

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET  
DEPARTAMENTO DE COMPUTAÇÃO– DC  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

**Conrado dos Santos Alves Saud**

**Uma infraestrutura computacional para  
a identificação de estudantes  
universitários com possível perfil  
depressivo usando dados de sensores  
móveis**



**Conrado dos Santos Alves Saud**

**Uma infraestrutura computacional para  
a identificação de estudantes  
universitários com possível perfil  
depressivo usando dados de sensores  
móveis**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Interação Humano-Computador

Orientador: Vânia Paula de Almeida Neris

São Carlos

2023



*Dedicado aos estudantes universitários, pelas suas experiências e força na jornada  
contra a depressão.*



---

# Agradecimentos

---

Expresso minha profunda e sincera gratidão à minha orientadora, Dra. Vania Neris, cuja orientação, sabedoria e apoio foram fundamentais em toda a minha jornada no programa de mestrado. Graças a ela, tive a oportunidade de mergulhar no mundo da ciência e da pesquisa acadêmica, o que me proporcionou o privilégio de conhecer e ser inspirado por pessoas extraordinárias ao longo deste percurso, aos quais também devo minha gratidão.

Acima de tudo, a concretização deste trabalho deve-se imensamente à minha mãe, Dulcelei Maria dos Santos, a fonte de minha força e inspiração diária. É na sua dedicação e amor que encontro o impulso para superar desafios e seguir crescendo e evoluindo continuamente.



*“É difícil assumir minhas fraquezas, minhas angústias, meus anseios. A minha maior batalha sempre vai ser vencer minha cabeça. [...]. Quando eu me sinto assim, só há um jeito de me levantar, mesmo sendo difícil de controlar: eu precisei me desligar do mundo pra me reconectar comigo mesmo.”*  
*(John Wayne. "Reconectar")*



---

# Resumo

---

A depressão, um dos transtornos mentais mais prevalentes, impacta significativamente a vida de milhões de pessoas em todo o mundo. Este transtorno mental não só causa graves problemas de saúde, mas também interfere negativamente na vida daqueles que sofrem com a doença, muitos dos quais não procuram tratamento. Na população universitária, a prevalência da depressão é ainda maior do que na população em geral. Este trabalho tem como objetivo a proposição de uma infraestrutura computacional que utiliza dados de fenotipagem digital coletados por sensores móveis e vestíveis, como celulares e relógios inteligentes, para identificar estudantes universitários brasileiros com um possível perfil depressivo (PPD). Adotando a metodologia de *Design Science Research* (DSR), o estudo desenvolve uma solução computacional cuja eficácia é validada por meio de uma análise exploratória descritiva dos dados coletados. Esta abordagem busca não apenas validar a viabilidade da infraestrutura proposta, mas também facilitar a identificação de padrões comportamentais associados ao PPD. A abordagem inclui o uso do método “Human in the Loop” para a remoção de *outliers* dos dados e emprega a técnica t-SNE (t-distributed Stochastic Neighbor Embedding) para redução de dados de alta dimensão e identificação de padrões na interpretação dos resultados. Este trabalho se alinha com os princípios da Interação Humano-Computador (IHC) com o intuito de contribuir para os campos da saúde mental e da computação.

**Palavras-chave:** depressão, sensores móveis, vestíveis, fenotipagem digital, infraestrutura computacional, estudantes universitários.



---

# Abstract

---

Depression, one of the most prevalent mental disorders, significantly impacts the lives of millions worldwide. This mental disorder not only causes severe health problems but also negatively interferes with the lives of those who suffer from it, many of whom do not seek treatment. In the university population, the prevalence of depression is even higher than in the general population. This work proposes a computing infrastructure that uses digital phenotyping data collected by mobile and wearable sensors, such as smartphones and smartwatches, to identify Brazilian university students with a possible depressive profile (PPD). Adopting the Design Science Research (DSR) methodology, the study develops a computational solution whose effectiveness is validated through a descriptive exploratory analysis of the collected data. This approach seeks not only to validate the feasibility of the proposed infrastructure but also to facilitate the identification of behavioral patterns associated with the PPD. The approach includes the use of the "Human in the Loop" method for removing data outliers and employs the t-SNE (t-distributed Stochastic Neighbor Embedding) technique for reducing high-dimensional data and identifying patterns in interpreting the results. This work aligns with the principles of Human-Computer Interaction (HCI) intending to contribute to the areas of mental health and computing.

**Keywords:** Depression, mobile sensors, wearables, digital phenotyping, computing infrastructure, college students.



---

# Lista de ilustrações

---

Figura 1 – Estimativas globais e regionais de prevalência de depressão . . . . .	34
Figura 2 – Processo de seleção dos artigos . . . . .	45
Figura 3 – Número de publicações por ano relacionadas a identificação do PPD com uso da tecnologia . . . . .	46
Figura 4 – Algoritmos de aprendizado de máquina mais utilizados . . . . .	47
Figura 5 – Generalização da coleta e tráfego de informações entre os dispositivos dos estudos analisados . . . . .	49
Figura 6 – Infraestrutura do projeto Amive . . . . .	54
Figura 7 – Protótipo do aplicativo Amive . . . . .	56
Figura 8 – Interface de usuário final do aplicativo Amive . . . . .	58
Figura 9 – Visão geral dos dados dos participantes no Amive Web . . . . .	62
Figura 10 – Dados detalhados de um único usuário no Amive Web . . . . .	63
Figura 11 – Relação de dados de um usuário no Amive Web . . . . .	64
Figura 12 – Relação de dados de um usuário no Amive Web . . . . .	66
Figura 13 – Histograma de frequência cardíaca . . . . .	72
Figura 14 – Boxplot de frequência cardíaca . . . . .	72
Figura 15 – Gráfico de densidade de frequência cardíaca . . . . .	73
Figura 16 – Histograma de tempo de sono . . . . .	73
Figura 17 – Boxplot de tempo de sono . . . . .	74
Figura 18 – Gráfico de densidade de tempo de sono . . . . .	74
Figura 19 – Histograma de contagem de passos . . . . .	75
Figura 20 – Boxplot de contagem de passos . . . . .	76
Figura 21 – Gráfico de densidade de contagem de passos . . . . .	77
Figura 22 – Histograma de exercícios físicos . . . . .	77
Figura 23 – Boxplot de exercícios físicos . . . . .	78
Figura 24 – Gráfico de densidade de exercícios físicos . . . . .	78
Figura 25 – t-SNE não normalizado . . . . .	82

Figura 26 – t-SNE normalizado . . . . . 83

---

## Lista de tabelas

---

Tabela 1 – Relação de sensores do relógio inteligente com sintomas descritos no DSM-5 . . . . .	37
Tabela 2 – Quantidade de artigos selecionados por base de dados . . . . .	46
Tabela 3 – Comparação entre os trabalhos selecionados do mapeamento sistemático	51
Tabela 4 – Total de envio de dados dos sensores . . . . .	61
Tabela 5 – Correlação de Spearman aplicado ao grupo geral . . . . .	79
Tabela 6 – Correlação de Spearman aplicado ao grupo com PPD . . . . .	80
Tabela 7 – Correlação de Spearman aplicado ao grupo não PPD . . . . .	81
Tabela 8 – Resumo dos grupos da correlação de Spearman . . . . .	82



---

# Lista de siglas

---

**Amive** Amigo Virtual Especializado

**BDI** *Beck's Depression Inventory*

**CEP** Comitê de Ética em Pesquisa

**DSR** *Design Science Research*

**GPS** *Global Positioning System*

**HITL** *Human in the Loop*

**HDRS** *Hamilton Depression Rating Scale*

**IHC** Interação Humano-Computador

**ID** Identificador Único

**IA** Inteligência Artificial

**PPD** Perfil Possivelmente Depressivo

**PLN** Processamento de Linguagem Natural

**PHQ-9** *Patient Health Questionnaire-9*

**QIDS** *Quick Inventory of Depressive Symptomatology*

**TDM** Transtorno Depressivo Maior

**UI** Interface de Usuário

**UX** Experiência do Usuário

**WHODAS-II** *WHO Disability Assessment Schedule 2.0*

---

# Sumário

---

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>21</b>
1.1	Contexto e motivação . . . . .	21
1.2	Problemática . . . . .	22
1.3	Objetivo . . . . .	24
1.4	Síntese da metodologia . . . . .	26
1.5	Síntese do resultado . . . . .	28
1.5.1	Disponibilização da base de dados . . . . .	29
1.6	Organização do documento . . . . .	30
<b>2</b>	<b>CONCEITOS FUNDAMENTAIS . . . . .</b>	<b>33</b>
2.1	Depressão . . . . .	33
2.2	Sensores móveis . . . . .	35
2.3	Análise estatística . . . . .	39
2.3.1	Análise de variáveis . . . . .	39
2.3.2	Estudo de correlação . . . . .	40
2.3.3	Redução de dimensionalidade . . . . .	40
<b>3</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>43</b>
3.1	Levantamento bibliográfico . . . . .	43
3.1.1	Termo de busca . . . . .	43
3.1.2	Fonte de dados . . . . .	44
3.1.3	Critério de seleção . . . . .	44
3.1.4	Análise . . . . .	45
3.1.5	Resultados . . . . .	46
3.1.6	Discussão . . . . .	48
<b>4</b>	<b>INFRAESTRUTURA . . . . .</b>	<b>53</b>

<b>5</b>	<b>ESTUDO COM USUÁRIOS</b> . . . . .	<b>59</b>
<b>5.1</b>	<b>Método de coleta</b> . . . . .	<b>59</b>
<b>5.2</b>	<b>Síntese dos dados coletados</b> . . . . .	<b>61</b>
<b>5.3</b>	<b>Disponibilização de dados</b> . . . . .	<b>61</b>
<b>5.4</b>	<b>Human in the loop</b> . . . . .	<b>65</b>
<b>6</b>	<b>ANÁLISE ESTATÍSTICA</b> . . . . .	<b>69</b>
<b>6.1</b>	<b>Conjunto de dados</b> . . . . .	<b>69</b>
<b>6.2</b>	<b>Métodos</b> . . . . .	<b>70</b>
<b>6.3</b>	<b>Análise univariada</b> . . . . .	<b>71</b>
6.3.1	Frequência cardíaca . . . . .	71
6.3.2	Tempo de sono . . . . .	73
6.3.3	Contagem de passos . . . . .	75
6.3.4	Exercícios físicos . . . . .	76
<b>6.4</b>	<b>Análise bivariada</b> . . . . .	<b>79</b>
6.4.1	Grupo geral . . . . .	79
6.4.2	Grupo PPD . . . . .	79
6.4.3	Grupo não PPD . . . . .	80
6.4.4	Resumo dos grupos . . . . .	81
<b>6.5</b>	<b>Análise multivariada</b> . . . . .	<b>82</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>85</b>
<b>7.1</b>	<b>Limitações</b> . . . . .	<b>86</b>
<b>7.2</b>	<b>Acesso aos Recursos da Pesquisa</b> . . . . .	<b>87</b>
<b>7.3</b>	<b>Considerações finais</b> . . . . .	<b>88</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>91</b>

---

# Capítulo 1

## Introdução

---

### 1.1 Contexto e motivação

A depressão afeta mais de 320 milhões de pessoas em todo o mundo (WHO, 2017; AL., 2018), com uma estimativa de aumento de 25% mais casos durante a pandemia (WHO, 2021). É uma das condições que ocasionam mais custos em termos de tratamento da qualidade de vida do indivíduo, podendo resultar em automutilação, ou mesmo morte. Só no Brasil, em média 11,3% dos brasileiros relatam ter recebido um diagnóstico médico da doença (VIGITEL, 2022).

Os episódios depressivos são mais prevalentes entre jovens de 18 a 25 anos (BOSE SARRA L. HEDDEN; PARK-LEE, 2018), o que coincide com o período em que a maioria dos estudantes ingressa na faculdade. A depressão afeta negativamente indivíduos de diversos níveis socioeconômicos, idade, sexo, mas é especialmente prevalente entre populações vulneráveis, ou seja, aqueles que vivem em condições de carência econômica, carecem de apoio social ou sofrem de outras condições médicas. Embora existam métodos e opções de tratamento para transtornos mentais como a depressão, aproximadamente 76% a 85% das pessoas em países de baixa e média renda não recebem tratamento (WANG et al., 2007).

No contexto do ambiente universitário, a prevalência da depressão parece ser ainda maior do que o restante da população. De acordo com Lauckner, Hill e Ingram (2018), “os estudantes universitários suportam um fardo desproporcional de depressão quando comparados com a população em geral”. Uma revisão sistemática realizada por Ibrahim et al. (2013) em estudos sobre a prevalência da depressão em estudantes universitários ao redor do mundo descobriu que a média da prevalência de sintomas depressivos era de 30%. Além disso, Pacheco et al. (2017), em um estudo similar usando a escala de

*Beck's Depression Inventory* (BDI) (BECK, 1961), também encontrou uma prevalência de 30,6% de sintomas depressivos entre estudantes de medicina brasileiros. Essas estatísticas influenciaram a motivação do estudo desta da pesquisa focado no ambiente universitário.

Tratamentos de saúde mental, como no caso da depressão, demandam tempo e monitoramento por parte de um profissional especializado. Há um consenso de que a identificação precoce auxilia no tratamento de problemas de saúde mental. No entanto, a identificação do problema não ocorre pontualmente, por meio de um único episódio de desconforto, mas requer a observação do comportamento do indivíduo ao longo do tempo, incluindo suas reações físicas e psicológicas (DSM, 2022). Nesse cenário, a computação desempenha um papel relevante devido à possibilidade de utilizar tecnologias, especialmente aquelas acrescidas de recursos de Inteligência Artificial (IA) para monitorar o indivíduo em suas ações diárias, coletando dados que podem apoiar a identificação de situações de cuidados com a saúde mental, como por exemplo, letargia ou agitação, fadiga e qualidade do sono, ou até mesmo processar automaticamente dados de autorrelato.

Com avanço no poder de processamento e o acesso facilitado a sensores, novos trabalhos acadêmicos têm usado abordagens tecnológicas para estudar a coleta, diagnóstico, monitoramento e as possibilidades de tratamento da depressão graças à oferta de tecnologias em dispositivos móveis (ZHENG; MOTTI, 2018; WANG et al., 2014; HARARI et al., 2016). Um estudo realizado no Colégio de Dartmouthk, Wang et al. (2016) utilizou sensores de dispositivos móveis para tentar identificar sintomas depressivos em estudantes universitários, e apresentou uma série de correlações significativas entre vários comportamentos depressivos que puderam ser inferidos a partir de dados dos sensores. Além de possibilitar a coleta de dados de forma contínua e automática, tais dispositivos, como celulares e relógios inteligentes, quando integrados a algoritmos de reconhecimento de padrão de comportamento, permitem fornecer intervenções personalizadas aos usuários em tempo real (EVMENOVA et al., 2019).

Neste trabalho, considera-se de que o hardware disponível atualmente pode ajudar significativamente na identificação de sintomas depressivos, aprimorando o diagnóstico precoce e auxiliando no apoio ao tratamento da depressão. Destacam-se as oportunidades oferecidas por sensores e pela inteligência artificial, que embasam a motivação de se explorar como dispositivos móveis e algoritmos de reconhecimento de padrões podem ser empregados para monitorar indivíduos de forma contínua e não invasiva.

## 1.2 Problemática

Atualmente, muitas pessoas têm sensores móveis ao seu alcance, como aqueles presentes em celulares e relógios inteligentes (*smartwatches*). Esses dispositivos estão equipados com uma variedade de sensores que podem desempenhar um papel fundamental na coleta de dados fisiológicos e comportamentais do indivíduo (SHEIKH; QASSEM; KYRIACOU,

2021). Esses dispositivos oferecem a capacidade de monitorar de forma contínua e não invasiva uma ampla gama de sinais vitais e padrões de comportamento do indivíduo, abrindo novas oportunidades para o monitoramento da saúde mental (LEE et al., 2021; ALAMOUDI et al., 2023). Este campo emergente, conhecido como **fenotipagem digital** (ONNELA, 2020), envolve a coleta de dados a partir de dispositivos vestíveis e móveis para analisar diversos aspectos da saúde e do comportamento humano, inclusive a identificação de sintomas depressivos.

Além do uso de sensores móveis como instrumento na detecção do comportamento depressivo, ainda é preciso haver uma comparação a fim de diferenciar o que é um comportamento depressivo e o que é um comportamento padrão (normal) do indivíduo, criando assim uma classificação. A estratégia mais comum encontrada na literatura é o uso de questionários de saúde mental (FARHAN et al., 2016; GARCIA-CEJA et al., 2018; WARE et al., 2020), que permitem obter um diagnóstico simples, servindo como rotulagem para classificar se os dados associados a determinados comportamento estão alinhados com padrões depressivos ou não. Apesar de ser a estratégia mais comum, existem outras formas de classificação, como Dai et al. (2022), que usou grupos de controle para comparar participantes depressivos e não depressivos, sem questionários de saúde mental.

O tema de pesquisa deste trabalho é a identificação de padrões e características do Perfil Possivelmente Depressivo (PPD) com o uso de sensores presentes em tecnologias móveis. Isso se faz necessário, uma vez que os tratamentos usuais, como psicoterapia e medicamentos, são custosos (CUIJPERS et al., 2019; KNAPP; WONG, 2020), e demandam da disposição do indivíduo para acompanhamento clínico. É importante ressaltar que esse tipo de abordagem não substitui o profissional da saúde e também não diminui o custo do tratamento, e sim, favorece o acompanhamento do profissional e do próprio indivíduo. O acompanhamento também pode ser útil para evitar o agravamento dos quadros, favorecendo intervenções precoces. A existência de uma tecnologia capaz de prever e auxiliar o tratamento da depressão de forma pervasiva com objetos que as pessoas usam no dia a dia (como relógios e celulares) pode auxiliar as pessoas de uma forma menos invasiva. Na literatura foram encontrados estudos que seguem essa linha de pesquisa na busca da identificação de sinais de depressão (GARCIA-CEJA et al., 2018; CHIKERSAL et al., 2021; LIU et al., 2021).

Para que as pessoas possam se beneficiar do uso dessas tecnologias, se faz necessário investigar tal comportamento. Por conta disso, nesta dissertação, elegeu-se pensar em uma solução computacional para a identificação do possível perfil depressivo na população brasileira, usando para isso, dispositivos móveis com sensores capazes de coletar dados de fenotipagem digital, como celulares e relógios inteligentes. O cenário de interesse escolhido para estudo nesta pesquisa foi o ambiente universitário, onde a prevalência de depressão em estudantes parece ser maior do que no restante da população (LAUCKNER; HILL; INGRAM, 2018; IBRAHIM et al., 2013).

### 1.3 Objetivo

O objetivo desta dissertação é a proposição de uma infraestrutura computacional<sup>1</sup> para coleta e armazenamento de dados que permitam a identificação de tendências de comportamento de estudantes universitários com possível perfil depressivo. Isso é feito através da análise de sinais de fenotipagem digital advindos de sensores móveis que coletam dados no cotidiano de indivíduos com e sem depressão. Este trabalho também visou avaliar se os dados permitem uma distinção entre os grupos, de forma que a análise respalde a criação de uma base de conhecimento científico sobre a depressão no contexto universitário brasileiro, apoiando pesquisas futuras.

O interesse de estudo com a depressão voltada para a população universitária brasileira também se deve a outros trabalhos que constataram que a doença afeta um número significativo de pessoas no Brasil. Ibrahim et al. (2013) realizaram uma revisão sistemática de estudos sobre prevalência de depressão em estudantes universitários em todo o mundo e encontram que a prevalência média de sintomas depressivos foi de 30,6%, no Brasil. Pacheco et al. (2017), em um estudo similar utilizando a escala BDI, também encontraram prevalência de 30,6% de sintomas depressivos entre os estudantes universitários brasileiros.

O Plano Nacional de Assistência Estudantil (PNAES) estabelece como um dos seus eixos a atenção à saúde do estudante de graduação (Decreto n 7.234, 2010). No entanto, o campo da Atenção Psicossocial<sup>2</sup> nas universidades, de forma geral, é incipiente (BLEICHER; OLIVEIRA, 2016). No mais, apesar de se ter na literatura alguns estudos sobre o uso de tecnologias como apoio à Atenção Psicossocial no ambiente universitário, pouco se vê sobre a utilização de tecnologias computacionais no auxílio deste cuidado no Brasil. Diante do exposto, compreende-se que a pesquisa em questão favorecerá a identificação e a busca ativa de casos possivelmente depressivos entre estudantes universitários brasileiros, favorecendo a intervenção precoce, na intenção de maior eficácia na atenção à saúde desta população.

Além do uso dos recursos de sensores móveis, entende-se que as opções de integração de tais dispositivos, quando em uso conjunto com algoritmos de reconhecimento de padrão de comportamento, permitem fornecer intervenções personalizadas aos usuários em tempo real. Um exemplo é o trabalho de (EVMENOVA et al., 2019) que usou relógios inteligentes para apoiar a participação de estudantes com Deficiência Intelectual e Desenvolvimento (DID) em ambientes acadêmicos do ensino superior. No contexto da depressão, esses padrões podem ser reconhecidos com o auxílio de escalas de avaliação psicométricas<sup>3</sup>, que

<sup>1</sup> De acordo com o Glossário de Tecnologia da Informação do National Institute of Standards and Technology (NIST), uma infraestrutura computacional abrange equipamentos interconectados, sistemas, software e procedimentos usados na aquisição, armazenamento, processamento, controle e transmissão automática de dados. Isso inclui o uso direto por agências executivas ou contratados, abrangendo computadores, equipamentos auxiliares, serviços e recursos relacionados.

<sup>2</sup> Atendimentos para pessoas com problemas mentais graves, incluindo uso de álcool e outras drogas.

<sup>3</sup> Estabelece uma relação entre os sintomas (no caso, a depressão) e o comportamento do indivíduo. Geralmente aplicado como forma de questionários ou lista de itens que formam uma escala de avaliação.

permitem criar uma relação direta com os sinais depressivos possíveis de se detectar com esses dispositivos (SWARTZ, 2005).

Este trabalho compõe uma parte do modelo multifatorial de identificação de PPD de um projeto maior, denominado Amigo Virtual Especializado (Amive). O projeto Amive visa investigar principalmente aspectos da Interação Humano-Computador (IHC), apoiando a reabilitação de usuários depressivos, usando para isso, dispositivos com sensores móveis (tema desta pesquisa) e textos publicados pelo indivíduo nas redes sociais. O Amive é um sistema *opt-in*<sup>4</sup>, materializado em um aplicativo para celulares, pelo qual o usuário tem o conhecimento de que seus dados estão sendo analisados e aceita receber intervenções geradas automaticamente por ele. De forma objetiva, este trabalho de mestrado visa contribuir com o projeto do Amive entregando a infraestrutura computacional para coleta e estudos dos dados dos participantes universitários, juntamente a investigação e análise do comportamento PPD a partir de dados dos sensores móveis obtidos para a pesquisa.

O projeto Amive é financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), teve início em março de 2021 e término em setembro de 2023. A primeira fase de coleta de dados com estudantes ocorreu na metade de setembro de 2022 e terminou na metade de outubro do mesmo ano. Durante esse período de 30 dias, foi possível coletar dados dos participantes voluntários da pesquisa usando a infraestrutura computacional construída para essa fase do projeto.

Todos os dados com as informações coletadas pelos participantes da pesquisa durante o período estão armazenados em um banco de dados, que visando a contribuição científica deste trabalho, estão disponibilizadas para que pesquisadores possam estudar e usar para pesquisas futuras (Capítulo 7.2). Ainda, tais dados foram processados e analisados com o objetivo de prover uma conclusão para este estudo, como é detalhado no capítulo de resultados (6).

Para isso, foi desenvolvido um aplicativo dedicado ao projeto, nomeado aplicativo Amive. O aplicativo coletou dados de fenotipagem digital por meio de sensores móveis, como os fornecidos pelo próprio celular do usuário e também de relógios inteligentes que foram disponibilizados pelo projeto Amive para um número limitado de participantes. Ainda, o aplicativo usou escalas psicométricas em formato de questionários e também realizou coleta de postagens das redes sociais do Facebook e Twitter para Processamento de Linguagem Natural (PLN). Os dados armazenados foram submetidos à análise estatística, passando por um processo de filtragem de dados, usando como base os algoritmos mencionados no levantamento bibliográfico (Capítulo 3.1).

Esta dissertação de mestrado identifica padrões e comportamentos caracterizados pelo PPD usando dados advindos de sensores móveis. Seus resultados têm como objetivo apoiar

---

<sup>4</sup> Quando o usuário aceita conscientemente fornecer seus dados e concorda com as possíveis intervenções às quais estará sujeito (COPPERSMITH et al., 2018).

as ações de pesquisa do projeto Amive, que por sua vez, tem como objetivo construir uma infraestrutura computacional usando um modelo multifatorial para identificação do PPD.

## 1.4 Síntese da metodologia

Este estudo adota a abordagem *Design Science Research (DSR)* (HEVNER; CHATTERJEE, 2010) como método de pesquisa, que auxilia o desenvolvimento tecnológico em pesquisas científicas por seguir princípios e critérios científicos no processo de design de forma sistemática e rigorosa. É particularmente relevante na seção que discute o levantamento bibliográfico (Capítulo 3) e a infraestrutura da pesquisa (Capítulo 4), onde o desenvolvimento do aplicativo e a estrutura para o servidor são detalhados, além de sua aplicação na seção de análise estatística (Capítulo 6) permitindo a justificativa das decisões de seleção de algoritmos com base em evidências prévias e teóricas.

O DSR é uma abordagem metodológica proeminente área da tecnologia da informação, que se concentra na criação e avaliação de artefatos destinados a solucionar problemas práticos. Tais artefatos, no contexto do DSR, referem-se a entidades criadas pelo homem, como modelos, métodos, ferramentas, ou sistemas, que são desenvolvidos para resolver problemas específicos. O processo envolve um ciclo iterativo de desenvolvimento, começando pela compreensão dos problemas e necessidades, seguida pela definição de objetivos de design. Em seguida, ocorre a construção do artefato, que pode ser um sistema, modelo, método ou teoria. Após a construção, o artefato é avaliado em relação aos critérios científicos estabelecidos. Nesta pesquisa, tal metodologia fornece um caminho sistemático para desenvolver e justificar as soluções tecnológicas propostas, sendo essencial para a criação da infraestrutura computacional do trabalho, que por sua vez, permitiu o desenvolvimento do aplicativo usado para coleta de dados e a análise resultante da pesquisa.

