



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
Fevereiro, 2022

Ivan Rodrigues de Moura

Identificação de Padrões Multimodais de Comportamento Humano Utilizando Fenotipagem Digital

Artigo de Qualificação

Artigo de Qualificação apresentado 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
Coorientador: Prof. Dr. Luciano Reis Coutinho

São Luís - MA
Fevereiro, 2022

Ivan Rodrigues de Moura

Identificação de Padrões Multimodais de Comportamento Humano Utilizando Fenotipagem Digital

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

Aprovada pela banca examinadora:

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. Antonio Marcus de Andrade Paes

Universidade Federal do Maranhão

Prof. Dr. Ricardo de Andrade Lira Rabêlo

Universidade Federal do Piauí

Prof. Dr. Fábio Kon

Universidade de São Paulo

Prof. Dr. Markus Endler

Pontifícia Universidade Católica do Rio de Janeiro

São Luís - MA, 10 de Fevereiro de 2022

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 coorientado, 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 FAPEMA - Código de Financiamento BD-01075/20.

"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 objetivas 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 multimodais de múltiplas situações de interesse. 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.

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 objective 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 multimodal patterns of multiple situations of interest. 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.

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

Lista de ilustrações

Figura 1 – Associação entre situações e sensores ubíquos.	19
Figura 2 – Representação dos níveis de dados de contexto	28
Figura 3 – Representação da quantificação momento a momento do comportamento humano.	28
Figura 4 – Modelo de dados de fenótipos digitais.	29
Figura 5 – Modelo de dados de padrões de comportamento.	30
Figura 6 – Fluxo de trabalho de detecção de mudança de padrão comportamento.	33
Figura 7 – Fluxo de trabalho de detecção de comportamento anormal.	33
Figura 8 – Exemplo de conjuntos difusos para modelar mudanças comportamentais.	34
Figura 9 – Regras fuzzy para mudança de comportamento.	34
Figura 10 – Arquitetura de software do framework proposto.	35
Figura 11 – Diagrama de atividades da solução proposta.	37
Figura 12 – Estrutura do fluxo de eventos comportamentais gerado.	38
Figura 13 – Desenho experimental para avaliar o número de observação apropriado para projetar padrões.	40
Figura 14 – Nível médio de previsão do comportamento.	40
Figura 15 – Desenho experimental para avaliar a estabilidade da rotina comportamental.	40
Figura 16 – Nível médio de estabilidade da rotina comportamental.	40
Figura 17 – Correlação (Corr) entre a capacidade média de modelagem e a estabilidade média de rotina para cada CA.	41
Figura 18 – Mesclagem dos dados <i>HomeStay</i> dos usuários <i>u42</i> e <i>u58</i>	42
Figura 19 – Fluxo de detecção de mudança no comportamento <i>HomeStay</i>	42

Lista de tabelas

Tabela 1 – Definição dos tipos de estudos revisados.	24
Tabela 2 – Critérios de seleção.	55

Lista de abreviaturas e siglas

DPMH	<i>Fenotipagem Digital da Saúde Mental</i>
FPM	<i>Mineração de Padrões Frequentes</i>
CA	<i>Atributos de contexto</i>
ML	<i>Aprendizado de Máquina</i>
RMSE	<i>Raiz do Erro Quadrático Médio</i>
BPRS	<i>Escala de Avaliação Psiquiátrica Breve</i>
FIS	<i>Sistema de Inferência Fuzzy</i>
SJB	<i>Scheduler and Job Management</i>
OCS	<i>Observation Control Service</i>
CDS	<i>Change Detection Service</i>
DP	<i>Desvio Padrão</i>

Sumário

1	INTRODUÇÃO	13
1.1	Questões de Pesquisa	15
1.1.1	Quais aspectos do comportamento humano podem ser inferidos considerando sociabilidade, mobilidade, atividade física e sono? . . .	15
1.1.2	(RQ2) Como identificar mudanças no padrão comportamental de indivíduos?	15
1.2	Objetivos	15
1.3	Metodologia	16
1.4	Estrutura do Artigo	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Fenotipagem Digital da Saúde Mental	19
2.2	Fontes de Dados de Contexto Usadas para Detectar Situações de Interesse	19
3	TRABALHOS RELACIONADOS	23
3.1	Resultados	23
3.1.1	Identificação de Informações de Alto Nível	23
3.1.2	Classificação e Predição de Estados Mentais	24
3.1.3	Associação entre Características Comportamentais e Estados Mentais	25
3.2	Discussão	26
4	SOLUÇÃO PROPOSTA	27
4.1	Modelo de Dados para Fenotipagem Digital	27
4.2	Algoritmo para Identificação de Padrões Comportamentais	30
4.2.1	Padrões Comportamentais Sensíveis ao Contexto	32
4.3	Deteção de Mudança Comportamental	32
4.3.1	Modelagem de Conhecimento Especializado	33
4.4	Framework para a Fenotipagem Digital de Comportamentos	34
4.4.1	Fluxo de Processamento	35
5	AVALIAÇÃO EXPERIMENTAL PRELIMINAR	38
5.1	Descrição dos Dados	38
5.2	Capacidade de Modelar Rotina Comportamental	39
5.3	Sensibilidade nas Deteções de Mudanças de Padrões Comportamentais	41
6	DISCUSSÃO	43

6.1	Cronograma	44
7	CONCLUSÃO	46
7.1	Publicações	46
	REFERÊNCIAS	48
A	APÊNDICE A - PROTOCOLO DE PESQUISA DA RSL CONDUZIDA	54
A.1	Metodologia	54
A.1.1	Estratégia de Busca	54
A.1.2	Seleção de Estudos	54
A.1.3	Extração de Dados	55

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). Este 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). Esta 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).

Em geral, as soluções para Fenotipagem Digital da Saúde Mental (DPMH) 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 mentais (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 mental, ainda existem questões em aberto nesta área de pesquisa, a saber: (i) é preciso desenvolver novas soluções capazes de fornecer informações objetivas, como padrões comportamentais que refletem o hábito de indivíduos, para profissionais de saúde, uma vez que as soluções de fenotipagem digital presentes na literatura se concentram no diagnóstico de um determinado estado de saúde; (ii) é necessário que as soluções considerem informações contextuais (por exemplo, dia da semana, dados climáticos, dados ambientais) na identificação e no processamento de fenótipos digitais para viabilizar a modelagem adequada da variação comportamental normal dos indivíduos; (iii) a detecção de comportamentos anômalos e mudanças comportamentais também é uma questão a ser explorada para permitir que profissionais especializados investiguem se há relação entre a mudança identificada e o estado de saúde do paciente; e (iv) é necessário investigar o desenvolvimento de soluções capazes de aprender continuamente ao longo do tempo, uma vez que as soluções de fenotipagem digital presentes na literatura projetam modelos de aprendizado de máquina estáticos a partir de dados históricos.

1.1 Questões de Pesquisa

Neste projeto pretendemos abordar as seguintes questões de pesquisa:

1.1.1 Quais aspectos do comportamento humano podem ser inferidos considerando sociabilidade, mobilidade, atividade física e sono?

Esta questão de pesquisa busca entender quais aspectos dos comportamentos de sociabilidade, mobilidade, atividade física e sono podem ser inferidos a partir de dados de contexto provenientes de *smartphones* e dispositivos vestíveis. Um aspecto importante a ser investigado é o desenvolvimento de um modelo de dados para fenotipagem digital que permita representar estes comportamentos humano e acompanhá-los ao longo do tempo. Será investigado ainda o aprendizado de padrões comportamentais de indivíduos a partir de dados de sensores. Além disso, também será explorado uma forma flexível para a especificação do contexto a ser utilizado na construção destes padrões (por exemplo, padrão de sociabilidade nos fins de semana e nos dias de trabalho).

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. Estes modelos podem se tornar ineficazes à medida que os indivíduos mudem o comportamento de sua rotina, necessitando de modelos capazes de aprender continuamente. Por outro lado, pesquisadores tem investigado técnicas computacionais para reconhecer mudanças na distribuição dos fluxos de dados oriundos de sensores em outros contextos, como a mobilidade urbana ([Ta-Shma et al. 2017](#)). A detecção destas mudanças é realizada através do monitoramento de propriedades estatísticas apresentadas pela distribuição dos dados. Assim, esta questão de pesquisa pretende investigar abordagens capazes de monitorar de maneira automática mudanças de padrões comportamentais.

1.2 Objetivos

O objetivo geral desta pesquisa é a 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.

São objetivos específicos deste projeto de pesquisa:

- A especificação de um modelo de dados para a representação de fenótipos digitais e padrões de comportamentos que considerem múltiplos aspectos do comportamento humano.
- A realização de uma revisão sistemática da literatura sobre a fenotipagem digital de comportamentos humanos para o monitoramento de saúde.
- O desenvolvimento de abordagens para o aprendizado incremental de padrões de comportamentos humanos a partir de fenótipos digitais em contextos estabelecidos (e.g. dias de trabalho, finais de semana).
- O desenvolvimento de uma abordagem para a identificação de mudanças no padrão de comportamento dos indivíduos.
- O desenvolvimento de uma abordagem para permitir profissionais especializados especificar previamente padrões comportamentais de interesse para serem monitorados continuamente no decorrer do processamento realizado pela solução proposta.
- O desenvolvimento de um aplicativo móvel para gerenciar a solução e notificar e visualizar as detecções de padrões e mudanças comportamentais.

1.3 Metodologia

O ponto de partida desta pesquisa foi a realização de uma Revisão Sistemática da Literatura (RSL) 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 seguintes bibliotecas digitais: *Science Direct*, *PubMed*, *Scopus*, *ACM Digital Library*, *Web of Science*, e *IEEE Xplore*. A partir de critérios de inclusão e exclusão, foram selecionados os estudos que apresentam escopo adequado ao objetivo da RSL conduzida. Por fim, foi realizada uma análise aprofundada dos estudos selecionados, consolidando uma profunda compreensão do estado da arte nesta área de pesquisa. Especificamente, os resultados desta revisão fornecem as seguintes contribuições: (i) caracterizamos e resumimos os estudos selecionados; (ii) respondemos a questões fundamentais para o esclarecimento desta área de investigação, tais como uma análise aprofundada dos processos de caracterização de situações de interesse através de dados de contexto e a utilização da consciência destas situações para apoiar os profissionais de saúde mental;

e (iii) discutimos tendências e identificamos questões em aberto, que contribuem para delinear perspectivas para pesquisas futuras.

Em seguida, explorou-se a Mineração de Padrões Frequentes (FPM) para projetar um algoritmo capaz de detectar padrões comportamentais. FPM está relacionado a descoberta de padrões que ocorrem com frequência significativa em diferentes tipos de coleção de dados, como bancos de dados, arquivos de texto e fluxos de dados. Assim, o desenvolvimento do algoritmo proposto inspirou-se na ideia de encontrar conjuntos de itens que se repetem em uma quantidade suficiente para representar um padrão. Também explorou-se a possibilidade de utilizar informações contextuais para enriquecer o algoritmo proposto com Atributos de contexto (CAs). Adicionalmente, utilizou-se 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.

