



UNIVERSIDADE FEDERAL DO MARANHÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
Doutorado em Ciência da Computação Associação
UFMA/UFPI

Ivan Rodrigues de Moura

**Identificação de Padrões Multimodais de
Comportamento Humano Utilizando Fenotipagem Digital**

Orientador: Prof. Dr. Francisco José da Silva e Silva
Coorientador: Prof. Dr. Luciano Reis Coutinho

São Luís - MA
Dezembro, 2023

Ivan Rodrigues de Moura

Identificação de Padrões Multimodais de Comportamento Humano Utilizando Fenotipagem Digital

TESE DE DOUTORADO

Tese apresentada como requisito parcial para obtenção do título de Doutor em Ciência da Computação, ao Doutorado em Ciência da Computação, Associação UFMA/UFPI.

Orientador: Prof. Dr. Francisco José da Silva e Silva
Co-orientador: Prof. Dr. Luciano Reis Coutinho

São Luís - MA
Dezembro, 2023

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Moura, Ivan.

Identificação de Padrões Multimodais de Comportamento Humano Utilizando Fenotipagem Digital / Ivan Moura. - 2023.

113 f.

Coorientador(a): Luciano Coutinho.

Orientador(a): Francisco Silva.

Tese (Doutorado) - Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI, Universidade Federal do Maranhão, São Luís - Maranhão, 2023.

1. Comportamento humano. 2. Dispositivos ubíquos. 3. Fenotipagem digital. 4. Padrões comportamentais. I. Coutinho, Luciano. II. Silva, Francisco. III. Título.

Ivan Rodrigues de Moura

Identificação de Padrões Multimodais de Comportamento Humano Utilizando Fenotipagem Digital

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Francisco José da Silva e Silva

Orientador

Universidade Federal do Maranhão

Prof. Dr. Luciano Reis Coutinho

Co-orientador

Universidade Federal do Maranhão

Prof. Dr. Fábio Kon

Examinador Externo

Universidade de São Paulo

Prof. Dr. Markus Endler

Examinador Externo

Pontifícia Universidade Católica do Rio de Janeiro

Prof. Dr. Ricardo de Andrade Lira Rabêlo

Examinador Interno

Universidade Federal do Piauí

Prof. Dr. Davi Viana dos Santos

Examinador Interno

Universidade Federal do Maranhão

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequada para obtenção do título de Doutora em Ciência da Computação.

Prof. Dr. Francisco José da Silva e Silva

Orientador

Prof. Dr. Rodrigo de Melo Souza Veras

Coordenador

São Luís - MA, 08 de Dezembro de 2023

Aos meus pais, meus irmãos, meus amigos e meus professores.

Agradecimentos

Ao bom Deus em primeiro lugar, pela minha vida e por todas as maravilhas.

Ao meu orientador, o Prof. Francisco José da Silva e Silva, e coorientador, o Prof. Luciano Reis Coutinho, pela orientação, compreensão e por se fazerem presente em todas as etapas do processo de concepção deste estudo

À FAPEMA, à UFMA e UFPI, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos membros da banca examinadora, por aceitarem a missão de avaliar este trabalho.

Aos meus familiares, que representam minha sustentação, pois forneceram o apoio necessário para me manter firme nesta jornada árdua. Aos meus irmãos Lucas Rodrigues e Valéria Rodrigues pela companhia, apoio, compreensão e por todos os momentos compartilhados na minha vida. A minha mãe Marlene Rodrigues e ao meu pai Raimundo Moura, pois sem eles nada disso seria possível, uma vez que estes são responsáveis pelo que sou atualmente, pois tanto minha índole quanto capacidade profissional são frutos de seus valiosos esforços para me educar.

Aos meus colegas do LSDi e da UFMA, com os quais compartilhei alegrias, tristezas, e conhecimento.

A todos os professores e funcionários do DCCMAPI pelos ensinamentos e pela colaboração.

A todo mundo que de uma forma ou de outra me estimulou ou me ajudou.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão - FAPEMA (BD-01075/20); INCT da Internet do Futuro para Cidades Inteligentes (CNPq 465446/2014-0, CAPES 88887.136422/2017-00, e FAPESP 14/50937-1 e 15/24485-9); e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Código de Financiamento 001).

"O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis"

(José de Alencar)

Resumo

A fenotipagem digital é uma área de pesquisa que propõe a coleta automática de dados de contexto por meio de sensores disponíveis em dispositivos pervasivos, permitindo que técnicas computacionais processem esses dados para detectar automaticamente comportamentos humanos (por exemplo, sociabilidade, atividade física). Essas informações podem subsidiar profissionais especializados no acompanhamento e tratamento da saúde dos indivíduos. Com base nesse cenário, este estudo propõe uma solução capaz de processar inferências comportamentais para reconhecer padrões comportamentais. Esses padrões são projetados com base em atributos de contexto para modelar o comportamento dos indivíduos em situações específicas, como finais de semana e dias úteis. Além disso, a solução proposta reconhece mudanças comportamentais por meio da modelagem do conhecimento do especialista em saúde a partir de conceitos da lógica fuzzy. Os experimentos preliminares identificaram que a estabilidade rotineira dos indivíduos apresenta uma alta correlação positiva com a habilidade da solução em reconhecer padrões comportamentais multimodais capazes de modelar a rotina comportamental. Esta avaliação também reconhece que a solução tem sensibilidade para identificar mudanças comportamentais. Por fim, apresentamos uma análise da influência dos hiperparâmetros da solução na aprendizagem de padrões de comportamento humano enriquecidos com o contexto. A partir desta análise, projetamos diretrizes para auxiliar o processo de parametrização da solução proposta.

Palavras-chave: Padrões comportamentais, Fenotipagem digital, Comportamento humano, Dispositivos ubíquos.

Abstract

Digital phenotyping is a research area that proposes the automatic collection of context data through sensors available in pervasive devices, allowing computational techniques to process this data to automatically detect human behaviors (e.g., sociability, physical activity). This information can support professionals specialized in monitoring and treating the health of individuals. Based on this scenario, this study proposes a solution capable of processing behavioral inference to recognize behavior patterns. These patterns are designed based on context attributes to model individuals' behavior in specific situations, such as weekends and working days. Also, the proposed solution recognizes behavioral changes through knowledge modeling of the health specialist from fuzzy logic concepts. The preliminary experiments identified that the routine stability of individuals presents a high positive correlation with the solution's ability to recognize multimodal behavioral patterns capable of modeling the behavioral routine. This evaluation also recognizes that the proposed solution is sensitive to identifying behavioral changes. Finally, we present an analysis of the influence of the solution's hyperparameters on learning context-enriched human behavior patterns. Based on this analysis, we designed guidelines to support the parameterization process of the proposed solution.

Keywords: Behavioral patterns, Digital phenotyping, Human behavior, Ubiquitous devices.

Lista de ilustrações

Figura 1 – Conjunto fuzzy para avaliar a velocidade de um veículo.	26
Figura 2 – Arquitetura de um sistema fuzzy.	27
Figura 3 – Processo de seleção.	33
Figura 4 – Tópicos de saúde mental visados por estudos selecionados.	34
Figura 5 – Número de estudos que usaram cada tipo de sensor.	35
Figura 6 – Número de soluções propostas que reconheceram cada tipo de situação.	35
Figura 7 – Fluxo de trabalho para realizar sensoriamento multimodal em soluções de fenotipagem digital da saúde mental.	38
Figura 8 – Número de artigos publicados por tipo de estudo.	47
Figura 9 – Representação da quantificação momento a momento do comportamento humano.	54
Figura 10 – Representação dos níveis de dados de contexto	56
Figura 11 – Representação visual do modelo de dados de uma observação comportamental.	58
Figura 12 – Representação visual de observações comportamentais.	59
Figura 13 – Representação visual de slots candidatos a formar padrão comportamental.	60
Figura 14 – Representação visual do padrão comportamental.	61
Figura 15 – Fluxo de trabalho de detecção de comportamento anormal.	63
Figura 16 – Fluxo de trabalho de detecção de mudança de padrão comportamento.	63
Figura 17 – Exemplo de conjuntos difusos para modelar mudanças comportamentais.	64
Figura 18 – Regras fuzzy para mudança de comportamento.	65
Figura 19 – Arquitetura de software do <i>framework</i> proposto.	66
Figura 20 – Diagrama de atividades da solução proposta.	67
Figura 21 – Desenho experimental para avaliar o número de observação apropriado para projetar padrões.	72
Figura 22 – Nível médio de previsão do comportamento.	73
Figura 23 – Desenho experimental para avaliar a estabilidade da rotina comportamental.	73
Figura 24 – Nível médio de estabilidade da rotina comportamental.	74
Figura 25 – Associação entre a capacidade de modelagem apresentada pelos padrões e a estabilidade média da rotina.	74
Figura 26 – Correlação entre a capacidade média de modelagem e a estabilidade média de rotina para cada CA.	75

Figura 27 – Conjuntos e regras fuzzy utilizadas nos experimentos de sensibilidade de mudança.	76
Figura 28 – Mesclagem dos dados de estado estacionário dos usuários u_{08} e u_{44}	76
Figura 29 – Fluxo de detecção de mudança no comportamento de estado estacionário.	77
Figura 30 – Mesclagem dos dados de sociabilidade dos usuários u_{01} e u_{05}	78
Figura 31 – O fluxo de detecção de mudanças nos comportamentos de sociabilidade.	79
Figura 32 – Mesclagem dos dados de movimentação dos usuários u_{00} e u_{31}	80
Figura 33 – O fluxo de detecção de mudanças nos comportamentos de movimentação.	80
Figura 34 – Configuração do experimento para avaliar o tamanho do slot de tempo.	83
Figura 35 – Similaridade média/mediana entre a verdade básica e observações de sociabilidade.	84
Figura 36 – Similaridade média/mediana entre a verdade básica e observações de estado estacionário.	85
Figura 37 – Configuração do experimento para avaliar o valor do parâmetro theta (θ).	85
Figura 38 – Similaridade entre observações comportamentais de sociabilidade e slots candidatos usando slots de 5, 10 e 30 minutos.	86
Figura 39 – Similaridade entre observações comportamentais de estado estacionário e slots candidatos usando slots de 5, 10 e 30 minutos.	87
Figura 40 – Configuração do experimento para avaliar o valor do parâmetro phi (φ).	88
Figura 41 – Similaridade entre slots de candidatos e padrões de sociabilidade	89
Figura 42 – Similaridade entre slots candidatos e padrões de estado estacionário.	89

Lista de tabelas

Tabela 1 – Palavras-chave usadas para criar a string de busca.	30
Tabela 2 – Critérios de seleção.	31
Tabela 3 – Itens usados no processo de extração de dados.	32
Tabela 4 – Descrição de situações, fontes de dados de contexto e abordagens para detectar situações.	36
Tabela 5 – Definição dos tipos de estudos revisados.	39
Tabela 6 – Principais resultados de estudos que visam reconhecer informações de alto nível.	39
Tabela 7 – Principais resultados de estudos que visam classificar estados/transtornos mentais.	40
Tabela 8 – Principais resultados de estudos que visam prever estados/transtornos mentais.	42
Tabela 9 – Principais resultados de estudos que visam associar características a estados/transtornos mentais.	44
Tabela 10 – Conceitos usados no fluxo de trabalho de soluções de fenotipagem digital.	51
Tabela 11 – Estatísticas descritivas de eventos de conversação.	82
Tabela 12 – Estatísticas descritivas de eventos de estado estacionário.	83
Tabela 13 – Estatísticas sobre a similaridade entre observações comportamentais de sociabilidade e slots candidatos.	86
Tabela 14 – Estatísticas sobre a similaridade entre observações comportamentais de estado estacionário e slots candidatos	87
Tabela 15 – Diretrizes para definir valores para os parâmetros da solução proposta.	91
Tabela 16 – Resumo dos estudos revisados.	108

Lista de abreviaturas e siglas

CA	<i>Atributos de contexto</i>
AUC	<i>Área Sob a Curva ROC</i>
BPRS	<i>Escala de Avaliação Psiquiátrica Breve</i>
CDS	<i>Change Detection Service</i>
CQ	<i>Critério de Qualidade</i>
DP	<i>Desvio Padrão</i>
FIS	<i>Sistema de Inferência Fuzzy</i>
FCL	<i>Linguagem de Controle Fuzzy</i>
IoT	<i>Internet das Coisas</i>
ML	<i>Aprendizado de Máquina</i>
MDD	<i>Transtorno Depressivo Maior</i>
OCS	<i>Observation Control Service</i>
PHQ-4	<i>Questionário de Saúde do Paciente-4</i>
RSL	<i>Revisão Sistemática da Literatura</i>
RBF	<i>Função de Base Radial</i>
RMSE	<i>Raiz do Erro Quadrático Médio</i>
QP	<i>Questão de Pesquisa</i>
SJB	<i>Scheduler and Job Management</i>
SVM	<i>Maquina de Vetor de Suporte</i>
TB	<i>Transtorno Bipolar</i>

Sumário

1	INTRODUÇÃO	16
1.1	Questões de Pesquisa	17
1.1.1	(RQ1) Como Identificar Padrões Comportamentais Enriquecidos por Contexto?	17
1.1.2	(RQ2) Como Identificar Mudanças no Padrão Comportamental de Indivíduos?	18
1.2	Objetivos	18
1.3	Metodologia	18
1.4	Contribuições	20
1.5	Estrutura da Tese	21
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Fenotipagem Digital	23
2.2	Computação Positiva	24
2.3	Mineração de Padrões Frequentes	24
2.4	Lógica Fuzzy	25
2.4.1	Sistemas de Inferência Fuzzy	26
3	TRABALHOS RELACIONADOS	28
3.1	Metodologia	28
3.1.1	Objetivo e Questões de Pesquisa	29
3.1.2	Estratégia de Pesquisa	29
3.1.3	Critérios de Seleção	30
3.1.4	Avaliação de Qualidade	31
3.1.5	Extração de Dados	32
3.2	Visão Geral dos Estudos Revisados	32
3.2.1	Seleção de Estudos	32
3.2.2	Tópicos de Saúde Mental Visados	33
3.2.3	Sensores Físicos e Virtuais	33
3.2.4	Situações de Interesse Identificadas	34
3.3	Respondendo as Questões de Pesquisa	35
3.3.1	Fontes de Dados de Contexto Usadas para Detectar cada Tipo de Situação (QP1)	36
3.3.2	Fluxo de Trabalho da Fenotipagem Digital da Saúde Mental Usando Detecção Multimodal (QP2)	37
3.3.3	Apoiando Profissionais de Saúde Mental (QP3)	38

3.3.3.1	Identificação de Informações de Alto Nível	39
3.3.3.2	Classificação de Estados Mentais	40
3.3.3.3	Previsão de Estados Mentais	42
3.3.3.4	Associação entre Características e Estados Mentais	44
3.4	Resultados	46
3.4.1	Tendências	46
3.4.2	Oportunidades de Pesquisas Identificadas	47
3.5	Considerações Finais	48
3.5.1	Publicações	49
4	SOLUÇÃO PROPOSTA	50
4.1	Modelo Formal para Soluções de Fenotipagem Digital	50
4.1.1	Comportamento Humano	50
4.1.2	Eventos Comportamentais	51
4.1.3	Eventos Comportamentais Enriquecidos por Contexto	52
4.1.4	Observações Comportamentais	53
4.1.5	Padrão Comportamental	54
4.2	Modelo de Dados para Fenotipagem Digital	55
4.2.1	Modelo de Dados de Eventos Comportamentais	56
4.2.2	Modelo de Dados de Observações Comportamentais	56
4.2.3	Modelo de Dados de Padrões Comportamentais	58
4.3	Algoritmo para Identificação de Padrões Comportamentais	59
4.3.1	Padrões Comportamentais Sensíveis ao Contexto	61
4.4	Deteção de Mudança Comportamental	62
4.4.1	Comportamentos Anormais	62
4.4.2	Mudanças de Padrão Comportamental	62
4.4.3	Modelagem de Conhecimento Especializado	63
4.5	Framework para a Fenotipagem Digital de Comportamentos	65
4.5.1	Fluxo de Processamento	66
4.6	Considerações Finais	68
4.6.1	Publicações	68
5	AVALIAÇÃO EXPERIMENTAL	70
5.1	Descrição dos Dados	70
5.2	Capacidade de Modelar Rotina Comportamental	71
5.3	Sensibilidade em Mudanças de Padrões Comportamentais	75
5.3.1	Mudanças nos Padrões de Estado Estacionário	75
5.3.2	Mudanças nos Padrões de Sociabilidade	77
5.3.3	Mudanças nos Padrões de Movimentação	79
5.4	Considerações Finais	81

6	ANÁLISE DA INFLUÊNCIA DOS HIPERPARÂMETROS	82
6.1	Características dos Eventos Comportamentais	82
6.2	Análise do Tamanho de Slot de Tempo	83
6.3	Análise do Limiar de Slot de Candidato	84
6.4	Análise de Limiar de Intervalo Frequente	88
6.5	Considerações Finais	90
7	CONCLUSÕES E TRABALHOS FUTUROS	92
7.1	Limitações	94
7.2	Trabalhos Futuros	94
7.3	Publicações	94
	REFERÊNCIAS	96
A	APÊNDICE A - CARACTERIZAÇÃO DOS ESTUDOS SELECIONADOS NA RSL	108

1 Introdução

Computação ubíqua é um termo que se refere à onipresença da computação na vida diária, em que dispositivos fornecem conexão em todos os lugares e de forma transparente para os usuários (Weiser 1993). Esse tema de pesquisa está relacionada a duas áreas: computação pervasiva e computação móvel (Krumm 2018). A computação pervasiva indica que o ambiente de computação é distribuído de forma transparente ao usuário, tornando seu uso quase imperceptível. A computação móvel está relacionada à possibilidade de ter poder computacional em qualquer lugar usando dispositivos móveis que trocam informações por meio de interfaces sem fio. Portanto, a computação ubíqua é composta de muitos dispositivos espalhados de forma transparente na vida diária com estruturas capazes de coletar, processar e compartilhar dados em tempo real.

Atualmente, a grande adoção de dispositivos ubíquos (por exemplo, *smartphones* e tecnologias vestíveis) tem sido explorada para desenvolver novas soluções para detecção automática de comportamentos humano em tempo real (Reinertsen e Clifford 2018, Mohr, Zhang e Schueller 2017). Esses dispositivos possuem sensores capazes de gerar um fluxo crescente de dados de contexto que caracteriza aspectos comportamentais (Mehta et al. 2020). Assim, técnicas computacionais podem processar esse fluxo de dados para inferir informações de alto nível (Liang, Zheng e Zeng 2019, Philip et al. 2021). Por exemplo, a mineração de dados do microfone pode identificar aspectos sociais (por exemplo, conversas) (Eskes et al. 2016), e o processamento de dados do acelerômetro pode reconhecer atividades físicas (por exemplo, correr, caminhar) (Wahle et al. 2016).

A detecção automática de comportamentos humano com base em dados de contexto faz parte dos objetivos da fenotipagem digital (Torous et al. 2016). O termo Fenotipagem Digital refere-se à “quantificação momento a momento do fenótipo humano em nível individual in-situ usando dados de smartphones e outros dispositivos digitais pessoais” (Torous et al. 2016). Essa área de pesquisa objetiva monitorar padrões ao longo do tempo que caracterizam comportamentos de indivíduos (por exemplo, atividades físicas, interações sociais e mobilidade) com base em dados de contexto derivados de dispositivos de computação móvel e vestível (Liang, Zheng e Zeng 2019). Portanto, a fenotipagem digital investiga novos mecanismos computacionais capazes de realizar a detecção contínua e discreta de múltiplos comportamentos humanos que podem apoiar o desenvolvimento de novas ferramentas de apoio a decisão para serviços de saúde.

A fenotipagem digital está sendo testada em várias áreas da saúde para complementar e estender as fontes tradicionais de dados clínicos com dados ecologicamente válidos e longitudinais. Por exemplo, pesquisas tem sido conduzidas para investigar a aplicabilidade da fenotipagem digital para apoiar o monitoramento e tratamento de doenças crônicas não transmissíveis (por exemplo, doença cardiovascular (Teo et al. 2019), obesidade (Kim et al. 2021) e diabetes (Yang et al. 2020)), doenças neurológicas (por exemplo, doença de Parkinson (Bhidayasiri e Mari 2020) e doença de Alzheimer (Kourtis et al. 2019)) e transtornos mentais (por exemplo, depressão (Wang et al. 2018), esquizofrenia (Barnett et al. 2018) e ansiedade social (Gong et al. 2019)). Em especial, grande parte das pesquisas de fenotipagem digital tem se concentrado no desenvolvimento de novas soluções ubíquas para complementar os métodos tradicionais de avaliação e monitoramento da saúde mental através de informações objetivas comportamentais (Liang, Zheng e Zeng 2019), uma vez que estes métodos tradicionais são baseados em autorrelatos subjetivos que estão expostos a vieses cognitivos, como viés de memória e deseabilidade social (Mortel et al. 2008).

1.1 Questões de Pesquisa

Em geral, soluções de fenotipagem digital visam detectar vários aspectos do comportamento humano (Cornet e Holden 2018, Liang, Zheng e Zeng 2019). Essas soluções projetam modelos de aprendizado de máquina com base em características extraídas de situações identificadas para associar, classificar e prever estados de saúde (Moura et al. 2020, Liang, Zheng e Zeng 2019, Barnett et al. 2018). Apesar do potencial de aplicação dessas soluções para detectar e monitorar estados de saúde, ainda existem questões em aberto nesta área de pesquisa. Portanto, apresentamos nas subseções a seguir as questões de pesquisa que pretendemos abordar.

1.1.1 (RQ1) Como Identificar Padrões Comportamentais Enriquecidos por Contexto?

Esta questão de pesquisa busca investigar novas soluções capazes de fornecer informações de interesse para profissionais de saúde, como padrões comportamentais que refletem o hábito de indivíduos. Além disso, objetiva-se analisar a viabilidade de projetar padrões comportamentais baseado em informações contextuais (por exemplo, padrão de sociabilidade nos fins de semana e nos dias de trabalho) para modelar a variação normal de comportamento. Um aspecto importante a ser investigado é o desenvolvimento de um modelo de dados para fenotipagem digital que permita representar esses padrões e acompanhá-los ao longo do tempo.

1.1.2 (RQ2) Como Identificar Mudanças no Padrão Comportamental de Indivíduos?

Atualmente, a maioria das soluções de fenotipagem digital não consideram que o comportamento dos indivíduos possa mudar ao longo do tempo (Liang, Zheng e Zeng 2019). Geralmente os pesquisadores projetam modelos estáticos que aprendem padrões em dados históricos, os quais não sofrem atualizações posteriores. Portanto, nesta questão de pesquisa, investigamos a viabilidade de implementação de abordagens capazes de monitorar de maneira automática comportamentos anormais e mudanças de padrões.

1.2 Objetivos

Desenvolver métodos de monitoramento de comportamento humano através da investigação de uma nova abordagem para a aprendizagem de padrões de comportamentos de indivíduos a partir de fenótipos digitais oriundos de dispositivos de computação ubíqua que considere múltiplos aspectos do comportamento humano.

Os objetivos específicos desta pesquisa são:

- Projetar um modelo de dados para a representação de fenótipos digitais e padrões de comportamentos que considerem múltiplos aspectos do comportamento humano.
- Conduzir uma revisão sistemática da literatura sobre a fenotipagem digital de comportamentos humanos para o monitoramento de saúde.
- Desenvolver e avaliar abordagens para o aprendizado incremental de padrões de comportamentos humanos a partir de fenótipos digitais em contextos estabelecidos (por exemplo, dias de trabalho, finais de semana).
- Desenvolver e avaliar abordagens para a identificação de mudanças de padrão e comportamentos anormais dos indivíduos.

1.3 Metodologia

O ponto de partida desta pesquisa foi a realização de uma revisão sistemática da literatura relativa ao sensoriamento multimodal para o monitoramento de comportamentos humanos relacionados a saúde mental a partir da fenotipagem digital. Para tanto, foi conduzida uma pesquisa bibliográfica abrangente nas bibliotecas digitais *Science Direct*, *PubMed*, *Scopus*, *ACM Digital Library*, *Web of Science*, e *IEEE Xplore*, que resultou em 1865 artigos publicados recuperados. Em seguida, foi realizado um

processo sistemático de seleção de estudos com base em critérios de inclusão e exclusão, que selecionou 59 estudos para a fase de extração de dados. Primeiramente, com base na análise dos dados extraídos dos estudos selecionados, descreveu-se uma visão geral deste campo de pesquisa, apresentando em seguida as características dos estudos selecionados, os principais temas de saúde mental visados, os sensores físicos e virtuais utilizados e os comportamentos identificados. Em seguida, delineou-se respostas às questões de pesquisa desta revisão, descrevendo as fontes de dados de contexto usadas para detectar comportamentos, o fluxo de trabalho usado para detecção multimodal de comportamentos e a aplicação das soluções no processo de avaliação e monitoramento de saúde mental. Além disso, reconheceram-se tendências apresentadas pelos estudos revisados, como o desenvolvimento de soluções para reconhecimento de informações de alto nível, associação de características com estados mentais, classificação de estados mentais e predição de estados mentais.

Através dos resultados da revisão sistemática da literatura conduzida, identificaram-se questões em aberto, como a necessidade de novas soluções capazes de reconhecer padrões e mudanças comportamentais. Por esse motivo, o próximo passo desta pesquisa foi projetar um algoritmo capaz de detectar padrões multimodais de comportamentos humanos a partir de dados de contexto provenientes de dispositivos ubíquos. Explorou-se a abordagem de mineração de padrões frequentes para desenvolver o algoritmo proposto, a qual consiste na ideia de encontrar conjuntos de itens que se repetem em uma quantidade suficiente para representar um padrão. Com base nesta lógica, o algoritmo proposto visa reconhecer a rotina diária de indivíduos (ou seja, intervalos de tempo em que os indivíduos repetem rotineiramente comportamentos). Também explorou-se a possibilidade de utilizar informações contextuais para enriquecer o algoritmo proposto com Atributos de Contexto (CAs). Adicionalmente, foram utilizados os conceitos da lógica fuzzy para detectar mudanças de padrões comportamentais. A motivação para o uso da lógica fuzzy nesta tarefa é a possibilidade de representar o conhecimento humano impreciso e qualitativo por meio de conjuntos fuzzy ao invés de conjuntos nítidos (McNeill e Thro 2014). Outra razão é a notação de regras fuzzy por meio de variáveis linguísticas (ou seja, linguagem natural) e conectores lógicos (Ross 2005), que apresenta uma semântica de fácil compreensão para os usuários.

Adicionalmente, projetou-se um modelo formal para representar os principais conceitos que envolvem a solução proposta, como comportamento humano, evento comportamental, observação comportamental e padrão comportamental enriquecido por contexto. Os conceitos foram definidos de maneira simples, clara e objetiva usando símbolos contidos na teoria de conjuntos para facilitar o entendimento do modelo. Também projetou-se um modelo de dados que descreve a estrutura dos eventos comportamentais e dos padrões comportamentais reconhecidos pela solução proposta.

Em seguida, desenvolveu-se um *framework* que implementa o algoritmo proposto e estrutura os padrões e mudanças comportamentais com base no modelo formal e modelo de dados desenvolvidos. Esse *framework* apresenta interfaces que abstraem a complexidade inerente à solução proposta, facilitando a instanciação de estratégias de detecção de padrões. O processo de desenvolvimento deste componente seguiu os princípios dos chamados Métodos Ágeis de Desenvolvimento de Software que preveem a construção de software de forma incremental, interativa e adaptativa. Sempre que possível, o desenvolvimento incremental do software utilizou técnicas de refatoração.

Realizou-se a avaliação da solução desenvolvida por meio de fluxos de dados derivados de *datasets* com dados reais. A princípio, avaliou-se a solução em termos de sua capacidade de projetar padrões enriquecidos com contexto capazes de modelar rotinas comportamentais, isto é, analisou-se se os padrões gerados apresentam um alto nível de similaridade com rotinas comportamentais estáveis (ou seja, os comportamentos se repetem ao longo do tempo) e menos similaridade com rotinas instáveis. Também analisou-se a sensibilidade da abordagem de detecção de mudança de comportamentos através do processamento de fluxo de dados com mudanças comportamentos anormais e mudanças de padrões controlados de forma explícita (por exemplo, concatenando-se fluxos de dados de indivíduos com comportamentos diferentes), permitindo identificar com precisão o ponto nos quais ocorreram mudanças comportamentais que devem ser detectadas. Posteriormente, foram analisadas as detecções de mudanças comportamentais realizadas pela solução, comparando-as com mudanças comportamentais anotadas no fluxo de dados processado.

Por fim, conduziu-se uma análise sobre os hiperparâmetros da solução proposta com o objetivo de delinear diretrizes capazes de auxiliar na definição de valores adequados. A partir dessa análise, projetamos orientações para definição de valores com base no tipo de comportamento e nos objetivos do monitoramento realizado.

1.4 Contribuições

Esta pesquisa contribui com a literatura de fenotipagem digital de comportamento humano nos seguintes aspectos:

- Fornece uma revisão sistemática abrangente sobre estudos que propõem soluções de fenotipagem digital para reconhecer múltiplas situações de interesse para profissionais de saúde mental. Essa revisão fornece uma visão geral da literatura de fenotipagem digital, responde questões de pesquisa levantadas e reconhece tendências.
- Projeta um modelo formal sobre os principais conceitos utilizados em sistemas de fenotipagem digital. Esse modelo formal é projetado de modo que possa ser

reutilizado e estendido por outras pesquisas.

- Propõe um algoritmo baseado na mineração de padrões frequentes para reconhecer padrões comportamentais, isto é, os intervalos de tempo que um comportamento se repete rotineiramente. Esses padrões são projetados com base em atributos de contexto para modelar a variação normal de comportamento.
- Descreve uma abordagem baseada no uso de lógica fuzzy para reconhecer comportamentos anormais e mudanças de padrões comportamentais.
- Apresenta resultados de uma avaliação experimental da solução proposta utilizando um conjunto de dados coletados de *smartphones* de estudantes.

1.5 Estrutura da Tese

A organização desta tese é a seguinte:

- O Capítulo 2 apresenta a fundamentação teórica, abordando conceitos sobre a aplicação das tecnologias ubíquas no monitoramento da saúde, introduzindo conceitos sobre fenotipagem digital e computação positiva. É apresentado também conceitos sobre as tecnologias utilizadas para desenvolver a solução proposta, como mineração de padrões frequentes e lógica fuzzy.
- O Capítulo 3 apresenta uma revisão sistemática sobre o sensoriamento multimodal de comportamentos humanos para o monitoramento da saúde mental. Neste capítulo apresenta-se uma visão geral da literatura revisada, respondem-se as questões fundamentais sobre essa área de pesquisa e discutem-se os principais resultados da revisão sistemática da literatura conduzida.
- O Capítulo 4 apresenta a solução proposta. Essa apresentação contempla o modelo formal que representa os principais conceitos que envolvem a solução proposta, o modelo de dados da solução, o algoritmo concebido para identificar padrões comportamentais enriquecidos por contexto e a abordagem de detecção de mudanças de padrões e comportamento anormais. Esse capítulo também expõem o *framework* que implementa o algoritmo proposto e estrutura os padrões e mudanças comportamentais.
- O Capítulo 5 expõe a avaliação experimental da solução proposta usando um fluxo de dados do mundo real, no qual se avaliou a capacidade da solução para projetar padrões multimodais enriquecidos com contexto que modelem rotinas comportamentais e a sensibilidade nas detecções de mudanças de padrões comportamentais

- O Capítulo 6 apresenta uma análise da influência dos hiperparâmetros na aprendizagem de padrões comportamentais realizada pela solução proposta.
- O Capítulo 7 apresenta as conclusões obtidas a partir desta pesquisa e apresenta trabalhos futuros que podem ser desenvolvidos a partir deste esforço inicial.

2 Fundamentação Teórica

Este capítulo contém a fundamentação teórica, abordando conceitos sobre a aplicação das tecnologias ubíquas para monitorar a saúde e principais tecnologias utilizadas no desenvolvimento da solução proposta. Portanto, este capítulo apresenta conceitos sobre fenotipagem digital (Torous et al. 2016) e computação positiva (Calvo e Peters 2014), os quais são termos cunhados para designar a aplicação das tecnologias ubíquas ao bem-estar e a saúde. Além disso, também são apresentados os principais conceitos e características da mineração de padrões frequentes e da *lógica fuzzy*. Algumas considerações são apresentadas no final do capítulo.

2.1 Fenotipagem Digital

Os métodos tradicionais de avaliação de comportamentos relacionados a saúde são baseados em evidências clínicas e informações autorrelatadas por pacientes. Essas abordagens são tipicamente baseadas em relatos retrospectivos de eventos experimentados na rotina dos pacientes, nos quais o tempo de lembrança pode ser dias, semanas e até meses. Portanto, esses métodos estão expostos a um conjunto de vieses cognitivos que podem ocorrer na imprecisão da memória relatada (Althubaiti 2016). Por exemplo, os pacientes podem, consciente ou inconscientemente, mudar a verdade do autorrelato para alcançar um resultado desejável (Mortel et al. 2008). Outro motivo que limita a eficácia deste tipo de metodologia é que esta ocorre em ambientes clínicos, onde a terapia é coordenada pelo profissional de saúde. Portanto, não existe validade ecológica nesse método, uma vez que o ambiente clínico é significativamente diferente do contexto natural das pessoas (Trull e Ebner-Priemer 2013). Assim, é necessário criar novos métodos capazes de monitorar automaticamente comportamentos de interesse para profissionais de saúde.

A onipresente adoção dos dispositivos ubíquos, incluindo *smartphones* e *tablets*, pode permitir oportunidades mais eficientes para rastrear comportamentos humanos relacionados a saúde, como a fenotipagem digital, um termo definido por Torous et al. (Torous et al. 2016) que se refere à quantificação momento a momento do fenótipo humano em nível individual in situ usando dados de *smartphones* e outros dispositivos digitais pessoais. A fenotipagem digital é uma área de pesquisa que se concentra na investigação e no desenvolvimento de novas soluções pervasivas para detectar múltiplas situações de interesse para profissionais de saúde (Liang, Zheng e Zeng 2019). Essas situações representam não apenas comportamentos humanos (por exemplo, sociabilidade, atividade física) e hábitos (por exemplo, mobilidade,

sono), mas também outras informações de interesse (por exemplo, contexto ambiental, humor, dicas vocais) para profissionais. As soluções de fenotipagem digital realizam detecção multimodal (ou seja, usando dois ou mais tipos de dados de contexto) de vários comportamentos (Garcia-Ceja et al. 2018), permitindo assim realizar a fusão de dados de baixo nível para derivar informações de alto nível, como inferência de padrões comportamentais e estados mentais. Portanto, essas soluções podem fornecer informações objetivas para apoiar a tomada de decisão por especialistas em saúde.

2.2 Computação Positiva

O termo Computação Positiva foi criado por Calvo e Peters (Calvo e Peters 2014). É uma mudança de paradigma que defende o “design e desenvolvimento de tecnologia para apoiar o bem-estar psicológico e o potencial humano” (Calvo e Peters 2014). A noção de computação positiva surgiu da necessidade de enfrentar os efeitos negativos do ônus do uso de alguns tipos de tecnologia, que incluem, por exemplo, o estresse causado por notificações excessivas e o sentimento de perda de privacidade. Portanto, soluções computacionais positivas são desenvolvidas para promover o bem-estar mental e ajudar a tornar realidade todas as potencialidades humanas, respeitando as necessidades psicológicas dos indivíduos. Além disso, eles são concebidos para melhorar a eficiência e a eficácia dos trabalhadores do conhecimento.

A computação ubíqua e as tecnologias da Internet das Coisas (IoT) contribuíram para promover a computação positiva inteligente (Lee et al. 2019). Alguns benefícios de soluções que usam esse paradigma incluem a facilitação de novas maneiras de detectar mudanças no comportamento humano que podem indicar problemas de bem-estar ou o início de transtornos mentais, fornecer intervenções terapêuticas oportunas e possibilitar o rastreamento de respostas para avaliar a eficácia das intervenções (Lee et al. 2019).

2.3 Mineração de Padrões Frequentes

A mineração de padrões frequentes foi proposta por Agrawal et al. (Agrawal, Imieliński e Swami 1993) para conceber uma abordagem capaz de realizar a análise da cesta de compra de clientes através da mineração de regras de associação. Essa abordagem mapeia os hábitos de compra de clientes, vindo a identificar as associações entre os produtos adicionados às cestas de compra. Logo, através dessa metodologia foi possível identificar a probabilidade de os clientes comprarem certos tipos de produtos em conjunto. Posteriormente, essa solução deu origem a vários algoritmos que seguem essa lógica (Aggarwal, Bhuiyan e Hasan 2014), sendo aplicado em diversos domínios.

Especificamente, a mineração de padrões frequentes objetiva descobrir padrões

que ocorrem com uma frequência significativa em diferentes tipos de coleção de dados, como banco de dados relacionais e não relacionais, arquivos de texto e até fluxos de dados. Nessa abordagem, técnicas computacionais concentram-se em encontrar conjuntos de itens, subestruturas, sequência de itens, dentre outras estruturas que se repetem em uma quantidade suficiente para representar um padrão (Han et al. 2007). Por exemplo, se um conjunto de itens aparecem juntos em transações com quantidade superior a um limiar de suporte, então essa estrutura é um padrão frequente. Esta abordagem não objetiva apenas descobrir estruturas, mas também fornece as características dos dados processados, funcionando como uma ferramenta para muitos tipos de tarefas no domínio de mineração de dados, como tarefas de clusterização, classificação, predição e associação (Bhandari, Gupta e Das 2015, Antonelli et al. 2015, Homayoun et al. 2017).

Atualmente, a mineração de padrões frequentes tem sido explorada para minerar fluxos de dados contínuos, reconhecendo padrões frequentes de forma online e incremental (Nasreen et al. 2014). Nesse cenário, as soluções são capazes de detectar padrões continuamente, reconhecendo em tempo real estruturas frequentes, assim como quando estas estruturas não são consideradas mais padrões. Em geral, os algoritmos que exploram a mineração de padrões frequentes para processar fluxo de dados realizam apenas uma passagem sobre os dados (Lago, Roncancio e Jiménez-Guarín 2019), podendo manter apenas o resumo do fluxo processado ou utilizar janelamento para manter os padrões frequentes associados aos dados mais recentes.

2.4 Lógica Fuzzy

Em alguns domínios de aplicação, sistemas computacionais precisam realizar avaliações e decisões inerentemente humano de maneira automática, isto é, necessitam responder a tarefas de natureza imprecisa. Assim, as repostas dos sistemas não representam conjuntos numéricos precisos, mas sim respostas com julgamento de grau de crença a diversos grupos de respostas possíveis (Silva 2018). Por exemplo, um sistema que calcula rotas entre duas localizações pode necessitar responder se o percurso é curto ou longo, em que dependendo da resposta, o usuário irá percorrê-lo de carro ou a pé. Nesse exemplo, a tarefa de classificar o percurso como curto ou longo possui uma essência subjetiva e imprecisa, pois uma certa distância pode ser considerada perto para um indivíduo, ao passo que para outro é avaliada como longe. Portanto, para mitigar esse desafio, desenvolveu-se a técnica de inteligência computacional denominada *lógica fuzzy* (Zadeh 1988), que permite projetar modelos computacionais capazes de tratar certos níveis de incerteza. Através desta técnica, os sistemas são capazes de aproximar seu modelo de decisão e avaliação ao raciocínio humano, em vez de utilizar valores limítrofes rígidos.

Na *lógica fuzzy*, uma premissa pode implicar em várias conclusões com diferentes níveis de pertinência, isto é, um elemento pode estar contido parcialmente a um conjunto. Assim, o nível de pertinência de um elemento a um conjunto é definido por uma função $\mu(x)$, que resulta em valores contidos no intervalo $[0,1]$. Portanto, o nível de inclusão do elemento pode variar entre não pertinente (ou seja, $\mu(x) = 0$) até totalmente pertinente (ou seja, $\mu(x) = 1$). Por exemplo, para decidir se a velocidade de um veículo é baixa, média ou alta, um especialista poderia modelar esse problema utilizando conjuntos *fuzzy* conforme ilustrado na Figura 1. Nesse cenário, é possível observar que existe um certo nível de interseção entre os conjuntos, representando uma transição gradativa de pertinência. Por exemplo, o conjunto *Média* inicia com velocidade a partir de 10km (ou seja, não incluso, pois $\mu(10) = 0$), aumentando de maneira gradual até atingir 40km (ou seja, totalmente incluso, pois $\mu(40) = 1$), e depois o nível de pertinência decresce gradualmente conforme transita para o conjunto *Alta*.

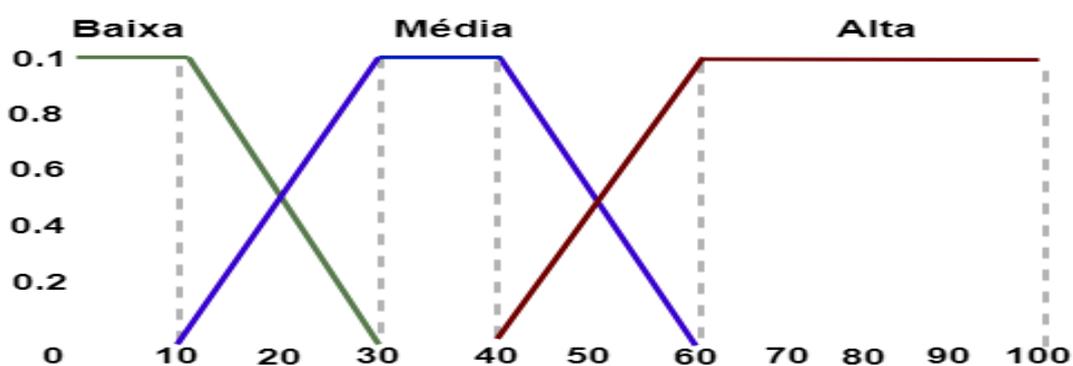


Figura 1 – Conjunto fuzzy para avaliar a velocidade de um veículo.

2.4.1 Sistemas de Inferência Fuzzy

Sistemas de Inferência Fuzzy (FIS) são modelos computacionais que objetivam representar incertezas e imprecisões por meio de regras, em vez de utilizar representações numéricas exatas (McNeill e Thro 2014). Em FIS, o conhecimento do especialista é representado em forma de implicações lógicas, que possuem a estrutura básica se (condição) então (consequência). Os especialistas também podem usar os conectores lógicos *AND* e *OR* para representar conhecimentos mais complexos, validando assim estruturas como a seguinte: <condição> *AND* <condição> *THEN* <consequência>. Por exemplo, no cenário em que o sistema necessita decidir se o usuário deverá fazer um percurso a pé ou de carro com base nas variáveis distância e condição climática, o especialista poderia modelar a seguinte preposição: se (distância é longa) e (condição climática é ruim) então (percurso de carro). Através dessa preposição, o FIS analisará as variáveis de entrada, que indicará percurso de carro nos casos de distâncias longas e climas ruins.

É importante destacar, que diferentemente da lógica clássica que define suas proposições como totalmente verdadeiras ou totalmente falsas, nos FIS as preposições assumem níveis de verdades, implicando em variáveis de saídas com diferentes graus de crenças. Portanto, quanto mais pertinentes as variáveis de entrada (por exemplo, distância e condição climática), maior será o grau de inclusão da variável de saída (por exemplo, percurso a pé ou de carro), viabilizando assim representar as incertezas inerentes a tarefa proposta. A Figura 2 apresenta a arquitetura geral de um FIS (McNeill e Thro 2014), composto pelos seguintes componentes:

- **Processador de entrada (fuzzyfier):** Derivam dados nebulosos a partir da entrada de dados precisos.
- **Base de conhecimento:** contém as variáveis, os conjuntos fuzzy e as regras definidas pelos especialistas.
- **Motor fuzzy:** processa as variáveis de entrada nebulosas com base nas regras definidas pelo especialista, vindo a realizar uma inferência de saída.
- **Processador de saída (defuzzyfier):** deriva um dado preciso a partir da saída do motor *fuzzy*.

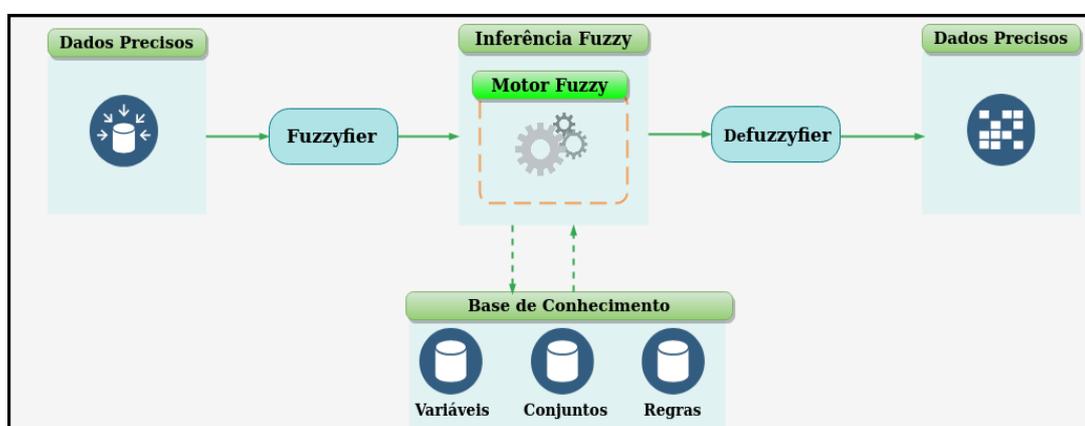


Figura 2 – Arquitetura de um sistema fuzzy.

Portanto, essa maneira de modelar a subjetividade humana é aplicável a várias áreas que necessitam representar o conhecimento impreciso, como sistemas especialistas e sistemas de apoio à decisão.

3 Trabalhos Relacionados

Muitos estudos usaram a fenotipagem digital para complementar os métodos tradicionais de avaliação e monitoramento da saúde mental. Assim, a literatura atual sobre fenotipagem digital da saúde mental contém muitos trabalhos que propõem soluções para monitorar múltiplos aspectos do comportamento humano, resultando em uma literatura abrangente (Liang, Zheng e Zeng 2019, Teles et al. 2020). Portanto, realizou-se uma Revisão Sistemática da Literatura (RSL) para reconhecer o estado da arte sobre o sensoriamento multimodal de situações de interesse para profissionais de saúde mental (Moura et al. 2022). Revisou-se estudos que propuseram soluções de fenotipagem digital da saúde mental capazes de realizar fusão de dados de contexto para identificar múltiplas situações de interesse, como sociabilidade, sono, pistas vocais, atividade física, humor, mobilidade e contexto ambiental. Neste capítulo, apresenta-se inicialmente a metodologia utilizada para a condução da presente revisão. Em seguida, apresenta-se uma visão geral dos estudos revisados, responde-se as questões fundamentais sobre este tema de pesquisa e discute-se os principais resultados da RSL conduzida. Ao final, apresenta-se as discussões sobre o capítulo.

3.1 Metodologia

RSL é um método que fornece meios para identificar, avaliar e interpretar o estado da arte sobre uma questão de pesquisa de interesse (Kitchenham e Charters 2007). Esses métodos sistemáticos utilizados para selecionar, avaliar, coletar e analisar pesquisas relevantes precisam seguir procedimentos validados presentes na literatura para aumentar a integridade da revisão realizada e permitir que outros pesquisadores reproduzam o método adotado. Assim, este estudo foi realizado com base nas diretrizes de revisão sistemática da literatura de engenharia de software proposta por Kitchenham e Charters (Kitchenham 2004, Kitchenham e Charters 2007). Essa RSL foi realizada de acordo com três fases principais: planejamento, execução e análise de resultados. A realização dessas fases foi gerenciada por meio do *Parsifal*¹, uma ferramenta de suporte online para executar RSLs no contexto da Engenharia de Software, que documenta os processos executados e fornece suporte aos pesquisadores por meio de um espaço de trabalho compartilhado. Nesta seção, apresentamos o protocolo de revisão adotado, que definiu objetivos, questões de pesquisa, estratégia de pesquisa, critérios de seleção e extração de dados.

¹ <<https://parsif.al/>>

3.1.1 Objetivo e Questões de Pesquisa

Esta RSL visa reconhecer o estado da arte sobre fenotipagem digital usando detecção multimodal de situações de interesse para profissionais de saúde mental. Especificamente, pretendemos revisar estudos que propuseram soluções de fenotipagem digital da saúde mental capazes de realizar fusão de dados de contexto para identificar múltiplas situações. A seguir, apresentamos as Questões de Pesquisa (QPs) e sua justificativa.

- (QP1) Quais são as fontes de dados de contexto usadas para detectar cada tipo de situação?
Justificativa: o objetivo desta questão é entender as fontes de dados de contexto exploradas para detecção automática de cada tipo de situação.
- (QP2) Qual é o fluxo de trabalho da fenotipagem digital da saúde mental usado para detecção multimodal de situações de interesse?
Justificativa: está focado em conhecer o fluxo de trabalho para realizar a fusão de dados de contexto multimodal para identificar várias situações de interesse adequadas para apoiar profissionais de saúde mental em seus tratamentos e terapias médicas.
- (QP3) Como as situações são usadas para apoiar os profissionais de saúde mental?
Justificativa: esta questão visa clarificar a aplicação das soluções identificadas no processo de avaliação e acompanhamento da saúde mental, compreendendo as principais tarefas de apoio à decisão.

3.1.2 Estratégia de Pesquisa

Uma pesquisa exaustiva dos estudos primários foi realizada em 29 de outubro de 2022, nas seguintes bibliotecas digitais: ACM Digital Library, IEEE Xplore, Web of Science, PubMed, ScienceDirect e Scopus. A busca teve como objetivo encontrar estudos que propuseram métodos de sensoriamento multimodal de múltiplas situações de interesse por meio de dispositivos ubíquos, cujo objetivo principal é apoiar profissionais de saúde mental. Para recuperar esses estudos, criamos uma string de busca com base em dois termos-chave, que possuem um conjunto de palavras-chave e sinônimos. A Tabela 1 mostra os termos selecionados. O primeiro termo (A) refere-se à saúde mental, enquanto o segundo termo (B) contém palavras relacionadas ao sensoriamento multimodal.

Projetamos uma string de busca com base nas palavras-chave apresentadas na Tabela 1. Também realizamos uma busca piloto na biblioteca digital Scopus para

Tabela 1 – Palavras-chave usadas para criar a string de busca.

Termo	Palavras-chave	Sinônimos
A	"Mental Health"	"Schizophrenia", "Depression", "Stress", "Anxiety", "Bipolar disorder", "Mental Disorder", "Mental State", "Psychological Health", "Psychiatry", "Clinical psychology"
B	"Multimodal Sensing"	"Smartphone Sensing", "Mobile Sensing", "Passive Sensing", "Passive Detection", "Digital Phenotyping", "Behavioral Monitoring", "Objective Sensor Data"

validação da string de busca, que consistiu na utilização dos estudos realizados por (Wang et al. 2017) e (Boukhechba et al. 2018) como artigos de controle. A string definida recuperou os artigos de controle, demonstrando sua capacidade de encontrar estudos adequados para esta RSL. Por fim, aplicamos a string de busca aos mecanismos de busca das demais bases de dados digitais para recuperar os estudos. Abaixo, apresentamos a estrutura da string de busca projetada.

<<(("Schizophrenia"OR "Depression"OR "Stress"OR "Anxiety"OR "Bipolar disorder"OR "Mental Health" OR "Mental Disorder" OR "Mental State" OR "Psychological Health" OR "Psychiatry"OR "Clinical psychology") AND ("Multimodal Sensing", "Smartphone sensing"OR "Mobile Sensing"OR "Passive Sensing"OR "Passive Detection"OR "Digital Phenotyp*"OR "Behavioral Monitoring"OR "Objective Sensor Sata"))>>

3.1.3 Critérios de Seleção

Realizamos o processo de seleção dos estudos com base nos critérios de inclusão e exclusão apresentados na Tabela 2. Especificamente, dois pesquisadores usaram esses critérios para realizar o processo de seleção de estudos de forma independente. Primeiro, os pesquisadores executaram o processo de triagem dos estudos com base no título, resumo e palavras-chave. Em um segundo momento, os pesquisadores leram e analisaram o texto completo dos estudos selecionados para identificar aqueles adequados ao escopo desta revisão. Em seguida, avaliamos a consistência do processo de seleção aplicando o coeficiente Kappa de Cohen (Cohen 1968), que mediu o nível de concordância entre as análises dos pesquisadores. Ao final, os pesquisadores conduziram discussões para resolver os conflitos de seleção e, nos casos em que não houve consenso, os juízes (isto é, outros dois pesquisadores) deliberaram sobre os conflitos.

Tabela 2 – Critérios de seleção.

Critério de inclusão	Critério de exclusão
(IC1) Estudos que utilizaram dados de contexto de sensores embutidos ou conectados em dispositivos ubíquos.	(EC1) Estudos que não utilizaram dados de sensores ubíquos para identificar situações de interesse.
(IC2) Estudos que realizaram sensoriamento multimodal de múltiplas situações de interesse.	(EC2) Artigos que não apresentam soluções para realizar o sensoriamento multimodal de múltiplas situações de interesse.
(IC3) Propostas de soluções que utilizaram sensoriamento multimodal para apoiar profissionais de saúde mental.	(EC3) Artigos que possuem outras publicações com uma versão mais atual e completa da solução proposta.
(IC4) Artigos completos.	(EC4) Literatura cinza: materiais como resenhas, relatórios, cartazes, relatórios breves, livros, comunicações breves, teses e dissertações.
(IC5) Artigos em língua inglesa.	(EC5) Artigos em um idioma diferente do inglês.

3.1.4 Avaliação de Qualidade

Em RSLs, o processo de avaliação de qualidade analisa se os estudos primários apresentam estratégias para mitigar ameaças à validade (Kitchenham 2004). Para isso, utilizamos o instrumento de avaliação de qualidade proposto por (Dybå e Dingsøyr 2008), que analisa a qualidade dos estudos selecionados em termos de rigor, credibilidade e pertinência dos resultados. Avaliamos os estudos selecionados de acordo com os critérios de qualidade, em que cada um dos estudos foi classificado em uma escala tricotômica (isto é, sim, parcialmente ou não) e pontuado com 1, 0,5 e 0, respectivamente. A seguir, apresentamos os Critérios de Qualidade (CQ) utilizados.

- (QC1) Is the paper based on research (or is it merely a "lessons learned" report based on expert opinion)?
- (QC2) Is there a clear statement of the aims of the research?
- (QC3) Is there an adequate description of the context in which the research was carried out?
- (QC4) Was the research design appropriate to address the aims of the research?
- (QC5) Was the recruitment strategy appropriate to the aims of the research?
- (QC6) Was there a control group with which to compare treatments?
- (QC7) Was the data collected in a way that addressed the research issue?
- (QC8) Was the data analysis sufficiently rigorous?
- (QC9) Has the relationship between researcher and participants been considered to an adequate degree?
- (QC10) Is there a clear statement of findings?
- (QC11) Is the study of value for research or practice?

3.1.5 Extração de Dados

Nesta etapa, foi realizada a leitura do texto completo dos artigos selecionados para extrair informações relevantes que permitissem sintetizar e caracterizar os estudos, responder às questões de pesquisa, identificar questões em aberto e traçar perspectivas futuras de pesquisa. Para isso, projetamos o formulário de extração de dados apresentado na Tabela 3.

Tabela 3 – Itens usados no processo de extração de dados.

ID	Questões
I1	Referência (autores e ano)
I2	Objetivo e principais resultados
I3	Estado mental investigado
I4	Dados de contexto coletados
I5	Situações identificadas
I6	Método, técnica ou processo utilizado para identificar situações
I7	Uso do sensoriamento multimodal no monitoramento da saúde mental
I8	Duração do estudo e quantidade de participantes
I9	Limitações do Estudo

3.2 Visão Geral dos Estudos Revisados

Os estudos revisados fornecem uma grande quantidade de informações relevantes para este tópico de pesquisa. Nesta seção, apresentamos informações sobre o processo de seleção de estudo, tópicos de saúde mental visados pelos trabalhos, sensores físicos e virtuais utilizados, e situações de interesse identificadas. Adicionalmente, resumimos as principais informações sobre os estudos revisados no Anexo A.

3.2.1 Seleção de Estudos

A Figura 3 apresenta o diagrama do processo de seleção. Primeiramente, foram identificados 1.865 artigos a partir da busca realizada nas seis bibliotecas digitais. Destes, 873 duplicados foram removidos. Em seguida, foi realizada a etapa de triagem, resultando em 63 estudos selecionados para a próxima etapa. Por fim, os pesquisadores executaram a etapa de elegibilidade, que consistiu na leitura dos textos completos para reconhecer estudos relevantes para esta revisão. Ao final dessa etapa, utilizamos o Kappa de Cohen para medir a concordância entre as seleções realizadas pelos pesquisadores. Esse procedimento apresentou um alto índice de concordância entre os pesquisadores: $\approx 0,85$, o que é considerado uma concordância quase perfeita (Viera e Garrett 2005). Por fim, os pesquisadores resolveram os conflitos por meio de discussões, resultando em 59 estudos selecionados.

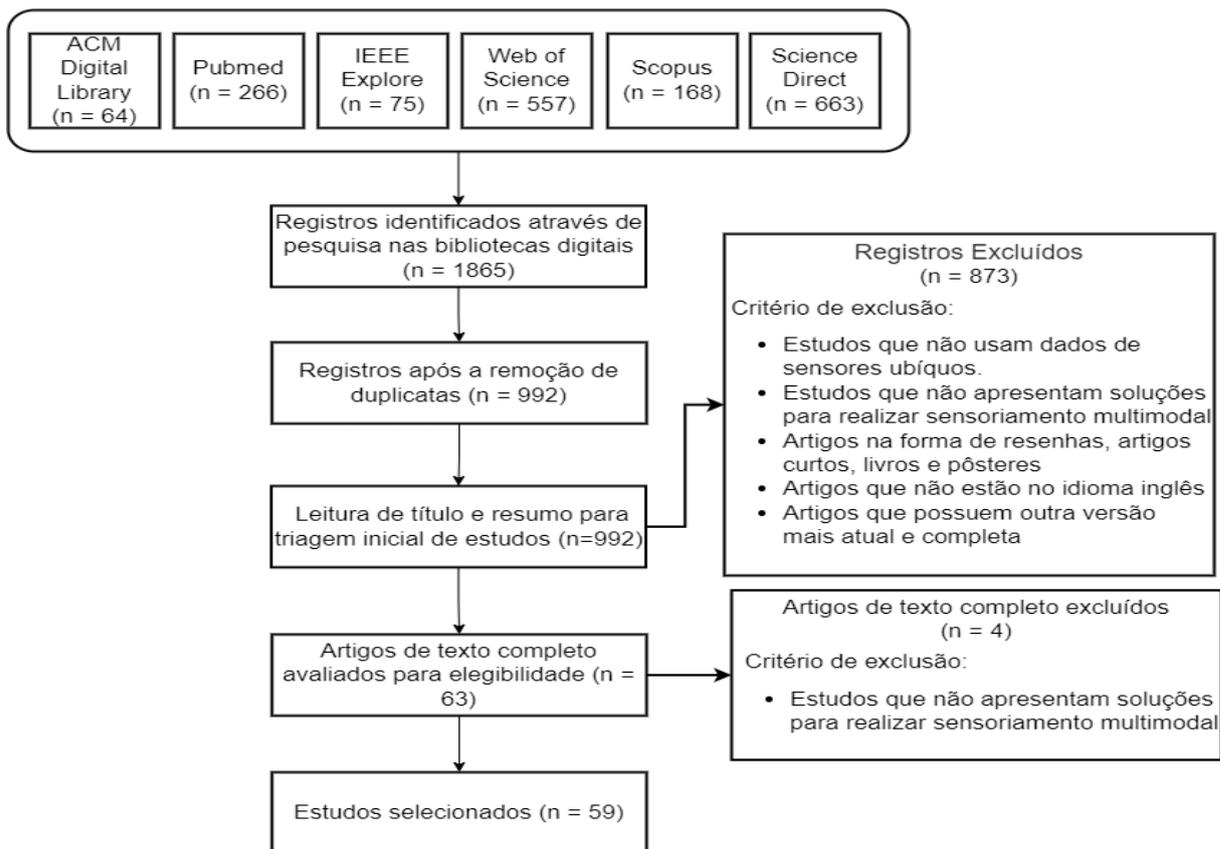


Figura 3 – Processo de seleção.

3.2.2 Tópicos de Saúde Mental Visados

Estudos propuseram soluções para monitorar diferentes estados/distúrbios mentais usando sensoriamento multimodal. A Figura 4 mostra os tópicos de saúde mental visados pelos estudos selecionados. Reconhecemos que a depressão ($n = 24$) e a esquizofrenia ($n = 10$) representam os principais transtornos de saúde mental investigados. Em seguida, identificamos que os estudos também exploraram temas como ansiedade ($n = 9$), bem-estar ($n = 7$), estresse ($n = 6$), humor ($n = 5$) e ansiedade social ($n = 5$). O item 'Outros' (Figura 4) agrega o número de estudos que abordaram os seguintes temas: estados emocionais ($n = 3$), solidão ($n = 3$), transtorno bipolar ($n = 3$), ideação suicida ($n = 2$), estado afetivo ($n = 1$) e comportamento impulsivo ($n = 1$).

3.2.3 Sensores Físicos e Virtuais

As soluções propostas pelos estudos selecionados usaram dados de contexto de sensores físicos e virtuais embutidos em dispositivos móveis e vestíveis. Os sensores virtuais são responsáveis por fornecer dados de componentes de software (por exemplo, chamadas telefônicas, SMS), enquanto os sensores físicos fornecem dados de componentes de hardware (por exemplo, Bluetooth, GPS). A Figura 5 mostra

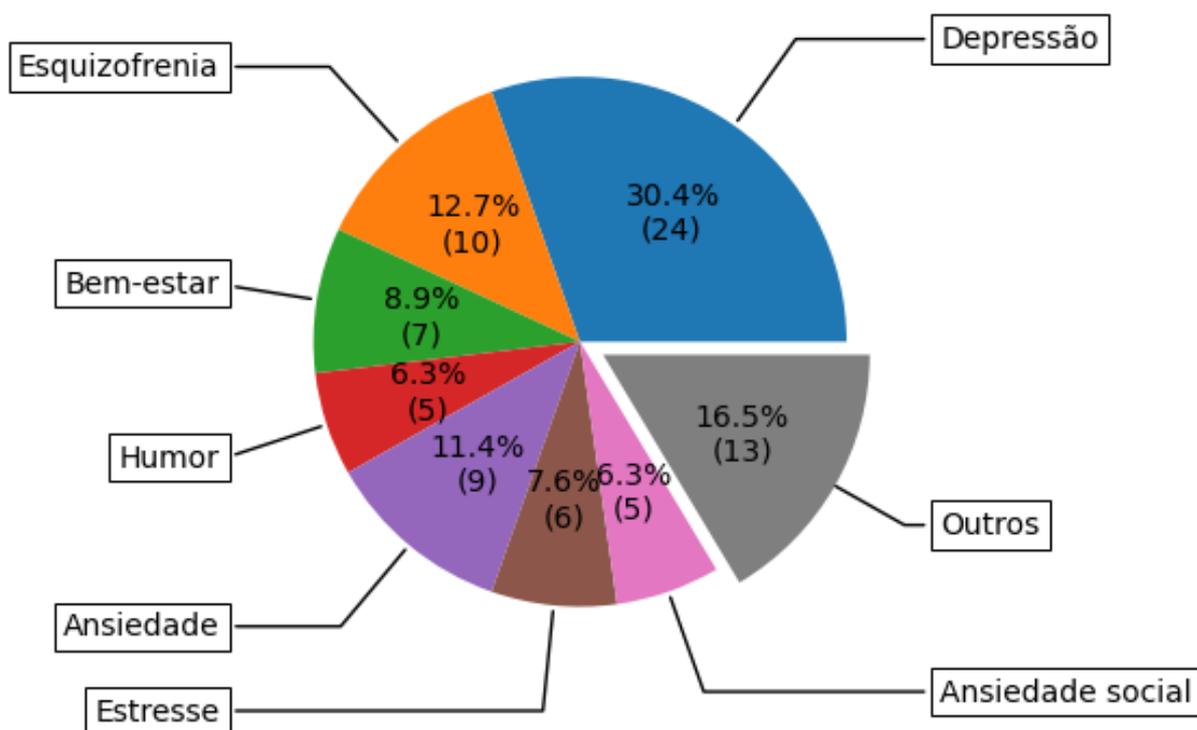


Figura 4 – Tópicos de saúde mental visados por estudos selecionados.

os sensores usados pelos estudos revisados. Reconhecemos que o acelerômetro ($n = 49$) e o GPS ($n = 44$) são os sensores mais explorados, seguidos pelos registros de chamadas ($n = 31$), microfone ($n = 30$), estado da tela ($n = 29$), sensor de luminosidade ($n = 23$) e SMS ($n = 20$). Finalmente, os estudos usaram com menos frequência os seguintes sensores: WiFi ($n = 10$), uso de aplicativo ($n = 8$), giroscópio ($n = 8$), frequência cardíaca ($n = 7$), magnetômetro ($n = 6$), Bluetooth ($n = 4$), mídia social ($n = 2$), pedômetro ($n = 1$) e temperatura da pele ($n = 1$).

3.2.4 Situações de Interesse Identificadas

As soluções revisadas usaram sensoriamento multimodal para detectar diferentes situações de interesse. A Figura 6 mostra o número de estudos que visaram reconhecer cada tipo de situação. Identificamos que atividade física ($n = 47$), mobilidade ($n = 45$) e sociabilidade ($n = 42$) foram as situações mais exploradas pelos estudos. Além disso, notamos que os pesquisadores também estavam focados em investigar o sono ($n = 26$), os padrões de uso do dispositivo ($n = 26$) e o contexto ambiental ($n = 17$). Finalmente, as soluções exploraram com menos frequência humor ($n = 3$), pistas vocais ($n = 3$) e movimento corporal ($n = 1$).

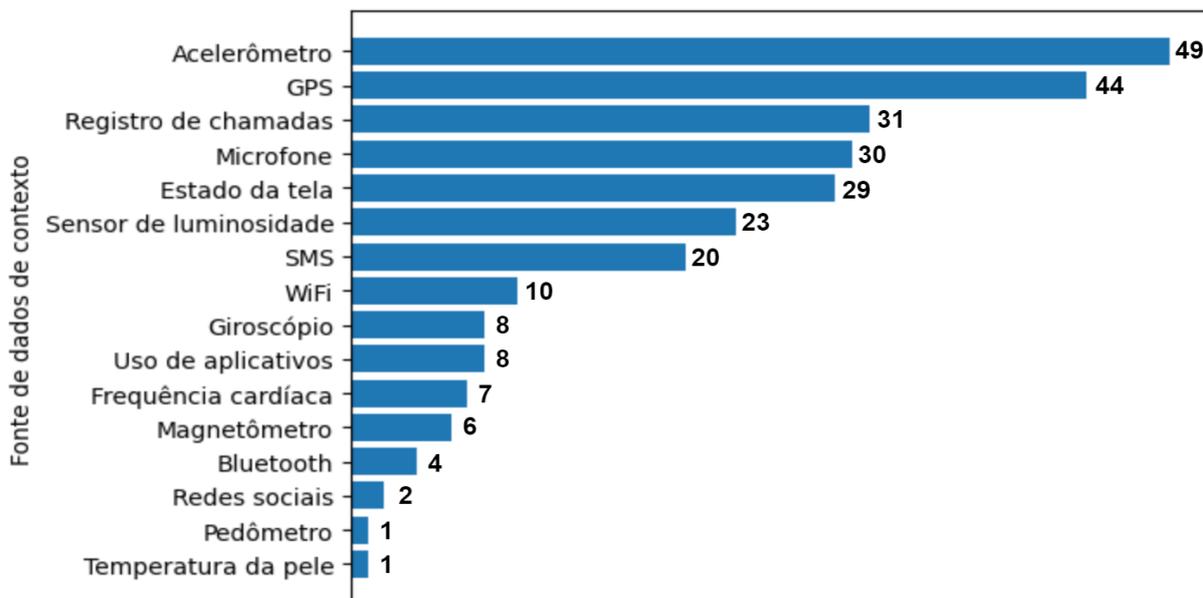


Figura 5 – Número de estudos que usaram cada tipo de sensor.

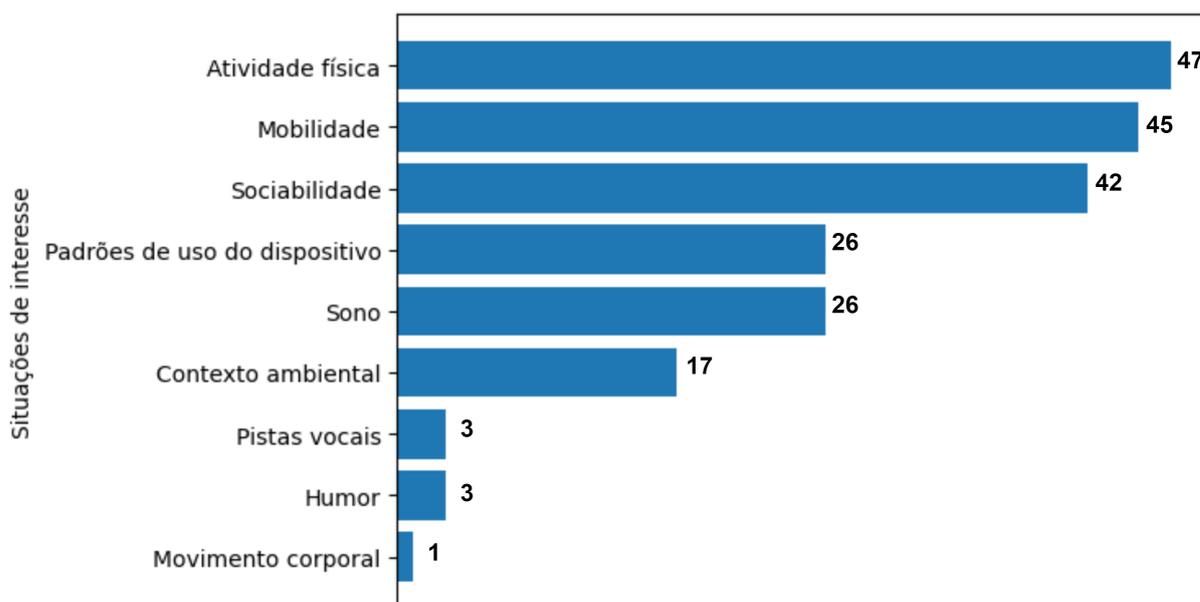


Figura 6 – Número de soluções propostas que reconheceram cada tipo de situação.

3.3 Respondendo as Questões de Pesquisa

Esta seção explora as informações extraídas dos artigos selecionados para delinear as discussões sobre as questões de pesquisa.

3.3.1 Fontes de Dados de Contexto Usadas para Detectar cada Tipo de Situação (QP1)

As soluções de fenotipagem digital da saúde mental detectam automaticamente diferentes situações de interesse usando detecção multimodal. Estudos selecionados usam informações identificadas passivamente para apoiar o monitoramento e avaliação da saúde mental. A Tabela 4 apresenta informações relevantes sobre as fontes de dados de contexto usadas para reconhecer cada situação de interesse. Além disso, esta tabela apresenta as abordagens utilizadas pelos estudos revisados para reconhecer situações de interesse.

Tabela 4 – Descrição de situações, fontes de dados de contexto e abordagens para detectar situações.

Situação	Fonte de dados de contexto	Abordagem de detecção da situação de interesse
Sociabilidade	Registro de chamadas, SMS, rede social, microfone e bluetooth	As interações sociais virtuais são reconhecidas pelo processamento de dados de redes sociais online e aplicativos de comunicação (por exemplo, registro de chamadas telefônicas e sms) (Dogruclu et al. 2020, Sano et al. 2018). Além disso, as interações sociais físicas são identificadas usando abordagens para reconhecer conversações por meio de dados de microfone (Harari et al. 2017, Farhan et al. 2016) e co-localizações de indivíduos por meio do escaneamento de dispositivos com interfaces <i>Bluetooth</i> (Wang et al. 2017, Doryab et al. 2019).
Atividade física	Acelerômetro, giroscópio e magnetômetro	Os pesquisadores treinam algoritmos de aprendizado de máquina usando dados de sensores inerciais para classificar diferentes atividades físicas (por exemplo, parado, caminhando, correndo) (Wang et al. 2018, Adler et al. 2020, Levine et al. 2020).
Mobilidade	GPS, WiFi, acelerômetro, giroscópio e magnetômetro	Soluções de fenotipagem digital processam amostras de GPS e dados de sensores inerciais para projetar padrões de mobilidade de indivíduos (por exemplo, distância total percorrida, o tempo gasto em pontos de interesse e a duração da viagem) (Wang et al. 2017, Farhan et al. 2016, Wahle et al. 2016). Além disso, eles identificam a semântica dos locais visitados usando dados de <i>OpenStreetMap</i> e <i>Foursquare APIs</i> (Cai et al. 2018, Boukhechba et al. 2018).

Sono	Microfone, sensor de luminosidade, acelerômetro, estado da tela	O monitoramento da qualidade do sono é realizada a partir da combinação de características extraídas de dados como luminosidade, movimento, estado da tela e som ambiente (Lane et al. 2014, Narziev et al. 2020, Demasi, Aguilera e Recht 2016, Sarda et al. 2019). Esta abordagem multimodal permite inferir informações ambientais e eventos de sono (por exemplo, hora de dormir e duração do sono) a partir dos padrões de uso do dispositivo.
Pistas vocais	Microfone	Pistas vocais representam características paralinguísticas que moldam diferentes aspectos da fala humana, como entonação, ritmo, energia do sinal, intensidade da fala e velocidade (Place et al. 2017, Dogrucu et al. 2020). As soluções de fenotipagem digital processam principalmente dados provenientes de microfone para identificar pistas vocais na voz de indivíduos durante conversações.
Movimento corporal	Acelerômetro, giroscópio e magnetômetro	As soluções de fenotipagem digital usam técnicas computacionais para processar dados de sensores inerciais para identificar movimentos corporais involuntários de indivíduos durante situações estressantes.
Humor	Frequência cardíaca e outras situações	Em geral, as soluções de fenotipagem digital realizam a fusão de dados para identificar esta situação de interesse (Narziev et al. 2020), visto que é difícil identificar humor a partir de uma única fonte de dados. Assim, a detecção de humor de indivíduos é realizada principalmente combinando características de outras situações (por exemplo, atividade física, sono, sociabilidade) com características adicionais coletada do sensor de frequência cardíaca de dispositivos vestíveis (Spathis et al. 2019, Wang et al. 2018).
Padrões de uso do dispositivo	Estado da tela e uso de aplicativo	Os sistemas de fenotipagem digital inferem essa situação a partir de registros de aplicativos móveis e do estado da tela (isto é, tela ligada/desligada). Essas informações são usadas para reconhecer os padrões de uso do dispositivo, como duração e frequência de uso (Levine et al. 2020, Sano et al. 2018).
Contexto ambiental	Microfone e sensor de luminosidade	Esta situação é derivada principalmente de sensores de luminosidade e dados de microfone. Especificamente, os pesquisadores usam estatísticas de resumo para medir o som ambiente e a amplitude da luz (Gu et al. 2017, Wang et al. 2017, Servia-Rodríguez et al. 2017).

3.3.2 Fluxo de Trabalho da Fenotipagem Digital da Saúde Mental Usando Detecção Multimodal (QP2)

O fluxo de trabalho fenotipagem digital da saúde mental usando sensoriamento multimodal apresenta etapas que visam transformar dados contextuais em informações úteis para profissionais de saúde mental. A Figura 7 mostra uma visão geral deste

fluxo de trabalho. A primeira etapa é responsável por coletar dados de contexto de sensores embarcados ou conectados a dispositivos pervasivos. A segunda etapa é responsável por pré-processar dados de contexto, executando tarefas como limpeza, transformação e enriquecimento de dados. A terceira etapa visa extrair recursos dos dados pré-processados, como contagens, variância e entropia. As soluções propostas identificam situações de interesse na quarta etapa, como sociabilidade, mobilidade e sono. Na quinta etapa, os pesquisadores se concentram em projetar características para descrever as situações identificadas. Finalmente, na sexta etapa, usa-se as características projetadas para desenvolver modelos computacionais capazes de apoiar o processo de avaliação e monitoramento da saúde mental, realizando tarefas como classificação de estados mentais e predição de estados mentais futuros.

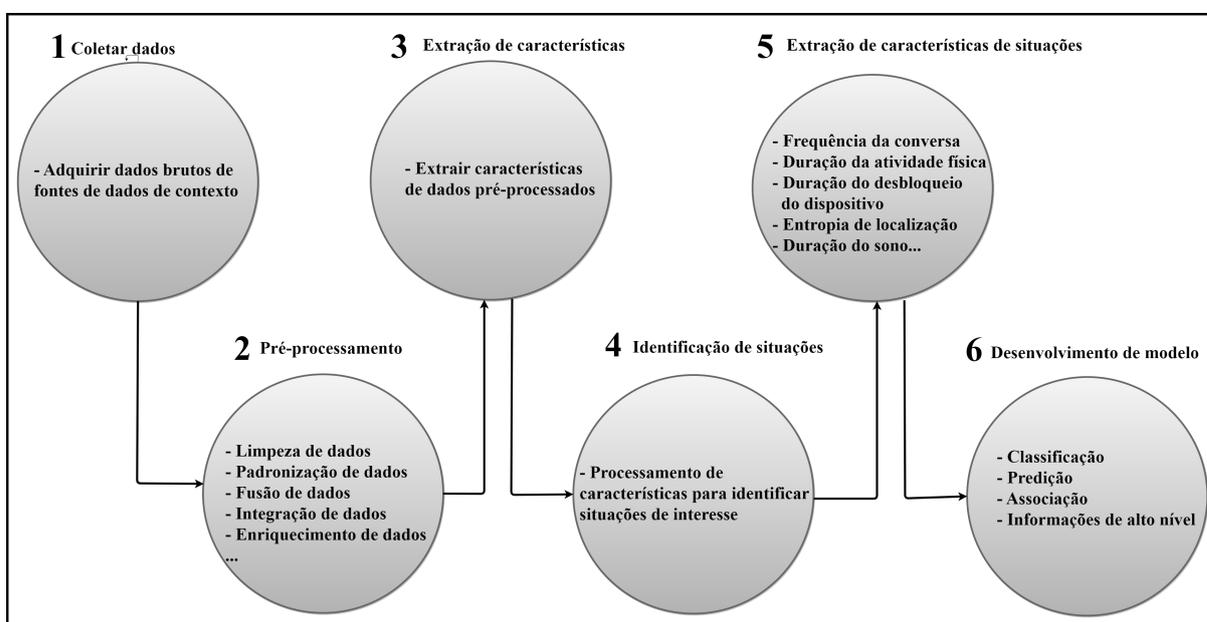


Figura 7 – Fluxo de trabalho para realizar sensoriamento multimodal em soluções de fenotipagem digital da saúde mental.

3.3.3 Apoiando Profissionais de Saúde Mental (QP3)

A partir da literatura revisada, identificamos que as soluções de fenotipagem digital realizam detecção multimodal de múltiplas situações de interesse dos profissionais, fornecendo informações objetivas sobre a saúde mental dos indivíduos. Embora os estudos revisados tenham esse objetivo em comum, eles diferem no tipo de solução projetada, que apoia o monitoramento e a avaliação da saúde mental de maneiras diferentes. Portanto, classificamos a literatura revisada em quatro categorias distintas: (i) reconhecimento de informações de alto nível; (ii) classificação de estados/transtornos mentais; (iii) predição de estados/transtornos mentais futuros; e (iv) associação de características com estados/transtornos mentais. A Tabela 5 apresenta

a definição de cada tipo de estudo revisado. No restante desta seção, descrevemos esses quatro tipos diferentes de estudos.

Tabela 5 – Definição dos tipos de estudos revisados.

Tipo de estudo	Definição
Reconhecimento de informações de alto nível	Nesse tipo de estudo, o objetivo é processar fenótipos digitais para reconhecer informações sobre aspectos comportamentais dos indivíduos, como padrões, mudanças comportamentais e quantificação comportamental (por exemplo, duração e frequência do comportamento).
Classificação de estados/transtornos mentais	Esse tipo de trabalho se concentra no treinamento de algoritmos de aprendizado de máquina para classificar estados/transtornos mentais.
Predição de estados/transtornos mentais	Trabalhos com foco em predição visam projetar modelos de aprendizado de máquina usando características comportamentais para prever estados/transtornos mentais futuros.
Associação de características com estados/transtornos mentais	Trabalhos que centram-se na análise da associação existente entre estados/transtornos mentais e características extraídas das situações identificadas.

3.3.3.1 Identificação de Informações de Alto Nível

A Tabela 6 apresenta os principais resultados de estudos que visam reconhecer informações de alto nível, ordenados por ano de publicação. Eles se concentram em reconhecer padrões relacionados a comportamentos humanos. Alguns sistemas quantificam comportamentos humanos, gerando pontuações com base na duração, intensidade e frequência de cada comportamento identificado (Lane et al. 2014). Os trabalhos também visam reconhecer mudanças comportamentais que podem ser indicadores de mudanças de estados mentais (Jagesar et al. 2021, Huckins et al. 2020, Barnett et al. 2018).

Tabela 6 – Principais resultados de estudos que visam reconhecer informações de alto nível.

Referência	Principais resultados
(Zhou et al. 2022)	Este estudo usa o modelo de mistura gaussiana e a partição em torno dos medoides para reconhecer diferentes padrões comportamentais (como clusters representando dias sedentários e dias ativos com baixa comunicação) em pacientes com esquizofrenia.
(Moura et al. 2022)	Este artigo propõe uma solução que combina mineração de padrões frequentes com processamento de eventos complexos para reconhecer padrões multimodais de sociabilidade, atividade física e mobilidade.

(Nepal et al. 2022)	Este artigo estudou as mudanças comportamentais de 180 estudantes universitários um ano antes da pandemia de COVID-19 como linha de base e depois durante o primeiro ano da pandemia usando detecção de <i>smartphone</i> e inferência comportamental.
(Filho et al. 2021)	Este estudo usou o algoritmo <i>Birch</i> para identificar pessoas com comportamentos relacionados à saúde mental semelhantes ao longo do tempo.
(Lai et al. 2021)	Este trabalho apresenta os resultados de um estudo de caso de um indivíduo monitorado por meio de uma abordagem multimodal de detecção de dados para entender padrões comportamentais (isto é, atividade física, sono e humor) que ajudam a explicar as mudanças no bem-estar psicológico do indivíduo durante a pandemia de COVID-19.
(Jagesar et al. 2021)	Este estudo identificou mudanças nos comportamentos humanos (mobilidade e sociabilidade) causadas pelas medidas de distanciamento social e bloqueio introduzidas para conter o surto de coronavírus COVID-19.
(Huckins et al. 2020)	Este estudo reconheceu mudanças nos padrões comportamentais (isto é, padrões de sono, mobilidade e uso de dispositivos) e na saúde mental (isto é, sintomas de depressão e ansiedade) de estudantes em resposta à pandemia de COVID-19 em comparação com períodos anteriores.
(Barnett et al. 2018)	Um método multivariado de detecção de <i>outliers</i> foi usado para avaliar as mudanças nos padrões de sociabilidade e mobilidade antes de recaídas esquizofrênicas.
(Gu et al. 2017)	Foi proposto um método computacional para detectar atividades familiares (por exemplo, comer, cozinhar, assistir TV, correr) com uma precisão de 80%.
(Harari et al. 2017)	Foram usados métodos de detecção para monitorar padrões de estabilidade e mudança que caracterizam a atividade física e os comportamentos de sociabilidade de alunos durante um período acadêmico de 10 semanas.
(Lane et al. 2014)	Foi proposto um aplicativo móvel para detectar e quantificar o nível de atividade física, sociabilidade e qualidade do sono.

3.3.3.2 Classificação de Estados Mentais

A tabela 7 apresenta os principais resultados dos estudos voltados à classificação, ordenados por ano de publicação. Em geral, as soluções propostas classificam estados emocionais (Sultana, Al-Jefri e Lee 2020, Buda, Khwaja e Matic 2021), ideação suicida (Haines-Delmont et al. 2020), depressão (Narziev et al. 2020, Masud et al. 2020, Wahle et al. 2016, Chikersal et al. 2021, Xu et al. 2021), recaídas esquizofrênicas (Wang et al. 2020), ansiedade (Levine et al. 2020), humor (Dogrucu et al. 2020, Spathis et al. 2019), solidão (Doryab et al. 2019) e estresse (Wang et al. 2019, Sano et al. 2018).

Tabela 7 – Principais resultados de estudos que visam classificar estados/transtornos mentais.

Referência	Principais resultados
------------	-----------------------

(Chikersal et al. 2022)	Este artigo apresenta uma abordagem de ML para detectar depressão com uma precisão de 82,5%, altas cargas globais de sintomas de esclerose múltipla com uma precisão de 90% e fadiga severa com uma precisão de 75,5%.
(Chikersal et al. 2021)	Foi projetada uma abordagem de ML usando dados de <i>smartphones</i> e dispositivos vestíveis de alunos para detectar sintomas depressivos pós-semester (isto é, ausência ou presença) com uma precisão de 85,7% e mudança na gravidade dos sintomas com uma precisão de 85,4%.
(Matteo et al. 2021)	Um algoritmo de regressão logística foi treinado para classificar transtorno de ansiedade social e depressão com área sob a curva ROC (AUC) de 64% e 72%, respectivamente.
(Buda, Khwaja e Matic 2021)	Um classificador de estado emocional (isto é, felicidade) foi projetado usando características baseadas em <i>outliers</i> , que mostraram melhorias de até 13% na AUC e 27% no F-score em comparação com as abordagens de modelagem tradicionais.
(Xu et al. 2021)	Um algoritmo de aprendizado baseado em memória foi usado para detectar depressão (isto é, ausência ou presença) entre estudantes universitários com uma precisão de 82% e um F-score de 85%.
(Haines-Delmont et al. 2020)	O algoritmo k-vizinhos mais próximos foi usado para classificar o risco de suicídio (isto é, baixo ou alto risco) com 68% de precisão.
(Narziev et al. 2020)	Um modelo de ML foi desenvolvido com base nas características extraídas das situações identificadas, que pode classificar os níveis de depressão (isto é, normal, leve, moderado ou grave) com uma precisão de 96%.
(Wang et al. 2020)	Um modelo de Máquina de Vetor de Suporte (SVM) com o kernel da Função de Base Radial (RBF) foi usado para prever recaídas esquizofrênicas a partir de dados de detecção passiva (precisão = 26,8% e <i>recall</i> = 28,4%).
(Sultana, Al-Jefri e Lee 2020)	A melhor AUC para detecção do estado emocional (isto é, discordante, satisfeito, dissuadido, excitado, submisso ou dominante) atingiu 60,55% nos modelos gerais (isto é, usando dados de todos os participantes) e 96,33% em modelos personalizados (isto é, usando dados de cada participante individualmente). A melhor AUC para detecção de transição emocional atingiu 90,5% nos modelos gerais e 88,73% nos modelos personalizados.
(Masud et al. 2020)	O algoritmo SVM foi usado para classificar os níveis de depressão (isto é, ausência, moderada ou grave) com uma precisão de 86,4%.
(Levine et al. 2020)	Características comportamentais foram usadas para projetar um modelo extra-árvores capaz de classificar os níveis diários de ansiedade e depressão (isto é, baixo, moderado, alto ou muito alto) com uma precisão de 76%.
(Dogrucu et al. 2020)	Algoritmos de ML foram treinados para serem capazes de classificar sintomas depressivos (sensibilidade de 75% e especificidade de 79,2%) e ideação suicida (sensibilidade de 86% e especificidade de 72,5%).
(Zhang et al. 2019)	Um modelo de classificação foi desenvolvido usando uma abordagem de <i>multi-view learning</i> para detectar instabilidade de humor com uma precisão de 78,33%.
(Doryab et al. 2019)	Um modelo de ML foi proposto para detectar níveis de solidão (isto é, baixo ou alto) usando dados de contexto coletados de <i>smartphones</i> e dispositivos vestíveis com uma precisão de 88,4%.

(Fukazawa et al. 2019)	Um modelo de floresta aleatória foi projetado para classificar os níveis de estresse relacionados à ansiedade usando características comportamentais extraídas dos dados de sensores de <i>smartphone</i> com um F-score de 74,2%.
(Sarda et al. 2019)	Um modelo de classificação foi projetado para detectar níveis de depressão (isto é, ausência ou presença) apresentados por indivíduos diabéticos (precisão de 81,1% e sensibilidade de 75%).
(Wang et al. 2019)	Foi desenvolvido um <i>framework</i> que pode realizar a detecção de estresse (isto é, ausência ou presença) com base na detecção de dados do <i>smartphone</i> , alcançando uma precisão de \approx 80%.
(Spathis et al. 2019)	Foi desenvolvido um modelo ML para detectar indivíduos cujo humor geral se desvia da característica relaxada comum da população em geral, alcançando AUC de \approx 75%.
(Sano et al. 2018)	Algoritmos SVM foram treinados a partir de características fisiológicas e comportamentais para classificar alto e baixo bem-estar mental (87% de precisão) e altos e baixos níveis de estresse (78,3% de precisão).
(Servia-Rodríguez et al. 2017)	Uma rede neural profunda foi treinada para classificar os humores positivos e negativos dos indivíduos com uma precisão de \approx 70%.
(Wahle et al. 2016)	Foi desenvolvido um modelo de classificação capaz de detectar indivíduos depressivos (isto é, ausência ou presença) com sensibilidade de 72,5% e especificidade de 60,8%.
(Farhan et al. 2016)	O algoritmo <i>multi-view bi-clustering</i> foi usado para reconhecer subgrupos de comportamento dos indivíduos. Além disso, o SVM foi usado para classificar os indivíduos em diferentes grupos com uma precisão de 87%.
(Grñerbl et al. 2015)	Um sistema baseado na detecção de <i>smartphones</i> foi proposto para reconhecer estados depressivos e maníacos com uma precisão de 76% e detectar mudanças de estado em pacientes com transtorno bipolar com uma precisão de cerca de 97%.

3.3.3.3 Previsão de Estados Mentais

A tabela 8 apresenta os principais resultados de estudos que visam prever estados/transtornos mentais, ordenados por ano de publicação. Em geral, os estudos se concentram em projetar modelos para prever recaídas esquizofrênicas (Barnett et al. 2018, Adler et al. 2020, Wang et al. 2017), depressão (Wang et al. 2018, Place et al. 2017, Demasi, Aguilera e Recht 2016, Chikersal et al. 2021), humor (Morshed et al. 2019), comportamento impulsivo (Wen et al. 2021), ansiedade (Moshe et al. 2021) e estados afetivos (Cai et al. 2018).

Tabela 8 – Principais resultados de estudos que visam prever estados/transtornos mentais.

Referência	Principais resultados
(Wang et al. 2022)	Um modelo de aprendizado profundo foi proposto para prever a saúde mental de alunos (pontuações PHQ-4) com <i>F1-score</i> de 71%.

(Zhou et al. 2022)	Um modelo de floresta aleatória balanceada foi projetado usando dados de detecção móvel para prever recaídas esquizofrênicas com um <i>F2 score</i> de 0,23.
(Chikersal et al. 2021)	Uma abordagem de ML foi desenvolvida usando dados de <i>smartphones</i> e dispositivos vestíveis de alunos para prever sintomas depressivos (isto é, ausência ou presença) 11 a 15 semanas antes do final do semestre, com uma precisão de 80%.
(Wen et al. 2021)	Um modelo de ML foi projetado para prever diferentes facetas do comportamento impulsivo (por exemplo, busca de sensações, planejamento, falta de perseverança, e falta de premeditação).
(Henson et al. 2021)	Uma abordagem de detecção de anomalia baseada em <i>smartphone</i> foi usada para prever recaídas esquizofrênicas com sensibilidade de 89% e especificidade de 75%.
(Adler et al. 2021)	Abordagens de estimativa de densidade foram desenvolvidas para prever indicadores de resiliência ao estresse (isto é, indivíduos que experimentaram alterações mínimas de saúde mental durante um período prolongado) a partir de dados de detecção móvel em uma grande população de médicos internos.
(Sükei et al. 2021)	Um modelo ML foi projetado para prever a valência emocional autorreferida, cujo melhor desempenho alcançou AUC de 81%.
(Moshe et al. 2021)	Modelos multiníveis foram usados para prever pontuações de ansiedade, estresse e depressão usando dados de <i>smartphones</i> e dispositivos vestíveis.
(Jacobson, Summers e Wilhelm 2020)	Um conjunto de <i>Extreme Gradient Boosting</i> (XGBoost) foi treinado usando dados de sensores passivos para prever a gravidade dos sintomas de ansiedade social. Uma forte correlação foi identificada ($r = 0,702$) entre a gravidade dos sintomas previstos e observados.
(Adler et al. 2020)	Uma arquitetura de <i>autoencoder</i> de rede neural totalmente conectada foi projetada para prever sinais de alerta de recaída psicótica com uma especificidade de 88%.
(Jacobson e Chung 2020)	Uma combinação de modelos de aprendizado de máquina nomotéticos e ponderados ideograficamente foi usada para prever o humor deprimido de hora em hora.
(Rashid et al. 2020)	Este trabalho usa o algoritmo <i>catboost</i> para prever medidas subjetivas de ansiedade social a partir de dados passivos. Além disso, explora métodos de imputação para lidar com dados ausentes que reduzem o erro de previsão em até 22%.
(Morshed et al. 2019)	Um modelo de ML foi desenvolvido para prever a estabilidade do humor (isto é, diferenças sucessivas de respostas de valência e excitação) a partir de três semanas de dados coletados passivamente de sensores embutidos em <i>smartphones</i> e dispositivos vestíveis.
(Cho et al. 2019)	O algoritmo de floresta aleatória foi usado para prever o estado de humor para os próximos três dias de todos os pacientes, pacientes com transtorno depressivo maior (MDD), pacientes com transtorno bipolar (TB) I e pacientes com TB II com uma precisão de 65%, 65% , 64% e 65%, respectivamente.
(Wang et al. 2018)	Um modelo de ML foi proposto para prever os sintomas de depressão (isto é, ausência ou presença) apresentados por alunos (<i>recall</i> de 81,5% e precisão de 69,1%).

(Barnett et al. 2018)	Identificou-se que a taxa de anomalias no comportamento social e na mobilidade detectadas nas duas semanas anteriores às recaídas esquizofrênicas foi 71% maior do que a taxa de anomalias nos outros períodos.
(Cai et al. 2018)	Um modelo de ML foi projetado com base em características comportamentais de mobilidade e atividade física para prever estados afetivos negativos. Os modelos personalizados (isto é, projetados com dados de um indivíduo específico) alcançaram o erro quadrático médio (RMSE) ≈ 13 , e os modelos generalizados (isto é, projetados com dados de todos os indivíduos) atingiram o RMSE ≈ 22 .
(Place et al. 2017)	Foi identificado que os indicadores comportamentais podem prever sintomas avaliados clinicamente para transtorno de estresse pós-traumático e depressão (por exemplo, humor deprimido, fadiga, interesse em atividades e conexão social).
(Wang et al. 2017)	Um sistema foi desenvolvido para prever sintomas de esquizofrenia com base na Escala Breve de Avaliação Psiquiátrica (BPRS) com erro médio de $\pm 1,45$ usando características rastreadas automaticamente de <i>smartphones</i> .
(Wang et al. 2016)	Um modelo de ML foi desenvolvido a partir de características de comportamentos (isto é, atividade física, sociabilidade, mobilidade, padrões de uso de dispositivos, sono e contexto ambiental) para prever pontuações agregadas de indicadores de saúde mental em esquizofrenia com um erro médio de 7,6% da faixa de pontuação .
(Demasi, Aguilera e Recht 2016)	A regressão logística foi utilizada para prever mudanças no sintoma depressivo dos indivíduos medidos pelo Inventário de Depressão de Beck.

3.3.3.4 Associação entre Características e Estados Mentais

A tabela 9 apresenta os principais resultados de estudos que analisam a associação existente entre estados/transtornos mentais e características extraídas de situações identificadas. Ordenamos os estudos apresentados por ano de publicação. Especificamente, os pesquisadores usam técnicas estatísticas (por exemplo, correlação de *Spearman* e correlação de *Pearson*) para medir o nível de associação entre as variáveis analisadas. Esta abordagem investiga se as características extraídas dos dados de contexto representam marcadores comportamentais que se correlacionam com estados/transtornos mentais, como depressão (Masud et al. 2020, Wang et al. 2018, Moshe et al. 2021), comportamento impulsivo (Wen et al. 2021), esquizofrenia (Fulford et al. 2020, Wang et al. 2016), humor (Cai et al. 2020, Servia-Rodríguez et al. 2017), ansiedade social (Gong et al. 2019) e estado afetivo (Cai et al. 2018).

Tabela 9 – Principais resultados de estudos que visam associar características a estados/transtornos mentais.

Referência	Principais resultados
(Wang et al. 2022)	Foram identificadas correlações entre regularidade do sono, atividade física, e mobilidade com os escores do Questionário de Saúde do Paciente-4 (PHQ-4) de estudantes.

(Moshe et al. 2021)	Uma associação significativa foi identificada entre dados de <i>smartphones</i> e sintomas de ansiedade, estresse e depressão.
(Wen et al. 2021)	Correlações significativas foram identificadas entre dados passivos e diferentes facetas do traço de impulsividade (isto é, busca de sensações, planejamento, falta de perseverança, falta de premeditação e traços de atenção).
(Adler et al. 2021)	Foram encontradas correlações significativas entre resiliência ao estresse e atividade física (contagem de passos), duração do sono e humor diário.
(Masud et al. 2020)	Fortes correlações entre níveis de depressão e características comportamentais (mobilidade e atividade física) foram reconhecidas.
(He-Yueya et al. 2020)	Foi identificado que maior estabilidade na atividade social (por exemplo, chamadas telefônicas, sms) foi associada a sintomas mais baixos de transtornos do espectro da esquizofrenia, e menor nível de atividade física (por exemplo, ficar parado) indicou associação a sintomas elevados.
(Fulford et al. 2020)	Foram identificadas correlações entre mobilidade e sociabilidade com avaliações autorrelatadas do comprometimento social de pacientes com esquizofrenia.
(Cai et al. 2020)	Foram reconhecidas associações significativas entre o humor de pacientes com câncer de mama e suas atividades físicas e tempo gasto em locais visitados (por exemplo, casa, restaurante).
(Doryab et al. 2019)	Foram identificadas correlações entre os níveis de solidão e o comportamento dos alunos.
(Gong et al. 2019)	Identificou-se que indivíduos com ansiedade social apresentam aumento da movimentação corporal durante atividades sociais (isto é, ligações telefônicas), principalmente em locais desconhecidos como fora de casa e no trabalho.
(Sefidgar et al. 2019)	Foi identificado que os alunos que sofrem tratamento injusto tornam-se mais ativos fisicamente, interagem mais com o telefone pela manhã, fazem mais ligações à noite e passam mais tempo na cama no dia do tratamento injusto.
(Wang et al. 2018)	Foram reconhecidas correlações entre características comportamentais (isto é, sono, padrões de uso do dispositivo, mobilidade, sociabilidade, humor, atividade física) e níveis de depressão dos alunos.
(Boukhechba et al. 2018)	Foram identificadas correlações significativas entre as características comportamentais (isto é, padrões de mobilidade, sociabilidade e atividade física) e os níveis de depressão e ansiedade social apresentados pelos alunos.
(Cai et al. 2018)	Identificou-se que as flutuações no estado de afeto negativo dos indivíduos estão associadas à atividade física e à mobilidade.
(Wang et al. 2017)	Foram identificadas associações entre os dados de contexto coletados dos <i>smartphones</i> dos alunos e sua saúde mental (isto é, níveis de depressão, estresse e solidão).
(Servia-Rodríguez et al. 2017)	Associações entre características e humor autorrelatado foram encontradas em um conjunto de dados de grande escala.
(Wang et al. 2016)	Foram encontradas correlações significativas entre comportamentos rastreados automaticamente (por exemplo, sono, mobilidade, conversas) e indicadores autorrelatados de sintomas esquizofrênicos.
(Demasi, Aguilera e Recht 2016)	Foram identificadas correlações entre sintomas depressivos e nível de atividade física e duração do sono.

3.4 Resultados

Nesta seção, derivamos discussões relevantes dos resultados desta revisão sistemática da literatura. Precisamente, discutimos as principais tendências nesta área de pesquisa e apresentamos as oportunidades de pesquisas identificadas através da revisão conduzida.

3.4.1 Tendências

Analisando a literatura revisada, reconhecemos um conjunto de tendências apresentadas pelos estudos de fenotipagem digital da saúde mental. Primeiro, notamos uma tendência de usar tecnologias ubíquas para realizar a detecção multimodal de situações de interesse para profissionais de saúde mental. Esses trabalhos desenvolvem soluções que realizam principalmente a fusão de dados coletados de sensores embarcados em *smartphones*. Além disso, alguns estudos realizam a fusão de dados fornecidos por sensores disponíveis em *smartphones* e em dispositivos vestíveis.

Também reconhecemos uma tendência para identificar diferentes situações usando sensoriamento multimodal. Em geral, os trabalhos processam dados de contexto para conscientizar sobre situações que revelam indicadores de saúde mental. Os estudos se concentram principalmente na detecção passiva de informações sobre mobilidade, atividade física e sociabilidade. Também identificamos uma tendência na detecção de padrões de uso de dispositivos, qualidade do sono e contexto ambiental.

As soluções de fenotipagem digital da saúde mental possuem um fluxo de trabalho bem definido para transformar dados contextuais em informações relevantes para profissionais especializados. Precisamente, os trabalhos tendem a realizar as seguintes etapas: (i) coleta de dados de contexto; (ii) pré-processamento e limpeza de dados; (iii) extração de características; (iv) identificação de situações; (v) extração de características de situações; e (vi) desenvolvimento de modelos para apoiar o monitoramento e avaliação da saúde mental.

Conforme apresentado na Seção 3.3.3, classificamos os artigos revisados em quatro tipos de estudos. A Figura 8 mostra a frequência de cada tipo de estudo. Primeiro, reconhecemos um esforço substancial na concepção de modelos de ML para classificar ($n = 23$) e prever ($n = 21$) estados mentais e distúrbios como depressão, ansiedade, humor e esquizofrenia. Outra tendência é o uso de coeficientes de correlação para investigar a associação entre características e estados e transtornos mentais ($n = 18$), visando encontrar evidências de que essas características representam marcadores comportamentais relevantes. Finalmente, há um interesse em derivar informações de alto nível do sensoriamento multimodal de situações ($n = 11$).

A Figura 8 mostra a distribuição dos estudos selecionados por ano, considerando o tipo de estudo. Essa distribuição mostra um aumento de trabalhos publicados a partir de 2016, indicando o crescente interesse em realizar estudos nessa área de pesquisa. Também é possível verificar um aumento no número de estudos que visam classificar os estados mentais a partir de 2019. Além disso, reconhecemos que o número de estudos na literatura sobre fenotipagem digital da saúde mental aumentou desde 2019, o que pode estar relacionado a crise epidemiológica da pandemia da COVID-19. A pandemia da COVID-19 provocou alterações sociais significativas, impondo limitações à capacidade de interação física entre as pessoas. Essas mudanças representam oportunidades de pesquisa para estudos que visam desenvolver soluções para fenotipagem digital da saúde mental, pois essas soluções se propõem a fornecer informações sobre uma nova perspectiva sobre o comportamento humano.

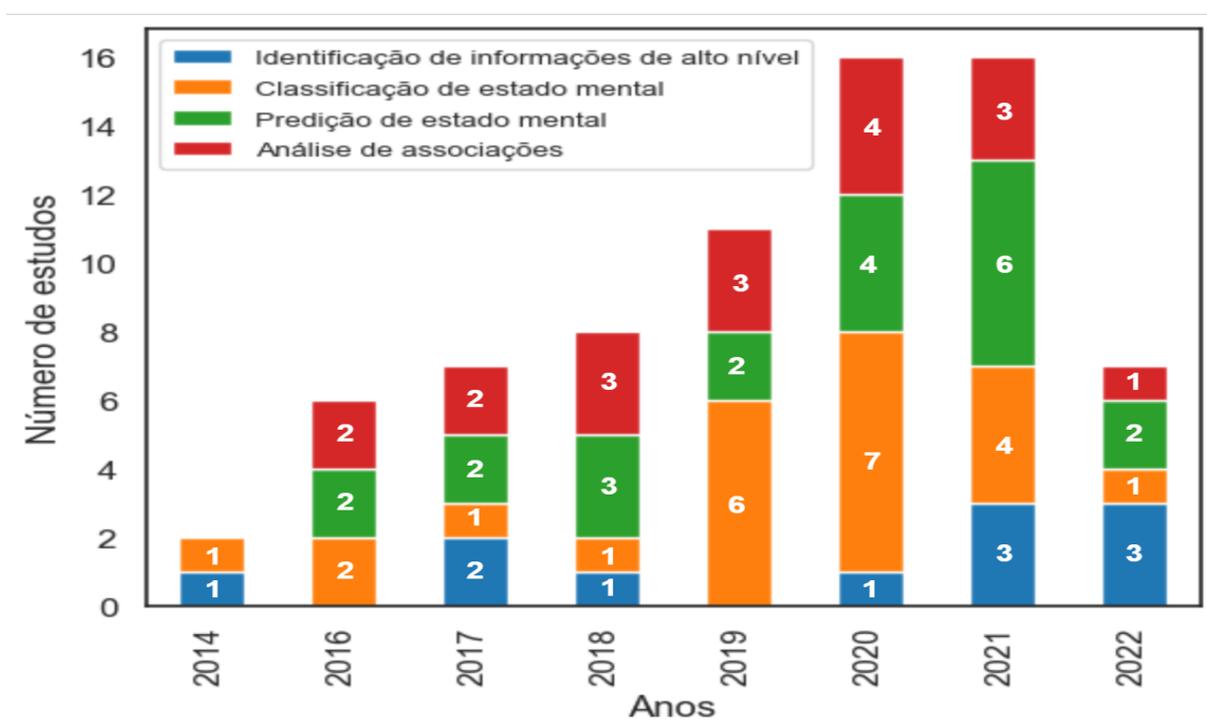


Figura 8 – Número de artigos publicados por tipo de estudo.

3.4.2 Oportunidades de Pesquisas Identificadas

A literatura sobre monitoramento pervasivo da saúde mental tem se concentrado na quantificação de comportamentos e na associação, classificação e predição de estados mentais. Apesar dos estudos revisados usarem a detecção automática de comportamento para projetar abordagens que apoiem o monitoramento e a avaliação de estados mentais, ainda existem questões em aberto neste campo de pesquisa. É necessário projetar abordagens para extrair informações de alto nível dos comportamentos identificados, como padrões capazes de descrever hábitos dos

indivíduos. Outra questão é a necessidade de projetar soluções para reconhecer mudanças comportamentais, pois essas informações podem indicar uma possível mudança do estado mental, permitindo então que especialistas verifiquem essas informações. Além disso, também é necessário utilizar informações de contexto no processo de detecção de padrões e mudanças comportamentais, como o dia da semana, dados climáticos, finais de semana, dias de trabalho, dentre outras informações que podem causar uma variação normal no comportamento humano.

Embora os trabalhos revisados enfoquem a fenotipagem digital de comportamentos humanos, eles diferem da solução proposta neste trabalho de doutorado em alguns aspectos. Primeiro, o trabalho proposto não é direcionado a um estado/transtorno mental específico, mas se concentra na identificação de padrões comportamentais a partir de situações de interesse. Segundo, a solução proposta apresenta um modelo de dados que pode ser facilmente estendido para apoiar a identificação de novos tipos de padrões de comportamento (por exemplo, mobilidade, sono). Terceiro, a solução proposta reconhece padrões comportamentais multimodais com base em informações contextuais, permitindo a identificação de variações comportamentais. Quarto, a solução proposta pode reconhecer comportamentos anormais e mudanças significativas no hábito comportamental de pacientes. Finalmente, a solução proposta aprende incrementalmente padrões comportamentais a partir de inferências de situações de interesse.

3.5 Considerações Finais

Este capítulo apresentou uma RSL sobre o monitoramento da saúde mental através do sensoriamento multimodal de múltiplas situações de interesse para profissionais da saúde. Essa é uma das contribuições desta pesquisa, uma vez que, de acordo com nosso conhecimento, não há uma profunda revisão recente sobre métodos capazes de reconhecer situações de interesse para melhorar os serviços de profissionais de saúde mental. Assim, os resultados desta RSL contribuem para uma melhor compreensão do uso dos dispositivos ubíquos para monitorar objetivamente o comportamento humano ([Moura et al. 2022](#)).

A partir da análise aprofundada dos estudos primários selecionados, foram identificadas oportunidades de pesquisas que motivam o desenvolvimento deste estudo. Portanto, este trabalho visa explorar a necessidade da concepção de novos estudos para reconhecer padrões e mudanças comportamentais para os profissionais de saúde. Especificamente, este estudo contribui com a literatura atual através do desenvolvimento de uma solução capaz de reconhecer padrões comportamentais projetados com base em informações contextuais, assim como comportamentos anormais e mudanças de padrões dos indivíduos monitorados.

3.5.1 Publicações

A revisão sistemática apresentada neste capítulo se encontra publicada ([Moura et al. 2022](#)). Abaixo informa-se o tipo de publicação, status em que ela se encontra, o Qualis (conforme classificação quadriênio 2017-2020) do periódico e o fator de impacto.

- Moura, Ivan; Teles, Ariel; Viana, Davi; Marques, Jean; Coutinho, Luciano; Silva, Francisco. **Digital Phenotyping of Mental Health Using Multimodal Sensing of Multiple Situations of Interest: A Systematic Literature Review** Journal of Biomedical Informatics, 2022.

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A1. **Fator de impacto:** 4.5

4 Solução Proposta

Neste capítulo, apresentamos inicialmente o modelo formal para soluções de fenotipagem digital e o modelo de dados projetados para apoiar o desenvolvimento da solução proposta. Em seguida introduzimos a solução proposta para detectar padrões e mudanças comportamentais sensíveis ao contexto. O objetivo dessa solução é fornecer informações sobre padrões comportamentais, comportamentos anômalos e mudanças de padrões comportamentais com base no processamento de fenótipos digitais. Esse objetivo busca abordar as questões em aberto identificadas na literatura, como a necessidade de gerar informações objetivas sobre padrões e mudanças comportamentais, enriquecer os modelos computacionais com informações contextuais sobre as situações identificadas e processar incrementalmente fenótipos digitais.

4.1 Modelo Formal para Soluções de Fenotipagem Digital

Nesta seção, apresentamos um modelo formal para soluções de fenotipagem digital que visam identificar padrões comportamentais enriquecidos por contexto. Nossa contribuição é um modelo formal para representar os principais conceitos desse tipo de solução, como comportamento humano, evento comportamental, observação comportamental e padrão comportamental. Baseamos o modelo formal proposto na experiência de nossos trabalhos ([Rodrigues de Moura et al. 2020](#), [Moura et al. 2021](#), [Moura et al. 2022](#)), nos quais reconhecemos a importância de entender os padrões e suas variações com base no contexto. Usamos símbolos contidos na teoria de conjuntos para projetar o modelo formal, permitindo a reutilização por outros sistemas de fenotipagem digital. A Tabela 10 mostra os conceitos cobertos pelo modelo formal proposto. Para cada conceito, apresentamos um cenário ilustrativo de uma pessoa chamada Eliza, possibilitando assim traçar a aplicabilidade de cada conceito. Consideramos que a rotina comportamental dessa pessoa consiste em uma sequência de atividades que representam comportamentos como sociabilidade, mobilidade e atividade física. Além disso, essas atividades ocorrem em diferentes contextos, como segunda, sexta, dias úteis e finais de semana.

4.1.1 Comportamento Humano

Formalizamos o comportamento humano como um conjunto de ações físicas associadas a um indivíduo, como interações sociais, atividades físicas e mobilidade.

$$B = \{b_1, b_2, b_3, \dots, b_n\}$$

Tabela 10 – Conceitos usados no fluxo de trabalho de soluções de fenotipagem digital.

Conceito	Significado
Comportamento Humano	Refere-se a um conjunto de comportamentos físicos em que os indivíduos se envolvem.
Eventos Comportamentais	Representa comportamentos identificados pelo processamento de dados coletados de sensores embutidos ou conectados a dispositivos ubíquos.
Eventos Comportamentais Enriquecidos por Contexto	Representa eventos comportamentais contendo informações sobre o contexto do evento, como dia da semana, local e dados meteorológicos.
Observações Comportamentais	Representa a quantificação momento a momento do comportamento humano.
Padrão Comportamental	Representa um conjunto de intervalos de tempo em que o indivíduo repete rotineiramente um comportamento.

Onde:

- B é o conjunto de comportamentos associados a um indivíduo.

Cenário ilustrativo 1. Apresentamos a seguir os comportamentos monitorados de Eliza, com comportamentos de caminhada, estado estacionário, mobilidade ao ar livre, interação face a face e sociabilidade virtual.

- $B = \{\text{walking, stationary state, outdoor mobility, face-to-face interaction, virtual socialization}\}$.

4.1.2 Eventos Comportamentais

Eventos comportamentais representam comportamentos identificados pelo processamento de dados coletados de sensores embutidos ou conectados a dispositivos ubíquos. Esses eventos contêm, além da identificação do comportamento, informações temporais (horário de início e término do evento) e identificação do proprietário do *smartphone*. Assim, formalizamos um evento comportamental como o seguinte quádruplo:

$$Event = (uid, b_i, T_s, T_e)$$

Onde:

- uid é o identificador do indivíduo;
- b_i é uma instância de comportamento;
- T_s é o timestamp de início do evento comportamental;
- T_e é o timestamp final do evento comportamental.

Cenário ilustrativo 2. Ao acordar, Eliza socializa com seus amigos interagindo em sua rede social online entre 6h00 e 6h30. Ela toma café da manhã com a família às 6h50 e participa de conversas até às 7h20. Ela então decide praticar a caminhada matinal às 7h30, com 1 hora e 20 de duração. Ao chegar em casa, ela se senta para assistir ao seu programa favorito às 9h, com 50 minutos de duração. Nesse cenário, a rotina de Eliza gera os seguintes eventos comportamentais:

- $Event_1 = (u23, \text{virtual socialization}, 6:00, 6:30)$;
- $Event_2 = (u23, \text{face-to-face interaction}, 6:50, 7:20)$;
- $Event_3 = (u23, \text{walking}, 7:30, 8:50)$;
- $Event_4 = (u23, \text{stationary state}, 9:00, 9:50)$.

4.1.3 Eventos Comportamentais Enriquecidos por Contexto

Ao realizar a identificação de eventos comportamentais, é essencial enriquecê-lo com atributos de contexto relacionados ao comportamento reconhecido, como informações meteorológicas (por exemplo, tempo chuvoso e nublado), informações temporais (dias da semana e fins de semana) e estados de humor (por exemplo, humor positivo e negativo). Além disso, essas informações contextuais são essenciais para modelar a variação comportamental normal, viabilizando projetar padrões comportamentais baseados nesses atributos de contexto. Assim, formalizamos atributos de contexto como uma propriedade que caracteriza entidades, ambientes ou situações, representando assim qualquer informação relacionada ao contexto em que ocorreu o evento identificado.

$$CA = \{c_1, c_2, \dots, c_n\}$$

Usamos o conceito de atributo de contexto para formalizar eventos comportamentais enriquecidos por contexto, estendendo então o quádruplo para um quádruplo, em que o quinto elemento é um conjunto de atributos de contexto vinculados ao evento comportamental identificado. Assim, formalizamos eventos comportamentais enriquecidos por contexto, conforme apresentado a seguir.

$$Event_{er} = (uid, b_i, T_s, T_e, \{c_1, c_2, \dots, c_n\})$$

Cenário ilustrativo 3. A rotina de Eliza varia em função de alguns contextos. Por exemplo, em dias úteis (isto é, de segunda a sexta), Eliza costuma conversar com seus colegas de trabalho às 8h, com duração média de 40 minutos. Ao sair do trabalho, pratica caminhada às 18h30, com duração de 1h30. Aos fins-de-semana (sábado e domingo), pratica atividade física das 9h00 às 10h00 e sai à noite com os colegas para confraternização das 21h00 às 23h30. Em dias de chuva, Eliza fica em casa o dia todo, vindo assistir TV sentada (estado estacionário) no sofá das 14h às 18h30. Nesse cenário, a rotina de Eliza gera o seguinte fluxo de eventos comportamentais enriquecidos por contexto:

- $Event_{er1} = (u23, \text{face-to-face interaction}, 8:00, 8:40, \{\text{tuesday, workday}\});$
- $Event_{er2} = (u23, \text{walking}, 18:30, 20:00, \{\text{tuesday, workday}\});$
- $Event_{er3} = (u23, \text{walking}, 9:00, 10:00, \{\text{saturday, sunday, weekend}\});$
- $Event_{er4} = (u23, \text{face-to-face interaction}, 21:00, 23:30, \{\text{saturday, sunday, weekend}\});$
- $Event_{er5} = (u23, \text{stationary state}, 14:00, 18:30, \{\text{rainy day}\}).$

4.1.4 Observações Comportamentais

Uma observação comportamental representa a quantificação momento a momento do comportamento humano. Primeiro, subdividi-se o dia (isto é, um período de 24 horas) em slots de tempo para representar um momento. Em seguida, quantifica-se o comportamento monitorado (isto é, número de eventos comportamentais) para cada slot de tempo. Portanto, a contagem de eventos comportamentais em cada slot de tempo de um dia representa uma observação comportamental. A Figura 9 ilustra a representação de uma observação comportamental, na qual, cada slot de tempo possui um identificador e a respectiva contagem de eventos. Definimos uma janela de tempo de 24 horas, devido ao ritmo diário do comportamento humano que se repete regularmente por aproximadamente 24 horas. Por fim, formalizamos a observação comportamental conforme apresentado a seguir.

$$obs = \{(s_1, count), (s_2, count), \dots, (s_n, count)\}$$

Onde

- s = identificador sequencial do slot de tempo;;
- $count$ = contagem de eventos que ocorreram em um determinado slot de tempo.

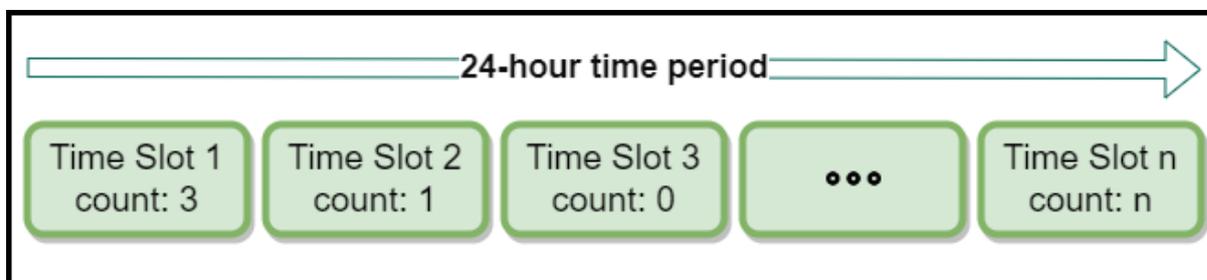


Figura 9 – Representação da quantificação momento a momento do comportamento humano.

4.1.5 Padrão Comportamental

Analisar um conjunto de observações comportamentais torna possível encontrar padrões comportamentais. Identificamos um padrão comportamental quando um comportamento é repetido com frequência suficiente em um determinado contexto. Definimos padrões de comportamento como uma rotina diária. A rotina diária representa os intervalos de tempo do dia em que um comportamento específico acontece habitualmente. Os intervalos de tempo na rotina diária possuem um conjunto de slots de tempo adjacentes. Formalizamos padrões comportamentais como o seguinte quádruplo:

$$P = (uid, b_i, dr, ca)$$

Onde:

- uid é o identificador do indivíduo;
- b_i é uma instância de comportamento;
- $ti = \{s_1, s_2, \dots, s_n\}$ representa um intervalo de tempo composto por um conjunto de slots de tempo adjacentes;
- $dr = \{ti_1, ti_2, \dots, ti_n\}$ representa a rotina diária composta por um conjunto de intervalos de tempo;
- ca é o atributo de contexto usado ao projetar um padrão.

Cenário ilustrativo 4. Eliza faz uma longa caminhada nos finais de semana, habitualmente das 6h às 8h30. Às segundas-feiras, faz caminhadas habitualmente das 17h às 18h. Em dias de chuva, Eliza costuma fazer apenas 30 minutos de caminhada dentro de seu condomínio às 18h. Assim, considerando slots de tempo de 30 minutos (isto é, dividindo o dia em 48 slots de tempo), os padrões de atividade física de Eliza são:

- $P_1 = (u23, (\text{physical activity, walking}), \{s13, s14, s15, s16, s17\}, \text{weekend});$
- $P_2 = (u23, (\text{physical activity, walking}), \{s35, s36\}, \text{monday});$
- $P_3 = (u23, (\text{physical activity, walking}), \{s37\}, \text{rainy day}).$

4.2 Modelo de Dados para Fenotipagem Digital

Modelos de dados permitem demonstrar a organização das estruturas de dados e seus relacionamentos, facilitando assim o entendimento do desenvolvimento de software (Simsion e Witt 2004). O projeto de modelo de dados depende do nível de dados utilizado, em que pode-se considerar o nível de dados brutos, eventos comportamentais e padrões comportamentais. A Figura 10 demonstra uma representação desses níveis de dados. O nível de dados brutos representa dados coletados de sensores sem nenhum processamento adicional, isto é, dados brutos provenientes de sensores, como acelerômetro, GPS, e sensor de luminosidade. O nível de eventos comportamentais processa dados brutos para detectar situações de interesse, como atividade física, mobilidade e sociabilidade. Por sua vez, o nível de padrões comportamentais processa eventos comportamentais para reconhecer informações de mais alto nível, como padrões que representam o hábito comportamental, mudanças de padrões e comportamentos anormais. Especificamente, a solução proposta neste estudo não trabalha no nível de dados brutos, estando focada na geração de padrões comportamentais a partir do processamento de eventos comportamentais.

A partir do modelo formal proposto neste estudo, projetamos um modelo de dados para apoiar o desenvolvimento da solução proposta. Nesse modelo, especifica-se as estruturas dos eventos comportamentais, observações comportamentais, padrões comportamentais e mudanças de padrões comportamentais. Nas seções seguintes, apresenta-se o modelo de dados projetado usando a Notação de objeto *JavaScript* (JSON).

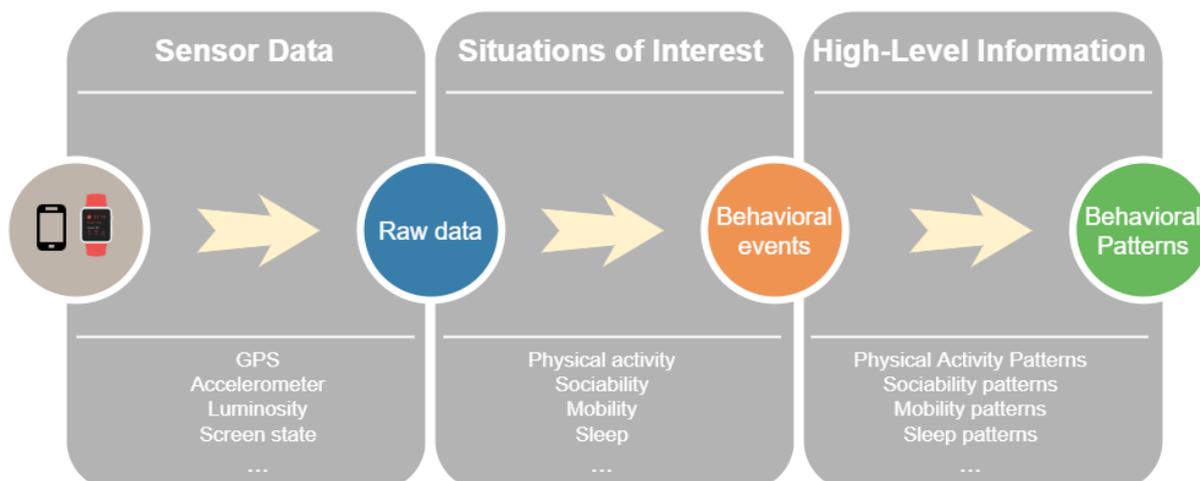


Figura 10 – Representação dos níveis de dados de contexto

4.2.1 Modelo de Dados de Eventos Comportamentais

Apesar deste estudo não focar na inferência de eventos comportamentais, é necessário projetar um modelo de dados para estes eventos, uma vez que a solução proposta processa-os para identificar padrões comportamentais. Abaixo apresentamos o modelo de dados de eventos comportamentais. O primeiro atributo é o identificador do usuário. O segundo e terceiro atributo representam o timestamp de início e fim do evento, respectivamente. O quarto atributo contém a indicação do comportamento inferido. Por fim, o quinto atributo é um array que contém os atributos de contexto relacionados ao evento.

```

1 {
2   "uid" : "u00",
3   "startDateTime": ISODate("2022-12-27T10:50:33.000Z"),
4   "endDateTime": ISODate("2022-12-27T10:51:34.000Z")
5   "behavior" : "walking",
6   "contexts" : [
7     "WEDNESDAY",
8     "WEEK"
9   ]
10 }
```

4.2.2 Modelo de Dados de Observações Comportamentais

Conforme formalizado na Seção 4.1.4, uma observação comportamental é uma representação da quantificação momento a momento do comportamento humano.

Desta forma, estruturamos observações comportamentais subdividindo o dia (isto é, tempo circular de 24 horas) em slots de tempo e contabilizando o número de eventos que aconteceram em cada slot. Apresentamos abaixo o modelo de dados de uma observação comportamental considerando slots de tempo de 1 hora (isto é, dividiu-se o dia em 24 slots de tempo). Especificamente, uma observação é um array de slots com a respectiva contagem de eventos ocorridos. A Figura 11 ilustra visualmente o modelo de dados de uma observação comportamental.

```
1 {
2   "observation" : [
3     {slot : 1, count: 0}
4     {slot : 2, count: 1},
5     {slot : 3, count: 0},
6     {slot : 4, count: 0},
7     {slot : 5, count: 0},
8     {slot : 6, count: 0},
9     {slot : 7, count: 0},
10    {slot : 8, count: 0},
11    {slot : 9, count: 0},
12    {slot : 10, count: 1},
13    {slot : 11, count: 2},
14    {slot : 12, count: 3},
15    {slot : 13, count: 3},
16    {slot : 14, count: 4},
17    {slot : 15, count: 0},
18    {slot : 16, count: 0}
19    {slot : 17, count: 0},
20    {slot : 18, count: 0},
21    {slot : 19, count: 1},
22    {slot : 20, count: 0},
23    {slot : 21, count: 0},
24    {slot : 22, count: 2},
25    {slot : 23, count: 1},
26    {slot : 24, count: 3}
27  ]
28 }
```

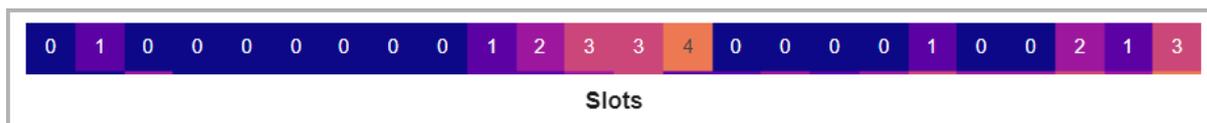


Figura 11 – Representação visual do modelo de dados de uma observação comportamental.

4.2.3 Modelo de Dados de Padrões Comportamentais

Os padrões comportamentais são projetados usando um conjunto de observações comportamentais, modelando os intervalos de tempo (isto é, conjunto de slots de tempo adjacentes) em que os indivíduos repetem rotineiramente comportamentos. Apresentamos abaixo o modelo de dados de um padrão comportamental. O atributo “uid” representa o identificador do usuário, enquanto o atributo “detectionDate” representa a data e hora da identificação do padrão. O atributo “behaviors” identifica a qual comportamento o padrão projetado pertence. O atributo “contextAttribute” indica o atributo de contexto do padrão. Por fim, o atributo “dailyRoutine” estrutura o conjunto de intervalos de tempo que representam a rotina comportamental do indivíduo.

```

1 {
2   "uid" : "u22",
3   "detectionDate" : ISODate("2013-04-09T03:00:00.000Z"),
4   "behaviors" : [
5     "convesartion",
6     "sociability"
7   ],
8   "contextAttribute" : "TUESDAY",
9   "dailyRoutine" : [
10    {
11      "startTime" : "00:00",
12      "endTime" : "01:00"
13    },
14    {
15      "startTime" : "13:30",
16      "endTime" : "17:30"
17    },
18    {
19      "startTime" : "18:00",
20      "endTime" : "00:00"

```


das propriedades do evento (por exemplo, registro de data e hora) ou recuperadas de fontes externas (por exemplo, APIs de clima). Cada atributo de contexto é usado como uma dimensão de segmentação de dados para identificar variação comportamental devido a situações de contexto específicas. Portanto, particiona-se os dados com base em atributos de contexto e realiza-se a aprendizagem incremental de padrões comportamentais para cada conjunto de dados derivado.

4.4 Detecção de Mudança Comportamental

Implementamos a detecção de mudança comportamental através do processamento de janelas de observações, possibilitando comparar a similaridade entre os padrões extraídos em diferentes janelas de observações. Comparamos padrões e comportamentos usando o coeficiente de similaridade de Jaccard (Equação 4.4), que quantifica a sobreposição entre os intervalos de tempo de padrões comportamentais. Assim, mudanças comportamentais ocorrem quando a aplicação do coeficiente de Jaccard resulta em um baixo nível de similaridade.

$$Jaccard(X, Y) = \frac{X \cap Y}{X \cup Y} \quad (4.4)$$

4.4.1 Comportamentos Anormais

Comportamentos anormais ocorrem quando os indivíduos se comportam de maneira diferente de seus hábitos. A Figura 15 apresenta o fluxo de trabalho de detecção de comportamento anormal. O fluxo de trabalho de detecção de comportamentos anormais consiste na aplicação do coeficiente de Jaccard para identificar a similaridade entre o padrão atual do indivíduo e suas observações comportamentais subsequentes. Como consequência, se a similaridade resultante estiver abaixo do limite, nossa solução reconhece o comportamento anormal.

4.4.2 Mudanças de Padrão Comportamental

Mudanças de padrões comportamentais ocorrem quando os indivíduos alteram efetivamente sua rotina comportamental. A Figura 16 mostra o fluxo de trabalho de detecção de alteração de padrão. Especificamente, a solução aplica o coeficiente de Jaccard entre os padrões para identificar mudança de padrão. Esse processo consiste em armazenar um padrão de referência e compará-lo com os padrões subsequentes. Nos casos em que a similaridade resultante é menor que o limiar de mudança, nossa solução atualiza o padrão de referência com o novo padrão identificado e notifica as partes interessadas sobre a mudança de padrão comportamental.

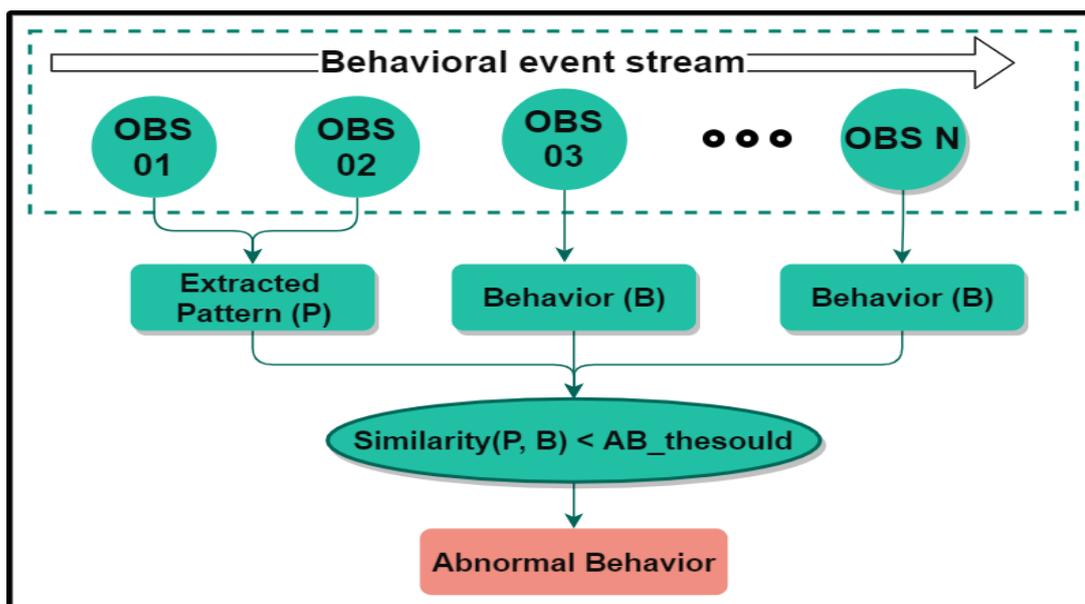


Figura 15 – Fluxo de trabalho de detecção de comportamento anormal.

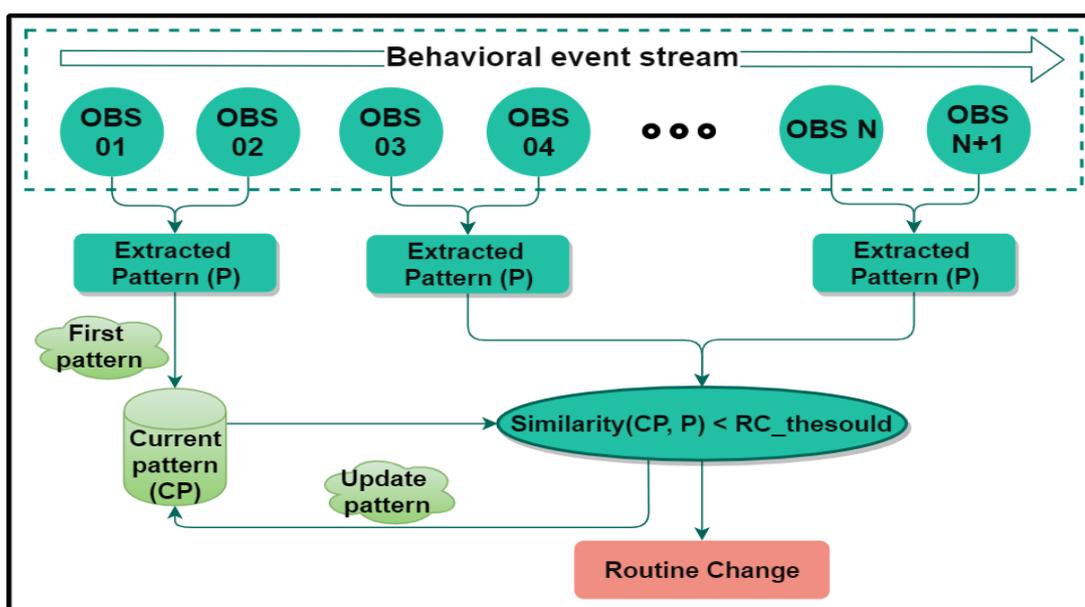


Figura 16 – Fluxo de trabalho de detecção de mudança de padrão comportamento.

4.4.3 Modelagem de Conhecimento Especializado

A detecção de mudança comportamental depende de limiares de similaridade especificados por especialistas. Essa solução é limitada pela necessidade de especificar limiares rígidos, o que implica inconsistência na modelagem da mudança comportamental. Por exemplo, definir o limite de similaridade em 50% detecta comportamento anormal se a similaridade entre padrões e observações estiver abaixo do limite especificado, mesmo que seja próximo (por exemplo, 49,65%). Mitigamos essa limitação usando lógica fuzzy, que representa conhecimento impreciso por

meio de conjuntos fuzzy (Zadeh 1996). Usamos um FIS para detectar mudanças de comportamento considerando saídas que variam de sem mudança a mudança gradualmente. Para isso, utilizamos a biblioteca Java *jFuzzyLogic*¹, que fornece suporte para a implementação de conjuntos e regras fuzzy por meio da Linguagem de Controle Fuzzy (FCL).

Utilizamos a lógica fuzzy nesta tarefa porque ela permite modelar computacionalmente o raciocínio humano, impreciso e ambíguo, por meio de conjuntos fuzzy, variáveis linguísticas e regras fuzzy (Zadeh 1996). O uso de conjuntos fuzzy permite que os elementos tenham graus de pertinência, ou seja, os valores das variáveis podem ser qualquer número real entre 0 (absolutamente falso) e 1 (absolutamente verdadeiro) (Yager e Zadeh 2012). Variáveis linguísticas são substantivos cujos valores nomeiam conjuntos fuzzy. Por exemplo, “similaridade” pode representar uma variável linguística que assume os atributos “baixo” ou “alto”. As regras fuzzy combinam variáveis linguísticas por meio de conectores lógicos. Especificamente, o motor fuzzy usa regras fuzzy para inferir o valor da variável de saída. Esses conceitos são modelados por meio de notações de fácil entendimento, facilitando assim a especificação do conhecimento especializado necessário para identificar a mudança comportamental.

A solução de modelagem de conhecimento especializado requer a definição de três conjuntos *fuzzy*: (i) sensibilidade de detecção de mudanças; (ii) similaridade entre padrões e comportamentos; e (iii) *drift*, que representa o valor de saída do FIS. A subdivisão e atribuição de variáveis linguísticas a cada conjunto *fuzzy* são baseadas no conhecimento de especialistas. A Figura 17 apresenta um exemplo visual da modelagem desses conjuntos fuzzy. Por exemplo, é possível subdividir o conjunto de similaridade em três níveis (isto é, baixo, moderado e alto) com uma transição gradual de pertinência entre os intervalos.

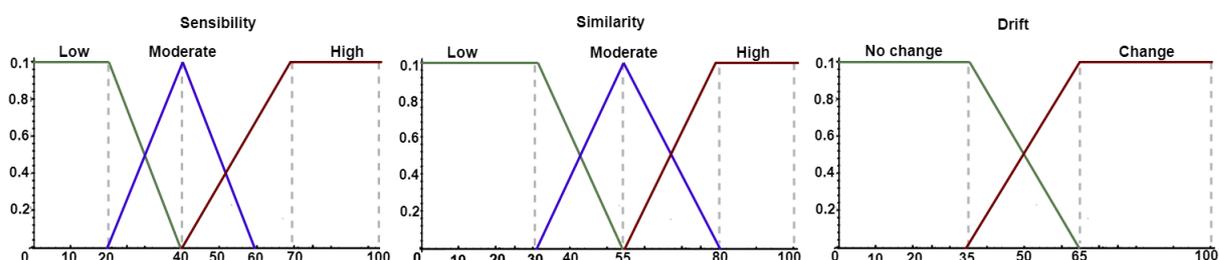


Figura 17 – Exemplo de conjuntos difusos para modelar mudanças comportamentais.

A solução proposta também exige a especificação de regras que norteiam a decisão do FIS. Essas regras possuem a estrutura básica < condição > *AND* < condição > *THEN* < consequência >. Assim, as regras definidas pelo especialista

¹ <<http://jfuzzylogic.sourceforge.net/>>

formarão a base de conhecimento utilizada pelo FIS para decidir o conjunto da variável de saída, assim como seu respectivo nível de pertinência. Demonstramos na Figura 18 um exemplo de implementação das regras baseadas nos conjuntos fuzzy especificados na Figura 17.

```
RULEBLOCK rules
  RULE 1 : IF sensibility IS low AND similarity IS low THEN drift IS change;
  RULE 2 : IF sensibility IS low AND similarity IS moderate THEN drift IS no_change;
  RULE 3 : IF sensibility IS low AND similarity IS high THEN drift IS no_change;
  RULE 4 : IF sensibility IS moderate AND similarity IS low THEN drift IS change;
  RULE 5 : IF sensibility IS moderate AND similarity IS moderate THEN drift IS no_change;
  RULE 6 : IF sensibility IS moderate AND similarity IS high THEN drift IS no_change;
  RULE 7 : IF sensibility IS high AND similarity IS low THEN drift IS change;
  RULE 8 : IF sensibility IS high AND similarity IS moderate THEN drift IS change;
  RULE 9 : IF sensibility IS high AND similarity IS high THEN drift IS no_change;
END_RULEBLOCK
```

Figura 18 – Regras fuzzy para mudança de comportamento.

4.5 Framework para a Fenotipagem Digital de Comportamentos

Nesta seção, apresenta-se o *framework* que implementa o algoritmo proposto e estrutura os padrões e mudanças comportamentais com base no modelo de dados desenvolvido. O *framework* projetado utiliza o *Apache Spark* para implementar o algoritmo proposto por meio de funções disponíveis para realizar seleções, derivações, junções, agregações e filtragens (Chambers e Zaharia 2018). O Spark foi escolhido devido a sua capacidade de processar grandes conjuntos de dados de forma paralela e distribuída, viabilizando utilizar a solução proposta em cenários com muitos usuários. Também integra-se a este *framework* um motor de inferência fuzzy para implementar a estratégia de detecção de comportamentos anômalos e mudanças de padrões. A ideia principal desse componente de software é baseada na estratégia de criação de jobs para realizar a detecção de padrões e mudanças comportamentais. Cada job criado é responsável por identificar padrões comportamentais de um determinado indivíduo, considerando uma situação de interesse e um conjunto de atributos de contexto. Os jobs criados são agendados para serem executado ao final de cada dia, quando serviços de detecção de padrões e mudanças comportamentais são acionados para processar a nova observação comportamental disponível (isto é, conjunto de eventos comportamentais do dia).

A Figura 19 apresenta a arquitetura de software do *framework* proposto. O primeiro componente é o *Scheduler and Job Management (SJB)*, responsável por

tarefas de gerenciamento dos jobs criados na solução, efetuando operações como criação, agendamento, pausa, atualização, e exclusão de jobs. Os componentes internos do SJB são: (i) *Instance ID Generator* atribui identificadores aos jobs criados; (ii) *Scheduler Job Info* contém informações sobre o job criado, como seu nome e grupo, além de dados referentes a detecção de padrões, como identificador do usuário, situação de interesse, conjunto de atributos de contexto e parâmetros do algoritmo; (iii) *Schedule Job Factory* é responsável por registrar e criar os jobs; e (iv) *Scheduler Job Service* fornece um conjunto de serviços, como iniciar, pausar, atualizar, deletar e verificar se um job está em execução. O segundo componente é o *Job Executer*, responsável pela execução efetiva dos jobs criados, orquestrando a execução dos serviços disponibilizados pela solução para detectar padrões e mudanças comportamentais. O terceiro componente é o *Pattern Inference and Change Detector*, composto por um conjunto de serviços responsáveis por executar tarefas específicas da solução, como detecção de padrões e mudanças de comportamentos e controle de observações. Por fim, o último componente é responsável por persistir os padrões e mudanças de comportamento, além de notificar as partes interessadas.

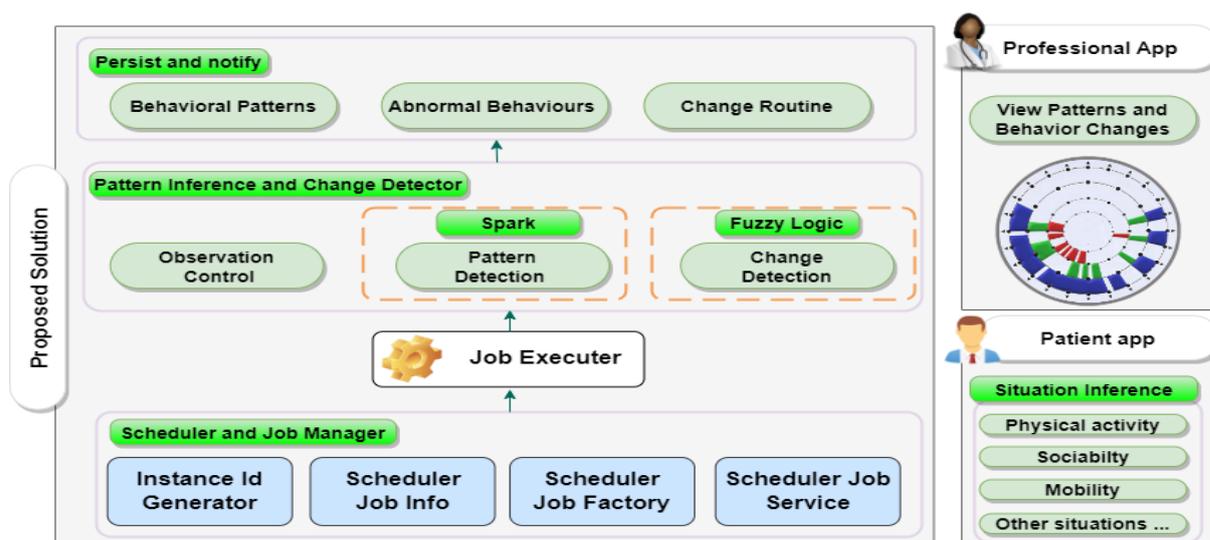


Figura 19 – Arquitetura de software do *framework* proposto.

4.5.1 Fluxo de Processamento

A solução proposta consiste em um fluxo de processamento bem definido para detectar padrões e mudanças comportamentais sensíveis ao contexto. A Figura 20 apresenta o diagrama de atividades que descreve o fluxo de processamento da solução. Primeiramente, o *Scheduler Job Management* agenda um novo job, com as respectivas informações de configuração (por exemplo, CAs, situação, parâmetros do algoritmo e número de observações para projetar um padrão), para ser executado toda vez que uma

observação ocorrer. Uma observação consiste no conjunto de eventos comportamentais que ocorreram em um determinado dia. Portanto, o job especificado será executado ao final de cada dia.

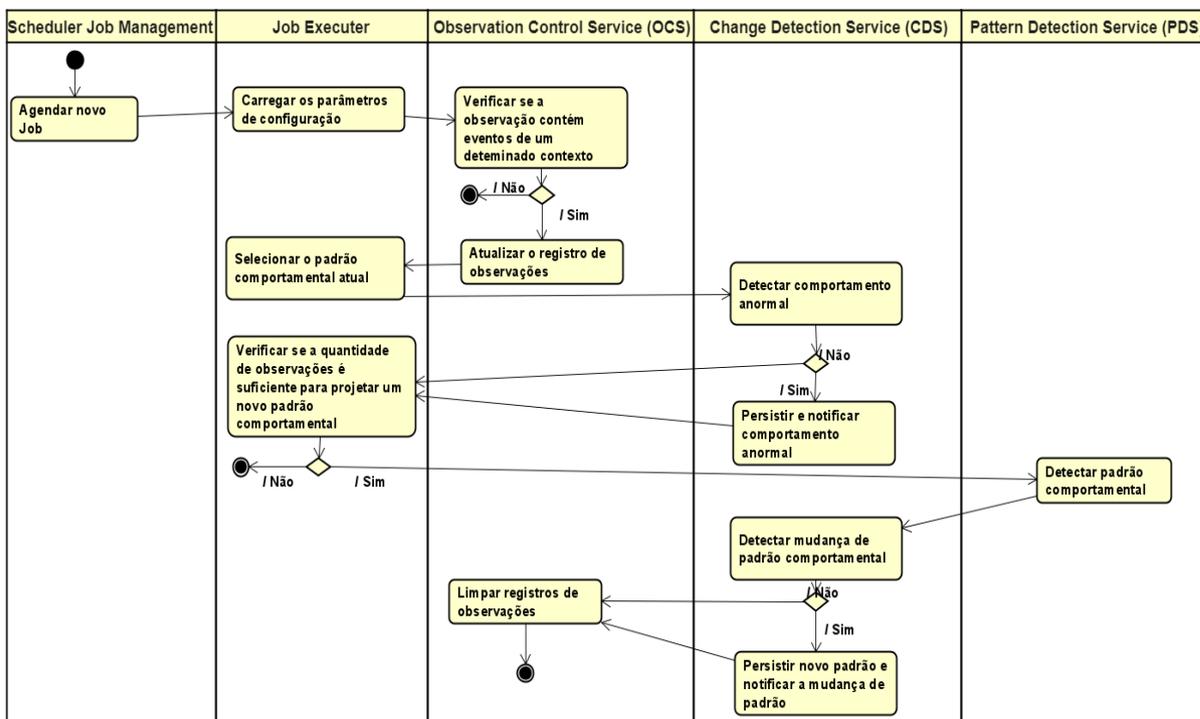


Figura 20 – Diagrama de atividades da solução proposta.

A princípio, o *Job Executer* carrega os parâmetros definidos no *Scheduler Job Management*. Em seguida, o *Observation Control Service (OCS)* verifica se o contexto da observação atual é de interesse para o job em execução, isto é, verifica se o contexto da observação está contido nos atributos de contexto especificados. Caso o contexto da observação não seja de interesse para o job, a execução é finalizada, caso contrário registra a ocorrência da observação. Esse registro é necessário, pois o job projeta um novo padrão comportamental quando um determinado número de observações ocorrer.

Neste momento, o *Job Executer* seleciona o padrão comportamental atual e encaminha-o para o *Change Detection Service (CDS)*, que por sua vez verifica a ocorrência de comportamento anormal comparando a similaridade entre esse padrão e a observação processada. Se o comportamento anormal for detectado, o CDS persiste e notifica esta informação. Logo após, o *Job Executer* consulta o registro de observações para identificar se existe quantidade de observações suficientes para projetar um novo padrão. Caso não exista, a execução é finalizada, caso contrário, o *Pattern Detection Service* projeta um novo padrão a partir dos eventos comportamentais das observações significantes registradas e encaminha-o para o CDS. Agora, o CDS compara a similaridade entre o padrão projetado e o padrão atual para detectar a

mudança de padrão comportamental. Se essa mudança for detectada, o CDS persiste e notifica essa informação. Por fim, o OCS limpa os registros de observações e finaliza a execução do job.

4.6 Considerações Finais

Este capítulo apresentou uma solução para a identificação de padrões comportamentais usando a fenotipagem digital. A princípio, foi apresentado o modelo formal para soluções de fenotipagem digital que visam identificar padrões comportamentais enriquecidos por contexto. Esse modelo utilizou símbolos contidos na teoria de conjuntos para formalizar as principais estruturas de soluções de fenotipagem digital. Em seguida, foi apresentado o modelo de dados projetado para apoiar o desenvolvimento da solução proposta. Além disso, foi introduzido o algoritmo proposto para realizar a detecção de padrões comportamentais enriquecidos por contexto através do processamento do fluxo eventos comportamentais. Também apresentou-se a abordagem de detecção de comportamento anormal e mudança de padrão comportamental, assim como a modelagem do conhecimento especialista necessário para executar essa tarefa. Por fim, apresentou-se o *framework* proposto para implementar o algoritmo proposto e automatizar o processo de detecção de padrões e mudanças comportamentais.

Considerando a classificação dos estudos relacionados apresentada na Seção 3.3.3, este trabalho está inserido na categoria que objetiva identificar informações de alto nível, uma vez que este utiliza eventos comportamentais derivados dos dispositivos ubíquos para reconhecer informações relevantes sobre o comportamento de indivíduos. Portanto, o conjunto de soluções apresentado neste capítulo pode ser utilizados para apoiar o processo de decisão médica fornecendo informações objetivas sobre o comportamento humano, podendo ser utilizado para complementar métodos clássicos de avaliação comportamental baseados em autorrelatos subjetivos.

4.6.1 Publicações

A solução proposta apresentada neste capítulo resultou em um conjunto de artigos científicos publicados e submetidos em periódicos e em conferência internacionais (Rodrigues de Moura et al. 2020, Moura et al. 2021, Moura et al. 2022). Abaixo, informa-se o tipo de publicação, status em que ela se encontra, o Qualis do periódico ou conferência (conforme classificação quadriênio 2017-2020), e fator de impacto se disponível.

- Moura, Ivan; Silva, Francisco; Coutinho, Luciano; Teles, Ariel. **Mental Health Ubiquitous Monitoring: Detecting Context-Enriched Sociability Patterns**

Through Complex Event Processing. IEEE 33rd International Symposium on Computer Based Medical Systems (CBMS), 2020 ([Rodrigues de Moura et al. 2020](#)).

Situação: Publicado. **Tipo de publicação:** Conferência internacional **Qualis:** A3.

- Moura, Ivan; Teles, Ariel; Endler, Markus; Coutinho, Luciano; Silva, Francisco. **Recognizing Context-Aware Human Sociability Patterns Using Pervasive Monitoring for Supporting Mental Health Professionals.** Sensors, 2021 ([Moura et al. 2021](#)).

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A2. **Fator de impacto:** 3.9.

- **Periódico:** Moura, Ivan; Teles, Ariel; Coutinho, Luciano; Silva, Francisco. **Towards Identifying Context-Enriched Multimodal Behavioral Patterns for Digital Phenotyping of Human Behaviors** Future Generation Computer Systems, 2022 ([Moura et al. 2022](#)).

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A1. **Fator de impacto:** 7.5.

5 Avaliação Experimental

Propomos neste trabalho de doutorado uma solução para reconhecer padrões de comportamento humano multimodal enriquecidos com contexto (Moura et al. 2022, Moura et al. 2021, Rodrigues de Moura et al. 2020). Especificamente, a solução proposta realiza aprendizado não supervisionado de padrões de comportamento humano. Essa característica dificulta o uso de métricas padrão para avaliar algoritmos de aprendizado de máquina, pois não há informações básicas disponíveis para medir os resultados. Portanto, pretendemos avaliar a solução proposta em termos de sua capacidade de projetar padrões enriquecidos por contexto capazes de modelar rotina comportamental, bem como sua sensibilidade para reconhecer mudanças comportamentais. Nessas avaliações, usou-se um conjunto de dados público para detectar padrões e mudanças de comportamento considerando vários aspectos do comportamento humano (ou seja, sociabilidade, estado estacionário e movimentação).

5.1 Descrição dos Dados

Usou-se o conjunto de dados público StudentLife (Wang et al. 2017) para criar os eventos comportamentais necessários para os experimentos propostos. Este conjunto de dados contém dados de sensores multimodais coletados de *smartphones* de 48 alunos de graduação e pós-graduação durante 66 dias. Além de dados brutos de sensores (por exemplo, localização GPS, registro de varredura de Bluetooth), o aplicativo também fez inferências de comportamentos, como atividades físicas e conversas. Processamos este conjunto de dados para gerar eventos de sociabilidade (*SociabilityEvent*) e atividade física (*ActivityEvent*) com base no modelo formal (Seção 4.1) e modelo de dados (Seção 4.2) propostos neste estudo.

Representamos cada amostra no conjunto de dados como um evento comportamental. Estruturamos *ActivityEvent* em dois tipos de eventos: *stationary* projetados a partir de inferências de estado estacionário e *moving* projetados a partir de inferências de estados andando e correndo. Projetamos *SociabilityEvent* utilizando eventos do tipo *conversation* projetados a partir de inferência de conversação. Por fim, usamos como atributo de contexto o dia da semana reconhecido a partir dos registros de data/hora do evento. Abaixo, apresentamos exemplos de eventos comportamentais gerados.

```

1 {
2     "uid" : "u00",
3     "startDateTime": ISODate("2023-03-22T10:50:33.000Z"),
4     "endDateTime": ISODate("2023-03-22T10:57:34.000Z")

```

```
5     "behavior" : "conversation",
6     "contexts" : [
7         "WEDNESDAY",
8         "WEEK"
9     ]
10 }
11 {
12     "uid" : "u00",
13     "startDateTime": ISODate("2023-03-23T14:00:22.000Z"),
14     "endDateTime": ISODate("2023-03-23T18:57:12.000Z")
15     "behavior" : "stationary",
16     "contexts" : [
17         "THURSDAY",
18         "WEEK"
19     ]
20 }
21 {
22     "uid" : "u00",
23     "startDateTime": ISODate("2023-03-24T20:00:15.000Z"),
24     "endDateTime": ISODate("2023-03-24T20:20:43.000Z")
25     "behavior" : "moving",
26     "contexts" : [
27         "FRIDAY",
28         "WEEK"
29     ]
30 }
```

5.2 Capacidade de Modelar Rotina Comportamental

Conduziu-se este experimento para avaliar a capacidade da solução proposta para gerar padrões comportamentais multimodais enriquecidos com contexto que modelem a rotina comportamental. O comportamento esperado para a solução proposta é gerar padrões comportamentais que apresentem um alto nível de similaridade com rotinas comportamentais estáveis (ou seja, os comportamentos se repetem ao longo do tempo) e menor nível de similaridade com rotinas instáveis. Portanto, este experimento consistiu em medir a associação entre a habilidade dos padrões multimodais em modelar rotinas comportamentais e a estabilidade da rotina dos indivíduos. Para isso, foi utilizado o coeficiente de correlação de Pearson ([Benesty et al. 2009](#)) para quantificar o nível de associação entre essas variáveis. Consequentemente, níveis mais altos de correlações positivas indicam uma maior capacidade da solução para modelar a rotina comportamental.

Inicialmente, conduziu-se um experimento para identificar o número de observações comportamentais necessárias para extrair padrões consistentes com

os hábitos comportamentais dos indivíduos. Nessa etapa, projetaram-se os padrões multimodais com uma, duas, três e quatro observações considerando como atributo de contexto cada dia da semana (por exemplo, sexta-feira, sábado e domingo) para reconhecer a semelhança média entre esses padrões e observações subsequentes. A Figura 21 apresenta o desenho experimental usando duas observações para extrair padrões. Especificamente, utilizou-se o coeficiente de *Jaccard* para medir a semelhança entre padrões e observações e, em seguida, calculamos a média dos resultados.

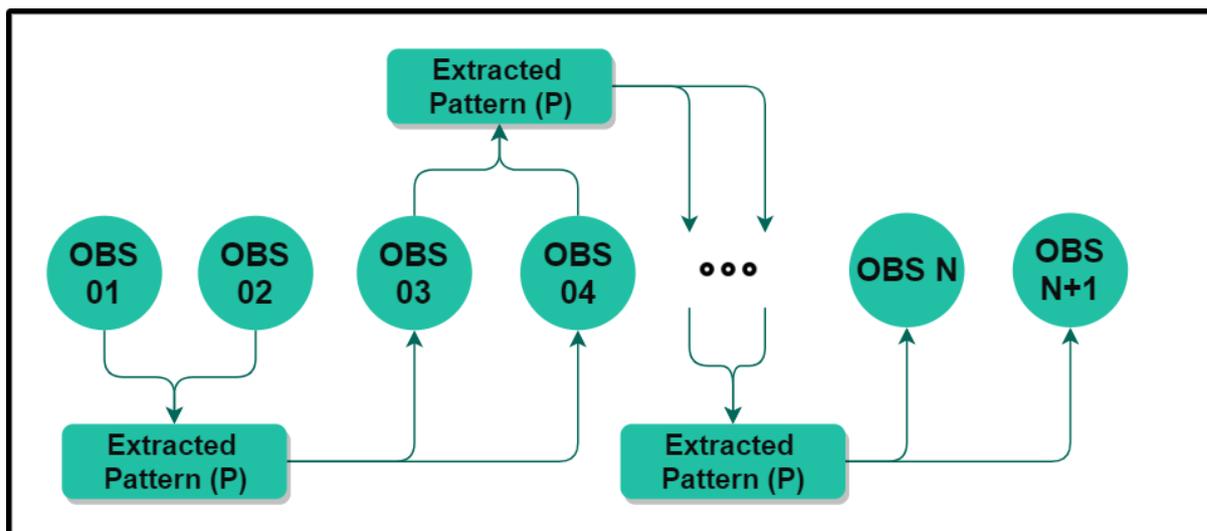


Figura 21 – Desenho experimental para avaliar o número de observação apropriado para projetar padrões.

A Figura 22 apresenta a similaridade média resultante para cada tipo de comportamento. Esses resultados demonstraram que a identificação de padrões usando uma observação resulta em uma baixa capacidade de modelar a rotina em comparação com outros cenários. Além disso, descobrimos que os outros cenários têm valores médios semelhantes. Com base nesses resultados, decidiu-se extrair padrões comportamentais multimodais de duas observações, uma vez que foi possível identificar padrões em menor quantidade de dias e com capacidade de modelagem semelhante ao usar mais do que duas observações.

Em uma etapa seguinte, calculou-se a estabilidade de rotina dos indivíduos. Especificamente, aplicou-se o coeficiente de similaridade de Jaccard entre observações comportamentais consecutivas e mensurou-se a média dos valores resultantes. Neste experimento, considera-se cada dia da semana como um CA. A Figura 23 mostra o desenho experimental desta avaliação. A realização deste experimento permitiu medir a estabilidade da rotina (ou seja, o nível médio de similaridade entre as observações) e, portanto, possibilitou sua correlação com a capacidade de modelagem apresentada pelos padrões.

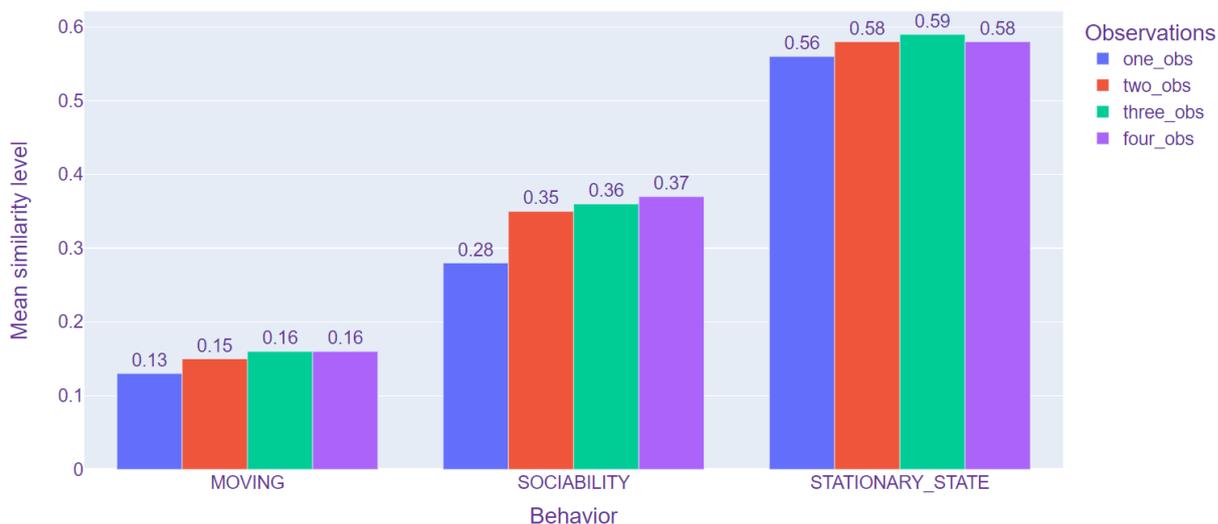


Figura 22 – Nível médio de previsão do comportamento.

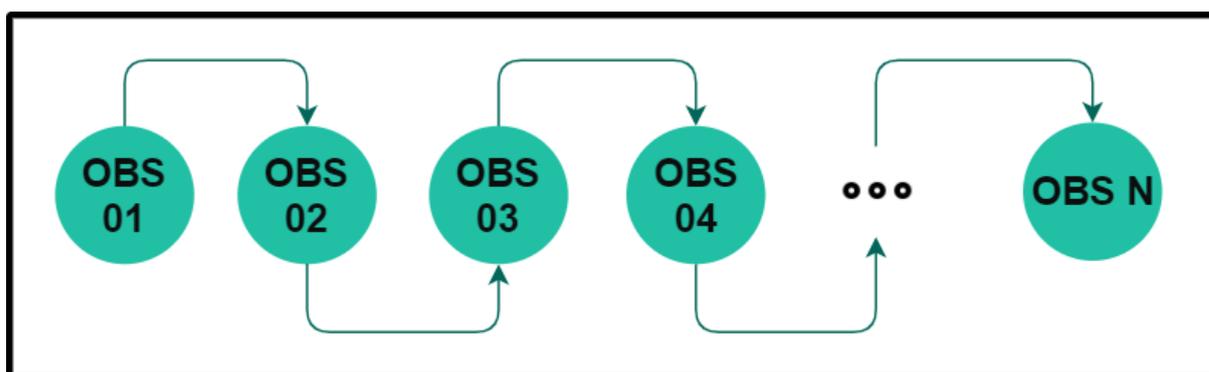


Figura 23 – Desenho experimental para avaliar a estabilidade da rotina comportamental.

A Figura 24 mostra a estabilidade média da rotina dos indivíduos para cada CA. Identificou-se alta estabilidade de rotina para estado estacionário, que apresentou uma média de $\approx 59\%$. Os comportamentos de movimentação e de sociabilidade mostraram uma estabilidade de rotina inferior, média de $\approx 17\%$ e $\approx 32\%$, respectivamente. Portanto, esperava-se que os padrões multimodais gerados pudessem modelar consistentemente estados estacionários e ter mais dificuldade ao modelar estados de movimentação e sociabilidade.

A Figura 25 mostra o gráfico de dispersão que relaciona o nível médio de similaridade (ou seja, a capacidade dos padrões de modelar rotinas) e a estabilidade média da rotina. Esta visualização permite inferir a existência de uma correlação linear entre as variáveis analisadas. Em seguida, quantificamos essa associação aplicando o coeficiente de correlação de Pearson, o que nos permitiu identificar a capacidade de

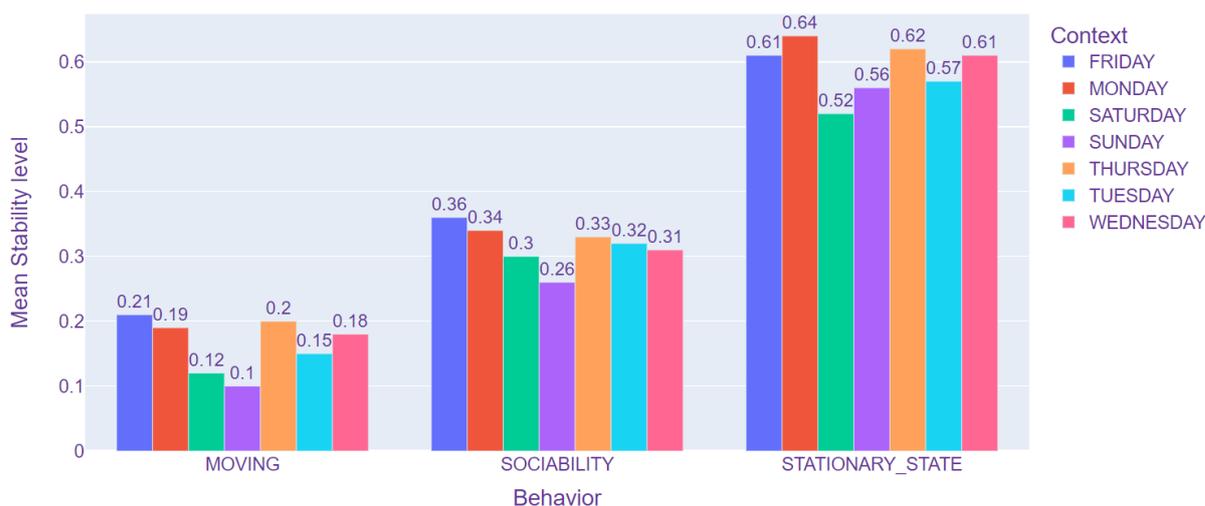


Figura 24 – Nível médio de estabilidade da rotina comportamental.

nossa solução em modelar rotinas comportamentais. Os resultados desta experiência demonstraram uma forte associação positiva para todos os comportamentos analisados (*Sociability* $\approx +0.92$, *Stationary State* $\approx +0.67$, *Moving* $\approx +0.8$), indicando que a solução modela adequadamente as rotinas comportamentais dos indivíduos.

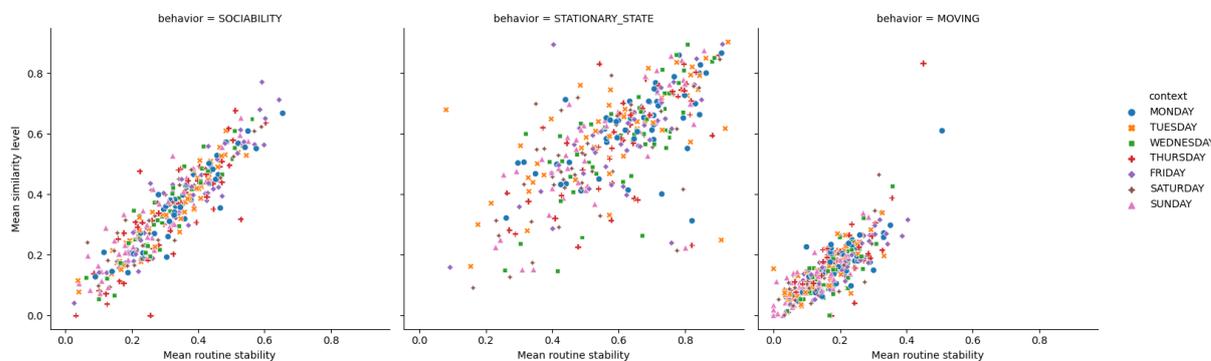


Figura 25 – Associação entre a capacidade de modelagem apresentada pelos padrões e a estabilidade média da rotina.

A Figura 26 mostra o nível de correlação resultante para cada CA. Identificamos que os comportamentos analisados apresentaram forte correlação positiva em todos os contextos. Esses resultados reforçam a conclusão de que nossa solução pode modelar adequadamente as rotinas comportamentais dos indivíduos.

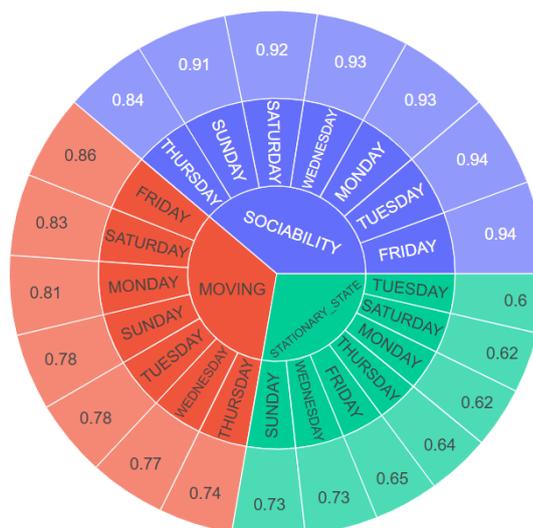


Figura 26 – Correlação entre a capacidade média de modelagem e a estabilidade média de rotina para cada CA.

5.3 Sensibilidade em Mudanças de Padrões Comportamentais

Nesta seção, apresentamos o experimento para avaliar a sensibilidade de detecção de mudanças de comportamento. Para isso, projetamos fluxos de dados com amostras de comportamentos anormais e mudanças de rotina. A Figura 27 apresenta os conjuntos fuzzy modelados e as regras fuzzy instanciadas no sistema de inferência fuzzy da solução proposta. Esta parte da solução será responsável por detectar os comportamentos anormais e mudanças de padrões ao processar os fluxos de dados. Ao final, analisamos se a solução apresentou sensibilidade adequada para reconhecer corretamente todas as mudanças de padrões e comportamentos anormais.

5.3.1 Mudanças nos Padrões de Estado Estacionário

Nesta etapa, projetamos o fluxo de eventos de estado estacionário combinando os dados do usuário u08 (u = usuário; número = identificação do participante no conjunto de dados StudentLife) no contexto *Monday* e u44 no contexto *Wednesday*, que possuem rotinas estáveis e significativamente diferentes. A Figura 28 mostra o número de eventos em cada slot de tempo (ou seja, slots de 10 minutos) no fluxo de dados gerado. Como a rotina do u08 é estável, nossa solução não deve detectar alterações ao processar esses dados do usuário. Entretanto, ao começar a processar dados do usuário u44, esperamos que a solução detecte observação que represente um comportamento anormal e, em seguida, detecte a mudança de padrão. Após reconhecer o novo padrão dos dados do u44, a solução não deve identificar comportamentos

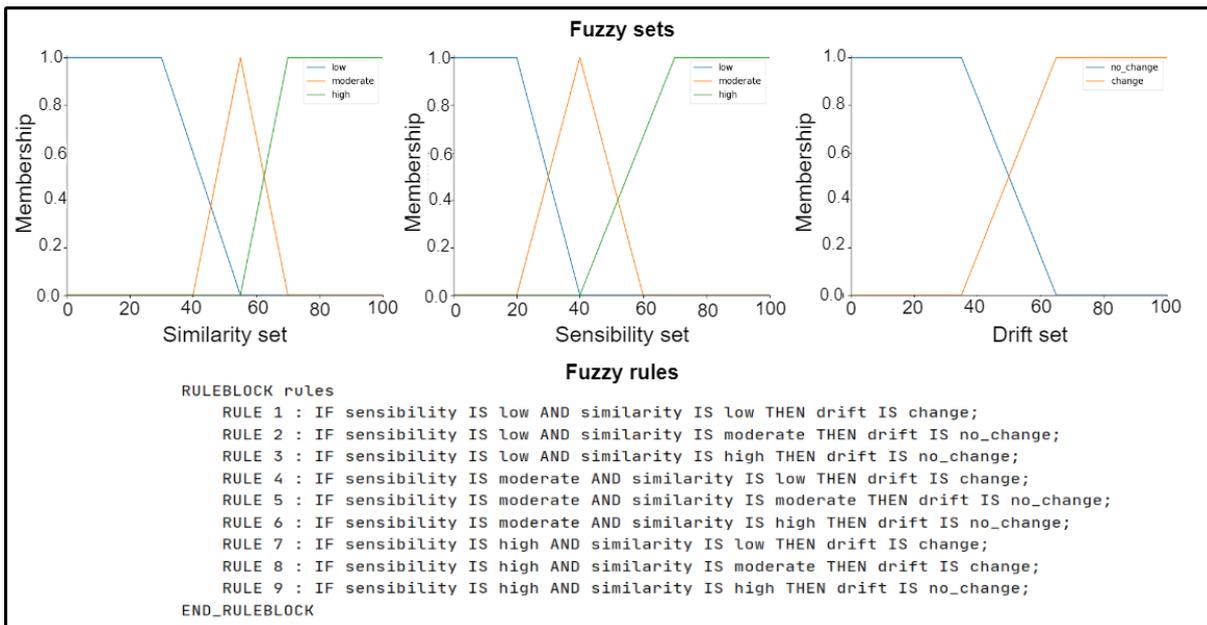


Figura 27 – Conjuntos e regras fuzzy utilizadas nos experimentos de sensibilidade de mudança.

anormais ou alterações de padrões porque os dados do segundo usuário têm uma rotina estável.

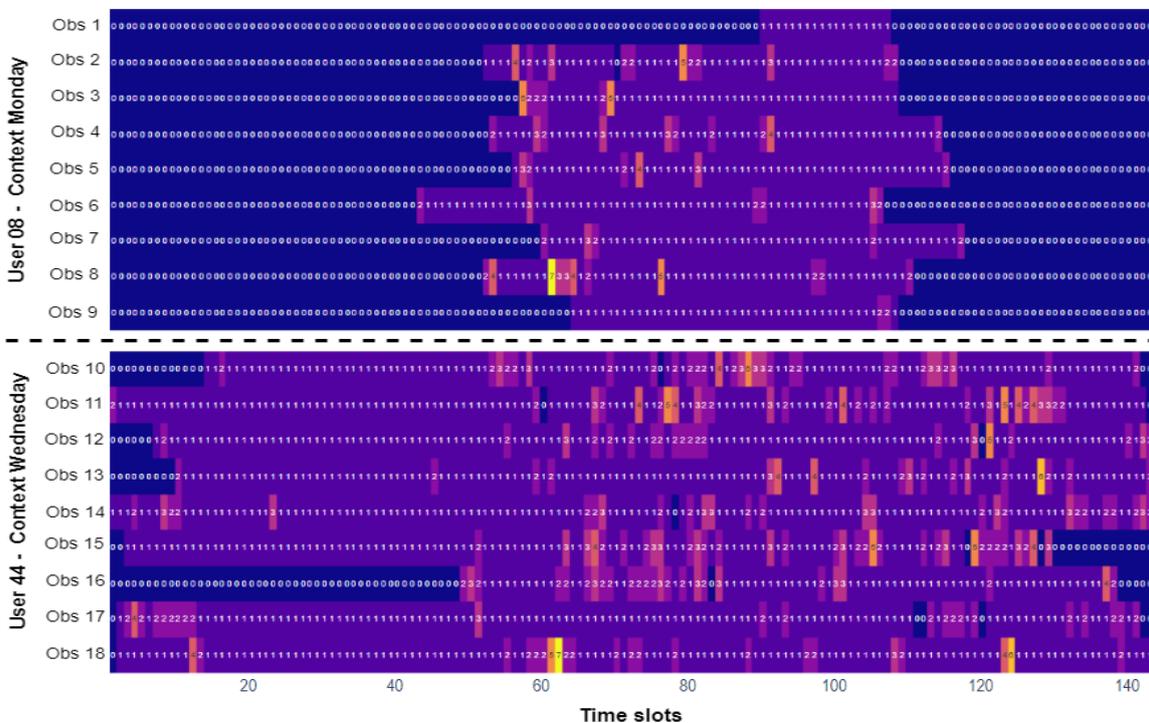


Figura 28 – Mesclagem dos dados de estado estacionário dos usuários u08 e u44.

A Figura 29 mostra as detecções de mudanças de comportamento realizadas

pela solução ao processar o fluxo de dados gerado. Ao processar os dados do u08, a solução não detectou nenhum comportamento anormal ou mudança de padrão. Especificamente, todas as detecções apresentaram grau de pertinência de 100% ao intervalo “no change” do conjunto fuzzy drift, representando alta confiança na detecção de não mudança de comportamento. Ao iniciar o processamento dos dados do u44, a solução identificou um comportamento anormal na Obs10 (Obs = observação; número = identificação da sequência), pois a similaridade entre essa observação e o padrão atual resultou em uma pertinência de 83% ao conjunto “change”. Ao projetar o padrão comportamental a partir da Obs 9 e Obs 10 (ou seja, 5º padrão), a solução reconheceu que houve uma mudança de padrão, uma vez que a similaridade resultante foi de 44%, o que representa 73% de pertinência ao intervalo “change” do conjunto drift. Esse novo padrão apresentou alta similaridade com as observações subsequentes, não implicando, portanto, em nenhuma nova detecção de comportamento anormal (ou mudança de padrão). Essas detecções coincidem com as expectativas ao analisar o fluxo de dados gerado, mostrando assim que a solução proposta tem sensibilidade para identificar comportamentos anormais e mudanças nas rotinas de estado estacionário.

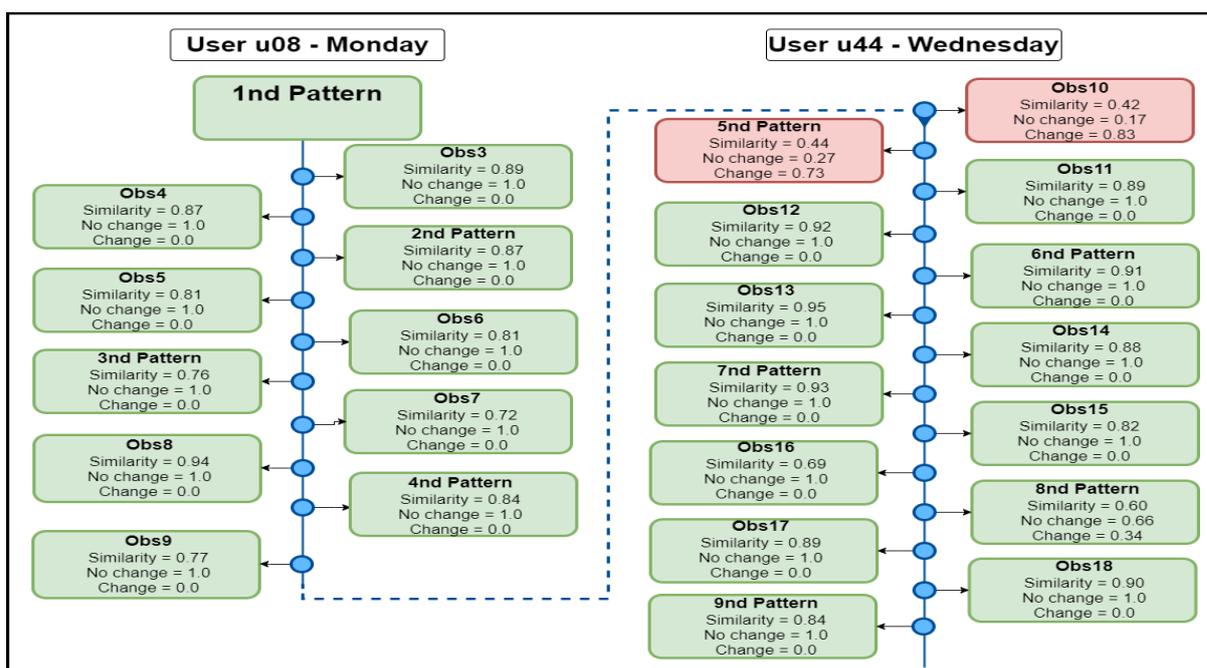


Figura 29 – Fluxo de detecção de mudança no comportamento de estado estacionário.

5.3.2 Mudanças nos Padrões de Sociabilidade

A Figura 30 mostra o fluxo de eventos de sociabilidade gerado a partir da combinação dos dados do usuário u01 no contexto *Monday* e u05 no contexto *Tuesday*. O comportamento de sociabilidade do usuário u01 é estável, enquanto o usuário u05

possui uma rotina instável. Analisando o fluxo de dados gerado, é possível reconhecer que, apesar da rotina do segundo usuário ser instável, as duas últimas observações apresentam uma similaridade significativa. Assim, neste experimento, objetivamos simular uma mudança de comportamento em que o indivíduo tenha uma rotina estável e, posteriormente, passe a variar seu comportamento com frequência, apresentando comportamentos anormais e mudanças de padrões.

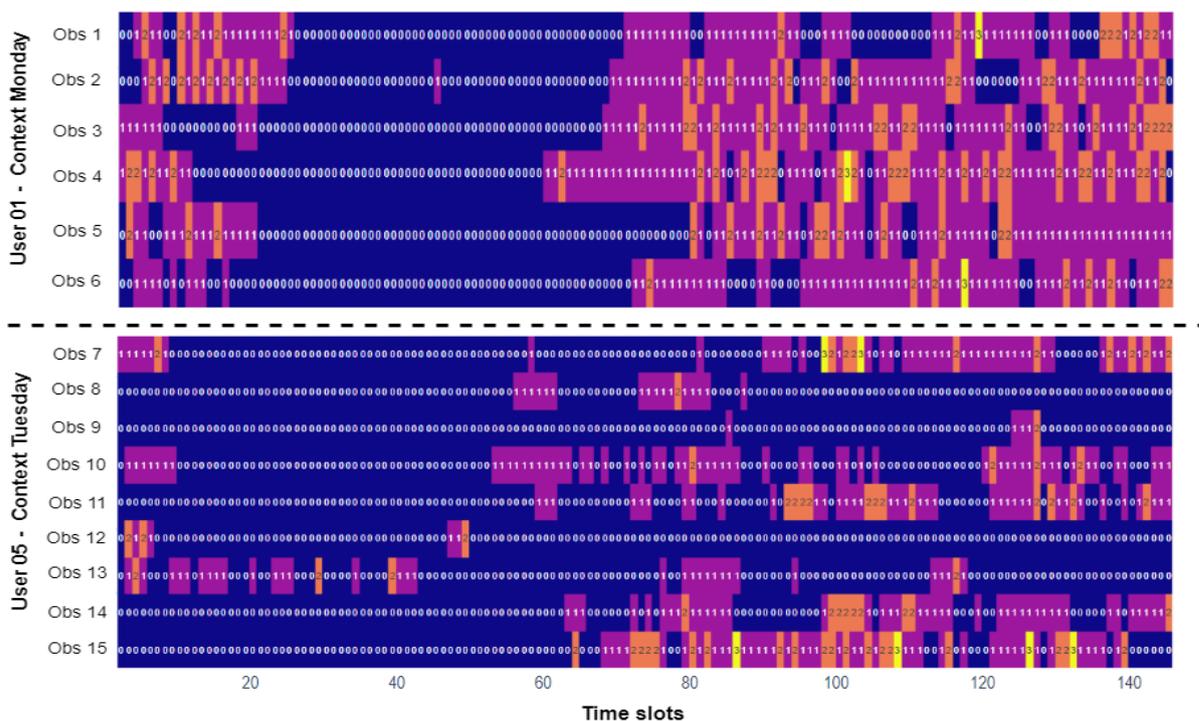


Figura 30 – Mesclagem dos dados de sociabilidade dos usuários u01 e u05.

A Figura 31 mostra comportamentos anormais e mudanças de rotina identificadas pela solução proposta ao processar o fluxo de dados de sociabilidade projetado (Figura 30). Reconhecemos que ao processar os dados do u01, a solução não reconheceu nenhum comportamento anormal ou mudança de padrão. Em seguida, o processamento dos dados do u05 resultou em muitas detecções de mudanças comportamentais, nas quais apenas a Obs15 apresentou similaridade adequada com seu padrão de referência (ou seja, 7º padrão). A maioria das detecções de comportamentos anormais apresentaram 100% de pertinência ao conjunto “change”, em que apenas a detecção da Obs7 apresentou pertinência de 57% a esse conjunto. Em relação às mudanças de padrões, ao analisar a detecção do 4º e 7º padrão, identificou-se pertinência ao conjunto “change” abaixo de 100%, resultando em 71% e 88% de pertinência, respectivamente. Além disso, a Obs15 representa um comportamento normal, uma vez que apresentou uma similaridade de 49% com o seu padrão de referência, representando pertinência de 66% ao intervalo “no change”.

Esses resultados mostram que nossa solução tem sensibilidade para identificar mudanças nos comportamentos de sociabilidade.

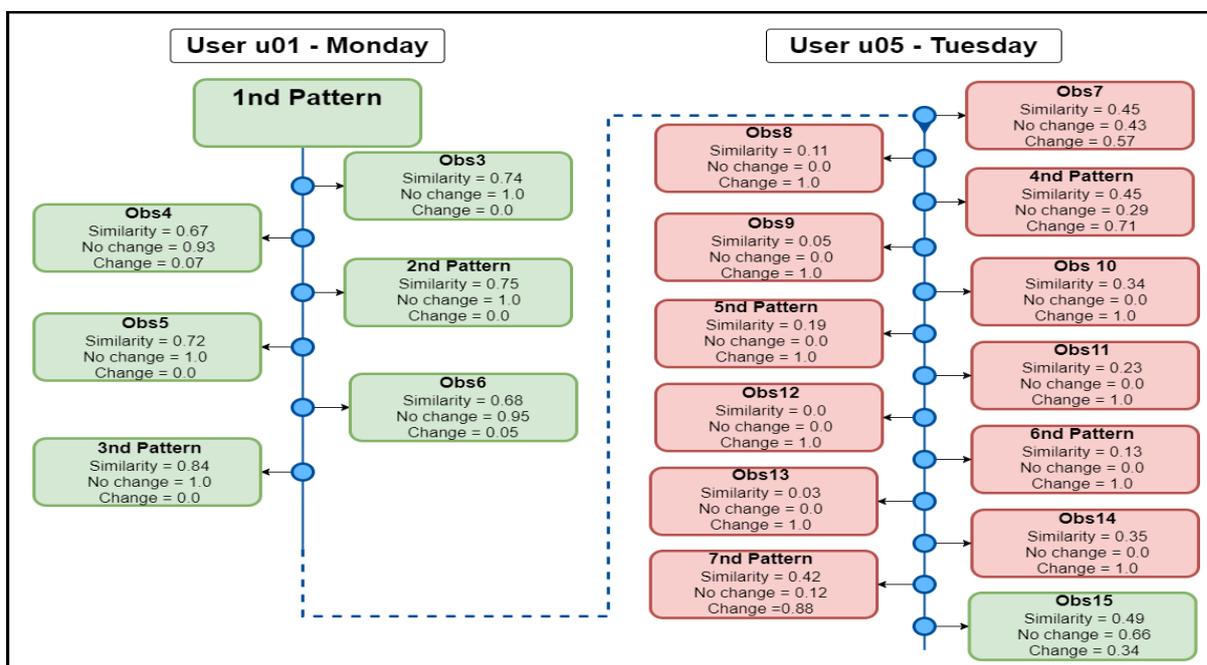


Figura 31 – O fluxo de detecção de mudanças nos comportamentos de sociabilidade.

5.3.3 Mudanças nos Padrões de Movimentação

Neste experimento projetamos um fluxo de dados a partir dos eventos de movimentação do u00 no contexto *Tuesday* e u31 no contexto *Tuesday*, os quais possuem rotinas com alta instabilidade. Usamos esse fluxo de dados para avaliar a sensibilidade da solução para detectar mudanças nos padrões de movimentação. A Figura 32 mostra a mesclagem dos dados de movimentação dos dois indivíduos. Assim, a partir desta visualização, é possível reconhecer uma clara instabilidade na rotina dos usuários. Neste cenário, espera-se que a solução proposta seja capaz de detectar todos os comportamentos anormais e mudanças de padrões que acontecem constantemente no fluxo de dados projetado.

A Figura 33 mostra as mudanças de padrões e comportamentos anormais detectadas pela solução ao processar o fluxo de dados gerado. Ao avaliar as detecções realizadas, reconheceu-se que a solução proposta se comportou de maneira consistente, reconhecendo comportamentos anormais e mudanças de rotinas de movimentação de maneira precisa. Especificamente, a solução reconheceu que todas as observações representam comportamentos anormais e todos os padrões projetados representam uma nova rotina do usuário. Em relação às pertinências no conjunto drift, todas as detecções representaram 100% de pertinência ao intervalo “change”.

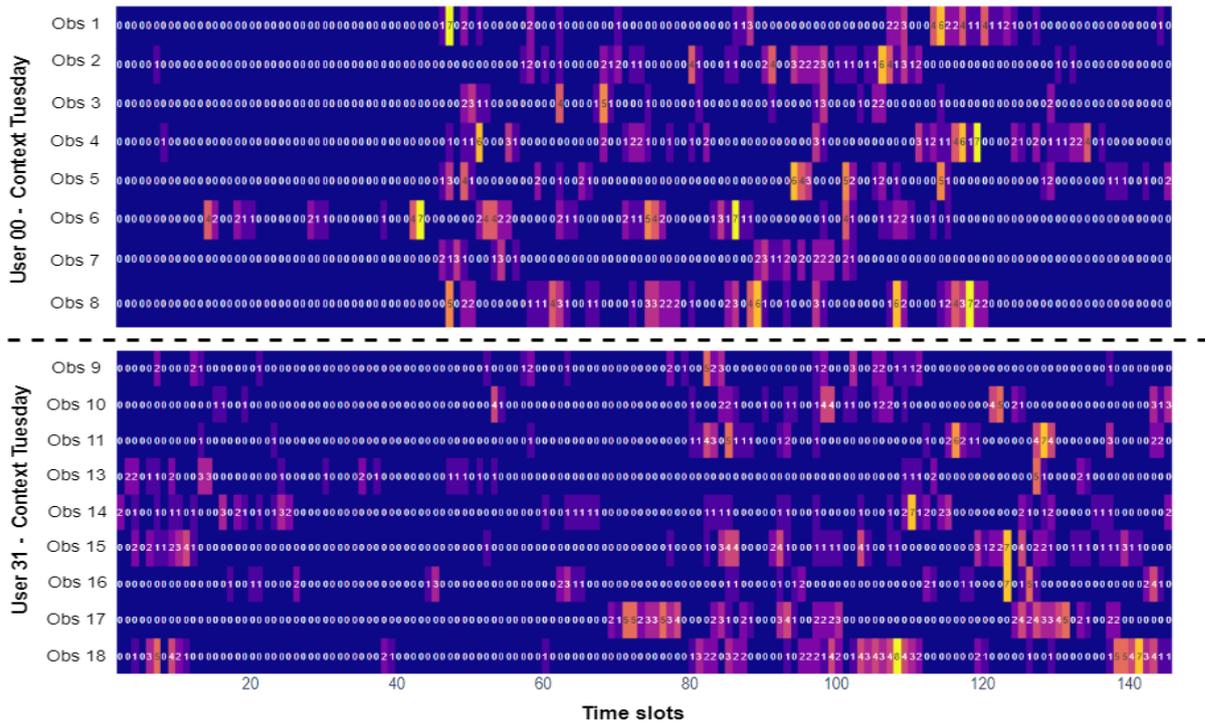


Figura 32 – Mesclagem dos dados de movimentação dos usuários u00 e u31.

Portanto, baseado nesse resultado, reconhecemos que a solução apresentou uma sensibilidade adequada para reconhecer mudanças comportamentais nesse cenário de comportamentos instáveis.

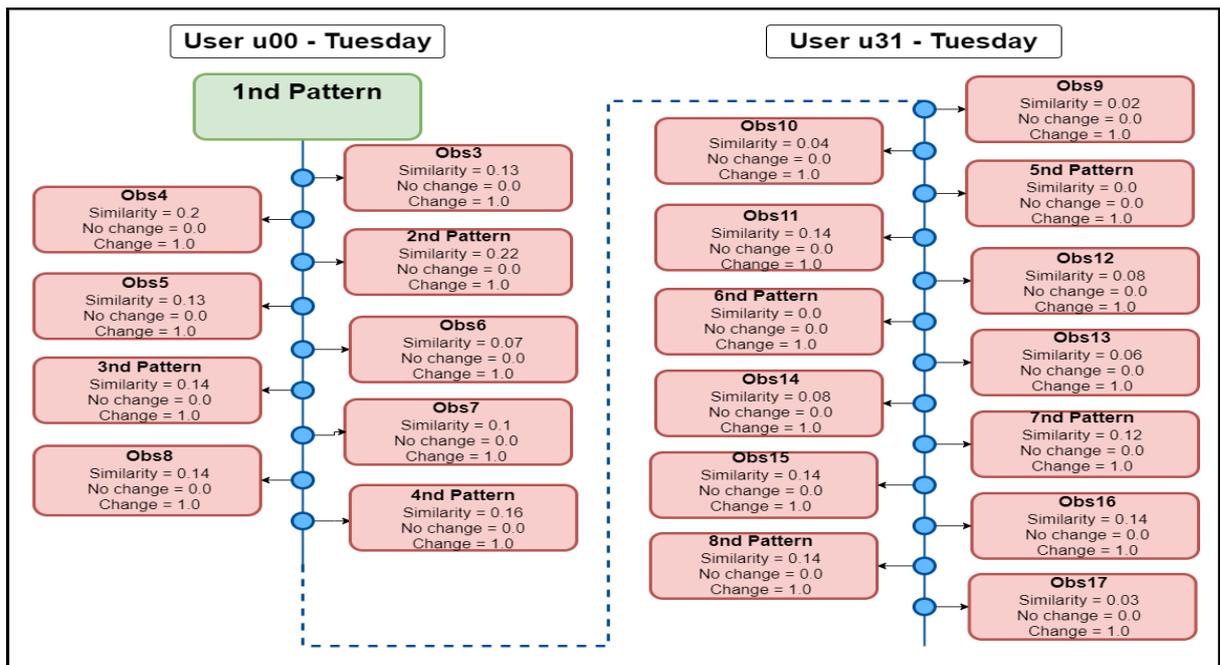


Figura 33 – O fluxo de detecção de mudanças nos comportamentos de movimentação.

5.4 Considerações Finais

Este capítulo apresentou uma avaliação experimental da solução proposta. Primeiramente analisamos a capacidade da solução proposta para reconhecer padrões que modelam adequadamente a rotina dos indivíduos. Basicamente, correlacionamos a similaridade entre padrões e observações comportamentais com a estabilidade de rotina dos indivíduos. Os resultados dessa análise destacam alta correlação positiva entre a capacidade do padrão de explicar rotinas e a estabilidade de rotina para todos os atributos de contexto analisados (*Sociability* $\approx +0.92$, *Stationary state* $\approx +0.67$, *Moving* $\approx +0.8$). Por fim, avaliamos a sensibilidade da detecção de mudança comportamental. Projetamos fluxos de dados a partir dos eventos comportamentais de usuários que apresentam diferentes tipos de rotinas (isto é, rotina estáveis e instáveis) e, em seguida, executamos a solução proposta para processar esses fluxos de dados. A solução proposta foi capaz de detectar corretamente todos os comportamentos anormais e mudanças nas rotinas para eventos de estado estacionário, sociabilidade e movimentação. Esses resultados evidenciam uma sensibilidade adequada para a detecção de mudanças comportamentais.

6 Análise da Influência dos Hiperparâmetros

Neste capítulo, pretendemos analisar a influência dos hiperparâmetros do algoritmo proposto na identificação de padrões de comportamento humano. Especificamente, visamos gerar diretrizes para que o usuário defina os melhores valores. A análise apresenta as seguintes etapas: (i) análise das características dos eventos comportamentais dos indivíduos; (ii) análise do tamanho do slot de tempo; (iii) análise da parametrização do limiar de slots candidatos; e (iv) análise da parametrização do limiar de intervalos frequentes.

6.1 Características dos Eventos Comportamentais

Nesta etapa, analisamos os eventos comportamentais gerados a partir do conjunto de dados. Especificamente, buscamos reconhecer a duração média dos eventos comportamentais de cada indivíduo. A tabela 11 apresenta duração média, máxima e mínima, desvio padrão e primeiro, segundo e terceiro quartis considerando todos os eventos de conversação. A duração média é de $\approx 10,48$ minutos, o evento mais longo é de $\approx 21,14$ minutos e o evento mais curto é de $\approx 4,50$ minutos. O primeiro quartil indica que 25% dos eventos duram até $\approx 7,96$ minutos, o segundo quartil indica que 50% dos eventos duram até $\approx 9,93$ minutos e o terceiro quartil indica que 75% dos eventos duram até $\approx 12,71$ minutos. Essas estatísticas reforçam a conclusão de que os eventos de conversação têm durações curtas. Assim, usamos eventos de conversação nas etapas posteriores desta análise para avaliar o comportamento da solução ao processar eventos de curta duração.

Tabela 11 – Estatísticas descritivas de eventos de conversação.

Estatística	Valor
Média	10.485678
Desvio Parão	3.280530
1° quartil (25%)	7.968035
2° quartil (50%)	9.931746
3° quartil (75%)	12.714046
Valor mínimo	4.504613
Valor máximo	21.145767

A Tabela 12 apresenta estatísticas descritivas de eventos de estado estacionário. A duração média é de $\approx 37,51$ minutos, o evento mais longo é de $\approx 271,86$ minutos e o evento mais curto é de $\approx 2,38$ minutos. O primeiro quartil indica que 25% dos eventos

duram até $\approx 19,87$ minutos, o segundo quartil indica que 50% dos eventos duram até $\approx 27,53\%$ e o terceiro quartil indica que 75% dos eventos são de até $\approx 44,43$ minutos. Esses valores sugerem que eventos de estado estacionário possuem longa duração e podem ser usados para analisar o comportamento da solução ao processar eventos de longa duração.

Tabela 12 – Estatísticas descritivas de eventos de estado estacionário.

Estatística	Valor
Média	37.519340
Desvio padrão	32.924682
1º quartil (25%)	19.873567
2º quartil (50%)	27.534785
3º quartil (75%)	44.433644
Valor mínimo	2.385085
Valor máximo	271.867949

6.2 Análise do Tamanho de Slot de Tempo

Nesta etapa, pretende-se reconhecer o comportamento da solução usando diferentes tamanhos de slot para detectar padrões comportamentais. A Figura 34 mostra a configuração do experimento para avaliar o tamanho do slot. Primeiro, projetamos observações comportamentais usando slot de tempo de 1 minuto, representando a verdade básica, pois essas observações se ajustam nos eventos comportamentais. Em seguida, aplicamos o coeficiente de similaridade de Jaccard entre a verdade básica e as observações comportamentais projetadas usando slots com 5, 10 e 30 minutos.

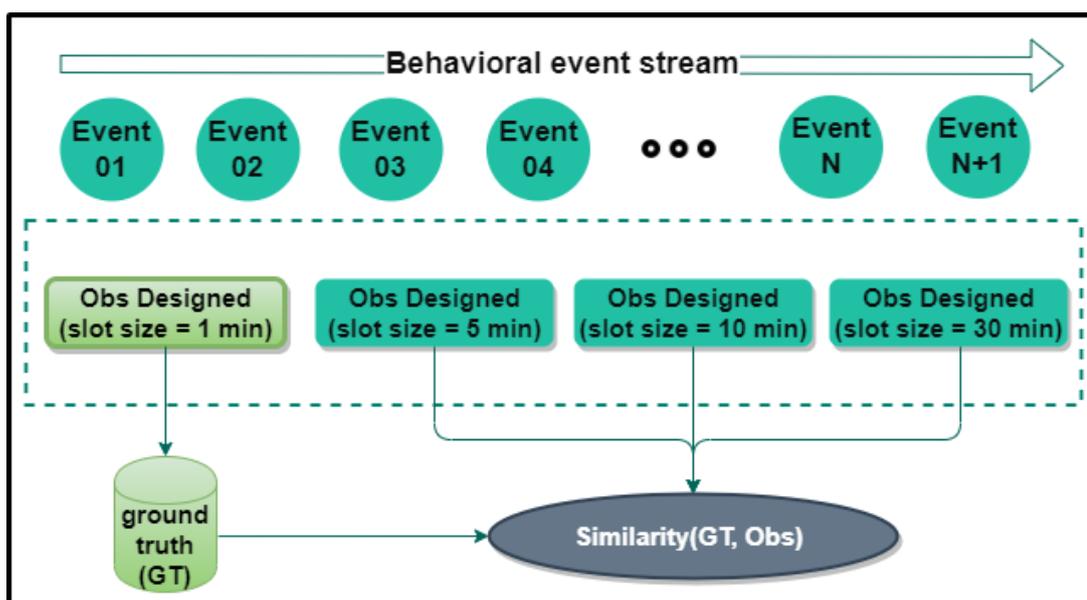


Figura 34 – Configuração do experimento para avaliar o tamanho do slot de tempo.

A Figura 35 apresenta a similaridade média e mediana entre a verdade básica e observações de sociabilidade projetadas usando slots de 5, 10 e 30 minutos. A partir desses resultados identificamos que usar tamanho de slot próximo à duração média dos eventos comportamentais (isto é, slots de 5 e 10 minutos) resulta em observações com alta similaridade com a verdade básica (isto é, similaridade $\geq 70\%$), ao passo que usar slot maior que a média de duração dos eventos resulta em baixa similaridade com a verdade básica (isto é, similaridade $\leq 60\%$).

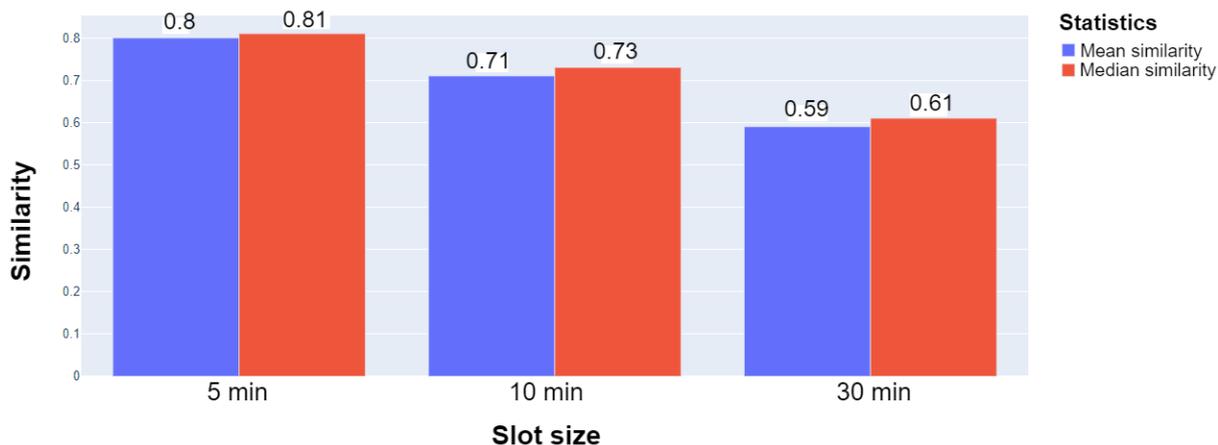


Figura 35 – Similaridade média/mediana entre a verdade básica e observações de sociabilidade.

A Figura 36 apresenta a similaridade média e mediana entre a verdade básica e as observações de estado estacionário projetadas usando slots com 5, 10 e 30 minutos. Esses resultados mostram uma alta similaridade em todos os tamanhos de intervalo de tempo analisados. Portanto, ao considerar que os eventos de estado estacionário têm uma duração média ($\approx 37,51$ minutos) maior que o tamanho dos slots utilizados, observamos que, em tal cenário, o tamanho do slot não interferiu na capacidade das observações de modelar o que aconteceu adequadamente.

Em relação ao tamanho de slot de tempo, observamos que quando os eventos comportamentais têm durações curtas, usar um tamanho de slot próximo à duração média dos eventos resulta em observações que se ajustam à verdade básica, enquanto usar valores maiores que a média da duração resulta em observações que diferem significativamente da verdade básica. Além disso, quando os eventos comportamentais têm durações longas, o tamanho do slot não interfere significativamente na capacidade das observações projetadas de modelar a verdade básica.

6.3 Análise do Limiar de Slot de Candidato

Nesta etapa, objetivamos identificar o comportamento da solução ao usar diferentes valores para o parâmetro theta (i.e., θ). A Figura 37 apresenta a configuração

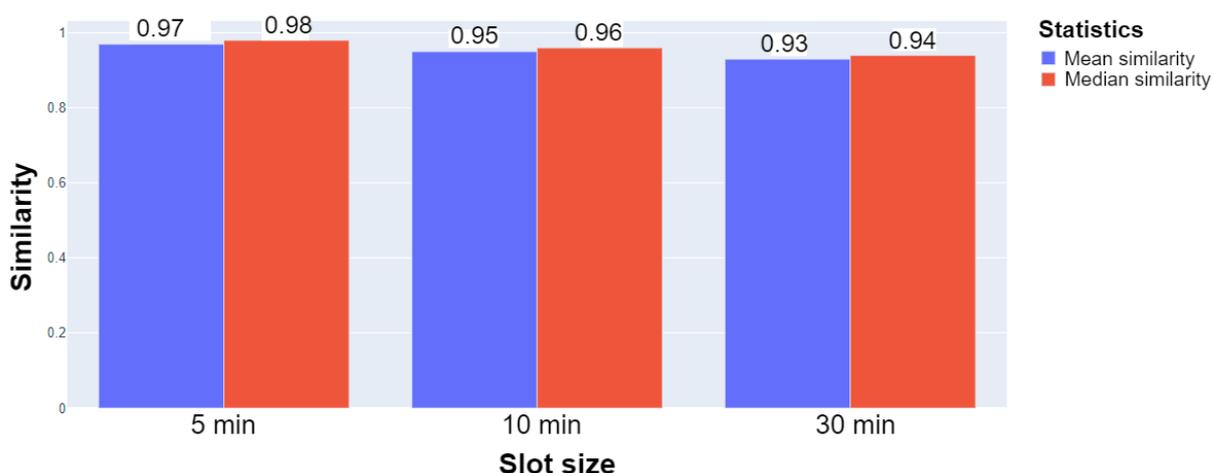


Figura 36 – Similaridade média/mediana entre a verdade básica e observações de estado estacionário.

do experimento conduzido que consiste em calcular a similaridade entre observação comportamental e slots candidatos (i.e., slots que satisfazem o limiar de slot candidato). Executamos esse experimento com θ igual a 0.6, 0.8 e 1.0 e tamanho de slots de 5, 10 e 30 minutos. No final, calculamos estatísticas descritivas sobre as similaridades resultantes entre observações e slots candidatos.

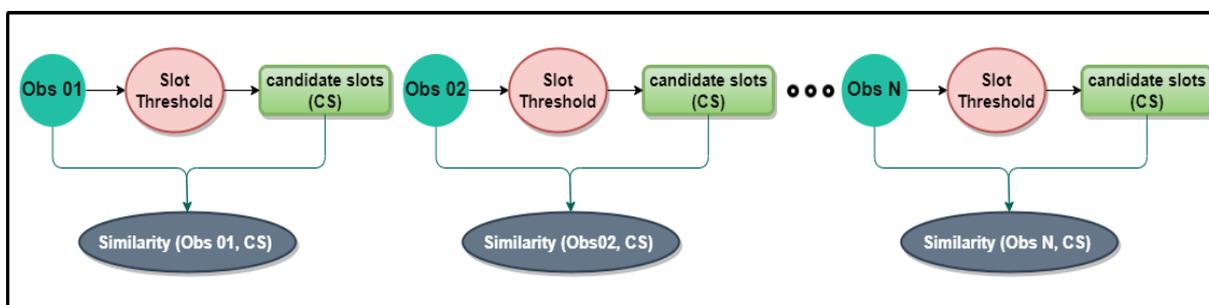


Figura 37 – Configuração do experimento para avaliar o valor do parâmetro theta (θ).

A Tabela 13 apresenta estatísticas descritivas sobre a similaridade entre observações de sociabilidade e slots candidatos usando valores θ iguais a 0.6, 0.8 e 1.0. O cenário com θ igual a 0.6 e 0.8 resultou em similaridade média de $\approx 97\%$ e 93% , respectivamente. Assim, é possível concluir que esses valores resultam em limiares que removem poucos slots, uma vez que a maioria dos slots nas observações são selecionados como slots candidatos para formar um padrão comportamental. Além disso, observamos que cenários com θ igual a 1.0 possuem similaridade média de $\approx 86\%$. Com base nesse resultado, observamos que o uso de θ igual a 1.0 gera limiares que filtram mais slots do que valores de θ iguais a 0.6 e 0.8.

A Figura 38 apresenta a similaridade média entre as observações de sociabilidade e slots candidatos considerando 5, 10 e 30 minutos. Esses resultados

Tabela 13 – Estatísticas sobre a similaridade entre observações comportamentais de sociabilidade e slots candidatos.

Estadística	$\theta = 0.6$	$\theta = 0.8$	$\theta = 1.0$
Média	0.971896	0.935568	0.869967
Desvio padrão	0.072559	0.121307	0.179707
1° quartil (25%)	1.000000	0.916667	0.719531
2° quartil (50%)	1.000000	1.000000	1.000000
3° quartil (75%)	1.000000	1.000000	1.000000
Valor mínimo	0.517037	0.313244	0.190244
Valor máximo	1.000000	1.000000	1.000000

mostram que quanto maior o tamanho do slot e o valor θ , mais slots são removidos pelo limiar de slot candidato resultante. Por exemplo, observamos que usar slots com 30 minutos e θ igual a 1.0 resultou em uma similaridade de $\approx 76\%$ enquanto usar slots com 5 minutos e θ igual a 0.6 resultou em uma similaridade de 100% (isto é, todos os slots foram selecionados como candidatos). No entanto, considerando esse cenário com eventos de curta duração (ou seja, sociabilidade), a similaridade média permaneceu alta em todos os cenários (ou seja, acima de 75%), indicando que, para eventos de curta duração, os limiares selecionam um número moderado de slots candidatos.

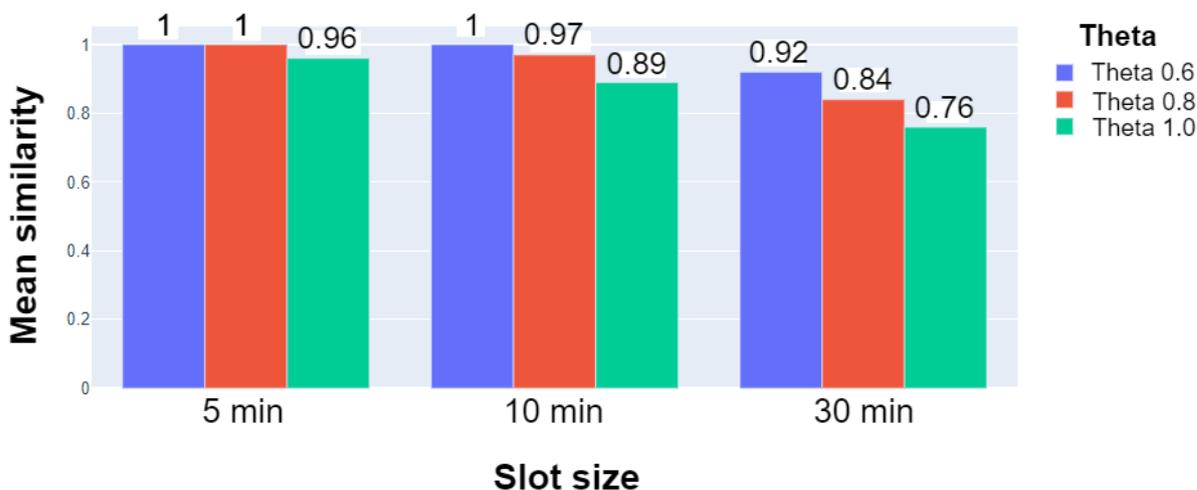


Figura 38 – Similaridade entre observações comportamentais de sociabilidade e slots candidatos usando slots de 5, 10 e 30 minutos.

A Tabela 14 apresenta estatísticas descritivas sobre a similaridade entre observações de estado estacionário e slots candidatos. O cenário com θ igual a 0.6 resultou em similaridade média de $\approx 90\%$, representando limiares que elegem muitos slots candidatos. O cenário com θ igual a 0.8 resultou em similaridade média de $\approx 76\%$, representando limiares que selecionam um número moderado de slots candidatos. Por fim, o cenário com θ igual a 1.0 resultou em uma similaridade média de $\approx 62\%$, representando limiares que selecionam uma quantidade menor de slots de candidatos. Dessa forma, concluímos que quanto maior o valor de θ , menor o número de slots

selecionados como candidatos a formar um padrão comportamental.

Tabela 14 – Estatísticas sobre a similaridade entre observações comportamentais de estado estacionário e slots candidatos

Estatística	$\theta = 0.6$	$\theta = 0.8$	$\theta = 1.0$
Média	0.905228	0.767727	0.627535
Desvio Padrão	0.148191	0.221396	0.266922
1° quartil (25%)	0.858863	0.559122	0.404542
2° quartil (50%)	1.000000	0.833333	0.595348
3° quartil (75%)	1.000000	1.000000	0.880053
Mínimo	0.361959	0.207783	0.028090
Máximo	1.000000	1.000000	1.000000

A Figura 39 apresenta a similaridade média entre observações de estados estacionário e slots candidatos considerando slots de 5, 10 e 30 minutos. A partir dos resultados obtidos, identificamos que quanto maior o valor de θ e o tamanho do slot, menor o número de slots candidatos selecionados pelo limiar resultante. Por exemplo, no cenário com θ igual a 1.0 e slot com 30 minutos, a similaridade média é de $\approx 55\%$, enquanto no cenário com θ igual a 0.6 e slot com 5 minutos, a similaridade média é de $\approx 96\%$. Observamos também que, nesse cenário com eventos de longa duração (estados estacionários), os limiares resultantes removem mais slots do que no cenário com eventos de curta duração (sociabilidade), pois a similaridade média tem valores menores neste cenário com eventos estacionários em relação ao cenário com eventos de sociabilidade.

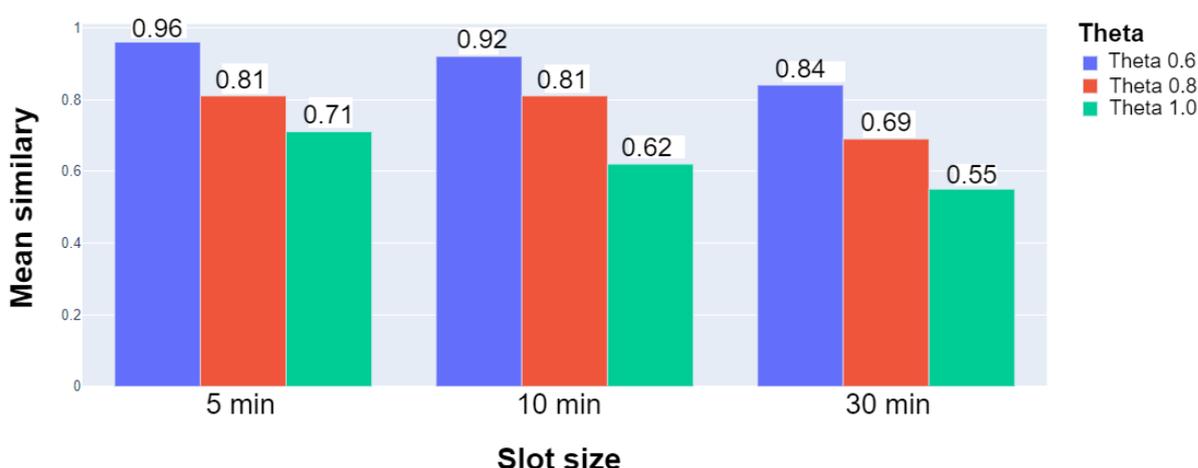


Figura 39 – Similaridade entre observações comportamentais de estado estacionário e slots candidatos usando slots de 5, 10 e 30 minutos.

Em relação ao valor do parâmetro θ , identificamos que quanto maior o seu valor, maior o número de slots removidos, considerando evento de curta duração. Além disso, identificamos que esse resultado é acentuado quando utilizamos tamanhos de slots maiores. Por exemplo, entre os cenários do experimento com eventos de

sociabilidade, a configuração com θ igual a 1.0 e tamanho de slot de 30 minutos resultou em limiares que selecionam um número menor de slots candidatos. Esse resultado também foi evidenciado em experimentos com eventos de longa duração. A partir desses resultados, concluímos que automatizar a atribuição de valores ao parâmetro θ é difícil, pois é uma tarefa subjetiva em que o profissional deve atribuir o valor θ com base nas características do comportamento e do indivíduo monitorado. Assim, fornecemos a seguinte orientação ao usuário da solução proposta: o θ deve ser definido com um valor menor ou igual a 0.8 quando o objetivo é selecionar um número maior de slots candidatos, e o θ deve ser definido com um valor maior que 0.8 quando o objetivo é remover um número maior de slots.

6.4 Análise de Limiar de Intervalo Frequente

Nesta etapa, visamos reconhecer o comportamento da solução ao utilizar diferentes valores para o parâmetro phi (ou seja, φ), responsável por regular o valor do limiar do intervalo frequente. A Figura 40 apresenta a configuração do experimento executado, que consiste em calcular a similaridade entre os slots candidatos e os padrões comportamentais projetados após a aplicação do limiar de intervalo. Especificamente, realizamos os experimentos com valores de φ iguais a 0.05, 0.1 e 0.2. Além disso, também usamos θ igual a 0.6, 0.8 e 1.0. Ao final, calculamos a similaridade média entre slots candidatos e padrões comportamentais para analisar o comportamento da solução proposta nesses cenários.

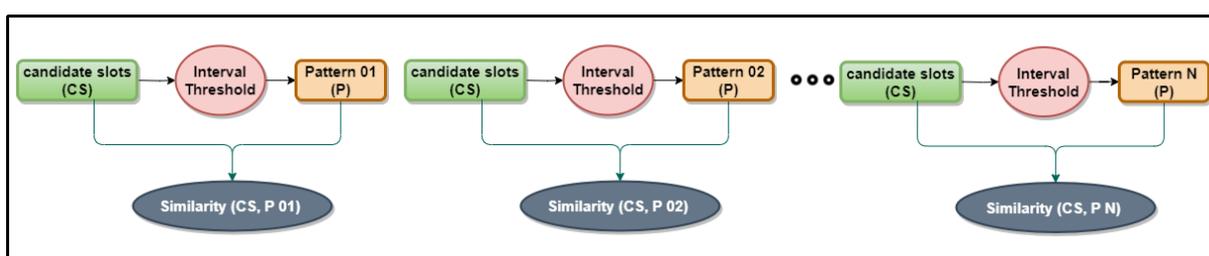


Figura 40 – Configuração do experimento para avaliar o valor do parâmetro phi (φ).

A Figura 41 apresenta a similaridade média entre os slots candidatos e os padrões de sociabilidade. No cenário com φ igual a 0.05, a similaridade permaneceu alta (ou seja, similaridade $\geq 81\%$) considerando todos os valores de θ utilizados. No cenário com φ igual a 0.1, a similaridade média foi de 75%, 72% e 67%, considerando valores de θ iguais a 0.6, 0.8 e 1.0, respectivamente. O cenário com φ igual a 0.2 apresentou baixa média de similaridade (ou seja, similaridades abaixo de 58%) considerando todos os valores θ utilizados. Analisando esses resultados, observamos que quanto maior o valor de φ utilizado, menor o número de intervalos frequentes selecionados.

A Figura 42 apresenta a similaridade média entre slots candidatos e padrões de

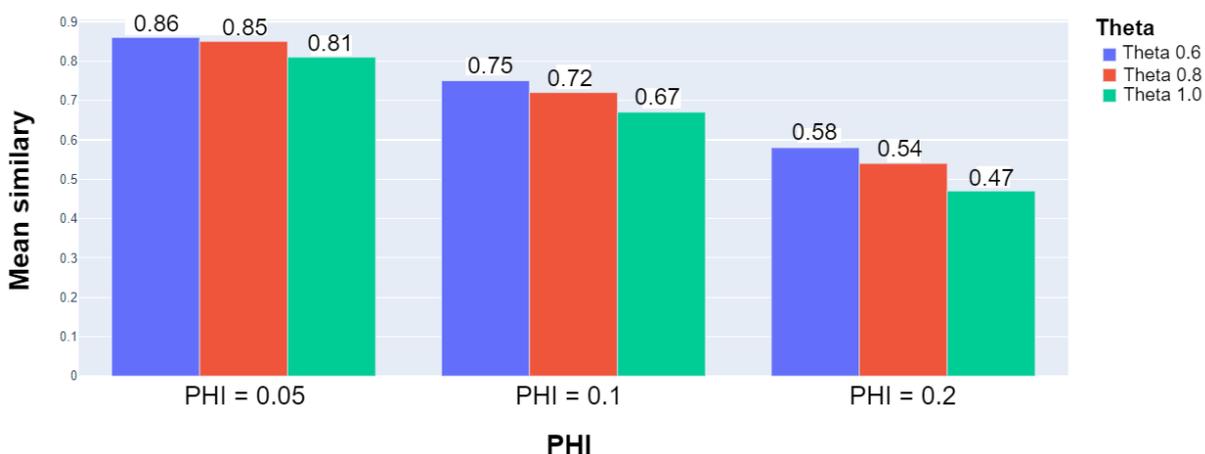


Figura 41 – Similaridade entre slots de candidatos e padrões de sociabilidade

estado estacionário. Com base nesses resultados, observamos uma alta similaridade média (ou seja, $\geq 80\%$) para todos os valores de φ considerando valores de θ iguais a 0,6 e 0,8. No entanto, considerando todos os valores de φ , identificamos uma baixa similaridade média (ou seja, $\leq 77\%$) ao usar θ igual a 1.0. Por exemplo, ao usar φ igual a 0.2 e θ igual a 1.0, a similaridade média resultante foi de 62%, enquanto ao usar φ igual a 0.2 e θ igual a 0.8, a similaridade média foi de 70%.

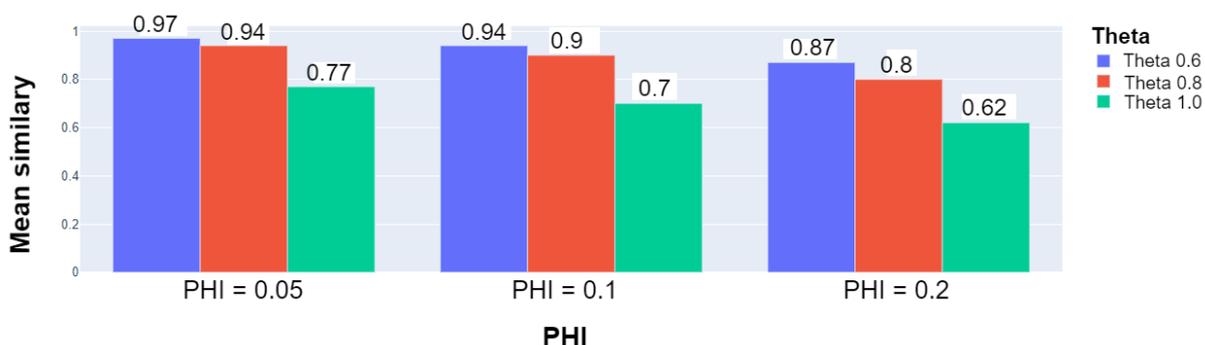


Figura 42 – Similaridade entre slots candidatos e padrões de estado estacionário.

Esse resultado pode ser explicado porque o valor de θ igual a 1.0 resulta em limiares que removem muitos slots, conforme apresentado na Seção 6.3. Conseqüentemente, os slots candidatos neste cenário formam um número maior de intervalos (slots candidatos adjacentes) contendo uma contagem menor de eventos por intervalo do que cenários com valor de θ igual a 0.8. Considerando que θ igual a 1.0 resulta em menos eventos por intervalo, é possível reconhecer que mais intervalos são removidos ao aplicar o limiar de intervalo frequente. Para reforçar essa explicação, aplicamos o coeficiente de correlação de Pearson entre o número de intervalos e a similaridade entre os slots candidatos e os padrões comportamentais. A realização dessa operação resultou em uma correlação de $\approx -0,82$, representando uma forte

associação negativa entre essas variáveis. A partir desse resultado, observamos que, quanto maior a quantidade de intervalos, menor será a similaridade resultante (isto é, maior quantidade de intervalos são removidos).

Em relação ao valor do parâmetro φ , identificamos que quanto maior o seu valor, maior o número de intervalos removidos. Além disso, observamos uma forte relação entre os valores de θ e φ , pois valores maiores de θ resultam em intervalos menores e, conseqüentemente, quando usados em conjunto com valores maiores de φ , tendem a remover muitos intervalos. Nesse cenário, também identificamos que atribuir valores para o parâmetro φ é uma tarefa subjetiva que depende do conhecimento do profissional e da finalidade de monitorar o comportamento. Assim, para auxiliar o usuário da solução proposta, fornecemos as seguintes orientações para a definição do valor do parâmetro φ : (i) deve-se utilizar um valor φ igual ou inferior a 0.05 quando se deseja manter a maioria dos intervalos; (ii) deve-se usar um valor θ maior que 0.05 e menor ou igual a 0.1 quando quiser remover um número moderado de intervalos; (iii) deve-se usar um valor φ maior que 0.1 quando se deseja remover um número considerável de intervalos; e (iv) quando o valor de θ for maior que 0.8, não se deve usar um valor de φ muito alto, mantendo-o abaixo de 0.1.

6.5 Considerações Finais

Primeiramente, realizamos uma análise da influência dos hiperparâmetros na detecção de padrões comportamentais. Especificamente, foram realizadas análises sobre o tamanho do slot de tempo, parametrização do limiar de slot candidato (θ) e parametrização do limiar de intervalo frequente (φ). Os resultados dessa análise resultou em um conjunto de diretrizes que apoiam a definição dos melhores valores para os parâmetros da solução.

A definição de valores para os parâmetros da solução deve ser baseada no tipo de comportamento e no objetivo de monitoramento. Quanto ao tipo de comportamento, deve-se observar a duração média dos eventos gerados. Por exemplo, eventos relacionados ao sono e mobilidade tendem a ter maior duração média do que eventos de sociabilidade. Essa característica tem relação com o tamanho do slot de tempo usado para identificar padrões comportamentais, pois definir tamanho de slot próximo a duração média dos eventos comportamentais processados resulta em observações que melhor se ajustam ao comportamento do indivíduo.

Quanto ao objetivo do monitoramento, deve-se observar a necessidade de modelar rotinas comportamentais mais consistentes ou mais flexíveis. Os parâmetros θ e φ estão relacionados ao objetivo do monitoramento. Quando o objetivo for modelar rotinas flexíveis, deve-se definir um valor para θ menor ou igual a 0.8 e um valor para φ menor ou igual a 0.05. Essa combinação configurará a solução para identificar rotinas

flexíveis (i.e., manterá a maioria dos intervalos analisados). Entretanto, caso o objetivo seja reconhecer rotinas consistentes, deve-se definir um valor para θ maior do que 0.8 e um valor para φ maior do que 0.05. Essa combinação resultará em padrões que modelam rotinas consistentes (i.e., manterá os intervalos com maior repetição comportamental). Além disso, caso seja definido um valor maior do que 0.8 para o parâmetro θ , deve-se manter o valor de φ abaixo de 0.1, pois, conforme os resultados da análise realizada, a combinação de θ maior do que 0.8 e φ maior ou igual a 0.1 tende a remover muitos intervalos. Por exemplo, no Capítulo 5, avaliamos a solução proposta projetando padrões usando θ igual a 0.8 e φ igual a 0.05, resultando em padrões mais flexíveis.

Baseado nos resultados da análise realizada, apresentamos na Tabela 15 as diretrizes projetadas para apoiar os usuários na definição dos valores mais adequados. Portanto, os usuários da solução proposta podem usar essas diretrizes para definir os melhores valores para parametrizar a solução proposta.

Tabela 15 – Diretrizes para definir valores para os parâmetros da solução proposta.

Informação base	Diretriz
Mediana da duração dos eventos comportamentais	configure o tamanho do slot próximo à duração mediana dos eventos.
Rotina comportamental flexível	defina valor θ menor ou igual a 0.8 e valor de φ menor ou igual a 0.05.
Rotina comportamental consistente	defina valor de θ maior do que 0.8 e valor de φ maior do que 0.05.
Orientação Geral	caso defina valor de θ maior do que 0.8, mantenha o valor de φ abaixo de 0.1.

7 Conclusões e Trabalhos Futuros

Este trabalho apresenta uma solução capaz de aprender padrões e mudanças comportamentais, representando uma valiosa ferramenta de apoio para diversas áreas da saúde. Primeiramente, essa solução tem potencial para complementar os métodos tradicionais de avaliação e monitoramento da saúde mental através de informações objetivas. Especificamente, é possível utilizar a solução proposta como uma ferramenta de apoio à tomada de decisões médicas, fornecendo informações sobre mudanças comportamentais que podem ser utilizadas para propor intervenções terapêuticas oportunas, permitindo o rastreamento de respostas para avaliar a eficácia das intervenções. Além disso, a solução pode monitorar situações de interesse relacionadas a doenças crônicas não transmissíveis, como obesidade e doenças cardiovasculares, fornecendo informações como padrões e mudanças de comportamento na atividade física, mobilidade e sono. A solução tem ainda potencial para apoiar o monitoramento comportamental de trabalhadores ao longo do seu período laboral, fornecendo informações valiosas sobre situações comportamentais que possam afetar a saúde do trabalhador, como alterações na mobilidade, sociabilidade e padrões de atividade física.

Esta pesquisa apresenta múltiplas contribuições para a literatura de fenotipagem digital de comportamentos humanos. Primeiramente, conduzimos uma revisão sistemática da literatura para reconhecer, caracterizar e analisar o estado da arte sobre fenotipagem digital da saúde mental usando sensoriamento multimodal de múltiplas situações de interesse para os profissionais. Os resultados desta revisão fornecem as seguintes contribuições: (i) caracterizamos e sintetizamos os estudos relacionados a esta pesquisa; (ii) respondemos a questões fundamentais para a clarificação desta área de investigação, como a análise aprofundada dos processos de caracterização de situações de interesse por meio de dados de contexto e a utilização desta consciência de situação para apoiar os profissionais de saúde mental; e (iii) discutimos tendências e identificamos questões em aberto na literatura.

Outra contribuição relevante desta pesquisa é o modelo formal e o modelo de dados projetado para suportar o desenvolvimento da solução proposta. O modelo formal proposto formaliza os principais conceitos que sistemas de fenotipagem digital geralmente requer, como comportamento humano, evento comportamental, observação comportamental e padrão comportamental enriquecido por contexto. A formalização consiste em um modelo de alto nível que permite a reutilização por outros estudos de fenotipagem digital. O modelo de dados projetado especifica as principais estruturas de eventos comportamentais, observações comportamentais, padrões comportamentais e

mudanças de padrões comportamentais.

Baseado no modelo formal e no modelo de dados, foi desenvolvido uma solução para o monitoramento ubíquo de padrões comportamentais multimodais enriquecidos com contexto. Essa solução realiza o aprendizado incremental de padrões e mudanças comportamentais a partir do processamento de fluxo de eventos comportamentais. Especificamente, a solução é composta por um algoritmo baseado na mineração de padrões frequentes capaz de detectar padrões comportamentais e por um mecanismo que utiliza a lógica fuzzy para detectar comportamentos anormais e mudanças de padrões. Também concebeu-se um *framework* que implementa o algoritmo proposto por meio de funções do spark e integra um módulo fuzzy para implementar o mecanismo de detecção de mudança comportamental.

Este estudo também apresenta experimentos executados para avaliar a solução proposta. O primeiro experimento avaliou a capacidade da solução proposta para reconhecer padrões capazes de modelar a rotina comportamental. A partir dos resultados desta avaliação, identificamos que a solução detecta padrões capazes de modelar adequadamente a rotina dos indivíduos. Essa conclusão é baseada nos níveis de correlação entre a capacidade do padrão de explicar rotinas e a estabilidade da rotina, que encontrou altas associações positivas (*Sociability* > 0.92, *Stationary state* > 0.67, *Moving* > 0.8). Por fim, avaliamos a sensibilidade da detecção de mudança comportamental. Essa avaliação mostrou que nossa solução tem sensibilidade para reconhecer comportamentos anormais e mudanças nas rotinas para comportamentos de sociabilidade, estado estacionário e movimentação.

Por fim, conduzimos uma análise da influência dos hiperparâmetros na aprendizagem de padrões comportamentais realizada pela solução proposta. Os resultados dessa análise viabilizaram a concepção de diretrizes capazes de auxiliar os usuários na parametrização adequada da solução proposta. Portanto, fornecemos orientações para a definição dos melhores valores para os parâmetros da solução com base no tipo de comportamento e nos objetivos do monitoramento realizado.

Em conclusão, abordamos neste trabalho de doutorado as questões de pesquisa apresentadas na Seção 1.1. Abordamos a questão de pesquisa sobre a identificação de padrões comportamentais por meio da investigação de uma nova solução para a aprendizagem de padrões comportamentais baseada na abordagem de mineração de padrões frequentes. Especificamente, projetamos essa solução com base no modelo formal e modelo de dados propostos neste trabalho, resultando na concepção de um algoritmo capaz de reconhecer os intervalos de tempo nos quais um comportamento ocorre rotineiramente. Além disso, abordamos a questão de pesquisa sobre o reconhecimento de mudanças comportamentais investigando uma nova abordagem para identificar mudanças de padrões e comportamentos anormais baseada na lógica

fuzzy. Especificamente, a partir dos conceitos da lógica fuzzy, a abordagem concebida é capaz de reconhecer mudança de rotina e comportamentos anormais considerando valores de pertinência que variam de sem mudança a mudança gradualmente.

7.1 Limitações

A solução depende de parâmetros definidos pelo usuário (ou seja, tamanho do slot, φ e θ) para detectar padrões de comportamento. Especificamente, Esses parâmetros influenciam diretamente nos limiares apresentados no algoritmo projetado para aprender padrões comportamentais. Apesar desta limitação, neste estudo apresentamos diretrizes para apoiar a definição de valores para esses parâmetros conforme o objetivo do monitoramento. Outra limitação é que a solução proposta aprende padrões considerando considerando slots de tempo de cada dia, não modelando padrões que se repetem em outra granularidade de tempo, como dias e semanas.

7.2 Trabalhos Futuros

O presente trabalho apresenta uma janela de oportunidades de pesquisas que podem contribuir com a melhoria e complementação da solução proposta. Os próximos passos desta pesquisa vão em direção aos seguintes objetivos: (i) estender a solução para suportar padrões mais flexíveis, que considerem, além de slots de tempo de cada dia, comportamentos que se repetem em outra granularidade de tempo, como dias, semanas e meses; (ii) desenvolver aplicações móveis para realizar a inferência de eventos comportamentais, uma vez que nossa solução não infere eventos, mas trabalha na identificação de padrões e mudanças comportamentais processando-os; (iii) desenvolver um modelo de segurança e privacidade para a solução proposta; (iv) estender a solução para adicionar a capacidade de identificar se uma mudança de padrão reconhecida reflete o mesmo hábito em horários diferentes (mudança de cronograma); e (v) investigar a aplicabilidade da lógica fuzzy do tipo 2 para a detecção de comportamentos anormais e mudanças de padrões.

7.3 Publicações

Para divulgação dos resultados desta pesquisa, foram escritos alguns artigos, que são apresentadas a seguir. Para cada publicação, informa-se o tipo de publicação, status em que ela se encontra e o Qualis do periódico ou conferência (conforme classificação quadriênio 2017-2020), se disponível.

- Moura, Ivan; Silva, Francisco; Coutinho, Luciano; Marques, Jean; Teles, Ariel.

Detectando Padrões de Sociabilidade de Seres Humanos Através do Processamento de Eventos Complexos. ERCEMAPI, 2020 ([Rodrigues et al. 2020](#)).

Situação: Publicado. **Tipo de publicação:** Capítulo de Livro.

- Moura, Ivan; Silva, Francisco; Coutinho, Luciano; Teles, Ariel. **Mental Health Ubiquitous Monitoring: Detecting Context-Enriched Sociability Patterns Through Complex Event Processing.** IEEE 33rd International Symposium on Computer Based Medical Systems (CBMS), 2020 ([Rodrigues de Moura et al. 2020](#)).

Situação: Publicado. **Tipo de publicação:** Conferência internacional **Qualis:** A3.

- Moura, Ivan; Teles, Ariel; Endler, Markus; Coutinho, Luciano; Silva, Francisco. **Recognizing Context-Aware Human Sociability Patterns Using Pervasive Monitoring for Supporting Mental Health Professionals.** Sensors, 2021 ([Moura et al. 2021](#)).

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A2. **Fator de impacto:** 3.9.

- **Periódico:** Moura, Ivan; Teles, Ariel; Coutinho, Luciano; Silva, Francisco. **Towards Identifying Context-Enriched Multimodal Behavioral Patterns for Digital Phenotyping of Human Behaviors** Future Generation Computer Systems, 2022 ([Moura et al. 2022](#)).

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A1. **Fator de impacto:** 7.5.

- Moura, Ivan; Teles, Ariel; Viana, Davi; Marques, Jean; Coutinho, Luciano; Silva, Francisco. **Digital Phenotyping of Mental Health Using Multimodal Sensing of Multiple Situations of Interest: A Systematic Literature Review** Journal of Biomedical Informatics, 2022.

Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A1. **Fator de impacto:** 4.5.

Referências

ADLER, D. A.; BEN-ZEEV, D.; TSENG, V. W.-S.; KANE, J. M.; BRIAN, R.; CAMPBELL, A. T.; HAUSER, M.; SCHERER, E. A.; CHOUDHURY, T. Predicting early warning signs of psychotic relapse from passive sensing data: An approach using encoder-decoder neural networks. **JMIR Mhealth Uhealth**, v. 8, n. 8, p. e19962, Aug 2020. ISSN 2291-5222. Disponível em: <<https://mhealth.jmir.org/2020/8/e19962>>. Citado 4 vezes nas páginas 36, 42, 43 e 110.

ADLER, D. A.; TSENG, V. W.-S.; QI, G.; SCARPA, J.; SEN, S.; CHOUDHURY, T. Identifying mobile sensing indicators of stress-resilience. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 5, n. 2, jun. 2021. Disponível em: <<https://doi.org/10.1145/3463528>>. Citado 3 vezes nas páginas 43, 45 e 109.

AGGARWAL, C. C.; BHUIYAN, M. A.; HASAN, M. A. Frequent pattern mining algorithms: A survey. In: _____. **Frequent Pattern Mining**. Cham: Springer International Publishing, 2014. p. 19–64. Citado na página 24.

AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **SIGMOD Rec.**, ACM, New York, NY, USA, v. 22, n. 2, p. 207–216, jun. 1993. ISSN 0163-5808. Disponível em: <<http://doi.acm.org/10.1145/170036.170072>>. Citado na página 24.

ALTHUBAITI, A. Information bias in health research: Definition, pitfalls, and adjustment methods. **Journal of Multidisciplinary Healthcare**, v. 9, p. 211, 2016. Citado na página 23.

ANTONELLI, M.; DUCANGE, P.; MARCELLONI, F.; SEGATORI, A. A novel associative classification model based on a fuzzy frequent pattern mining algorithm. **Expert Systems with Applications**, Elsevier, v. 42, n. 4, p. 2086–2097, 2015. Citado na página 25.

BARNETT, I.; TOROUS, J.; STAPLES, P.; SANDOVAL, L.; KESHAVAN, M.; ONNELA, J.-P. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. **Neuropsychopharmacology**, Nature Publishing Group, v. 43, n. 8, p. 1660, 2018. Citado 6 vezes nas páginas 17, 39, 40, 42, 44 e 111.

BENESTY, J.; CHEN, J.; HUANG, Y.; COHEN, I. Pearson correlation coefficient. In: **Noise reduction in speech processing**. [S.l.]: Springer, 2009. p. 1–4. Citado na página 71.

BHANDARI, A.; GUPTA, A.; DAS, D. Improved apriori algorithm using frequent pattern tree for real time applications in data mining. **Procedia Computer Science**, Elsevier, v. 46, p. 644–651, 2015. Citado na página 25.

BHIDAYASIRI, R.; MARI, Z. Digital phenotyping in parkinson's disease: Empowering neurologists for measurement-based care. **Parkinsonism & Related Disorders**, v. 80, p. 35–40, 2020. ISSN 1353-8020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1353802020307124>>. Citado na página 17.

BOUKHECHBA, M.; DAROS, A. R.; FUA, K.; CHOW, P. I.; TEACHMAN, B. A.; BARNES, L. E. Demonic salmon: Monitoring mental health and social interactions of college students using smartphones. **Smart Health**, v. 9-10, p. 192 – 203, 2018. ISSN 2352-6483. CHASE 2018 Special Issue. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2352648318300400>>. Citado 4 vezes nas páginas 30, 36, 45 e 112.

BUDA, T. S.; KHWAJA, M.; MATIC, A. Outliers in smartphone sensor data reveal outliers in daily happiness. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 5, n. 1, mar. 2021. Disponível em: <<https://doi.org/10.1145/3448095>>. Citado 3 vezes nas páginas 40, 41 e 109.

CAI, L.; BOUKHECHBA, M.; GERBER, M. S.; BARNES, L. E.; SHOWALTER, S. L.; COHN, W. F.; CHOW, P. I. An integrated framework for using mobile sensing to understand response to mobile interventions among breast cancer patients. **Smart Health**, v. 15, p. 100086, 2020. ISSN 2352-6483. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2352648319300509>>. Citado 3 vezes nas páginas 44, 45 e 110.

CAI, L.; BOUKHECHBA, M.; WU, C.; CHOW, P. I.; TEACHMAN, B. A.; BARNES, L. E.; GERBER, M. S. State affect recognition using smartphone sensing data. In: **Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies**. New York, NY, USA: Association for Computing Machinery, 2018. (CHASE '18), p. 120–125. ISBN 9781450359580. Disponível em: <<https://doi.org/10.1145/3278576.3284386>>. Citado 5 vezes nas páginas 36, 42, 44, 45 e 112.

CALVO, R. A.; PETERS, D. **Positive Computing: Technology for Well-Being and Human Potential**. [S.l.]: The MIT Press, 2014. Citado 2 vezes nas páginas 23 e 24.

CHAMBERS, B.; ZAHARIA, M. **Spark: The definitive guide: Big data processing made simple**. [S.l.]: "O'Reilly Media, Inc.", 2018. Citado na página 65.

CHIKERSAL, P.; DORYAB, A.; TUMMINIA, M.; VILLALBA, D. K.; DUTCHER, J. M.; LIU, X.; COHEN, S.; CRESWELL, K. G.; MANKOFF, J.; CRESWELL, J. D.; GOEL, M.; DEY, A. K. 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>>. Citado 5 vezes nas páginas 40, 41, 42, 43 e 109.

CHIKERSAL, P.; VENKATESH, S.; MASOWN, K.; WALKER, E.; QURAIISHI, D.; DEY, A.; GOEL, M.; XIA, Z. Predicting multiple sclerosis outcomes during the covid-19 stay-at-home period: Observational study using passively sensed behaviors and digital phenotyping. **JMIR Ment Health**, v. 9, n. 8, p. e38495, Aug 2022. ISSN 2368-7959. Disponível em: <<https://mental.jmir.org/2022/8/e38495>>. Citado 2 vezes nas páginas 41 e 108.

CHO, C.-H.; LEE, T.; KIM, M.-G.; IN, H. P.; KIM, L.; LEE, H.-J. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes

based on the circadian rhythm: Prospective observational cohort study. **J Med Internet Res**, v. 21, n. 4, p. e11029, Apr 2019. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2019/4/e11029/>>. Citado 2 vezes nas páginas 43 e 111.

COHEN, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. **Psychological bulletin**, v. 70, n. 4, 1968. Citado na página 30.

CORNET, V. P.; HOLDEN, R. J. Systematic review of smartphone-based passive sensing for health and wellbeing. **Journal of Biomedical Informatics**, v. 77, p. 120 – 132, 2018. Citado na página 17.

Demasi; Aguilera; Recht. Detecting change in depressive symptoms from daily wellbeing questions, personality, and activity. In: **2016 IEEE Wireless Health (WH)**. [S.l.: s.n.], 2016. p. 1–8. Citado 5 vezes nas páginas 37, 42, 44, 45 e 113.

DOGRUCU, A.; PERUCIC, A.; ISARO, A.; BALL, D.; TOTO, E.; RUNDENSTEINER, E. A.; AGU, E.; DAVIS-MARTIN, R.; BOUDREAUX, E. Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. **Smart Health**, v. 17, p. 100118, 2020. ISSN 2352-6483. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2352648319300273>>. Citado 5 vezes nas páginas 36, 37, 40, 41 e 110.

DORYAB, A.; VILLALBA, D. K.; CHIKERSAL, P.; DUTCHER, J. M.; TUMMINIA, M.; LIU, X.; COHEN, S.; CRESWELL, K.; MANKOFF, J.; CRESWELL, J. D.; DEY, A. K. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: Statistical analysis, data mining and machine learning of smartphone and fitbit data. **JMIR Mhealth Uhealth**, v. 7, n. 7, p. e13209, Jul 2019. Citado 5 vezes nas páginas 36, 40, 41, 45 e 111.

DYBa, T.; DINGSØYR, T. Strength of evidence in systematic reviews in software engineering. In: **Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement**. New York, NY, USA: ACM, 2008. (ESEM '08), p. 178–187. Citado na página 31.

ESKES, P.; SPRUIT, M.; BRINKKEMPER, S.; VORSTMAN, J.; KAS, M. J. The sociability score: App-based social profiling from a healthcare perspective. **Computers in Human Behavior**, Elsevier, v. 59, p. 39–48, 2016. Citado na página 16.

Farhan, A. A.; Lu, J.; Bi, J.; Russell, A.; Wang, B.; Bamis, A. Multi-view bi-clustering to identify smartphone sensing features indicative of depression. In: **2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)**. [S.l.: s.n.], 2016. p. 264–273. Citado 3 vezes nas páginas 36, 42 e 112.

FILHO, J. D. P. R.; TELES, A. S.; SILVA, F. J.; COUTINHO, L. R. Towards clustering human behavioral patterns based on digital phenotyping. In: **2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.: s.n.], 2021. p. 271–276. Citado 2 vezes nas páginas 40 e 109.

FUKAZAWA, Y.; ITO, T.; OKIMURA, T.; YAMASHITA, Y.; MAEDA, T.; OTA, J. Predicting anxiety state using smartphone-based passive sensing. **Journal of Biomedical**

Informatics, v. 93, p. 103151, 2019. ISSN 1532-0464. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1532046419300693>>. Citado 2 vezes nas páginas 42 e 111.

FULFORD, D.; MOTE, J.; GONZALEZ, R.; ABPLANALP, S.; ZHANG, Y.; LUCKENBAUGH, J.; ONNELA, J.-P.; BUSSO, C.; GARD, D. E. Smartphone sensing of social interactions in people with and without schizophrenia. **Journal of Psychiatric Research**, 2020. ISSN 0022-3956. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S002239562031058X>>. Citado 3 vezes nas páginas 44, 45 e 110.

GARCIA-CEJA, E.; RIEGLER, M.; NORDGREEN, T.; JAKOBSEN, P.; OEDEGAARD, K. J.; TØRRESEN, J. Mental health monitoring with multimodal sensing and machine learning: A survey. **Pervasive and Mobile Computing**, v. 51, p. 1 – 26, 2018. Citado na página 24.

GONG, J.; HUANG, Y.; CHOW, P. I.; FUA, K.; GERBER, M. S.; TEACHMAN, B. A.; BARNES, L. E. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. **Information Fusion**, v. 49, p. 57–68, 2019. Citado 4 vezes nas páginas 17, 44, 45 e 111.

GRÑERBL, A.; MUAREMI, A.; OSMANI, V.; BAHLE, G.; ÖHLER, S.; TRÖSTER, G.; MAYORA, O.; HARING, C.; LUKOWICZ, P. Smartphone-based recognition of states and state changes in bipolar disorder patients. **IEEE Journal of Biomedical and Health Informatics**, v. 19, n. 1, p. 140–148, 2015. Citado 2 vezes nas páginas 42 e 113.

Gu, F.; Niu, J.; He, Z.; Jin, X.; Rodrigues, J. J. P. C. Smartbuddy: An integrated mobile sensing and detecting system for family activities. In: **GLOBECOM 2017 - 2017 IEEE Global Communications Conference**. [S.l.: s.n.], 2017. p. 1–7. Citado 3 vezes nas páginas 37, 40 e 112.

HAINES-DELMONT, A.; CHAHAL, G.; BRUEN, A. J.; WALL, A.; KHAN, C. T.; SADASHIV, R.; FEARNLEY, D. Testing suicide risk prediction algorithms using phone measurements with patients in acute mental health settings: Feasibility study. **JMIR Mhealth Uhealth**, v. 8, n. 6, p. e15901, Jun 2020. ISSN 2291-5222. Disponível em: <<https://mhealth.jmir.org/2020/6/e15901>>. Citado 3 vezes nas páginas 40, 41 e 110.

HAN, J.; CHENG, H.; XIN, D.; YAN, X. Frequent pattern mining: current status and future directions. **Data mining and knowledge discovery**, Springer, v. 15, n. 1, p. 55–86, 2007. Citado na página 25.

HARARI, G. M.; GOSLING, S. D.; WANG, R.; CHEN, F.; CHEN, Z.; CAMPBELL, A. T. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. **Computers in Human Behavior**, v. 67, p. 129 – 138, 2017. ISSN 0747-5632. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0747563216307282>>. Citado 3 vezes nas páginas 36, 40 e 112.

HE-YUEYA, J.; BUCK, B.; CAMPBELL, A.; CHOUDHURY, T.; KANE, J. M.; BEN-ZEEV, D.; ALTHOFF, T. Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. **NPJ schizophrenia**, Nature Publishing Group, v. 6, n. 1, p. 1–8, 2020. Citado 2 vezes nas páginas 45 e 109.

HENSON, P.; D'MELLO, R.; VAIDYAM, A.; KESHAVAN, M.; TOROUS, J. Anomaly detection to predict relapse risk in schizophrenia. **Translational psychiatry**, Nature Publishing Group, v. 11, n. 1, p. 1–6, 2021. Citado 2 vezes nas páginas 43 e 108.

HOMAYOUN, S.; DEHGHANTANHA, A.; AHMADZADEH, M.; HASHEMI, S.; KHAYAMI, R. Know abnormal, find evil: frequent pattern mining for ransomware threat hunting and intelligence. **IEEE transactions on emerging topics in computing**, IEEE, 2017. Citado na página 25.

HUCKINS, J. F.; DASILVA, A. W.; WANG, W.; HEDLUND, E.; ROGERS, C.; NEPAL, S. K.; WU, J.; OBUCHI, M.; MURPHY, E. I.; MEYER, M. L.; WAGNER, D. D.; HOLTZHEIMER, P. E.; CAMPBELL, A. T. Mental health and behavior of college students during the early phases of the covid-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. **J Med Internet Res**, v. 22, n. 6, p. e20185, Jun 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/6/e20185/>>. Citado 3 vezes nas páginas 39, 40 e 110.

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**, v. 20, n. 12, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/12/3572>>. Citado 2 vezes nas páginas 43 e 110.

JACOBSON, N. C.; SUMMERS, B.; WILHELM, S. Digital biomarkers of social anxiety severity: Digital phenotyping using passive smartphone sensors. **J Med Internet Res**, v. 22, n. 5, p. e16875, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e16875/>>. Citado 2 vezes nas páginas 43 e 109.

JAGESAR, R. R.; ROOZEN, M. C.; van der Heijden, I.; IKANI, N.; TYBOROWSKA, A.; PENNINX, B. W.; RUHE, H. G.; SOMMER, I. E.; KAS, M. J.; VORSTMAN, J. A. Digital phenotyping and the covid-19 pandemic: Capturing behavioral change in patients with psychiatric disorders. **European Neuropsychopharmacology**, v. 42, p. 115 – 120, 2021. ISSN 0924-977X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0924977X20309457>>. Citado 3 vezes nas páginas 39, 40 e 109.

KIM, M.; YANG, J.; AHN, W.-Y.; CHOI, H. J. Machine learning analysis to identify digital behavioral phenotypes for engagement and health outcome efficacy of an mhealth intervention for obesity: Randomized controlled trial. **J Med Internet Res**, v. 23, n. 6, p. e27218, Jun 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/6/e27218/>>. Citado na página 17.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Department of Computer Science, Keele University, UK, 2004. Citado 2 vezes nas páginas 28 e 31.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for Performing Systematic Literature Reviews in Software Engineering**. [S.I.], 2007. Citado na página 28.

KOURTIS, L. C.; REGELE, O. B.; WRIGHT, J. M.; JONES, G. B. Digital biomarkers for alzheimer's disease: the mobile/wearable devices opportunity. **NPJ digital medicine**, Nature Publishing Group, v. 2, n. 1, p. 1–9, 2019. Citado na página 17.

KRUMM, J. **Ubiquitous computing fundamentals**. [S.I.]: CRC Press, 2018. Citado na página 16.

- LAGO, P.; RONCANCIO, C.; JIMÉNEZ-GUARÍN, C. Learning and managing context enriched behavior patterns in smart homes. **Future Generation Computer Systems**, v. 91, p. 191 – 205, 2019. ISSN 0167-739X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167739X18307180>>. Citado na página 25.
- LAI, J.; RAHMANI, A.; YUNUSOVA, A.; RIVERA, A. P.; LABBAF, S.; HU, S.; DUTT, N.; JAIN, R.; BORELLI, J. L. Using multimodal assessments to capture personalized contexts of college student well-being in 2020: Case study. **JMIR Form Res**, v. 5, n. 5, p. e26186, May 2021. ISSN 2561-326X. Disponível em: <<https://formative.jmir.org/2021/5/e26186>>. Citado 2 vezes nas páginas 40 e 109.
- LANE, N. D.; LIN, M.; MOHAMMOD, M.; YANG, X.; LU, H.; CARDONE, G.; ALI, S.; DORYAB, A.; BERKE, E.; CAMPBELL, A. T.; CHOUDHURY, T. Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing. **Mob. Netw. Appl.**, Springer-Verlag, Berlin, Heidelberg, v. 19, n. 3, p. 345–359, jun. 2014. ISSN 1383-469X. Disponível em: <<https://doi.org/10.1007/s11036-013-0484-5>>. Citado 4 vezes nas páginas 37, 39, 40 e 113.
- LEE, U.; HAN, K.; CHO, H.; CHUNG, K.-M.; HONG, H.; LEE, S.-J.; NOH, Y.; PARK, S.; CARROLL, J. M. Intelligent positive computing with mobile, wearable, and iot devices: Literature review and research directions. **Ad Hoc Networks**, Elsevier, v. 83, p. 8–24, 2019. Citado na página 24.
- LEVINE, L. M.; GWAK, M.; KÄRKKÄINEN, K.; FAZELI, S.; ZADEH, B.; PERIS, T.; YOUNG, A. S.; SARRAFZADEH, M. Anxiety detection leveraging mobile passive sensing. In: ALAM, M. M.; HÄMÄLÄINEN, M.; MUCCHI, L.; NIAZI, I. K.; MOULLEC, Y. L. (Ed.). **Body Area Networks. Smart IoT and Big Data for Intelligent Health**. Cham: Springer International Publishing, 2020. p. 212–225. ISBN 978-3-030-64991-3. Citado 5 vezes nas páginas 36, 37, 40, 41 e 110.
- LIANG, Y.; ZHENG, X.; ZENG, D. D. A survey on big data-driven digital phenotyping of mental health. **Information Fusion**, Elsevier, v. 52, p. 290–307, 2019. Citado 5 vezes nas páginas 16, 17, 18, 23 e 28.
- MASUD, M. T.; MAMUN, M. A.; THAPA, K.; LEE, D.; GRIFFITHS, M. D.; YANG, S.-H. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. **Journal of Biomedical Informatics**, v. 103, p. 103371, 2020. ISSN 1532-0464. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046419302916>>. Citado 5 vezes nas páginas 40, 41, 44, 45 e 110.
- MATTEO, D. D.; FOTINOS, K.; LOKUGE, S.; MASON, G.; STERNAT, T.; KATZMAN, M. A.; ROSE, J. Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: Cross-sectional study. **J Med Internet Res**, v. 23, n. 8, p. e28918, Aug 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/8/e28918>>. Citado 2 vezes nas páginas 41 e 108.
- MCNEILL, F. M.; THRO, E. **Fuzzy logic: a practical approach**. [S.l.]: Academic Press, 2014. Citado 3 vezes nas páginas 19, 26 e 27.
- MEHTA, Y.; FATEHI, S.; KAZAMEINI, A.; STACHL, C.; CAMBRIA, E.; EETEMADI, S. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In: **2020 IEEE International Conference on Data Mining (ICDM)**. [S.l.: s.n.], 2020. p. 1184–1189. Citado na página 16.

- MOHR, D. C.; ZHANG, M.; SCHUELLER, S. M. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. **Annual review of clinical psychology**, Annual Reviews, v. 13, p. 23–47, 2017. Citado na página 16.
- MORSHED, M. B.; SAHA, K.; LI, R.; D’MELLO, S. K.; CHOUDHURY, M. D.; ABOWD, G. D.; PLÖTZ, T. Prediction of mood instability with passive sensing. Association for Computing Machinery, New York, NY, USA, v. 3, n. 3, set. 2019. Disponível em: <<https://doi.org/10.1145/3351233>>. Citado 3 vezes nas páginas 42, 43 e 111.
- MORTEL, T. F. Van de et al. Faking it: social desirability response bias in self-report research. **Australian Journal of Advanced Nursing, The**, Australian Nursing Federation, v. 25, n. 4, p. 40, 2008. Citado 2 vezes nas páginas 17 e 23.
- MOSHE, I.; TERHORST, Y.; ASARE, K. O.; SANDER, L. B.; FERREIRA, D.; BAUMEISTER, H.; MOHR, D. C.; PULKKI-RÅBACK, L. Predicting symptoms of depression and anxiety using smartphone and wearable data. **Frontiers in Psychiatry**, v. 12, p. 43, 2021. ISSN 1664-0640. Disponível em: <<https://www.frontiersin.org/article/10.3389/fpsy.2021.625247>>. Citado 5 vezes nas páginas 42, 43, 44, 45 e 109.
- MOURA, I.; TELES, A.; COUTINHO, L.; SILVA, F. Towards identifying context-enriched multimodal behavioral patterns for digital phenotyping of human behaviors. **Future Generation Computer Systems**, v. 131, p. 227–239, 2022. ISSN 0167-739X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167739X22000310>>. Citado 7 vezes nas páginas 39, 50, 68, 69, 70, 95 e 108.
- MOURA, I.; TELES, A.; SILVA, F.; VIANA, D.; COUTINHO, L.; BARROS, F.; ENDLER, M. Mental health ubiquitous monitoring supported by social situation awareness: A systematic review. **Journal of Biomedical Informatics**, v. 107, p. 103454, 2020. ISSN 1532-0464. Citado na página 17.
- MOURA, I.; TELES, A.; VIANA, D.; MARQUES, J.; COUTINHO, L.; SILVA, F. Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review. **Journal of Biomedical Informatics**, p. 104278, 2022. ISSN 1532-0464. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1532046422002830>>. Citado 3 vezes nas páginas 28, 48 e 49.
- MOURA, I. R. de; TELES, A. S.; ENDLER, M.; COUTINHO, L. R.; SILVA, F. J. da Silva e. Recognizing context-aware human sociability patterns using pervasive monitoring for supporting mental health professionals. **Sensors**, v. 21, n. 1, 2021. ISSN 1424-8220. Citado 5 vezes nas páginas 50, 68, 69, 70 e 95.
- NARZIEV, N.; GOH, H.; TOSHNAZAROV, K.; LEE, S. A.; CHUNG, K.-M.; NOH, Y. Std: Short-term depression detection with passive sensing. **Sensors**, v. 20, n. 5, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/5/1396>>. Citado 4 vezes nas páginas 37, 40, 41 e 110.
- NASREEN, S.; AZAM, M. A.; SHEHZAD, K.; NAEEM, U.; GHAZANFAR, M. A. Frequent pattern mining algorithms for finding associated frequent patterns for data streams: a survey. **Procedia Computer Science**, Elsevier, v. 37, p. 109–116, 2014. Citado na página 25.

NEPAL, S.; WANG, W.; VOJDANOVSKI, V.; HUCKINS, J. F.; DASILVA, A.; MEYER, M.; CAMPBELL, A. Covid student study: A year in the life of college students during the covid-19 pandemic through the lens of mobile phone sensing. In: **Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2022. (CHI '22). ISBN 9781450391573. Disponível em: <<https://doi.org/10.1145/3491102.3502043>>. Citado 2 vezes nas páginas 40 e 108.

PHILIP, N. Y.; RODRIGUES, J. J. P. C.; WANG, H.; FONG, S. J.; CHEN, J. Internet of things for in-home health monitoring systems: Current advances, challenges and future directions. **IEEE Journal on Selected Areas in Communications**, v. 39, n. 2, p. 300–310, 2021. Citado na página 16.

PLACE, S.; BLANCH-HARTIGAN, D.; RUBIN, C.; GORROSTIETA, C.; MEAD, C.; KANE, J.; MARX, B. P.; FEAST, J.; DECKERSBACH, T.; NIERENBERG, A. et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. **Journal of medical Internet research**, JMIR Publications Inc., Toronto, Canada, v. 19, n. 3, p. e75, 2017. Citado 4 vezes nas páginas 37, 42, 44 e 112.

RASHID, H.; MENDU, S.; DANIEL, K. E.; BELTZER, M. L.; TEACHMAN, B. A.; BOUKHECHBA, M.; BARNES, L. E. Predicting subjective measures of social anxiety from sparsely collected mobile sensor data. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 4, n. 3, sep 2020. Disponível em: <<https://doi.org/10.1145/3411823>>. Citado 2 vezes nas páginas 43 e 110.

REINERTSEN, E.; CLIFFORD, G. D. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. **Physiological Measurement**, IOP Publishing, v. 39, n. 5, p. 05TR01, may 2018. Citado na página 16.

Rodrigues de Moura, I.; José da Silva e Silva, F.; Reis Coutinho, L.; Soares Teles, A. Mental health ubiquitous monitoring: Detecting context-enriched sociability patterns through complex event processing. In: **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.: s.n.], 2020. p. 239–244. Citado 5 vezes nas páginas 50, 68, 69, 70 e 95.

RODRIGUES, I.; SILVA, F.; COUTINHO, L.; MARQUES, J.; TELES, A. S. Detectando padrões de sociabilidade de seres humanos através do processamento de eventos complexos. **Sociedade Brasileira de Computação**, 2020. Citado na página 95.

ROSS, T. J. **Fuzzy logic with engineering applications**. [S.l.]: John Wiley & Sons, 2005. Citado na página 19.

SANO, A.; TAYLOR, S.; MCHILL, A. W.; PHILLIPS, A. J.; BARGER, L. K.; KLERMAN, E.; PICARD, R. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. **J Med Internet Res**, v. 20, n. 6, p. e210, Jun 2018. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2018/6/e210/>>. Citado 5 vezes nas páginas 36, 37, 40, 42 e 111.

SARDA, A.; MUNUSWAMY, S.; SARDA, S.; SUBRAMANIAN, V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: Cross-sectional observational study. **mHealth and uHealth**, JMIR, v. 7, n. 1, p. e11041, 2019. Citado 3 vezes nas páginas 37, 42 e 111.

SEFIDGAR, Y. S.; SEO, W.; KUEHN, K. S.; ALTHOFF, T.; BROWNING, A.; RISKIN, E.; NURIUS, P. S.; DEY, A. K.; MANKOFF, J. Passively-sensed behavioral correlates of discrimination events in college students. **Proc. ACM Hum.-Comput. Interact.**, Association for Computing Machinery, New York, NY, USA, v. 3, n. CSCW, nov. 2019. Disponível em: <<https://doi.org/10.1145/3359216>>. Citado 2 vezes nas páginas 45 e 111.

SERVIA-RODRÍGUEZ, S.; RACHURI, K. K.; MASCOLO, C.; RENTFROW, P. J.; LATHIA, N.; SANDSTROM, G. M. Mobile sensing at the service of mental well-being: A large-scale longitudinal study. In: **Proceedings of the 26th International Conference on World Wide Web**. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017. (WWW '17), p. 103–112. ISBN 9781450349130. Disponível em: <<https://doi.org/10.1145/3038912.3052618>>. Citado 5 vezes nas páginas 37, 42, 44, 45 e 112.

SILVA, C. W. D. **Intelligent control: fuzzy logic applications**. [S.l.]: CRC press, 2018. Citado na página 25.

SIMSION, G.; WITT, G. **Data modeling essentials**. [S.l.]: Elsevier, 2004. Citado na página 55.

SPATHIS, D.; SERVIA-RODRIGUEZ, S.; FARRAHI, K.; MASCOLO, C.; RENTFROW, J. Passive mobile sensing and psychological traits for large scale mood prediction. In: **Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare**. New York, NY, USA: Association for Computing Machinery, 2019. (PervasiveHealth'19), p. 272–281. ISBN 9781450361262. Disponível em: <<https://doi.org/10.1145/3329189.3329213>>. Citado 4 vezes nas páginas 37, 40, 42 e 111.

SÜKEI, E.; NORBURY, A.; PEREZ-RODRIGUEZ, M. M.; OLMOS, P. M.; ARTÉS, A. Predicting emotional states using behavioral markers derived from passively sensed data: Data-driven machine learning approach. **JMIR Mhealth Uhealth**, v. 9, n. 3, p. e24465, Mar 2021. ISSN 2291-5222. Disponível em: <<https://mhealth.jmir.org/2021/3/e24465>>. Citado 2 vezes nas páginas 43 e 109.

SULTANA, M.; AL-JEFRI, M.; LEE, J. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: Exploratory study. **JMIR Mhealth Uhealth**, v. 8, n. 9, p. e17818, Sep 2020. ISSN 2291-5222. Disponível em: <<http://mhealth.jmir.org/2020/9/e17818/>>. Citado 3 vezes nas páginas 40, 41 e 110.

TELES, A.; BARROS, F.; RODRIGUES, I.; BARBOSA, A.; SILVA, F.; COUTINHO, L.; TEIXEIRA, S. Internet of things applied to mental health: Concepts, applications, and perspectives. In: _____. **IoT and ICT for Healthcare Applications**. [S.l.]: Springer International Publishing, 2020. cap. 4, p. 33–58. Citado na página 28.

TEO, J. X.; DAVILA, S.; YANG, C.; HIL, A. A.; PUA, C. J.; YAP, J.; TAN, S. Y.; SAHLÉN, A.; CHIN, C. W.-L.; TEH, B. T. et al. Digital phenotyping by consumer wearables

identifies sleep-associated markers of cardiovascular disease risk and biological aging. **Communications biology**, Nature Publishing Group, v. 2, n. 1, p. 1–10, 2019. Citado na página 17.

TOROUS, J.; KIANG, M. V.; LORME, J.; ONNELA, J.-P. New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. **JMIR Mental Health**, JMIR Publications, v. 3, n. 2, 2016. Citado 2 vezes nas páginas 16 e 23.

TRULL, T. J.; EBNER-PRIEMER, U. Ambulatory assessment. **Annual Review of Clinical Psychology**, v. 9, n. 1, p. 151–176, 2013. Citado na página 23.

VIERA, A. J.; GARRETT, J. M. Understanding interobserver agreement: the kappa statistic. **Family Medicine**, v. 37 5, p. 360–3, 2005. Citado na página 32.

WAHLE, F.; KOWATSCH, T.; FLEISCH, E.; RUFER, M.; WEIDT, S. Mobile sensing and support for people with depression: a pilot trial in the wild. **JMIR mHealth and uHealth**, JMIR Publications Inc., Toronto, Canada, v. 4, n. 3, p. e111, 2016. Citado 5 vezes nas páginas 16, 36, 40, 42 e 112.

Wang, F.; Wang, Y.; Wang, J.; Xiong, H.; Zhao, J.; Zhang, D. Assessing mental stress based on smartphone sensing data: An empirical study. In: **2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)**. [S.l.: s.n.], 2019. p. 1031–1038. Citado 3 vezes nas páginas 40, 42 e 111.

WANG, R.; AUNG, M. S. H.; ABDULLAH, S.; BRIAN, R.; CAMPBELL, A. T.; CHOUDHURY, T.; HAUSER, M.; KANE, J.; MERRILL, M.; SCHERER, E. A.; TSENG, V. W. S.; BEN-ZEEV, D. Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. In: **Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: Association for Computing Machinery, 2016. (UbiComp '16), p. 886–897. ISBN 9781450344616. Disponível em: <<https://doi.org/10.1145/2971648.2971740>>. Citado 3 vezes nas páginas 44, 45 e 112.

WANG, R.; CHEN, F.; CHEN, Z.; LI, T.; HARARI, G.; TIGNOR, S.; ZHOU, X.; BEN-ZEEV, D.; CAMPBELL, A. T. Studentlife: Using smartphones to assess mental health and academic performance of college students. In: _____. **Mobile Health: Sensors, Analytic Methods, and Applications**. Cham: Springer International Publishing, 2017. cap. 2, p. 7–33. Citado 5 vezes nas páginas 30, 36, 45, 70 e 112.

WANG, R.; WANG, W.; AUNG, M. S. H.; BEN-ZEEV, D.; BRIAN, R.; CAMPBELL, A. T.; CHOUDHURY, T.; HAUSER, M.; KANE, J.; SCHERER, E. A.; WALSH, M. Predicting symptom trajectories of schizophrenia using mobile sensing. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 1, n. 3, set. 2017. Disponível em: <<https://doi.org/10.1145/3130976>>. Citado 5 vezes nas páginas 36, 37, 42, 44 e 112.

WANG, R.; WANG, W.; DASILVA, A.; HUCKINS, J. F.; KELLEY, W. M.; HEATHERTON, T. F.; CAMPBELL, A. T. 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>>. Citado 8 vezes nas páginas 17, 36, 37, 42, 43, 44, 45 e 111.

Wang, R.; Wang, W.; Obuchi, M.; Scherer, E.; Brian, R.; Ben-Zeev, D.; Choudhury, T.; Kane, J.; Hauser, M.; Walsh, M.; Campbell, A. On predicting relapse in schizophrenia using mobile sensing in a randomized control trial. In: **2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)**. [S.l.: s.n.], 2020. p. 1–8. Citado 3 vezes nas páginas 40, 41 e 110.

WANG, W.; NEPAL, S.; HUCKINS, J. F.; HERNANDEZ, L.; VOJDANOVSKI, V.; MACK, D.; PLOMP, J.; PILLAI, A.; OBUCHI, M.; DASILVA, A.; MURPHY, E.; HEDLUND, E.; ROGERS, C.; MEYER, M.; CAMPBELL, A. First-gen lens: Assessing mental health of first-generation students across their first year at college using mobile sensing. **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/10.1145/3543194>>. Citado 3 vezes nas páginas 42, 44 e 108.

WEISER, M. Ubiquitous computing. **Computer**, IEEE, n. 10, p. 71–72, 1993. Citado na página 16.

WEN, H.; SOBOLEV, M.; VITALE, R.; KIZER, J.; POLLAK, J. P.; MUENCH, F.; ESTRIN, D. mpulse mobile sensing model for passive detection of impulsive behavior: Exploratory prediction study. **JMIR Ment Health**, v. 8, n. 1, p. e25019, Jan 2021. ISSN 2368-7959. Disponível em: <<http://mental.jmir.org/2021/1/e25019/>>. Citado 4 vezes nas páginas 42, 43, 45 e 109.

XU, X.; CHIKERSAL, P.; DUTCHER, J. M.; SEFIDGAR, Y. S.; SEO, W.; TUMMINIA, M. J.; VILLALBA, D. K.; COHEN, S.; CRESWELL, K. G.; CRESWELL, J. D.; DORYAB, A.; NURIUS, P. S.; RISKIN, E.; DEY, A. K.; MANKOFF, J. Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students. **Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.**, Association for Computing Machinery, New York, NY, USA, v. 5, n. 1, mar. 2021. Disponível em: <<https://doi.org/10.1145/3448107>>. Citado 3 vezes nas páginas 40, 41 e 109.

YAGER, R. R.; ZADEH, L. A. **An introduction to fuzzy logic applications in intelligent systems**. [S.l.]: Springer Science & Business Media, 2012. v. 165. Citado na página 64.

YANG, Q.; HATCH, D.; CROWLEY, M. J.; LEWINSKI, A. A.; VAUGHN, J.; STEINBERG, D.; VORDERSTRASSE, A.; JIANG, M.; SHAW, R. J. Digital phenotyping self-monitoring behaviors for individuals with type 2 diabetes mellitus: Observational study using latent class growth analysis. **JMIR Mhealth Uhealth**, v. 8, n. 6, p. e17730, Jun 2020. ISSN 2291-5222. Disponível em: <<https://mhealth.jmir.org/2020/6/e17730>>. Citado na página 17.

Zadeh, L. A. Fuzzy logic. **Computer**, v. 21, n. 4, p. 83–93, April 1988. ISSN 1558-0814. Citado na página 25.

ZADEH, L. A. Fuzzy sets. In: **Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh**. [S.l.]: World Scientific, 1996. p. 394–432. Citado na página 64.

ZHANG, X.; ZHUANG, F.; LI, W.; YING, H.; XIONG, H.; LU, S. Inferring mood instability via smartphone sensing: A multi-view learning approach. In: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 1401–1409. ISBN 9781450368896. Disponível em: <<https://doi.org/10.1145/3343031.3350957>>. Citado 2 vezes nas páginas 41 e 111.

ZHOU, J.; LAMICHHANE, B.; BEN-ZEEV, D.; CAMPBELL, A.; SANO, A. Predicting psychotic relapse in schizophrenia with mobile sensor data: Routine cluster analysis. **JMIR Mhealth Uhealth**, v. 10, n. 4, p. e31006, Apr 2022. ISSN 2291-5222. Disponível em: <<https://mhealth.jmir.org/2022/4/e31006>>. Citado 3 vezes nas páginas 39, 43 e 108.

A Apêndice A - Caracterização dos Estudos Seleccionados na RSL

Resumimos as principais informações dos estudos revisados na Tabela 16, que apresenta dados sobre o tópico de saúde mental visado, tipo de dispositivo usado, situações identificadas por meio de detecção passiva, objetivo do estudo e pontuação de qualidade. Medimos a pontuação de qualidade com base no procedimento de avaliação conforme descrito na Seção 3.1.4.

Tabela 16 – Resumo dos estudos revisados.

Referência	Tópico de saúde mental	Dispositivo	Situações	Objetivo	Pontuação
(Zhou et al. 2022)	Esquizofrenia	Smartphone	Mobilidade, atividade física, sociabilidade e contexto ambiental	Identificação de informações de alto nível e predição do estado mental	9.5
(Chikersal et al. 2022)	Depressão	Smartphone e dispositivo vestível	Mobilidade, atividade física, sociabilidade, padrão de uso de dispositivos e sono	Classificação do estado mental	10
(Moura et al. 2022)	Bem-estar	Smartphone	Sociabilidade, atividade física e mobilidade	Identificação de informações de alto nível	9.5
(Nepal et al. 2022)	Estresse	Smartphone	Sono, mobilidade, padrões de uso de dispositivos e atividade física	Identificação de informações de alto nível	9.5
(Wang et al. 2022)	Depressão e ansiedade	Smartphone	Atividade física, padrões de uso de dispositivos, mobilidade e sono	predição do estado mental e análise de associação	9.5
(Matteo et al. 2021)	Ansiedade, ansiedade social e depressão	Smartphone	Sociabilidade, sono, mobilidade, contexto ambiental e padrão de uso do dispositivo	Classificação do estado mental	8
(Henson et al. 2021)	Esquizofrenia	Smartphone	Mobilidade, sociabilidade, padrão de uso de dispositivos e sono	Predição do estado mental	9

(Filho et al. 2021)	bem-estar	Smartphone	Sociabilidade, atividade física, contexto ambiental e padrão de uso do dispositivo	Identificação de informações de alto nível	9
(Xu et al. 2021)	Depressão	Smartphone e dispositivo vestível	Sociabilidade, atividade física, mobilidade, sono e padrões de uso de dispositivos	Classificação do estado mental	9.5
(Lai et al. 2021)	Depressão, ansiedade e solidão	Smartphone e dispositivo vestível	Atividade física, sono e humor	Identificação de informações de alto nível	8
(Buda, Khwaja e Matic 2021)	Estados emocionais	Smartphone	Atividade física, mobilidade, contexto ambiental e padrões de uso de dispositivos	Classificação do estado mental	8.5
(Sükei et al. 2021)	Estados emocionais	Smartphone	Atividade física, mobilidade, sono e padrões de uso de dispositivos	Predição do estado mental	9
(Moshe et al. 2021)	Depressão, ansiedade e estresse	Smartphone e dispositivo vestível	Atividade física, mobilidade, sono e padrões de uso de dispositivos	predição do estado mental e análise de associação	8
(Chikersal et al. 2021)	Depressão	Smartphone	Atividade física, sociabilidade, sono e padrões de uso de dispositivos	Classificação do estado mental e análise de associação	9.5
(Wen et al. 2021)	Comportamento impulsivo	Smartphone	Sociabilidade e padrões de uso de dispositivos	predição do estado mental e análise de associação	8.5
(Adler et al. 2021)	Estresse	Smartphone e dispositivo vestível	Atividade física e sono	predição do estado mental e análise de associação	9
(Jagesar et al. 2021)	Depressão, esquizofrenia, transtorno bipolar	Smartphone	Mobilidade e sociabilidade	Identificação de informações de alto nível	9
(He-Yueya et al. 2020)	Esquizofrenia	Smartphone	Sociabilidade, atividade física, contexto ambiental e padrão de uso de dispositivos	Análise de associação	9.5
(Jacobson, Summers e Wilhelm 2020)	Ansiedade social	Smartphone	Atividade física e sociabilidade	Predição do estado mental	8

(Rashid et al. 2020)	Ansiedade social	Smartphone	Mobilidade, Atividade Física e Sociabilidade	Predição do estado mental	9.5
(Haines-Delmont et al. 2020)	Ideação Suicida	Smartphone e dispositivo vestível	Sociabilidade, atividade física e sono	Classificação do estado mental	8.5
(Jacobson e Chung 2020)	Depressão	Smartphone	Mobilidade, contexto ambiental e sociabilidade	Predição do estado mental	8.5
(Narziev et al. 2020)	Depressão	Smartphone e dispositivo vestível	Atividade física, sono, sociabilidade e humor	Classificação do estado mental	8.5
(Wang et al. 2020)	Esquizofrenia	Smartphone	Atividade física, contexto ambiental, mobilidade, sociabilidade	Classificação do estado mental	8.5
(Huckins et al. 2020)	Depressão e ansiedade	Smartphone	Padrões de sono, mobilidade e uso de dispositivos	Identificação de informações de alto nível	9.5
(Sultana, Al-Jefri e Lee 2020)	estados emocionais	Smartphone e dispositivo vestível	Atividade física, contexto ambiental, mobilidade e padrões de uso de dispositivos	Classificação do estado mental	8.5
(Adler et al. 2020)	Esquizofrenia	Smartphone	Atividade física, mobilidade, sociabilidade, padrões de uso de dispositivos e sono	Predição do estado mental	9.5
(Masud et al. 2020)	Depressão	Smartphone	Atividade física e mobilidade	Classificação do estado mental e análise de associação	9
(Fulford et al. 2020)	Esquizofrenia	Smartphone	Mobilidade e sociabilidade	Análise de associação	10
(Levine et al. 2020)	Depressão e ansiedade	Smartphone	Atividade física, contexto ambiental, mobilidade, sociabilidade e padrões de uso de dispositivos	Classificação do estado mental	8
(Cai et al. 2020)	Humor, depressão e ansiedade	Smartphone	Atividade física e mobilidade	Análise de associação	9
(Dogrucu et al. 2020)	Depressão e ideação suicida	Smartphone	Sociabilidade, mobilidade e pistas vocais	Classificação do estado mental	9

(Fukazawa et al. 2019)	Ansiedade	Smartphone	Mobilidade, contexto ambiental, sociabilidade, atividade física	Classificação do estado mental	9
(Cho et al. 2019)	Transtorno bipolar e depressão	Smartphone e dispositivo vestível	Contexto ambiental, atividade física e sono	Predição do estado mental	10
(Zhang et al. 2019)	Humor	Smartphone	Sociabilidade, atividade física, mobilidade, contexto ambiental e padrões de uso de dispositivos	Classificação do estado mental	9
(Doryab et al. 2019)	Solidão	Smartphone e dispositivo vestível	Atividade física, mobilidade, sociabilidade e sono	Classificação do estado mental e análise de associação	9.5
(Morshed et al. 2019)	Humor	Smartphone e dispositivo vestível	Atividade física, sociabilidade, mobilidade e sono	Predição do estado mental	9.5
(Gong et al. 2019)	Ansiedade social	Smartphone	Movimento corporal e sociabilidade	Análise de associação	9.5
(Sarda et al. 2019)	Depressão	Smartphone	Atividade física, mobilidade, sociabilidade e sono	Classificação do estado mental	9
(Sefidgar et al. 2019)	Bem-estar	Smartphone e dispositivo vestível	Atividade física, sono, sociabilidade, padrões de uso de dispositivos e mobilidade	Análise de associação	9.5
(Wang et al. 2019)	Estresse	Smartphone	Atividade física, mobilidade e sociabilidade	Classificação do estado mental	8
(Spathis et al. 2019)	Humor	Smartphone	Atividade física e contexto ambiental	Classificação do estado mental	9
(Wang et al. 2018)	Depressão	Smartphone e dispositivo vestível	Sono, sociabilidade, padrões de uso de dispositivos, mobilidade, atividade física e humor	Previsão do estado mental e análise de associação	9.5
(Sano et al. 2018)	Estresse e bem-estar	Smartphone e dispositivo vestível	Sono, sociabilidade, padrões de uso de dispositivos e mobilidade	Classificação do estado mental	9.5
(Barnett et al. 2018)	Esquizofrenia	Smartphone	Mobilidade e sociabilidade	Predição do estado mental e identificação de informações de alto nível	9.5

(Boukhechba et al. 2018)	Depressão e ansiedade social	Smartphone	Mobilidade, atividade física e sociabilidade	Análise de associação	9.5
(Cai et al. 2018)	Estado afetivo	Smartphone	Mobilidade e atividade física	Predição do estado mental e análise de associação	9
(Wang et al. 2017)	Depressão, estresse e solidão	Smartphone	Atividade física, mobilidade, sociabilidade e sono	Análise de associação	9.5
(Gu et al. 2017)	Bem-estar	Smartphone e dispositivo vestível	Atividade física e contexto ambiental	Identificação de informações de alto nível	5.5
(Place et al. 2017)	Depressão e ansiedade	Smartphone	Atividade física, sociabilidade, mobilidade, padrões de uso de dispositivos e pistas vocais	Predição do estado mental	9
(Wang et al. 2017)	Esquizofrenia	Smartphone	Atividade física, sono, sociabilidade, mobilidade, padrões de uso de dispositivos e contexto ambiental	Predição do estado mental	10.5
(Harari et al. 2017)	Bem-estar	Smartphone	Atividade física e sociabilidade	Identificação de informações de alto nível	9.5
(Servia-Rodríguez et al. 2017)	Humor	Smartphone	Atividade física, sociabilidade, mobilidade e contexto ambiental	Classificação do estado mental e análise de associação	9.5
(Wang et al. 2016)	Esquizofrenia	Smartphone	Atividade física, sociabilidade, mobilidade, padrões de uso de dispositivos, sono e contexto ambiental	Previsão do estado mental e análise de associação	10.5
(Wahle et al. 2016)	Depressão	Smartphone	Atividade física, mobilidade, sociabilidade e padrões de uso de dispositivos	Classificação do estado mental	8.5
(Farhan et al. 2016)	Depressão	Smartphone	Atividade física, padrões de uso de dispositivos, contexto ambiental, sociabilidade e mobilidade	Classificação do estado mental	9

(Demasi, Aguilera e Recht 2016)	Depressão	Smartphone	Atividade física e sono	Predição do estado mental e análise de associação	9
(Grñerbl et al. 2015)	Transtorno bipolar	Smartphone	Sociabilidade, mobilidade e pistas vocais	Classificação do estado mental	9
(Lane et al. 2014)	Bem-estar	Smartphone	Atividade física, sociabilidade, sono	Identificação de informações de alto nível	10