A condução e aplicação prática da metodologia do DSR neste estudo, partiu do problema que o projeto Amive queria resolver: a crescente prevalência de depressão em ambientes acadêmicos brasileiro. Com isso veio a identificação do problema, que se dá pela necessidade e falta de métodos para sua detecção. Com base nisso, o objetivo do artefato foi estabelecido em desenvolver uma infraestrutura tecnológica, composta por sensores em dispositivos móveis, de forma que um algoritmo de análise de dados poderia inferir indicadores relevantes do comportamento depressivo. No estágio de projeto e desenvolvimento, o sistema foi concretizado com a integração de sensores de celulares e relógios inteligentes, coletando dados de fenotipagem digital e questionários de saúde mental. A fase de demonstração foi realizada através da coleta de dados com participantes durante um período de 30 dias, evidenciando a funcionalidade prática do sistema. Por fim, a avaliação dos resultados coletados permitiu verificar a eficácia dos dados coletados pela infraestrutura, com algoritmos que permitiram a análise estatística identificar padrões de comportamento associados ao perfil depressivo.

A estratégia adotada para coleta e análise de dados é dado de acordo com conhecimento adquirido no levantamento do estado da arte realizado no mapeamento sistemático desta dissertação e apresentado no Capítulo 3. De forma geral, os dados dos participantes são coletados passivamente sem a necessidade de interação direta do usuário com o dispositivo. Os dados são enviados para um serviço de armazenamento em um servidor seguro e posteriormente submetidos ao aprendizado de máquina para análise e identificação dos sinais de depressão. Esse padrão pode ser identificado em Chikersal et al. (2021), Lu et al. (2018), Ware et al. (2020) e em outros trabalhos referenciados no mapeamento. Isso se deve por ser um modelo que aproveita outros recursos disponíveis no próprio dispositivo de coleta que o participante está usando. O mesmo celular que coleta e envia informações de forma passiva para o servidor, também pode ser usado como um instrumento de aplicação prática dos questionários de saúde mental.

Para esta pesquisa, que trabalhou em conjunto ao projeto Amive, adotou-se o uso dos celulares e relógios inteligentes para a coleta de dados. O aplicativo Amive, desenvolvido no âmbito deste projeto foi instalado no celular dos participantes e recebeu dados de fenotipagem digital de frequência cardíaca, tempo de sono, contagem de passos e atividades físicas (como exercícios e esportes) realizadas no dia. Esses dados foram originalmente extraídos de um aplicativo auxiliar, Samsung Health, antes de serem enviados ao aplicativo Amive. O aplicativo Amive também fez o uso de escalas psicométricas com questionários de saúde mental, sentimento subjetivo e avaliação do usuário em relação ao que ele entendia sobre os seus dados coletados. O sistema também coletou dados das redes sociais do Twitter e Facebook. Apesar do projeto ter envolvido a coleta de informações sobre redes sociais, esse sistema compõe um dos módulos do modelo multifatorial do projeto Amive como um todo, que não abrange os interesses deste trabalho, e portanto, não será abordado aqui.

Esta pesquisa realizou o empréstimo de 28 relógios inteligentes aos participantes. O modelo dos relógios foi o Samsung Galaxy Watch 4. Fora orientado que os participantes enviassem seus dados duas vezes ao dia, enquanto os questionários eram disponibilizados em períodos específicos pelo aplicativo. O envio de postagem das redes sociais também faziam parte da rotina do participante ao usar o aplicativo (abordado com detalhes no Capítulo 4). Embora não faça parte da proposta deste estudo, o aplicativo também coletou postagens realizadas pelos participantes nas redes sociais, sendo elas o Twitter e o Facebook. As postagens do Twitter eram coletadas todos os dias de forma automática. Dados de postagens do Facebook deveriam ser baixados e enviados para o Amive pelos participantes sempre que tivessem feito alguma nova atividade na rede, exigindo assim uma interação a mais do usuário com o aplicativo.

Para compreender os sinais de depressão em participantes da pesquisa e regerar os resultados finais da pesquisa, validando o modelo proposto da infraestrutura desenvolvida para o trabalho, foi escolhido a abordagem de análise exploratória descritiva. Essa me-

metodologia envolve uma investigação detalhada dos dados coletados por meio de sensores móveis e questionários respondidos pelos participantes durante o uso do aplicativo, com o objetivo de identificar indicadores e padrões associados à depressão. Nesse contexto, a proposição da infraestrutura desenvolvida para o estudo viabilizou a coleta e o processamento eficiente dos dados, e também foi essencial na execução da análise exploratória. A seguir, é detalhado as etapas e estratégias adotados para essa metodologia, que parte desde a análise univariada até a redução de dimensionalidade dos dados, com o objetivo de criar uma análise para padrões de usuários com PPD e não PPD.

## 1.5 Síntese do resultado

A proposta central deste trabalho consistiu na proposição e o desenvolvimento de uma infraestrutura computacional, servida como um modelo para coletar dados de fenotipagem digital provenientes de sensores móveis, como celulares e relógios inteligentes, no contexto universitário brasileiro. Essa infraestrutura se concretiza como um aplicativo para celulares, apresentada aos participantes como um aplicativo de saúde mental. A análise exploratória descritiva dos dados coletados foi usada como um instrumento para avaliar a eficácia da infraestrutura proposta. A análise demonstra a capacidade da infraestrutura de capturar e interpretar padrões comportamentais de sintomas depressivos nos dados coletados.

Foram usados os dados de 89 participantes para o desenvolvimento desta etapa da pesquisa, sendo esses dados coletados pelo relógio inteligente, celular e também pelo questionário *Patient Health Questionnaire-9* (PHQ-9) enviados pelos participantes. Os resultados apresentados foram obtidos a partir dos dados de um grupo de 33 participantes, proveniente de um processo de imputação de dados (detalhado no Capítulo 6) para lidar com quaisquer valores ausentes, garantindo uma precisão maior nos resultados gerados.

A análise exploratória foi direcionada a visualização de três grupos:

1. Todos os participantes do estudo
2. Grupo identificado com PPD
3. Grupo identificado como não PPD

Os grupos foram classificados de acordo com resultado do PHQ-9.

O estudo da análise exploratória foi realizado em três etapas, passando pelas abordagens univariada, seguindo para análise bivariada com a correlação entre as variáveis e, finalmente, uma análise multivariada. Essas etapas ajudaram a compreender e identificar possíveis diferenças entre indivíduos com depressão e sem depressão com base nas variáveis estudadas. Foram elas:

1. Análise exploratória univariada: o objetivo foi estudar apenas um grupo e analisar os resultados usando gráficos de histograma, bloxplot e de densidade para cada variável dos sensores.
2. Correlação entre as variáveis (análise bivariada): foi construído uma matriz de correlação entre as variáveis da análise univariada, gerada através da correlação de Spearman e, em seguida, filtrado e comparado as matrizes com os dados por perfil depressivo e não depressivo.
3. Análise exploratória multivariada: foi aplicado a técnica t-SNE para reduzir a dimensionalidade dos dados, a fim de compactar variáveis em duas dimensões e gerar um gráfico para visualizar mais facilmente os padrões dos indivíduos.

A conclusão da análise exploratória gerou um relatório final no Capítulo 6, na qual foi possível documentar todos resultados. Este relatório inclui gráficos relevantes e conclusões, destacando os padrões identificados e as relações observadas entre as variáveis dos participantes identificados como PPD e não PPD. Os resultados e o processo desenvolvido durante a análise têm o propósito de servir como orientação para as próximas etapas das ações da comunidade científica que buscam a compreensão e inferência dos sinais e diagnósticos da depressão por meio da computação. As informações obtidas a partir desta análise exploratória fornecem conhecimentos para direcionar estudos adicionais, intervenções e compreensão sobre saúde mental, contribuindo para o avanço no entendimento da depressão e para o desenvolvimento de abordagens mais eficazes em termos de prevenção e possível tratamento.

### 1.5.1 Disponibilização da base de dados

Este estudo também toma a iniciativa de tornar a base de dados (banco de dados) resultante deste estudo acessível à comunidade científica, juntamente ao código-fonte com o conjunto de dados dos participantes usados para gerar resultados da análise exploratória. Em suma, o banco de dados (PostgreSQL) contém os dados de todos 89 participantes, anonimizados, com suas informações enviadas pelos sensores móveis e pelos questionários usados na pesquisa.

Além da transparência com a pesquisa, a disponibilização da base de dados permite que outros pesquisadores examinem e verifiquem os métodos aplicados e os resultados obtidos. Os dados disponibilizados podem servir como um recurso para o desenvolvimento de novas hipóteses e estudos relacionados à depressão. Outros pesquisadores podem explorar esses dados sob diferentes ângulos e contextos, levantando questões e investigando aspectos que talvez não tenham sido considerados inicialmente. A disponibilização desses recursos pode ajudar na busca de conhecimentos inovadores tanto para a área da computação quanto para a de saúde mental.

## 1.6 Organização do documento

A segunda parte deste documento (Capítulo 2) apresenta os conceitos da Fundamentação Teórica deste estudo, sendo introduzido o contexto dos aspectos da doença mental da depressão e como a tecnologia pode ser usada para ajudar as pessoas. Neste mesmo capítulo é apresentada, de forma objetiva, a computação vestível, e como ela pode ser usada como meio de integração para ajudar a identificar os sinais de alerta em relação a saúde mental, como a depressão.

A seção que diz respeito aos Trabalhos Relacionados (Capítulo 3) detalha a pesquisa feita por meio de um mapeamento sistemático para conhecer o estado da arte na detecção de sinais depressivos utilizando sensores de dispositivos móveis e vestíveis. Essa seção aborda a explicação e os motivos do termo de busca escolhido para o mapeamento e seu processo de amadurecimento, juntamente das bases de dados e o porquê de que terem sido escolhidas. É demonstrado também o processo de seleção dos artigos provenientes das bases de dados, seguido da análise dos artigos, resultados e a discussão sobre as estratégias de abordagem escolhida por cada trabalho. Por fim, os trabalhos são comparados. A seção do levantamento bibliográfico apresenta gráficos e tabelas de estatísticas sobre o levantamento de informações de estudos que buscam a identificação de depressão usando dispositivos móveis e quais foram as metodologias utilizadas para chegar a este resultado.

A tecnologia empregada neste estudo também é apresentada nas seções seguintes. Em Infraestrutura (Capítulo 4), são apresentadas as estratégias adotadas, limitações e desafios enfrentados na construção da infraestrutura. Este capítulo detalha o desenvolvimento do aplicativo para *smartphones* que seria usado mais tarde pelos participantes da pesquisa, e segue detalhando a estrutura especificada para o servidor, e o planejamento para receber, tratar e processar os dados provenientes dos participantes.

O Capítulo 5 relata como foi realizado do Estudo com Usuários. Detalha-se todas as etapas do processo de interação com os participantes com a pesquisa, indo desde a orientação dos voluntários sobre como usar o aplicativo e o relógio inteligente oferecido para o estudo, até a coleta e armazenamento dos dados. Este capítulo aborda as diretrizes da pesquisa, o processo de anonimização e proteção de dados dos participantes, problemas e adversidades enfrentadas durante a fase de coleta, suporte e apoio aos participantes da pesquisa, encerramento do estudo com usuários e o procedimento para organizar e armazenar os dados coletados.

Esta pesquisa traz conclusões apresentando a infraestrutura desenvolvida como resultado deste trabalho de mestrado (Capítulo 6), em conjunto da análise estatística como um ferramenta para avaliar a eficácia dos dados coletados. Assim, o capítulo é dedicado à análise estatística detalhada dos algoritmos e estratégias selecionadas para esta avaliação, com justificativas dos resultados obtidos a partir da análise.

Em seguida, é retomado os objetivos gerais e específicos da proposta deste estudo, são sumarizados os dados e resultados, e apresentados a conclusão e expectativa de contri-

buição científica da obra. Por fim, esta dissertação se encerra com uma síntese de todo o trabalho realizado e também uma breve discussão sobre as limitações encontradas durante o processo de estudo e desenvolvimento da pesquisa. Ainda no capítulo de Conclusão (7) serão disponibilizados os recursos usados em cada etapa da pesquisa para serem baixados.



---

## Capítulo 2

# Conceitos fundamentais

---

### 2.1 Depressão

O Transtorno Depressivo Maior (TDM), comumente conhecido como **depressão**, é um problema de saúde comum, mas muito grave, que segundo o DSM-5 (DSM, 2022), envolve sintomas como humor deprimido e a perda do interesse ou prazer em grande parte do dia. Somado ao humor deprimido, ainda pode haver a presença de desânimo ou anedonia, redução ou aumento do apetite (pode provocar a perda/ganho de peso), alterações no sono (hipersonia ou insônia), sentimento de culpa e/ou inutilidade, diminuição da autoestima, agitação psicomotora ou retardo, perda de concentração e dificuldades para pensar. Isso afeta uma pessoa física, emocional e socialmente. Além disso, depressão e ansiedade são os principais contribuintes para suicídios no mundo todo (WHO, 2017) por gerar ideação suicida e pensamentos constantes sobre a morte. Ainda, a depressão pode ser confundida com a tristeza, seja ela sazonal ou proveniente do luto. Nesses casos, luto e tristeza podem induzir um grande sofrimento, mas não costumam provocar um episódio de transtorno depressivo maior e se esvai em alguns dias. Embora exista tratamento para a depressão, incluindo uma variedade de métodos, como psicoterapia e medicação, a maior parte das pessoas afetadas pela doença não procura por tratamento (EISENBERG; GOLBERSTEIN; GOLLUST, 2007).

As diferentes teorias do psiquismo e modelos de cuidado resultam em diversos caminhos terapêuticos para o cuidado da depressão. Eles incluem a medicalização de problemas sociais, econômicos, políticos e de diferentes contextos. Além disso, há uma profunda lacuna na prestação de cuidados entre diferentes países (KOHN et al., 2004). As razões para esta lacuna incluem, entre outras, a falta de consciência de que existe um problema.

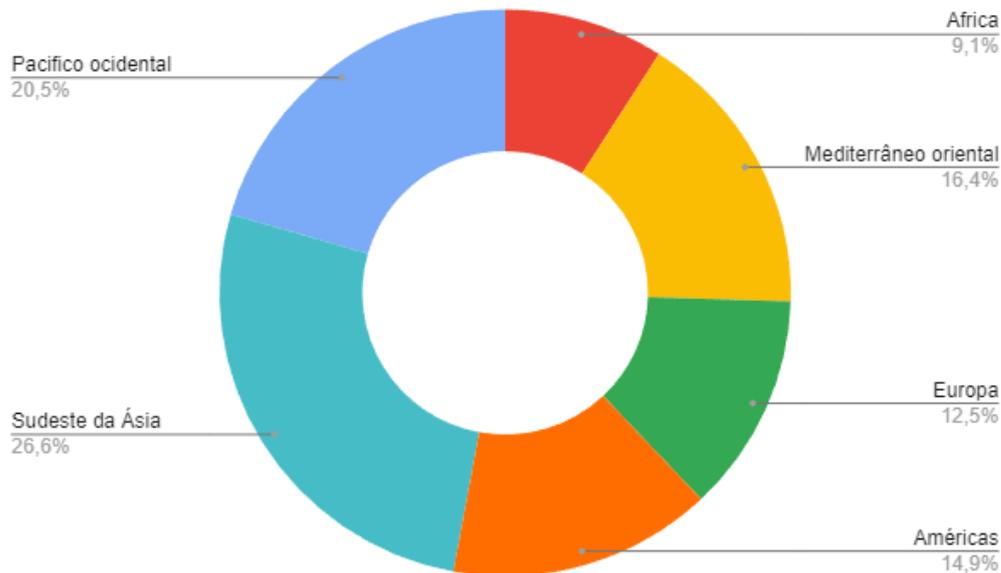
Segundo dados da Organização Mundial da Saúde (OMS), a proporção da população

global com depressão em 2015 foi estimada em 4,4%. A doença da depressão é considerada um grande problema para a população em todo o mundo, levando a oito mil mortes por suicídio a cada ano (AL., 2018). De acordo com dados levantados por Nguyen et al. (2021) (citação com tradução adaptada pelo autor):

Nos Estados Unidos, mais de 20% da população experimentaram uma doença de saúde mental em 2016. Em 2015, quase 11% da população na Europa morreu por ter uma doença de depressão grave. Na última pesquisa nacional, havia 9,3% de pacientes com depressão grave entre a população chinesa. Na Austrália, uma pesquisa de saúde mental e bem-estar (*Australian Bureau of Statistics*), realizada em 2007, mostrou que 45% da população sofreu um transtorno mental em toda a sua vida e 4,8 milhões de pessoas (20,1%) com um transtorno comportamental ou mental em 2017-2018, o que aumenta de 2,6% pontos de 2014 a 2015. Na África, os problemas de saúde mental estão aumentando também. De 2000 a 2015, os anos perdidos com incapacidade aumentaram em 52% devido a transtornos mentais e por uso de substâncias, enquanto a população do continente cresceu em 49%.

Isso demonstra que a depressão tem um impacto mundial e que afeta todos os países. A Figura 1 demonstra o impacto da depressão ao redor do globo.

Figura 1 – Estimativas globais e regionais de prevalência de depressão



Fonte: (WHO, 2017)

Além das questões de saúde, a depressão também tem um impacto econômico mundial. A depressão está diretamente associada a um enorme fardo econômico, cujo maior componente deriva da perda de produtividade no trabalho pelo doente (WANG; SIMON; KESSLER, 2003). Mesmo nas sociedades economicamente mais favorecidas, decisões difíceis precisam ser tomadas sobre como alocar recursos limitados de saúde, como tratamentos hospitalares e ambulatoriais,

além dos custos indiretos que podem surgir por conta da mortalidade atribuível à doença. Essas decisões são ainda mais complicadas em países menos desenvolvidos socioeconomicamente.

Diante deste problema, pesquisadores em computação vêm buscando criar soluções que possam apoiar o diagnóstico e intervenções para pessoas que podem possivelmente estar sofrendo com sintomas depressivos. Nos países desenvolvidos, houve progresso na detecção de sintomas corporais por meio de tecnologia vestível, que pode auxiliar os sistemas de prestação de cuidados de saúde, visando fornecer atendimento precoce e facilitar o monitoramento remoto (COSTA et al., 2019; LIN et al., 2012). Parece haver potencial em empregar tecnologias móveis para lidar com problemas de saúde mental para entender melhor a influência do contexto e do comportamento na formação de sintomas depressivos. Kumar et al. (2021) faz um mapeamento sobre os sensores encontrados em tecnologias móveis e como eles podem se relacionar para identificar sintomas relacionados a problemas de saúde mental (incluindo a depressão). Tais dispositivos também podem contribuir para identificar como as respostas biofisiológicas estão associadas aos sintomas depressivos.

Para auxílio na identificação e rotulação dos sintomas depressivos, as escalas psicométricas são amplamente utilizadas pelos trabalhos estudados nesta dissertação, presentes em 17 dos 19 artigos estudados no mapeamento sistemático. As escalas de psicométricas são escalas de avaliação ou questionários já consolidados que trazem um diagnóstico rápido da situação de saúde mental do participante. Dos estudos, é uma prática comum e eficiente calcular as características dos sintomas a partir dos dados de detecção passiva, coletados pelos dispositivos móveis e realizar uma análise de correlação entre as características dos sintomas e os resultados dos questionários de depressão (PEDRELLI et al., 2020; WARE et al., 2020; WANG et al., 2018a). As escalas mais utilizadas pelos trabalhos foram o PHQ-9, *Quick Inventory of Depressive Symptomatology* (QIDS) e *Hamilton Depression Rating Scale* (HDRS).

Trazendo esse contexto para o objetivo deste trabalho, unido com a proposição de uma infraestrutura tecnológica para coletar e analisar dados relacionados à saúde mental, este trabalho também visa a detecção de padrões comportamentais associados ao perfil depressivo. Tal infraestrutura é integrado a um aplicativo que se comporta como um sistema de monitoramento baseado em sensores de celulares e também de relógios inteligentes, com o intuito de identificar um PPD, gerando informações que possam apoiar o diagnóstico da depressão.

## 2.2 Sensores móveis

Nos dias de hoje, pode-se usar sensores móveis para monitorar a saúde pessoal, seja com celulares ou mesmo relógios inteligentes. Por conta desses dispositivos, algumas pessoas estão gerando grandes quantidades de dados todos os dias, seja para fins de melhoria de sua qualidade de vida, seja para supervisionar seus níveis de condicionamento físico ou até mesmo melhorar seus hábitos e ter uma rotina mais saudável. Embora comumente esses dados sejam usados para medir a quantidade de passos diários, calorias queimadas e repetições de exercícios realizados no dia, eles têm muito potencial para inferir um estado de saúde (mental e físico) do indivíduo, uma vez que são geralmente coletados registros contínuos de frequência cardíaca, resposta galvânica

da pele (GSR)<sup>1</sup>, monitoramento e qualidade do sono, e diversos outros tipos de níveis de atividade corporal. Há uma crescente conscientização no campo da psiquiatria sobre como esses dados de atividade se relacionam com vários problemas de saúde mental, como mudanças de humor e personalidade, incapacidade de lidar com problemas diários, estresse, afastamento de amigos e perda interesse/prazer em atividades físicas ou sociais (PENEDO; DAHN, 2005; GARCIA-CEJA et al., 2018). Isso demonstra que, se antes era necessário que uma pessoa comum fosse até um consultório realizar uma grande bateria de exames com equipamentos clínicos para conseguir um diagnóstico dos seus problemas de saúde físico e mental, hoje temos tecnologias acessíveis na palma da nossa mão e na superfície de nossos pulsos, capazes de não apenas identificar esses possíveis sintomas, como também prever e oferecer intervenções. Apesar de eficientes e práticos, essas tecnologias não substituem os equipamentos clínicos. Sua grande vantagem é o acompanhamento do indivíduo em situações do cotidiano dele.

Essas tecnologias pervasivas, têm a vantagem de coletar dados dos usuários continuamente, de maneira objetiva. Apesar de cara para algumas pessoas, tal abordagem é discreta e não estigmatizante, pois requer dispositivos convencionais que não se assemelham a equipamentos médicos. Além disso, geralmente é menos propensa a erros, por confiar na coleta automatizada de dados em vez de autorrelatos ou *recall* (retorno) individual do paciente. Há um grande potencial para entender melhor a influência do contexto e do comportamento na formação de sintomas depressivos ao empregar tecnologias com sensores móveis para lidar com problemas de saúde mental. Esses sensores têm sido explorados no monitoramento de estudantes universitários para detecção passiva de comportamentos que possam indicar sintomas depressivos (WANG et al., 2016).

Os dispositivos com sensores móveis, como celulares e relógios inteligentes, permitem coletar informações de forma passiva e discreta, como localização, comunicação, ambiente, uso do telefone, atividade física, sono e outros. Trabalhos anteriores mostraram que tais informações estão ligadas a sintomas depressivos, como isolamento social e distúrbios do sono (DSM, 2022). Usando esses sensores, pode-se medir a gravidade dos sintomas depressivos com a possibilidade de desenvolver um modelo computacional com o potencial para intervenções em tempo real, mediadas por tecnologia que apoiam o diagnóstico, tratamento e prevenção da depressão.

Por se tratar de uma tecnologia comumente presente nos dispositivos que as pessoas usam no seu dia a dia, muitos pesquisadores da indústria e da academia têm investigado soluções baseadas em tais sensores para monitorar a pessoa que vive de forma independente e extrair características que podem se traduzir em sintomas de depressão. Entre eles, um estudo foi realizado na Universidade de Connecticut e eles acompanharam a rotina diária de 79 estudantes por 5 meses (FARHAN et al., 2016). Nesse estudo, apenas dados de GPS (localização) foram coletados, durante 214 dias, para extrair recursos de movimento, e o uso desse recurso mostrou uma correlação com a gravidade dos sintomas de depressão descritos no PHQ-9 (SAEB et al., 2016). Farhan et al. (2016) afirmam “Descobrimos que dados comportamentais de smartphones podem prever depressão clínica com boa precisão”.

Conforme os critérios do DSM-5 (DSM, 2022), um dos principais sintomas que dão indícios de que uma pessoa pode estar sofrendo de depressão é o humor deprimido ou perda de

<sup>1</sup> A resposta galvânica corresponde ao nível de condutância da pele, indicando ansiedade e estresse.

interesse/prazer. Juntamente a eles, também devem estar presentes os sintomas secundários do transtorno depressivo, que incluem: aumento ou diminuição de peso, alterações de sono provenientes de insônia ou hipersonia, agitação ou retardo psicomotor, fadiga ou perda de energia, diminuição da capacidade de pensar ou concentrar-se, sentimentos de inutilidade ou culpa excessiva, e por fim, pensamentos de morte com tendências suicidas. Esses sintomas podem ser identificados por meio de sensores móveis e vestíveis, como relógios inteligentes. Todos os trabalhos da Tabela 3 identificaram, ao menos, um dos sintomas descritos. A Tabela 1 apresenta uma possível relação dos sensores de um relógio inteligente (Samsung Galaxy Watch4<sup>2</sup>, usado nesta pesquisa) e os sintomas do DSM-5 que podem ser identificados com estes sensores.

