

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE  
CENTRO DE BIOCÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOBIOLOGIA

**STEPHANIA RUTH BASILIO SILVA GOMES**

ANÁLISE PREDITIVA DE VARIÁVEIS DE SONO E BIOMARCADORES DE  
COMORMIDADES NA DETECÇÃO DE SINTOMAS DEPRESSIVOS EM ADULTOS DE  
MEIA-IDADE E IDOSOS: UMA ABORDAGEM DE APRENDIZAGEM DE MÁQUINAS

Dissertação apresentada à Universidade Federal  
do Rio Grande do Norte, para obtenção do título  
de Mestre em Psicobiologia.

Área de concentração: Psicologia Fisiológica

Natal  
2022

**STEPHANIA RUTH BASILIO SILVA GOMES**

**ANÁLISE PREDITIVA DE VARIÁVEIS DE SONO E BIOMARCADORES DE  
COMORRIDADES NA DETECÇÃO DE SINTOMAS DEPRESSIVOS EM ADULTOS DE  
MEIA-IDADE E IDOSOS: UMA ABORDAGEM DE APRENDIZAGEM DE MÁQUINAS**

Dissertação apresentada à Universidade Federal  
do Rio Grande do Norte, para obtenção do título  
de Mestre em Psicobiologia.

Área de concentração: Psicologia Fisiológica

**Orientador:** Prof. Dr. Mario Andre Leocadio Miguel

Natal

2022

Universidade Federal do Rio Grande do Norte - UFRN  
Sistema de Bibliotecas - SISBI

Catálogo de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Leopoldo Nelson - -Centro de Biociências - CB

Gomes, Stephania Ruth Basilio Silva.

Análise preditiva de variáveis de sono e biomarcadores de comorbidades na detecção de sintomas depressivos em adultos de meia-idade e idosos: uma abordagem de aprendizagem de máquinas / Stephania Ruth Basilio Silva Gomes. - 2022.

74 f.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Norte, Centro de Biociências, Programa de Pós-graduação em Psicobiologia. Natal/RN, 2022.

Orientador: Prof. Dr. Mario Andre Leocadio Miguel.

1. Sintomas depressivos - Dissertação. 2. Multimorbidades - Dissertação. 3. Machine Learning - Dissertação. I. Miguel, Mario Andre Leocadio. II. Universidade Federal do Rio Grande do Norte. III. Título.

RN/UF/BSCB

CDU 616.89-008.454

**Título:**

Análise preditiva de variáveis de sono e biomarcadores de comorbidades na detecção de sintomas depressivos em adultos de meia-idade e idosos: uma abordagem de aprendizagem de máquinas

**Autor:** Stephania Ruth Basilio Silva Gomes

**Data de defesa:** 26/07/2022

**Banca Examinadora:**

Prof. Dr. Felipe Beijamini

Universidade Federal da Fronteira do sul (UFFS)

Prof<sup>a</sup>. Dr<sup>a</sup>. Carolina Virginia Macedo de Azevedo

Universidade Federal do Rio Grande do Norte (UFRN)

Prof. Dr. Mario Andre Leocadio Miguel

Universidade Federal do Rio Grande do Norte (UFRN)

## AGRADECIMENTOS

Agradeço primeiramente com todo meu amor e afeto minha mãe, Maria Cícera da Silva Gomes, que sempre esteve presente me apoiando em toda minha trajetória e que muitas vezes renunciou a muitas coisas para me proporcionar o melhor que estava ao alcance. Também agradeço minhas irmãs (Leise Raquel Basilio Silva Gomes, Lívia da Silva) e sobrinhos (Bruno de França Silva Rocha e Theo Bernard da Silva Rocha), que também estabeleceram uma rede de apoio muito importante durante o processo do curso de mestrado em Psicobiologia, no qual ocorreu inclusive em um difícil cenário de pandemia.

Agradeço minha outra rede de apoio mais próxima que são minhas amigas pessoais: Juliana Leroy, Camila Marina, Jéssica Paiva, Hannah Lima, Natsumi Hamada e Luana Duarte. Companheiras estas as quais me deram suporte nos momentos mais difíceis de aflições e angústias.

Agradeço meu orientador, Mário André Leocadio Miguel, pela oportunidade de iniciar os estudos no Laboratório de Neurobiologia e Ritmicidade Biológica (LNRB) e desta forma experienciar um primeiro contato com uma área que estava pouco familiarizada, por ter sempre me incentivado a melhorar meu desempenho quando necessário, por ter acreditado no meu potencial e por me proporcionar momentos magníficos de aprendizagens científicas. Além disso, sou imensamente grata pelos aconselhamentos, ensinamentos e orientações dadas durante todo esse percurso de trabalho juntos.

Agradeço a colaboração de professores como da Rovena Clara Galvão Januário Engelberth, da Carolina Virginia Macedo de Azevedo e do Jerferson de Souza Cavalcante que foram extremamente solícitos em sanar dúvidas, em compartilhar conhecimentos e me auxiliar no meu processo formativo.

Agradeço também à toda equipe do LNRB que me proporcionou muitos momentos ricos de aprendizagens. Especialmente sou grata à Catarina Lira, Bruna Ribeiro, Antônia Líria, Lucas Alves, Letícia Barradas, Gian Siebra, Júlio Leal, Karen Pugliane, Breno Carneiro e John Fontenelle.

Por fim, agradeço ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). O presente trabalho foi realizado com auxílio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## RESUMO

**Introdução:** Estudos têm evidenciado altas incidências da depressão em escala mundial e sua co-ocorrência com várias condições de saúde importantes, principalmente em sujeitos de meia-idade e idosos. Neste cenário de multimorbidades, a depressão encontra-se comumente associada às doenças relacionadas à síndrome metabólica, como obesidade e diabetes. Alterações crônicas nos ritmos circadianos de sono-vigília representam uma relação com o desenvolvimento da depressão e de suas respectivas comorbidades associadas, ao favorecerem a quebra da organização temporal interna de processos fisiológicos e metabólicos essenciais. Atualmente, fazer diagnósticos e triagens clínicas precisas têm sido um desafio persistente na área de saúde mental, em virtude do uso de ferramentas tradicionais limitadas que não incluem características adicionais de dados clínicos importantes do paciente, inclusive observações objetivas biomarcadoras de doenças. **Objetivo:** Dessa forma, o objetivo do presente estudo foi detectar a sintomatologia depressiva a partir de biomarcadores gerais de obesidade e diabetes, além de variáveis relacionadas ao sono e atividade física, em adultos de meia-idade e idosos, através de uma abordagem de aprendizagem de máquinas. **Método:** Dados do Questionário de Atividade Física Global (GPAQ - nível de atividade física), do Questionário de Saúde do Paciente (PHQ-9), e do questionário sobre hábitos de sono foram extraídos da base de dados do National Health and Nutrition Examination Survey (NHANES) no período de 2015-2016. Outras variáveis foram acessadas e utilizadas como recursos previsores, como medidas antropométricas e biomarcadores plasmáticos da obesidade e diabetes. Um total de 2907 sujeitos adultos de meia-idade e idosos foram elegíveis para o estudo. Três algoritmos de aprendizagem supervisionada foram implementados: Regressão Logística penalizada com Lasso (RL), Random Forest (RF) e Extreme Gradient Boosting (XGBoost). **Resultados:** O modelo XGBoost proporcionou maior acurácia e precisão (87%), com proporção de acertos nos casos com sintomas depressivos acima de 80%. Além disso, a sonolência diurna foi a variável preditora que melhor contribuiu para prever sintomas depressivos. **Conclusões:** As variáveis de sono e atividade física, além dos biomarcadores de obesidade e diabetes, em conjunto assumem significativa importância para prever, com acurácia e precisão de 87%, a ocorrência de sintomas depressivos em indivíduos de meia-idade e idosos.

**Palavras-chave:** Sintomas depressivos; Multimorbidades; Sono; Atividade física; Machine Learning

## ABSTRACT

**Introduction:** studies have shown high incidences of depression worldwide and its co-occurrence with several important medical conditions, especially in middle-aged and elderly subjects. In this multimorbidity scenario, depression is commonly associated with diseases related to metabolic syndrome, such as obesity and diabetes. Chronic alterations in circadian sleep-wake rhythm represent a relationship with the development of depression and its associated comorbidities, as they favor the breakdown of the internal temporal organization of essential physiological and metabolic processes. Currently, making accurate clinical diagnoses and screenings have been a persistent challenge in mental health, due to the use of limited traditional tools that do not include additional characteristics of important clinical data of the patient, including objective observations of disease biomarkers. **Objective:** Thus, the objective of the present study was to detect depressive symptomatology from general biomarkers of obesity and diabetes, as well as variables related to sleep and physical activity, in middle-aged and elderly adults, through a learning approach of machines. **Method:** Data from the Global Physical Activity Questionnaire (GPAQ - physical activity level), from the Patient Health Questionnaire (PHQ-9), and from the sleep habits questionnaire were extracted from the National Health and Nutrition Examination Survey database (NHANES) in the period 2015-2016. Other variables were accessed and used as predictive resources, such as anthropometric measurements and plasmatic biomarkers of obesity and diabetes. A total of 2907 middle-aged and elderly adult subjects were eligible for the study. Three supervised learning algorithms were implemented: Penalized Logistic Regression with Lasso (RL), Random Forest (RF) and Extreme Gradient Boosting (XGBoost). **Results:** The XGBoost model provided greater accuracy and precision (87%), with a proportion of correct answers in cases with depressive symptoms above 80%. In addition, daytime sleepiness was the predictor variable that best contributed to predicting depressive symptoms. **Conclusions:** Sleep and physical activity variables, in addition to obesity and diabetes biomarkers, together assume significant importance in predicting, with an accuracy and precision of 87%, the occurrence of depressive symptoms in middle-aged and elderly individuals.

**Keywords:** Depressive symptoms; Multimorbidities; Sleep; Physical activity; machine learning

## SUMÁRIO

<b>1. Introdução Geral</b> .....	<b>8</b>
1.1. Características gerais do sono .....	8
1.2. Aspectos gerais e regulatórios do ciclo sono-vigília .....	9
1.3. Aspectos gerais da depressão .....	12
1.4. Importância do sono na depressão .....	13
1.5. A natureza multicomorbida da depressão .....	15
1.6. Estilo de vida sedentário e depressão .....	17
1.7. Os desafios da detecção da depressão na atenção primária em saúde .....	18
1.8. Aprendizagem de Máquinas ( <i>Machine learning</i> ) voltada à problemas de saúde .....	19
<b>2. Objetivos</b> .....	<b>21</b>
2.1. Objetivos gerais .....	21
2.2. Objetivos específicos .....	22
<b>3. Hipóteses e predições</b> .....	<b>22</b>
<b>Capítulo 1-</b> Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: a machine learning approach	
Abstract .....	24
Introduction .....	27
Methods .....	29
Results .....	33
Discussion .....	39
Conclusion .....	42
References .....	43
<b>4. Discussão Geral</b> .....	<b>50</b>
<b>5. Conclusão Geral</b> .....	<b>53</b>
<b>6. Implicações do estudo</b> .....	<b>53</b>
<b>7. Referências Gerais</b> .....	<b>54</b>
<b>8. Apêndices</b> .....	<b>70</b>



## **1. Introdução Geral**

### **1.1. Características gerais do sono**

O sono já foi definido como “uma inconsciência reversível”, caracterizado principalmente pela ausência de mobilidade e menor resposta aos estímulos externos (Carskadon & Dement, 2011). Entretanto, o sono é um processo fisiológico ativo, envolvendo a interação de múltiplos componentes do sistema nervoso central e autônomo, segundo Perrotta, 2019. A avaliação dos critérios fisiológicos do sono, por sua vez, deve ocorrer a partir da combinação de registros eletrofisiológicos do eletroencefalograma (EEG), eletro-oculograma (EOG) e o eletromiograma (EMG), para observar as ocorrências de atividade cerebral, oculares e musculares durante as fases do sono, respectivamente (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957; Jouvet et al., 1959).

Ainda, o sono é subdividido em duas fases principais, conhecidas como sono NREM (sono de movimentos não rápidos dos olhos), na qual podemos observar a presença de três estágios (N1, N2, N3), e como sono REM, ou paradoxal, (sono de movimento rápido dos olhos), de acordo com o Manual da Academia Americana de Medicina do sono (AASM, 2007). No estágio 1 do sono NREM (N1), o padrão de EEG evidencia atividades de baixa amplitude e frequência mista, enquanto no estágio 2 (N2), observa-se a presença de ondas teta e delta de alta voltagem, além de formas específicas de ondas, tais como os fusos do sono e complexos K. Já no estágio 3 (N3), considerado a fase de sono mais profundo, encontra-se um sono caracterizado por ondas lentas (SWS) de baixa frequência e presença de ondas deltas e tetas de alta amplitude (Soldatos & Paparrigopoulos, 2005; Chokroverty, 2010; Patel et al., 2021). O sono REM, por outro lado, está associado à ocorrência dos sonhos em humanos e é marcado por um padrão no EEG semelhante à vigília, constituído de ondas cerebrais com uma amplitude reduzida e frequência alta (Siegel, 2011; Peever & Fuller, 2017). Além disso, o sono REM é considerado um “sono paradoxal” devido a esta semelhança com a fase de

vigília e ao mesmo tempo por desencadear a atonia muscular de alguns músculos (Peever & Fuller, 2017).

Importantemente, a expressão usual do ciclo sono-vigília, caracterizada pela alternância cíclica entre as suas fases, bem como a ocorrência consolidada de um bloco diurno de vigília e um bloco noturno de sono, depende de processos regulatórios para ser expresso adequadamente (Borbély et al., 2016).

## **1.2. Aspectos gerais e regulatórios do ciclo sono-vigília**

O ciclo sono-vigília (CSV) é um ritmo biológico de caráter circadiano, ou seja, com um período de aproximadamente 24h. A sincronização deste ritmo ocorre através de pistas ambientais temporais, conhecidas como agentes arrastadores ou doadores de tempo (*Zeitgebers*), os quais podem ser a informação luminosa do ciclo claro-escuro ambiental e estímulos não-fóticos tais como o convívio social e o regime alimentar, por exemplo (Mistlberger & Skene, 2004; Pardini & Kaeffer, 2006). O CSV é parcialmente regulado por um Sistema de Temporização Circadiana (STC) que é também responsável pela organização temporal interna do organismo e de outros processos fisiológicos importantes, assim como a antecipação de comportamentos essenciais para a sobrevivência (Dibner et al., 2010).

Em mamíferos, a arquitetura do STC é constituída por vias de entradas fotorreceptoras localizadas na retina, compostas por células fotossensíveis denominadas cones e bastonetes, além das células ganglionares da retina intrinsecamente fotossensíveis (ipRGCs) as quais expressam melanopsina, um pigmento sensível à luz (Rea et al., 2005). A informação sobre a variação da intensidade luminosa é conduzida até os Núcleos Supraquiasmáticos (NSQ), marcapasso central, principalmente através do trato retino-hipotalâmico (TRH) (Dibner et al., 2010). Entretanto, existem duas outras vias sincronizadoras do NSQ, ambas não-fóticas e fóticas indiretas, tratando-se então do trato genículo-hipotalâmico (TGH), que surge no folheto intergenículado (FIG) e a entrada serotoninérgica (5-HT), que se origina do núcleo

dorsal e mediano da rafe (Cavalcante et al., 2006; Dibner et al., 2010). O NSQ, por sua vez, coordena a sincronização de osciladores secundários distribuídos em diversos tecidos (Rosenwasser & Turek, 2005; Dibner et al., 2010).

Neste contexto, a regulação circadiana do ciclo sono-vigília ocorre principalmente através da estimulação de circuitos neurais, constituídos por células especializadas na indução e manutenção do despertar e do sono (REM e NREM), pelo NSQ (Mistlberger, 2005). Assim, o relógio central promove o despertar na fase ativa e o sono na fase de repouso, coincidindo com a fase de claro e escuro ambiental, respectivamente.

Adicionalmente à regulação circadiana, o sono sofre influência e interage com um componente homeostático, segundo o modelo de dois processos de regulação do sono (Processo C e S) proposto inicialmente por Borbéley (1982). Dessa forma, o processo homeostático (S), é um mecanismo dependente do tempo de vigília anterior, representando um aumento na propensão ao sono quanto maior o tempo desperto, decorrente do acúmulo de moléculas de adenosina neste período (Beersma & Gordijn, 2007; Fuller et al., 2006). Este mesmo impulso homeostático diminui durante as horas de sono e, em contrapartida, para manter o estado de vigília, a propensão ao sono é neutralizada pelo impulso circadiano durante o dia, demonstrando a atuação de ambos os processos na regulação do ciclo sono-vigília (Borbély et al., 2016).

A atividade de ondas lentas (SWA), por sua vez, constitui o principal marcador da homeostase do sono durante a fase NREM e em situações de privação de sono essa perda costuma ser compensada principalmente através do aumento da SWA, possibilitando assim um processo de intensificação e aprofundamento do sono NREM (Borbély et al., 2016; Deboer, 2018). Dessa forma, o acúmulo de SWA muda de acordo com o processo S, podendo ser verificado uma taxa maior nos primeiros episódios de sono NREM decorrente da alta pressão homeostática presente ao iniciar o sono (Acheram et al., 1993).

No entanto, durante o processo de envelhecimento alguns destes aspectos na regulação homeostática do sono podem sofrer alterações. Assim, embora idosos acumulem níveis mais altos de adenosina em comparação aos adultos mais jovens, eles costumam sofrer com um menor sinal de pressão do sono por existir uma redução dos receptores de adenosina (Mander et al., 2017). Além disso, estudos com sujeitos de meia-idade que avaliaram o impacto da privação do SWS durante episódios de recuperação do sono iniciada pela manhã, sob influência de uma propensão circadiana contrária ao sono, evidenciaram um menor efeito rebote de densidade de ondas lentas (SW) e uma recuperação de SWS e SWA menos pronunciada na meia-idade em comparação com indivíduos mais jovens (Gaudreau et al., 2001; Lafortune et al., 2012). Um padrão similar é encontrado em idosos, havendo também uma menor predominância frontal de atividade delta no cérebro após privação do sono quando comparado a adultos jovens (Münch et al., 2004; Taillard et al., 2021).

