precision}$

Table 4: Performance of the ML algorithms (in percentage).

Algorithm	Precision	Recall	F1	Class
DT	96	74	83	No diagnosis
AUC 0.74	24	74	37	Diagnosed
RF	96	73	83	No diagnosis
AUC 0.83	24	75	37	Diagnosed
GradientBoost	96	75	85	No diagnosis
AUC 0.83	27	76	39	Diagnosed
XGBoost	97	76	85	No diagnosis
AUC 0.77	28	78	41	Diagnosed
SVM	96	75	84	No diagnosis
AUC 0.75	26	74	38	Diagnosed
LR	96	77	85	No diagnosis
AUC 0.81	26	70	38	Diagnosed
MLP	96	76	85	No diagnosis
AUC 0.81	27	75	39	Diagnosed

with complementary exams overdiagnosis would become evident.

Although, in this context, a more sensitive model with a greater probability of detecting an individual with a depressive disorder is desired, the low Precision values draw attention. Due to the high complexity of the database and the attempt to infer a diagnosis of depression based on risk factors related to the disorder and not the patients' typical symptoms, the precise separation between older adults with and without a diagnosis does not seem so trivial. Figure 3 shows the base instances with dimensionality reduced to 2 dimensions using non-linear dimensionality reduction technique, *t-distributed stochastic neighbor embedding* (t-SNE) (van der Maaten and Hinton, 2008), for visualization[5].

We performed partial visualizations to verify the spatial proximity between the predicted instances in the test set, considering categories of related attributes in each image. Figures 4 and 5 show the categories and the positioning of the instances according to each one. Some categories better separate instances, such as the attributes of the health and chronic disease perception module in Figure 4, while in other modules, we could not perceive the groupings of each type, as in the finance and education modules in Figure 5.

Associated with the high complexity of the database, through the attribute selected for the label ("Has a doctor ever given you a diagnosis of depression?"), it is impossible to say with absolute certainty that people without a diagnosis are not suffering from this disorder. Our objective is precisely to analyze the instances mistakenly considered positive by the learning algorithms but which officially do not have a diagnosis.

First, we selected the correctly classified instances and those that all the trained models wrongly classified during the testing stage. This intersection produced a subset of instances with four types (positive correctly classified, negative correctly classified, false negative, and false positive). From this new reclassified set, we built the following visualizations. Figure 6 represents the amounts of each type.

Furthermore, we also investigated the consistency of the test set classification predicted by the ML algorithms for the two groups of instances (diagnosed and undiagnosed). To do this, we calculated the Si of each instance of the test set. We evaluated whether the correctly classified ones had positive values, indicating that they belonged to their classification group, or negative values, suggesting they were closer to the neighboring group. Figure 7 shows parts of the complete graph, which contains each value of the Si for each instance predicted by the ML algorithms.

In Figure 7, each bar represents an instance in the dataset with its respective silhouette index value. For undiagnosed instances correctly predicted by all trained ML algorithms, the value of Si, although very close to 0, is positive, indicating that, despite such instances being very close to the border with the neighboring (diagnosed) group, they are more relative to the group they belong to. The same does not occur with diagnostic instances correctly predicted by ML algorithms. Some instances have positive Si values, while others have negative Si values very close to 0. For this group, the average Si was -0.017, indicating they are on the border between the two groups (with and without diagnosis) and tending very slightly into the non-diagnosis class. However, they were correctly identified by the trained ML algorithms.

A more careful investigation of this group of diagnosed-class instances shows that those with higher mean values in health perception attributes have positive silhouette values, indicating that they are in the correct group. In contrast, those with negative silhouette values, although they belong to the diagnosed class, have much lower average values than their group mates, so they tend to get closer to the border of the opposite class (no diagnosis). Figure 8 presents the heat map with the average values of all instances correctly classified as "diagnosed" but with opposite silhouette index values.

The third group of instances, mistakenly identified by the algorithms as diagnosed (false positives), have negative Si values. The average Si value for this group reached -0.084. Again, a value close to the border. Here, there is a slightly stronger tendency that these instances should be in another group.