Tabela 1 – Possível relação de sensores do relógio inteligente com sintomas descritos no DSM-5

Sintomas DSM-5	Dados coletados por sensores
Humor depressivo	Captura de áudio, mensagens de texto, frequência cardíaca
Perda de prazer/interesse	Registros de chamadas telefônicas, mensagens de texto, GPS, contagem de passos, tempo de uso do dispositivo
Perda/ganho de peso	Monitoramento de ingestão de alimentos
Insônia ou hipersonia	Monitoramento de sono, sensores de movimento (acelerômetro, giroscópio)
Agitação ou retardo psicomotor	GPS, contagem de passos, atividade física, frequência cardíaca
Fadiga, perda de energia	GPS, contagem de passos, atividade física, frequência cardíaca
Sentir-se inútil ou culpa excessiva	Captura de áudio, mensagens de texto
Diminuição/perda da concentração	Tempo de uso do dispositivo e seus aplicativos
Pensamentos de morte/suicídio	Captura de áudio, mensagens de texto

Fonte: própria

Cientistas da computação e da área da saúde (muitos casos em equipes multidisciplinares) se dedicaram para desenvolver trabalhos que estudassem a relação dos sintomas com os dados coletados por sensores. A seguir, é exemplificado a relação entre os sintomas específicos dos itens do DSM-5 (e também da Tabela 1) e as características dos dados extraídos por alguns estudos que foram selecionados para o mapeamento sistemático deste trabalho (Tabela 3).

- a) **Insônia ou hipersonia:** Dai et al. (2022) utilizou a captura de dados do monitoramento de sono de um relógio inteligente para identificar sintomas depressivos nos participantes. Além dos sensores que fazem a identificação direta dos sintomas depressivos, também é possível fazê-lo de maneira indireta, como (KIM; HONG; CHOI, 2021) identificou os

<sup>2</sup> Ficha técnica, com informações sobre os sensores, disponível em: <<https://www.samsung.com/br/watches/galaxy-watch/?galaxy-watch4>>

mesmos sintomas através do bloqueio/desbloqueio de tela do celular dos participantes de sua pesquisa;

- b) **Agitação ou retardo psicomotor:** Ware et al. (2020) fez uso do *Global Positioning System (GPS)*<sup>3</sup> de celulares e de redes Wi-Fi para identificar diferentes informações de localização dos participantes, como locais visitados, tempo gasto se movimentando, distância percorrida e outros. Com essas informações aplicadas ao aprendizado de máquina para comparar dados de participantes não depressivos, o autor conclui “Nossos resultados, usando dados coletados de 182 estudantes universitários, demonstraram que um rico conjunto de sintomas depressivos pode ser previsto com precisão usando dados de smartphones.”;
- c) **Fadiga, perda de energia:** Masud et al. (2020b) identificou tais sintomas por meio do número reduzido de atividades físicas, entropia (muito tempo parado, com falta de energia ou indisposição para se locomover), tempo despendido em casa, transição entre ambientes, entre outros, e para isso, usou apenas o celular dos participantes. Esses mesmos dados também foram usados para identificação de sintomas do item anterior (agitação ou retardo psicomotor), assim como *Ware et al.* também utilizou mais dados de sua pesquisa para identificação de fadiga e perda de energia;
- d) **Diminuição/perda da concentração:** Wang et al. (2018b) através do bloqueio e desbloqueio da tela do celular dos participantes (estudantes universitários) em períodos específicos, como na sala de aula ou em áreas de estudo, identificou a perda da concentração em atividades. Segundo Wang “Quando um aluno está na sala de aula ou nas áreas de estudo, ele deve se concentrar no trabalho ou no estudo. Nesses locais, presumimos que o uso maior do celular (ou seja, passar mais tempo em seus celulares) pode indicar que eles estão tendo dificuldade em se concentrar em seu trabalho.”

Os demais sintomas, como humor depressivo, perda de prazer/interesse, perda/ganho de peso, sentimento de inutilidade ou culpa excessiva, e pensamentos de morte/suicídio, não foram obtidos dados de forma direta por meio dos sensores móveis, embora possa ser possível obtê-los através da frequência cardíaca (HARTMANN et al., 2019) e até mesmo pelo clima do tempo (KLIMSTRA et al., 2011). Como auxílio na identificação dos demais sintomas que são difíceis de serem obtidos por meio da interpretação dos dados dos sensores, é comum o uso de escalas psicométricas. Outros trabalhos também estudaram os padrões de interações sociais baseadas em dispositivos móveis usando o número de chamadas telefônicas e mensagens de textos (PEDRELLI et al., 2020; LIU et al., 2021). O uso do processamento de linguagem natural pode identificar padrões de um PPD por meio do processamento de textos e mensagens (ZUORBA; OLAN; CANTARA, 2017), porém, esses são dados muito invasivos e por motivos de privacidade, nenhuma informação de uso do telefone foi registrada, incluindo conteúdo de chamadas/textos, nomes de aplicativos e conteúdo de uso da Internet. Ao mesmo tempo em que ter acesso a mais informações sobre o indivíduo poderia ajudar na identificação de sinais depressivos, também é

<sup>3</sup> Sistema de posicionamento global, amplamente usado para localização geográfica de um objetos por meio de satélites.

preciso manter os limites da privacidade do usuário e, por conta disso, é necessário adequar as estratégias de coleta e de pesquisa para que a privacidade dos participantes seja preservada.

Todavia, é importante salientar que os dados brutos que são coletados por esses dispositivos móveis não necessariamente apresentam uma correlação direta com os sintomas depressivos. É necessário que esses dados sejam processados e relacionados para que eles possam ser interpretados. No trabalho realizado com o aplicativo Amive, todos os dados obtidos pelos sensores móveis eram em formatos numéricos que estariam relacionados com o dado específico daquele sensor e sua data de criação. Por exemplo, dados de frequência cardíaca são fornecidos pelo sensor no formato: “75.5,2022-09-18T18:05:53” (dados em formato *CSV*<sup>4</sup>), onde o primeiro valor representa a taxa de amostragem da frequência cardíaca (75.5) e o segundo valor (2022-09-18T18:05:53) a data e hora em que o dado foi criado no dispositivo, respectivamente. Esse é um padrão que se repete em outros dados coletados, como contagem de passos (exemplo “852,2022-10-01T05:08:23” para o número de passos contados e a data de criação), exercícios (exemplo “412,2022-09-18T17:35:11” para o tipo do exercício praticado e data de criação), e também para o monitoramento de sono (exemplo “2022-09-18T22:49:53,2022-09-19T07:00:02,80” para data de início do sono, data de término e nota da qualidade do sono). Com os dados em formato CSV, é possível separá-los por seus respectivos índices e armazenar todos os valores no banco de dados para serem treinados mais tarde. Esse formato de obtenção de dados é muito comum e possível observar em outros estudos (KIM; HONG; CHOI, 2021; DAI et al., 2022; FUKUDA et al., 2020).

## 2.3 Análise estatística

Esta seção proporciona uma compreensão teórica das técnicas utilizadas no estudo, servindo para estabelecer uma base conceitual sólida para as análises estatísticas que serão realizadas posteriormente no Capítulo 6. A análise exploratória descritiva aplicada a este trabalho aborda três etapas de análise, sendo a análise individual das variáveis (abordado neste estudo como análise exploratória univariada), estudo de correlação entre as variáveis (análise bivariada) e a redução de dimensionalidade (análise multivariada). O objetivo é entender como os dados coletados dos participantes da pesquisa podem estar relacionados, de forma que nos tragam percepções sobre o comportamento de indivíduos e grupos.

### 2.3.1 Análise de variáveis

A investigação de cada variável isoladamente é abordada neste estudo como análise exploratória univariada. Nesse processo, cada variável em um conjunto de dados é examinada individualmente, sem considerar suas interações com outras variáveis. Essa abordagem utiliza técnicas estatísticas e visualizações gráficas para descrever e resumir as propriedades de cada variável. O objetivo dessa análise é obter as características das variáveis e entender sua distribuição, para logo em seguida realizar uma comparação de centralidade ou dispersão entre os grupos da análise.

---

<sup>4</sup> Tipo de documento que padroniza a separação de dados por vírgula.

Essa análise gera tipos diferentes de gráficos que ajudam a interpretar e entender os dados de forma visual. Para isso, foram usados os gráficos de histograma, de boxplot e de densidade. Os histogramas permitem visualizar a distribuição dos valores da variável, enquanto os boxplots revelam a dispersão e assimetria de dados em quartis<sup>5</sup>, e por fim, o gráfico de densidade ilustra a probabilidade da variável em diferentes pontos do seu intervalo. Essas representações visuais nos auxiliam na identificação de como os dados estão dispostos, se há assimetrias, onde a maior concentração de dados ocorre, e se existem valores pertencentes a um padrão.

### 2.3.2 Estudo de correlação

Uma das formas de entender como uma variável pode depender da outra, é realizando um estudo de correlação. No estudo de correlação, o objetivo é compreender se as variáveis estão relacionadas entre si e como influenciam umas às outras, dessa forma, se tornando uma análise bivariada. A análise bivariada é uma abordagem estatística que envolve a avaliação simultânea de duas variáveis em um conjunto de dados, como no caso onde temos quatro tipos de variáveis para análise em três grupos diferentes.

A análise de correlação envolve técnicas estatísticas que quantificam a relação entre duas ou mais variáveis em um conjunto de dados. Para quantificar a correlação entre as variáveis, o coeficiente de correlação de Pearson<sup>6</sup> foi o utilizado nos trabalhos relacionados desta dissertação para avaliar relações lineares entre variáveis. No entanto, os dados desta pesquisa não seguem uma distribuição normal, tornando-os não lineares, e para isso, a correlação de Spearman<sup>7</sup> se demonstrou mais apropriada para ser aplicada neste estudo.

O resultado da correlação é representada em uma tabela, que deve ser interpretada com os valores das variáveis entre -1 e 1. Se um valor estiver perto de 1, isso significa que as duas variáveis têm uma correlação positiva; quando uma aumenta, a outra também aumenta. Se estiver perto de -1, há uma correlação negativa; quando uma variável aumenta, a outra diminui. Se o valor estiver próximo de 0, a correlação é fraca ou inexistente.

### 2.3.3 Redução de dimensionalidade

Com o objetivo de tornar essa uma análise exploratória multivariada, no contexto dos dados analisados, é possível fornecer uma visualização melhor da correlação entre todas as variáveis em um gráfico tridimensional (3D). No entanto, gráficos 3D podem distorcer as percepções, tornando a análise complicada, e para simplificar e tornar a visualização mais clara, pode-se usar técnicas de redução de dimensionalidade. Para isso, será introduzida a técnica t-SNE ((t-distributed Stochastic Neighbor Embedding)<sup>8</sup>(MAATEN; HINTON, 2008), uma abordagem que permite a identificação e a representação de padrões não tão aparentes na análise bivariada dos

<sup>5</sup> Divisão de um conjunto de dados em quatro partes iguais, destacando a mediana e a dispersão.

<sup>6</sup> Medida estatística que avalia a relação linear entre duas variáveis contínuas, indicando a força e a direção dessa relação.

<sup>7</sup> Medida estatística que avalia a relação monotônica entre duas variáveis, fornecendo informações sobre a tendência geral da relação, independentemente da sua linearidade.

<sup>8</sup> Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

dados, possibilitando a visualização das relações complexas e das interdependências entre as variáveis.

A interpretação de um gráfico t-SNE envolve a análise das posições relativas dos pontos de dados, e não dependem das etiquetas dos eixos X e Y. Essa técnica preserva as relações de similaridade local entre os pontos, de forma que os pontos próximos possuem características semelhantes, enquanto aglomerados de pontos podem indicar diferentes classes ou grupos de dados.

Na função algorítmica usada para gerar o t-SNE, há um hiperparâmetro chamado *perplexity* (perplexidade), crucial na configuração do algoritmo, pois controla a escolha do número de vizinhos considerados para cada ponto de dados durante a construção do espaço de menor dimensão. A perplexidade é uma medida que reflete a incerteza na distribuição de probabilidades conjuntas nos dados originais e no espaço de menor dimensão e, de acordo com a medida escolhida, os gráficos podem trazer resultados diferentes.

Ainda, é possível aplicar técnicas de normalização de dados para verificar o comportamento dos dados em análises multivariadas. A normalização envolve a transformação das variáveis originais para uma escala comum, a fim de garantir que todas as variáveis tenham o mesmo peso na análise. Por exemplo, transformar medidas, como os passos de caminhada — que podem variar de 0 até 10.000 — em um valor de mesmo peso que as horas de sono — que variam entre números 4 a 8. No entanto, a normalização pode não influenciar significativamente os resultados do t-SNE, uma vez que o mesmo realiza uma redução de dimensionalidade de maneira intrínseca, que permite identificar padrões complexos nos dados, independentemente das escalas das variáveis.



---

## Capítulo 3

# Trabalhos relacionados

---

### 3.1 Levantamento bibliográfico

O levantamento bibliográfico realizado neste trabalho visou analisar estudos recentes pautados pela área da saúde. Foram analisados os aspectos metodológicos, éticos e técnicos de informática voltados para a detecção do possível perfil depressivo por meio do uso de sensores móveis e vestíveis. O método de pesquisa adotado foi o **mapeamento sistemático** (KITCHE-NHAM, 2004). Um mapeamento sistemático da literatura visa reunir publicações científicas relacionadas ao tema de pesquisa específico, de modo a identificar o estado da arte e abranger os conhecimentos do tema em que a pesquisa está sendo feita. Este mapeamento sistemático também responde às perguntas pré-definidas sobre o tema:

1. O número de publicações relacionados a identificação de características depressivas por meio de dispositivos móveis aumentou ao longo do tempo?
2. Que abordagens (dados coletados, relação com a depressão e algoritmos empregados) têm sido utilizadas para a coleta de dados?
3. Qual foi o tamanho da amostra? Qual foi a duração dos estudos?
4. Os profissionais de saúde mental foram envolvidos no estudo (equipes multidisciplinares)?
5. Que precauções e considerações foram tomadas para prevenir riscos e danos aos usuários e seus dados?

#### 3.1.1 Termo de busca

O seguinte termo de busca foi usado para encontrar trabalhos relacionados ao tema de depressão e sensores móveis: *(depression OR depressive) AND (smartwatch OR wearable OR sensing OR monitoring)*

Visando ser inclusivo, foram buscados artigos que mencionassem tecnologias pervasivas atuais como *smartwatch* (relógio inteligente) e *wearable* (tecnologia vestível), em geral. Também foi adicionado *monitoring* (monitoramento) para incluir qualquer outra tecnologia abrangente criada para esse fim. Não foi usado a palavra-chave *pervasive* (pervasivo) ou *passive* (passivo) no termo de pesquisa, pois são termos abrangentes, portanto, optou-se por termos específicos, como *wearable*, *monitoring* e *smartwatch*. Ainda, na primeira parte do termo de busca, inicialmente estava sendo usado *mental health* (saúde mental) para aumentar a área de busca com sintomas parecidos com a depressão, como estresse ou ansiedade, porém essas palavras foram removidas na *string* final, pois muitos artigos não condiziam com o propósito da pesquisa.

### 3.1.2 Fonte de dados

Os artigos foram buscados em cinco bases de dados diferentes, sendo elas: ACM Digital Library, IEEE Digital Library, PubMed Central (PMC), Scopus e também Springer Link. Essas bibliotecas digitais incluem artigos relacionados a ciência da computação (ACM), engenharia (IEEE), medicina (PubMed) e também áreas de pesquisas multidisciplinares (Spring e Scopus).

### 3.1.3 Critério de seleção

Após pesquisar e encontrar os artigos buscados nas bases de dados utilizando o termo de busca, foi necessário filtrá-los através de critérios que visam selecionar ou descartar artigos relacionados a pesquisa. Esses são os critérios de inclusão e exclusão.

Para os critérios de inclusão, foram considerados estudos que:

- a) Apresentam dados de usuário que são usados em um modelo de classificação do PPD;
- b) Discutem e apresentam algoritmos para classificação do PPD;
- c) Apresentam dados relacionados a saúde mental e que ajudem na classificação do PPD;

Esses critérios foram adotados para incluir o maior número possível de artigos que trabalharam com a identificação do possível perfil depressivo usando dados reais de participantes, priorizando estudos que coletaram diretamente esses dados ao invés de usar informações de alguma base de dados pronta. O critério C permite a inclusão de artigos que podem ter estudado outro tipo de saúde mental, como o estresse (MISHRA, 2019), mas que esteja relacionado com o perfil depressivo de forma que se possa usar isso como base de estudo para este trabalho.

Para os critérios de exclusão, foram adotadas as seguintes regras:

- a) O trabalho não aborda depressão ou aspecto que não pode ser relacionado com depressão;
- b) Estudo fora do escopo;
- c) Não utiliza tecnologias móveis ou vestíveis de uso comum;

Estudos fora do escopo representam os trabalhos não completos (como resumos) ou não-primários (revisões) e também artigos que não tinham nenhuma relação com a área de pesquisa (ruídos de má formação do termo de busca em algumas bases de pesquisa). Ainda, o critério de exclusão C teve que ser adicionado após verificar-se que havia tipos de tecnologias móveis

e vestíveis que não se relacionavam com o que pessoas têm usado no dia a dia ou tecnologias comercialmente indisponíveis (camisetas, cintas, chapéus inteligentes, etc.).

### 3.1.4 Análise

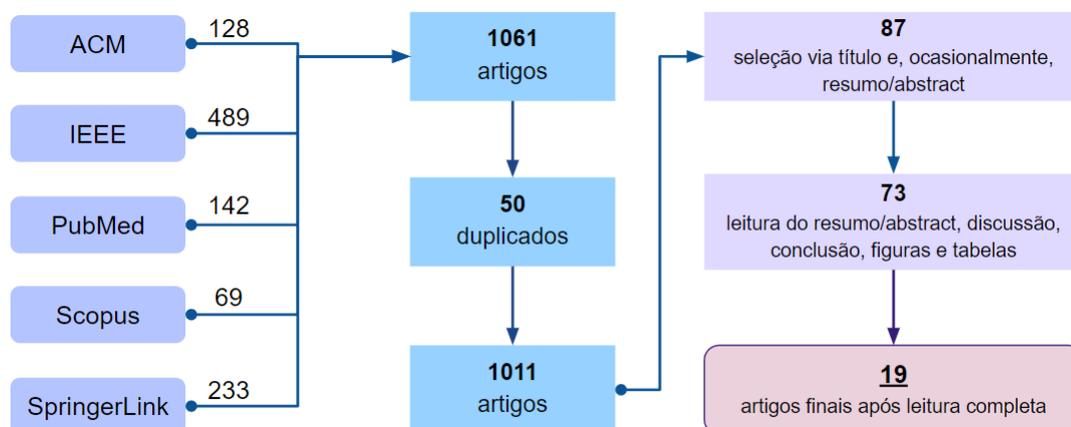
De todos os artigos selecionados, foi feita uma primeira leitura e seleção pelo título e resumo da publicação e, caso o artigo atendesse aos critérios de inclusão para este trabalho, também seria feita a leitura do artigo na íntegra. A ferramenta Parsifal<sup>1</sup> foi usada para gerenciar a busca, organizar a análise dos dados e acompanhar as etapas seguintes.

Com base no termo de busca, filtrando apenas por título, resumo e palavras-chave, foram extraídos os estudos primários que foram publicados entre 2015 e 2022 nas bases de dados citadas anteriormente. As buscas foram realizadas em 1 de outubro de 2022. Nesta primeira fase da pesquisa, 1.061 estudos foram recuperados e, em seguida, 50 artigos duplicados foram removidos. Após a remoção dos duplicados, a primeira fase constituiu em analisar os artigos aplicando os critérios de inclusão e exclusão pela leitura do título e, ocasionalmente, do resumo/abstract. Com essa análise, dos 1.011 artigos selecionados, 924 foram descartados e 87 foram aceitos e seguiram para a próxima etapa.

A segunda etapa consistiu em uma leitura do resumo/abstract dos artigos, juntamente das seções de conclusão, discussão e análise de figuras e tabelas. Dos 87 artigos remanescentes, foram descartados 14, restando ainda 73 estudos que seguiram para a terceira etapa. Por fim, na terceira etapa, foi feita uma leitura completa de 73 artigos, sendo 54 descartados, chegando a conclusão de 19 artigos finais selecionados para relato neste mapeamento sistemático.

A Figura 2 ilustra o processo de seleção por sequência de cada etapa e o número de artigos considerados em cada uma, enquanto a Tabela 2 demonstra o número de artigos obtidos e selecionados em cada base de dados.

Figura 2 – Processo de seleção dos artigos



Fonte: própria

<sup>1</sup> Disponível em <https://parsif.al/>

Tabela 2 – Quantidade de artigos selecionados por base de dados

Base de dados	Aceitos	Descartados	Duplicados	Total
ACM	8	114	6	128
IEEE	6	473	10	489
PubMed	2	124	16	142
Scopus	3	48	18	69
Springer Link	0	233	0	233

Fonte: própria

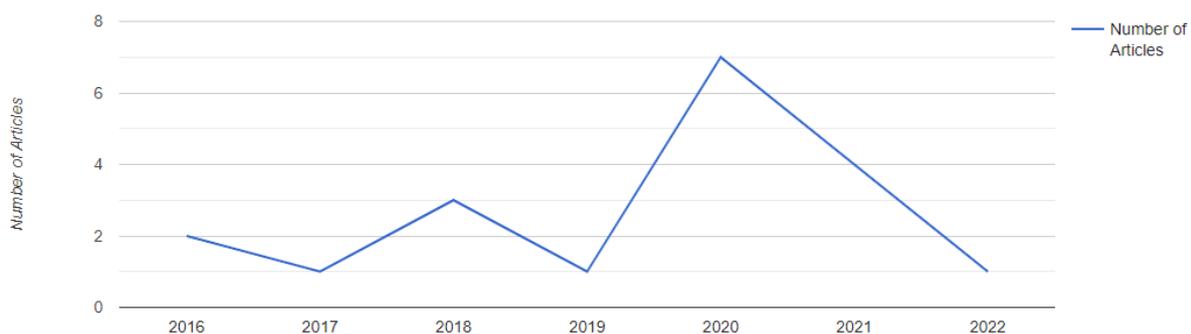
### 3.1.5 Resultados

Com todos os 19 artigos analisados e estudados, as perguntas da pesquisa puderam então ser respondidas:

- a) **O número de publicações relacionadas a identificação de características depressivas através de dispositivos móveis aumentou ao longo do tempo?**

Sim. Houve um aumento nos estudos relacionados ao tema a partir do período de 2018 (com exceção de 2019), com um pico em 2020, como mostra a Figura 3. Como essa pesquisa foi feita em outubro de 2022, pode ser que os números finais aumentem até o início do próximo ano.

Figura 3 – Número de publicações por ano relacionadas a identificação do PPD com uso da tecnologia



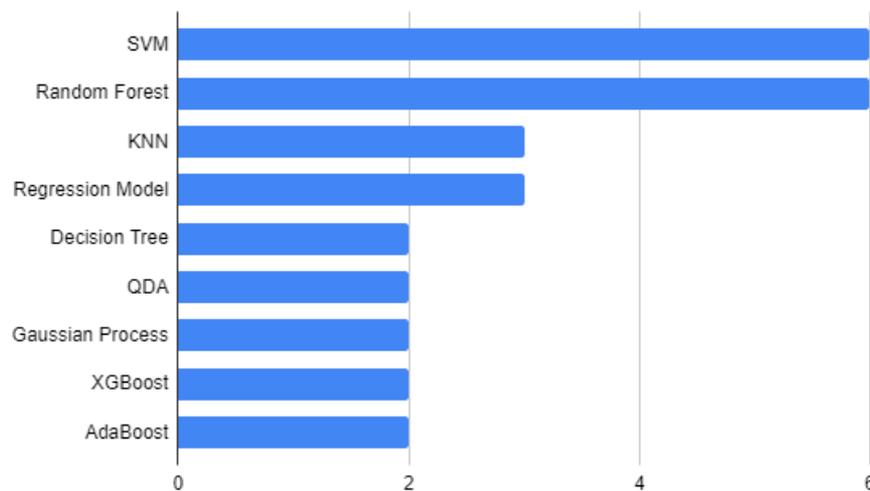
Fonte: gráfico de análise de dados gerado automaticamente pelo Parsifal representando o número de artigos por ano com base nos artigos selecionados.

- b) **Que abordagens (dados coletados, relação com a depressão e algoritmos empregados) têm sido utilizadas para a coleta de dados?**

Os estudos utilizaram estratégias muito similares, onde os dados dos indivíduos foram coletados em determinados períodos específicos do dia, principalmente por conta da limitação de uso da bateria dos relógios, e de forma passiva pelos relógios inteligentes,

sem a interação direta do usuário com o dispositivo. Após a coleta, os dados foram enviados para o celular do indivíduo (geralmente via Bluetooth, mas alguns trabalhos utilizaram a transferência por Wi-Fi), e uma vez no celular, os dados foram enviados para um servidor seguro onde foram processados de acordo com o interesse da pesquisa.

Figura 4 – Algoritmos de aprendizado de máquina mais utilizados



Fonte: própria

Outros algoritmos de aprendizado de máquina, além dos citados, foram utilizados em estudos de forma pontual (SSA, ANN, LOOCV, Boruta, Naïve Bayes, entre outros). Os algoritmos SVM e Random Forest foram usados em conjunto em metade dos trabalhos. A Figura 4 demonstra os algoritmos de aprendizado de máquina que foram utilizados nos trabalhos do mapeamento. Os estudos utilizaram o aprendizado de máquina nos dados coletados dos participantes para identificar os sintomas de depressão, geralmente, acompanhado de um questionário de saúde mental. Pedrelli et al. (2020) fez uma correlação direta dos modelos de predição resultante dos algoritmos em relação aos resultados do questionário de saúde mental escolhido para o estudo, usando para isso, as próprias métricas de score do questionário para criar uma visualização de desempenho de cada algoritmo. Kilaskar et al. (2021) fez um estudo analisando diferentes algoritmos de aprendizado de máquina, e os comparou selecionando vários parâmetros e mostrando qual algoritmo é mais preciso para prever a depressão. Os melhores resultados se deram aos algoritmos XGBoost e regressão logística. No entanto, outros algoritmos também foram aplicados e tiveram bons resultados, como AdaBoost, Random Forest, Decision Tree e SVM. Esses outros algoritmos também são amplamente utilizados no treinamento de dados para identificação da depressão, e também são citados na figura 4.

c) **Qual foi o tamanho da amostra? Qual foi a duração dos estudos?**

Em modo geral, os estudos apresentam um número médio de 50 participantes (variando entre 20 a 80 pessoas geralmente), e o tempo de coleta e amostragem de dados segue

também o número de participantes, sendo uma amostragem maior para os estudos com maior número de participantes, e uma menor amostragem para um estudo com menor número. Vale ressaltar que dos demais estudos que não entraram nos critérios de inclusão, mas que foram lidos durante o processo, diversos deles não coletaram necessariamente dados diretos de participantes, mas fizeram uso de aplicativos e outras bases de dados para apenas testar modelos de treinamento com algoritmos de aprendizado de máquina.