Ainda, podemos observar características similares no sono tanto no envelhecimento saudável quanto na depressão, sendo comum em ambas as condições mudanças relacionadas à continuidade e arquitetura do sono (Espiritu, 2008; Steiger & Pawlowski, 2019). Inicialmente, para explicar as alterações do sono na depressão foi proposto uma hipótese da deficiência no processo S, a qual diz que o nível do processo S é inferior ao alcançado por sujeitos saudáveis no transcorrer da vigília, ocasionando, por exemplo, um sono mais superficial e curto (Borbély, 1987). Entretanto, um estudo evidencia o oposto, com mulheres depressivas jovens apresentando níveis consideravelmente mais elevados de SWA principalmente na região frontal em situações de privação de sono quando comparadas a mulheres jovens saudáveis (Frey et al., 2012). Além disso, este mesmo estudo indica diferenças no padrão da regulação homeostática do sono entre o envelhecimento e a condição depressiva ao ser observado na amostra idosa níveis de SWA mais atenuados e menos predominantes na região frontal cerebral em comparação às mulheres jovens depressivas.

### **1.3. Aspectos gerais da depressão**

O transtorno depressivo compreende uma das disfunções mentais mais prevalentes e incapacitantes na contemporaneidade, afetando aproximadamente 322 milhões de pessoas em extensão global, segundo dados da Organização Mundial de Saúde (OMS, 2017). Em termos etiológicos, vários agentes intrínsecos e extrínsecos podem contribuir para sua prevalência e consequente previsão, o que inclui fatores biológicos, ambientais, psicossociais ou sociodemográficos (Saveanu & Nemeroff, 2012).

De acordo com o Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5), a sintomatologia depressiva envolve a presença de humor deprimido, diminuição significativa do interesse ou prazer em tudo ou em quase todas as atividades, perda ou ganho de peso, insônia ou hipersonia, agitação ou retardo psicomotor, fadiga ou perda de energia, sentimentos de inutilidade ou culpa excessiva, diminuição da capacidade de pensar ou se concentrar e a presença recorrente de pensamentos de morte, ideias ou tentativa de suicídio. Entretanto, para ser considerado um transtorno depressivo maior (TDM) cinco ou mais destes sintomas devem estar presentes durante pelo menos duas semanas.

Existem algumas hipóteses que tentam explicar a origem etiológica da depressão. Na década de 60 foi proposta a hipótese da catecolaminas dos transtornos afetivos, na qual diz que alguns tipos de depressão estão associadas a uma diminuição das catecolaminas, especialmente a noradrenalina disponíveis nos receptores adrenérgicos centrais (Schildkraut, 1965). Outro estudo, no entanto, evidenciou que a depressão pode surgir pela deficiência de serotonina no cérebro (Lapin & Oxenkrug, 1969). Esses estudos, baseados em mecanismos de ação de algumas drogas antidepressivas, impulsionaram o que conhecemos hoje como a teoria clássica monoaminérgica da depressão, propondo que a origem da depressão é proveniente de um déficit na neurotransmissão monoaminérgica envolvendo principalmente os neurotransmissores noradrenalina e serotonina.

Alguns estudos têm evidenciado o estresse como um dos principais fatores associados ao transtorno depressivo (Qin et al., 2016). Em situações de estresse crônico, por exemplo, ocorre como resposta fisiológica uma hiperestimulação do eixo hipotálamo-hipófise-adrenal (HPA), desencadeado principalmente uma elevação da expressão do fator liberador de corticotropina (CRF), o qual estimula a secreção pela hipófise do hormônio adrenocorticotrófico (ACTH), que é essencial para estimular a secreção de glicocorticóides (Sapolske, 2000; Berton e Nestler, 2006). Nesse contexto, a hipersecreção de cortisol tem sido considerado um importante marcador do estresse em sujeitos depressivos (Bhagwagar et al., 2005).

#### **1.4. Importância do sono na depressão**

Distúrbios do sono frequentemente coocorrem com o transtorno depressivo, sejam eles distúrbios de continuidade do sono envolvendo principalmente o aumento de despertares noturnos e eficiência do sono reduzida, distúrbios do sono de ondas lentas, destacando-se sua redução, ou distúrbios do sono REM, caracterizados sobretudo pela latência encurtada e o aumento da sua atividade (Berger et al., 2003; Wichniak et al., 2013). Além disso, na condição depressiva existe alta prevalência de casos de insônia e apneia obstrutiva do sono (Khurshid, 2018; Vanek et al., 2020).

Estudos indicam uma relação bidirecional entre sono e depressão, compartilhando mecanismos regulatórios comuns (Pandi-Perumal et al., 2020). Os sistemas de neurotransmissão colinérgica e serotoninérgica podem representar os substratos compartilhados na regulação do humor, assim como do sono e da vigília (Adrien, 2002). Neste sentido, a depressão está associada a uma diminuição da neurotransmissão serotoninérgica e um aumento da colinérgica. Em contrapartida, o sistema colinérgico é capaz de promover atividades oscilatórias de alta frequência durante a vigília e sono REM através de neurônios colinérgicos localizados no prosencéfalo basal, disparando mais rapidamente

nestas duas fases em virtude da ativação cortical. A serotonina, por sua vez, facilita a vigília e desempenha papel na supressão do sono REM (Brown et al., 2012; Adrien, 2002).

Quanto à gravidade da depressão, ela está atrelada a diversas alterações na fase e na amplitude dos ritmos circadianos. Dessa forma, diferentes estudos têm mostrado uma associação entre o cronotipo vespertino e a maior gravidade dos sintomas de humor na depressão (Gaspar-Barba et al., 2009; Au e Reece, 2017; Antypa et al., 2017), evidenciando que quanto mais atrasada a fase de sono, maior é a gravidade do estado depressivo. Outros estudos, por exemplo, revelam uma redução na amplitude dos ritmos de temperatura corporal e melatonina em pacientes com sintomas depressivos, além de uma menor liberação de melatonina noturna (Souêtre et al., 1989; Soria & Urretavizcaya 2009; Lorenz et al., 2019).

A sonolência diurna excessiva (SDE), por sua vez, é altamente prevalente entre os casos de depressão, chegando a ser superior a 50%, como expresso em estudo conduzido com 703 sujeitos depressivos (Hein et al., 2019). Por outro lado, a SDE geralmente é mais frequente entre as mulheres, estando mais associada à depressão atual e àquela desenvolvida ao longo da vida, do que ao transtorno de ansiedade, por exemplo (Hayley et al., 2013). Além disso, independente do sexo, o maior risco de desenvolver SED está associado a um histórico de depressão preexistente (Gonsalvez et al., 2021). Estudos também têm evidenciado uma associação entre SED e maior gravidade da depressão (Hublin et al., 1996; Chellappa & Araújo, 2006; Mume, 2010). Ainda, a SED constitui um marcador importante da depressão ao ser capaz de discriminar níveis da gravidade do transtorno depressivo em sujeitos com esta condição (Nuyen et al., 2016).

A SDE pode surgir de uma duração do sono menor que a necessidade individual de tempo de sono (Johns & Hocking, 1997; Keneita et al., 2005). Na condição depressiva, entretanto, tanto a ocorrência da curta duração do sono, quanto da longa, representa um fator de risco importante para seu desenvolvimento (Zhai et al., 2015; Dong et al., 2022).

Especialmente em idosos, um estudo longitudinal mostrou uma forte associação entre depressão e a curta duração do sono, ao evidenciá-la como fator de risco importante para a depressão incidente e recorrente (Sun et al., 2018).

### **1.5. A natureza multicomorbida da depressão**

A coexistência de outras condições médicas é bem característica no transtorno de humor depressivo (Smith et al., 2014). Estudos têm apresentado um aumento no risco de desenvolver depressão, de até 2 vezes ou mais, em sujeitos que apresentam multimorbidades (Read et al., 2017), com a prevalência da sintomatologia depressiva e respectivo nível de gravidade da depressão aumentando de acordo com o acréscimo no número de doenças crônicas (Gunn et al., 2012; Tong et al., 2021). Em contrapartida, a depressão também pode representar um fator de risco para o desenvolvimento de doenças físicas (Li et al., 2021).

A obesidade compreende uma das doenças crônicas mais comuns que coexistem com a depressão. Os indivíduos que apresentam esta condição são 32% mais propensos a desenvolverem o transtorno depressivo, quando comparados a sujeitos com peso adequado, segundo meta-análise realizada por Pereira-Miranda et al. (2017). Deste modo, um IMC alto ( $\geq 30 \text{ kg/m}^2$ ), que faz referência a uma obesidade geral, tem sido associado ao aumento nas chances de depressão, tanto em mulheres quanto em homens, além de estar associado à maior ocorrência de sintomatologia moderada à grave em mulheres (Dong et al., 2004; Ma & Xiao, 2010; Zavala et al., 2018). Já a obesidade abdominal, revelada por medidas de circunferência da cintura elevadas, é consideravelmente mais alta em sujeitos depressivos, em relação a sujeitos saudáveis, associando-se positivamente com a ocorrência de sintomas depressivos em mulheres (Ma & Xiao, 2010; Kaner et al., 2015).

Neste cenário de obesidade a depressão pode estar associada principalmente à alterações metabólicas, como a síndrome metabólica (SM), e por um consequente estado inflamatório de baixo grau (Hryhorczuk et al., 2013; Jantaratnotai et al., 2017).



Importantemente, vale ressaltar que a obesidade não representa um risco elevado no desenvolvimento da sintomatologia depressiva em sujeitos metabolicamente saudáveis, como evidenciado pelos resultados obtidos por Hamer et al. (2012) em estudo longitudinal. Na SM é verificada uma condição de dislipidemia, caracterizada principalmente pela redução nos níveis séricos de Lipoproteína de Alta Densidade (HDL), sendo este um atributo presente em sujeitos deprimidos (Lehto et al., 2008). Além disso, o HDL cai progressivamente com a manutenção dos sintomas deprimidos por tempo prolongado (Lehto et al., 2010). Sobre outras características do perfil lipídico também podemos observar o colesterol total em níveis reduzidos associado à sintomatologia depressiva, especialmente em homens de meia-idade e idosos (Morgan et al., 1993; Steegmans et al., 2000). Quanto ao estado inflamatório, a sintomatologia depressiva e a obesidade recebem mediação dos níveis de Proteína C Reativa (PCR), que estão aumentados nos processos inflamatórios, assim como na condição depressiva e de obesidade (Ford & Erliger, 2004; Daly, 2013).

Outra condição crônica que assume grande importância na depressão é a diabetes, podendo as chances de desenvolver depressão serem duas vezes maiores em pacientes com diabetes mellitus tipo I e II, em relação à sujeitos não diabéticos (Anderson et al., 2001). As estimativas de prevalência da depressão em estudo envolvendo indivíduos iranianos com diabetes foi superior a 50%, indicando uma forte relação entre estas duas condições (Pashaki et al., 2019). Além disso, um IMC  $\geq 30$  kg/m<sup>2</sup> e concentrações de PCR  $\geq 4,5$  mg/l representam fatores de riscos para a depressão comórbida em sujeitos com diabetes tipo II (Hein et al., 2018). Um dos biomarcadores mais utilizados para avaliar a associação entre diabetes e depressão é a hemoglobina glicada (HbA1c), a qual indica o controle glicêmico. Existe uma propensão maior de pessoas com concentrações de HbA1c alta ter sintomas depressivos quando comparadas às que têm um melhor controle da glicemia (Li et al., 2019). Outros estudos evidenciam uma relação recíproca entre níveis de HbA1c e sintomas

depressivos, mostrando, por exemplo, uma redução significativa da hemoglobina glicada decorrente da diminuição da sintomatologia depressiva após cuidados de 24 meses (Schmitz et al., 2016; Ali et al., 2020).

Neste contexto, sujeitos de meia idade e idosos são frequentemente acometidos por inúmeras comorbidades. A prevalência de multimorbidade, considerando mais de uma condição clínica em uma população idosa, chega a ser de 58% e aumenta significativamente de acordo com a idade (Kirchberger et al., 2012). Em centros comunitários de saúde, por exemplo, tanto adultos mais velhos quanto de meia idade apresentam padrões de comorbidades que englobam dimensões somáticas e de saúde mental (Quinõnes et al., 2022). Ainda, neste mesmo público, é comum uma associação entre multimorbidades cardiometabólicas e sintomas depressivos, com a obesidade e o diabetes representando comorbidades importantes (Huang et al., 2022).

### **1.6. Estilo de vida sedentário e depressão**

O comportamento sedentário (CS) e o menor engajamento em atividades físicas de intensidade moderada a vigorosa (AFMV) são característicos na condição depressiva e estão associados ao maior risco de desenvolver depressão (Zhai e Zhang, 2015; Schuch et al., 2017). Além disso, em estudo com adultos e idosos americanos, foi evidenciado que o CS se associa com a ocorrência de sintomas depressivos de forma mediada pela redução em AFMV, bem como diminuição da duração do sono (del Pozo Cruz et al., 2020). Por outro lado, Pengpid & Peltzer (2019), enfatizaram que um tempo superior ou igual a 8h gasto com atividades sedentárias não aumenta as chances de depressão, apesar da simples prática de AFMV reduzir essas chances. Ainda, em adultos de meia idade e mais velhos, a relação do comportamento sedentário com multimorbidades se estende além da causa depressiva, sendo problemas com mobilidade, dor e desconforto uns dos fatores importantes nesta relação (Vancampfort et al., 2017).

A prática de atividade física, por sua vez, reduz a sintomatologia depressiva e está relacionada ao estímulo de vários processos neuroplásticos envolvidos na depressão. Dessa forma, associa-se, por exemplo, ao crescimento de regiões cerebrais que geralmente são encontradas reduzidas em pacientes depressivos, como o hipocampo, o córtex pré-frontal e o córtex cingulado anterior (Kandola et al., 2019). Em camundongos é evidenciado que a prática de exercícios em intensidade alta está associada a redução de comportamentos depressivos ao reverter alterações regulatórias negativas na expressão da proteína do receptor de glicocorticóides presentes no hipocampo, córtex pré-frontal medial e amígdala (Liu et al., 2022). Em humanos, a prática aeróbica de exercícios também é eficaz na diminuição de sintomas depressivos, sobretudo naqueles sujeitos com sintomatologia mais grave e que sofrem com déficits no processamento do sistema de recompensa (Brush et al., 2022). Além disso, estudos com modelos animais mostraram que o fator neurotrófico derivado do cérebro (BDNF), que dentre suas várias funções atua como um importante modulador de neurotransmissores, pode ser eficaz no tratamento da depressão através do aumento do RNA mensageiro do BDNF estimulado pelo exercício físico, por exemplo (Russo-Neustadt et al. 2000; Russo-Neustadt et al. 2001). Ainda, outros estudos sugerem que o exercício físico tem ação antidepressiva ao aumentar a transmissão aminérgica, a exemplo inclusive do aumento da sensibilização de receptores de serotonina após um período de treinamento físico encontrado em ratos (Ransford. 1982; Dey, 1994).

### **1.7. Os desafios da detecção da depressão na atenção primária em saúde**

Embora a taxa de detecção da depressão seja considerada baixa no ambiente de atenção primária em saúde, em adultos de modo geral e especialmente em idosos que apresentam depressão comórbida, muitos casos conseguem ser identificados e encaminhados por médicos não psiquiatras (Craven & Bland, 2013; Akincigil & Matthews, 2017). Geralmente são utilizadas escalas e questionários como, por exemplo, o Inventário de

Depressão de Beck (BDI), o Questionário de Saúde do Paciente (PHQ-9), a Escala de Depressão Geriátrica e dentre outros que costumam ser instrumentos adequados e eficientes para detectar e acompanhar a evolução do transtorno depressivo (Rapp et al., 1988; Robison et al., 2002; Craven & Bland, 2013).

Entretanto, comumente existe uma dificuldade dos profissionais de saúde na administração destes recursos em virtude do maior tempo necessário para a sua aplicação e interpretação de resultados mais complexos. Além disso, pode existir uma interferência direta do paciente, o qual pode omitir informações sobre o estado atual de saúde na ausência de dados clínicos objetivos, principalmente do público idoso que tem uma tendência maior de carregar crenças negativas e estigmas sociais sobre a doença e muitas vezes negligencia o tratamento (Jang et al., 2007; Jang et al., 2009; Nair et al., 2020). Dessa forma, é comum encontrarmos problemas relativos a um diagnóstico precoce da depressão que interfere diretamente no tratamento eficaz desta condição.

Assim, torna-se urgente desenvolver ferramentas baseadas em aprendizagem de máquinas para detectar a depressão considerando biomarcadores importantes que incluam dados clínicos objetivos, possibilitando assim uma detecção eficiente da depressão, intencionando maior precisão e menor tempo de administração, além de menor interferência do paciente no processo avaliativo.

### **1.8. Aprendizagem de Máquinas (*Machine Learning*) voltada à problemas de saúde**

Atualmente, contamos com a disponibilidade de uma ampla variedade de banco de dados nacionais distribuídos pelo mundo que armazenam um grande volume de informações sobre saúde da população, conhecidos como biobancos, que têm como um dos objetivos apoiar a comunidade médica através de uma abordagem personalizada na identificação de biomarcadores importantes que possam estabelecer diagnósticos com segurança (Coppola et al., 2019). Dois grandes conhecidos biobancos de base populacional são o UK Biobank no

Reino Unido, no qual conta com dados de 500.000 participantes recrutados a longo prazo entre 2006-2010, e a Pesquisa Nacional de Exame de Saúde e Nutrição (NHANES) nos Estados Unidos que anualmente avalia uma amostra populacional de aproximadamente 5.000 indivíduos (Allen et al., 2012; NHANES, 2022).

Essas diversas informações biológicas, coletadas em grande velocidade e armazenadas em larga escala ao longo do tempo, conecta os biobancos ao universo do *Big Data* (Kinkorová & Topolčan, 2020). Por *Big Data* entende-se os aspectos voltados para o volume, velocidade e variedade dos dados, de acordo com Laney, 2001. Dessa forma, o volume diz respeito ao tamanho dos dados, geralmente em terabytes e petabytes, que exigem grande capacidade de processamento e armazenamento computacional. Enquanto a velocidade representa a rapidez com que esses dados são coletados e a variedade significa a diversidade de formatos possíveis, incluindo dados estruturados, não estruturados e até semiestruturados (Furht & Villanustre, 2016; Grable & Lyons, 2018).