Delving deeper into this group of individuals, we can compare some of their critical attributes with their

[5]Conversion performed using the *TSNE* package from the *l*ibrary sklearn.manifold language *Python v3*

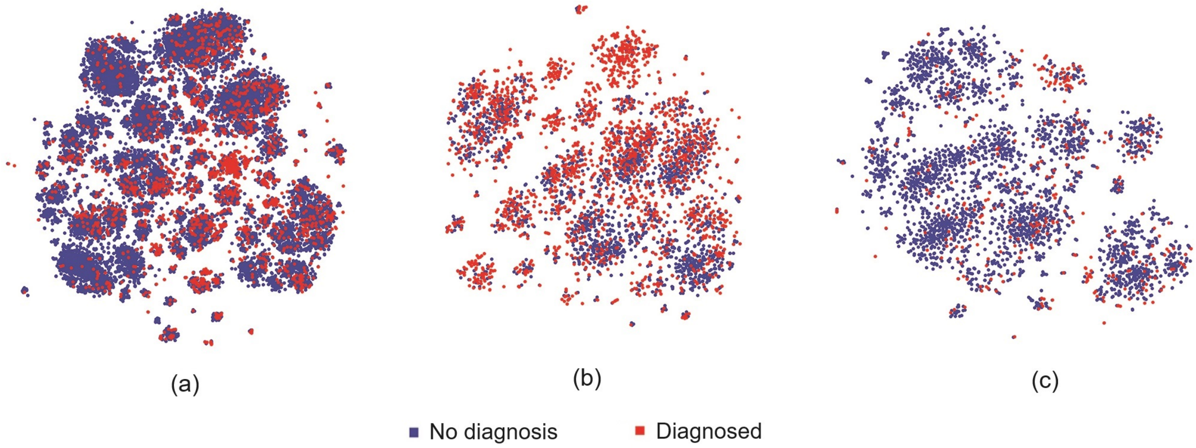

Figure 3: Spatial representation of training and test sets. (a) Unbalanced train set. (b) Balanced train set. (c) Test set.

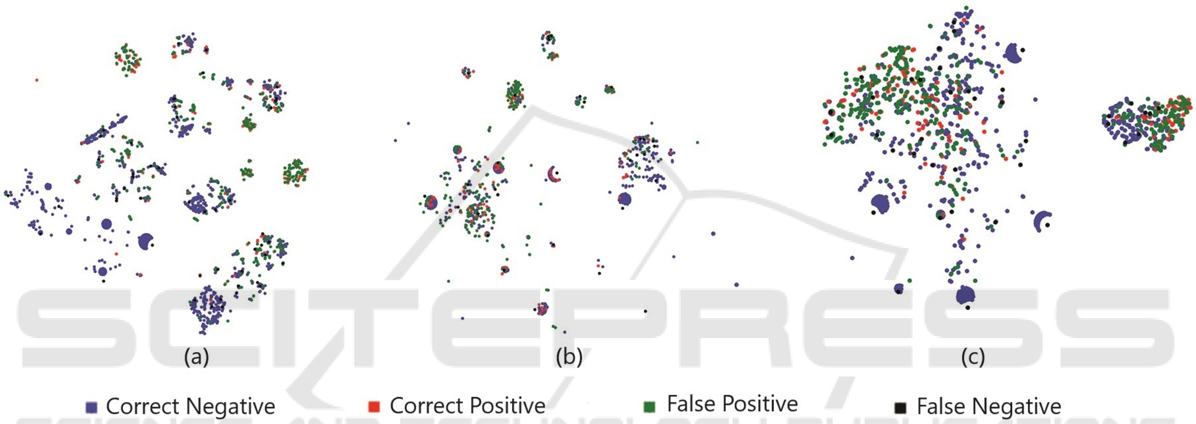

Figure 4: Database attributes show the proximity between correctly identified diagnosed instances and false positives. (a) Relating to chronic diseases module. (b) Relating to impairments (c) Relating to health status perception module.