Através dos resultados da RSL conduzida, foram identificados os principais aspectos do comportamento humano que são de interesse para os profissionais especializados e que possam ser inferidos a partir de fluxos de dados de sensores. Deste modo, a partir deste levantamento, foi projetado um modelo de dados baseado no conceito de fenótipo digital capaz de representar os diversos aspectos do comportamento humano. Formalizou-se as estruturas de dados para modelar os fenótipos digitais identificados pelos dispositivos ubíquos, assim como os padrões e mudanças de comportamentos reconhecido 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 de dados desenvolvido. Este framework provém interfaces que abstraem a complexidade inerente a 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. Esta escolha foi tomada devido a necessidade de utilizar um processo ágil de desenvolvimento, que seja razoavelmente documentado e fácil de ser gerenciado. Sempre que possível, o desenvolvimento incremental do software utilizou técnicas de refatoração.

A princípio, avaliou-se o algoritmo proposto através de fluxos de dados derivados de *datasets* com dados reais. Primeiro, Avaliou-se o algoritmo em termos de sua capacidade de projetar padrões multimodais 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 validou-se a abordagem de detecção de mudança de comportamentos através do processamento de fluxo de dados com mudanças na distribuição dos dados controladas de forma explícita (e.g., concatenando-se fluxos de dados de indivíduos com comportamentos diferentes), permitindo identificar com precisão o ponto no qual a mudança deve ser detectada, assim como a adaptação ao novo padrão a ser identificado. Verificou-se se os pontos de detecção de mudanças da solução proposta refletem a verdade fundamental apresentada pelas anotações de mudanças.

1.4 Estrutura do Artigo

O resto deste trabalho está organizado como segue. O Capítulo 2 apresenta a fundamentação teórica, abordando conceitos sobre a fenotipagem digital da saúde mental. O Capítulo 3 apresenta uma revisão sistemática sobre o sensoriamento multimodal de comportamentos humanos para o monitoramento da saúde mental. O Capítulo 4 apresenta a versão atual da solução proposta. O Capítulo 5 expõe a avaliação experimental da versão atual da solução proposta usando um fluxo de dados do mundo real. O Capítulo 6 expõe a discussão sobre as contribuições já obtidas, as limitações atuais do trabalho, as atividades futuras e o cronograma das próximas atividades. No final, apresenta-se as conclusões no Capítulo 7.

2 Fundamentação Teórica

2.1 Fenotipagem Digital da Saúde Mental

A Fenotipagem Digital da Saúde Mental 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 mental (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 (e.g., contexto ambiental, humor, dicas vocais) para profissionais. As soluções DPMH realizam detecção multimodal (isto é, usando dois ou mais tipos de dados de contexto) de várias situações (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, as soluções DPMH podem fornecer informações objetivas para apoiar a tomada de decisão por especialistas em saúde mental.

2.2 Fontes de Dados de Contexto Usadas para Detectar Situações de Interesse

O sensoriamento multimodal pode detectar automaticamente diferentes situações de interesse, possibilitando utilizar essa consciência de situação para apoiar o monitoramento e a avaliação da saúde mental. A Figura 1 mostra as fontes de dados de contexto utilizadas no processo de identificação de cada situação.

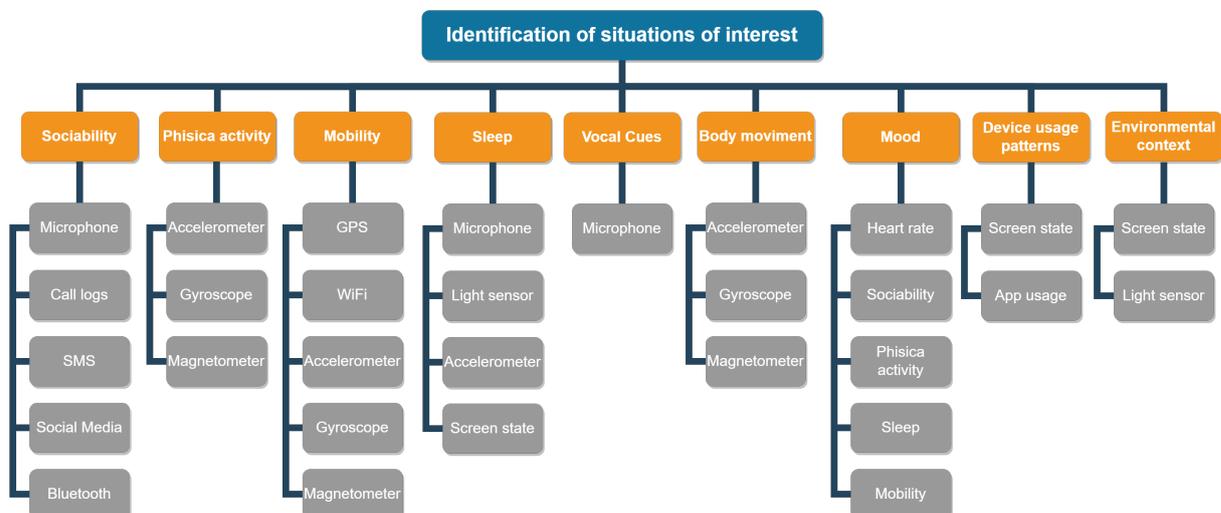


Figura 1 – Associação entre situações e sensores ubíquos.

Sociabilidade. Interações sociais físicas e virtuais são identificadas através do processamento de dados de contexto (Adler et al. 2020, Levine et al. 2020). As soluções de DPMH identificam interações virtuais através do processamento de dados de redes sociais online (por exemplo, *Twitter*, *Instagram* e *Facebook*) e de aplicativos de comunicação (por exemplo, chamadas e SMS) (Dogrucu et al. 2020, Sano et al. 2018). Esses sensores virtuais geram um alto fluxo de dados comportamentais das interações dos indivíduos em tais ambientes virtuais (por exemplo, estabelecendo novas amizades, escrevendo comentários e compartilhando fotos e vídeos), o que permite o registro de laços sociais, frequência e duração das interações. As interações físicas são identificadas principalmente usando dados de microfone e interfaces de rede sem fio (Doryab et al. 2019, Wang et al. 2017, Wang et al. 2018, Lane et al. 2014). A maioria dos estudos que abordam esse tipo de interação social usa algoritmos de Aprendizado de Máquina (ML) para reconhecer a fala humana em amostras de áudio (Wang et al. 2017, Harari et al. 2017, Farhan et al. 2016). Além disso, os pesquisadores usam interfaces *Bluetooth* para reconhecer co-localizações de indivíduos ao escanear dispositivos, inferindo informações como o número de pessoas co-localizadas (Wang et al. 2017, Doryab et al. 2019).

Mobilidade. Os padrões de mobilidade representam informações de alto nível derivadas da sequência de movimento (ou falta de movimento) dos indivíduos. Esses padrões são projetados principalmente a partir do processamento de dados GPS, que permite identificar os traços de mobilidade dos participantes (Wang et al. 2017, Farhan et al. 2016, Wahle et al. 2016). Em princípio, os pesquisadores aplicam algoritmos de agrupamento (por exemplo, *K-means*, *DBScan*) a amostras de GPS para reconhecer locais visitados ao longo do tempo. Além disso, as soluções revisadas identificam a semântica dos locais visitados usando dados de *OpenStreetMap* e *Foursquare APIs* (Cai et al. 2018, Boukhechba et al. 2018), reconhecendo assim informações relacionadas a pontos de interesse (por exemplo, trabalho, alimentação, lazer). Em seguida, as técnicas estatísticas (por exemplo, média, variância, desvio padrão, entropia) descrevem os padrões de mobilidade, como a distância total percorrida, o tempo gasto em pontos de interesse e a duração da viagem.

Atividade física. As atividades físicas apresentam forte associação com estados mentais (Paluska e Schwenk 2000), representando situações de interesse para profissionais de saúde mental. Por esta razão, pesquisadores usam técnicas computacionais para reconhecer essas situações por meio do processamento de dados coletados de sensores inerciais (por exemplo, acelerômetro, magnetômetro, giroscópio). As soluções de DPMH usam algoritmos de ML para classificar diferentes atividades físicas (por exemplo, estacionário, caminhando, correndo) (Wang et al. 2018, Adler et al. 2020). Estes modelos de classificação são projetados a partir de características (por exemplo, média, desvio padrão, curtose, assimetria, raiz quadrada média) extraídas de dados de

sensores inerciais. Os pesquisadores também usam a *Google Activity Recognition API* e a *API Apple CMMotionActivity* para detectar diferentes tipos de atividades realizadas por indivíduos (Cai et al. 2020, Levine et al. 2020).

Qualidade do Sono. O monitoramento da qualidade do sono pode ser 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. Por exemplo, (Narziev et al. 2020) implementaram um classificador de sono com base na verificação das seguintes condições: quando o smartphone está estático, não em uso e o ambiente está escuro. Além disso, alguns trabalhos processam dados fornecidos pelo *smartwatch Fitbit* para reconhecer informações como duração e início do sono, minutos para adormecer e acordar (Haines-Delmont et al. 2020, Adler et al. 2021).

Pistas vocais. Essa situação representa variações significativas no som da voz devido às alterações cognitivas, afetivas e fisiológicas causadas pela degradação do estado de saúde. 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 microfones para identificar pistas vocais na voz de indivíduos durante conversações. Estas pistas vocais são mensuradas em quadros sobrepostos e, em seguida, agregados por meio de estatística de resumo para projetar características comportamentais relevantes para soluções de fenotipagem digital.

Movimento corporal. Esta situação representa os movimentos corporais involuntários dos indivíduos durante situações estressantes. Esses movimentos corporais são reconhecidos por meio de técnicas computacionais para processar dados de sensores inerciais, como acelerômetros, giroscópios e magnetômetros. Especificamente, os pesquisadores usam essa situação para identificar os movimentos corporais de indivíduos socialmente ansiosos durante os períodos em que eles se envolvem em alguma forma de interação social e como esses movimentos diferem em função do papel social da localização (Gong et al. 2019).

Humor. Esta situação é um estado afetivo que representa o grau de bem-estar psicológico e emocional de um indivíduo (isto é, estado de ânimo positivo ou negativo). 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). Estas características multimodais anotadas com autorrelatos de humor viabilizam os pesquisadores projetar modelos de ML para classificação de humor positivo e negativo.