Assim, diante deste cenário de complexidade nos dados foram necessárias inserções de novas metodologias direcionadas para análise de dados envolvendo Inteligência Artificial (IA), como os próprios modelos de *Machine Learning* (ML), para encontrar possíveis padrões escondidos nos dados e também fazer previsões mais robustas, possibilitando assim a melhora da saúde pública através do desenvolvimento de algoritmos que auxiliem em diagnósticos e prognósticos (Syeda-Mahmood, 2018; Silahtaroglu & Yilmaztür, 2021). Além disso, tornar viável a aplicação de políticas públicas mais adequadas e eficientes que permitam o controle e prevenção de doenças na população ao se obter uma visão geral sobre a previsão de risco de uma determinada doença, a exemplo do risco cardiovascular, ou a mortalidade do indivíduo com base nos preditores utilizados, por exemplo (Alaa et al., 2019; Soffer et al., 2021).

Neste contexto, o campo da psicologia clínica e a psiquiatria têm grande potencial de se beneficiar dos modelos preditivos de ML, através do auxílio nas tomadas de decisões em diagnósticos de depressão ou escolhas de melhor tratamento para esta condição, por exemplo. Diferentes estudos têm evidenciado isso ao testar o poder preditivo de um conjunto altamente diverso de biomarcadores associados à depressão para prever casos de transtorno depressivo ou casos de depressão resistente ao tratamento utilizando métodos de aprendizagem supervisionada e encontrando precisão relevante nos modelos de ML com apoio dos recursos previsores, assim como na identificação das variáveis mais relevantes para a previsão (Dipnall et al., 2016; Kautzky et al., 2017; Sharma & Verbeke, 2020).

Embora incipiente, o campo da cronobiologia tem contribuído com a criação de modelos de aprendizagem de máquina para o diagnóstico da depressão, a previsão dos episódios de exacerbação dos sintomas depressivos, os episódios de reinternação por transtorno depressivo maior e a gravidade da doença. Em predominância, estes estudos têm utilizado dados objetivos de detecção passiva do ciclo atividade-reposo, dados clínicos e biomarcadores com associações claras com a condição depressiva (Cho et al., 2019; Cearns et al., 2019; Narziev et al., 2020; Rodriguez-Ruiz et al., 2020). Nesta perspectiva, muitos aspectos cronobiológicos mesclados com outras variáveis fisiológicas ainda precisam ser explorados a partir desta metodologia para melhor entendermos a natureza complexa da depressão, bem como facilitar o processo de triagem dos casos, sobretudo em ambientes de atenção primária à saúde, característico em apresentar uma maior dificuldade na identificação de indivíduos depressivos (de Souza Filho et al., 2021).

## **2. Objetivos**

### **2.1. Objetivos Gerais**

Prever a sintomatologia depressiva com base em biomarcadores gerais de obesidade e diabetes, além de variáveis de sono e atividade física, em um público de adultos de meia idade e mais velhos, utilizando métodos de aprendizagem de máquinas.

## **2.2. Objetivos Específicos**

**I-** Predizer a ocorrência de sintomas depressivos em adultos de meia idade e idosos através de variáveis de sono, presença de SDE, nível de atividade física, medidas antropométricas, biomarcadores de obesidade e da diabetes, sob uma abordagem de Aprendizagem de Máquinas (MA).

**II-** Avaliar o potencial dos recursos previsores, em conjunto e individualmente, na previsão de ocorrência da sintomatologia depressiva.

## **3. Hipóteses e predições**

**Hipótese 1:** Modelos de aprendizagem de máquinas são eficientes para a prever a ocorrência da sintomatologia depressiva com base em variáveis preditoras relacionadas à hábitos de sono, atividade física e comorbidades (obesidade e diabetes). **Predição 1:** O modelo XGBoost possui melhor capacidade preditiva para a sintomatologia depressiva em relação aos modelos Random forest e regressão logística.

**Hipótese 2:** Sujeitos com presença e ausência de sintomatologia depressiva diferem quanto aos hábitos de sono e à frequência de ocorrência da SDE. **Predição 2:** Sujeitos com depressão dormem e despertam mais tardiamente durante a semana em relação aos indivíduos saudáveis e apresentam com maior frequência SDE.

**Hipótese 3:** Sujeitos com presença e ausência de sintomatologia depressiva diferem quanto às medições antropométricas de IMC e CC. **Predição 3:** O IMC e a CC são mais elevados em indivíduos com sintomas depressivos.

**Hipótese 4:** A ocorrência de sintomas depressivos é influenciada pelo sexo. **Predição 4:** Sujeitos com sintomatologia depressiva são em predominância representados pelo sexo feminino.

**Hipótese 5:** Sujeitos saudáveis e com sintomatologia depressiva diferem quanto aos níveis de PCR e glicohemoglobina. **Predição 5:** Os sujeitos que apresentam sintomas depressivos possuem níveis de PCR e glicohemoglobina superiores.

**Hipótese 6:** Existem diferenças nos níveis plasmáticos de HDL e CT entre sujeitos com e sem sintomas depressivos. **Predição 6:** Os indivíduos com sintomatologia depressiva apresentam medidas plasmáticas de HDL e CT inferiores.

**Hipótese 7:** A frequência para a ocorrência de atividade física recreativa (moderada ou vigorosa) difere entre sujeitos com presença e ausência de sintomatologia depressiva.

**Predição 7:** Sujeitos com sintomas depressivos costumam realizar com menor frequência atividade física de caráter recreativo em nível moderado a vigoroso.

**Hipótese 8:** O tempo gasto em comportamentos sedentários difere entre sujeitos saudáveis e com presença de sintomas depressivos. **Predição 8:** Indivíduos com sintomatologia depressiva gastam mais tempo em comportamentos considerados sedentários.



## Capítulo 1

### **Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: a machine learning approach**

**Autores:** Stephania Ruth Basilio Silva Gomes, Malcolm von Schantz, Mario André Leocadio-Miguel

**A ser submetido:** Sleep Medicine Journal (Fator de impacto: 4.84- 2021)

### **Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: a machine learning approach**

Stephania Ruth Basilio Silva Gomes <sup>a</sup>, Malcolm von Schantz <sup>b</sup>, Mario Leocadio-Miguel <sup>a</sup>

<sup>a</sup> Department of Physiology and Behavior, Federal University of Rio Grande Do Norte, Natal, Rio Grande do Norte, Brazil.

<sup>b</sup> Faculty of Health and Life Sciences, Northumbria University, Newcastle, UK.

#### **Highlights**

- Sleep data facilitates the prediction of depressive symptoms in middle-aged and older adults.
- Daytime sleepiness is an important predictor of depressive symptoms.
- XGBoost outperforms other machine learning (ML) models in the prediction of depressive symptoms.
- ML models to predict depressive symptoms facilitate the screening of depression.

#### **Abstract**

**Objectives:** Comorbid depression is a highly prevalent and debilitating condition in middle-aged and elderly adults, particularly when associated with obesity, diabetes, and sleep disturbances. In this context, there is a growing need to develop efficient screening methods for cases based on clinical health markers for these comorbidities and sleep data. Thus, our objective was to detect depressive symptoms in these subjects, considering general biomarkers of obesity and diabetes and variables related to sleep and physical exercise through a machine learning approach. **Methods:** A total of 2907 middle-aged and elderly subjects were eligible for the study. National Health and Nutrition Examination Survey (NHANES) 2015-2016 data were used and eighteen variables on self-reported physical activity, self-reported sleep habits, sleep disturbance indicative, anthropometric

measurements, sociodemographic characteristics and plasma biomarkers of obesity and diabetes were selected as predictors. Supervised learning algorithms such as Lasso penalized Logistic Regression (LR), Random forest (RF) and Extreme Gradient Boosting (XGBoost) were implemented. **Results:** XGBoost provided greater accuracy and precision (87%), with a proportion of hits in cases with depressive symptoms above 80%. In addition, daytime sleepiness was the most significant predictor variable for predicting depressive symptoms. **Conclusions:** Sleep and physical activity variables, in addition to obesity and diabetes biomarkers, together assume significant importance to predict, with accuracy and precision of 87%, the occurrence of depressive symptoms in middle-aged and elderly individuals.

**Keywords:** Depressive symptomatology; Obesity; Diabetes; Sleep variables; Physical activity; Machine Learning

### Abbreviations

AC

Accuracy

AUC

Area under curve

ACTH

Adrenocorticotrophic hormone

BMI

Body mass index

CRP

C-reactive protein

CORT

Corticosterone

EDS

Excessive daytime sleepiness

FA

Frequency absolute

FR

Frequency relative

GPAQ

Global physical activity questionnaire

HPA

Hypothalamic-pituitary-adrenal

HDL

High-density lipoprotein

Hb A1c

Glycated hemoglobin

LR

Logistic regression

MDD

Major depressive disorders

ML

Machine learning

NPV

Negative predictive value

NHANES

National Health and Nutrition Examination Survey

NCHS

National Center for Health Statistic

PC

Precision

PA

Physical activity

PHQ-9

Patient health questionnaire

PPV  
Positive predictive value

RF  
Random forest

SA  
Sedentary activity

SD  
Sleep duration

TC  
Total cholesterol

XGBoost  
Extreme gradient boosting

5-HT1A  
Serotonin receptors

## **1. Introduction**

Depression is one of the most prevalent mental disorders, affecting a third of a billion people in the world [1]. Importantly, the prevalence of depressive symptoms in middle-aged and elderly adults is high, affecting their quality of life and increasing healthcare spending [2]. The coexistence of multiple comorbidities is common among older adults and there is evidence of a bidirectional association between multimorbidity and depression. For example, the prevalence of depression increases by 37.97% in elderly and middle-aged individuals facing more than three chronic conditions when compared to healthy volunteers [3].

Obesity is an example of a chronic condition that not only shares a common biological mechanism, but also displays a bidirectional relationship with depression, particularly in women [4]. Anthropometric measurements, such as the Body Mass Index (BMI) and the Waist Circumference (WC), are widely adopted to assess general and abdominal obesity, respectively. Both BMI and WC are increased in subjects presenting depressive symptoms when compared to healthy individuals [5]. C-Reactive Protein (CRP), an indicator of peripheral inflammatory processes, is also an indirect marker of obesity with the potential to

predict depression. CRP serum concentration is higher in both obese and depressed individuals [6]. Moreover, in a predictive scenario, CRP has the ability to differentiate cases of unipolar and bipolar depression [7]. Furthermore, higher levels of CRP are significantly associated with the severity of depression in women [8].

In addition, the lipid profile, represented by high concentrations of total serum cholesterol (TC) and reduced levels of high-density lipoprotein (HDL), is an important marker of obesity in the metabolic syndrome, associated with a higher risk of developing cardiovascular diseases [9]. The same pattern in HDL levels was associated with long-term depression and long-term depressive symptomatology [10,11]. However, reduced levels of TC are associated with depressive symptoms, especially in middle-aged and elderly men [12,13].

Again, when considering the multicomorbid nature of the depressive disorder, researchers have shown that diabetes is also an important risk factor for the development of depression [14]. Moreover, an improvement in glycemic control from 10% to 25% is associated with a decrease in the prevalence of depression [15]. Finally, the elevation of glycated hemoglobin (HbA1c), frequently used as a biomarker of diabetes and pre-diabetes, also predicts depression [16].

Delayed sleep phase [17], instability in sleep onset and/or offset [18], and reduction of the amplitude of multiple circadian rhythms [19,20] are related to depression. In addition, individuals experiencing depression are also often concurrently affected by sleep disorders, such as excessive daytime sleepiness (EDS), which appears to be a common symptom of depression [21]. Interestingly, however, the reduction in nocturnal sleep duration, which is usually associated with EDS, seems to be irrelevant to predicting depression. However, a recent meta-analysis showed that short sleep duration, per se, is associated with the risk of depression [22]. Furthermore, reduced mean sleep duration can predict an increase in depressive symptoms when presented over a period of at least two weeks [23].

Importantly, EDS and insufficient sleep duration result in increased sedentary behavior and reduced physical activity (PA) levels. PA levels are reduced as a function of sleep duration in subjects with EDS [24]. Furthermore, a more sedentary lifestyle is associated with a higher risk of developing the major depressive disorder (MDD) and this relationship is mediated by sleep disorders in general [25]. In addition, subjects with depression tend to perform recreational physical activities less frequently and spend less time on moderate to vigorous activities [26].

Given the increasing difficulty and importance of detecting depression, particularly in middle-aged adults and the elderly, fast and accurate screening methods are needed [27]. Depression screening is usually based on questionnaires and self-report scales, covering questions about sociodemographic characteristics, risk factors and comorbid medical conditions associated with depression [28]. Moreover, interpretation issues and the lack of more objective clinical data may explain the low detection rate of depression found in primary care settings, for example [29]. In this scenario, the application of state of art statistics based on artificial intelligence, such as machine learning (ML), promises to revolutionize clinical practice in psychiatry by enabling the differential diagnosis of depression based on objective, ordinary clinical data [30]. Rather than solving inference problems and understanding the relationship between variables, ML algorithms can learn directly from data and search for complex, non-linear, patterns to enable prediction [31].

Therefore, the objective of the study was to predict the occurrence of depressive symptoms in middle-aged and elderly adults through sleep-related variables, levels of physical activity, anthropometric measures, obesity and diabetes biomarkers, using a machine learning approach (ML). In addition, we aim to evaluate the potential of multiple predictive variables, together and individually, in predicting the occurrence of depressive symptoms.

## **2. Methods**

### **Database**

Data was retrieved from the National Health and Nutrition Examination Survey (NHANES), biennial period (2015-2016). NHANES is a study program of the National Center for Health Statistics (NCHS) aimed at assessing the health and nutritional status of individuals in the United States, compiling data from a combination of questionnaires and clinical examinations, supporting researchers around the world with an interest in epidemiological studies and aimed at understanding public health problems [32].

### **Eligible Sample**

A total of 2907 middle-aged and elderly subjects, participants of NHANES (2015-2016), presenting depressive symptoms or not, were eligible for the present study.

### **Measures of depressive symptoms**

To assess the occurrence of depressive symptoms, we extracted responses from the Patient Health Questionnaire (PHQ-9). This instrument delimits different cut-off points to classify the degree of severity of depressive symptoms into different levels: no depression (0-4), mild (5-9), moderate (10-14), moderately severe (15-19) and severe (20-27) [33]. In the study, two cut-off points were determined to classify individuals according to the absence or presence of depressive symptoms. Thus, the established scoring ranges were 0-4 and 5-27, respectively.

### **Selection of predictor variables (features)**

The variable selection was guided by the literature review of depression and its comorbidities. In total, 18 variables were selected, distributed into 7 distinct resource axes: (1) self-reported physical activity (the occurrence of moderate or vigorous recreational activities and sedentary activity (SA)), (2) self-reported sleep habits (sleep onset, sleep offset and sleep duration (SD)), (3) indicative of sleep disturbance (excessive daytime sleepiness), (4) anthropometric measurements (weight, WC and BMI), (5) plasma biomarkers of obesity (C-reactive protein (CRP), total cholesterol (TC) and high-density lipoprotein (HDL), (6) biomarker of diabetes (glycated hemoglobin (HbA1c) and (7) sociodemographic characteristics (sex and age).

### **Measures of predictor variables**

The GPAQ - Global Physical Activity Questionnaire [34] provided data on the prevalence of physical activity in the recreational context, from moderate to vigorous level (at least 10 minutes in a typical week), as well as the time allocated in sedentary activities on a normal day, especially sitting and resting.

The assessment of sleep habits and EDS was established through general self-report questionnaires. Participants reported their usual bedtime, wake-up time and sleep duration on weekdays and workdays. EDS was based on the volunteer's perception of the frequency of diurnal excessive sleepiness in the last 30 days. Through a reference scale the EDS was classified as: never, rarely (once a month), sometimes (2-4 times a month), often (5-15 times a month) and almost always (16-30 times a month).

Data on the age and sex of participants were acquired from a sociodemographic survey. The anthropometric variables included weight (kg), waist circumference (cm) and BMI (kg/m<sup>2</sup>).

To assess plasma biomarkers, such as high-sensitivity CRP, CT and HDL, minimum blood detection limits were determined: 0.11 mg/l, 3mg/dl and 4mg/dl, respectively. Glycated hemoglobin data was obtained from accurate measurements of its stable form (HbA1c) in relative percentages. More details on data collection and processing can be found in the NHANES website (<https://www.cdc.gov/nchs/nhanes>).

### **Descriptive and inferential analysis of the eligible sample**

Data were submitted to a descriptive statistical routine, considering measures of central tendency (mean and median) and standard deviation for quantitative variables. Furthermore, the normality of these features was verified using the Shapiro-Wilk test. Absolute and relative frequencies for all levels of nominal categorical predictor variables were calculated, as well as for the binary target class of prediction.

Spearman's correlation test was applied to assess the existence of a linear association between predictor variables related to sleep habits, anthropometric measurements, obesity/diabetes plasma biomarkers, in addition to age and the level of depressive symptoms. To investigate whether the groups of participants with and without depressive symptoms differ in terms of the frequency of daytime sleepiness, physical activity at different levels in a recreational context, and gender, the chi-square test of association (independence) was used and the frequency of observed and expected participants at each attribute level was calculated, assuming independence between these variables. In addition, the same test was used to compare possible differences in sleep onset and sleep offset time frequencies between the groups.

To assess whether individuals with and without depressive symptoms vary in terms of age, BMI, TC, CRP, HbA1c, HDL, SA, SD, sleep onset, sleep offset, WC and weight, the Kruskal-Wallis H test was applied.

### **Supervised machine learning in predicting depressive symptoms**

To predict the occurrence of depressive symptoms, based on the predictive features, we developed different supervised machine learning models following four steps [35,36,37]: (1) data pre-processing, (2) implementation of classification algorithms, (3) model validation and (4) evaluation of predictive performance (Fig. 1).

#### ***Data pre-processing***



This step requires the treatment of missing variables and we decided to remove every row with unknown values. We substituted qualitative predictor variables by creating dummy variables, represented by binary values of 0 and 1. Also, we performed normalization and standardization processes for adjustments related to the scale of each variable.

We assessed the collinearity between predictors using Spearman's correlation test (the threshold for exclusion:  $\rho > 0.80$ ). Problems with the imbalance in the distribution of classes were solved by oversampling the minority class, where new instances were randomly produced by duplication, based on the input data, according to the methodology used by the Data Sampler function in the Orange software (version 3.29.3).

### ***Classification Algorithms***

We implemented Lasso penalized Logistic Regression (LR), Random forest (RF) and Extreme Gradient Boosting (XGBoost). These are supervised learning algorithms often used to solve classification problems. All implementations were performed in Orange software.

### ***Data Partitioning***

The final dataset consisted of 18 predictor variables and the outcome labels (presence (1) and absence (2) of depressive symptoms) that was partitioned into two data subsets for training and testing the models (Fig.1), 66% training data the 34% reserved to test the final performance of the algorithms and verify their respective generalization capacity after the insertion of new data.