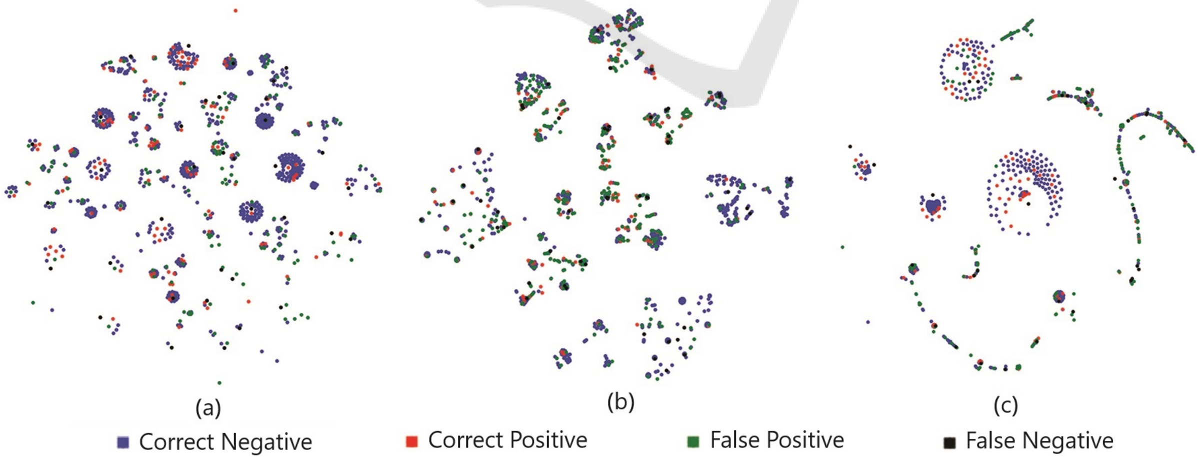

Figure 5: Characteristics illustrating the challenge in classification stemming from the absence of clearly defined instance groups (a) Relating to education module. (b) Relating to older adults health module (c) Relating to finance module.

class neighbors to better investigate which group they are most similar to (diagnosed or undiagnosed). Figure 9 presents the average values of the attributes of 3 relevant categories in the context of depression:

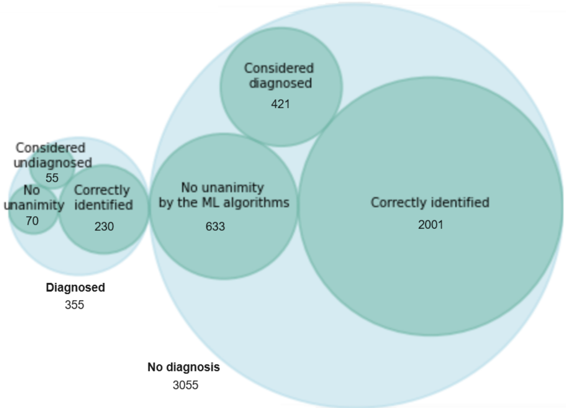

Figure 6: Diagram of the collective prediction of all ML algorithms during the testing stage.

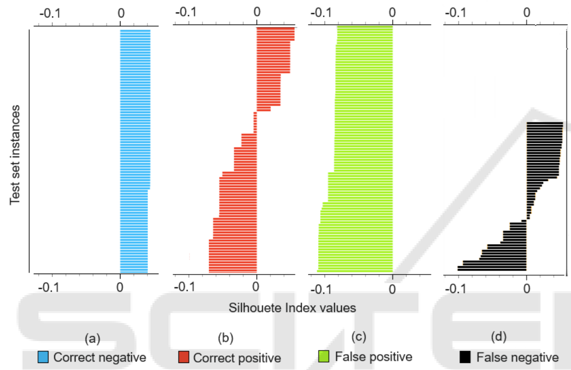

Figure 7: Silhouette index samples of each predicted instance group by the ML algorithms. (a) No diagnosis instances correctly classified. (b) Diagnosed instances correctly classified. (c) No diagnosis instances predicted as diagnosed (false positive). (d) Diagnosed instances predicted as no diagnosis (false negative). In (d), the silhouette values of all 55 instances identified as FN were represented. There were no more instances to describe, so the above space was kept blank.