Padrões de uso do dispositivo. Os sistemas de DPMH inferem essa situação a partir de registros de aplicativos móveis e do estado da tela (ou seja, 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. Esta situação é derivada principalmente de sensores de luz 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 Trabalhos Relacionados

Muitos estudos usaram a DPMH para complementar os métodos tradicionais de avaliação e monitoramento da saúde mental. Assim, a literatura atual sobre DPMH 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 RSL para reconhecer o estado da arte sobre o sensoriamento multimodal de situações de interesse para profissionais de saúde mental. Revisou-se estudos que propuseram soluções de DPMH 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. Apresenta-se em detalhes o protocolo de pesquisa desta RSL no Apêndice A. A seguir, apresenta-se os resultados da RSL conduzida.

3.1 Resultados

A partir da literatura revisada, identificou-se que as soluções de DPMH realizam o sensoriamento multimodal de múltiplas situações de interesse para os 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 apoiam o monitoramento e a avaliação da saúde mental de diferentes formas. Portanto, é possível classificar 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; e (iv) associação de características comportamentais com estados/transtornos mentais. A Tabela 1 apresenta a definição de cada tipo de estudo revisado. No restante desta seção, descreve-se esses quatro tipos diferentes de estudos.

3.1.1 Identificação de Informações de Alto Nível

(Barnett et al. 2018) aplicou um método de detecção de outlier multivariado inspirado na abordagem de *Filzmoser* (Filzmoser 2004) para analisar mudanças nos padrões de mobilidade e sociabilidade dias antes das recaídas esquizofrênicas. (Harari et al. 2017) usaram um aplicativo móvel para identificar padrões de estabilidade e mudanças no comportamento social e físico dos alunos. (Gu et al. 2017) propôs um sistema para detectar atividades familiares (por exemplo, ocorrência de refeição, assistir televisão, conversas) a partir de processamento de dados de *smartphones* e *smartwatch*. O sistema utiliza essa consciência de situação para propor atividades ao

Tabela 1 – Definição dos tipos de estudos revisados.

Tipo de estudo	Definição
Reconhecimento de informações de alto nível	Neste tipo de estudo, visa-se processar fenótipos digitais para reconhecer informações sobre aspectos comportamentais dos indivíduos, como padrões, mudanças comportamentais e quantificações comportamentais (por exemplo, duração e frequência do comportamento).
Classificação de estados/transtornos mentais	Este tipo de trabalho se concentra no treinamento de algoritmos de ML para classificar estados/transtornos mentais
Predição de estados/transtornos mentais	Trabalhos focados em predição têm como objetivo projetar modelos de ML usando características extraídas das situações identificadas para prever estados/transtornos mentais.
Associação de características comportamentais com estados/transtornos mentais	Neste tipo de estudo, as soluções analisam a associação existente entre estados/transtornos mentais e características extraídas das situações identificadas.

usuário, que contribuam para o bem-estar físico e mental.

Alguns estudos se concentraram na quantificação de aspectos do comportamento humano. (Wahle et al. 2016) demonstraram um aplicativo móvel capaz de medir o envolvimento dos usuários em atividades físicas, socialização e mobilidade. Além disso, este aplicativo recomenda atividades com base no nível de envolvimento do usuário, hora do dia e local. (Lane et al. 2014) apresentou um sistema que monitora a atividade física, o sono e a sociabilidade do usuário. Os pesquisadores usaram regressão linear para quantificar o nível de envolvimento do usuário nesses comportamentos humanos.

3.1.2 Classificação e Predição de Estados Mentais

(Narziev et al. 2020) treinou um modelo de classificação usando características extraídas de atividade física, humor, sociabilidade e sono para reconhecer sintomas de depressão (ou seja, normal, leve, moderado e grave) que atingiu 96% precisão. (Dogrucu et al. 2020) usaram sensores de *smartphone* e redes sociais para treinar modelos de aprendizado de máquina para classificar sintomas depressivos e suicidas. O modelo proposto para classificar a depressão atingiu uma sensibilidade de 75% e especificidade de 79,2%, enquanto a tarefa de triagem de ideação suicida atingiu uma sensibilidade de 86% e especificidade de 72,5%. (Doryab et al. 2019) propuseram um pipeline de aprendizado de máquina para detectar o nível de solidão e a mudança no nível de solidão que atingiu 80,2% e 88,4% de precisão, respectivamente. (Sano et al. 2018) usaram medidas fisiológicas e comportamentais para detectar estresse (ou seja, alto ou baixo) com uma precisão de 78,3% e classificar o bem-estar (ou seja, pontuação de bem estar mensurada a partir de uma pesquisa de saúde curta de 12 itens) com uma precisão de 87%.

Alguns estudos têm como objetivo prever estados mentais. (Morshed et al. 2019)

desenvolveram um modelo de aprendizado de máquina capaz de prever instabilidade de humor usando três semanas de dados detectados passivamente de *smartphones* e sensores vestíveis. (Barnett et al. 2018) identificaram que a taxa de anomalias nos comportamentos de sociabilidade e mobilidade detectadas nas 2 semanas anteriores à recaída esquizofrênica foi 71% maior do que a taxa de anomalias durante outros períodos de tempo. (Cai et al. 2018) desenvolveram modelos de predição baseados em características dos comportamentos de mobilidade e atividade física para prever estados afetivos negativos, em que modelos personalizados (isto é, projetado usando dados de um indivíduo específico) atingiram Raiz do Erro Quadrático Médio (RMSE) ≈ 13 e modelos generalizados (isto é, projetado usando dados de todos indivíduos) atingiram RMSE ≈ 22 . (Wang et al. 2018) usaram regressão logística para prever o estado depressivo de alunos com 81,5% de *recall* e 69,1% de precisão. (Wang et al. 2017) desenvolveram um sistema chamado *CrossCheck* para prever sintomas psiquiátricos em pacientes esquizofrênicos. Este sistema pode prever a pontuação da Escala de Avaliação Psiquiátrica Breve (BPRS) de 7 itens com base em características dos comportamentos de atividade física, sociabilidade, mobilidade, sono, padrão de uso de dispositivo e contexto ambiental.

3.1.3 Associação entre Características Comportamentais e Estados Mentais

(Cai et al. 2020) identificou correlações entre características comportamentais coletadas passivamente (por exemplo, tempo gasto em cada localização semântica, nível de atividade física) e humor autorrelatado por pacientes com câncer de mama. (Doryab et al. 2019) encontraram associações entre os níveis de solidão dos alunos com seus padrões de comportamento (por exemplo, diminuição da atividade física, menos tempo de socialização). (Boukhechba et al. 2018) identificaram correlações significativas entre mobilidade, atividade física e padrões de comunicação com relatos de alunos sobre ansiedade social e sintomas depressivos. (Cai et al. 2018) reconheceram que as flutuações no estado afetivo negativo dos participantes estão associadas à mobilidade e à atividade física. (Servia-Rodríguez et al. 2017) aplicou o coeficiente de correlação de *Pearson* a um conjunto de dados em grande escala para encontrar associações entre as rotinas dos indivíduos (ou seja, mobilidade, atividade física, sociabilidade) com seu humor autorrelatado. (Wang et al. 2017) encontraram uma forte correlação entre sociabilidade, sono, atividade física e mobilidade de alunos com autorrelatos de bem-estar mental. (Gong et al. 2019) identificaram que indivíduos socialmente ansiosos aumentam seus movimentos corporais (reconhecidos por dados do acelerômetro) em períodos que interagem socialmente, principalmente quando essas interações ocorrem no ambiente de trabalho e fora de casa.

3.2 Discussão

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.

4 Solução Proposta

Nesta seção, apresentamos a solução proposta para detectar padrões e mudanças comportamentais sensíveis ao contexto. O objetivo desta solução é fornecer informações sobre padrões comportamentais, comportamentos anômalos e mudanças de padrões comportamentais com base no processamento de fenótipos digitais. Este 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 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 2 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 proveniente 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 especificação do modelo de dados para a solução proposta foi baseada na definição de fenotipagem digital (Torous et al. 2016). As estruturas de dados da solução representam a quantificação momento a momento de situações de interesse para os profissionais especializados. A Figura 3 mostra uma representação da quantificação momento a momento do comportamento humano. Primeiro, subdividimos o dia (ou seja, tempo circular de 24 horas) em intervalos de tempo para representar um momento, no qual quantificamos o comportamento do indivíduo monitorado (por exemplo, número de eventos de sociabilidade e atividade física) para cada intervalo de tempo. Portanto, a sequência de eventos comportamentais em cada intervalo de tempo representa um

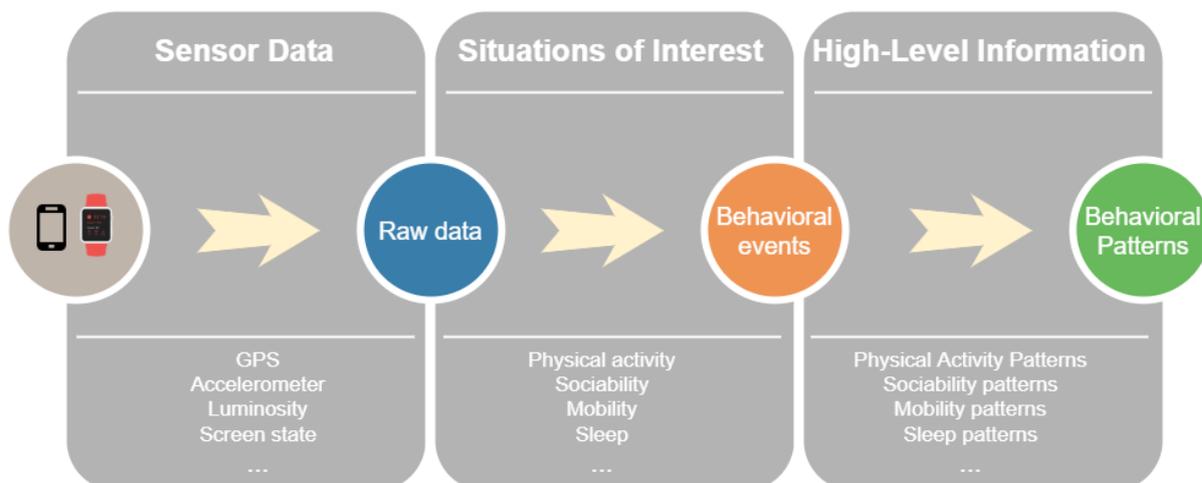


Figura 2 – Representação dos níveis de dados de contexto

fenótipo digital. Usamos Atributos de Contextos (CAs) para modelar a variação no comportamento de acordo com situações específicas, como finais de semana, feriados e dias chuvosos.