### ***Training and Validation***

Validation with the training data was performed to adjust the combination of the best hyperparameters, necessary to control the complexity of the models (overfitting) and to optimize their performance, maintaining the balance between bias and variance. We used a 10-fold cross-validation method (k-folds) to validate the predictive performance of our models (Fig. 1).

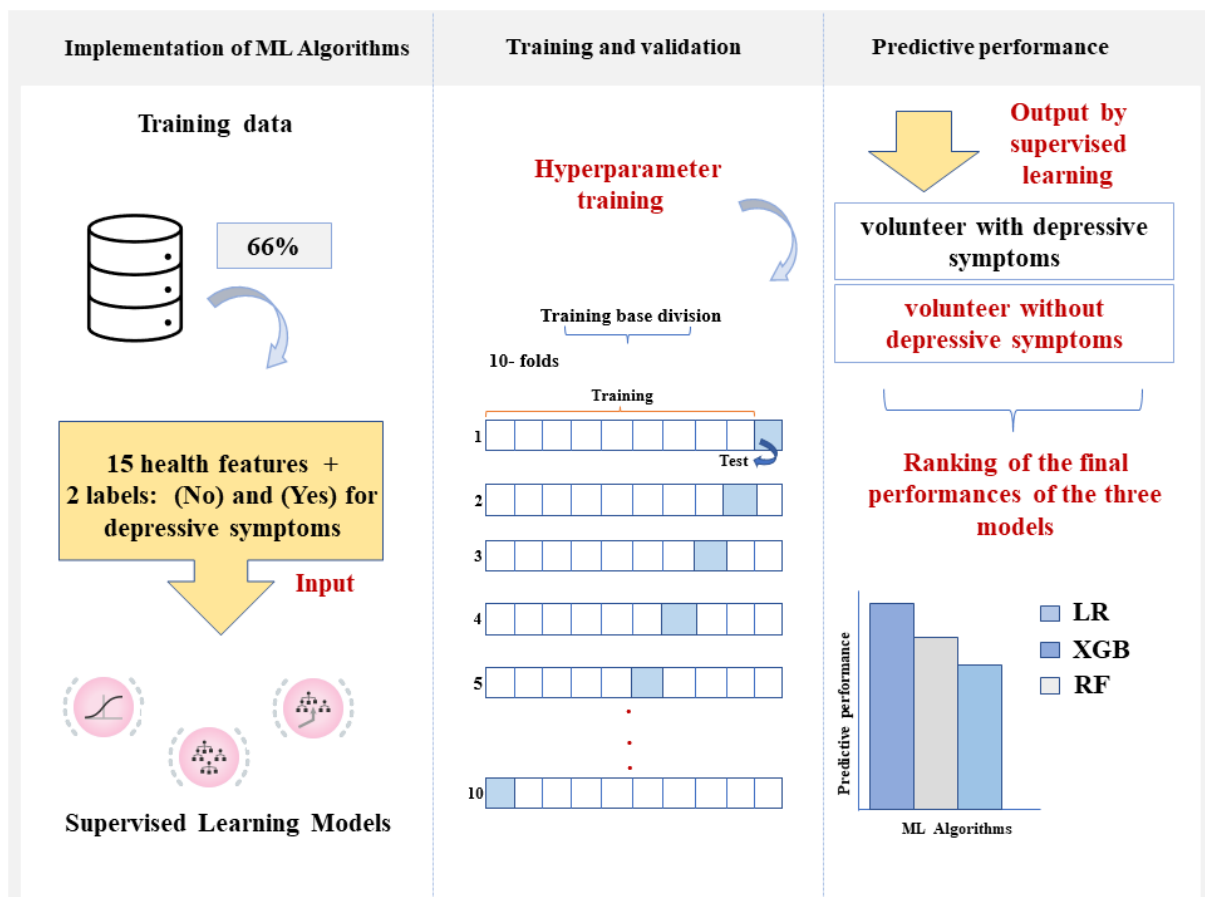
### ***Predictive performance***

To assess the predictive performance of each model, we used: Area Under the Curve (AUC), Accuracy (AC), Precision (PC), Sensitivity, Specificity, F1 Score, and Recall (RC). Furthermore, we created confusion matrices to estimate the performance of the classifiers considering the different classes, enabling us to extract the Positive Predictive Value (PPV)

and the Negative Predictive Value (NPV). The PPV was calculated as the number of cases that actually have depressive symptoms, divided by the number of cases that the model classifies as having this condition. The NPV was calculated as the number of cases that actually show the absence of depressive symptoms, divided by the number of cases that the classifier identifies as healthy. The percentage of the individual proportion of predicted classes was also verified.

### Variable importance

To assess the importance of the predictors in the detection of depressive symptoms, the "information gain" of each respective variable was used as a parameter. Gain refers to the improvement of accuracy in a particular branch of the decision tree in which the variable in question is inserted.



**Figure 1** - Schematic representation of the methodology used to implement supervised machine learning models to predict depressive symptoms.

## 3. Results

### Descriptive and inferential analysis of the eligible sample

### *Descriptive analysis*

The sample consisted of 2,907 eligible subjects, most of whom were women (51.5%) and with a mean age of 59.4 years. According to the PHQ-9 score, 25.1% were labelled with 'yes', representing the presence of these symptoms (Table 2). Additional information about the descriptive statistics of the other predictor variables can be accessed in tables 1 and 2.

**Table 1** - Summary of descriptive statistics for all numerical predictors.

<b>Feature</b>	<b>Mean</b>	<b>Median</b>	<b>Standard deviation</b>
Age	59.4	59	11.9
Body Mass Index (kg/m <sup>2</sup> )	29.9	28.9	6.7
Total Cholesterol (mg/dl)	195	193	42.2
C-Reactive Protein (mg/l)	4.4	2.1	8.4
Glycohemoglobin (%)	6.06	5.7	1.2
High Density Lipoprotein (mg/dl)	54.6	51	18.2
Sedentary Activity (min.)	365	360	198
Sleep Duration (min.)	460	480	90.7
Sleep Onset (h)	22:45	22:30	118 min.
Sleep Offset (h)	06:01	06:00	155 min.
Waist Circumference (cm)	103	101	15.6
Weight (kg)	82.3	79.4	20.5

**Table 2** - Absolute and relative frequency of levels referring to all predictive categorical variables and the target class of the prediction.

<b>Name</b>	<b>Role</b>	<b>Levels</b>	<b>AF</b>	<b>RF(%)</b>
Gender	feature	female	1497	51.5%
		male	1410	48.5%
Vigorous Recreational Activity	feature	no	2405	82.7%
		yes	502	17.3%
Moderate Recreational Activity	feature	no	1741	59.9%
		yes	1166	40.1%
Daytime Sleepiness	feature	sometimes	927	31.9%

		rarely	711	24.4%
		never	580	19.9%
		often	464	15.9%
		almost always	225	7.7%
Depressive symptoms	target	no	2176	74.9%
		yes	731	25.1%

Note. (FA)= Frequency Absolute and (FR)= Frequency Relative in percentage.

### ***Inferential analysis***

When considering the relationship between the PHQ-9 score and the quantitative predictor variables, a positive correlation was observed between the PHQ-9 score and BMI ( $\rho = 0.137$ ;  $p < 0.05$ ), PHQ-9 score and WC ( $\rho = 0.14$ ;  $p < 0.05$ ) and between PHQ-9 score and CRP ( $\rho = 0.125$ ;  $p < 0.05$ ). The other variables showed a correlation coefficient below 0.1. Thus, these findings suggest a non-linear interaction between the predictor variables in question and the PHQ-9 score, which represents the severity of depressive symptoms (Appendix fig. 1).

We observed an association between depressive symptoms and daytime sleepiness ( $X^2(4) = 279$ ;  $p < 0.05$ ), vigorous recreational activity ( $X^2(1) = 23.9$ ;  $p < 0.05$ ), moderate recreational activity ( $X^2(1) = 29.4$ ;  $p < 0.05$ ) and sex ( $X^2(1) = 22.6$ ;  $p < 0.05$ ), evidencing a relationship of dependence between the variables. According to the contingency table (Table S1), we observed more participants with depressive symptoms within the groups of individuals classified as experiencing daytime sleepiness (almost always or often) and more participants without depressive symptoms in the groups of individuals who never presented daytime sleepiness or rarely, when compared to the expected values. Still, in contrast to the expected values, we observed more men in the group of participants without depressive symptoms and more women in the group with such symptoms. For vigorous and moderate recreational activity, we observed more people who practiced these activities in the group of individuals classified with no symptoms and more participants who did not perform them in the group characterized with the presence of symptoms. Furthermore, participants who presented depressive symptoms spent more time in sedentary activities in relation to individuals without depressive symptoms (Table 3), 60 minutes longer (median=360 min) in relation to the group without depressive symptoms.

Depressive symptoms were also associated with sleep onset ( $X^2(52) = 104$ ;  $p < 0.05$ ) and sleep offset ( $X^2(77) = 142$ ;  $p < 0.05$ ). We observed a higher frequency of participants belonging to the group with symptoms initiating sleep later, as well as waking up later, in comparison with those without symptoms.

**Table 3** - Differences between the medians of the characteristics of the groups, verified by the Kruskal-Wallis H test.

Characteristics	Group without depressive symptoms (median)	Group with depressive symptoms (median)	$X^2$	df	$p$ - value
Age	59	59	0.006	1	0.938
BMI (kg/m <sup>2</sup> )	28.6	30.2	33.65	1	< .001
TC (mg/dl)	192	193	0.04	1	0.828
CRP (mg/l)	1.9	2.7	31.3	1	< .001
HbA1c (%)	5.7	5.8	17.9	1	< .001
HDL (mg/dl)	52	50	9.3	1	0.002
SA (min.)	300	360	9.13	1	0.003
SD (h)	7:30	8:00	6.3	1	0.012
Sleep on. (min.)	1350	1350	5.5	1	0.018
Sleep off. (min.)	1800	1800	13.9	1	< .001
WC (cm)	100	104	33.5	1	< .001
Weight (kg)	78.6	81.7	11.2	1	< .001

Note. (BMI)- Body Mass Index; (TC)- Total cholesterol; (CRP)- C-Reactive Protein; (HbA1c)- Glycohemoglobin; (HDL)- High-Density Lipoprotein; (SA)- Sedentary Activity; (SD)- Sleep Duration; (Sleep on)- Sleep onset; (Sleep off)- Sleep offset and (WC)- Waist Circumference.

Regarding the biomarkers for obesity, BMI ( $X^2(1)= 33.65$ ;  $p<0.05$ ), CRP ( $X^2(1)= 31.3$ ;  $p<0.05$ ), WC ( $X^2(1) = 33.5$ ;  $p<0.05$ ), HDL ( $X^2(1)= 9.3$ ;  $p<0.05$ ) and Weight ( $X^2(1)= 11.2$ ;  $p<0.05$ ) differed significantly between groups (Table 3). No significant differences were observed in the serum total cholesterol level between the groups (Table 3).

Sleep habits also differed significantly between groups (Table 3). Individuals with depressive symptoms went to sleep 16 minutes earlier (10:57 pm, on average) and woke up 47 minutes later (at 6:11 am, on average). In addition, on average, the group with depressive symptoms slept 30 minutes longer than the volunteers without depressive symptoms.

## Supervised machine learning in predicting depressive symptoms

### *Data pre-processing*

After adjusting the imbalance between the target classes, the final sample number increased from 2907 to 3638 volunteers, with the insertion of new instances. The variables

related to anthropometric measurements (weight, BMI, and WC) were the only ones that showed a high correlation with each other, with  $\rho$  above 0.8 (Figure S1). Among them, weight was excluded from the dataset, as studies indicate that both general obesity, assessed by the BMI, and abdominal obesity, assessed by the WC, are the most important metrics, both having an important association with depressive symptoms in middle-aged adults and elderly [38]. Thus, the final dataset had 15 predictor variables, after filtering out highly correlated predictors (Table S2).

### *Definition of hyperparameters*

**Table 4** - Hyperparameter definition with 10-fold cross-validation.

Model	Hyperparameter	Fixed values (All features)
<b>XGBoost</b>	Maximum tree depth	6
	Boosting learning rate	0.3
	Number of trees	100
	Regularization strength	1
	Fraction of training instances	1
	Fraction of features for each tree	1
	Fraction of features for each level	1
	Fraction of features for each split	1
<b>Random Forest</b>	Number of trees	10
	Maximal number of considered features	7
	Maximal tree depth	18
	Stop splitting nodes with maximum instances	5
<b>Logistic Regression</b>	Regularization type	Lasso (L1)
	Regularization strength	1

### *Comparison of the three models*

XGBoost was the one with the highest predictive performance, according to the main evaluation metrics (AUC= 0.92; Accuracy= 0.87 and Precision= 0.87) (Table 5). Furthermore, this was the model that best managed to predict cases considered truly positive for depressive symptoms, with a PPV equal to 0.80, when compared to RF (0.77) and LR (0.64) (Table 4). Moreover, XGBoost and Random Forest had similar proportions of correct answers in classifying cases for the presence of depressive symptoms, with 80.6% and 77%, respectively (Fig. 2).

**Table 5** - Evaluation metrics.

Model	AUC	AC	PC	Recall	F1	PPV	NPV
XGBoost	0.92	0.87	0.87	0.87	0.87	0.80	0.92
Random Forest	0.89	0.82	0.83	0.82	0.83	0.77	0.87
Logistic Regression	0.71	0.67	0.67	0.67	0.66	0.64	0.68

Note. (AUC)- Area under the curve; (AC)- Accuracy; (PC)- Precision; (F1)- F1 Score; (PPV)- Positive Predictive Value and (NPV)- Negative Predictive Value.

		Predicted		
		no	yes	$\Sigma$
A	Actual no	92.9%	19.4%	2176
	Actual yes	7.1%	80.6%	1462
	$\Sigma$	2001	1637	3638

		Predicted		
		no	yes	$\Sigma$
B	Actual no	87.3%	23.0%	2176
	Actual yes	12.7%	77.0%	1462
	$\Sigma$	2083	1555	3638

		Predicted		
		no	yes	$\Sigma$
C	Actual no	68.9%	35.7%	2176
	Actual yes	31.1%	64.3%	1462
	$\Sigma$	2639	999	3638

**Figure 2** - Confusion matrix of tested models with the predicted proportion for each class. In (a): XGBoosting; (b) Random Forest and (c) Logistic Regression.**Variable importance**

The most significant predictive variable for depressive symptoms was daytime sleepiness when compared to the other features in descending order of importance (Table 6).

**Table 6** - Importance of the variable according to information gain. (#)- stands for the number of levels of a given categorical variable.

Attributes	#	Information gain
Daytime Sleepiness	5	0.079
Waist Circumference		0.013

Body Mass Index		0.012
Sleep onset		0.011
Sleep offset		0.011
Sleep Duration		0.011
C-Reactive Protein		0.010
Moderate Recreational Activity	2	0.010
Vigorous Recreational Activity	2	0.008
Gender		0.007
Glycohemoglobin		0.006
High-Density Lipoprotein		0.004
Age		0.003
Sedentary Activity		0.003
Total Cholesterol		0.002

---

#### 4. Discussion

Recently, machine learning has been successfully applied to predict depression, from the diagnosis to treatment response, through the use of multimodal wearable devices [39], 24-h locomotor activity data [40], and predictive features extracted from neuroimaging data [41]. However, none of these studies is solely based on ordinary, widely available clinical data.

Here, we revealed that excessive daytime sleepiness is the most important feature to predict the presence of depressive symptoms in middle-aged and elderly Americans. Community-based studies focused on the elderly population of different countries, which investigated the association or prevalence of EDS in depression, support our findings by showing a significant association between EDS, depressive symptoms and the severity of depression [42]. Also, an increased risk of depression is related to daytime sleepiness compared to other sleep-wake disorders and medical conditions such as obesity and diabetes [43]. Furthermore, longitudinal research revealed that a decrease in depressive symptoms is one of the main predictors of remission of EDS [44].

EDS is characterized by the difficulty of maintaining wakefulness and alertness during the daytime, being a cause of traffic and workplace accidents, leading to damage to life and health [45]. It partially results from chronic sleep deprivation [46], unrested and insufficient [47]. Particularly in older adults, fragmented sleep, insomnia, obstructive sleep apnea and other sleep-related disorders also reduce total sleep time [48].



Chronic sleep deprivation affects the serotonergic system and the hypothalamic-pituitary-adrenal (HPA) axis responsible for the physiological response to stress, whose action causes changes in serotonin receptors (5-HT<sub>1A</sub>) and in the neuroendocrine response of adrenocorticotrophic hormone (ACTH) and corticosterone (CORT) release [49]. These alterations are similar to the underlying neuronal processes of depression, which present hyperactivity of the HPA axis marked by the hypersecretion of ACTH and glucocorticoids [50] and a decrease in the availability of serotonin [51]. Animal model studies described that sleep deprivation for up to eight days causes gradual postsynaptic desensitization in the serotonin receptor (5-HT<sub>1A</sub>), and this persists even after sleep recovery [52]. Moreover, chronic sleep deprivation sensitized the HPA axis, resulting in elevated levels of ACTH and CORT after every 20-hour period of sleep deprivation for seven days [53].

When evaluating total self-reported sleep duration, we observed that individuals with symptoms of depression slept on average longer than those with no symptoms. Most studies, however, have reported a strong association between short sleep duration and depression [22], with participants suffering from depression reporting up to 40 minutes less sleep [54]. The association between the level of depressive symptoms and sleep duration in the elderly seems to be non-linear [55], although extremely short sleep duration (less than 4 hours of sleep) is associated with a higher risk of presenting symptoms of depression. Other studies including middle-aged and elderly participants, emphasized that short sleep duration is a risk factor for both incident and recurrent depression, with equal risk probability for both age groups [56]. On the other hand, long sleep duration ( $\geq 9$ ) in the elderly is associated with EDS, which may reflect poor sleep quality in these individuals (57).

In addition to an increased sleep duration in symptomatic participants, they also exhibited a delay in sleep onset and offset when compared with asymptomatic participants. This result is consistent with the literature and a typical finding in depression [58].