**d) Os profissionais de saúde mental foram envolvidos no estudo (equipes multi-disciplinares)?**

Exceto pelos trabalhos de Mishra (2019), Ware et al. (2020), que contou com psicólogos para o estudo clínico, os demais artigos não citaram o envolvimento direto de profissionais da saúde para estudo e avaliação dos dados coletados pelos dispositivos. A forma mais comum de avaliar e diagnosticar o perfil de saúde mental dos participantes nos estudos foi por meio de questionários que autoavaliam e classificam o estado de saúde mental do indivíduo. Alguns outros trabalhos também usaram estudos clínicos, mas apenas para acompanhamento dos resultados de escalas psicométricas e não para avaliação dos resultados provenientes dos dispositivos móveis (FARHAN et al., 2016; LU et al., 2018).

**e) Que precauções e considerações foram tomadas para prevenir riscos e danos aos usuários e seus dados?**

Não foi citado nenhum tipo de plano de contingência para prevenção de riscos ou falhas nos estudos analisados, tanto para os estudos selecionados através dos critérios de inclusão, quanto dos estudos lidos e removidos pelos critérios de exclusão. Porém, durante alguns dos estudos analisados, houve pequenas falhas na coleta de dados, como o que ocorreu com Wang et al. (2018a), no qual os dados coletados de temperatura da pele não foram precisos por conta da influência do clima, uma vez que o estudo foi realizado no inverno. Nesse caso, os dados simplesmente não puderam ser utilizados e foram excluídos das análises da pesquisa.

### 3.1.6 Discussão

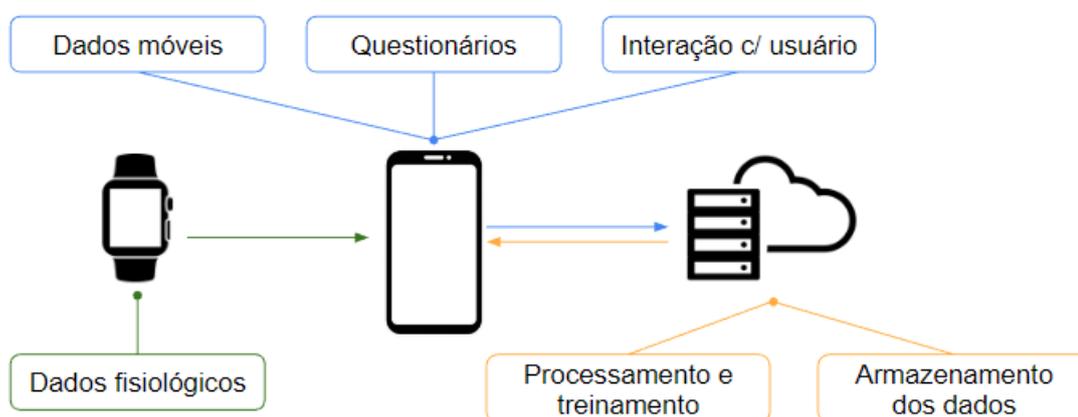
Com a leitura e análise dos artigos, foi possível elencar alguns pontos de reflexão em relação ao estado da arte e verificar possibilidades de estudos para trabalhos futuros. Um dos pontos a se citar, é a limitação do tempo de uso dos relógios inteligentes. Dos estudos selecionados, todos aqueles que trabalharam com relógios precisaram gerir de forma inteligente e eficaz a medição do tempo de uso dos dispositivos em relação a sua bateria, uma vez que quanto mais o dispositivo coleta dados, mais rápido sua bateria é consumida, diminuindo assim a taxa de amostragem de dados coletados. Ainda se tratando dos relógios inteligentes, foi possível perceber a partir dos estudos analisados, que apenas alguns dados são usados de forma recorrente, entre elas, a frequência cardíaca, monitor de atividades físicas e sono são os mais comuns. Além disso, nos estudos, em nenhum momento foi citado que houve a interação dos indivíduos com o relógio em si, e sim, apenas uma coleta de dados passiva por meio do dispositivo. De forma geral, os dados extraídos pelos relógios eram enviados para o celular, para que dessa forma, o celular pudesse

transferir as informações para algum servidor, onde os dados eram processados e traziam as informações de volta para os celulares. Essa foi a estratégia comum utilizada pelos estudos para a coleta de dados.

Todos os estudos que utilizaram o método de coleta de dados dos participantes para identificação e predição do tratamento das doenças mentais, se basearam em questionários que realizam o autodiagnóstico de forma individual da pessoa para treinar e avaliar seus modelos preditivos com base nos dados coletados (exceto os artigos P06 e P07). A maior parte dos artigos utilizaram os questionários derivados do PHQ, seguido do QIDS.

Com a análise dos artigos selecionados para o mapeamento sistemático que compõe a base desta pesquisa, foi possível chegar à identificação de alguns padrões usados pelos trabalhos que buscam identificar sinais de depressão com o uso de dispositivos móveis. O primeiro deles é que os relógios inteligentes estão sendo utilizados para coleta de dados passivos sem a intervenção dos participantes, e que os sensores dos relógios inteligentes podem ser associados aos comportamentos depressivos. Os questionários de saúde mental são amplamente utilizados para se ter uma medida de comparação entre o diagnóstico resultado pelas escalas psicométricas e os resultados obtidos após o processamento dos dados coletados pelos dispositivos. Os celulares foram utilizados como dispositivo intermediador para envio das informações do vestível para o servidor, e também são os responsáveis pela entrega dos questionários de saúde mental para os participantes (isso pode ser observado na Figura 5 a seguir). O processamento de dados é realizado usando aprendizado de máquina, testando mais de um algoritmo para uma avaliação da acurácia de predição, com o objetivo de identificar padrões quando comparados aos dados resultantes das escalas psicométricas.

Figura 5 – Generalização da coleta e tráfego de informações entre os dispositivos dos estudos analisados



Fonte: própria

Em síntese, a infraestrutura desenvolvida para pesquisa realizada neste trabalho, levou em consideração o aprendizado e o conhecimento obtido ao investigar os padrões usados por outros estudos que se comprometeram a identificar o possível perfil depressivo, tanto na etapa de coleta de dados, que já foi realizada, quanto nas etapas de processamento de dados. Também

se investigou os padrões comportamentais no contexto universitário brasileiro na intenção de favorecer a atenção à saúde desta população.

Tabela 3 – Comparação entre os trabalhos selecionados do mapeamento sistemático

ID	Referência	Participantes	Dias	Dispositivos	Sensores	Sintomas	Algoritmos	Escalas
P01	(KIM; HONG; CHOI, 2021)	48	70	Smartphone	Desbloqueio de tela	Alteração do sono	SVM, Random Forest e XGBoost	PHQ-9 e DSM-4
P02	(FARHAN et al., 2016)	79	214	Smartphone	GPS	Agitação ou retardo; fadiga e perda de energia	SVM	PHQ-9
P03	(NGUYEN et al., 2021)	60	70	Smartphone	Ativ. física, áudio, GPS, Bluetooth, bateria, desbloqueio de tela e Wi-Fi	Humor depressivo; agitação ou retardo; fadiga e perda de energia	MVBC algorithm	PHQ-9
P04	(GARCIA-CEJA et al., 2018)	23	12.6	Smartwatch (Activewatch AW4)	Freq. cardíaca e acelerômetro	Agitação ou retardo; fadiga e perda de energia	KNN, SVM, Gaussian Process, Decision Tree, Random Forest, ANN, AdaBoost, Naive Bayes, QDA e ZeroR	MADRS
P05	(CHIKERSAL et al., 2021)	138	112	Smartphone e smartwatch (Fitbit Flex 2)	Contagem de passo, monitor de sono, Bluetooth, GPS, desbloqueio de tela, chamadas e áudio	Agitação ou retardo; fadiga e perda de energia; alteração do sono; perda de interesse ou prazer;	Regression Model e AdaBoost	BDI-II
P06	(MISHRA, 2019)	26	3	Smartwatch (Polhar H7)	GSR	<b>Estresse</b>	-	EMA
P07	(LIU et al., 2021)	711	364	Smartphone e smartwatch (FitBit)	Contagem de passos, monitor de sono e mensagens SMS	Perda de interesse ou prazer; agitação ou retardo; fadiga e perda de energia;	-	-
P08	(LU et al., 2018)	103	98	Smartphone e smartwatch (FitBit)	Freq. cardíaca, monitor de sono, ativ. física e GPS	Agitação ou retardo; fadiga e perda de energia; perda de interesse ou prazer; alteração do sono	-	QIDS
P09	(PEDRELLI et al., 2020)	21	63	Smartphone e smartwatch (E4 Empatica)	EDA, GSR, freq. cardíaca, acelerômetro, monitor de sono, chamada, mensagens SMS, ativ. físicas e uso do celular	Humor depressivo; perda de interesse ou prazer; agitação ou retardo; fadiga e perda de energia; diminuição e perda de concentração	Boruta algorithm	HDRS-17 e DSM-4
P10	(DAI et al., 2022)	106	367	Smartwatch (Fitbit Alta HR)	Freq. cardíaca, ativ. física e monitor de sono	Agitação ou retardo; fadiga e perda de energia; alteração do sono	SSA	-

P11	(MASUD et al., 2020b)	33	77	Smartphone	Mensagens SMS, chamadas, desbloqueio de tela, acelerômetro, giroscópio e GPS	Agitação ou retardo; fadiga e perda de energia; diminuição e perda de concentração	QDA, Regression Model e LOOCV	QJDS
P12	(GHANDEHARIOUN et al., 2017)	12	56	Smartphone e smartwatch (Empatica E4)	Acelerômetro e EDA, chamadas, mensagens SMS, uso do celular, desbloqueio de tela e GPS	Humor depressivo; agitação ou retardo; fadiga e perda de energia; diminuição e perda da concentração; alteração do sono; perda de interesse ou prazer	PCA, SVD, Regression Model, Random Forest, Gaussian Process	HDRS
P13	(JACOBSON; CHUNG, 2020)	31	-	Smartphone	GPS, freq. cardíaca e chamadas	Humor depressivo; perda de interesse ou prazer; agitação ou retardo; fadiga e perda de energia	XGBoost	DASS-14
P14	(FUKUDA et al., 2020)	60	14 21	Smartwatch (FitBit Charge 3)	Monitor de sono	Humor depressivo; agitação ou retardo; fadiga e perda de energia; alteração do sono	Random Forest	DAMS
P15	(WARE et al., 2020)	182	518	Smartphone	GPS e Wi-Fi	Agitação ou retardo; fadiga e perda de energia;	SVM	PHQ-9, QJDS e DSM-5
P16	(NARZIEV et al., 2020)	20	28	Smartphone e smartwatch (Gear S3 Frontier)	Acelerômetro, contador de passos, ativ. física, freq. cardíaca, monitor de sono, chamadas, mensagens de SMS	Humor depressivo; perda de interesse ou prazer; alteração do peso; agitação ou retardo; fadiga e perda de energia;	Random Forest e SVM	PHQ-9
P17	(MEHROTRA; HENDLEY; MUSOLESI, 2016)	25	30	Smartphone	Notificações e uso do celular	Diminuição ou perda de concentração	-	PHQ-8
P18	(WANG et al., 2018a)	83	63	Smartphone e smartwatch (Microsoft Band 2)	Ativ. físicas, monitor de sono, áudio, desbloqueio da tela, freq. cardíaca, GSR e ativ. física	Humor depressivo; perda de interesse ou prazer; agitação ou retardo; fadiga e perda de energia	Modelo de regressão	PHQ-8 e DSM 5
P19	(MASUD et al., 2020a)	33	77	Smartphone	Acelerômetro, giroscópio e GPS	Agitação ou retardo; fadiga e perda de energia;	SVM, KNN e ANN	PHQ-9

Fonte: própria

---

# Capítulo 4

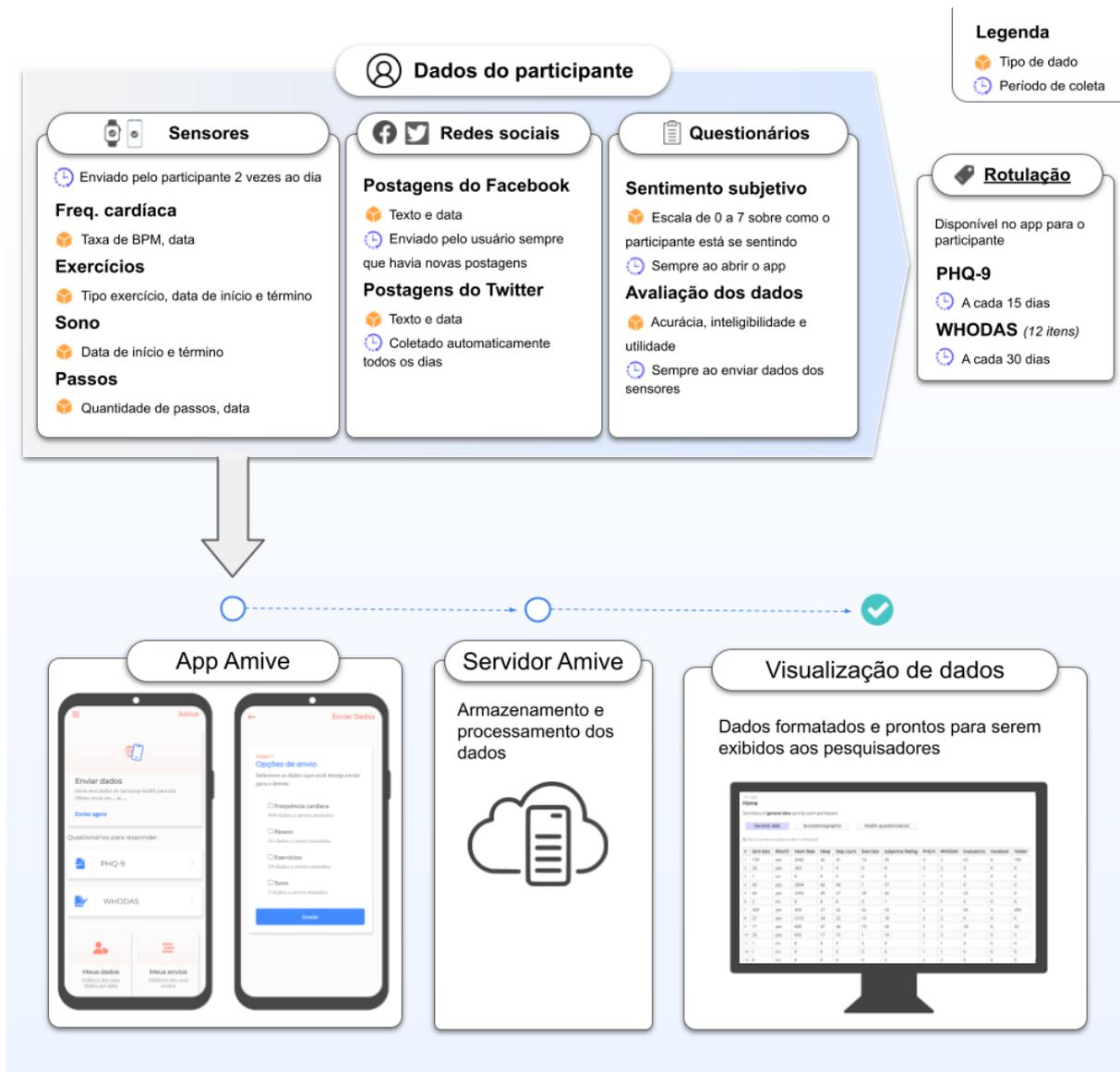
## Infraestrutura

---

A infraestrutura do projeto Amive foi composta por diversos sistemas que se integram para entregar a solução proposta pela pesquisa. Foi desenvolvido um aplicativo para celulares Android para que os participantes da pesquisa pudessem enviar seus dados. O aplicativo realiza uma integração com o aplicativo Samsung Health, que é um aplicativo móvel desenvolvido pela Samsung que oferece recursos para monitorar a saúde e o bem-estar dos usuários, permitindo o acompanhamento de atividades físicas, sono, frequência cardíaca, etc. As informações geradas pelo monitoramento do Samsung Health ficam armazenados no dispositivo do usuário e podem ser baixados a qualquer momento, e foi usando este recurso que a pesquisa obteve os dados de fenotipagem digital dos participantes. O aplicativo Samsung Health foi instalado nos celulares de todos os participantes, e também nos 28 relógios inteligentes que foram emprestados para a pesquisa. Dessa forma, o aplicativo Amive se tornava responsável por receber os dados do relógio e do celular (incluindo os questionários de saúde mental) e enviar essas informações para serem armazenadas no servidor do Amive. Os dados armazenados foram usados nesta dissertação para gerar o capítulo final de avaliação dos resultados. Tais dados também foram disponibilizados em um outro sistema interno chamado Amive Web, onde as informações ficam disponíveis para os pesquisadores visualizarem e estudarem de forma simplificada a amostra dos dados coletados na primeira fase da pesquisa.

Inicialmente, a ideia principal da solução seria focada apenas no relógio inteligente vestido pelo participante, de forma que ele pudesse tanto coletar como também enviar os dados e interagir com o usuário. No entanto, diversos obstáculos apareceram no desenvolvimento de um aplicativo dedicado ao relógio que tornaria limitado a coleta e transferência de dados de acordo com o planejado para o projeto. Após realizado o estudo com os trabalhos relacionados, foi eleita uma nova solução de como a infraestrutura do projeto poderia ser montada, como ilustra a Figura 6. Com essa estrutura, o relógio inteligente atua apenas como um dispositivo de coleta passiva dos dados de fenotipagem digital do participante, enquanto seu celular se torna responsável por ter acesso a essas informações armazenadas no relógio e enviá-las para ao servidor

Figura 6 – Infraestrutura do projeto Amive



Fonte: própria

por meio do aplicativo Amive. O servidor Amive também pode fazer envios de informações e respostas para o celular do participante, como por exemplo, informar se um novo questionário já está disponível ou não. A solução de tornar a interação do participante unicamente via celular, se demonstrou ainda mais eficaz do que a ideia inicial, onde o foco seria no relógio inteligente, pois dessa forma, o celular ainda nos ofereceu a flexibilidade de utilizar ele mesmo como um dispositivo secundário de coleta passivo, removendo a limitação de que apenas participantes com relógio poderiam participar da pesquisa.

O aplicativo Amive desempenhou um papel central na condução da pesquisa. Sua principal função incluiu a integração com o relógio inteligente, permitindo a coleta e transferência de dados armazenados no dispositivo para o servidor do Amive. Além disso, o aplicativo promoveu

a interação com os participantes, onde o usuário poderia consultar seu histórico de envios, avaliar os dados coletados e receber orientações sobre o envio correto dos dados. O aplicativo também foi responsável pela disponibilização dos questionários de saúde mental aos participantes, sendo eles, PHQ-9, *WHO Disability Assessment Schedule 2.0* (WHODAS-II) e sentimento subjetivo, com uma interface adaptada para celulares, na intenção de tornar o processo de preenchimento e envio dos questionários simples e rápido. Por último, o aplicativo permitiu o envio dos dados textuais postados pelos usuários no Facebook. Em suma, todas as interações foram feitas por meio do aplicativo Amive, de forma que o participante não precisasse configurar ou interagir diretamente com o relógio, bastando que o mesmo apenas o vestisse durante o dia pelo maior tempo que pudesse.

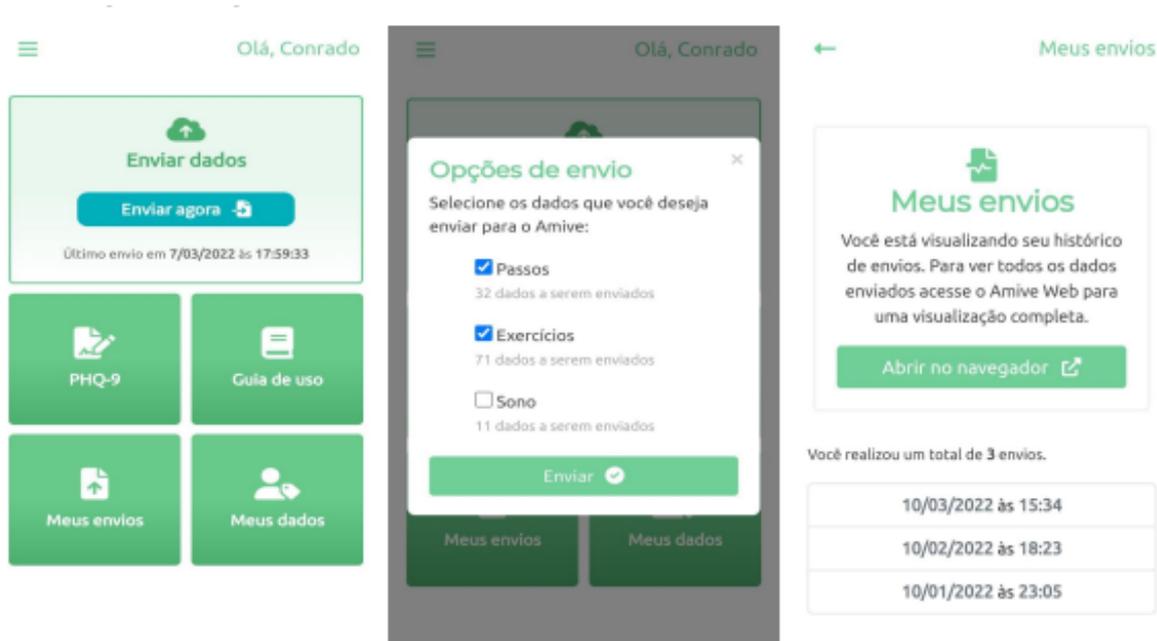
O desenvolvimento do aplicativo Amive passou por diversas etapas, que foram desde as decisões de design adotadas para o aplicativo até as dificuldades e limitações encontradas no momento de extração e coleta de dados, que alteraram a forma de funcionamento do software e, conseqüentemente, sua interface. Inicialmente, pensava-se em duas plataformas, sendo o aplicativo apenas para entrada de dados do participante, e uma solução web, onde o participante acessaria um site para visualizar suas informações enviadas. Essa estrutura foi abandonada, e decidiu-se integrar tudo dentro de um mesmo aplicativo, onde o participante pudesse tanto enviar seus dados como visualizá-los.

Para isso, um dos pontos de atenção foi o desenvolvimento de uma interface com boa usabilidade e interação simplificada do usuário com o aplicativo, pois era sabido que o aplicativo poderia se tornar muito difícil de ser compreendido e usado dependendo da forma que fosse apresentada para o participante, principalmente por conta de que os aplicativos de saúde, em geral, tendem a despejar muitas informações textuais para o usuário ou oferecer um gama de funcionalidades que podem torná-lo confuso e complexo (NANOU, 2022). A preocupação de se criar algo esteticamente agradável e fácil de usar, é mencionado no meio da tecnologia como UI/UX. Interface de Usuário (UI) se refere à aparência e ao layout de um aplicativo ou sistema, incluindo elementos visuais, botões e a maneira como as informações são organizadas na tela, enquanto Experiência do Usuário (UX) se concentra em como os usuários interagem com o aplicativo e como se sentem ao usá-lo.

Para solução dos possíveis problemas com a UI, a paleta de cores pensada para o aplicativo, inicialmente, continha tons leves e suaves, que remetessem a saúde e trouxesse uma sensação de calma e paz (BIANCHI R. G. ; NERIS, 2015), visando modelar o ambiente terapêutico. Para isso, adotou-se a cor verde no esquema monocromático. Enquanto isso, em relação a UX, decidiu-se reduzir significativamente a quantidade de texto necessária para explicar como usar as funções do aplicativo, com botões que transmitissem uma mensagem clara e simples de sua funcionalidade, tornando intuitivo o que o usuário deveria/poderia fazer ao abrir o aplicativo. Nesse caso, os botões para envio de questionários, relatório de informações e ajuda, ficaram padronizados na parte inferior do aplicativo, enquanto a parte superior dava destaque para o principal objetivo do aplicativo Amive, que era o envio dos dados coletados pelo celular e relógio. O menu possuía algumas poucas opções, como informações sobre o projeto Amive, acesso à política de privacidade de dados, preferências do aplicativo e sair da conta.

Outra preocupação que acompanhou o desenvolvimento do aplicativo até os momentos finais,

Figura 7 – Protótipo do aplicativo Amive



Fonte: própria

foi sobre como seria feito a autenticação do usuário. Por questões de conforto dos participantes, não optou-se por fazer o uso de credenciais tradicionais como usuário, e-mail ou senha. Para isso, adotou-se a autenticação utilizando a conta do Facebook. Mais tarde, essa estratégia foi repensada e, para que o usuário pudesse se autenticar, bastava inserir um código de acesso único que cada participante teria. Em ambos os casos, após autenticado, o aplicativo se lembraria da conta do participante e o conectava automaticamente todas as outras vezes que abrisse o aplicativo novamente, ou até que ele saísse de sua conta (opção disponível no menu). Essa solução, dissociada de outro sistema computacional, também apresenta vantagem com relação à privacidade, dado que o código de acesso é único para o aplicativo Amive.