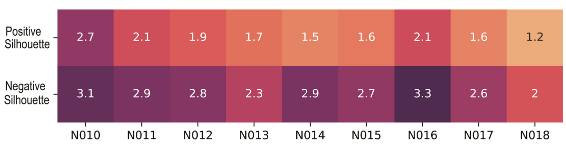

Figure 8: Comparison of attributes related to health perception in diagnosed-class instances with positive and negative silhouette index.

limitations caused by chronic diseases, perception of health status, and lifestyle. These categories were chosen because they better separated the classes in the spatial visualizations presented previously. In the heatmap mentioned above, in 74% of the attributes (23 of the 31 evaluated), the average values of false positives are closer to the average values of the diagnosed class (true positive) than the no diagnosis class (true negative).

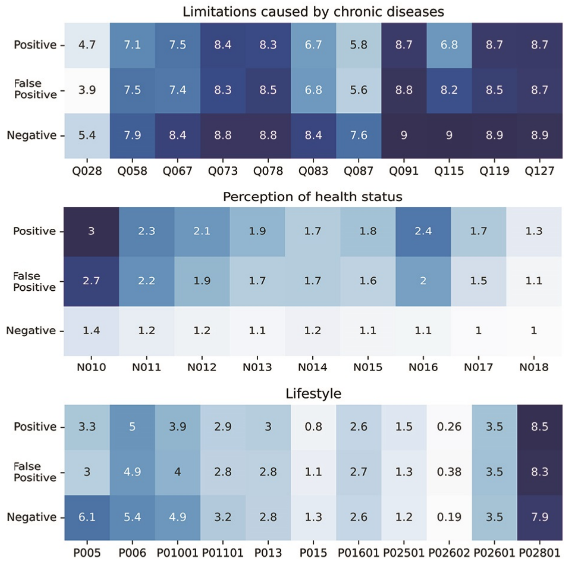

Figure 9: Comparison of attributes by category between instances erroneously classified as positive (false positive), diagnosed (true positive), and undiagnosed (true negative).

The average Si of the false positive group and the analysis of this comparative heatmap provide some clues that these people might be underdiagnosed or at risk of developing depressive disorder, needing a more detailed investigation and appropriate treatment, if necessary.

The last group, instances considered undiagnosed by the algorithms but have been diagnosed with depression (false negative), has almost equivalent amounts of instances with positive and negative Si, moving around the border between classes. The average Si of this false-negative group was 0.004, suggesting that they are in the correct group and that there was an inaccuracy in the models, but this value is too low to be entirely sure. Generally speaking, in almost all groups, the silhouette values are positioned on the borders of the classes, making it difficult to draw more objective conclusions about the individuals.

Corroborating the difficulty in diagnosing older people, Sjöberg et al. (2017) claims that the complex spectrum of Late-Life Depression (LLD), for example, goes beyond the main diagnostic entities of unipolar depressive disorders, such as Major Depressive Disorder (MDD) and persistent depressive disorder. Relevant depressive symptoms that do not fulfill the criteria for a diagnosis of depression have nevertheless a significant clinical relevance because of their association with poorer quality of life and increased disability; however, they are often undetected and untreated despite a minimal chance of spontaneous remission.

Moving now to the interpretability of the trained

ML models, we used the SHapley Additive exPlanations (SHAP) method (Lundberg and Lee, 2017) to explain the ML models[6]. Figure 10 presents an example of the Gradient Boost's explained output . We can visualize features' importance and impact on the class forecast. This graph classifies resources by the sum of the magnitudes of SHAP values across all samples. It also uses SHAP values to show the distribution of impacts for each feature. The color represents the attribute's value (red indicating high value and blue showing low). Thus, high values for attribute ("In the last two weeks, how often did you feel depressed or without perspective?"), for example, contribute to an exit positive (diagnosed class), while low values are the opposite. Attribute ("In the last two weeks, did you use any sleeping medication? 1-Yes, 2-No") contributes negatively to the positive output. That is high values (2) tend to pull the class to "no diagnosis", while low values (1) contribute to the "diagnosed" class.