Higher BMI and WC were observed in our sample of individuals with depressive symptoms, compared to healthy volunteers, which is in accordance with the literature. American women have been reported to show a greater probability of the occurrence of depressive symptoms when their BMI is above 30kg/m<sup>2</sup>. Moreover, a recent meta-analysis study reported that volunteers with obesity are 1.18 times more likely to manifest depressive symptoms than those with a healthy weight range [59]. We also found lower levels of HDL in individuals with depressive symptoms compared to those without symptoms, despite being within the normal range for the age group. By contrast, there was no significant difference between groups in the levels of total cholesterol, with both having levels within the range

considered normal. A growing body of evidence supports that low levels of HDL are associated with the occurrence of depressive symptoms [10,11]. Other studies have shown that reduced levels of total cholesterol are not associated with depression and more severe symptoms in the elderly [60,61]. However, there are studies showing, in both middle-aged and elderly men, an association between reduced levels of TC and depressive symptomatology, representing an increased risk of developing it [12,13].

Significant differences between groups were also observed in the level of C-reactive protein (CRP). Several studies have reported results supporting our findings by showing an association between increasing levels of plasma CRP with an increased risk of depression [62]. Moreover, higher levels of CRP have been significantly associated with greater severity of depressive symptoms [8]. However, it is important to emphasize that the mean CPR values verified in both groups in the present study suggest systemic inflammation, which demonstrates how the general population shows signs of inflammation.

Physical activities of a recreational nature, from moderate to vigorous intensity levels, were less frequent among subjects with depressive symptoms. Similar results were obtained in a study that sought to characterize groups with depression through cluster solutions, evidencing a propensity of depressive subjects to avoid moderate recreational activities [63]. There is evidence among middle-aged and older adults that recreational physical activities, both vigorous and moderate, contribute to a decrease in more severe depression [64]. For instance, physical activity is generally associated with a reduced proportion of subjects exceeding a PHQ-9 score  $> 10$ , which indicates greater severity of symptoms of depression [65]. These results are in line with expectations since it has been shown that depressive patients avoid recurrent practices of moderate to vigorous physical activities, dedicating less time to them when compared with subjects without depression [26]. This trend is even more frequent among older individuals, with a higher risk of low physical activity [66]. It can be explained mainly by limitations in mobility, pain and discomfort. In addition, the association between low levels of physical activity and depression is also due to somatic conditions related to increased functional limitations, increased use of medication and psychosocial factors in individuals over 60 years of age [67].

Furthermore, sedentary behavior was prevalent in the group with symptoms of depression. These data corroborate results that show an association between depression and decreased minutes of physical activity [68]. Furthermore, even the chance of developing depression is increased by sedentarism [69].

Some important limitations in the present study are worth mentioning. For example, only weekday data for self-reported sleep onset, sleep offset and sleep duration data were available, which prevented us to analyse the instability in the sleep-wake behavior between weekdays and weekends. In addition, we did not have access to objective data, such as actigraphy, which would provide more accurate information about both the sleep-wake cycle and an estimate of the total time spent during the week in moderate to vigorous physical activity.

## **5. Conclusion**

The variables of sleep and physical activity, in addition to the biomarkers of obesity and diabetes, jointly assume significant importance to predict, with accuracy and precision of 87%, the occurrence of depressive symptoms in middle-aged and elderly individuals. Daytime sleepiness has greater power to predict depressive symptoms compared to other predictors. Individuals with depressive symptoms often experience excessive daytime sleepiness, tend to fall asleep and wake up later, have higher BMI and WC, have poorer glycemic control, and avoid moderate and vigorous recreational physical activity. Compared to psychometric analysis tools that require specialisation and proper time, machine learning predictive models facilitate the screening of depressive symptoms in elderly and middle-aged patients suffering from depression, minimizing morbidity and mortality through early treatment.

## **Funding sources**

This work was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) to Stephania Ruth Basilio Silva Gomes (Postgraduate scholarship) and Mario Leocadio-Miguel (CAPES-PRINT 88887.465857/2019-00).

## **Declaration of competing interest**

The authors have no conflicts of interest to disclose.

## **CRedit authorship contribution statement**

**Stephania Ruth Basilio Silva Gomes:** Contributed to the study design, formal data analysis, interpretation and writing of the manuscript. **Malcolm von Schantz:** Contributed to the writing and critical review of the manuscript. **Mario Leocadio-Miguel:** Contributed to the

design and supervision of the study, in addition to the interpretation and writing of the manuscript.

## References

- [1] World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.
- [2] Mojtabai, R., & Olfson, M. (2004). Major depression in community-dwelling middle-aged and older adults: prevalence and 2-and 4-year follow-up symptoms. *Psychological medicine*, 34(4), 623-634. <https://doi.org/10.1017/S0033291703001764>
- [3] Jiang, C. H., Zhu, F., & Qin, T. T. (2020). Relationships between Chronic Diseases and Depression among Middle-aged and Elderly People in China: A Prospective Study from CHARLS. *Current Medical Science*, 40(5), 858-870. <https://doi.org/10.1007/s11596-020-2270-5>
- [4] Milaneschi, Y., Simmons, W. K., van Rossum, E. F., & Penninx, B. W. (2019). Depression and obesity: evidence of shared biological mechanisms. *Molecular psychiatry*, 24(1), 18-33. <https://doi.org/10.1038/s41380-018-0017-5>
- [5] Zhao, G., Ford, E. S., Li, C., Tsai, J., Dhingra, S., & Balluz, L. S. (2011). Waist circumference, abdominal obesity, and depression among overweight and obese US adults: National Health and Nutrition Examination Survey 2005-2006. *BMC psychiatry*, 11(1), 1-9. <https://doi.org/10.1186/1471-244X-11-130>
- [6] Visser, M., Bouter, L. M., McQuillan, G. M., Wener, M. H., & Harris, T. B. (1999). Elevated C-reactive protein levels in overweight and obese adults. *Jama*, 282(22), 2131-2135. <https://doi.org/10.1001/jama.282.22.2131>
- [7] Poletti, S., Vai, B., Mazza, M. G., Zanardi, R., Lorenzi, C., Calesella, F., Cazzetta, S., Branchi, I., Colombo, C., Furlan, R., & Benedetti, F. (2020). A peripheral inflammatory signature discriminates bipolar from unipolar depression: A machine 15 learning approach. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 105, 110136. <https://doi.org/10.1016/j.pnpbp.2020.110136>
- [8] Köhler-Forsberg, O., Buttenschøn, H. N., Tansey, K. E., Maier, W., Hauser, J., Dernovsek, M. Z., ... & Mors, O. (2017). Association between C-reactive protein (CRP) with depression symptom severity and specific depressive symptoms in major depression. *Brain, behavior, and immunity*, 62, 344-350. <https://doi.org/10.1016/j.bbi.2017.02.020>
- [9] Wang, H., & Peng, D. Q. (2011). New insights into the mechanism of low high-density lipoprotein cholesterol in obesity. *Lipids in health and disease*, 10(1), 1-10. <https://doi.org/10.1186/1476-511X-10-176>
- [10] Lehto, S. M., Hintikka, J., Niskanen, L., Tolmunen, T., Koivumaa-Honkanen, H., Honkalampi, K., & Viinamäki, H. (2008). Low HDL cholesterol associates with major depression in a sample with a 7-year history of depressive symptoms. *Progress in Neuro-*

*Psychopharmacology and Biological Psychiatry*, 32(6), 1557-1561.  
<https://doi.org/10.1016/j.pnpbp.2008.05.021>

[11] Lehto, S. M., Niskanen, L., Tolmunen, T., Hintikka, J., Viinamäki, H., Heiskanen, T., ... & Koivumaa-Honkanen, H. (2010). Low serum HDL-cholesterol levels are associated with long symptom duration in patients with major depressive disorder. *Psychiatry and clinical neurosciences*, 64(3), 279-283. <https://doi.org/10.1111/j.1440-1819.2010.02079.x>

[12] Morgan, R. E., Palinkas, L. A., Barrett-Connor, E. L., & Wingard, D. L. (1993). Plasma cholesterol and depressive symptoms in older men. *The Lancet*, 341(8837), 75-79. [https://doi.org/10.1016/0140-6736\(93\)92556-9](https://doi.org/10.1016/0140-6736(93)92556-9)

[13] Steegmans, P. H., Hoes, A. W., Bak, A. A., van der Does, E., & Grobbee, D. E. (2000). Higher prevalence of depressive symptoms in middle-aged men with low serum cholesterol levels. *Psychosomatic medicine*, 62(2), 205-211. <https://doi.org/10.1097/00006842-200003000-00009>

[14] Rotella, F., & Mannucci, E. (2013). Diabetes mellitus as a risk factor for depression. A meta-analysis of longitudinal studies. *Diabetes research and clinical practice*, 99(2), 98-104. <https://doi.org/10.1016/j.diabres.2012.11.022>

[15] Chireh, B., Li, M., & D'Arcy, C. (2019). Diabetes increases the risk of depression: a systematic review, meta-analysis and estimates of population attributable fractions based on prospective studies. *Preventive medicine reports*, 14, 100822. <https://doi.org/10.1016/j.pmedr.2019.100822>

[16] Almutairi, I., Al-Rashdi, M., & Almutairi, A. (2020). Prevalence and predictors of depression, anxiety and stress symptoms in paramedics at Saudi Red Crescent Authority. *Saudi Journal of Medicine & Medical Sciences*, 8(2), 105. [https://doi.org/10.4103/sjmms.sjmms\\_227\\_18](https://doi.org/10.4103/sjmms.sjmms_227_18)

[17] Au, J., & Reece, J. (2017). The relationship between chronotype and depressive symptoms: a meta-analysis. *Journal of affective disorders*, 218, 93-104. <https://doi.org/10.1016/j.jad.2017.04.021>

[18] Cho, C. H., Lee, T., Kim, M. G., In, H. P., Kim, L., & Lee, H. J. (2019). Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *Journal of medical Internet research*, 21(4), e11029. <https://doi.org/10.2196/11029>

[19] Souêtre, E., Salvati, E., Belugou, J. L., Pringuey, D., Candito, M., Krebs, B., Ardisson, J-L., & Darcourt, G. (1989). Circadian rhythms in depression and recovery: evidence for blunted amplitude as the main chronobiological abnormality. *Psychiatry research*, 28(3), 263-278. [https://doi.org/10.1016/0165-1781\(89\)90207-2](https://doi.org/10.1016/0165-1781(89)90207-2)

[20] Vadnie, C. A., & McClung, C. A. (2017). Circadian rhythm disturbances in mood disorders: insights into the role of the suprachiasmatic nucleus. *Neural plasticity*, 2017. <https://doi.org/10.1155/2017/1504507>

- [21] Hawley, C. J., Gale, T. M., Sivakumaran, T., Paul, S., Kondan, V. R. G., Farag, A., & Shahzad, J. (2010). Excessive daytime sleepiness in psychiatric disorders: Prevalence, correlates and clinical significance. *Psychiatry research*, 175(1-2), 138-141. <https://doi.org/10.1016/j.psychres.2008.10.037>
- [22] Zhai, L., Zhang, H., & Zhang, D. (2015). Sleep duration and depression among adults: A meta-analysis of prospective studies. *Depression and anxiety*, 32(9), 664-670. <https://doi.org/10.1002/da.22386>
- [23] Hamilton, J. L., Stange, J. P., Burke, T. A., Franzen, P. L., & Alloy, L. B. (2019). Sleep disturbance and physiological regulation among young adults with prior depression. *Journal of psychiatric research*, 115, 75-81. <https://doi.org/10.1016/j.jpsychires.2019.05.016>
- [24] McClain, J. J., Lewin, D. S., Laposky, A. D., Kahle, L., & Berrigan, D. (2014). Associations between physical activity, sedentary time, sleep duration and daytime sleepiness in US adults. *Preventive medicine*, 66, 68-73. <https://doi.org/10.1016/j.ypmed.2014.06.003>
- [25] Hallgren, M., Vancampfort, D., Owen, N., Rossell, S., Dunstan, D. W., Bellocco, R., & Lagerros, Y. T. (2020). Prospective relationships of mentally passive sedentary behaviors with depression: mediation by sleep problems. *Journal of affective disorders*, 265, 538-544. <https://doi.org/10.1016/j.jad.2019.11.088>
- [26] Schuch, F., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P., Reichert, T., Bagatini, N. C., Bgeginski, R., & Stubbs, B. (2017). Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. *Journal of affective disorders*, 210, 139-150. <https://doi.org/10.1016/j.jad.2016.10.050>
- [27] Richardson, L., & Brahmabhatt, A. (2021). Depression in Primary Care. *The Journal for Nurse Practitioners*, 17(1), 37-43. <https://doi.org/10.1016/j.nurpra.2020.09.002>
- [28] Maurer, D. M., Raymond, T. J., & Davis, B. N. (2018). Depression: screening and diagnosis. *American family physician*, 98(8), 508-515.
- [29] Craven, M. A., & Bland, R. (2013). Depression in primary care: current and future challenges. *The Canadian Journal of Psychiatry*, 58(8), 442-448. <https://doi.org/10.1177/070674371305800802>
- [30] Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223- 230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- [31] Baştanlar, Y., & Özuysal, M. (2014). Introduction to machine learning. *miRNomics: MicroRNA Biology and Computational Analysis*, 105-128. [https://doi.org/10.1007/978-1-62703-748-8\\_7](https://doi.org/10.1007/978-1-62703-748-8_7)
- [32] Johnson, C. L., Dohrmann, S. M., Burt, V. L., & Mohadjer, L. K. (2014). *National health and nutrition examination survey: sample design, 2011-2014 (No. 2014)*. US

Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

[33] Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606-613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>

[34] Armstrong, T., & Bull, F. (2006). Development of the world health organization global physical activity questionnaire (GPAQ). *Journal of Public Health*, 14(2), 66-70. <https://doi.org/10.1007/s10389-006-0024-x>

[35] Stefanowski, J., & Wilk, S. (2008). Selective pre-processing of imbalanced data for improving classification performance. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 283-292). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-85836-2\\_27](https://doi.org/10.1007/978-3-540-85836-2_27)

[36] Shrivastava, H., & Sridharan, S. (2013). Conception of data preprocessing and partitioning procedure for machine learning algorithm. *International Journal of Recent Advances in Engineering & Technology (IJRAET)*, 1(3), 2347-2812.

[37] Tohka, J., & van Gils, M. (2021). Evaluation of machine learning algorithms for Health and Wellness applications: a tutorial. *Computers in Biology and Medicine*, 104324. <https://doi.org/10.1016/j.compbiomed.2021.104324>

[38] Mulugeta, A., Zhou, A., Power, C., & Hyppönen, E. (2018). Obesity and depressive symptoms in mid-life: a population-based cohort study. *BMC psychiatry*, 18(1), 1-10. <https://doi.org/10.1186/s12888-018-1877-6>

[39] Tazawa, Y., Liang, K. C., Yoshimura, M., Kitazawa, M., Kaise, Y., Takamiya, A., Kishi, A., Horigome, T., Mitsukura, Y., Mimura, M., & Kishimoto, T. (2020). Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Helvion*, 6(2), e03274. <https://doi.org/10.1016/j.helivon.2020.e03274>

[40] Rodríguez-Ruiz, J. G., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., ... & Soto-Murillo, M. A. (2020). Comparison of night, day and 24 h motor activity data for the classification of depressive episodes. *Diagnostics*, 10(3), 162. <https://doi.org/10.3390/diagnostics10030162>

[41] Shi, Y., Zhang, L., Wang, Z., Lu, X., Wang, T., Zhou, D., & Zhang, Z. (2021). Multivariate machine learning analyses in identification of major depressive disorder using resting-state functional connectivity: A multicenter study. *ACS Chemical Neuroscience*, 12(15), 2878-2886. <https://doi.org/10.1021/acscchemneuro.1c00256>

[42] Lima, C. A., Soares, W. J. D. S., Bilton, T. L. Dias, R. C., Ferrioll, E., & Perracini, M. R. (2015). Correlates of excessive daytime sleepiness in community-dwelling older adults: an exploratory study. *Revista Brasileira de Epidemiologia*, 18, 607-617. <https://doi.org/10.1590/1980-5497201500030007>



- [43] Lai, H. C., Hsu, N. W., Chou, P., & Chen, H. C. (2020). The associations between various sleep-wake disturbances and depression in community-dwelling older adults-the Yilan study, Taiwan. *Aging & mental health*, 24(5), 717-724. <https://doi.org/10.1080/13607863.2019.1582006>
- [44] Jaussent, I., Morin, C. M., Ivers, H., & Dauvilliers, Y. (2020). Natural history of excessive daytime sleepiness: a population-based 5-year longitudinal study. *Sleep*, 43(3), zsz249. <https://doi.org/10.1093/sleep/zsz249>
- [45] Zwahlen, D., Jackowski, C., & Pfäffli, M. (2016). Sleepiness, driving, and motor vehicle accidents: a questionnaire-based survey. *Journal of forensic and legal medicine*, 44, 183-187. <https://doi.org/10.1016/j.jflm.2016.10.014>
- [46] Carskadon, M. A., & Dement, W. C. (1981). Cumulative effects of sleep restriction on daytime sleepiness. *Psychophysiology*, 18(2), 107-113. <https://doi.org/10.1111/j.1469-8986.1981.tb02921.x>
- [47] Ebben, M. R. (2017). Nonpharmacologic management of excessive daytime sleepiness. *Sleep medicine clinics*, 12(3), 479-487. <https://doi.org/10.1016/j.jsmc.2017.03.020>
- [48] Cooke, J. R., & Ancoli-Israel, S. (2006). Sleep and its disorders in older adults. *Psychiatric Clinics*, 29(4), 1077-1093. <https://doi.org/10.1016/j.psc.2006.08.003>
- [49] Longordo, F., Kopp, C., & Lüthi, A. (2009). Consequences of sleep deprivation on neurotransmitter receptor expression and function. *European Journal of Neuroscience*, 29(9), 1810-1819. <https://doi.org/10.1111/j.1460-9568.2009.06719.x>
- [50] Barden, N. (2004). Implication of the hypothalamic–pituitary–adrenal axis in the physiopathology of depression. *Journal of Psychiatry and Neuroscience*, 29(3), 185.
- [51] Fakhoury, M. (2016). Revisiting the serotonin hypothesis: implications for major depressive disorders. *Molecular neurobiology*, 53(5), 2778-2786. <https://doi.org/10.1007/s12035-015-9152-z>
- [52] Román, V., Hagewoud, R., Luiten, P. G., & Meerlo, P. (2006). Differential effects of chronic partial sleep deprivation and stress on serotonin-1A and muscarinic acetylcholine receptor sensitivity. *Journal of sleep research*, 15(4), 386-394. <https://doi.org/10.1111/j.1365-2869.2006.00555.x>
- [53] Meerlo, P., Koehl, M., Van der Borght, K., & Turek, F. W. (2002). Sleep restriction alters the hypothalamic-pituitary-adrenal response to stress. *Journal of neuroendocrinology*, 14(5), 397-402. <https://doi.org/10.1046/j.0007-1331.2002.00790.x>
- [54] Cepeda, M. S., Stang, P., Blacketer, C., Kent, J. M., & Wittenberg, G. M. (2016). Clinical relevance of sleep duration: results from a cross-sectional analysis using NHANES. *Journal of Clinical Sleep Medicine*, 12(6), 813-819. <https://doi.org/10.5664/jcsm.5876>