Mesmo que o aplicativo já tivesse uma interface agradável e funcional, faltava ainda uma identidade visual própria do projeto. Essa identidade foi criada por um profissional especializado e, com ela, além de uma nova logo, as paletas do Amive foram totalmente revisadas e alteradas. Dessa forma, a cor verde principal foi substituída pela nova paleta: azul bebê e salmão (apesar de as cores não terem uma definição exata de nome, esses são os nomes de cores que mais se aproximam). Embora a nova paleta de cores se comunicasse bem com o projeto Amive, ela não combinava com design do protótipo, que era monocromático. A solução foi repensar o design do aplicativo e utilizar o branco com sombras cinza-claro como a cor principal de toda interface, enquanto os tons de azul e salmão seriam voltados para botões, ícones, destaques e imagens. Essa nova identidade visual foi adicionada após os todos testes internos realizados com o protótipo do aplicativo. A Figura 8 são capturas de telas do aplicativo Amive, com as alterações mencionadas. Essa foi a interface final do aplicativo da forma que ele foi usado pelos

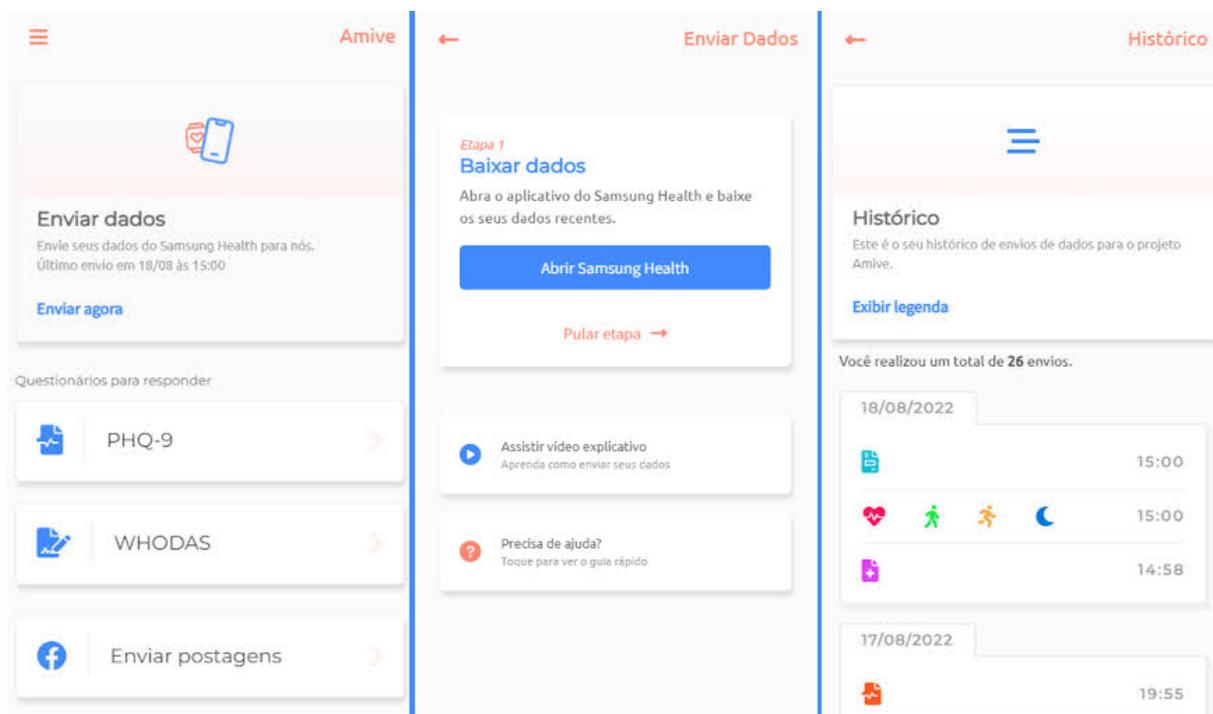
participantes durante a fase 1 do projeto Amive.

Como mencionado, a coleta de dados pelo aplicativo Amive demanda a interação com um outro aplicativo, o Samsung Health. Por conta disso, o participante precisa transicionar entre os dois aplicativos antes de enviar os seus dados para o servidor do projeto Amive. Para orientar o usuário sobre a interação entre dois aplicativos, foi necessário mudar o simples botão de “Enviar dados” para criar o que foi chamado de “Fluxo de envio”. A solução de design contempla então várias etapas sequenciais que guiam o usuário até o final da ação de enviar seus dados. Na solução final, ao clicar para enviar os dados, o participante é direcionado para uma tela com cinco etapas, onde cada etapa apresenta o que o usuário deve fazer.

A etapa 1 do fluxo de envio de dados, orienta o participante a abrir o aplicativo Samsung Health e, após baixar seus dados por lá, voltar para o Amive para continuar o fluxo. A etapa 2 era desbloqueada automaticamente ao voltar para o Amive, com uma mensagem de confirmação de que os dados haviam sido baixados. Na etapa 3, os participantes escolhem os dados que querem enviar. Na etapa de número 4, os usuários enviariam os seus dados de redes sociais, que mais tarde foi substituído por um botão específico na tela inicial do aplicativo, reduzindo o número de etapas seguidas pelo usuário para enviar seus dados. A 5.ª e última etapa é a conclusão do envio dos dados, que orienta o participante de que tudo ocorreu bem e apresenta um lembrete de repetir esse procedimento no dia seguinte. Esse fluxo foi desenvolvido com o intuito de facilitar o processo de aprendizado do participante sobre como enviar os seus dados. Outras questões também foram repensadas. Foram adicionados tutoriais escritos e em vídeo que estavam distribuídos pelo fluxo de envio de dados para que o participante pudesse se orientar em qualquer momento sobre como prosseguir. Os botões que abriam os questionários de saúde mental também foram substituídos e exibidos somente quando estavam disponíveis na interface inicial. Os botões de relatório (não capturados na figura) ainda continuaram os mesmos, sendo eles os de consultar o histórico de envios e visualizar gráficos de um determinado período.

As decisões de projeto relacionadas a coleta de dados dos sensores móveis demandaram estudos da literatura e das tecnologias e foram realizadas pelo autor desta dissertação. O aplicativo Amive foi implementado pelo autor desta dissertação utilizando programação nativa para aplicativos Android, empregando a linguagem Java, e incorporando uma página local HTML programada em JavaScript por meio do componente WebView. A estrutura e arquitetura computacional disponíveis no servidor do projeto Amive, incluindo questões de segurança dos dados e privacidade do usuário foram especificadas pela estudante Amanda de Oliveira (OLIVEIRA, 2022), sob orientação do professor Hélio Guardia. A arquitetura especificada foi implementada pelo bolsista do projeto Amive, Franco Garcia.

Figura 8 – Interface de usuário final do aplicativo Amive



Fonte: própria

---

# Capítulo 5

## Estudo com usuários

---

Foram coletados dados de três diferentes fontes: dados de fenotipagem digital (do próprio celular e do relógio inteligente), redes sociais (postagens do Facebook e *tweets* do Twitter) e questionários, sendo eles as escalas psicométricas (PHQ-9, WHODAS-II e sentimento subjetivo) e um questionário onde o participante poderia avaliar as informações enviadas por ele. Com exceção do Twitter (diretamente coletado de forma automática pelo sistema no servidor), os demais dados foram coletados usando o aplicativo Amive como intermediário. Para o capítulo de análise do uso dos resultados deste trabalho, serão usados os dados advindos dos sensores móveis, as avaliações dos usuários sobre esses dados e as escalas psicométricas que foram usadas para a rotulação dos dados.

### 5.1 Método de coleta

Os dados de fenotipagem digital, provenientes do celular e do relógio inteligente, foram coletados utilizando o aplicativo do Samsung Health. O aplicativo registra e gerencia as atividades físicas, comportamentos do indivíduo, padrões de sono e outras características gravadas, e as detecta automaticamente enquanto instalado no dispositivo. Essa solução foi adotada pelo projeto por conta de outras dificuldades encontradas durante o estudo de análise de estratégias para coleta de dados, descritos no Capítulo 7.1. Em modo geral, o Samsung Health, além de obter essas informações de forma automática, isto é, sem que o usuário precise interagir com o aplicativo ou lembrar de habilitá-las, também oferece a opção do usuário baixar todas as suas informações armazenadas pelo aplicativo. Foi através dessa opção de baixar os dados pessoais que o Amive conseguiu obter os dados de cada participante. Sendo assim, os participantes faziam suas atividades diárias normalmente, enquanto o Samsung Health as registrava e, ao final do dia, os participantes baixavam seus dados no aplicativo da Samsung Health e os enviavam para o servidor do projeto Amive usando o aplicativo Amive.

O Samsung Health obtém e registra diversos tipos de dados<sup>1</sup>, e o aplicativo Amive coletou e armazenou apenas os dados relevantes para o projeto, sendo eles: frequência cardíaca, monitoramento de sono, contagem de passos e atividades físicas. Quaisquer outros dados não relacionados a esses, eram ignorados pelo aplicativo antes do envio para o servidor. Todos os dados coletados selecionados foram enviados diretamente ao servidor, sem intervenções do aplicativo, exceto pela contagem de passos, que eram unificados todos os passos ao longo do dia para um único registro. Desses dados, a frequência cardíaca foi o único registro que só era possível ser coletado por participantes que estavam usando o relógio inteligente. As demais informações poderiam ser capturadas apenas pelo celular do participante, porém com menor precisão, principalmente em relação aos registros de sono, que precisavam ser confirmados pelo próprio usuário dentro do aplicativo da Samsung Health para serem efetivamente registrados.

Após o participante baixar suas informações do Samsung Health e enviá-los para o aplicativo Amive, chega a hora de escolher quais dos seus dados serão enviados ao servidor para pesquisa. O projeto Amive adotou uma abordagem ética de proteção de dados e anonimato dos participantes. O estudo com usuários teve aprovação do Comitê de Ética em Pesquisa (CEP). No projeto avaliado, há a especificação de todos os dados que seriam coletados e a forma que isso seria feita. Respeitando a privacidade do participante, era uma opção dele enviar ou não os dados daquele dia.

As informações das escalas psicométricas também faziam parte do envio do aplicativo para o servidor. Os questionários do PHQ-9 e do WHODAS-II ficavam disponíveis na interface a cada 15 e 30 dias, respectivamente, e após respondidos não apareciam mais na tela até a próxima rodada. Ainda, havia o questionário de avaliação dos dados coletados. Após cada envio dos dados de fenotipagem digital do participante, ou então, caso o participante optasse por visualizar os gráficos de uma determinada data, um questionário com três perguntas sobre acurácia, inteligibilidade e utilidade eram exibidos. Dessa forma, após o participante visualizar suas informações coletadas pelos sensores móveis, ele poderia também avaliar se elas condiziam com a realidade e com o que ele viveu. Todas as respostas dos questionários foram enviadas para o servidor, juntamente da data de quando o participante o respondeu.

Em síntese, a rotina do participante da pesquisa consistia em, basicamente, viver o seu dia a dia normalmente, sendo um participante com relógio ou não, e no final da tarde e/ou da noite, abrir o aplicativo Amive para enviar seus dados. Nesse momento, o questionário de sentimento subjetivo seria respondido, juntamente do envio de seus dados fisiológicos. Esporadicamente, os questionários PHQ-9 e WHODAS-II eram exibidos para que o participante os respondesse, se disponíveis, e o mesmo poderia eventualmente consultar seu histórico e envio de informações através das consultas disponíveis no aplicativo. Essa foi a rotina proposta para os participantes da pesquisa ao longo do dia, abrindo e usando o aplicativo entre uma e duas vezes ao dia, com o tempo médio de um a três minutos de interação.

---

<sup>1</sup> Disponível em: <https://samsunghealth.com/privacy>

## 5.2 Síntese dos dados coletados

Com o término da fase de coleta de dados, inicia-se o trabalho de desenvolver a análise exploratória para a identificação de estudantes universitários brasileiros com PPD. Sendo assim, é possível já criar uma síntese desses dados sem expor a confidencialidade deles.

Ao todo foram 89 estudantes universitários da UFSCar que se voluntariaram a participar da coleta de dados, sendo que 61 deles fizeram *login* ao menos uma vez durante o período, descartando então 28 participantes que se inscreveram para a pesquisa, mas não usaram o aplicativo. Desses 61 usuários, 54 deles fizeram ao menos um envio do questionário do PHQ-9 e 53 deles também fizeram ao menos um envio do WHODAS-II. No total, foram 35 usuários que enviaram quaisquer dados dos sensores móveis. Em uma visão geral, pode-se estimar que 57% das pessoas que usaram o aplicativo fizeram o envio de dados de fenotipagem digitais dos sensores. A Tabela 4 demonstra o número de dados coletados por sensores móveis enviados pelos participantes.

Tabela 4 – Total de envio de dados dos sensores

Contador de passos	1.055
Exercícios	882
Frequência cardíaca	33.651
Dados de sono	736

Cada informação enviada pelo participante utilizando o aplicativo foi contabilizada como um envio (questionários, postagens de redes sociais), exceto pelos dados dos sensores dos celulares e relógios, que são contabilizados como um conjunto de um único envio. Por exemplo, caso o participante envie dados de sono e de passos, conjuntamente, tudo é contabilizado como um único envio, embora sejam armazenados separadamente no banco. Sabendo disso, ao todo, foram feitos 3.988 envios, e em média, cada usuário fez cerca de 65 envios para o aplicativo (o número máximo de envios feito por um usuário foi de 339). Isso significa que, em média, foram feitos 2 envios por dia para cada participante ativo da pesquisa (para 30 dias de pesquisa). Esses 2 envios são, em maioria, do sentimento subjetivo (que aparecem sempre ao abrir o aplicativo) e dos dados de sensores. Também houve usuários que não enviaram nenhum tipo de dado. A média de envios diários demonstra que a maioria dos participantes não tivera a disponibilidade de, conforme o orientado, realizar dois envios de dados dos sensores por dia, e sim, apenas um. Houve dificuldades (abordadas no Capítulo 7.1) que impactaram a relação entre os participantes ativos e o número de envios únicos dos sensores móveis.

## 5.3 Disponibilização de dados

Com a primeira fase da coleta de dados finalizada, os dados armazenados em nosso banco foram trabalhados de forma que pudessem ser visualizados de forma fácil para os demais pesquisadores da equipe. O Amive é um projeto multidisciplinar que, além da computação, conta

também com pesquisadores da área da saúde que contribui com toda parte técnica e ética relacionada aos estudos com saúde mental sobre depressão. A fim de disponibilizar os dados para análise e pesquisa, todo o montante de dados armazenados no banco foi organizado em tabelas e disponibilizados online em um site interno do servidor Amive, nomeado de Amive Web, de forma que qualquer pesquisador autorizado pudesse acessar, estudar e baixar os dados organizados em tabelas, disponível tanto em português quanto em inglês.

O Amive Web disponibiliza os dados de todos os participantes organizados em tabelas com informações gerais (Figura 9), como a visualização geral dos dados enviados pelo relógio e celular, dados demográficos de cada participante — os dados demográficos foram enviados pelo participante previamente antes do início da pesquisa — e um resumo das datas e quantidade de envio dos questionários de saúde mental.

Figura 9 – Visão geral dos dados dos participantes no Amive Web

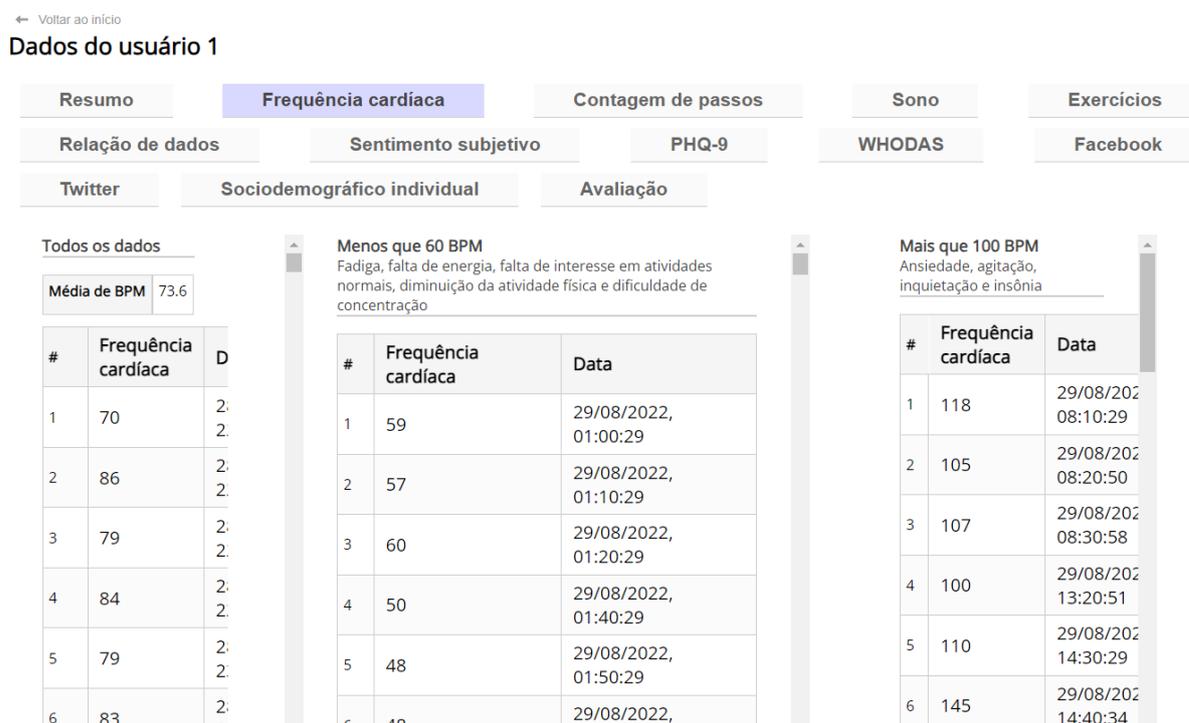
#	Dados enviados	Relógio?	Frequência cardíaca	Sono	Contagem de passos	Exercícios	Sentimento subjetivo	PHQ-9	WHODAS	Avaliações
1	159	yes	2040	42	41	74	29	3	2	62
2	20	yes	283	3	4	0	8	3	2	9
3	1	no	0	0	0	0	0	1	1	0
4	45	yes	2894	43	40	1	27	3	3	0
5	66	yes	2392	39	27	39	42	3	2	22
6	2	no	0	0	0	0	1	1	1	0
7	309	yes	439	27	42	62	24	3	2	66
8	27	yes	2172	24	22	16	18	3	2	0
9	77	yes	608	37	46	74	20	3	2	24
10	26	yes	656	17	15	1	10	2	2	3

Fonte: própria

Com o propósito de uma investigação mais detalhada, o site também oferece a visualização dos dados de um único participante em específico. Dessa forma é possível ver em detalhes tudo que um usuário enviou durante a pesquisa. Todas as informações são munidas com datas de envio (do aplicativo) e coleta (do celular), vide Figura 10.

Além da visualização dos dados isolados do usuário (sensores, questionários e redes sociais), também há a disposição uma tabela que demonstra a relação direta entre os dados (Figura 11).

Figura 10 – Dados detalhados de um único usuário no Amive Web



Fonte: própria

É possível observar que os dados de uma frequência cardíaca alta podem estar relacionada a momentos em que houve atividade física pelo usuário, da mesma forma baixos dados de batimento cardíaco baixo pode estar relacionado a momentos em que o usuário estava dormindo. Essas relações entre os dados foram identificadas antes do início do estudo por meio da análise exploratória, e posteriormente, foram confirmadas com os resultados da correlação de Spearman (abordado no Capítulo 6.2). A partir das informações coletadas e organizadas no Amive Web, que tornam a visualização mais simples do que dentro de tabelas de um banco de dados, foi possível traçar conexões entre as variáveis e identificar tendências significativas de padrões comportamentais dos participantes. No próximo capítulo, são apresentados os resultados que emergiram dessa análise, revelando como esses dados forneceram uma compreensão maior das questões relacionadas aos sintomas depressivos.

Apesar da disponibilização dos dados, o projeto Amive seguiu rigorosamente as normas do CEP. Uma das diretrizes mais impactantes é a anonimização dos dados. Tanto por questões de segurança quanto por questões éticas, os pesquisadores não têm conhecimento de informações que identifiquem os participantes. Os dados pessoais foram ou ocultados ou substituídos por outros que impossibilitassem a identificação.

Durante o recrutamento de participantes, os interessados responderam a um questionário sociodemográfico e forneceram informações pessoais de contato. Após o término da fase de recrutamento, um programa foi executado sobre os dados para gerar um arquivo final que removia as informações pessoais dos participantes e os tornavam usuários anônimos, dando a

Figura 11 – Relação de dados de um usuário no Amive Web

Resumo	Frequência cardíaca	Contagem de passos	Sono
Exercícios	Relação de dados	Sentimento subjetivo	PHQ-9
WHODAS	Facebook	Twitter	Sociodemográfico individual
Avaliação			

**Frequência cardíaca alta X dados de exercício**  
Momentos em que a frequência cardíaca alta pode estar relacionada a momentos em que houve atividade física

#	Frequência cardíaca	Registro da frequência cardíaca	Início do exercício
1	118	29/08/2022, 08:10:29	24/08/2022, 12:06:15
2	118	29/08/2022, 08:10:29	24/08/2022, 13:54:50
3	118	29/08/2022, 08:10:29	24/08/2022, 14:49:28
4	118	29/08/2022, 08:10:29	25/08/2022, 12:25:15
5	105	29/08/2022, 08:20:50	24/08/2022, 12:06:15

**Frequência cardíaca baixa X dados de sono**  
Momentos em que o batimento cardíaco baixo pode estar relacionado a momentos em que o usuário estava dormindo

#	Frequência cardíaca	Registro da frequência cardíaca	Início do sono
1	48	29/08/2022, 01:50:29	25/08/2022, 01:21:00
2	48	29/08/2022, 01:50:29	26/08/2022, 01:02:00
3	48	29/08/2022, 01:50:29	26/08/2022, 06:48:00
4	48	29/08/2022, 01:50:29	27/08/2022, 01:00:00
5	48	29/08/2022, 01:50:29	29/08/2022, 01:40:00

Fonte: própria

eles um Identificador Único (ID), um e-mail aleatório (exemplo, 05c909@amive.ufscar.br) e uma senha randômica. A partir desse ponto, não era mais possível identificar qualquer participante a não ser pelo seu ID ou e-mail gerado aleatoriamente, e foram com esses dados que toda a pesquisa foi conduzida.

Ao anonimizar as informações pessoais identificáveis, como nomes, endereços e contatos, e substituí-las por um ID, garante-se que os dados sejam utilizados apenas para fins de pesquisa, sem comprometer a privacidade dos participantes. A anonimização é uma abordagem comum usada nos estudos que criaram aplicativos dedicados para coleta de dados de usuários. A estratégia de anonimização escolhida para este trabalho, também levou em consideração os estudos dos trabalhos relacionados (FARHAN et al., 2016; LU et al., 2018; AHMED et al., 2020), que também priorizaram a anonimização de informações sensíveis dos participantes, focando apenas nos dados relevantes para a pesquisa.

Deve-se destacar que os participantes que tiveram dificuldades para acessar o aplicativo e configurar o relógio puderam contar com o suporte técnico do Amive. Durante esses momentos, houve troca de e-mails dos participantes com os pesquisadores responsáveis pelo desenvolvimento do aplicativo e alguns encontros pessoais dos pesquisadores com os participantes no campus da universidade para auxiliar em outros problemas relacionados a pesquisa. Apesar do contato com alguns poucos participantes, a anonimização das informações pessoais dentro da infraestrutura

do Amive tornava impossível identificá-los dentro dos dados.

## 5.4 Human in the loop

Durante o desenvolvimento e estudo da pesquisa, foi tomado o conhecimento de uma abordagem de IHC chamado de *Human in the Loop* (HITL) (humano no loop), abordado por Munro (MUNRO, 2021). O HITL envolve os próprios usuários no treinamento de algoritmos de dados. Os dados do usuário são usados para melhorar sistemas de forma que os próprios usuários recebam essas melhorias e avaliam se são úteis ou corretas. Isso ajuda a personalizar a experiência do usuário e garantir que os sistemas funcionem melhor para cada pessoa.

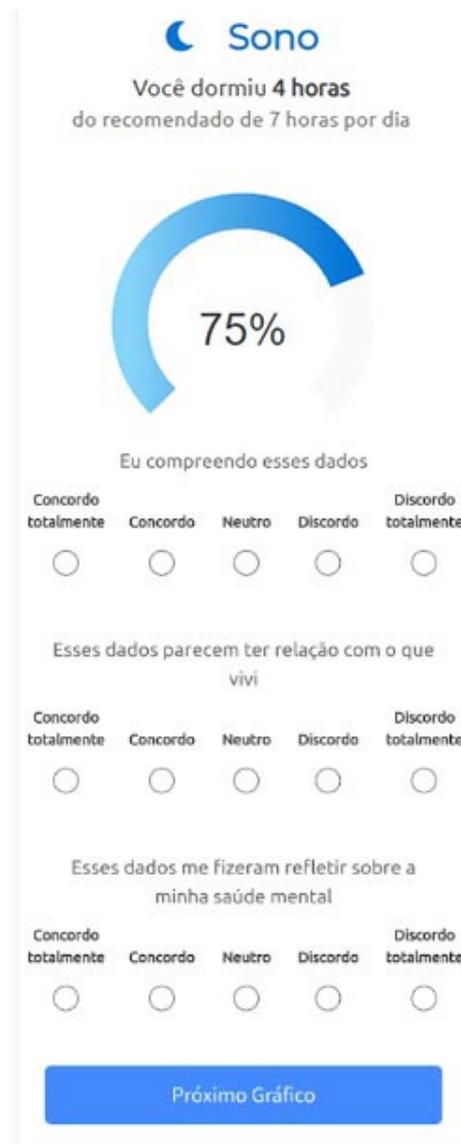
Esse conceito foi explorado durante a pesquisa, pois a perspectiva da identificação de sintomas depressivos exige abordagens que vão além da coleta de dados de fenotipagem digital ou identificações padrões. A abordagem HITL propõe uma evolução dos modelos de computação atuais para considerar outros dados relacionados a aspectos sociais, com uma abordagem mais centrada no ser humano, uma vez que diferentes vulnerabilidades contribuem para o sofrimento da depressão. Nessa perspectiva, os participantes da pesquisa podem desempenhar um papel mais ativo nos seus próprios algoritmos de classificação, interagindo com a solução para informar sobre seus sentimentos subjetivos e participando ativamente no processo de avaliação dos dados.