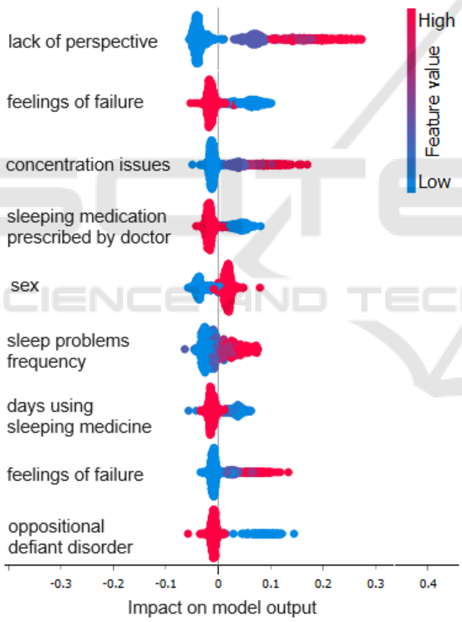

Figure 10: SHAP explanations for Gradient Boost model.

We also used SHAP explanations to extract the top 10 features of the four best-performing models and build a diagram to analyze the most significant ones. Figure 11 shows that all ML models incorporated attributes related to concentration issues (N013), lack of perspective (N016), feelings of failure (N017), and sleep problems (N010, Q133), as well as the use of medication for this purpose (Q132, Q134). In addition, chronic physical or mental health conditions, were selected by at least three models (J007). Furthermore, problems with feeling tired and without energy (N011), the number of biological children (Z00102, Z00101), and lack of interest or pleasure in doing things (N012) were considered very relevant in the prediction by at least 2 of the 4 best models. Finally, at least 1 model assumed attributes referring to slowness or restlessness when moving or speaking (N015), diagnosis of another mental illness (Q111, Q11009), and sex (C006) in its list of the 10 most relevant characteristics.

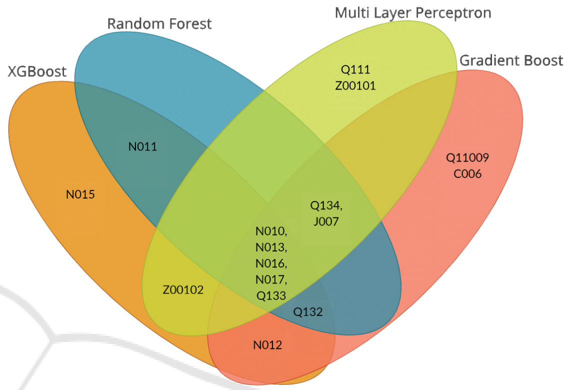

Figure 11: Main characteristics obtained by applying SHAP in the interpretability of the four best-trained ML models.

6 FINAL CONSIDERATIONS

Our study aimed to analyze the population of older adults participating in the NHS 2019. We used many different ML techniques to highlight the main factors related to depression in older people. We identified that all ML models incorporated attributes related to concentration issues, lack of perspective, feelings of failure, and sleep problems. Furthermore, we pointed out some evidence of underdiagnosis and called attention to more outstanding care in perceiving signs of depression in this target audience. As limitations of our work, it is necessary to mention the need for analysis of the results by a professional in the field of mental disorders in order to validate or not the signs of depression indicated. In future work, we can extend the analysis to other target audiences, such as children and adolescents whose diagnosis and early recognition of the disorder are not trivial. We also intend to investigate different types of normalization for different attributes, analyzing their influence on the classifier results (Singh and Singh, 2020).

[6]Interpretability performed using the *SHAP* package from *Python v3*

ACKNOWLEDGEMENTS

The authors thank the National Council for Scientific and Technological Development of Brazil (CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico – Code: 311573/2022-3), the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas, the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES – Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - Informática and Finance Code 001), and the Foundation for Research Support of Minas Gerais State (FAPEMIG – Code: APQ-03076-18).

REFERENCES

Alexopoulos, G. S. (2005). Depression in the elderly. *The lancet*, 365(9475):1961–1970.

Belo, P., Navarro-Pardo, E., Pocinho, R., Carrana, P., and Margarido, C. (2020). Relationship between mental health and the education level in elderly people: Mediation of leisure attitude. *Frontiers in Psychology*, 11.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Chang, H.-H. and Yen, S. T. (2012). Association between obesity and depression: evidence from a longitudinal sample of the elderly in taiwan. *Aging & mental health*, 16(2):173–180.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.