- [55] Maglione, J. E., Ancoli-Israel, S., Peters, K. W., Paudel, M. L., Yaffe, K., Ensrud, K. E., & Stone, K. L. (2012). Depressive symptoms and subjective and objective sleep in community-dwelling older women. *Journal of the American Geriatrics Society*, *60*(4), 635-643. <https://doi.org/10.1111/j.1532-5415.2012.03908.x>
- [56] Ouyang, P., & Sun, W. (2019). Depression and sleep duration: findings from middle-aged and elderly people in China. *Public health*, *166*, 148-154. <https://doi.org/10.1016/j.puhe.2018.10.007>
- [57] Mesas, A. E., López-García, E., León-Muñoz, L. M., Graciani, A., Guallar-Castillón, P., & Rodríguez-Artalejo, F. (2011). The association between habitual sleep duration and sleep quality in older adults according to health status. *Age and ageing*, *40*(3), 318-323. <https://doi.org/10.1093/ageing/afr004>
- [58] Sakamoto, N., Nanri, A., Kochi, T., Tsuruoka, H., Pham, N. M., Kabe, I., Matsuda, S., & Mizoue, T. (2013). Bedtime and sleep duration in relation to depressive symptoms among Japanese workers. *Journal of occupational health*, 13-0074. <https://doi.org/10.1539/joh.13-0074-OA>
- [59] De Wit, L., Luppino, F., van Straten, A., Penninx, B., Zitman, F., & Cuijpers, P. (2010). Depression and obesity: a meta-analysis of community-based studies. *Psychiatry research*, *178*(2), 230-235. <https://doi.org/10.1016/j.psychres.2009.04.015>
- [60] Ergün, U. Ö., Uguz, S., Bozdemir, N. A. F. İ. Z., Güzel, R., Burgut, R., Saatci, E. S. R. A., & Akpınar, E. (2004). The relationship between cholesterol levels and depression in the elderly. *International journal of geriatric psychiatry*, *19*(3), 291-296. <https://doi.org/10.1002/gps.1078>
- [61] Shin, J. Y., Suls, J., & Martin, R. (2008). Are cholesterol and depression inversely related? A meta-analysis of the association between two cardiac risk factors. *Annals of Behavioral Medicine*, *36*(1), 33-43. <https://doi.org/10.1007/s12160-008-9045-8>
- [62] Ford, D. E., & Erlinger, T. P. (2004). Depression and C-reactive protein in US adults: data from the Third National Health and Nutrition Examination Survey. *Archives of internal medicine*, *164*(9), 1010-1014. <https://doi.org/10.1001/archinte.164.9.1010>
- [63] Dipnall, J. F., Pasco, J. A., Berk, M., Williams, L. J., Dodd, S., Jacka, F. N., & Meyer, D. (2017). Why so GLUMM? Detecting depression clusters through graphing lifestyle-environments using machine-learning methods (GLUMM). *European Psychiatry*, *39*, 40-50. <https://doi.org/10.1016/j.eurpsy.2016.06.003>
- [64] Mumba, M. N., Nacarrow, A. F., Cody, S., Key, B. A., Wang, H., Robb, M., Jurczyk, A., Ford, C., Kelley, M. A., & Allen, R. S. (2021). Intensity and type of physical activity predicts depression in older adults. *Aging & mental health*, *25*(4), 664-671. <https://doi.org/10.1080/13607863.2020.1711861>
- [65] Zulyniak, S., Williams, J. V., Bulloch, A. G., Lukmanji, A., & Patten, S. B. (2020). The association of recreational and non-recreational physical activity with mental health: A Canadian cross-sectional analysis. *Journal of Affective Disorders Reports*, *1*, 100021. <https://doi.org/10.1016/j.jadr.2020.100021>

- [66] Stubbs, B., Koyanagi, A., Schuch, F. B., Firth, J., Rosenbaum, S., Veronese, N., Solmi, M., Mugisha, J., & Vancampfort, D. (2016). Physical activity and depression: a large cross-sectional, population-based study across 36 low-and middle-income countries. *Acta Psychiatrica Scandinavica*, *134*(6), 546-556. <https://doi.org/10.1111/acps.12654>
- [67] Wassink-Vossen, S., Collard, R. M., Voshaar, R. C. O., Comijs, H. C., de Vocht, H. M., & Naarding, P. (2014). Physical (in) activity and depression in older people. *Journal of affective disorders*, *161*, 65-72. <https://doi.org/10.1016/j.jad.2014.03.001>
- [68] Van Gool, C. H., Kempen, G. I., Penninx, B. W., Deeg, D. J., Beekman, A. T., & Van Eijk, J. T. (2003). Relationship between changes in depressive symptoms and unhealthy lifestyles in late middle aged and older persons: results from the Longitudinal Aging Study Amsterdam. *Age and ageing*, *32*(1), 81-87. <https://doi.org/10.1093/ageing/32.1.81>
- [69] Teychenne, M., Ball, K., & Salmon, J. (2010). Sedentary behavior and depression among adults: a review. *International journal of behavioral medicine*, *17*(4), 246-254. <https://doi.org/10.1007/s12529-010-9075-z>

#### 4. Discussão Geral

Durante o estudo encontramos algumas limitações ao trabalhar com o banco de dados disponibilizado pelo NHANES no período entre 2015-2016, principalmente relacionadas à estrutura da coleta de dados sobre sono. Neste sentido, os participantes forneceram respostas sobre hábitos de sono, tal como sua duração, horário de início e de despertar referente apenas aos dias úteis, além de informações indicando alguns tipos específicos de distúrbios do sono, como a presença de sonolência diurna, através de um questionário simples de autorrelato, não formulado de forma complexa como alguns questionários clássicos que promoveriam a análise mais aprofundada acerca do sono dos participantes.

Neste contexto, exemplos de questionários pertinentes que poderiam ser incorporados neste e em outros biobancos na metodologia de coleta destes dados são em primeiro lugar o Questionário de matutuidade-vespertuidade de Horne e Östberg (1976), no qual podemos acessar as diferenças interindividuais (cronotipos), expressas através das preferências matutinas ou vespertinas para realizar atividades distintas que inclui desde a fase de sono e de despertar, até o período ideal para a realização de atividade física, por exemplo. Em segundo lugar, temos o Questionário de Cronotipo de Munique que pode fornecer informações a respeito do tempo de sono individual, exposição à luz e cronotipo tanto nos dias úteis quanto nos dias livres, além da ocorrência e quantificação do jet lag social (Roenneberg et al., 2003; Roenneberg et., 2019). Por fim, para avaliar a sonolência diurna, a escala de sonolência de Epworth (ESS) é uma excelente sugestão de ferramenta baseada em 8 questões direcionadas para a ocorrência de cochilos e sono durante situações diárias bem específicas (Johns, 1991). Entretanto, embora essas opções não sejam adotadas, a simples medida de adicionar questionamentos sobre o tempo de sono e os horários de início do sono e despertar também referentes aos finais de semana, possibilitaria avaliar a existência de alguma instabilidade nestes horários, com base em diferenças encontradas entre os dias úteis e livres.

Quando nos voltamos para a detecção do transtorno depressivo na prática clínica é frequente o uso de ferramentas, como escalas e questionários de difícil interpretação, as quais podem demandar tempo de manipulação, exigir a recuperação da memória do paciente sobre comportamentos passados e apresentam algumas limitações de atualização sobre os novos avanços da doença (Mitchell & Coyne, 2009). Nestas circunstâncias, tornou-se frequente a busca por alternativas que visem substituir o modo como as informações sobre a saúde mental do paciente são coletadas, processadas e avaliadas, objetivando minimizar os custos médicos, reduzir o tempo para o diagnóstico, além de intervir sobre possíveis omissões do paciente a respeito da sua condição depressiva atual durante o processo de entrevistas convencionais. Assim, surgem os dados de detecção passiva oriundos de dispositivos como actígrafo e smartwatch, originando os chamados biomarcadores digitais da depressão, caracterizados por alguns comportamentos e aspectos fisiológicos do usuário (De Angel et al., 2022). Essas ferramentas podem ser utilizadas no ambiente natural do paciente durante dias sem causar desgastes ou grandes modificações na vida diária, coletando dados referentes principalmente ao ciclo atividade-reposo. Além disso, neste mesmo contexto de ambiente real, é possível avaliar o estado de humor durante vários dias consecutivos através de uma avaliação ecológica momentânea (AEM), a qual pode ser realizada a partir do auto-relato do indivíduo sobre o humor em horários específicos, solicitado por aplicativos em smartphones desenvolvidos para esta finalidade, por exemplo (Marzano et al., 2015).

Estudos já comprovam o alto valor preditivo de alguns biomarcadores digitais para a depressão com o desenvolvimento de algoritmos de aprendizagem de máquinas capazes de discriminar sujeitos depressivos dos controles saudáveis, detectar episódios e sintomatologia com precisão considerável a partir desses dados. Por exemplo, em estudo conduzido com idosos que vivem sozinhos, as melhores variáveis consideradas ao final do modelo de classificação de grupos depressivos, atingindo uma precisão de 92%, foram o escore médio

diário da AEM, o nível médio diário de atividade, exposição a luz branca e RGB no intervalo entre 16:00 e 20:00 horas, além da eficiência diária do sono (Kim et al., 2019). Por outro lado, Narziev et al. (2020) ao construírem um modelo para a detecção da depressão de curto prazo com capacidade de rastrear cinco grupos de sintomas depressivos, constataram que a variável relacionada ao sono foi a que melhor contribuiu para o desempenho final do modelo, com contribuição de 34%. Em busca de identificar episódios depressivos, Rodríguez-Ruiz et al. (2020) avaliaram as melhores combinações de 23 recursos provenientes dos dados de atividade motora em três condições distintas (atividade do dia, noite e dia inteiro) e observaram que para cada uma delas a combinação de nove recursos apresentaram os melhores desempenhos, com a precisão variando de 98,45% a 99,72%. Já outros trabalhos que testaram a capacidade preditiva dos fenótipos digitais, constataram outros recursos importantes para a previsão da depressão, tal como variáveis de temperatura da pele e frequência cardíaca (Tazawa et al., 2020; Rykov et al., 2021).

Geralmente para desenvolver modelos de AM generalizáveis o suficiente para casos não acessados anteriormente têm-se como uma das estratégias aumentar o número de dados para proporcionar mais exemplos durante o processo de treinamento, tornando o aprendizado mais consistente e ao mesmo tempo possibilitando um melhor ajuste entre *overfitting* e *underfitting*, embora exista uma variedade de outras técnicas (Mackenzie, 2015). No entanto, o que observamos nos estudos supracitados é um treinamento de modelos com bases em características de poucos sujeitos, em virtude do baixo número amostral. Dessa forma, a existência de biobancos que disponibilizem dados de detecção passiva provenientes de actigrafia, em grande escala e de forma adequada, é de extrema importante para pesquisadores que busquem reproduzir e desenvolver novos modelos que auxiliem no diagnóstico e detecção da depressão com base inclusive em séries temporais de variáveis cronobiológicas, ainda tão pouco exploradas.

## **5. Conclusão Geral**

As variáveis de sono e atividade física, além dos biomarcadores de obesidade e diabetes, em conjunto assumem significativa importância para predizer, com acurácia e precisão de 87%, a ocorrência de sintomas depressivos em indivíduos de meia-idade e idosos. Sujeitos com sintomas depressivos frequentemente apresentam sonolência diurna excessiva, tendem a adormecer e acordar mais tarde, apresentam IMC e CC mais elevados, pior controle glicêmico e evitam realizar atividades físicas recreativas de intensidade moderada à vigorosa em comparação a sujeitos saudáveis. Modelos de machine learning constituem ferramentas preditivas que podem assumir um caráter importante e auxiliar na prática clínica, principalmente em ambientes de atenção primária à saúde, direcionada para a detecção da sintomatologia depressiva neste público-alvo, tendo como base estes biomarcadores específicos da depressão.

## **6. Implicações do estudo**

Implicações clínicas podem ser percebidas com o presente modelo de ML, desenvolvido para a detecção da sintomatologia depressiva, mas se utilizado de forma a melhorar a rotina de triagem para posterior direcionamento diagnóstico com médico psiquiatra, por exemplo. Desta forma, tem grande potencial de causar impactos positivos no sistema de saúde, principalmente o de atenção primária, ao possibilitar maior facilidade de triagem dos casos, menor intervenção direta do paciente, maiores taxas de detecção dos sintomas e maior possibilidade de tratamento precoce da condição depressiva. Além disso, implicações para a pesquisa também podem ser observadas, dando o entendimento para a necessidade de explorar em futuros modelos preditivos de Machine Learning recursos preditores pouco investigados que têm forte relação com o desencadeamento da depressão, como variáveis cronobiológicas envolvendo, por exemplo, aspectos do ciclo sono-vigília.

## 7. Referências Gerais

- Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, *118*(3062), 273-274. <https://doi.org/10.1126/science.118.3062.273>
- Achermann, P., Dijk, D. J., Brunner, D. P., & Borbély, A. A. (1993). A model of human sleep homeostasis based on EEG slow-wave activity: quantitative comparison of data and simulations. *Brain research bulletin*, *31*(1-2), 97-113. [https://doi.org/10.1016/0361-9230\(93\)90016-5](https://doi.org/10.1016/0361-9230(93)90016-5)
- Adrien, J. (2002). Neurobiological bases for the relation between sleep and depression. *Sleep medicine reviews*, *6*(5), 341-351. <https://doi.org/10.1053/smr.2001.0200>
- American Academy of Sleep Medicine. (2007). The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. *Westchester, IL: American Academy of Sleep Medicine*, 23.
- Allen, N., Sudlow, C., Downey, P., Peakman, T., Danesh, J., Elliott, P., ... & Collins, R. (2012). UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology*, *1*(3), 123-126. <https://doi.org/10.1016/j.hlpt.2012.07.003>
- American Psychiatric Association. (2014). *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Artmed Editora.
- Au, J., & Reece, J. (2017). The relationship between chronotype and depressive symptoms: a meta-analysis. *Journal of affective disorders*, *218*, 93-104. <https://doi.org/10.1016/j.jad.2017.04.021>
- Antypa, N., Verkuil, B., Molendijk, M., Schoevers, R., Penninx, B. W., & Van Der Does, W. (2017). Associations between chronotypes and psychological vulnerability factors of depression. *Chronobiology international*, *34*(8), 1125-1135. <https://doi.org/10.1080/07420528.2017.1345932>
- Akincigil, A., & Matthews, E. B. (2017). National rates and patterns of depression screening in primary care: results from 2012 and 2013. *Psychiatric services*, *68*(7), 660-666. <https://doi.org/10.1176/appi.ps.201600096>

- Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H., & Van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PloS one*, *14*(5), e0213653. <https://doi.org/10.1371/journal.pone.0213653>
- Ali, M. K., Chwastiak, L., Poongothai, S., Emmert-Fees, K. M., Patel, S. A., Anjana, R. M., ... & INDEPENDENT Study Group. (2020). Effect of a collaborative care model on depressive symptoms and glycated hemoglobin, blood pressure, and serum cholesterol among patients with depression and diabetes in India: the INDEPENDENT randomized clinical trial. *Jama*, *324*(7), 651-662. <https://doi.org/10.1001/jama.2020.11747>
- Borbély, A. A. (1982). Sleep regulation: circadian rhythm and homeostasis. In *Sleep* (pp. 83-103). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-68333-6\\_3](https://doi.org/10.1007/978-3-642-68333-6_3)
- Borbély, A. A. (1987). The S-deficiency hypothesis of depression and the two-process model of sleep regulation. *Pharmacopsychiatry*, *20*(01), 23-29. <https://doi.org/10.1055/s-2007-1017069>
- Berger, M. V., Van Calker, D., & Riemann, D. (2003). Sleep and manipulations of the sleep-wake rhythm in depression. *Acta Psychiatrica Scandinavica*, *108*, 83-91. <https://doi.org/10.1034/j.1600-0447.108.s418.17.x>
- Bhagwagar, Z., Hafizi, S., & Cowen, P. J. (2005). Increased salivary cortisol after waking in depression. *Psychopharmacology*, *182*(1), 54-57. <https://doi.org/10.1007/s00213-005-0062-z>
- Berton, O., & Nestler, E. J. (2006). New approaches to antidepressant drug discovery: beyond monoamines. *Nature Reviews Neuroscience*, *7*(2), 137-151. <https://doi.org/10.1038/nrn1846>
- Beersma, D. G., & Gordijn, M. C. (2007). Circadian control of the sleep-wake cycle. *Physiology & behavior*, *90*(2-3), 190-195. <https://doi.org/10.1016/j.physbeh.2006.09.010>
- Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., & McCarley, R. W. (2012). Control of sleep and wakefulness. *Physiological reviews*. <https://doi.org/10.1152/physrev.00032.2011>



- Besedovsky, L., Lange, T., & Born, J. (2012). Sleep and immune function. *Pflügers Archiv-European Journal of Physiology*, 463(1), 121-137. <https://doi.org/10.1007/s00424-011-1044-0>
- Borbély, A. A., Daan, S., Wirz-Justice, A., & Deboer, T. (2016). The two-process model of sleep regulation: a reappraisal. *Journal of sleep research*, 25(2), 131-143. <https://doi.org/10.1111/jsr.12371>
- Brush, C. J., Hajcak, G., Bocchine, A. J., Ude, A. A., Muniz, K. M., Foti, D., & Alderman, B. L. (2022). A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets. *Psychological medicine*, 52(5), 893-903. <https://doi.org/10.1017/S0033291720002573>
- Cavalcante, J. S., Nascimento Júnior, E. S., & Costa, M. S. M. O. (2006). Componentes centrais do sistema de temporização circadiana: o núcleo supraquiasmático e o folheto intergeniculado. *Neurociências*, 3, 1-10.
- Chellappa, S. L., & Araújo, J. F. (2006). Excessive daytime sleepiness in patients with depressive disorder. *Brazilian Journal of Psychiatry*, 28, 126-129. <https://doi.org/10.1590/S1516-44462006000200010>
- Chokroverty, S. (2010). Overview of sleep & sleep disorders. *Indian J Med Res*, 131(2), 126-140.
- Courtet, P., & Olié, E. (2012). Circadian dimension and severity of depression. *European Neuropsychopharmacology*, 22,S476-S481 <https://doi.org/10.1016/j.euroneuro.2012.07.009>
- Craven, M. A., & Bland, R. (2013). Depression in primary care: current and future challenges. *The Canadian Journal of Psychiatry*, 58(8), 442-448. <https://doi.org/10.1177/070674371305800802>
- Cearns, M., Opel, N., Clark, S., Kaehler, C., Thalamuthu, A., Heindel, W., ... & Baune, B. T. (2019). Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Translational psychiatry*, 9(1), 1-9. <https://doi.org/10.1038/s41398-019-0615-2>