A ideia principal seria fazer com que o HITL acontecesse em tempo real durante a fase de processamento de dados do participante, de forma que o *feedback* do participante pudessem influenciar a forma que o aplicativo Amive se comportaria com o mesmo. Embora tenha havido um interesse significativo em integrar essa abordagem durante a fase de estudos com os usuários, sua implementação ocorreu de forma efetiva no capítulo de avaliação dos resultados desta dissertação. Durante essa etapa, os dados foram analisados pelo algoritmo, levando em consideração as contribuições dos participantes para determinar a precisão das informações e gerar os resultados finais, revelando os padrões de comportamento entre usuários com e sem PPD.

A abordagem do HITL foi implementada através dos questionários avaliativos que os usuários respondiam de forma opcional. O questionário era apresentado para o usuário em dois momentos, sendo um deles, sempre que o participante enviava algum dado do sensor móvel, dessa forma, o participante poderia visualizar de forma gráfica as informações que foram enviadas pelo sensor e, em seguida, avaliar se os dados estavam condizentes com o que ele viveu naquele dia. Segundo, quando o participante consultava seu histórico de envios. O questionário avaliativo aparecia acompanhado dos gráficos e suas informações do dia consultado.

A Figura 12 ilustra o processo onde o usuário enviou dados de sono. As informações desse dado são exibidas em formato de gráfico, que varia de acordo com o tipo de dado enviado, sendo dados de sono e passos um meio círculo (como na figura), exercícios em formato textual informando o tempo decorrido em atividade, e a frequência cardíaca em gráficos de linhas. Logo após o gráfico com as informações dos dados, o formulário de avaliação sempre é apresentado. Esse processo acontecia com cada um dos dados enviados (sono, frequência cardíaca, passos e exercícios), e o usuário poderia responder a um, todos, ou seguir sem responder nenhuma opção

Figura 12 – Relação de dados de um usuário no Amive Web



Fonte: própria

de avaliação. Os gráficos dos dados enviados e os formulários de avaliação eram apresentados um por vez, dessa forma o usuário poderia ir seguindo um passo a passo de avaliação e validação dos seus dados de forma sequenciada.

O questionário avaliativo era composto de três perguntas, com opções de resposta como concordo, concordo totalmente, discordo, discordo totalmente e neutro. As perguntas foram computadas com a acurácia ("Esses dados parecem ter relação com o que eu vivi"), inteligibilidade ("Eu compreendo esses dados") e utilidade ("Esses dados me fizeram refletir sobre a minha saúde mental"). Todos os usuários que enviaram dados dos sensores, responderam ao menos um questionário avaliativo, mas não a todos os questionários que foram apresentados durante os envios dos sensores.

Durante a observação dos dados dos participantes, foram encontrados vários *outliers*<sup>2</sup> proveniente do envio dos sensores, e com a intenção de reduzir o número de informações incorretas dos participantes, foram consideradas as resposta enviadas pelo formulário de acurácia de todos usuários. Sendo assim, qualquer resposta marcada como discordo ou discordo totalmente, foi considerada uma informação incorreta na base de dados e, portanto, foi removida para a análise final do capítulo de avaliação dos resultados. Dessa forma, a técnica HITL foi aplicada para considerar a avaliação do próprio participante em relação a remoção de outliers e, assim, melhorando a qualidade do conjunto de dados final apresentado no Capítulo 6.

---

<sup>2</sup> Valores atípicos que se desviam significativamente do padrão geral de um conjunto de dados.



---

# Capítulo 6

## Análise estatística

---

Com os dados dos participantes disponíveis no banco de dados ao final da primeira fase do projeto Amive, é chegada a hora de investigar tais dados a fim de encontrar informações relacionados a padrões de comportamento entre indivíduos que possuem ou não o possível perfil depressivo, aplicando a estratégia da análise exploratória. O objetivo dessa análise é, por meio de um estudo apenas descritivo, conseguir responder se com a infraestrutura computacional construída durante a pesquisa é possível identificar ou não grupos com PPD.

### 6.1 Conjunto de dados

A primeira etapa da análise de dados foi separar quais seriam os usuários que iriam fazer parte do *dataset*<sup>1</sup> dessa análise. Apesar de se ter dados de 89 participantes, a maioria desses dados não puderam ser incluídos na análise, pois existiam inconsistências nos dados que os tornam inapropriados para compor o conjunto e também dados ausentes. Por exemplo, assim como existem participantes que não chegaram a usar efetivamente o aplicativo, enviando poucos ou nenhum dado, também tiveram participantes que não tiveram um fluxo de envio suficiente de dados para serem interpretados e compreendidos pelo algoritmo escolhido para gerar resultados mais precisos. Ainda, após a análise dos gráficos disponibilizadas no Amive Web, foram percebidos outliers em dados de alguns participantes.

Com o intuito de ter um dataset sólido com dados preenchidos e informações relevantes, foi realizado uma limpeza nos dados do banco para gerar o dataset final. Essa limpeza passou primeiramente pelo processo de remoção de outliers, seguido de uma imputação dos dados.

A remoção dos outliers foi feita usando a abordagem do HITL (abordada no Capítulo 5.4). Foram removidos todos dados dos participantes onde a acurácia estivesse avaliada como "Dis-

---

<sup>1</sup> Conjunto de dados organizados que podem ser usados para análises, pesquisas ou treinamento de modelos de aprendizado de máquina.

cordo"ou "Discordo totalmente". Esses dados foram considerados outliers pelos próprios participantes, uma vez que eles mesmos identificaram que os dados não estavam condizentes com o que eles viveram naquele dia.

A imputação de dados é um processo de preencher ou estimar valores ausentes ou faltantes em um conjunto de dados. Isso foi feito para permitir uma análise mais completa e precisa, garantindo que os dados possam ser usados nas técnicas estatísticas da pesquisa mesmo quando algumas informações estão ausentes. Esse processo se fez necessário uma vez que a maior parte dos participantes possuíam um ou mais dados faltantes em um dia, como por exemplo, porque estava sem o relógio (como em momentos que precisavam carregar). O pacote `mice.impute.rf`<sup>2</sup> da linguagem R foi usado para imputação de dados.

Após a limpeza dos dados, permaneceram 33 participantes com dados completos para se trabalhar com os algoritmos de reconhecimento de padrões, sendo 11 com PPD e 22 não PPD. Esses participantes e seus dados foram escolhidos para compor o dataset final da análise exploratória.

Os métodos aplicados na análise exploratória foram orientados pelo especialista em análise de dados, Professor Anderson Ara, da Universidade Federal do Paraná (UFPR), que sugeriu que os testes executados e os resultados gerados fossem feitos sempre em três grupos, sendo eles: grupo geral de indivíduos, grupo classificado como PPD e grupo classificado como não PPD. Desta forma, é possível ter um panorama de todos os dados da pesquisa e a comparação entre grupos. A classificação dos grupos tomou como base os resultados do questionário PHQ-9 respondido pelos participantes da pesquisa. A classificação corresponde aos 14 dias anteriores do último envio do questionário, sendo rotulado como PPD os indivíduos que obtiveram a pontuação maior ou igual a nove pontos no questionário.

## 6.2 Métodos

Para que se possa elucidar a forma com que os grupos se relacionam, podem ser aplicados diversos métodos da análise exploratória. A ideia da análise exploratória em forma geral, vai tender ao termo de reconhecimento de padrões. Muitas das técnicas envolvem a entrada de dados, fazendo uma descritiva inicial e depois usar algumas funções para ajustar o modelo de acordo com as variáveis. Para esta análise, isso foi feito por indivíduo, trabalhando com a média dos dados, sendo possível gerar gráficos de histograma, bloxplot, densidade e outros. Esse processo foi feito iniciando pelo total de indivíduo do grupo geral e depois para os outros dois tipos de grupos, com e sem PPD.

Entre o processo de análise univariável (com cada variável separadamente) e antes de avançar para a análise multivariada (com todas as variáveis juntas), foi incluído um estudo de correlação entre as variáveis, construído por meio de uma matriz de correlação. Embora a correlação de Pearson tenha sido uma escolha comum em trabalhos anteriores, optamos pelo uso da correlação de Spearman para esta pesquisa, uma vez que os dados apresentam características que não seguem uma distribuição linear. A escolha pelo coeficiente de Spearman se justifica devido à sua robustez em lidar com dados não paramétricos e a ausência da necessidade de pressupostos

<sup>2</sup> Disponível em: <https://www.rdocumentation.org/packages/mice/versions/3.16.0/topics/mice.impute.rf>

quanto à normalidade da distribuição dos dados, tornando-o uma escolha mais apropriada para a análise em questão.

Os resultados deste estudo também abrangem a fase de análise multivariada, onde todas as variáveis dos indivíduos são combinadas para obter uma visão ampla do comportamento dos grupos. Nessa etapa, optou-se por aplicar a técnica do t-SNE, que tem a capacidade de simplificar o panorama complexo dos dados, condensando as informações em um gráfico bidimensional, facilitando significativamente a visualização e a análise dos padrões que surgem quando se unem todos os dados.

A análise exploratória conduzida neste estudo segue uma abordagem sistematizada em três etapas. Inicialmente, na análise exploratória univariada, foram gerados gráficos de histograma, boxplot e densidade para cada variável, acompanhados de uma análise descritiva dos dados. Em seguida, na etapa de correlação entre as variáveis, uma matriz de correlação foi construída, sendo utilizado o coeficiente de Spearman. Os dados foram posteriormente filtrados de acordo com o perfil depressivo e não depressivo, permitindo a comparação das matrizes de correlação entre esses grupos, dessa forma foi possível observar as diferenças nas correlações entre os grupos depressivos e não depressivos. Por fim, na análise exploratória multivariada, a técnica t-SNE foi aplicada para reduzir a dimensionalidade dos dados, compactando as variáveis em duas dimensões. Isso resultou na criação de um gráfico bidimensional que facilita a visualização dos padrões dos indivíduos. Essas estratégias de análise exploratória permitem uma compreensão mais aprofundada das possíveis diferenças entre os indivíduos com depressão e sem depressão com base nas variáveis examinadas.

## 6.3 Análise univariada

O processamento dos dados foi feito usando a linguagem de programação Python, com o auxílio das bibliotecas `matplotlib`<sup>3</sup> e `seaborn`<sup>4</sup>. Essas bibliotecas possuem uma variedade de recursos projetados para facilitar a análise de dados, incluindo algoritmos específicos para geração dos gráficos requeridos para essa pesquisa, tais como histograma, boxplot e gráficos de densidade.

As variáveis analisadas correspondem aos campos provenientes dos sensores móveis, que são: frequência cardíaca, contagem de passos, tempo de sono e frequência de atividade física. O propósito desta análise é obter as características das variáveis e compreender sua distribuição, visando em seguida comparar a centralidade ou dispersão entre os grupos com PPD e não PPD.

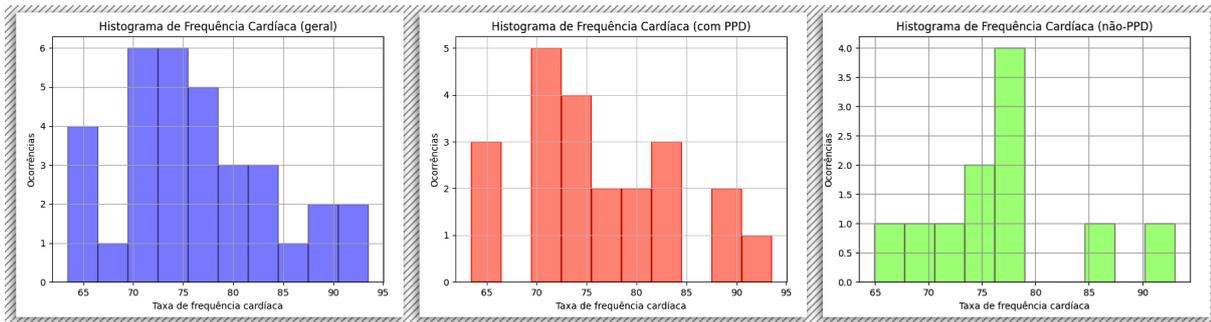
### 6.3.1 Frequência cardíaca

Começando pela frequência cardíaca, é notável que a maioria dos participantes apresentam uma taxa média de batimentos cardíacos situada entre 70 e 80 por minuto, o que é considerado o padrão esperado para adultos na faixa etária de 18 a 25 anos. No entanto, é importante destacar que, como os cálculos do histograma se baseiam na média de valores registrados ao longo

<sup>3</sup> Disponível em: <https://matplotlib.org>

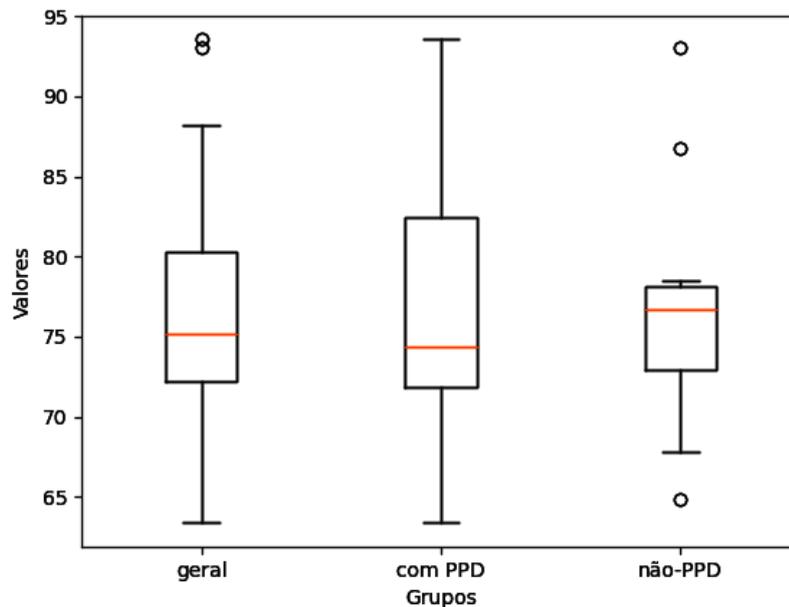
<sup>4</sup> Disponível em: <https://seaborn.pydata.org>

Figura 13 – Histograma de frequência cardíaca



Fonte: própria

Figura 14 – Boxplot de frequência cardíaca

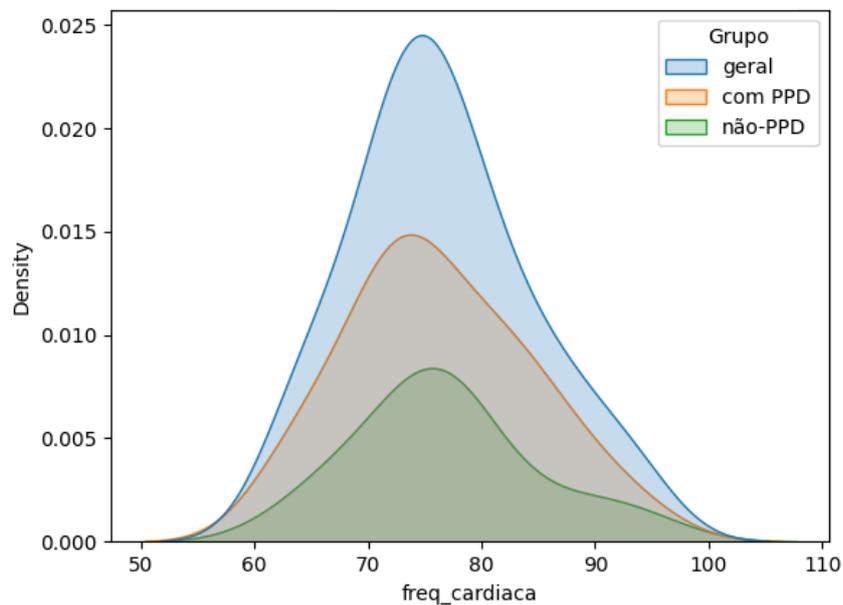


Fonte: própria

do dia, os resultados indicam que alguns participantes possuem taxas de frequência cardíaca significativamente acima ou abaixo dessa média com maior frequência do que o normal.

Valores abaixo ou acima dessa média podem sugerir a presença de problemas de saúde. No entanto, pode-se estabelecer correlações entre as variáveis e compreender melhor por que certos participantes apresentam taxas mais altas ou mais baixas do que a média. Por exemplo, taxas de frequência cardíaca em torno de 65 a 70 são comuns durante o sono, o que pode indicar que participantes que dormem mais ao longo do dia ou em longos períodos, conseqüentemente, possuem uma média de frequência cardíaca menor. Da mesma forma, valores acima de 90 podem ser facilmente alcançados pelo nosso metabolismo durante atividades físicas intensas, indicando uma alta taxa de frequência cardíaca diretamente associada a participantes que se envolvem em atividades físicas mais frequentemente ao longo do dia.

Figura 15 – Gráfico de densidade de frequência cardíaca

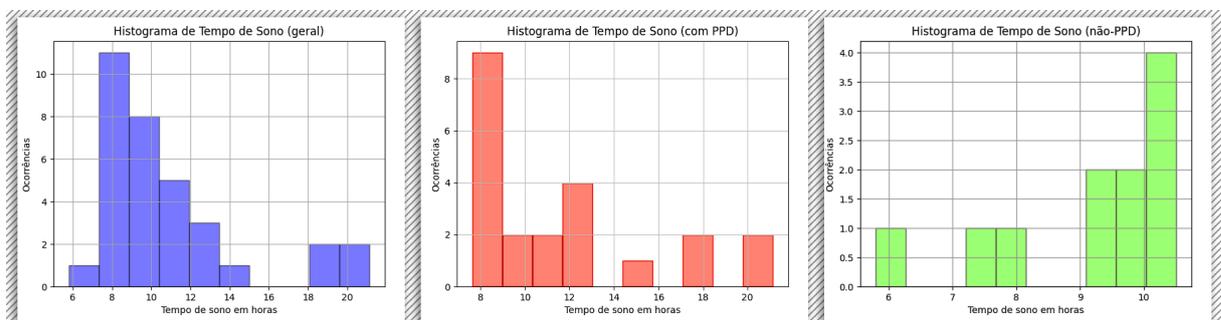


Fonte: própria

Na análise dos grupos, inicialmente observa-se uma contradição entre os valores e a teoria mencionada anteriormente. O grupo com PPD apresenta taxas de frequência cardíaca mais elevadas, em contraste com o grupo sem PPD, que registra valores mais baixos. Apesar disso, ainda é necessário considerar as outras variáveis para entender melhor como os dados se comportam e compreender suas relações com as demais variáveis.

### 6.3.2 Tempo de sono

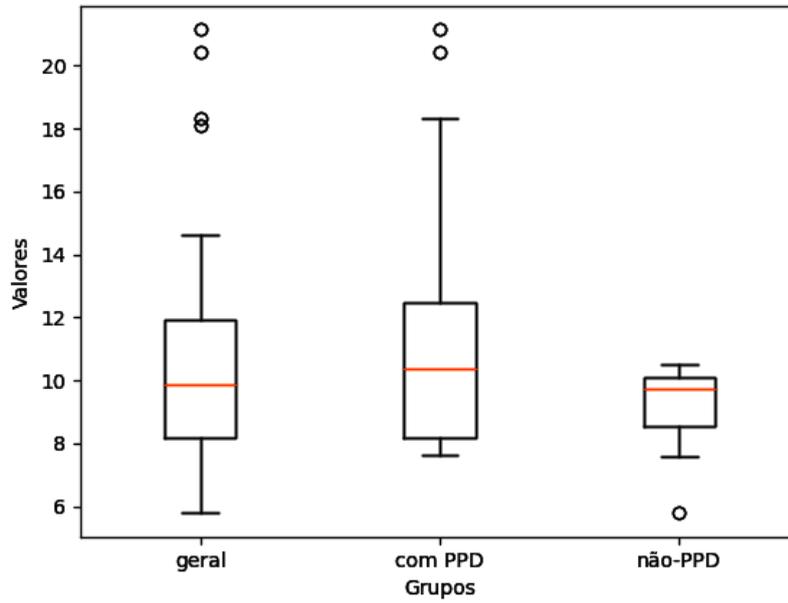
Figura 16 – Histograma de tempo de sono



Fonte: própria

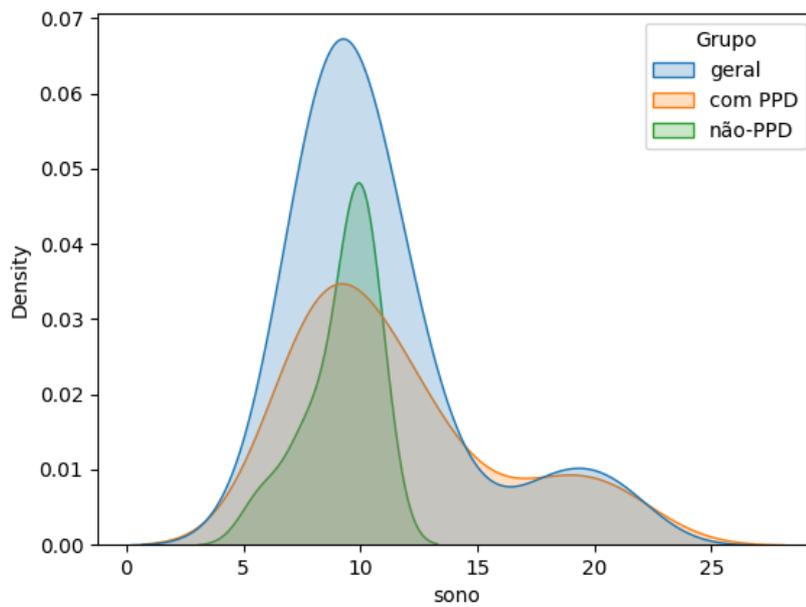
O grupo com PPD tem uma média de sono significativamente mais alta (11.7 horas) em comparação com o grupo não PPD, que tem uma média de sono mais baixa (9.2 horas). Isso sugere que, em média, as pessoas com depressão na amostra estão dormindo mais do que aquelas

Figura 17 – Boxplot de tempo de sono



Fonte: própria

Figura 18 – Gráfico de densidade de tempo de sono



Fonte: própria

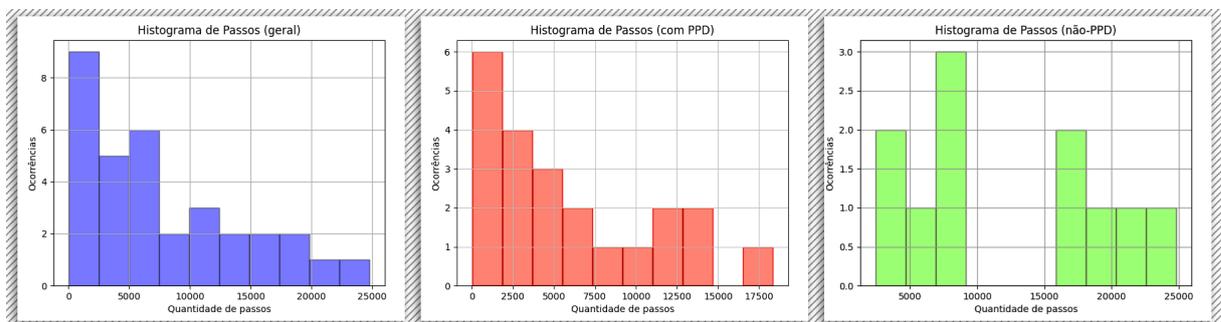
sem depressão. Além disso, o grupo com PPD possui uma variação maior no tempo de sono, o que pode sugerir que o grupo com depressão apresenta uma maior dispersão nos hábitos de sono, com algumas pessoas dormindo muito mais do que a média e outras dormindo menos.

Um dos sintomas caracterizados pela depressão é a insônia ou hipersonia. Embora a insônia não seja tão evidente nesses grupos de dados, a hipersonia por sua vez pode ser notada, com alguns indivíduos dormindo muito mais do que o padrão médio. Isso pode ser interpretado como um alerta para o sinal de fadiga excessiva e letargia, características frequentes da depressão.

Os resultados dos gráficos de sono é um exemplo de como os dados podem ser observados para identificação de correlações entre as variáveis. Isso pode explicar, por exemplo, o motivo pela qual os usuários com PPD tem uma taxa de frequência cardíaca mais alta. Isso pode ser devido ao fato de que os usuários sem PPD dormem mais, resultando em uma média de frequência cardíaca menor durante o período de repouso do sono. Por outro lado, o grupo com PPD mantém taxas mais elevadas ao longo do dia, possivelmente devido a períodos de sono de menor duração.

### 6.3.3 Contagem de passos

Figura 19 – Histograma de contagem de passos

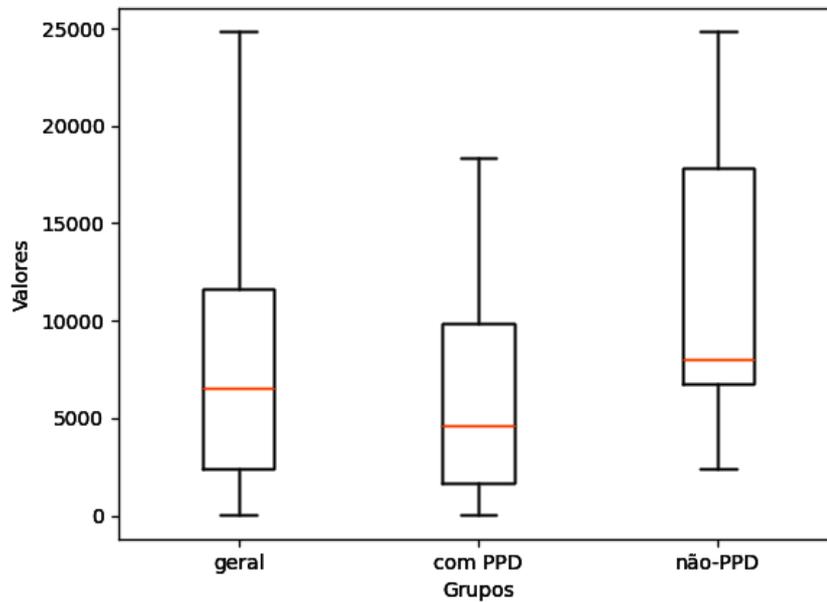


Fonte: própria

Os resultados dos gráficos de contagem de passos ao longo do dia também mostram uma clara diferença entre os grupos, onde indivíduos do grupo não PPD tendem a dar mais passos em média por dia em comparação com os usuários com depressão. Além disso, no grupo com PPD há uma variabilidade maior nos dados, conforme indicado pelo desvio padrão mais alto.

Esses valores apoiam o fato de pessoas com sintomas depressivos tenderem a dar menos passos, devido a fatores como baixa motivação, falta de energia e isolamento social. Além disso, a depressão pode levar a distúrbios do sono que afetam a disposição para se exercitar, criando uma correlação com a variável de sono que, por sua vez, também afeta o quanto a taxa de frequência cardíaca de um grupo pode ser mais alta ou mais baixa. Portanto, essa relação é bidirecional, com o número de passos afetando a depressão e o possível perfil depressivo afetando o número de passos.

Figura 20 – Boxplot de contagem de passos



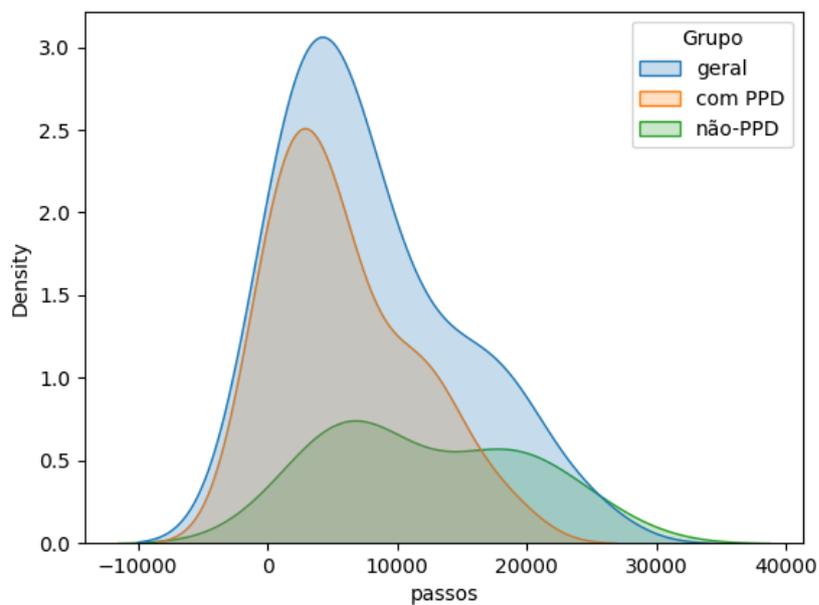
Fonte: própria

### 6.3.4 Exercícios físicos

Os resultados sobre exercício dizem respeito a quantos exercícios físicos o indivíduo realiza por dia. Há uma diferença bem clara de que o grupo não PPD é detentor da maior parte dos que realizam atividades físicas, enquanto o grupo com PPD tende a realizar pouco ou nenhum exercício. Os sintomas depressivos frequentemente afetam o ânimo, a motivação e o interesse por atividades em geral, levando as pessoas a se engajarem menos em exercícios.

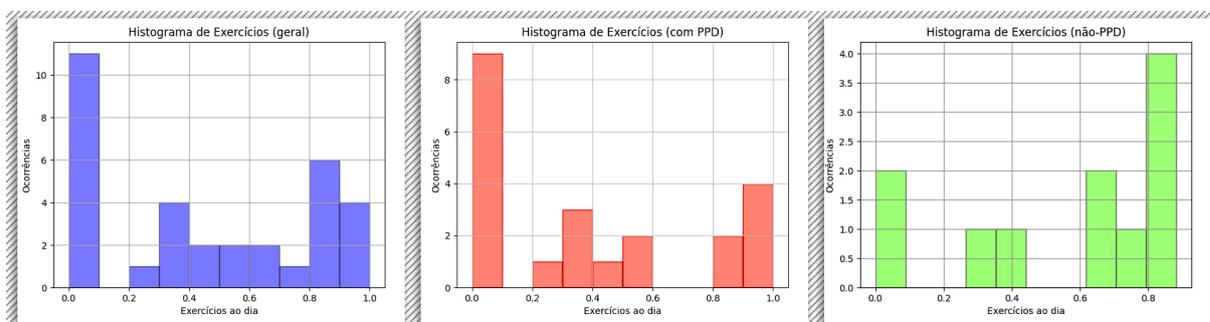
Os dados de exercícios também podem criar relações com as demais variáveis. Quanto mais exercícios realizados, mais passos um indivíduo dá (ou ao contrário), mesmo que o sensor não compute passos como uma parte da atividade física, o indivíduo que tem disposição e energia para fazê-lo também tende a percorrer o caminho até seu ambiente de recreação todos os dias.

Figura 21 – Gráfico de densidade de contagem de passos



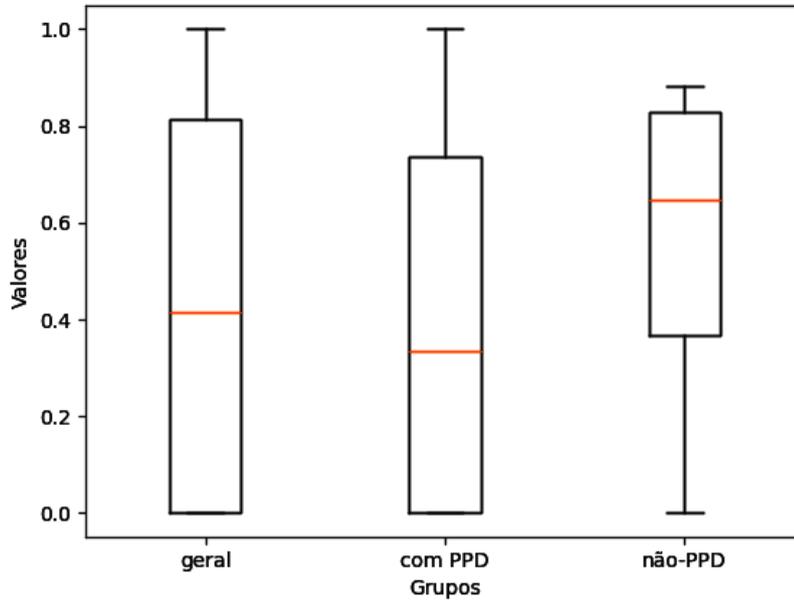
Fonte: própria

Figura 22 – Histograma de exercícios físicos



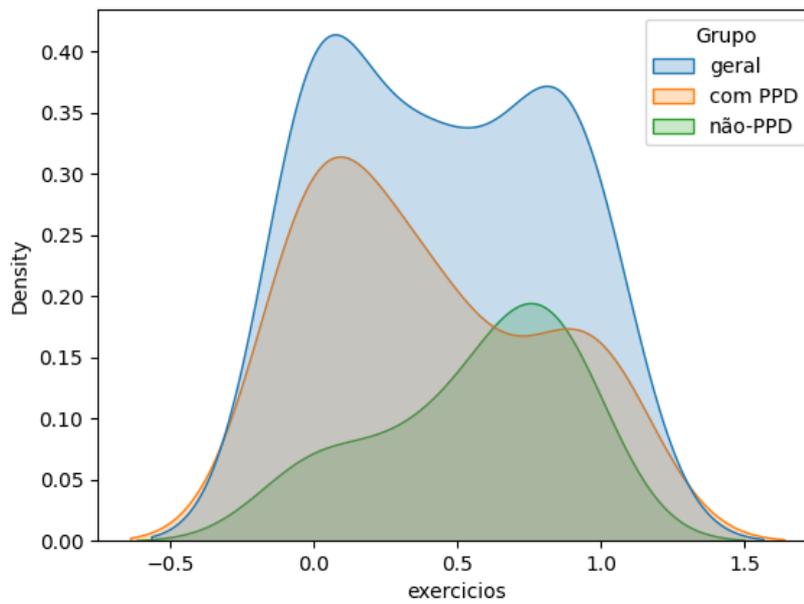
Fonte: própria

Figura 23 – Boxplot de exercícios físicos



Fonte: própria

Figura 24 – Gráfico de densidade de exercícios físicos



Fonte: própria

## 6.4 Análise bivariada

Na seção anterior, ainda na análise univariável, foi possível avaliar cada variável separadamente e criar especulações de relação entre os grupos e os sintomas depressivos. Nesta seção, o intuito será avaliar se as variáveis dos sensores possuem alguma dependência entre elas, como se o sono e a frequência cardíaca variam de forma positiva ou negativa de acordo com algum padrão. Esses resultados permitem investigar as associações entre as variáveis da pesquisa, como frequência cardíaca, contagem de passos, tempo de sono e a quantidade de atividades física praticadas no dia, em relação aos grupos com PPD e sem PPD.

### 6.4.1 Grupo geral

Tabela 5 – Correlação de Spearman aplicado ao grupo geral

	Freq. cardíaca	Sono	Passos	Exercícios
Freq. cardíaca		0.2964	0.0848	0.1214
Sono	0.2964		-0.3890	-0.0971
Passos	0.0848	-0.3890		0.5324
Exercícios	0.1214	-0.0971	0.5324	

Fonte: própria

Os resultados do grupo geral não demonstram muitas correlações fortes, nem positivas nem negativas. Uma correlação forte indica que uma variável está relacionada com outra, de forma que, se o valor de uma for alterada, a outra também será, demonstrando que ambas estão relacionadas de alguma forma. Valores de correlações significativos estão mais distante do 0, algo próximo entre -0.5 e 0.5. O motivo pela qual a tabela com o grupo geral não demonstrar grande correlação entre os valores, se deve a tabela incluir dados de todos os participantes, independentemente de terem ou não depressão. Isso pode resultar em uma grande variabilidade nos dados, tornando mais difícil identificar correlações significativas. Isso não ocorre nas tabelas específicas dos grupos, que têm uma amostra mais homogênea, o que torna mais fácil identificar padrões.

Apesar disso, os valores obtidos pela correlação levemente moderada entre passos e sono (-0.38) e exercícios e passos (0.53), no grupo geral, indica que são variáveis fortemente relacionadas independente dos grupos. Indivíduos com e sem depressão experimentam que um menor número de passos está associado a um aumento na qualidade do sono (correlação negativa), ou contrário, e que um maior número de passos está associado a um maior nível de atividade física (correlação positiva).

### 6.4.2 Grupo PPD

Passando para os grupos de classificadores, iniciando pelo grupo com PPD, nota-se que os valores de correlação tendem a não variar muito além do valor zero, demonstrando que as

amostras do grupo possuem menos correlações entre os dados, e que seu grupo aparenta ter características parecidas com o grupo geral.

Tabela 6 – Correlação de Spearman aplicado ao grupo com PPD

	Freq. cardíaca	Sono	Passos	Exercícios
Freq. cardíaca		0.2128	0.0355	0.1416
Sono	0.2128		-0.4511	-0.1503
Passos	0.0355	-0.4511		0.4486
Exercícios	0.1416	-0.1503	0.4486	

Fonte: própria

**Frequência cardíaca, sono e exercícios:** a frequência cardíaca possui uma correlação positiva fraca com sono (0.21) e exercícios (0.14), indicando que um aumento na frequência cardíaca está associado a um maior número de passos dados e a uma maior quantidade de exercícios realizados, de forma pouco significativa. Além disso, a correlação da frequência cardíaca com passos é quase nulo, o que sugere que esses dados não estão correlacionados, ou seja, o aumento ou diminuição de dado não afeta o outro.

**Sono e passos:** a correlação de sono com passos é negativa moderada (-0.45), sugerindo que quanto maior o número de passos, menor é o tempo de sono, ou contrário. Este dado cria uma concordância com os dados do grupo geral, onde ambos os grupos têm o mesmo comportamento.

**Passos e exercícios:** Por fim, a quantidade de passos tem uma correlação positiva moderada com exercícios físicos (0.44), sugerindo que um maior número de passos está associado a uma maior quantidade de exercícios realizados.

No grupo com PPD, as correlações são mais fracas e menos consistentes. Notavelmente, a frequência cardíaca mostra apenas uma correlações leves, que são com o sono e os exercícios, indicando que essas variáveis podem ter uma influência limitada sobre a frequência cardíaca nesse grupo. Além disso, a correlação quase nula entre a frequência cardíaca e os passos, sugere que um aumento ou diminuição na da quantidade de passos dados ao longo do dia não está associado com a alteração da frequência cardíaca. O sono, por sua vez, apresenta uma correlação moderada e negativa com os passos, indicando que um aumento no sono está associado a uma diminuição na quantidade de passos dados pelos estudantes com depressão. Essa relação pode indicar uma menor motivação ou pouca energia para se engajar em atividades físicas nesse grupo..

### 6.4.3 Grupo não PPD

Na análise do grupo não PPD (Tabela 7), é possível perceber que os valores de correlação começam flutuar um pouco mais para além do valor zero, demonstrando que as amostras do grupo possuem mais correlações entre os dados, e que seu grupo aparenta ter características próprias.

Tabela 7 – Correlação de Spearman aplicado ao grupo não PPD

	Freq. cardíaca	Sono	Passos	Exercícios
Freq. cardíaca		0.5181	0.3272	0.2596
Sono	0.5181		-0.1181	0.4373
Passos	0.3272	-0.1181		0.5603
Exercícios	0.2596	0.4373	0.5603	

Fonte: própria

**Frequência cardíaca:** possui correlações positivas significativas com todas as outras variáveis, indicando que o aumento da frequência cardíaca deste grupo está correlacionado com o aumento do valor das demais variáveis.

**Sono, frequência cardíaca e exercícios:** o sono possui uma correlação positiva moderada com frequência cardíaca (0.51), indicando que um aumento no sono pode levar a um aumento na frequência cardíaca. Ao mesmo tempo, a correlação do sono com exercícios também é positiva e moderada (0.43), indicando que um maior tempo de sono está associado a uma maior quantidade de exercícios realizados.

**Passos e exercícios** A quantidade de passos dados tem uma correlação positiva moderada com a frequência cardíaca (0.32) e uma correlação positiva moderada com exercícios físicos (0.59). Isso sugere que um maior número de passos está associado a um aumento na frequência cardíaca e a uma maior quantidade de exercícios realizados.

No grupo não PPD, as correlações demonstram um padrão mais consistente e revelam relações moderadas entre as variáveis. Especificamente, a frequência cardíaca exibe uma associação moderada com o número de passos dados e a prática de exercícios, sugerindo que um aumento na atividade física pode levar a um aumento na frequência cardíaca. Além disso, o sono apresenta correlações positivas com exercícios, indicando que uma melhor qualidade de sono está relacionada a um maior engajamento nessas atividades físicas. Essas análises sugerem uma relação saudável entre essas variáveis no grupo não PPD.

#### 6.4.4 Resumo dos grupos

A Tabela 8 mostra um resumo das principais características de correlação dos grupos. Em modo geral, o grupo com PPD não apresenta nenhuma correlação significativa com as demais variáveis além de sono, enquanto no grupo não PPD apresenta uma correlação moderada positiva com todas as demais variáveis, como passos (0.32) e exercícios (0.25).

Com base nos dados gerados pela análise, é possível afirmar que existem, sim, correlações entre as variáveis. Também temos características de comportamento diferentes para cada grupo e em como as variáveis se relacionam em cada um deles. Com a análise dos dados foi possível observar, mesmo que de forma preliminar, que há padrões entre os grupos. No entanto, o alcance dessa análise, embora informativo, é limitado pela natureza bivariada das correlações de Spearman, que não capturam a total complexidade das relações entre as variáveis.

Tabela 8 – Resumo dos grupos da correlação de Spearman

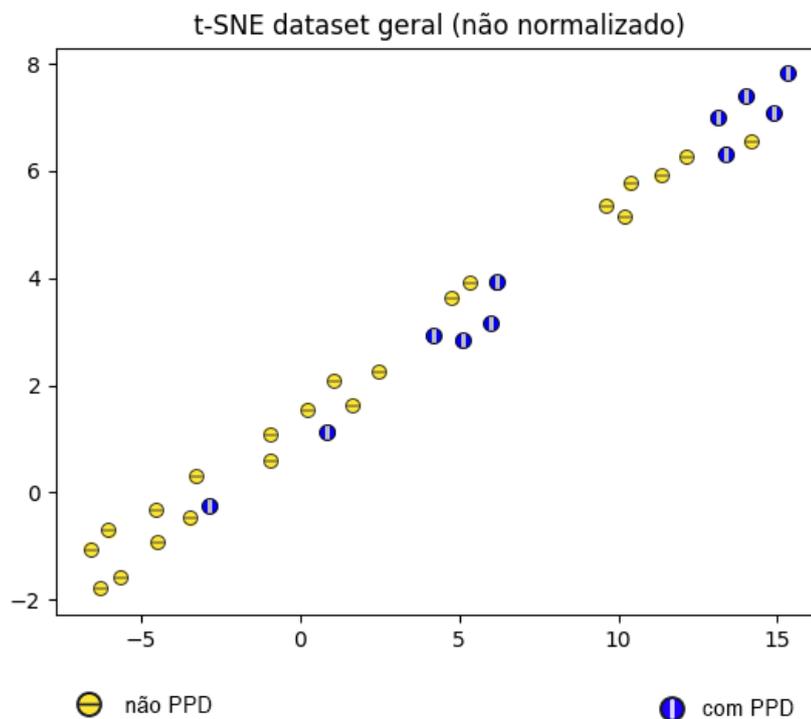
	Freq. cardíaca e sono	Sono e passos	Passos e exercícios	Sono e exercícios
<b>Grupo geral</b>	0.2964	-0.3890	0.5324	-0.0971
<b>PPD</b>	0.2128	-0.4511	0.4486	-0.1503
<b>Não PPD</b>	0.5181	-0.1181	0.5603	0.4373

Fonte: própria

## 6.5 Análise multivariada

O t-SNE explora as relações complexas entre as variáveis de frequência cardíaca, sono, exercícios e passos, preservando a estrutura das relações entre pontos de dados, concentrando-se em manter grupos de pontos semelhantes próximos e separar pontos de grupos distintos em um espaço de menor dimensão. Com isso é possível avaliar se os grupos com PPD e sem PPD demonstram comportamentos semelhantes ou distintos, com base nos dados coletados nesta pesquisa.

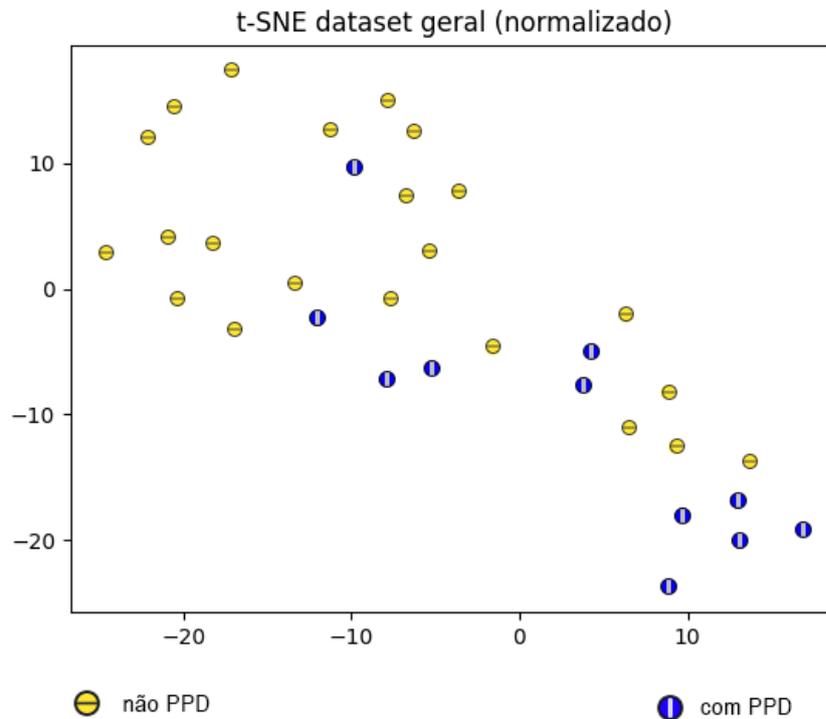
Figura 25 – t-SNE não normalizado



Fonte: própria

Os resultados gerados pelo t-SNE são representados nas Figuras 25 e 26. Pontos que estão próximos no gráfico t-SNE são aqueles que têm semelhanças significativas nas variáveis, enquanto pontos distantes indicam diferenças. Grupos ou aglomerados de pontos no gráfico t-SNE

Figura 26 – t-SNE normalizado



Fonte: própria

sugerem a presença de padrões ou *clusters* nos dados. Esses grupos podem ser facilmente encontrados nos resultados gerados. Os pontos em azul, representando o grupo com PPD, tendem a permanecer juntos no canto do gráfico ou no centro dele. Já a maior parte dos pontos amarelos, representando o grupo não PPD, embora tenha uma variação de posições, tende, na sua maioria, a estar próxima, no outro canto do gráfico ou também próxima do centro. Esse comportamento de grupos pode ser observado tanto no gráfico normalizado quanto no não normalizado. A normalização foi feita na padronização z-score, uma abordagem comum para normalização de dados que envolve a subtração da média dos dados e a divisão pelo desvio padrão.

Ao conduzir a análise de redução de dimensionalidade com o t-SNE, foram testados vários valores de perplexidade para avaliar seu impacto na visualização dos dados. Foi usada a medida de perplexidade 5 para amostra dos gráficos anteriores. Em todos os experimentos, mesmo com diferentes números de perplexidade, os gráficos resultantes mantiveram grupos muitos similares e padrões subjacentes. Esse resultado demonstra que os dados coletados na pesquisa estão alinhados com o objetivo fundamental de identificar e compreender os grupos com e sem PPD.

Ainda na análise dos dados do gráfico t-SNE, uma observação notável é a proximidade de pontos de diferentes grupos, incluindo a coexistência de pontos amarelos (não PPD) na mesma região onde se encontram pontos azuis (com PPD), e também o contrário. Isso sugere a existência de indivíduos que, embora classificados como pertencentes a um grupo específico com base no PHQ-9, exibem comportamentos associados a outro grupo. Por exemplo, pontos azuis, representando o grupo com PPD, encontrados na proximidade da região predominada pelos

pontos amarelos (grupo não PPD), indicam a presença de indivíduos que experimentam sintomas depressivos, mas podem passar despercebidos em um ambiente onde a maioria não percebe sua necessidade de apoio ou ajuda. Da mesma forma, identificamos indivíduos classificados como não PPD, mas que demonstram comportamentos semelhantes aos daqueles com PPD (visíveis na Figura 25 no canto superior direito), sugerindo um alerta para a possibilidade de sintomas depressivos não diagnosticados, o que permite ações preventivas antes da situação se agravar.

Com base nos dados comportamentais dos grupos e com os padrões formados pelo t-SNE, tomando como base a Figura 25, pode-se supor que se fosse adicionada uma linha diagonal no gráfico, criando uma divisão distintiva entre os indivíduos na parte superior e inferior, dos indivíduos que ficaram na parte superior, 70% estaria corretamente classificado.

Esta seção de análise estatística apresentou os resultados obtidos por meio da análise exploratória realizada com os dados de 33 estudantes universitários brasileiros participantes da primeira fase do projeto Amive. Com os resultados apresentados, é possível concluir que há diferenças significativas entre os grupos com e sem PPD. A análise univariável mostrou as variações de comportamento de cada variável dos sensores e como elas tendem a aumentar ou diminuir de acordo com cada grupo. A análise bivariada revelou a existência de correlações entre os dados dos participantes, demonstrando que cada grupo apresenta correlações positivas e negativas específicas para cada tipo de dado. Quando uma variável se altera, a outra tende a acompanhar essa mudança, e esse padrão de correlação difere entre os grupos. A análise multivariada conclui o processo, ilustrando o comportamento dos grupos e a formação de padrões distintos. Todo esse processo evidencia a existência de diferenças entre os grupos e, por meio do processamento dos dados de frequência cardíaca, contagem de passos, exercícios físicos e tempo de sono, torna-se possível identificar sintomas depressivos em indivíduos. Com o padrão de comportamento identificado, é viável implementar medidas de apoio e ajuda aos que sofrem de depressão, além de ações preventivas com identificação precoce.

---

## Capítulo 7

# Considerações Finais

---

Para a realização desta pesquisa, foi desenvolvida uma infraestrutura computacional que possibilitou a coleta de dados móveis de relógios inteligentes e celulares por meio de um aplicativo dedicado. O estudo foi conduzido com a participação de 89 indivíduos, sendo que 28 deles receberam empréstimo de relógios inteligentes para a pesquisa. Os dados incluíram informações diárias sobre a frequência cardíaca, tempo de sono, quantidade de passos e atividades físicas dos participantes, além de questionários de saúde mental, como sentimento subjetivo, questionário PHQ-9 e WHODAS-II. Além disso, os usuários participaram do processo por meio de um questionário avaliativo, usado conforme princípios da abordagem Human in the Loop para melhorar a qualidade dos dados usados na geração de resultados.

Os dados foram armazenados em um banco de dados e, após a conclusão da etapa de coleta de dados com os usuários, foi realizado o processamento desses dados, com o objetivo de identificar padrões de comportamento depressivo nos participantes. Os dados passaram por um processo de limpeza com posterior imputação que resultou em um dataset final com dados de 33 participantes. Para análise e geração dos resultados, foi empregado a abordagem de análise exploratória, dividida em três etapas: observação univariada dos dados, correlação entre variáveis e encerramento com análise multivariada das informações obtidas. Nesse processo, utilizou-se técnicas analíticas com a criação de histogramas, boxplots e gráficos de densidade. Além disso, foi aplicado o método do coeficiente de correlação de Spearman para investigação de correlação das variáveis em uma análise bivariada, e, por fim, explorou-se a redução de dimensionalidade em quatro variáveis usando a técnica de t-SNE, concluindo a análise exploratória multivariada. Essas técnicas permitiram visualizar e analisar o comportamento dos dados, gerando resultados que possibilitaram chegar em uma conclusão sobre os padrões de comportamento dos grupos classificados com e sem o possível perfil depressivo.

Os resultados da pesquisa indicaram que os grupos classificados como PPD e não PPD, com base nas respostas do PHQ-9, apresentaram comportamentos coerentes com as expectativas da literatura. Os participantes com PPD demonstraram variações no sono, menor atividade

física, menor número de passos e uma média de frequência cardíaca mais alta. Esses achados são consistentes com os sintomas associados à depressão, como alterações no sono, perda de interesse ou prazer, falta de energia e desânimo, que conseqüentemente pode induzir outros sintomas como humor depressivo.

## 7.1 Limitações

O aplicativo inicialmente pensado para a pesquisa foi idealizado de forma que os dados coletados pelos relógios fossem enviados de forma direta para o servidor, sem a necessidade da interferência de um celular para isso. Para realização desse feito, seria necessário desenvolver um aplicativo para o relógio inteligente, para que ele mesmo pudesse fazer o envio dos dados. Durante a fase de estudo para desenvolvimento da pesquisa, foi utilizado um primeiro modelo de relógio para iniciar os testes, o Samsung Galaxy Watch 2. Seu sistema operacional, Tizen, desenvolvido pela Samsung, possuía alguns limitantes em relação aos tipos de aplicativos que poderiam ser instalados. Além disso, houve a barreira do alto esforço de programação necessário para se desenvolver um software para esse sistema operacional, tornando inviável o desenvolvimento de um aplicativo próprio para o relógio. Dessa forma, mesmo buscando alternativas de aplicativos que pudessem coletar os dados do relógio sem a intervenção do usuário, nenhum outro aplicativo disponível na loja atendia a esse desejo. Sendo assim, ficou decidido o uso de aplicativos de terceiros que fizessem a tarefa. Vários aplicativos de saúde que possuíam integrações com relógios foram testados, e o Samsung Health foi o que se saiu melhor devido a facilitação de exportação do pacote de dados.

O Samsung Health possui uma API<sup>1</sup> que poderia permitir o projeto coletar e enviar os dados de forma passiva para o servidor, porém, o acesso a ela esteve restrito ao programa de parcerias, que foi fechado em 2020, impossibilitando o uso desse recurso no período de desenvolvimento da solução da pesquisa entre 2021 e 2022. Por fim, concluiu-se que a melhor solução seria o envio dos dados, contando com a participação ativa do usuário, passando pelo processo de abrir o aplicativo da Samsung Health, manualmente, baixando os pacotes de dados e enviando para o aplicativo da pesquisa. Esse foi um dos maiores problemas enfrentados durante o desenvolvimento de soluções para coleta de dados da pesquisa: a dificuldade de integração de serviços com o relógio para envio de dados passivos ao servidor.

Apesar das adversidades encontradas durante o desenvolvimento do aplicativo, todas elas foram contornadas e repensadas para que a solução final pudesse ser criada e usada pelos usuários. No final, a infraestrutura cumpriu o seu propósito de coletar os dados de sensores dos dispositivos móveis dos participantes (celulares e relógios), assim como também armazenar as respostas dos questionários de escalas de depressão.

Além do aplicativo, a pesquisa também enfrentou desafios referente a quantidade de dados enviados pelos participantes. Os dados variaram entre os 28 participantes que tinham relógios inteligentes e aqueles que dependiam exclusivamente de seus celulares, uma vez que os sensores

---

<sup>1</sup> Conjunto de funcionalidades para que diferentes sistemas de software possam se comunicar e trabalhar juntos

dos celulares não captavam certos dados com tanta precisão quanto os sensores do relógio. Como resultado, a limpeza e preparação dos dados levou à redução da amostra final de dados disponíveis para análise exploratória com 33 dos 89 participantes da pesquisa.

A falta de dados nos dispositivos móveis sem relógio ocorreu principalmente devido à ausência de coleta de frequência cardíaca e à necessidade de inserção manual do tempo de sono pelos participantes no aplicativo Samsung Health. Ainda assim, mesmo com participantes que possuíam os relógios, houve imputações de dados referentes ao tempo de sono, uma vez que os participantes relataram não dormir com o relógio. Outro problema estava relacionado com a curta duração de bateria do relógio Galaxy Watch 4, onde houve perda de dados nos momentos em que os participantes seguiram suas rotinas enquanto precisavam tirar o seu relógio para carregar.

A redução da amostra também se deu principalmente pela remoção de outliers indicados pelos próprios participantes nos questionários avaliativos. Os participantes relataram dados que não estavam condizentes com o que viveram ao longo do dia. Ainda, os pesquisadores do projeto Amive com acesso ao Amive Web (Capítulo 5.3) relataram diversos dados discrepantes com a realidade, como exercícios realizados durante dias seguidos e tempo de sono que ultrapassavam 24 horas. Vale ressaltar que os dados coletados e armazenados na pesquisa não passaram por nenhum tipo de filtragem prévia, ou seja, os dados obtidos pelos sensores vieram diretamente pelo aplicativo Samsung Health. Isso pode ser um indicador de que, apesar de poucas ocorrências relatadas, tanto o aplicativo quanto o próprio sensor criaram registros que não parecem fazer parte de uma rotina diária dos participantes. Apesar dos desafios, a pesquisa conseguiu trazer a contribuição esperada e gerar os resultados de forma satisfatória de acordo com seu objetivo principal através de uma análise descritiva.

## 7.2 Acesso aos Recursos da Pesquisa

Visando uma maior contribuição para a comunidade acadêmica e científica, este trabalho oferece a oportunidade de acesso aos recursos da pesquisa. Estão à disposição os dados do banco de dados com todas as informações de sensores e questionários enviados pelos participantes, que foram devidamente anonimizados para proteger a privacidade deles. Além disso, o repositório com o código-fonte usado para gerar os resultados da análise exploratória está disponível. Essa é uma forma de promover a pesquisa colaborativa e o progresso na compreensão da depressão, permitindo que outros pesquisadores explorem e utilizem essas informações para pesquisas posteriores.

Os dados da pesquisa estão armazenados em um banco de dados relacional do PostgreSQL, que durante a pesquisa foi hospedado na plataforma Supabase<sup>2</sup>, na sua versão gratuita. Este banco de dados abriga diversas tabelas, incluindo registros dos sensores, dados de fenotipagem digital e respostas aos questionários de saúde mental dos participantes. É relevante destacar que estão disponibilizados os dados brutos e não passaram por nenhum tipo de tratamento, ou seja, não foram submetidos à imputação ou à remoção de outliers como os realizadas ao longo

---

<sup>2</sup> Disponível em: <https://supabase.com>

da pesquisa, sendo o mesmo que os pesquisadores do Amive teve acesso à consulta no Amive Web na fase 1 do projeto.

Os dados usados no dataset final para gerar os resultados da pesquisa, já tratado com a remoção de outliers e imputação dos dados, também está disponível. O diretório com o dataset inclui também os scripts em Python usados para gerar os resultados da análise univariada, matriz de Spearman e o gráfico t-SNE.

Além da disponibilização do banco de dados e do código-fonte usado para gerar os resultados do dataset, o Amive Web, que apresenta todos os dados de forma visual para serem analisados e estudados por pesquisadores de qualquer área, também pode ser acessado.

Aos interessados em acessar os recursos disponibilizados na pesquisa, será necessário solicitar um pedido de acesso por meio de um formulário consentindo com os termos de compromisso e as questões éticas da pesquisa. Disponível em <https://forms.gle/Shz7bsHw6uUif1Dh6>.

### 7.3 Considerações finais

No decorrer deste trabalho, foi percorrido um caminho desafiador para explorar e analisar os sinais e comportamentos de indivíduos em relação ao possível perfil depressivo, almejando usufruir ao máximo o uso das capacidades que as tecnologias móveis e vestíveis têm a oferecer. As etapas que incluíram a coleta de dados, a imputação e a análise exploratória, forneceram resultados significativos sobre como é possível identificar e analisar padrões comportamentais associados a um possível perfil depressivo no ambiente universitário. Enquanto avança-se na era da tecnologia vestível e da análise de dados, é importante reconhecer as implicações que esses avanços podem ter na área da saúde e como se pode usar a computação para trazer benefícios para a vida das pessoas. À medida que os dispositivos do cotidiano se tornam ferramentas de monitoramento e diagnóstico, há um potencial imenso para melhorar o acompanhamento clínico e a intervenção precoce em questões de saúde mental, principalmente relacionado ao transtorno depressivo maior.

Os resultados gerados pela análise exploratória **sugerem que o objetivo da pesquisa foi alcançado** e a infraestrutura construída permitiu a coleta de dados que apoiam a identificação de padrões comportamentais relacionados ao possível perfil depressivo no contexto universitário brasileiro. Este trabalho também traz inovações no modelo de infraestrutura computacional ao empregar a integração de diversas funcionalidades em um único ambiente de estudo, como questionários de saúde mental integrado no próprio aplicativo usado pelos participantes do estudo, juntamente a consulta de histórico de envios, que possibilitou o uso da abordagem Human in the Loop para validação de dados dos próprios participantes, usados mais tarde para remoção de outliers do conjunto de dados para análise exploratória. Ainda, pode-se incluir o sistema de visualização de dados online para análise e estudo para pesquisadores de todas as áreas, e também a coleta de dados automatizado de redes sociais como o Twitter.

Uma vez que esses sinais possam ser identificados nas pessoas em tempo real, a tecnologia vestível junto ao processamento computacional podem favorecer o acompanhamento de profissionais de saúde na identificação de agravamento de quadros e no auxílio da intervenção precoce

---

da depressão. O próprio indivíduo pode se beneficiar com a comodidade do acompanhamento de forma passiva e não invasiva. Este trabalho contribui cientificamente na área da computação e da saúde com a criação de uma solução computacional que pode apoiar o diagnóstico e criar intervenções para pessoas com possível perfil depressivo, analisando o comportamento delas por meio da coleta de dados de fenotipagem digital advindos de sensores presentes em dispositivos móveis e vestíveis, como celulares e relógios inteligentes.



---

## Referências

---

AHMED, M. S. et al. Smartphone usage behavior between depressed and non-depressed students. In: **Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers**. ACM, 2020. Disponível em: <<https://doi.org/10.1145/3410530.3414441>>.

AL., J. S. et. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. **The Lancet**, Elsevier BV, v. 392, n. 10159, p. 1789–1858, nov 2018. Disponível em: <<https://doi.org/10.1016%2Fs0140-6736%2818%2932279-7>>.

ALAMOUDI, D. et al. The feasibility of using smartphone sensors to track insomnia, depression, and anxiety in adults and young adults: Narrative review. **JMIR mHealth and uHealth**, JMIR Publications Inc., v. 11, p. e44123, fev. 2023. Disponível em: <<https://doi.org/10.2196/44123>>.

BECK, A. T. An inventory for measuring depression. **Archives of General Psychiatry**, American Medical Association (AMA), v. 4, n. 6, p. 561, jun. 1961. Disponível em: <<https://doi.org/10.1001/archpsyc.1961.01710120031004>>.

BIANCHI R. G. ; NERIS, V. P. d. A. Com que cor eu vou? um estudo sobre cores e emoções na interação humano-computador. **XIV Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais**, 2015.

BLEICHER, T.; OLIVEIRA, R. C. N. de. Políticas de assistência estudantil em saúde nos institutos e universidades federais. **Psicologia Escolar e Educacional**, FapUNIFESP (SciELO), v. 20, n. 3, p. 543–549, dez. 2016. Disponível em: <<https://doi.org/10.1590/2175-3539201502031040>>.

BOSE SARRA L. HEDDEN, R. N. L. J.; PARK-LEE, E. **Key Substance Use and Mental Health Indicators in the United States: Results from the 2017 National Survey on Drug Use and Health**. [S.l.], 2018. 124 p.

CHIKERSAL, P. et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: A machine learning approach with robust feature selection. Association for Computing Machinery, New York, NY, USA, v. 28, n. 1, jan 2021. ISSN 1073-0516. Disponível em: <<https://doi.org/10.1145/3422821>>.

COPPERSMITH, G. et al. Natural language processing of social media as screening for suicide risk. **Biomedical Informatics Insights**, SAGE Publications, v. 10, p. 117822261879286, jan. 2018. Disponível em: <<https://doi.org/10.1177/1178222618792860>>.

- COSTA, J. et al. Boostmeup: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch. Association for Computing Machinery, New York, NY, USA, v. 3, n. 2, jun 2019. Disponível em: <<https://doi.org/10.1145/3328911>>.
- CUIJPERS, P. et al. Psychological treatment of depression in primary care: Recent developments. **Current Psychiatry Reports**, Springer Science and Business Media LLC, v. 21, n. 12, nov. 2019. Disponível em: <<https://doi.org/10.1007/s11920-019-1117-x>>.
- DAI, R. et al. Multi-task learning for randomized controlled trials: A case study on predicting depression with wearable data. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 6, n. 2, jul 2022. Disponível em: <<https://doi.org/10.1145/3534591>>.
- Decreto n 7.234. (BRASIL) Decreto n 7.234, de 19 de julho de 2010. Dispõe sobre o Programa Nacional de Assistência Estudantil - PNAES. 2010. <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2010/decreto/d7234.htm](http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7234.htm)>. Acesso em: 7 nov. 2022.
- DSM. Depressive disorders. American Psychiatric Association Publishing, mar. 2022. Disponível em: <[https://doi.org/10.1176/appi.books.9780890425787.x04\\_depressive\\_disorders](https://doi.org/10.1176/appi.books.9780890425787.x04_depressive_disorders)>.
- EISENBERG, D.; GOLBERSTEIN, E.; GOLLUST, S. E. Help-seeking and access to mental health care in a university student population. **Medical Care**, Ovid Technologies (Wolters Kluwer Health), v. 45, n. 7, p. 594–601, jul. 2007. Disponível em: <<https://doi.org/10.1097/mlr.0b013e31803bb4c1>>.
- EVMENOVA, A. S. et al. Designing a wearable technology intervention to support young adults with intellectual and developmental disabilities in inclusive postsecondary academic environments. **Journal of Special Education Technology**, v. 34, n. 2, p. 92–105, 2019. Disponível em: <<https://doi.org/10.1177/0162643418795833>>.
- FARHAN, A. A. et al. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. p. 1–8, 2016.
- FUKUDA, S. et al. Predicting depression and anxiety mood by wrist-worn sleep sensor. In: **2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)**. [S.l.: s.n.], 2020. p. 1–6.
- GARCIA-CEJA, E. et al. Depresjon: A motor activity database of depression episodes in unipolar and bipolar patients. In: **Proceedings of the 9th ACM Multimedia Systems Conference**. New York, NY, USA: Association for Computing Machinery, 2018. (MMSys '18), p. 472–477. ISBN 9781450351928. Disponível em: <<https://doi.org/10.1145/3204949.3208125>>.
- GHANDEHARIOUN, A. et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In: **2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)**. [S.l.: s.n.], 2017. p. 325–332.
- HARARI, G. M. et al. Using smartphones to collect behavioral data in psychological science. **Perspectives on Psychological Science**, SAGE Publications, v. 11, n. 6, p. 838–854, nov. 2016. Disponível em: <<https://doi.org/10.1177/1745691616650285>>.
- HARTMANN, R. et al. Heart rate variability as indicator of clinical state in depression. **Frontiers in Psychiatry**, Frontiers Media SA, v. 9, jan. 2019. Disponível em: <<https://doi.org/10.3389/fpsy.2018.00735>>.

- HEVNER, A.; CHATTERJEE, S. Design science research in information systems. In: **Integrated Series in Information Systems**. Springer US, 2010. p. 9–22. Disponível em: <[https://doi.org/10.1007/978-1-4419-5653-8\\_2](https://doi.org/10.1007/978-1-4419-5653-8_2)>.
- IBRAHIM, A. K. et al. A systematic review of studies of depression prevalence in university students. **Journal of Psychiatric Research**, Elsevier BV, v. 47, n. 3, p. 391–400, mar. 2013. Disponível em: <<https://doi.org/10.1016/j.jpsychires.2012.11.015>>.
- JACOBSON, N. C.; CHUNG, Y. J. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. **Sensors**, MDPI AG, v. 20, n. 12, p. 3572, jun. 2020. Disponível em: <<https://doi.org/10.3390/s20123572>>.
- KILASKAR, M. et al. Machine learning algorithms for analysis and prediction of depression. **SN Computer Science**, v. 3, n. 2, p. 103, Dec 2021. ISSN 2661-8907. Disponível em: <<https://doi.org/10.1007/s42979-021-00967-0>>.
- KIM, J.; HONG, J.; CHOI, Y. Automatic depression prediction using screen lock/unlock data on the smartphone. In: **2021 18th International Conference on Ubiquitous Robots (UR)**. IEEE, 2021. Disponível em: <<https://doi.org/10.1109/ur52253.2021.9494673>>.
- KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Department of Computer Science, Keele University, UK, 2004.
- KLIMSTRA, T. A. et al. Come rain or come shine: Individual differences in how weather affects mood. **Emotion**, American Psychological Association (APA), v. 11, n. 6, p. 1495–1499, 2011. Disponível em: <<https://doi.org/10.1037/a0024649>>.
- KNAPP, M.; WONG, G. Economics and mental health: the current scenario. **World Psychiatry**, Wiley, v. 19, n. 1, p. 3–14, jan. 2020. Disponível em: <<https://doi.org/10.1002/wps.20692>>.
- KOHN, R. et al. The treatment gap in mental health care. **Bull World Health Organ**, v. 82, n. 11, p. 858–866, dez. 2004.
- KUMAR, P. et al. A machine learning implementation for mental health care. application: Smart watch for depression detection. In: **2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence)**. [S.l.: s.n.], 2021. p. 568–574.
- LAUCKNER, C.; HILL, M.; INGRAM, L. A. An exploratory study of the relationship between social technology use and depression among college students. **Journal of College Student Psychotherapy**, Informa UK Limited, v. 34, n. 1, p. 33–39, set. 2018. Disponível em: <<https://doi.org/10.1080/87568225.2018.1508396>>.
- LEE, S. et al. Current advances in wearable devices and their sensors in patients with depression. **Frontiers in Psychiatry**, Frontiers Media SA, v. 12, jun. 2021. Disponível em: <<https://doi.org/10.3389/fpsy.2021.672347>>.
- LIN, M. et al. Bewell+: Multi-dimensional wellbeing monitoring with community-guided user feedback and energy optimization. In: **Proceedings of the Conference on Wireless Health**. New York, NY, USA: Association for Computing Machinery, 2012. (WH '12). ISBN 9781450317603. Disponível em: <<https://doi.org/10.1145/2448096.2448106>>.
- LIU, S. et al. Heterogeneous network approach to predict individuals' mental health. **ACM Transactions on Knowledge Discovery from Data**, Association for Computing Machinery (ACM), v. 15, n. 2, p. 1–26, abr. 2021. Disponível em: <<https://doi.org/10.1145/3429446>>.

- LU, J. et al. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. Association for Computing Machinery, New York, NY, USA, v. 2, n. 1, mar 2018. Disponível em: <<https://doi.org/10.1145/3191753>>.
- MAATEN, L. van der; HINTON, G. Visualizing data using t-sne. **Journal of Machine Learning Research**, v. 9, n. 86, p. 2579–2605, 2008. Disponível em: <<http://jmlr.org/papers/v9/vandermaaten08a.html>>.
- MASUD, M. T. et al. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. **Journal of Biomedical Informatics**, Elsevier BV, v. 103, p. 103371, mar. 2020. Disponível em: <<https://doi.org/10.1016/j.jbi.2019.103371>>.
- \_\_\_\_\_. Non-pervasive monitoring of daily-life behavior to access depressive symptom severity via smartphone technology. In: **2020 IEEE Region 10 Symposium (TENSYP)**. [S.l.: s.n.], 2020. p. 602–607.
- MEHROTRA, A.; HENDLEY, R.; MUSOLESI, M. Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction. In: **Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct**. ACM, 2016. Disponível em: <<https://doi.org/10.1145/2968219.2968299>>.
- MISHRA, V. From sensing to intervention for mental and behavioral health. In: . New York, NY, USA: Association for Computing Machinery, 2019. (UbiComp/ISWC '19 Adjunct), p. 388–392. ISBN 9781450368698. Disponível em: <<https://doi.org/10.1145/3341162.3349304>>.
- MUNRO, R. **Human-In-The-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI**. [S.l.: s.n.], 2021.
- NANOUE, E. **7 Useless Mental Health Apps You Should Avoid**. 2022. <<https://www.makeuseof.com/useless-mental-health-apps-avoid>>. Acessado em: outubro de 2023.
- NARZIEV, N. et al. STDD: Short-term depression detection with passive sensing. **Sensors**, MDPI AG, v. 20, n. 5, p. 1396, mar. 2020. Disponível em: <<https://doi.org/10.3390/s20051396>>.
- NGUYEN, B. et al. Clustering and feature analysis of smartphone data for depression monitoring. In: **2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)**. [S.l.: s.n.], 2021. p. 113–116.
- OLIVEIRA, A. B. d. Modelagem de arquitetura baseada em microsserviços com foco em segurança. **Repositório Institucional UFSCar**, Universidade Federal de São Carlos, p. 68, maio 2022. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/16036>>.
- ONNELA, J.-P. Opportunities and challenges in the collection and analysis of digital phenotyping data. **Neuropsychopharmacology**, Springer Science and Business Media LLC, v. 46, n. 1, p. 45–54, jul. 2020. Disponível em: <<https://doi.org/10.1038/s41386-020-0771-3>>.
- PACHECO, J. P. et al. Mental health problems among medical students in brazil: a systematic review and meta-analysis. **Revista Brasileira de Psiquiatria**, EDITORA SCIENTIFIC, v. 39, n. 4, p. 369–378, ago. 2017. Disponível em: <<https://doi.org/10.1590/1516-4446-2017-2223>>.
- PEDRELLI, P. et al. Monitoring changes in depression severity using wearable and mobile sensors. **Frontiers in Psychiatry**, Frontiers Media SA, v. 11, dez. 2020. Disponível em: <<https://doi.org/10.3389/fpsy.2020.584711>>.

- PENEDO, F. J.; DAHN, J. R. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. **Current Opinion in Psychiatry**, Ovid Technologies (Wolters Kluwer Health), v. 18, n. 2, p. 189–193, mar. 2005. Disponível em: <<https://doi.org/10.1097/00001504-200503000-00013>>.
- SAEB, S. et al. The relationship between mobile phone location sensor data and depressive symptom severity. **PeerJ**, PeerJ, v. 4, p. e2537, set. 2016. Disponível em: <<https://doi.org/10.7717/peerj.2537>>.
- SHEIKH, M.; QASSEM, M.; KYRIACOU, P. A. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. **Frontiers in Digital Health**, Frontiers Media SA, v. 3, abr. 2021. Disponível em: <<https://doi.org/10.3389/fdgth.2021.662811>>.
- SWARTZ, M. H. A. Recognition and treatment of depression. **AMA Journal of Ethics**, American Medical Association (AMA), v. 7, n. 6, jun. 2005. Disponível em: <<https://doi.org/10.1001/virtualmentor.2005.7.6.cpr11-0506>>.
- VIGITEL. **Vigitel Brasil 2021 - Estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas**. [S.l.], 2022. 131 p.
- WANG, P. S. et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. **Lancet**, v. 370, n. 9590, p. 841–850, set. 2007.
- WANG, P. S.; SIMON, G.; KESSLER, R. C. The economic burden of depression and the cost-effectiveness of treatment. **International Journal of Methods in Psychiatric Research**, Wiley, v. 12, n. 1, p. 22–33, fev. 2003. Disponível em: <<https://doi.org/10.1002/mpr.139>>.
- WANG, R. et al. StudentLife. In: **Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. ACM, 2014. Disponível em: <<https://doi.org/10.1145/2632048.2632054>>.
- \_\_\_\_\_. Smartgpa: How smartphones can assess and predict academic performance of college students. **GetMobile: Mobile Comp. and Comm.**, Association for Computing Machinery, New York, NY, USA, v. 19, n. 4, p. 13–17, mar 2016. ISSN 2375-0529. Disponível em: <<https://doi.org/10.1145/2904337.2904343>>.
- \_\_\_\_\_. Tracking depression dynamics in college students using mobile phone and wearable sensing. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, Association for Computing Machinery (ACM), v. 2, n. 1, p. 1–26, mar. 2018. Disponível em: <<https://doi.org/10.1145/3191775>>.
- \_\_\_\_\_. Tracking depression dynamics in college students using mobile phone and wearable sensing. Association for Computing Machinery, New York, NY, USA, v. 2, n. 1, mar 2018. Disponível em: <<https://doi.org/10.1145/3191775>>.
- WARE, S. et al. Predicting depressive symptoms using smartphone data. **Smart Health**, Elsevier BV, v. 15, p. 100093, mar. 2020. Disponível em: <<https://doi.org/10.1016/j.smhl.2019.100093>>.
- WHO. **Depression and other common mental disorders: global health estimates**. [S.l.], 2017. 24 p. p.
- \_\_\_\_\_. **Mental Health and COVID-19: Early evidence of the pandemic's impact**. [S.l.], 2021. 11 p.

ZHENG, H.; MOTTI, V. G. Assisting students with intellectual and developmental disabilities in inclusive education with smartwatches. In: **Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2018. (CHI '18), p. 1–12. ISBN 9781450356206. Disponível em: <<https://doi.org/10.1145/3173574.3173924>>.

ZUORBA, H. D.; OLAN, C. L. O.; CANTARA, A. D. A framework for identifying excessive sadness in students through twitter and facebook in the philippines. In: **Proceedings of the International Conference on Bioinformatics Research and Applications 2017**. New York, NY, USA: Association for Computing Machinery, 2017. (ICBRA 2017), p. 52–56. ISBN 9781450353823. Disponível em: <<https://doi.org/10.1145/3175587.3175600>>.