Cramer, J. S. (2004). The early origins of the logit model. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 35(4):613–626.

DeSouza, D. D., Robin, J., Gumus, M., and Yeung, A. (2021). Natural language processing as an emerging tool to detect late-life depression. *Frontiers in Psychiatry*, 12:719125.

Devita, M., De Salvo, R., Ravelli, A., De Rui, M., Coin, A., Sergi, G., and Mapelli, D. (2022). Recognizing depression in the elderly: Practical guidance and challenges for clinical management. *Neuropsychiatric Disease and Treatment*, pages 2867–2880.

Diniz, B. S., Reynolds III, C. F., Butters, M. A., Dew, M. A., Firmo, J. O., Lima-Costa, M. F., and Castro-Costa, E. (2014). The effect of gender, age, and symptom severity in late-life depression on the risk of all-cause mortality: The bambuí cohort study of aging. *Depression and anxiety*, 31(9):787–795.

Errazuriz, A., Avello-Vega, D., Ramirez-Mahaluf, J. P., Torres, R., Crossley, N. A., Undurraga, E. A., and Jones, P. B. (2023). Prevalence of depressive disorder in the adult population of latin america: a systematic review and meta-analysis. *The Lancet Regional Health–Americas*, 26.

Fernández-Niño, J. A., Bonilla-Tinoco, L. J., Manrique-Espinoza, B. S., Romero-Martínez, M., and Sosa-Ortiz, A. L. (2018). Work status, retirement, and depression in older adults: An analysis of six countries based on the study on global ageing and adult health (sage). *SSM-population health*, 6:1–8.

Foong, H. F., Hamid, T. A., Ibrahim, R., and Haron, S. A. (2021). The intersectional effects of ethnicity/race and poverty on health among community-dwelling older adults within multi-ethnic asian populace: a population-based study. *BMC geriatrics*, 21:1–10.

Gonyea, J., Curley, A., and Melekis, K. (2017). Correlates of depression among lower-income minority older adults living in urban subsidized housing. *Innovation in Aging*, 1(Suppl 1):1150.

IBGE (2021). População cresce, mas número de pessoas com menos de 30 anos cai 5,4% de 2012 a 2021. https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/34438-populacao-cresce-mas-numero-de-pessoas-com-menos-de-30-anos-cai-5-4-de-2012-a-2021. Accessed 11 Oct 2023.

Invernizzi, S., Simoes Loureiro, I., Kandana Arachchige, K. G., and Lefebvre, L. (2022). Late-life depression, cognitive impairment, and relationship with alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 50(5):414–424.

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., Kim, N., et al. (2019). Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. *JMIR mHealth and uHealth*, 7(10):e14149.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Kubat, M. (1999). Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7. *The Knowledge Engineering Review*, 13(4):409–412.

Lee, J. and Holtzer, R. (2021). Independent associations of apathy and depressive symptoms with perceived social support in healthy older adults. *Aging & Mental Health*, 25(10):1796–1802.

Li, D., Su, M., Guo, X., Liu, B., and Zhang, T. (2023). The association between chronic disease and depression in middle-aged and elderly people: The moderating effect of health insurance and health service quality. *Frontiers in Public Health*, 11:935969.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Lytle, A., Apriceno, M., Dyar, C., and Levy, S. R. (2018). Sexual orientation and gender differences in aging perceptions and concerns among older adults. *Innovation in Aging*, 2(3).

Mașstaleru, A., Abdulan, I. M., Ștefăniu, R., Lefter, N., Sandu, I. A., Pîslaru, A. I., Leon-Constantin, M. M., Alexa, I. D., and Ilie, A. C. (2022). Relationship between frailty and depression in a population from north-eastern romania. *International Journal of Environmental Research and Public Health*, 19(9):5731.

Mendes, F., Pereira, J., Zangão, O., Pereira, C., and Bravo, J. (2021). The relationship between depression and risk of violence in portuguese community-dwelling older people. *BMC Public Health*, 21(S2).

Palapinyo, S., Methaneethorn, J., and Leelakanok, N. (2021). Association between polypharmacy and depression: a systematic review and meta-analysis. *Journal of Pharmacy Practice and Research*, 51(4):280–299.

Pan, L., Li, L., Peng, H., Fan, L., Liao, J., Wang, M., Tan, A., and Zhang, Y. (2022). Association of depressive symptoms with marital status among the middle-aged and elderly in rural china–serial mediating effects of sleep time, pain and life satisfaction. *Journal of Affective Disorders*, 303:52–57.

Papageorgiou, A., Bakola, M., Kitsou, K., Mousafeiris, V., Mavridou, K., Kallianezos, P., Kampouraki, M., Charalambous, G., and Jelastopulu, E. (2022). The association between depression and quality of life in the elderly. *European Journal of Public Health*, 32(Supplement_3):ckac131–125.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Pigeon, W. R., Hegel, M., Unützer, J., Fan, M.-Y., Sateia, M. J., Lyness, J. M., Phillips, C., and Perlis, M. L. (2008). Is insomnia a perpetuating factor for late-life depression in the impact cohort? *Sleep*, 31(4):481–488.

Quittschalle, J., Pabst, A., Löbner, M., Luppa, M., Heser, K., Wagner, M., van den Bussche, H., Hajek, A., König, H.-H., Wiese, B., Angermeyer, M. C., Maier, W., Scherer, M., and Riedel-Heller, S. G. (2021). Association of alcohol and tobacco consumption with depression severity in the oldest old. results from the age different old age cohort platform. *International Journal of Environmental Research and Public Health*, 18(15):7959.

Rhee, T. G., Capistrant, B. D., Schommer, J. C., Hadsall, R. S., and Uden, D. L. (2018). Effects of the 2009 uspstf depression screening recommendation on diagnosing and treating mental health conditions in older adults: a difference-in-differences analysis. *Journal of managed care & specialty pharmacy*, 24(8):769–776.

Rogozinski, A. and Zisberg, A. (2020). Associations between depressive symptoms, appetite and food intake among hospitalized older adults. *Innovation in Aging*, 4(Suppl 1):919.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Salzberg, S. L. (1994). C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993.

Scazufca, M., Nakamura, C. A., Peters, T. J., Henrique, M. G., Seabra, A., La Rotta, E. G., Franzin, R. M., Martins, D. F., Van de Ven, P., Hollingworth, W., et al. (2020). A collaborative care psychosocial intervention to improve late life depression in socioeconomically deprived areas of guarulhos, brazil: the proactive cluster randomised controlled trial protocol. *Trials*, 21(1):1–14.

Singh, D. and Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524.

Sjöberg, L., Karlsson, B., Atti, A.-R., Skoog, I., Fratiglioni, L., and Wang, H.-X. (2017). Prevalence of depression: Comparisons of different depression definitions in population-based samples of older adults. *Journal of affective disorders*, 221:123–131.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Vapnik, V. N. (1997). The support vector method. In *International Conference on Artificial Neural Networks*, pages 261–271. Springer.

Wang, J., Li, R., Zhang, L., Gao, X., Zhou, M., Zhang, X., and Ma, Y. (2022). Associations between sedentary behaviour patterns and depression among people aged 60 and older in hebei province of china. *BMC public health*, 22(1):1–9.

Weyerer, S., Eifflaender-Gorfer, S., Wiese, B., Luppa, M., Pentzek, M., Bickel, H., Bachmann, C., Scherer, M., Maier, W., and Riedel-Heller, S. G. (2013). Incidence and predictors of depression in non-demented primary care attenders aged 75 years and older: results from a 3-year follow-up study. *Age and ageing*, 42(2):173–180.

WHO (2023). Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression. Accessed 14 Dec 2023.

Xu, Z., Zhang, Q., Li, W., Li, M., and Yip, P. S. F. (2019). Individualized prediction of depressive disorder in the elderly: a multitask deep learning approach. *International journal of medical informatics*, 132:103973.

Zhang, C., Chen, X., Wang, S., Hu, J., Wang, C., and Liu, X. (2021). Using catboost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry Research*, 306:114261.