- Coppola, L., Cianflone, A., Grimaldi, A. M., Incoronato, M., Bevilacqua, P., Messina, F., ... & Salvatore, M. (2019). Biobanking in health care: evolution and future directions. *Journal of translational medicine*, *17*(1), 1-18. <https://doi.org/10.1186/s12967-019-1922-3>
- Cho, C. H., Lee, T., Kim, M. G., In, H. P., Kim, L., & Lee, H. J. (2019). Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *Journal of medical Internet research*, *21*(4), e11029. <https://doi.org/10.2196/11029>
- Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and clinical neurophysiology*, *9*(4), 673-690. [https://doi.org/10.1016/0013-4694\(57\)90088-3](https://doi.org/10.1016/0013-4694(57)90088-3)
- Dey, S. (1994). Physical exercise as a novel antidepressant agent: possible role of serotonin receptor subtypes. *Physiology & behavior*, *55*(2), 323-329. [https://doi.org/10.1016/0031-9384\(94\)90141-4](https://doi.org/10.1016/0031-9384(94)90141-4)
- Dong, C., Sanchez, L. E., & Price, R. A. (2004). Relationship of obesity to depression: a family-based study. *International journal of obesity*, *28*(6), 790-795. <https://doi.org/10.1038/sj.ijo.0802626>
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114-126. <https://doi.org/10.1038/nrn2762>
- Dibner, C., Schibler, U., & Albrecht, U. (2010). The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annual review of physiology*, *72*, 517-549. <https://doi.org/10.1146/annurev-physiol-021909-135821>
- Daly, M. (2013). The relationship of C-reactive protein to obesity-related depressive symptoms: A longitudinal study. *Obesity*, *21*(2), 248-250. <https://doi.org/10.1002/oby.20051>
- Dipnall, J. F., Pasco, J. A., Berk, M., Williams, L. J., Dodd, S., Jacka, F. N., & Meyer, D. (2016). Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one*, *11*(2), e0148195. <https://doi.org/10.1371/journal.pone.0148195>

- Deboer, T. (2018). Sleep homeostasis and the circadian clock: Do the circadian pacemaker and the sleep homeostat influence each other's functioning?. *Neurobiology of sleep and circadian rhythms*, 5, 68-77. <https://doi.org/10.1016/j.nbscr.2018.02.003>
- del Pozo Cruz, B., Alfonso-Rosa, R. M., McGregor, D., Chastin, S. F., Palarea-Albaladejo, J., & del Pozo Cruz, J. (2020). Sedentary behaviour is associated with depression symptoms: compositional data analysis from a representative sample of 3233 US adults and older adults assessed with accelerometers. *Journal of affective disorders*, 265, 59-62. <https://doi.org/10.1016/j.jad.2020.01.023>
- de Souza Filho, E. M., Rey, H. C. V., Frajttag, R. M., Cook, D. M. A., de Carvalho, L. N. D., Ribeiro, A. L. P., & Amaral, J. (2021). Can machine learning be useful as a screening tool for depression in primary care?. *Journal of Psychiatric Research*, 132, 1-6. <https://doi.org/10.1016/j.jpsychires.2020.09.025>
- Dong, L., Xie, Y., & Zou, X. (2022). Association between sleep duration and depression in US adults: A cross-sectional study. *Journal of Affective Disorders*, 296, 183-188. <https://doi.org/10.1016/j.jad.2021.09.075>
- De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., ... & Hotopf, M. (2022). Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ Digital Medicine*, 5(1), 1-14. <https://doi.org/10.1038/s41746-021-00548-8>
- Espiritu, J. R. D. (2008). Aging-related sleep changes. *Clinics in geriatric medicine*, 24(1), 1-14. <https://doi.org/10.1016/j.cger.2007.08.007>
- Ford, D. E., & Erlinger, T. P. (2004). Depression and C-reactive protein in US adults: data from the Third National Health and Nutrition Examination Survey. *Archives of internal medicine*, 164(9), 1010-1014. <https://doi.org/10.1001/archinte.164.9.1010>
- Fuller, P. M., Gooley, J. J., & Saper, C. B. (2006). Neurobiology of the sleep-wake cycle: sleep architecture, circadian regulation, and regulatory feedback. *Journal of biological rhythms*, 21(6), 482-493. <https://doi.org/10.1177/0748730406294627>
- Frey, S., Birchler-Pedross, A., Hofstetter, M., Brunner, P., Götz, T., Münch, M., ... & Cajochen, C. (2012). Challenging the sleep homeostat: sleep in depression is not

- premature aging. *Sleep Medicine*, 13(7), 933-945.  
<https://doi.org/10.1016/j.sleep.2012.03.008>
- Furht, B., & Villanustre, F. (2016). Introduction to big data. In *Big data technologies and applications* (pp. 3-11). Springer, Cham. <https://doi.org/10.1007/978-3-319-44550-2>
- Gaudreau, H., Morettini, J., Lavoie, H. B., & Carrier, J. (2001). Effects of a 25-h sleep deprivation on daytime sleep in the middle-aged. *Neurobiology of aging*, 22(3), 461-468.  
[https://doi.org/10.1016/S0197-4580\(00\)00251-7](https://doi.org/10.1016/S0197-4580(00)00251-7)
- Gaspar-Barba, E., Calati, R., Cruz-Fuentes, C. S., Ontiveros-Uribe, M. P., Natale, V., De Ronchi, D., & Serretti, A. (2009). Depressive symptomatology is influenced by chronotypes. *Journal of affective disorders*, 119(1-3), 100-106.  
<https://doi.org/10.1016/j.jad.2009.02.021>
- Gunn, J. M., Ayton, D. R., Densley, K., Pallant, J. F., Chondros, P., Herrman, H. E., & Dowrick, C. F. (2012). The association between chronic illness, multimorbidity and depressive symptoms in an Australian primary care cohort. *Social psychiatry and psychiatric epidemiology*, 47(2), 175-184. <https://doi.org/10.1007/s00127-010-0330-z>
- Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. *Annual review of clinical psychology*, 10, 679-708. <https://doi.org/10.1146/annurev-clinpsy-032813-153716>
- Grable, J. E., & Lyons, A. C. (2018). An Introduction to Big Data. *Journal of financial service professionals*, 72(5).
- Gonsalvez, I., Li, J. J., Stevens, C., Chen, J. A., & Liu, C. H. (2021). Preexisting Depression and Daytime Sleepiness in Women and Men. *Behavioral Sleep Medicine*, 1-13.  
<https://doi.org/10.1080/15402002.2021.1924720>
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology*.

- Hublin, C., Kaprio, J., Partinen, M., Heikkilä, K., & Koskenvuo, M. (1996). Daytime sleepiness in an adult, Finnish population. *Journal of internal medicine*, 239(5), 417-423. <https://doi.org/10.1046/j.1365-2796.1996.475826000.x>
- Hamer, M., Batty, G. D., & Kivimaki, M. (2012). Risk of future depression in people who are obese but metabolically healthy: the English longitudinal study of ageing. *Molecular psychiatry*, 17(9), 940-945. <https://doi.org/10.1038/mp.2012.30>
- Hryhorczuk, C., Sharma, S., & Fulton, S. E. (2013). Metabolic disturbances connecting obesity and depression. *Frontiers in neuroscience*, 7, 177. <https://doi.org/10.3389/fnins.2013.00177>
- Hayley, A. C., Williams, L. J., Berk, M., Kennedy, G. A., Jacka, F. N., & Pasco, J. A. (2013). The relationship between excessive daytime sleepiness and depressive and anxiety disorders in women. *Australian & New Zealand Journal of Psychiatry*, 47(8), 772-778. <https://doi.org/10.1177/0004867413490036>
- Huang, Z. T., Luo, Y., Han, L., Wang, K., Yao, S. S., Su, H. X., ... & Xu, B. (2022). Patterns of cardiometabolic multimorbidity and the risk of depressive symptoms in a longitudinal cohort of middle-aged and older Chinese. *Journal of affective disorders*. <https://doi.org/10.1016/j.jad.2022.01.030>
- Hein, M., Lanquart, J. P., Loas, G., Hubain, P., & Linkowski, P. (2019). Prevalence and risk factors of excessive daytime sleepiness in major depression: a study with 703 individuals referred for polysomnography. *Journal of affective disorders*, 243, 23-32. <https://doi.org/10.1016/j.jad.2018.09.016>
- Kaneita, Y., Ohida, T., Uchiyama, M., Takemura, S., Kawahara, K., Yokoyama, E., ... & Akashiba, T. (2005). Excessive daytime sleepiness among the Japanese general population. *Journal of epidemiology*, 15(1), 1-8. <https://doi.org/10.2188/jea.15.1>
- Kirchberger, I., Meisinger, C., Heier, M., Zimmermann, A. K., Thorand, B., Autenrieth, C. S., ... & Döring, A. (2012). Patterns of multimorbidity in the aged population. Results from the KORA-Age study. *PloS one*, 7(1), e30556. <https://doi.org/10.1371/journal.pone.0030556>

- Kaner, G., Soylu, M., Yüksel, N., Inanç, N., Ongan, D., & Başmısırlı, E. (2015). Evaluation of nutritional status of patients with depression. *BioMed research international*, 2015. <https://doi.org/10.1155/2015/521481>
- Kautzky, A., Dold, M., Bartova, L., Spies, M., Vanicek, T., Souery, D., ... & Kasper, S. (2017). Refining prediction in treatment-resistant depression: results of machine learning analyses in the TRD III sample. *The Journal of clinical psychiatry*, 78(1), 14989.
- Khurshid, K. A. (2018). Comorbid insomnia and psychiatric disorders: an update. *Innovations in clinical neuroscience*, 15(3-4), 28.
- Kandola, A., Ashdown-Franks, G., Hendrikse, J., Sabiston, C. M., & Stubbs, B. (2019). Physical activity and depression: Towards understanding the antidepressant mechanisms of physical activity. *Neuroscience & Biobehavioral Reviews*, 107, 525-539. <https://doi.org/10.1016/j.neubiorev.2019.09.040>
- Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (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. <https://doi.org/10.2196/14149>
- Kinkorová, J., & Topolčan, O. (2020). Biobanks in the era of big data: objectives, challenges, perspectives, and innovations for predictive, preventive, and personalised medicine. *EPMA Journal*, 11(3), 333-341. <https://doi.org/10.1007/s13167-020-00213-2>
- Jouvet, M., Michel, F., & Courjon, J. (1959). Sur un stade d'activité électrique cérébrale rapide au cours du sommeil physiologique. *CR Soc Biol*, 153, 1024-1028.
- Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *sleep*, 14(6), 540-545. <https://doi.org/10.1093/sleep/14.6.540>
- Johns, M., & Hocking, B. (1997). Daytime sleepiness and sleep habits of Australian workers. *Sleep*, 20(10), 844-847. <https://doi.org/10.1093/sleep/20.10.844>
- Jang, Y., Kim, G., Hansen, L., & Chiriboga, D. A. (2007). Attitudes of older Korean Americans toward mental health services. *Journal of the American Geriatrics Society*, 55(4), 616-620. <https://doi.org/10.1111/j.1532-5415.2007.01125.x>

- Jang, Y., Chiriboga, D. A., & Okazaki, S. (2009). Attitudes toward mental health services: Age-group differences in Korean American adults. *Aging and Mental Health, 13*(1), 127-134. <https://doi.org/10.1080/13607860802591070>
- Jantarantotai, N., Mosikanon, K., Lee, Y., & McIntyre, R. S. (2017). The interface of depression and obesity. *Obesity research & clinical practice, 11*(1), 1-10. <https://doi.org/10.1016/j.orcp.2016.07.003>
- Lapin, I. P., & Oxenkrug, G. F. (1969). Intensification of the central serotonergic processes as a possible determinant of the thymoleptic effect. *The Lancet, 293*(7586), 132-136. [https://doi.org/10.1016/s0140-6736\(69\)91140-4](https://doi.org/10.1016/s0140-6736(69)91140-4)
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note, 6*(70), 1.
- Lehto, S. M., Hintikka, J., Niskanen, L., Tolmunen, T., Koivumaa-Honkanen, H., Honkalampi, K., & Viinamäki, H. (2008). Low HDL cholesterol associates with major depression in a sample with a 7-year history of depressive symptoms. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 32*(6), 1557-1561. <https://doi.org/10.1016/j.pnpbp.2008.05.021>
- Lehto, S. M., Niskanen, L., Tolmunen, T., Hintikka, J., Viinamäki, H., Heiskanen, T., ... & Koivumaa-Honkanen, H. (2010). Low serum HDL-cholesterol levels are associated with long symptom duration in patients with major depressive disorder. *Psychiatry and clinical neurosciences, 64*(3), 279-283. <https://doi.org/10.1111/j.1440-1819.2010.02079.x>
- Lafortune, M., Gagnon, J. F., Latreille, V., Vandewalle, G., Martin, N., Filipini, D., ... & Carrier, J. (2012). Reduced slow-wave rebound during daytime recovery sleep in middle-aged subjects. <https://doi.org/10.1371/journal.pone.0043224>
- Li, H., Wang, A., Feng, W., Zheng, D., Gao, Q., Tao, L., ... & Guo, X. (2019). Prospective study of glycated hemoglobin and trajectories of depressive symptoms: the China health and retirement longitudinal study. *Aging and disease, 10*(2), 249. <https://doi.org/10.14336/AD.2018.0410>

- Lorenz, N., Spada, J., Sander, C., Riedel-Heller, S. G., & Hegerl, U. (2019). Circadian skin temperature rhythms, circadian activity rhythms and sleep in individuals with self-reported depressive symptoms. *Journal of psychiatric research*, *117*, 38-44. <https://doi.org/10.1016/j.jpsychires.2019.06.022>
- Li, A., Rosella, L. C., Kurdyak, P., & Wodchis, W. P. (2021). Depression as a Risk Factor for Physical Illness and Multimorbidity in a Cohort with No Prior Comorbidity. *The Canadian Journal of Psychiatry*, *66*(8), 726-736. <https://doi.org/10.1177/0706743720974832>
- Mistlberger, R. E., & Skene, D. J. (2004). Social influences on mammalian circadian rhythms: animal and human studies. *Biological Reviews*, *79*(3), 533-556. <https://doi.org/10.1017/S1464793103006353>
- Münch, M., Knoblauch, V., Blatter, K., Schröder, C., Schnitzler, C., Kräuchi, K., ... & Cajochen, C. (2004). The frontal predominance in human EEG delta activity after sleep loss decreases with age. *European Journal of Neuroscience*, *20*(5), 1402-1410. <https://doi.org/10.1111/j.1460-9568.2004.03580.x>
- Mistlberger, R. E. (2005). Circadian regulation of sleep in mammals: role of the suprachiasmatic nucleus. *Brain research reviews*, *49*(3), 429-454. <https://doi.org/10.1016/j.brainresrev.2005.01.005>
- Mitchell, A. J., & Coyne, J. C. (2009). *Screening for depression in clinical practice: an evidence-based guide*. Oxford University Press.
- Ma, J., & Xiao, L. (2010). Obesity and depression in US women: results from the 2005–2006 National Health and Nutritional Examination Survey. *Obesity*, *18*(2), 347-353. <https://doi.org/10.1038/oby.2009.213>
- Mume, M. O. (2010). Excessive daytime sleepiness among depressed patients. *Libyan Journal of Medicine*, *5*(1). <https://doi.org/10.4176/091024>
- Morselli, L. L., Guyon, A., & Spiegel, K. (2012). Sleep and metabolic function. *Pflügers Archiv-European Journal of Physiology*, *463*(1), 139-160. <https://doi.org/10.1007/s00424-011-1053-z>



- Marzano, L., Bardill, A., Fields, B., Herd, K., Veale, D., Grey, N., & Moran, P. (2015). The application of mHealth to mental health: opportunities and challenges. *The Lancet Psychiatry*, 2(10), 942-948. [https://doi.org/10.1016/S2215-0366\(15\)00268-0](https://doi.org/10.1016/S2215-0366(15)00268-0)
- Mackenzie, A. (2015). The production of prediction: What does machine learning want?. *European Journal of Cultural Studies*, 18(4-5), 429-445. <https://doi.org/10.1177/1367549415577384>
- Mander, B. A., Winer, J. R., & Walker, M. P. (2017). Sleep and human aging. *Neuron*, 94(1), 19-36. <https://doi.org/10.1016/j.neuron.2017.02.004>
- Nuyen, B. A., Fox, R. S., Malcarne, V. L., Wachsman, S. I., & Sadler, G. R. (2016). Excessive daytime sleepiness as an indicator of depression in Hispanic Americans. *Hispanic Health Care International*, 14(3), 116-123. <https://doi.org/10.1177/1540415316660616>
- Narziev, N., Goh, H., Toshnazarov, K., Lee, S. A., Chung, K. M., & Noh, Y. (2020). STDD: Short-term depression detection with passive sensing. *Sensors*, 20(5), 1396. <https://doi.org/10.3390/s20051396>
- Nair, P., Bhanu, C., Frost, R., Buszewicz, M., & Walters, K. R. (2020). A systematic review of older adults' attitudes towards depression and its treatment. *The Gerontologist*, 60(1), e93-e104. <https://doi.org/10.1093/geront/gnz048>
- Olusi, S. O., & Fido, A. A. (1996). Serum lipid concentrations in patients with major depressive disorder. *Biological psychiatry*, 40(11), 1128-1131. [https://doi.org/10.1016/S0006-3223\(95\)00599-4](https://doi.org/10.1016/S0006-3223(95)00599-4)
- Partonen, T., Haukka, J., Virtamo, J., Taylor, P. R., & Lönnqvist, J. (1999). Association of low serum total cholesterol with major depression and suicide. *The British Journal of Psychiatry*, 175(3), 259-262. <https://doi.org/10.1192/bjp.175.3.259>
- Pardini, L., & Kaeffer, B. (2006). Feeding and circadian clocks. *Reproduction Nutrition Development*, 46(5), 463-480. <https://doi.org/10.1051/rnd:2006032>
- Peever, J., & Fuller, P. M. (2017). The biology of REM sleep. *Current Biology*, 27(22), R1237-R1248. <https://doi.org/10.1016/j.cub.2017.10.026>