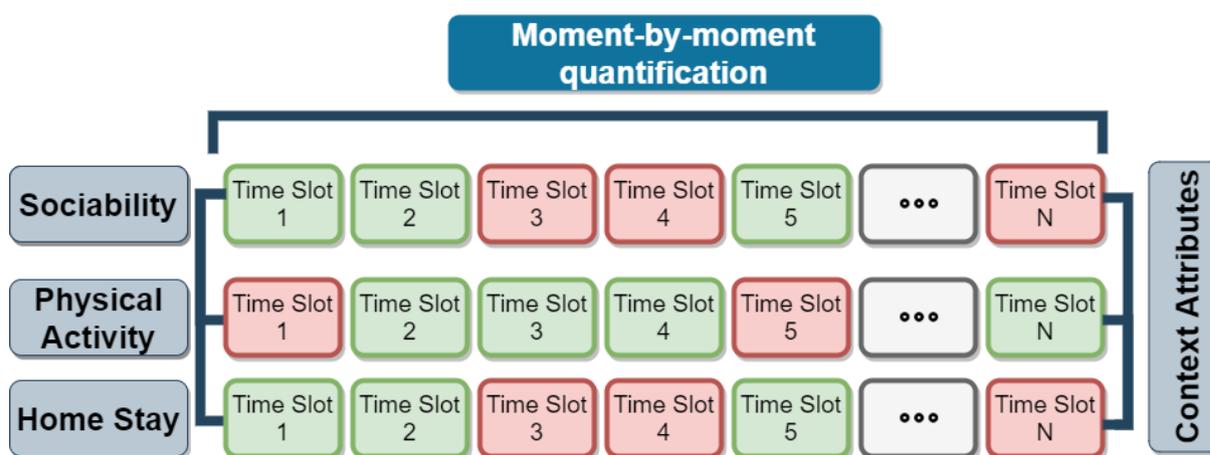


Figura 3 – Representação da quantificação momento a momento do comportamento humano.

A partir dessa representação de fenotipagem digital, especificou-se um modelo de dados que estrutura padrões de comportamento enriquecidos por contexto para representar os momentos (ou seja, intervalos de tempo) em que as atividades ocorrem rotineiramente. Primeiro, modelou-se a estrutura de dados dos fenótipos digitais que são processados pela solução proposta (Figura 4). A classe *Attribute* contém o nome, valor e o tipo de atributos específicos de um determinado evento comportamental, permitindo o desenvolvedor adicionar atributos personalizados aos eventos, como distância percorrida, duração total do evento e velocidade média. A classe *BehavioralEvent* estrutura eventos comportamentais, que contém informações sobre o tipo (por exemplo, sociabilidade e mobilidade) e subtipo (por exemplo, conversação presencial, conversação telefônica e conversação SMS) do comportamento identificado, data e hora de início

e fim do evento, lista de atributos específicos do comportamento identificado e a lista de atributos de contexto. Por fim, a classe *DigitalPhenotype* modela fenótipos digitais, que são composto pelo identificador do usuário e por um conjunto de eventos comportamentais que aconteceram em um determinado momento.

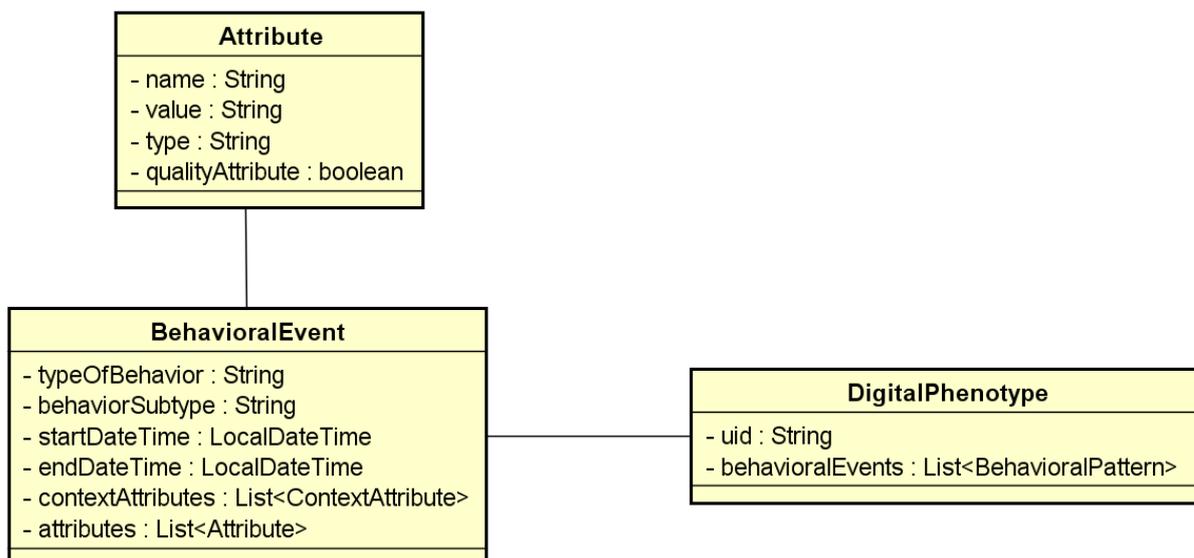


Figura 4 – Modelo de dados de fenótipos digitais.

A Figura 5 apresenta o modelo de dados proposto para estruturar padrões e mudanças comportamentais. Os padrões representam o conjunto de slots (ou seja, intervalos de tempo) que modelam a rotina do indivíduo monitorado. Portanto, definimos esse modelo de dados com base nas seguintes classes: (i) *TimeInterval* modela um conjunto de slots adjacentes que representa um intervalo de tempo [startTime, endTime]; (ii) *ContextAttribute* caracteriza CAs, incluindo tipo (por exemplo, temporal, espacial), nome e descrição; (iii) *BehavioralPattern* estrutura padrões comportamentais, que contém o identificador do usuário, a data e hora da detecção, o tipo de comportamento, um conjunto de subtipos de comportamento, o CA, e os intervalos de tempo que compõem o hábito comportamental; (iv) *BehavioralPatternChange* contém informações sobre mudanças de padrões, como o identificador do usuário, a data e hora da detecção da mudança, a similaridade, a pertinência da mudança, o novo padrão reconhecido e o antigo padrão; e (v) *AbnormalBehavior* modela comportamentos anormais, estruturando atributos como o identificador do usuário, a data e hora da detecção do comportamento anormal, a similaridade, a pertinência da anormalidade, o padrão comportamental e a observação comportamental anormal.

Este modelo de dados é utilizado para apoiar o processo de desenvolvimento da solução proposta, o que facilita o entendimento das estruturas de dados de padrões e mudanças de comportamento. Enfatizamos que o modelo de dados apresentado pode ser facilmente utilizado para várias situações (por exemplo, mobilidade, sono) e,

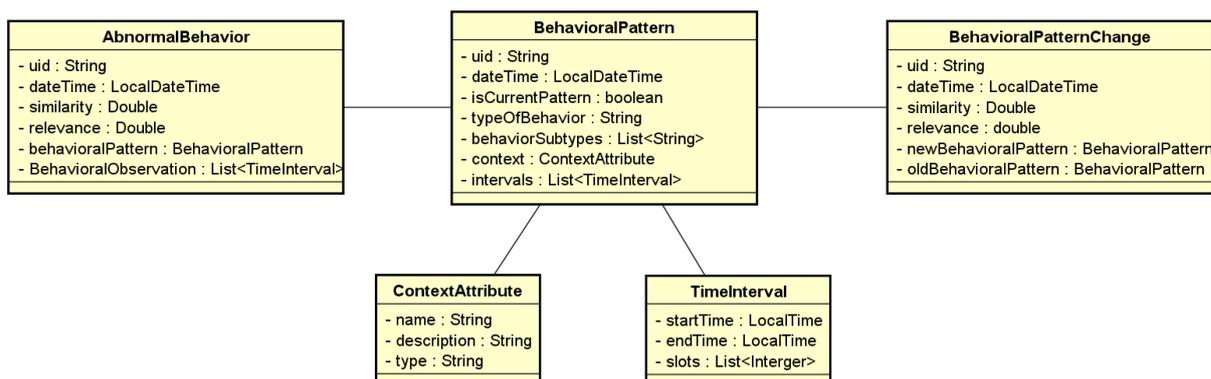


Figura 5 – Modelo de dados de padrões de comportamento.

consequentemente, permite que a solução suporte a detecção multimodal de várias situações de interesse.

4.2 Algoritmo para Identificação de Padrões Comportamentais

Neste trabalho, considera-se que, se um comportamento é detectado com frequência em um determinado intervalo de tempo, esse intervalo compõe o padrão de comportamental dos indivíduos monitorados. Assim, definimos padrões comportamentais como *períodos do dia em que o indivíduo costuma apresentar determinado comportamento, ou seja, o conjunto de intervalos de tempo em que um comportamento ocorre habitualmente*. Assim, inspirado na FPM, o algoritmo processa fluxo de dados para reconhecer intervalos de tempo $[T_{\text{início}}, T_{\text{fim}}]$ nos quais o número de ocorrências de eventos é superior a $\phi * |n|$. Nesse sentido, $|n|$ é considerado como o número total de eventos comportamentais processados em uma janela de dados definida para modelar um comportamento, e ϕ é um parâmetro a ser definido manualmente, que é responsável por regular a variação do valor do limiar de suporte. Por exemplo, valores maiores de ϕ resultam em limiares mais exigentes que requerem maior quantidade de eventos comportamentais para um intervalo de tempo ser considerado um padrão comportamental.

A entrada do algoritmo é um fluxo de eventos comportamentais que possui a hora de início e fim de cada situação identificada. A primeira etapa do algoritmo é determinar, com base no registro de data e hora, a qual período de tempo cada evento comportamental pertence. Para isso, o algoritmo segmenta o tempo em slots de tempo de tamanhos iguais. Cada slot representa uma fatia do dia e possui um identificador sequencial. Para definir o tamanho do slot (ou seja, em quantos períodos deve dividir o dia), especifica-se um valor para o parâmetro t , que é responsável por criar um vetor de contadores com o número total de slots. Por exemplo, se o dia for dividido em períodos de 30 minutos, o parâmetro t deve ser igual a 0,5, já que $\frac{24}{t} = 48$ slots. Essa equação é

responsável por criar a estrutura de armazenamento para contagem de ocorrências de eventos comportamentais em cada slot.

Após definir o tamanho dos slots, descreve-se agora a fase de contagem do algoritmo. Nesse estágio, o algoritmo usa o registro de data e hora de cada evento para definir seu slot. Ao identificar o slot do evento comportamental, o algoritmo incrementa o valor do contador que representa esse slot na estrutura responsável por armazenar essas estatísticas. Portanto, ao processar o fluxo de eventos, a frequência dos eventos comportamentais em cada slot é atualizada.

A próxima fase do algoritmo é a descoberta do padrão comportamental, que usa o resumo da fase de contagem para identificar intervalos frequentes de comportamento. Nesta fase, é necessário definir quais slots possuem um número suficiente de eventos, ou seja, uma quantidade que os habilite a serem slots candidatos para formar um período frequente. Para isso, a contagem do número de eventos comportamentais do slot analisado deve ser maior ou igual a S_{th} . O algoritmo usa a Equação 4.1 para definir o valor de S_{th} . Nesta equação, o parâmetro θ é inserido para regular a variação do valor do limiar S_{th} .

$$S_{th} = |n| * \theta * \frac{1}{24} \quad (4.1)$$

A equação 4.2 é responsável por percorrer o conjunto de slots C_s e verificar quais slots são candidatos a formar um padrão de comportamento, atribuindo zero à contagem dos slots não candidatos. Para isso, esta equação define a multiplicação entre a contagem de cada slot ($slot[i]$) e a função de passo unitário, que retorna valor zero em casos de argumentos negativos ($slot[i] - slot_{th} < 0$) e um para argumentos não negativos ($slot[i] - slot_{th} > 0$). No final, o array de slots C_s é enviado para o processo de identificação de intervalos frequentes.

$$C_s[i] := slot[i] * \mathbf{unit_step}(slot[i] - slot_{th}) \quad (4.2)$$

Por fim, após definir os requisitos para que um slot seja candidato, o próximo passo é identificar quais conjuntos de slots compõem um intervalo no qual o comportamento é habitual para o indivíduo monitorado. A equação 4.3 agrupa os slots candidatos não nulos adjacentes na matriz C_s em um padrão de comportamento. A função de etapa de unidade verifica se a soma das contagens de eventos para os slots agrupados (ou seja, intervalos de tempo) subtraídos de $\varphi * |n|$ resulta em um valor positivo. Se essa condição for satisfeita, o intervalo de tempo formado por esses conjuntos de slots adjacentes representa um padrão de comportamento. No final, o array P_s conterá os padrões de comportamento, ou seja, os intervalos de tempo em que as situações ocorrem rotineiramente para o indivíduo monitorado.

$$P_s[i : i + n] := \mathbf{unit_step}\left(\left(\sum_{j=i}^{i+n-1} C_s[j]\right) - \varphi * |n|\right) \quad (4.3)$$

4.2.1 Padrões Comportamentais Sensíveis ao Contexto

Até o momento, o algoritmo proposto permite identificar a rotina comportamental do indivíduo, mapeando os horários frequentes de início das situações de interesse. No entanto, essa análise livre de contexto pode resultar em ineficiência no delineamento do hábito comportamental, uma vez que o comportamento do indivíduo pode variar em função de contextos específicos, como dias de trabalho, fins de semana, dias chuvosos, entre outros. Para isso, utilizou-se a estratégia de CAs, na qual diversas escalas podem ser utilizadas para representá-los. Por exemplo, uma característica temporal pode ter várias escalas, como uma escala ampla, diferenciando dias da semana e finais de semana ou de forma mais específica, distinguindo cada dia da semana (por exemplo, segunda, terça, quarta-feira). Injeta-se esses CAs no fluxo de eventos comportamentais, que podem ser derivados diretamente das propriedades do evento (por exemplo, registro de data e hora) ou recuperadas de fontes externas (por exemplo, APIs de clima). Cada CA 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.

4.3 Detecção de Mudança Comportamental

Implementamos a detecção de mudança comportamental através do processamento de janelas de observações, tornando possível comparar a similaridade entre os padrões extraídos em diferentes janelas de observações. A Figura 6 apresenta o fluxo de trabalho de detecção de mudança de padrão comportamental. Especificamente, a solução aplica o coeficiente de Jaccard ($Jaccard(X,Y) = \frac{X \cap Y}{X \cup Y}$) entre os padrões para identificar mudanças comportamentais. Este 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 limite de mudança, a solução atualiza o padrão de referência com o novo padrão identificado e notifica as partes interessadas sobre a mudança de rotina. Similarmente, o fluxo de detecção de comportamentos anormais (Figura 7) consiste na aplicação do coeficiente de Jaccard para identificar semelhanças entre o padrão atual do indivíduo e suas observações comportamentais subsequentes.

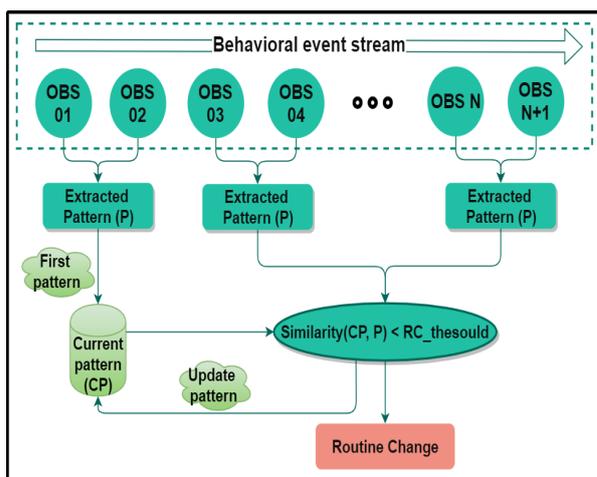


Figura 6 – Fluxo de trabalho de detecção de mudança de padrão comportamento.

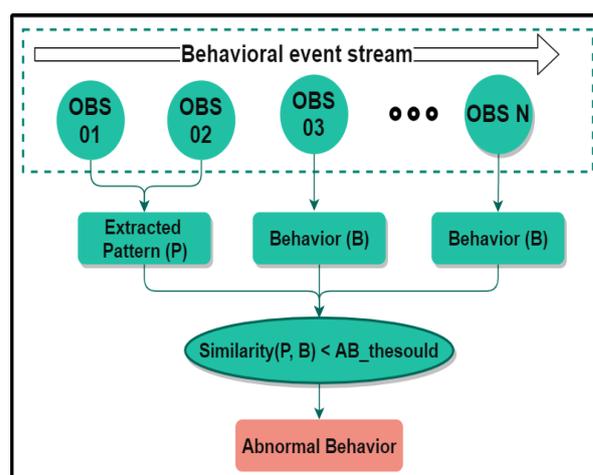


Figura 7 – Fluxo de trabalho de detecção de comportamento anormal.

4.3.1 Modelagem de Conhecimento Especializado

Analisando a solução para detecção de mudanças de comportamentos apresentada anteriormente, é possível identificar a necessidade de especialistas definirem um limiar de similaridade entre padrões. Assim, a especificação do limiar de mudança possui um caráter subjetivo, pois depende do conhecimento do especialista, tendo também uma essência imprecisa, pois um limiar limítrofe rígido pode não modelar adequadamente a mudança ocorrida. Assim, mitigou-se esta limitação através da implementação da modelagem do conhecimento especialista usando a logica *fuzzy*, que representa conhecimento impreciso por meio de conjuntos fuzzy (Zadeh 1996). Incorporou-se um Sistema de Inferência Fuzzy (FIS) na solução para detectar mudanças de comportamento considerando saídas que variam de *sem mudança* a *mudança* gradualmente.

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 8 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.

A solução proposta também exige a especificação de regras que norteiam a decisão do FIS. Estas regras possuem a estrutura básica $\langle \text{condição} \rangle \text{ AND } \langle \text{condição} \rangle \text{ THEN } \langle \text{consequência} \rangle$. Assim, as regras definidas pelo especialista 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 9

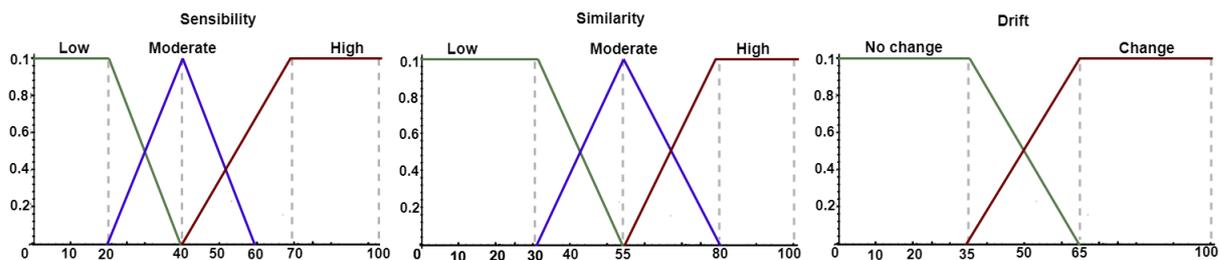


Figura 8 – Exemplo de conjuntos difusos para modelar mudanças comportamentais.

um exemplo de implementação das regras baseadas nos conjuntos fuzzy especificados na Figura 8.

```

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 9 – Regras fuzzy para mudança de comportamento.

4.4 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 através de funções disponíveis para realizar seleções, derivações, junções, agregações e filtragens (Chambers e Zaharia 2018). 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 deste 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 à zero hora de cada dia, quando serviços de detecção de padrões e mudanças comportamentais são acionados pelo job para processar a nova observação comportamental disponível (isto é, conjunto de eventos comportamentais do dia anterior).

A Figura 10 apresenta a arquitetura de software do framework proposto. O primeiro componente é o *Scheduler and Job Management (SJB)*, que é 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 CAs 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*, que é responsável pela execução efetiva dos jobs criados, vindo a orquestrar a execução dos serviços disponibilizados pela solução para detectar padrões e mudanças comportamentais. O terceiro componente é o *Pattern Inference and Change Detector*, que é composto por um conjunto de serviços responsáveis por executar tarefas específicas da solução, como detecção de padrões e mudanças de comportamentos, verificação de observações significativas e controle de observações. Neste estudo, definimos observações como o conjunto de eventos comportamentais registrados em um determinado dia. Por fim, o último componente é responsável por persistir os padrões e mudanças de comportamento, além de notificar as partes interessadas.

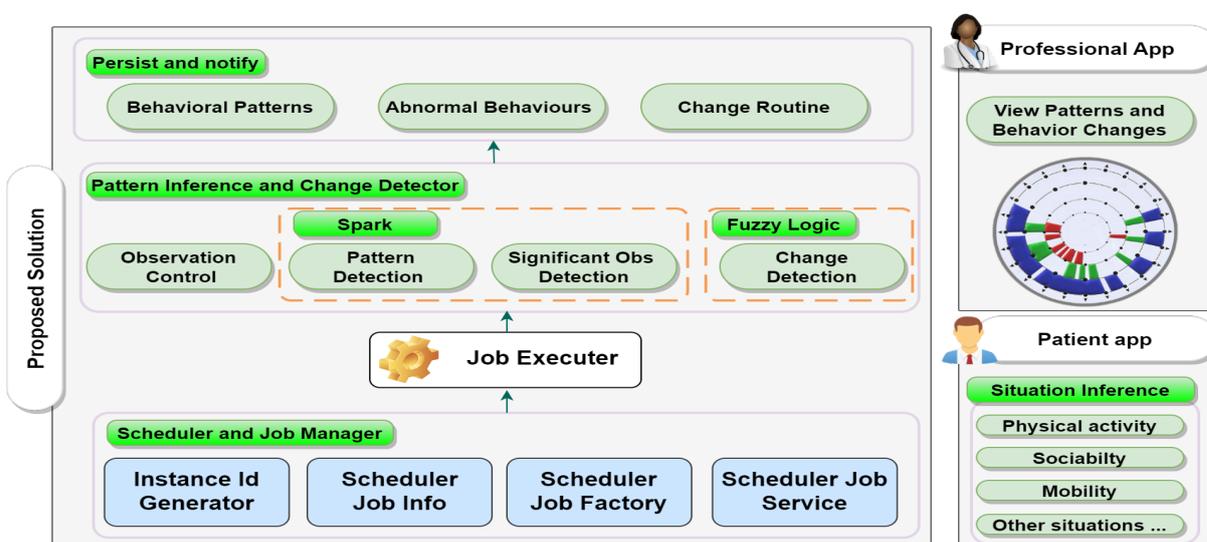


Figura 10 – Arquitetura de software do framework proposto.

4.4.1 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 11 apresenta o diagrama de atividades que descreve o fluxo de processamento da solução. Primeiramente, o SJB 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.

Primeiro, o *Job Executer* carrega os parâmetros definidos no SJB. 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 CAs especificados. Caso o contexto da observação não seja de interesse para o job, a execução é finalizada, caso contrário, o *Observation Significant Service* verifica se o conjunto de eventos da observação representa um comportamento significativo com base em regras especificadas pelos especialistas (por exemplo, média da duração dos eventos superior a um determinado limiar). Em seguida, o OCS registra a ocorrência da observação. Este 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 esta mudança for detectada, o CDS persiste e notifica esta informação. Por fim, o OCS limpa os registros de observações e finaliza a execução do job.

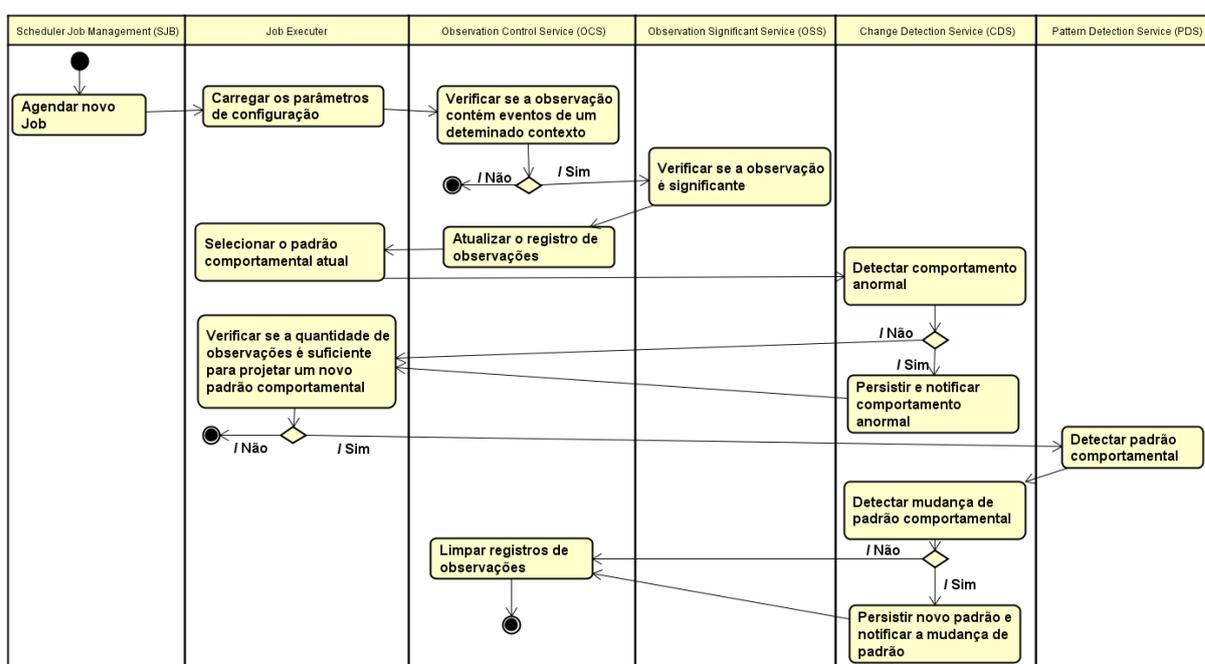


Figura 11 – Diagrama de atividades da solução proposta.

5 Avaliação Experimental Preliminar

Os experimentos realizados objetivaram avaliar a solução em termos de sua capacidade de projetar padrões multimodais enriquecidos com contexto capazes de modelar rotinas comportamentais, 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, atividade física e estadia em casa).

5.1 Descrição dos Dados

Usou-se o conjunto de dados público StudentLife (Wang et al. 2017) para criar o fluxo de dados necessário para avaliar a solução proposta. 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*), atividade física (*ActivityEvent*) e estadia em casa (*HomeStayEvent*) com base no modelo de dados apresentado na Seção 4.1.

Representou-se cada amostra no conjunto de dados como um evento comportamental que compõe um fluxo de dados. Especificamente, estruturou-se *ActivityEvent* a partir de inferências de atividades físicas (ou seja, caminhar e correr) e *SociabilityEvent* a partir de inferências de conversação. Também processou-se amostras de localização GPS para gerar o fluxo *HomeStayEvent*, representando eventos em que o indivíduo está em casa. Aplicou-se o algoritmo de agrupamento DBSCAN a amostras de GPS coletadas durante o período entre 23h e 5h para identificar a localização de residência dos indivíduos (ou seja, agrupamento com maior número de amostras). Portanto, *HomeStayEvent* representa as amostras que os indivíduos estão em casa (ou seja, localizados a no máximo 50 metros de sua casa). Por fim, usamos como CA o dia da semana reconhecido a partir dos registros de data/hora do evento. A Figura 12 mostra a estrutura do fluxo de eventos comportamentais gerado.

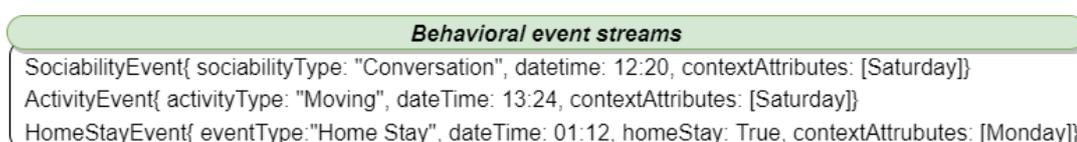


Figura 12 – Estrutura do fluxo de eventos comportamentais gerado.

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. Nesta etapa, projetou-se os padrões multimodais com uma, duas, três e quatro observações considerando como CA cada dia da semana (por exemplo, sexta-feira, sábado e domingo) para reconhecer a semelhança média entre esses padrões e observações subsequentes. A Figura 13 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.

A Figura 14 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 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 menos dias e com capacidade de modelagem semelhante ao usar mais 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 15 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

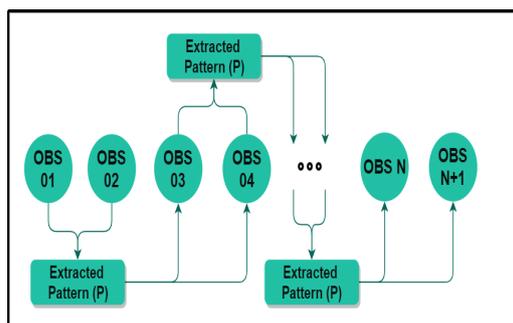


Figura 13 – Desenho experimental para avaliar o número de observação apropriado para projetar padrões.

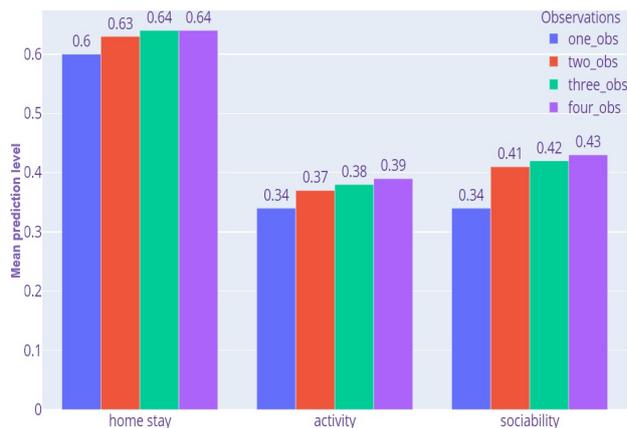


Figura 14 – Nível médio de previsão do comportamento.

pelos padrões.

A Figura 16 mostra a estabilidade média da rotina dos indivíduos para cada CA. Identificou-se alta estabilidade de rotina para *HomeStay*, que apresentou uma média de $\approx 60\%$. Os comportamentos *Activity* e *Sociability* mostraram uma estabilidade de rotina inferior, média de $\approx 0,33\%$ e $\approx 0,37\%$, respectivamente. Portanto, esperava-se que os padrões multimodais gerados pudessem modelar consistentemente *HomeStay* e ter mais dificuldade ao modelar *Activity* e *Sociability*.

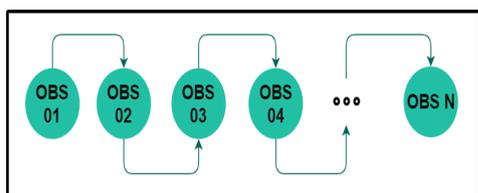


Figura 15 – Desenho experimental para avaliar a estabilidade da rotina comportamental.

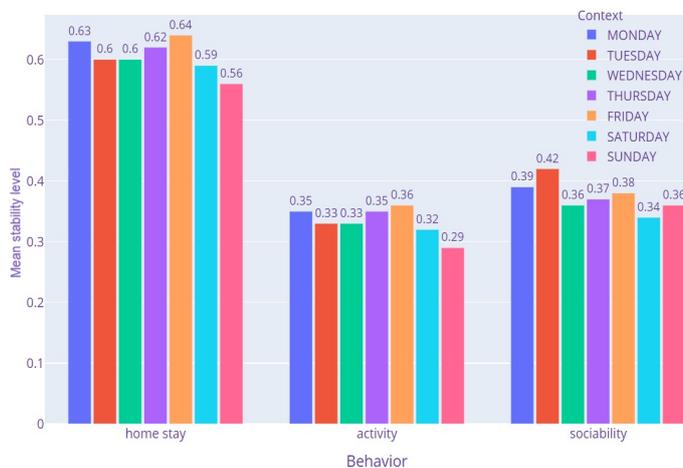


Figura 16 – Nível médio de estabilidade da rotina comportamental.

Em seguida, quantificou-se a associação entre o nível médio de similaridade (ou seja, a capacidade dos padrões de modelar rotinas) e a estabilidade média da rotina aplicando o coeficiente de correlação de Pearson, que permitiu identificar a capacidade da solução em modelar rotinas comportamentais. Os resultados deste experimento demonstraram uma forte associação positiva para todos os comportamentos analisados (*HomeStay* $\approx +0.91$, *Activity* $\approx +0.89$, *Sociability* $\approx +0,86$), indicando que a solução modela adequadamente as rotinas comportamentais dos indivíduos. A Figura 17 mostra

o nível de correlação resultante para cada CA. Identificou-se que os comportamentos analisados apresentaram forte correlação positiva em todos os contextos. Esses resultados reforçam a conclusão de que a solução proposta pode modelar adequadamente as rotinas comportamentais dos indivíduos.

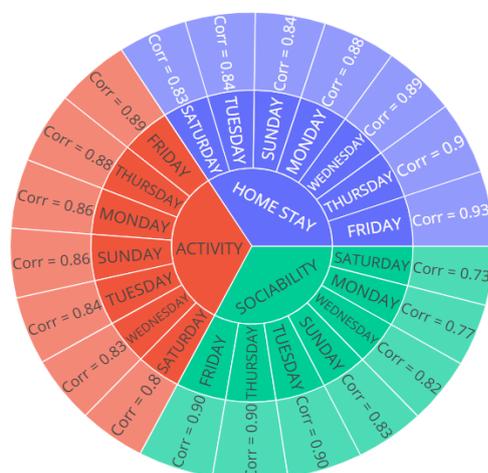


Figura 17 – Correlação (Corr) entre a capacidade média de modelagem e a estabilidade média de rotina para cada CA.

5.3 Sensibilidade nas Detecções de Mudanças de Padrões Comportamentais

Nesta seção, apresenta-se o experimento para avaliar a sensibilidade de detecção de mudanças de padrões comportamentais. Para isso, projetou-se um fluxo de dados com amostras de comportamentos anormais e mudanças de rotina. Assim, foi possível analisar o desempenho da solução proposta no processamento deste fluxo de dados. Em seguida, definiu-se os limiares para mudanças comportamentais. Calculou-se a média e o Desvio Padrão (DP) da estabilidade da rotina dos indivíduos, definindo os limiares de mudança em $[\mu + \sigma]$. Devido a limitação de espaço, neste projeto de pesquisa foca-se em uma situação específica para demonstrar a sensibilidade de detecção de mudança de padrões. Especificamente, processou-se dados *HomeStay*, no qual os usuários apresentam em média 59% de estabilidade em suas rotinas, apresentando também 10.5% de desvio padrão. Assim, especificou-se o limiar de mudança de comportamento *HomeStay* como 69.5%.

Nesta etapa, projetou-se o fluxo *HomeStay* combinando os dados do usuário u42 (u = usuário; número = identificação do participante no conjunto de dados StudentLife) e usuário u58, que têm rotinas estáveis e significativamente diferentes. A Figura 18 mostra o número de eventos em cada intervalo de tempo (ou seja, intervalos de 30 minutos) no fluxo de dados gerado. Como a rotina do u42 é estável, o comportamento esperado para a solução proposta ao processar o fluxo de dados deste usuário é

detectar apenas um comportamento anormal ao processar a Obs4 (Obs = observação; número = identificação da sequência), que é significativamente diferente das outras observações do u42. Também é esperado que a solução detecte comportamento anormal e mudança de padrão ao começar a processar dados do u58. Depois de reconhecer o novo padrão dos dados do u58, não é esperado que a solução identifique comportamentos anormais ou alterações de rotina, pois os dados do segundo usuário representa uma rotina estável.

A Figura 19 mostra as detecções de mudanças de comportamento realizadas pela solução ao processar o fluxo de dados gerado. Ao processar os dados do u42, a solução detectou apenas um comportamento anormal na Obs4. Então, ao iniciar o processamento dos dados do u58, a solução identificou comportamento anormal na Obs7 e na Obs8. Ao projetar o padrão de comportamento a partir dessas observações (ou seja, 4º padrão), a solução detectou uma mudança na rotina (ou seja, mudança de padrão). Este novo padrão apresentou grande similaridade com observações subsequentes, não implicando então em nenhuma nova detecção de comportamento anormal (ou mudança na rotina). Essas detecções coincidem com as expectativas ao se analisar o fluxo de dados gerado, mostrando que a solução proposta tem sensibilidade para identificar comportamentos anormais e mudanças de padrões.

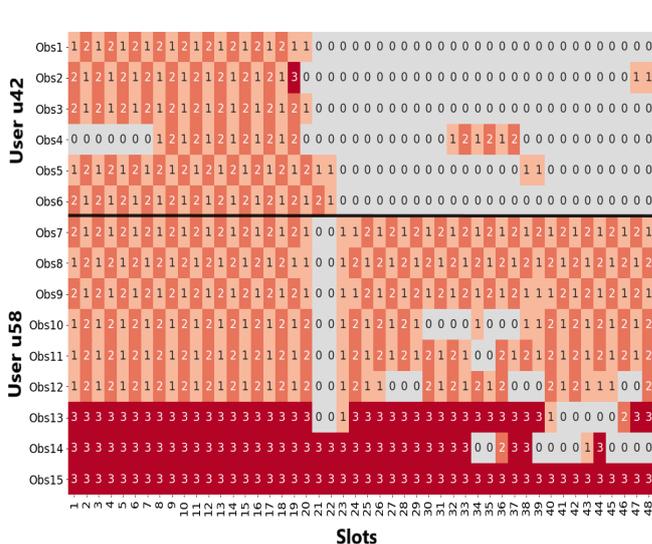


Figura 18 – Mesclagem dos dados *HomeStay* dos usuários u42 e u58.

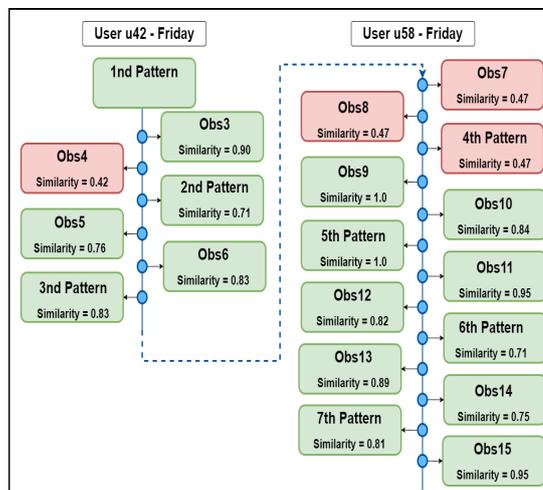


Figura 19 – Fluxo de detecção de mudança no comportamento *HomeStay*.

6 Discussão

A princípio, esta pesquisa concebeu uma primeira versão da solução proposta para detectar padrões e mudanças comportamentais processando eventos de alto nível gerados por dispositivos ubíquos. Esta solução foi desenvolvida com base em um modelo de dados projetado a partir dos conceitos de fenotipagem digital. Os experimentos preliminares conduzidos identificaram que a solução reconhece padrões multimodais capazes de modelar adequadamente a rotina dos indivíduos. Esta conclusão é baseada nos resultados dos níveis de correlação entre a capacidade do padrão de explicar rotinas e a estabilidade de rotina, que encontrou fortes associações positivas para todos os CAs analisados (*HomeStay* $\approx +0.91$, *Activity* $\approx +0.89$, *Sociability* $\approx +0,86$). Também avaliou-se a sensibilidade da detecção de mudança comportamental. Esta avaliação mostrou que a solução proposta pode reconhecer consistentemente comportamentos anormais e mudanças nas rotinas para *HomeStay*, *Activity* e *Sociability*. Em suma, com base nos resultados dos experimentos, reconhecemos que nossa proposta é uma solução em potencial para apoiar decisões de especialistas em saúde, pois pode fornecer informações objetivas sobre padrões multimodais que descrevem o hábito comportamental dos indivíduos.

Reconhecemos que nosso estudo tem limitações. A solução depende de parâmetros definidos empiricamente (ou seja, φ e θ) para detectar padrões de comportamento em vez de reconhecer os melhores valores com base no fluxo de dados processado. Outra limitação é que as mudanças de rotina não resultam necessariamente em mudanças comportamentais. Por exemplo, os indivíduos podem mudar suas atividades físicas da manhã para a tarde, mas manter a mesma intensidade. Apesar dessa limitação, a solução proposta pode informar ao profissional sobre as ocorrências de todas as mudanças de rotina, mesmo que o indivíduo mude apenas os períodos do dia de seu comportamento. Portanto, os profissionais podem investigar se essas mudanças comportamentais são decorrentes de mudanças do estado de saúde. Por fim, os participantes do estudo representam apenas estudantes universitários, portanto os resultados não podem ser generalizados para outras populações. Consequentemente, é necessário avaliar a solução em trabalhos futuros com populações mais heterogêneas.

Os planos futuros incluem várias frentes de trabalho. Em primeiro lugar, planeja-se estender a solução proposta para fornecer os seguintes recursos: (i) detectar padrões de comportamento multimodais pré-especificados por especialistas em saúde mental; (ii) identificar se uma mudança de padrão reconhecida reflete o mesmo hábito em horários diferentes (mudança de cronograma); (iii) configurar automaticamente os parâmetros do algoritmo da solução ou melhorar a semântica do valores definidos para

o usuário; (iv) integrar a solução desenvolvida com o aplicativo do paciente que irá gerar eventos comportamentais; e (v) desenvolver um mecanismo de segurança e privacidade para a solução proposta. Em segundo lugar, planeja-se desenvolver um aplicativo móvel para gerenciar a solução e notificar e visualizar as detecções de padrões e mudanças comportamentais. Este aplicativo também deve integrar as funcionalidades de uma ferramenta de avaliação momentânea ecológica. Também objetiva-se realizar uma avaliação aprofundada da utilidade da solução para profissionais especializados e seus pacientes. Especificamente, serão utilizadas estratégias metodológicas usualmente adotadas na avaliação da experiência do usuário.

6.1 Cronograma

Esta pesquisa de doutorado será conduzida de acordo com o cronograma apresentado abaixo.

1. Refatorar o código fonte do Framework.
2. Implementar uma abordagem para definição automática dos parâmetros da solução ou melhorar a semântica destes parâmetros.
3. Implementação da detecção de padrões pré-especificados por especialistas.
4. Estender o algoritmo para suportar padrões mais flexíveis, que considerem, além de slots de tempo de cada dia, comportamentos que se repetem em outra granularidade de tempo.
5. Refatorar o código fonte do framework.
6. Desenvolver um modelo de segurança e privacidade para a solução proposta.
7. Desenvolver um aplicativo móvel para gerenciar a solução e visualizar padrões e mudanças comportamentais.
8. Escrever artigo sobre o modelo de segurança e privacidade concebido.
9. Defender exame de proposta de tese.
10. Avaliar a experiência do usuário apresentada pela solução proposta.
11. Escrever artigo completo, contemplando o modelo de dados, o algoritmo projetado, o framework, o modelo de segurança, a visualização dos padrões e mudanças comportamentais, um estudo de caso para o modelo de segurança e privacidade, a avaliação da capacidade de projetar padrões e detectar mudanças comportamentais e a avaliação da experiência do usuário.

12. Escrever a tese de doutorado
13. Defender o doutorado
14. Realizar correções propostas pela banca

N°	Atividade	2022		2023	
		1° sem.	2 sem.	1 sem.	2 sem.
1	Refatoração de código				
2	Parametrização automática do algoritmo				
3	Padrões pré-especificados por especialistas				
4	Padrões com diferentes granularidade de tempo				
5	Refatoração de código				
6	Modelo de segurança e privacidade				
7	Visualização de padrões comportamentais				
8	Artigo sobre modelo de segurança e privacidade				
9	Exame de proposta de tese				
10	Avaliação da experiência do usuário				
11	Artigo completo				
12	Escrita da tese de doutorado				
13	Defesa do doutorado				
14	Realização das correções propostas				

7 Conclusão

Esta pesquisa propõe uma solução para o monitoramento ubíquo de padrões comportamentais multimodais enriquecidos com contexto. Esta 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 em FPM capaz de detectar padrões comportamentais e por um mecanismo que utiliza a lógica fuzzy para detectar comportamentos anômalos e mudanças de padrões. Também concebeu-se um framework que implementa o algoritmo proposto através de funções do spark e integra um módulo fuzzy para implementar o mecanismo de detecção de mudança comportamental. Os experimentos realizados objetivaram avaliar a solução em termos de sua capacidade de projetar padrões multimodais enriquecidos com contexto capazes de modelar rotinas comportamentais, bem como sua sensibilidade para reconhecer mudanças comportamentais. O primeiro experimento demonstrou uma alta correlação positiva entre os padrões gerados para modelar rotinas e a estabilidade rotineira dos indivíduos. O segundo experimento reconheceu que a solução proposta possui sensibilidade para identificar mudanças comportamentais. Assim, concluímos que a solução apresenta potencial para apoiar os especialistas em saúde, fornecendo informações objetivas (ou seja, padrões e mudanças comportamentais) sobre vários aspectos comportamentais, o que contribui para uma tomada de decisão mais eficaz.

7.1 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, se disponível.

- Moura, Ivan; Teles, Ariel; Silva, Francisco; Viana, Davi; Coutinho, Luciano; Barros, Flávio; Endler, Markus. **Mental Health Ubiquitous Monitoring Supported by Social Situation Awareness: A Systematic Review**. Journal of Biomedical Informatics, 2020 ([Moura et al. 2020](#));
Situação: Publicado. **Tipo de publicação:** Periódico. **Qualis:** A2. **Fator de impacto:** 6.317
- 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:** A1. **Fator de impacto:** 3.576

- **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:** 7.187

- 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** ACM Survey, 2022.

Situação: Submetido. **Tipo de publicação:** Periódico. **Qualis:** A1. **Fator de impacto:** 10.282

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 na página 20.

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 na página 21.

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 3 vezes nas páginas 14, 23 e 25.

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 39.

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 14.

BOUKHECHBA, M.; DAROS, A. R.; FUA, K.; CHOW, P. I.; TEACHMAN, B. A.; BARNES, L. E. DemonicSalmon: 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 na página 20.

BOUKHECHBA, M.; DAROS, A. R.; FUA, K.; CHOW, P. I.; TEACHMAN, B. A.; BARNES, L. E. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. **Smart Health**, v. 9-10, p. 192 – 203, 2018. CHASE 2018 Special Issue. Citado na página 25.

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. Citado 2 vezes nas páginas 21 e 25.

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: ACM, 2018. (CHASE '18), p. 120–125. Citado 2 vezes nas páginas 20 e 25.

CHAMBERS, B.; ZAHARIA, M. **Spark: The definitive guide: Big data processing made simple**. [S.l.]: "O'Reilly Media, Inc.", 2018. Citado na página 34.

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 55.

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 14.

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 na página 21.

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. Citado 3 vezes nas páginas 20, 21 e 24.

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**, JMIR Publications, v. 7, n. 7, p. e13209, 2019. Citado 3 vezes nas páginas 20, 24 e 25.

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 13.

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 na página 20.

FILZMOSER, P. **A multivariate outlier detection method**. [S.l.]: International Conference on Computer Data Analysis and Modeling, 2004. Citado na página 23.

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 19.

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 3 vezes nas páginas 14, 21 e 25.

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 2 vezes nas páginas 22 e 23.

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 na página 21.

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. Citado 2 vezes nas páginas 20 e 23.

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 14.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Department of Computer Science, Keele University, UK, 2004. Citado na página 54.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for Performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007. Citado na página 54.

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 14.

KRUMM, J. **Ubiquitous computing fundamentals**. [S.l.]: CRC Press, 2018. Citado na página 13.

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. **Mobile Networks and Applications**, Springer-Verlag, Berlin, Heidelberg, v. 19, n. 3, p. 345–359, 2014. Citado 3 vezes nas páginas 20, 21 e 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 3 vezes nas páginas 20, 21 e 22.

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 13, 14, 15, 19 e 23.

MCNEILL, F. M.; THRO, E. **Fuzzy logic: a practical approach**. [S.l.]: Academic Press, 2014. Citado na página 17.

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 13.

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 13.

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. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, ACM, New York, NY, USA, v. 3, n. 3, 2019. Citado na página 24.

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 na página 14.

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**, 2022. ISSN 0167-739X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167739X22000310>>. Citado na página 47.

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 2 vezes nas páginas 14 e 46.

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 na página 47.

NARZIEV, N.; GOH, H.; TOSHNAZAROV, K.; LEE, S. A.; CHUNG, K.-M.; NOH, Y. Stdd: Short-term depression detection with passive sensing. **Sensors**, v. 20, n. 5, 2020. Citado 2 vezes nas páginas 21 e 24.

PALUSKA, S. A.; SCHWENK, T. L. Physical activity and mental health. **Sports medicine**, Springer, v. 29, n. 3, p. 167–180, 2000. Citado na página 20.

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 13.

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, v. 19, n. 3, p. e75, 2017. Citado na página 21.

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 13.

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 na página 47.

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 46.

ROSS, T. J. **Fuzzy logic with engineering applications**. [S.l.]: John Wiley & Sons, 2005. Citado na página 17.

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. **Journal of Medical Internet Research**, JMIR Publications, v. 20, n. 6, p. e210, 2018. Citado 3 vezes nas páginas 20, 22 e 24.

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 Publications, v. 7, n. 1, p. e11041, 2019. Citado na página 21.

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. Citado 2 vezes nas páginas 22 e 25.

SIMSION, G.; WITT, G. **Data modeling essentials**. [S.l.]: Elsevier, 2004. Citado na página 27.

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 na página 22.

TA-SHMA, P.; AKBAR, A.; GERSON-GOLAN, G.; HADASH, G.; CARREZ, F.; MOESSNER, K. An ingestion and analytics architecture for iot applied to smart city use cases. **IEEE Internet of Things Journal**, PP, p. 1–1, 06 2017. Citado na página 15.

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 23.

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 14.

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 13 e 27.

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, v. 4, n. 3, p. e111, 2016. Citado 3 vezes nas páginas 13, 20 e 24.

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 3 vezes nas páginas 20, 25 e 38.

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, 2017. Citado 3 vezes nas páginas 20, 22 e 25.

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. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, ACM, New York, NY, USA, v. 2, n. 1, 2018. Citado 4 vezes nas páginas 14, 20, 22 e 25.

WEISER, M. Ubiquitous computing. **Computer**, IEEE, n. 10, p. 71–72, 1993. Citado na página 13.

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 14.

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 33.

A Apêndice A - Protocolo de Pesquisa da RSL Conduzida

A.1 Metodologia

Conduzimos esta RSL com base nas diretrizes de revisão sistemática da literatura de engenharia de software proposta por Kitchenham e Charters (Kitchenham 2004, Kitchenham e Charters 2007). Seguindo essas orientações, realizamos três fases distintas: (i) planejamos a realização da revisão, gerando um protocolo de pesquisa; (ii) realizamos então as atividades contidas no protocolo de pesquisa; e (iii) analisamos e sintetizamos os dados, a fim de extrair, resumir e relatar informações relevantes sobre esta área de pesquisa. Nesta seção, apresentamos os processos de planejamento e condução desta RSL.

A.1.1 Estratégia de Busca

Realizamos uma busca sistemática nas seguintes bibliotecas digitais: *ACM Digital Library*, *IEEE Xplore*, *Web of Science*, *PubMed*, *ScienceDirect* e *Scopus*. A busca teve como objetivo encontrar estudos que propusessem 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, projetamos uma *string* de busca com base em termos-chave, que possuem um conjunto de palavras-chave e sinônimos. A seguir, apresentamos a estrutura da *string* de busca projetada.

<<(("Mental Health" OR "Mental Disorder" OR "Mental State" OR "Psychological Health") AND ("Multimodal Sensing", "Smartphone sensing"OR "Mobile Sensing"OR "Passive Sensing"OR "Passive Detection"OR "Digital Phenotyp"OR "Behavioral Monitoring"OR "Objective Sensor Sata"))>>

A.1.2 Seleção de Estudos

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. Primeiramente, 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 por meio da aplicação do coeficiente *Kappa de Cohen* (Cohen 1968), que mediu o nível de concordância entre as análises dos pesquisadores. Ao final, os pesquisadores realizaram discussões para resolver os conflitos de seleção e, nos casos em que não houve consenso, os juízes (ou seja, dois outros autores) deliberaram sobre os conflitos.

Tabela 2 – Critérios de seleção.

Critérios de inclusão	Critérios de exclusão
(IC1) Estudos que usaram dados de contexto de sensores embutidos ou conectados em dispositivos ubíquos.	(EC1) Estudos que não usaram 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) Soluções propostas que utilizaram sensoriamento multimodal para apoiar os 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, pôsteres, relatórios resumidos, livros, comunicações breves, teses e dissertações.
(IC5) Artigos em língua inglesa.	(EC5) Artigos em um idioma diferente do inglês.

A.1.3 Extração de Dados

Nessa etapa, os pesquisadores realizaram a leitura do texto integral dos artigos selecionados para extrair informações relevantes que permitiram sintetizar e caracterizar os estudos, responder às questões de pesquisa, identificar questões em aberto e traçar perspectivas futuras de pesquisa. Para isso, elaboramos um formulário de extração de dados com os seguintes itens: (i) referência (autor e ano); (ii) objetivo e principais resultados; (iii) desordem/estado mental; (iv) dados de contexto coletado; (v) situações identificadas; (vi) método, técnica ou processo para identificar situações; (vii) uso do sensoriamento multimodal no monitoramento da saúde mental; (viii) duração do estudo e número de participantes; e (ix) limitações.