- Pereira-Miranda, E., Costa, P. R., Queiroz, V. A., Pereira-Santos, M., & Santana, M. L. (2017). Overweight and obesity associated with higher depression prevalence in adults: a systematic review and meta-analysis. *Journal of the American College of Nutrition*, *36*(3), 223-233. <https://doi.org/10.1080/07315724.2016.1261053>
- Perrotta, G. (2019). Sleep-wake disorders: Definition, contexts and neural correlations. *J Neurol Psychol*, *7*(09), 034-042. <https://doi.org/10.13188/2332-3469.1000044>
- Pengpid, S., & Peltzer, K. (2019). High sedentary behaviour and low physical activity are associated with anxiety and depression in Myanmar and Vietnam. *International Journal of Environmental Research and Public Health*, *16*(7), 1251. <https://doi.org/10.3390/ijerph16071251>
- Pandi-Perumal, S. R., Monti, J. M., Burman, D., Karthikeyan, R., BaHammam, A. S., Spence, D. W., ... & Narashimhan, M. (2020). Clarifying the role of sleep in depression: A narrative review. *Psychiatry Research*, *291*, 113239. <https://doi.org/10.1016/j.psychres.2020.113239>
- Patel, A. K., Reddy, V., & Araujo, J. F. (2021). Physiology, sleep stages. In *StatPearls [Internet]*. StatsPearls Publishing.
- Rea, M. S., Figueiro, M. G., Bullough, J. D., & Bierman, A. (2005). A model of phototransduction by the human circadian system. *Brain Research Reviews*, *50*(2), 213-228. <https://doi.org/10.1016/j.brainresrev.2005.07.002>
- Quiñones, A. R., Markwardt, S., & Botosaneanu, A. (2019). Diabetes-multimorbidity combinations and disability among middle-aged and older adults. *Journal of general internal medicine*, *34*(6), 944-951. <https://doi.org/10.1007/s11606-019-04896-w>
- Qin, D. D., Rizak, J., Feng, X. L., Yang, S. C., Lü, L. B., Pan, L., ... & Hu, X. T. (2016). Prolonged secretion of cortisol as a possible mechanism underlying stress and depressive behaviour. *Scientific reports*, *6*(1), 1-9. <https://doi.org/10.1038/srep30187>
- Ransford, C. P. (1982). A role for amines in the antidepressant effect of exercise: a review. *Medicine and science in sports and exercise*, *14*(1), 1-10. <https://doi.org/10.1249/00005768-198201000-00001>

- Rapp, S. R., Parisi, S. A., Walsh, D. A., & Wallace, C. E. (1988). Detecting depression in elderly medical inpatients. *Journal of consulting and clinical psychology*, *56*(4), 509. <https://doi.org/10.1037/0022-006X.56.6.851>
- Robison, J., Gruman, C., Gaztambide, S., & Blank, K. (2002). Screening for depression in middle-aged and older Puerto Rican primary care patients. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *57*(5), M308-M314. <https://doi.org/10.1093/gerona/57.5.M308>
- Roenneberg, T., Wirz-Justice, A., & Mellow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. *Journal of biological rhythms*, *18*(1), 80-90. <https://doi.org/10.1177/0748730402239679>
- Rosenwasser, A. M., & Turek, F. W. (2005). Physiology of the mammalian circadian system. *Principles and practice of sleep medicine*, *4*, 351-362. <https://doi.org/10.1016/B0-72-160797-7/50036-7>
- Read, J. R., Sharpe, L., Modini, M., & Dear, B. F. (2017). Multimorbidity and depression: a systematic review and meta-analysis. *Journal of affective disorders*, *221*, 36-46. <https://doi.org/10.1016/j.jad.2017.06.009>
- Roenneberg, T., Pilz, L. K., Zerbini, G., & Winnebeck, E. C. (2019). Chronotype and social jetlag: a (self-) critical review. *Biology*, *8*(3), 54. <https://doi.org/10.3390/biology8030054>
- Rodríguez-Ruiz, J. G., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., ... & Soto-Murillo, M. A. (2020). Comparison of night, day and 24 h motor activity data for the classification of depressive episodes. *Diagnostics*, *10*(3), 162. <https://doi.org/10.3390/diagnostics10030162>
- Rykov, Y., Thach, T. Q., Bojic, I., Christopoulos, G., & Car, J. (2021). Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, *9*(10), e24872. <https://doi.org/10.2196/24872>
- Schildkraut, J. J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *American journal of Psychiatry*, *122*(5), 509-522. <https://doi.org/10.1176/ajp.122.5.509>

- Sou tre, E., Salvati, E., Belugou, J. L., Pringuey, D., Candito, M., Krebs, B., ... & Darcourt, G. (1989). Circadian rhythms in depression and recovery: evidence for blunted amplitude as the main chronobiological abnormality. *Psychiatry research*, 28(3), 263-278. [https://doi.org/10.1016/0165-1781\(89\)90207-2](https://doi.org/10.1016/0165-1781(89)90207-2)
- Sapolsky, R. M. (2000). Stress hormones: good and bad. *Neurobiology of disease*, 7(5), 540-542. <https://doi.org/10.1006/nbdi.2000.0350>
- Soria, V., & Urretavizcaya, M. (2009). Circadian rhythms and depression. *Actas Esp Psiquiatr*, 37(4), 222-232.
- Siegel, J. M. (2011). REM sleep: a biological and psychological paradox. *Sleep medicine reviews*, 15(3), 139. <https://doi.org/10.1016/j.smr.2011.01.001>
- Saveanu, R. V., & Nemeroff, C. B. (2012). Etiology of depression: genetic and environmental factors. *Psychiatric Clinics*, 35(1), 51-71. <https://doi.org/10.1016/j.psc.2011.12.001>
- Smith, D. J., McLean, G., Martin, D., Martin, J. L., Guthrie, B., Gunn, J., & Mercer, S. W. (2014). Depression and multimorbidity: a cross-sectional study of 1,751,841 patients in primary care. *The Journal of clinical psychiatry*, 75(11), 4205. <https://doi.org/10.4088/JCP.14m09147>
- Schmitz, N., Desch nes, S., Burns, R., & Smith, K. J. (2016). Depressive symptoms and glycated hemoglobin A1c: a reciprocal relationship in a prospective cohort study. *Psychological Medicine*, 46(5), 945-955. <https://doi.org/10.1017/S0033291715002445>
- Schuch, F., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P., Reichert, T., ... & Stubbs, B. (2017). Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. *Journal of affective disorders*, 210, 139-150. <https://doi.org/10.1016/j.jad.2016.10.050>
- Syeda-Mahmood, T. (2018). Role of big data and machine learning in diagnostic decision support in radiology. *Journal of the American College of Radiology*, 15(3), 569-576. <https://doi.org/10.1016/j.jacr.2018.01.028>
- Sun, Y., Shi, L., Bao, Y., Sun, Y., Shi, J., & Lu, L. (2018). The bidirectional relationship between sleep duration and depression in community-dwelling middle-aged and elderly

- individuals: evidence from a longitudinal study. *Sleep medicine*, 52, 221-229. <https://doi.org/10.1016/j.sleep.2018.03.011>
- Steiger, A., & Pawlowski, M. (2019). Depression and sleep. *International journal of molecular sciences*, 20(3), 607. <https://doi.org/10.3390/ijms20030607>
- Sharma, A., & Verbeke, W. J. (2020). Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset (n= 11,081). *Frontiers in big Data*, 3, 15. <https://doi.org/10.3389/fdata.2020.00015>
- Silahtaroglu, G., & Yilmaztürk, N. (2021). Data analysis in health and big data: a machine learning medical diagnosis model based on patients' complaints. *Communications in Statistics-Theory and Methods*, 50(7), 1547-1556.
- Soffer, S., Klang, E., Barash, Y., Grossman, E., & Zimlichman, E. (2021). Predicting in-hospital mortality at admission to the medical ward: a big-data machine learning model. *The American journal of medicine*, 134(2), 227-234. <https://doi.org/10.1016/j.amjmed.2020.07.014>
- Tazawa, Y., Liang, K. C., Yoshimura, M., Kitazawa, M., Kaise, Y., Takamiya, A., ... & Kishimoto, T. (2020). Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Heliyon*, 6(2), e03274. <https://doi.org/10.1016/j.heliyon.2020.e03274>
- Tong, L., Pu, L., Guo, X., Sun, M., Guo, F., Zhao, S., ... & Jin, L. (2021). Multimorbidity study with different levels of depression status. *Journal of Affective Disorders*, 292, 30-35. <https://doi.org/10.1016/j.jad.2021.05.039>
- Vancampfort, D., Stubbs, B., & Koyanagi, A. (2017). Physical chronic conditions, multimorbidity and sedentary behavior amongst middle-aged and older adults in six low- and middle-income countries. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1), 1-13. <https://doi.org/10.1186/s12966-017-0602-z>
- Vanek, J., Prasko, J., Genzor, S., Ociskova, M., Kantor, K., Holubova, M., ... & Sova, M. (2020). Obstructive sleep apnea, depression and cognitive impairment. *Sleep Medicine*, 72, 50-58. <https://doi.org/10.1016/j.sleep.2020.03.017>

- Zhai, L., Zhang, H., & Zhang, D. (2015). Sleep duration and depression among adults: A meta-analysis of prospective studies. *Depression and anxiety*, 32(9), 664-670. <https://doi.org/10.1002/da.22386>
- Zhai, L., Zhang, Y., & Zhang, D. (2015). Sedentary behaviour and the risk of depression: a meta-analysis. *British journal of sports medicine*, 49(11), 705-709. <http://dx.doi.org/10.1136/bjsports-2014-093613>
- Zavala, G. A., Kolovos, S., Chiarotto, A., Bosmans, J. E., Campos-Ponce, M., Rosado, J. L., & Garcia, O. P. (2018). Association between obesity and depressive symptoms in Mexican population. *Social psychiatry and psychiatric epidemiology*, 53(6), 639-646. <https://doi.org/10.1007/s00127-018-1517-y>
- Wichniak, A., Wierzbicka, A., & Jernajczyk, W. (2013). Sleep as a biomarker for depression. *Internacional review of psychiatry*, 25(5), 632-645. <https://doi.org/10.3109/09540261.2013.812067>

## **8. Apêndices**

**Material Suplementar do artigo:**

Supplementary Table 1: Contingency table for different levels of categorical variables.

<b>DS</b>	<b>Depressive Symptoms</b>		<b>Total</b>	
	<b>no</b>	<b>yes</b>		
almost always	Count	92	133	225
	Expected count	168	56.6	225
never	Count	502	78	580
	Expected count	434	145.8	580
often	Count	271	193	464
	Expected count	347	116.7	464
rarely	Count	595	116	711
	Expected count	532	178.8	711
sometimes	Count	716	211	927
	Expected count	694	233.1	927
Total	Count	2176	731	2907
	Expected count	2176	731	2907
<b>MRA</b>				
no	Count	1241	500	1741
	Expected count	1303	438	1741
yes	Count	935	231	1166
	Expected count	873	293	1166
Total	Count	2176	731	2907
	Expected count	2176	731	2907
<b>VRA</b>				
no	Count	1757	748	2405
	Expected count	1800	605	2405
yes	Count	419	83	502
	Expected count	376	126	502
Total	Count	2176	731	2907
	Expected count	2176	731	2907

**Gender**



female	Count	1065	432	1497
	Expected count	1121	376	1497
male	Count	1111	299	1410
	Expected count	1055	355	1410
Total	Count	2176	731	2907
	Expected count	2176	731	2907

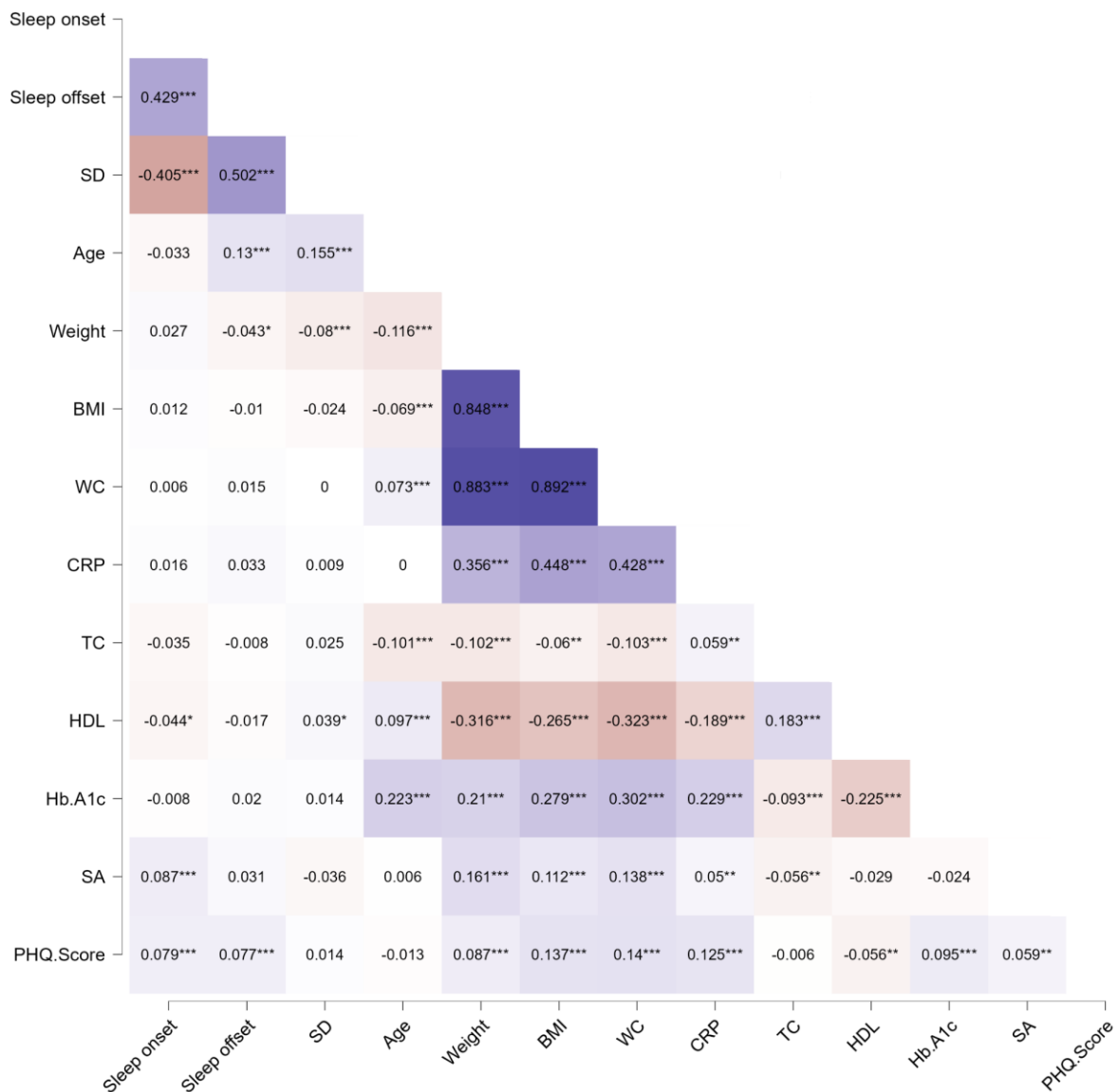
Note. (DS)- Daytime Sleepiness; (MRA)- Moderate Recreational Activity and (VRA)- Vigorous Recreational Activity.

### Supplementary Table 2: List of 15 predictor variables selected in the final dataset

Attributes	Categories
Moderate Recreational Activity	
Vigorous Recreational Activity	Self-reported physical activity
Sedentary Activity	
Sleep onset	
Sleep offset	Self-reported sleep habits
Sleep duration	
Excessive daytime sleepiness	Indicative of sleep disturbance
Waist circumference	
Body Mass Index	Anthropometric measurement
C-reactive protein	
Total cholesterol	Plasma biomarkers of obesity
High-density lipoprotein	
Glycated hemoglobin	Biomarker of diabetes
Sex	
Age	Sociodemographic characteristics

Supplementary Table 3: Glossary of key hyperparameters of machine learning models

<b>XGBoost model specifications</b>	<b>Maximum tree depth:</b> Specifies the maximum depth of the individual tree.
	<b>Boosting learning rate:</b> Specifies the boosting learning rate. The learning rate shrinks the contribution of each tree.
	<b>The number of trees:</b> Specifies how many gradient boosted trees will be included. A large number usually results in better performance.
	<b>Regularization strength:</b> Defines the cost strength of control over the model complexity to avoid overfitting.
	<b>Fraction of training instances:</b> Specifies the percentage of the training instances for fitting the individual tree.
	<b>Fraction of features for each tree:</b> Specifies the percentage of features to use when constructing each tree.
	<b>Fraction of features for each level:</b> Specifies the percentage of features to use for each level.
<b>Random Forest model specifications</b>	<b>Fraction of features for each split:</b> Specifies the percentage of features to use for each split.
	<b>Number of trees:</b> Specifies how many decision trees will be included in the forest.
	<b>Maximal number of considered features:</b> Number of attributes considered at each split.
<b>Logistic Regression Model Specifications</b>	<b>Maximal tree depth:</b> The depth at which the trees will be grown. <b>Stop splitting nodes with maximum instances:</b> Specifies not to split subsets smaller than requested.
	<b>Regularization type:</b> Regularization Lasso (L1). An alternative regularized version of the least squares method is Lasso (absolute minimum selection and shrinkage operator). <b>Regularization strength:</b> Sets the cost strength (default is C=1).



Supplementary Figure 1: Spearman's correlation between quantitative predictor variables and PHQ-9 score. (SD)- Sleep Duration; (BMI)- Body Mass Index; (WC)- Waist Circumference; (CRP)- C-Reactive Protein; (TC)- Total cholesterol; (HDL)- High Density Lipoprotein; (Hb. A1c)- Glycohemoglobin; (SA)- Sedentary Activity and (PHQ.Score)- PHQ-9 score. Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .