

# Integração de Questionários de Saúde e Dados Coletados de Dispositivos Vestíveis em Modelos de Aprendizado de Máquina: Uma Revisão Rápida da Literatura

Nadiana K. N. Mendes<sup>1</sup>, Rossana M. C. Andrade<sup>1</sup>, Pedro A. M. Oliveira<sup>2</sup>

<sup>1</sup> Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)  
Universidade Federal do Ceará (UFC)  
Fortaleza – CE – Brasil

<sup>2</sup> Laboratório de Inovação e Computação Científica (LJCC)  
Instituto Federal do Maranhão (IFMA)  
Pedreiras – MA – Brasil

nadianakelly@alu.ufc.br, rossana@ufc.br, pedro.oliveira@ifma.edu.br

**Resumo.** *O avanço dos dispositivos vestíveis e do aprendizado de máquina tem ampliado a integração entre dados fisiológicos passivos e questionários de saúde. Esta revisão rápida analisou 13 estudos publicados entre 2020 e 2025 sobre o uso combinado dessas abordagens em aplicações de saúde digital. Os resultados identificaram três estratégias principais: uso de questionários como variável-alvo, integração multimodal entre autorrelatos e dados fisiológicos e predição contínua de escores clínicos. Observou-se potencial promissor, especialmente em modelos multimodais e temporais. Contudo, persistem desafios como heterogeneidade metodológica, amostras reduzidas e validação externa limitada. Conclui-se que essa convergência representa uma tendência consistente, mas ainda demanda maior padronização e validação para adoção em larga escala.*

**Abstract.** *The advancement of wearable devices and machine learning has expanded the integration of passive physiological data and health questionnaires. This rapid review analyzed 13 studies published between 2020 and 2025 on the combined use of these approaches in digital health applications. The findings identified three main strategies: the use of questionnaires as target variables, multimodal integration of self-reports and physiological data, and continuous prediction of clinical scores. Promising potential was observed, particularly in multimodal and temporal models. However, challenges remain, including methodological heterogeneity, small sample sizes, and limited external validation. It is concluded that this convergence represents a consistent trend, but still requires greater standardization and validation for large-scale adoption.*

## 1. Introdução

O avanço das tecnologias digitais em saúde tem ampliado o uso de dispositivos vestíveis, smartphones e sensores IoT para monitoramento contínuo de variáveis fisiológicas e comportamentais, como frequência cardíaca, sono e atividade física. Esses dados vêm sendo

integrados a questionários clínicos tradicionais, como PHQ-8, PHQ-9 e escalas de ansiedade, que apesar de amplamente validados, apresentam limitações relacionadas à subjetividade e à incapacidade de captar variações ao longo do tempo.

Nesse contexto, técnicas de aprendizado de máquina têm sido utilizadas para complementar avaliações clínicas por meio da análise de dados coletados passivamente. Estudos recentes demonstram o potencial da fenotipagem digital para estimar sintomas depressivos, ansiedade, estresse e bem-estar a partir de dados de wearables e smartphones [Price et al. 2023, Price et al. 2024, Rykov et al. 2021, Sun et al. 2023]. Trabalhos também evidenciam a viabilidade do monitoramento longitudinal em populações específicas, como adultos em contextos de restrição de movimento [Moshe et al. 2021], e a predição de crises em transtornos de ansiedade com horizonte temporal de dias [Tsai et al. 2022], além da aplicação de modelos multitarefa e ensemble para predição simultânea de múltiplos desfechos em saúde mental [Jacobson and Feng 2022, Saylam and Durmaz İncel 2024]. Em larga escala, [Zhang et al. 2025] utilizaram dados do Fitbit combinados com PHQ-8 e GAD-7 em mais de 10.000 participantes para predição de severidade de depressão e ansiedade com modelos XGBoost.

Além das abordagens preditivas, estudos têm proposto arquiteturas IoT-cloud integrando questionários, wearables e algoritmos de ML para monitoramento contínuo de saúde mental e bem-estar [Machado-Jaimes et al. 2022]. Essas estratégias também vêm sendo aplicadas em condições físicas crônicas, como artrite reumatoide, utilizando dados passivos de atividade física para classificação de escores autorrelatados [Rao et al. 2023]. Apesar do crescimento da área, persistem desafios relacionados à heterogeneidade metodológica, interoperabilidade, privacidade e validação clínica em larga escala [Sun et al. 2023]. Assim, este trabalho realiza uma revisão rápida da literatura sobre a integração entre questionários de saúde, dispositivos vestíveis e aprendizado de máquina em aplicações de saúde e bem-estar entre 2020 e 2025.

## **2. Metodologia**

O estudo foi conduzido como uma Revisão Rápida da Literatura, seguindo a definição metodológica proposta por [Smela et al. 2023]. A escolha desse método foi baseada em três fatores principais. Primeiro, o objetivo do estudo é exploratório e de síntese. Busca-se mapear abordagens emergentes e identificar tendências, lacunas e padrões metodológicos na literatura recente. Nesse contexto, as revisões rápidas são consideradas apropriadas [Smela et al. 2023].

Segundo, o tema possui natureza interdisciplinar, envolvendo computação aplicada, ciência de dados e saúde digital. Essa característica dificulta a realização de uma revisão sistemática tradicional, pois os estudos disponíveis apresentam grande diversidade metodológica incluindo diferentes dispositivos, populações, métricas e pipelines, que inviabiliza uma meta-análise quantitativa. Terceiro, a rápida evolução dos dispositivos vestíveis e das técnicas de aprendizado de máquina exige sínteses mais ágeis e atualizáveis, o que é uma característica essencial das revisões rápidas.

De acordo com [Smela et al. 2023], esse tipo de revisão mantém princípios fundamentais como transparência, sistematização e reprodutibilidade, ao mesmo tempo em que adota simplificações metodológicas controladas, como limitação de bases de dados, recorte temporal e triagem com um número reduzido de revisores, para produzir evidências

em tempo oportuno. O objetivo desta revisão foi responder às seguintes questões de pesquisa:

- **Q1:** O que já existe na literatura sobre o uso de questionários de saúde combinados com técnicas de machine learning e dados de dispositivos vestíveis?
- **Q2:** Como dados de questionários autorrelatados têm sido combinados com dados provenientes de dispositivos vestíveis em modelos preditivos?

## 2.1. Bases de dados e recorte temporal

Conforme apresentado na Figura 1, a busca bibliográfica foi realizada nas seguintes bases de dados: Scopus, PubMed e IEEE Xplore. A escolha dessas bases considerou o caráter interdisciplinar do tema, envolvendo tanto a área da saúde quanto a área de computação aplicada. Foi adotado recorte temporal entre 2020 e 2025, considerando a evolução recente dos dispositivos vestíveis e das técnicas de aprendizado de máquina aplicadas à saúde digital. Não houve restrição de idioma.

## 2.2. Estratégia de busca

A estratégia de busca foi estruturada a partir de três eixos conceituais principais: questionários de saúde, dispositivos vestíveis e técnicas de inteligência artificial ou aprendizado de máquina. As buscas utilizaram operadores booleanos e combinações de termos livres. No caso da base PubMed, também foram empregados descritores indexados quando aplicável. Foram utilizadas as seguintes strings de busca:

1. ("health questionnaire"OR "patient survey"OR "health form"OR "self-reported questionnaire") AND ("wearable data"OR "fitness tracking"OR "smart device") AND ("machine learning"OR "artificial intelligence"OR "predictive model")
2. ("health questionnaire"OR "patient survey"OR "self-reported questionnaire") AND ("wearable data"OR "smart device"OR "fitness tracking"OR "Wearable Electronic Devices"[MeSH]) AND ("machine learning"OR "artificial intelligence"OR "predictive model"OR "Machine Learning"[MeSH])
3. ("Internet of Things") AND ("Health Questionnaire"OR "Health Assessment"OR "Health Survey") AND ("Machine Learning"OR "Artificial Intelligence"OR "Deep Learning") AND ("Wearable device"OR "Wearable technology"OR "Wearable sensor")
4. ("Internet of Things") AND ("Health Questionnaire") AND ("Machine Learning"OR "Deep Learning") AND ("Wearable device")
5. ("Artificial Intelligence") AND ("Machine Learning") AND ("Wearable Devices") AND ("Health Questionnaire"OR "Patient-Reported Outcomes")

Também foram exploradas variações envolvendo o termo *Internet of Things*, considerando sua recorrência em arquiteturas de monitoramento remoto em saúde.

Na base PubMed, foram incluídos descritores como "Wearable Electronic Devices" [MeSH] e "Machine Learning" [MeSH], quando disponíveis. As buscas foram aplicadas nos campos de título, resumo e palavras-chave.

### **2.3. Critérios de inclusão e exclusão**

- **Critérios de inclusão**

- Utilizaram questionários de saúde como variável de entrada ou desfecho;
- Integraram dados provenientes de dispositivos vestíveis;
- Aplicaram técnicas de machine learning ou deep learning;
- Apresentaram resultados quantitativos ou descrição metodológica clara;
- Foram publicados entre 2020 e 2025.

- **Critérios de exclusão**

- Eram puramente conceituais ou sem descrição metodológica detalhada e resultados;
- Utilizavam apenas questionários, sem integração com dados objetivos de dispositivos vestíveis;
- Empregavam exclusivamente métodos estatísticos tradicionais, sem técnicas de aprendizado de máquina;
- Eram revisões narrativas ou estudos sem modelo preditivo próprio;
- Estavam duplicados entre as bases consultadas.

### **2.4. Processo de seleção**

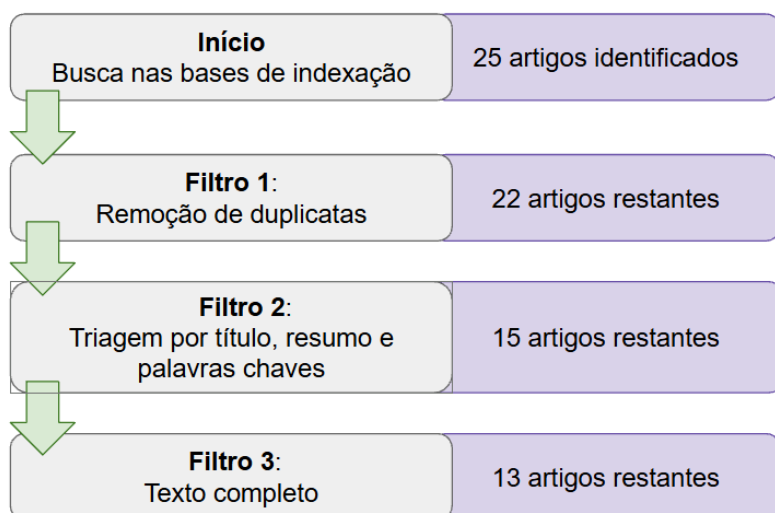
A aplicação das estratégias de busca resultou na seleção final de 13 estudos publicados entre 2020 e 2025, os quais abordam a integração entre questionários de saúde, dispositivos vestíveis e técnicas de aprendizado de máquina [Price et al. 2024, Price et al. 2023, Rykov et al. 2021, Jacobson and Feng 2022, Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024, Tsai et al. 2022, Mullick et al. 2022, Rao et al. 2023, Machado-Jaimes et al. 2022, Zhang et al. 2025, Moshe et al. 2021, Sun et al. 2023]

Os resultados das buscas foram exportados para planilha eletrônica para remoção de duplicatas. Em seguida, foi realizada triagem em três etapas: leitura de títulos, leitura de resumos e leitura completa dos artigos potencialmente elegíveis. A seleção foi conduzida por um único revisor, caracterizando uma das simplificações metodológicas típicas de revisões rápidas, conforme descrito por [Smela et al. 2023]. O fluxo de seleção seguiu estrutura semelhante ao modelo PRISMA adaptado para revisões rápidas.

## **3. Resultados**

### **3.1. Caracterização geral dos estudos**

Observa-se predominância de aplicações em saúde mental, abrangendo detecção e monitoramento de sintomas depressivos, ansiosos e de estresse. [Rykov et al. 2021], [Price et al. 2023] e [Price et al. 2024] exploram dados passivamente coletados por wearables, combinados com PHQ-9, para estimar presença ou variabilidade de sintomas depressivos em adultos, enquanto [Sun et al. 2023] adotam abordagem similar utilizando o PHQ-8, com foco nos desafios metodológicos de dados mHealth. Em populações específicas, [Moshe et al. 2021] investigaram, em estudo longitudinal de 30 dias com adultos, a capacidade preditiva de dados passivos de dispositivos vestíveis (Oura Ring) e de



**Figura 1. Processo de busca e seleção de artigos.**

smartphones para sintomas de depressão, ansiedade e estresse mensurados pelo DASS-21, enquanto [Mullick et al. 2022] investigaram a capacidade preditiva de dados passivos de dispositivos móveis e vestíveis para estimação de escores de depressão em adolescentes.

A cobertura clínica estende-se à ansiedade e ao estresse: [Jacobson and Feng 2022] predisseram severidade de transtorno de ansiedade generalizada a partir de dados passivos de movimento de uma coorte nacional, com AUC de 0,892; [Saylam and Durmaz İncel 2024] investigaram a predição simultânea de depressão, ansiedade e estresse por aprendizado multitarefa; [Saylam and Durmaz İncel 2023] quantificaram biomarcadores digitais associados a afeto positivo, negativo, estresse e ansiedade a partir de dados temporais de wearables; e [Tsai et al. 2022] desenvolveram modelo preditivo de crises de pânico com horizonte de sete dias, combinando dados fisiológicos de smartwatch com múltiplos questionários validados (BDI, BAI, STAI e PDSS-SR), alcançando acurácia entre 67,4% e 81,3% conforme o algoritmo empregado. Outros trabalhos ampliam as aplicações para condições físicas crônicas [Rao et al. 2023], bem-estar geral e saúde mental em larga escala [Zhang et al. 2025], além de aplicações de bem-estar geral com infraestrutura IoT-cloud [Machado-Jaimes et al. 2022].

### 3.2. Arquiteturas e infraestrutura tecnológica

Parte dos estudos propõe arquiteturas baseadas em IoT e computação integrada: [Machado-Jaimes et al. 2022] desenvolveram sistema IoT-cloud integrando questionários psicológicos, dados de smartwatch e algoritmos de classificação em uma única plataforma. Os demais estudos adotam arquiteturas mais simples, com dispositivos comerciais como Fitbit, Apple Watch, Garmin, Oura Ring e acelerômetros de pulso, e processamento em ambiente computacional convencional. [Zhang et al. 2025] e [Rao et al. 2023] empregaram o Fitbit como dispositivo de coleta passiva, integrando-o a questionários autorrelatados validados por meio de aplicativos móveis, sem dependência de infraestrutura IoT dedicada. [Moshe et al. 2021] utilizaram o Oura Ring em conjunto com aplicativo para iPhone para coleta de dados de sono, atividade física e HRV em ambiente naturalístico. [Tsai et al. 2022] empregaram um rastreador fitness integrado a aplicativo móvel para coleta contínua de dados fisiológicos ao longo de um ano.

### 3.3. Modelos de aprendizado de máquina

Os estudos empregam algoritmos tradicionais, como Random Forest, SVM e KNN [Machado-Jaimes et al. 2022, Mullick et al. 2022, Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024], e arquiteturas profundas, incluindo CNN, LSTM e modelos híbridos [Price et al. 2024]. Abordagens de aprendizado multitarefa e ensemble são recorrentes [Saylam and Durmaz İncel 2024, Saylam and Durmaz İncel 2023].

[Jacobson and Feng 2022] empregaram XGBoost em estrutura de ensemble composta por dois modelos de ordem superior e 100 modelos de ordem inferior treinados com features distintas; [Mullick et al. 2022] compararam estimação de valores contínuos e classificação de mudanças clínicas de depressão em adolescentes, empregando múltiplos algoritmos, incluindo Random Forest, Gradient Boosting e XGBoost; [Zhang et al. 2025] aplicaram modelos XGBoost para predição de severidade de depressão e ansiedade a partir de dados do Fitbit combinados com questionários PHQ-8 e GAD-7, em uma das maiores amostras populacionais da literatura, com mais de 10.000 participantes; e [Rao et al. 2023] compararam Random Forest e Hidden Markov Model para classificação semanal de escores de instrumentos autorrelatados (PROMIS) a partir de dados passivos de atividade física coletados por Fitbit em pacientes com artrite reumatoide, com AUC máximo de 0,750. [Moshe et al. 2021] aplicaram modelos multinível (MLM) e K-means para predição de sintomas de saúde mental a partir de dados combinados de wearable e smartphone, com o modelo combinado apresentando melhor desempenho preditivo para depressão. [Tsai et al. 2022] compararam seis algoritmos de ML — Random Forest, Decision Tree, Análise Discriminante Linear, AdaBoost, XGBoost e Regularized Greedy Forest para predição de crises de pânico, com Random Forest obtendo melhor desempenho (acurácia de 81,3%). Os melhores desempenhos associam-se a combinações multimodais e abordagens temporais, sendo que [Saylam and Durmaz İncel 2024] destacam que o aprendizado multitarefa melhora resultados para depressão e estresse especialmente quando a temporalidade dos dados é considerada, resultado convergente com a vantagem do Hidden Markov Model sobre modelos estáticos observada por [Rao et al. 2023].

### 3.4. Integração entre questionários e dados de wearables

Três estratégias principais de integração foram identificadas:

- Utilização de questionários como variável alvo para treinamento supervisionado [Price et al. 2024, Rykov et al. 2021, Jacobson and Feng 2022, Mullick et al. 2022, Zhang et al. 2025];
- Combinação de dados fisiológicos e autorrelatos como entrada multimodal [Sun et al. 2023, Moshe et al. 2021, Saylam and Durmaz İncel 2023, Tsai et al. 2022];
- Predição de escores de instrumentos autorrelatados a partir de dados passivos de wearables como estratégia de monitoramento contínuo e redução de carga ao paciente [Rao et al. 2023].

Os dados fisiológicos analisados incluem frequência cardíaca, variabilidade da frequência cardíaca, padrões de sono, atividade física, condutância da pele, temperatura periférica e acelerometria [Sun et al. 2023, Moshe et al. 2021, Jacobson and Feng 2022, Saylam and Durmaz İncel 2023, Mullick et al. 2022, Zhang et al. 2025, Rao et al. 2023, Tsai et al. 2022].

### 3.5. Síntese geral

De forma agregada, os estudos indicam avanço consistente na convergência entre dados subjetivos e objetivos para suporte à decisão em saúde. Entretanto, observa-se predominância de estudos exploratórios ou de prova de conceito, com heterogeneidade quanto a amostras, protocolos e validação externa [Sun et al. 2023, Saylam and Durmaz İncel 2024, Mullick et al. 2022]. Uma exceção relevante é o trabalho de [Zhang et al. 2025], que se destaca pelo volume amostral expressivo (n = 10.129), representando avanço em relação à tendência de amostras reduzidas dominante na área. Nessa mesma direção, [Tsai et al. 2022] adotaram delineamento prospectivo de coorte com seguimento de um ano, o que representa contribuição metodológica relevante quanto à validade longitudinal dos modelos preditivos.

**Tabela 1. Comparação entre os estudos incluídos na revisão rápida (2020–2025)**

Estudo	Domínio	Questionário	Dados do Wearable	Modelo ML	Principais Contribuições
Rykov et al. (2021)	Depressão	PHQ-9	Atividade, sono, FC	XGBoost (DART)	Biomarcadores digitais para triagem
Price et al. (2023)	Depressão	PHQ-9	Movimento, sono	Ensemble ML	Variabilidade sintomática longitudinal
Price et al. (2024)	Depressão	PHQ-9 (triagem MDD)	Movimento passivo (actigrafia)	ML supervisionado + CNN	Detecção populacional de MDD
Sun et al. (2023)	Depressão	PHQ-8	Sono, mobilidade, uso do telefone	Estatístico + ML	Confiabilidade de dados mHealth
Moshe et al. (2021)	Depressão, ansiedade e estresse	DASS-21	Atividade, sono, HRV (Oura Ring)	K-means + MLM	Fenotipagem digital com wearable comercial em contexto de restrição de movimento
Mullick et al. (2022)	Depressão	PHQ-9	Atividade, sono, uso do telefone	Modelos baseados em árvore (RF, GB, XGBoost) + regressão linear	Predição em adolescentes
Jacobson & Feng (2022)	Ansiedade	GAD-7	Movimento (acelerômetro)	XGBoost (Ensemble)	Predição de TAG (AUC 0,892)
Saylam & İncel (2023)	Bem-estar	PANAS, Omnibus Anxiety, MITRE Omnibus Stress	FC, EDA, acelerômetro, sono	RF, XGBoost, LSTM	Biomarcadores temporais de afeto e estresse
Saylam & İncel (2024)	Saúde mental	DASS	FC, sono, atividade	RF, XGBoost, LSTM	Predição simultânea de depressão, ansiedade e estresse
Tsai et al. (2022)	Transtorno do pânico	PDSS-SR, BDI, BAI, STAI	Sono, FC, atividade (Garmin Vivosmart 4)	RF, DT, LDA, AdaBoost, XGBoost, RGF	Predição de crises de pânico com 7 dias de antecedência; acurácia 67,4%–81,3%
Machado-Jaimes et al. (2022)	Bem-estar	DASS-21, Q-LES-Q e Ryff Psychological Well-Being Scale	Sono, atividade	RF, SVM, KNN, DT, NB e NN	Arquitetura IoT-Cloud integrada
Rao et al. (2023)	Condição crônica (AR)	PROMIS + Godin Leisure-Time Physical Activity Questionnaire	Atividade física (Fitbit)	RF, Hidden Markov Model	Classificação temporal de escores PRO; AUC 0,750
Zhang et al. (2025)	Depressão e ansiedade	PHQ-8, GAD-7	Sono, atividade, FC (Fitbit)	XGBoost; K-means; PCA	Fenotipagem digital em larga escala (n = 10.129); população geral do Reino Unido

A Tabela 1 apresenta uma síntese comparativa dos 13 estudos incluídos na revisão, organizados segundo domínio clínico, estratégia de integração entre questionários e dados de dispositivos vestíveis, tipos de dados utilizados, modelos de aprendizado de máquina empregados e principais contribuições. Observa-se predominância de aplicações voltadas à saúde mental, com cobertura que se estende da depressão à ansiedade, estresse, transtorno do pânico e bem-estar geral, abrangendo diferentes faixas etárias, de adolescentes a adultos. O uso recorrente de instrumentos como PHQ-8, PHQ-9, GAD-7, DASS-21 e PROMIS como variável alvo ou componente multimodal evidencia a centralidade dos questionários padronizados na validação dos modelos. Os estudos também evidenciam di-

versidade metodológica, variando entre modelos tradicionais, como Random Forest, SVM e KNN, e arquiteturas profundas baseadas em CNN, LSTM e abordagens multitarefa e ensemble. Apesar dos resultados promissores reportados, a maioria dos trabalhos apresenta limitações relacionadas a amostras reduzidas, validação externa restrita e heterogeneidade nos pipelines de processamento, aspectos que impactam a generalização e a reprodutibilidade dos achados. Destaca-se, contudo, a contribuição de [Zhang et al. 2025], cujo volume amostral ( $n = 10.129$ ) representa exceção relevante a esse padrão.

Cabe destacar que a soma total de ocorrências de algoritmos supera o número de estudos incluídos, uma vez que a maioria dos trabalhos empregou mais de um algoritmo simultaneamente, seja para fins de comparação de desempenho, seja por composição de arquiteturas ensemble ou multitarefa.

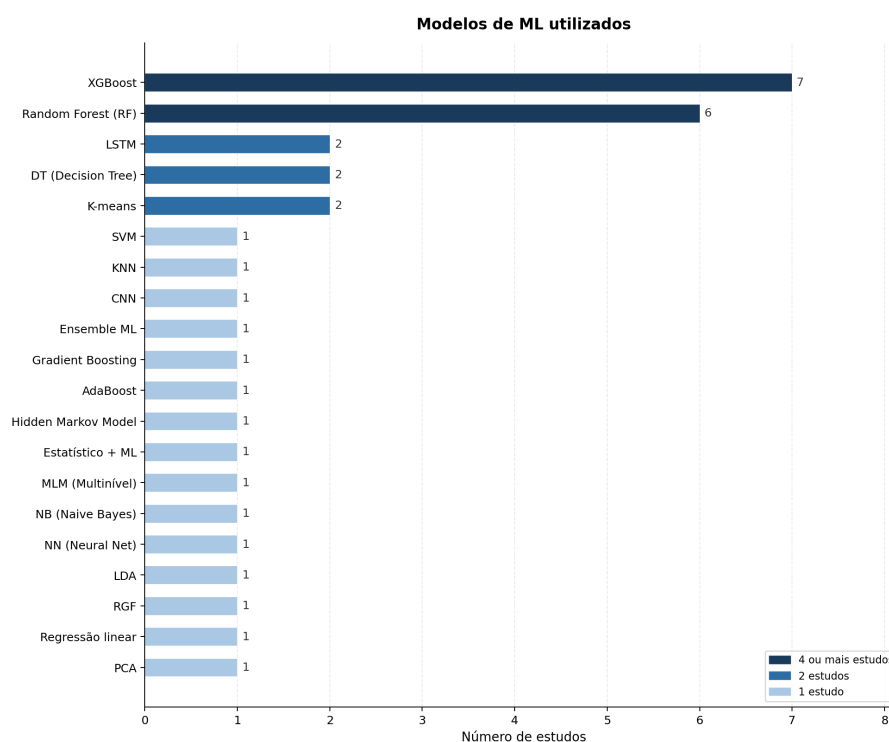
A Figura 2 ilustra a frequência de uso dos algoritmos de aprendizado de máquina identificados nos estudos analisados. O XGBoost foi o algoritmo mais recorrente, presente em sete trabalhos [Rykov et al. 2021, Mullick et al. 2022, Jacobson and Feng 2022, Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024, Tsai et al. 2022, Zhang et al. 2025], seguido pelo Random Forest, empregado em seis estudos [Mullick et al. 2022, Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024, Tsai et al. 2022, Machado-Jaimes et al. 2022, Rao et al. 2023]. LSTM foi identificado em dois estudos [Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024]. Decision Tree aparece em dois estudos [Tsai et al. 2022, Machado-Jaimes et al. 2022]. K-means foi empregado em dois estudos [Moshe et al. 2021, Zhang et al. 2025]. SVM, KNN, CNN, Gradient Boosting, Hidden Markov Model, AdaBoost, LDA e Regularized Greedy Forest foram cada um identificado em um único estudo [Machado-Jaimes et al. 2022, Machado-Jaimes et al. 2022, Price et al. 2024, Mullick et al. 2022, Rao et al. 2023, Tsai et al. 2022, Tsai et al. 2022, Tsai et al. 2022]. Abordagens de ensemble, modelagem multinível e métodos estatísticos combinados com ML foram identificadas em estudos isolados [Price et al. 2023, Moshe et al. 2021, Sun et al. 2023].

#### **4. Discussão**

Os resultados desta revisão rápida indicam que a integração entre questionários de saúde, dispositivos vestíveis e técnicas de aprendizado de máquina tem se consolidado como uma linha de investigação relevante no período entre 2020 e 2025. Observa-se, entretanto, que essa consolidação ocorre de forma heterogênea, tanto do ponto de vista metodológico quanto em relação aos contextos clínicos investigados.

Uma das principais tendências identificadas é a utilização de questionários padronizados como variável de referência para treinamento supervisionado. Instrumentos como PHQ-8, PHQ-9 e GAD-7 são frequentemente empregados como marcador clínico para validação dos modelos [Rykov et al. 2021, Price et al. 2023, Price et al. 2024, Jacobson and Feng 2022, Mullick et al. 2022, Zhang et al. 2025].

Nessa configuração, os dados provenientes de dispositivos vestíveis atuam como preditores de escores autorrelatados, estabelecendo uma ponte entre biomarcadores digitais objetivos e medidas subjetivas de saúde. Embora essa estratégia seja metodologicamente coerente, ela também evidencia dependência significativa de instrumentos autorrelatados como padrão de validação. O trabalho de [Rao et al. 2023] expande essa



**Figura 2. Frequência de uso dos modelos de aprendizado de máquina.**

lógica ao empregar escores PROMIS instrumentos padronizados amplamente utilizados em condições crônicas como variável-alvo para classificação por meio de dados passivos do Fitbit, sugerindo que a abordagem é generalizável para além do domínio da saúde mental.

A predominância de aplicações em saúde mental, especialmente no monitoramento de sintomas depressivos e ansiosos, reforça a afinidade entre fenotipagem digital e modelos preditivos baseados em comportamento. Estudos longitudinais analisados demonstram associação entre padrões de sono, variabilidade da frequência cardíaca e alterações nos escores de depressão e ansiedade [Price et al. 2023, Jacobson and Feng 2022, Moshe et al. 2021], sugerindo potencial para monitoramento contínuo e intervenções precoces. A extensão dessas abordagens para populações específicas, como adolescentes [Mullick et al. 2022] e pacientes com transtorno do pânico [Tsai et al. 2022], indica movimento em direção à personalização clínica, embora a heterogeneidade comportamental observada entre participantes sugira que modelos individualizados podem ser mais adequados do que abordagens generalistas.

A cobertura de condições além da depressão também merece destaque. [Jacobson and Feng 2022] demonstraram que dados passivos de acelerômetro são suficientes para discriminar indivíduos com sintomas elevados de transtorno de ansiedade generalizada com alto poder preditivo, enquanto [Saylam and Durmaz İncel 2024] evidenciaram que a predição simultânea de depressão, ansiedade e estresse por aprendizado multitarefa é viável e metodologicamente vantajosa. [Tsai et al. 2022] ampliaram essa cobertura ao demonstrar a viabilidade de predição de crises de pânico com sete dias de antecedência, combinando dados fisiológicos contínuos de smartwatch com múltiplos ques-

tionários clínicos validados. Esses achados sugerem que a integração multicondicional em um único pipeline preditivo pode ser uma direção promissora para futuras investigações. A extensão para condições físicas crônicas, como demonstrado por [Rao et al. 2023] no contexto da artrite reumatoide, indica que o paradigma questionário + wearable + ML possui potencial de generalização para além da saúde mental, abrindo perspectivas para monitoramento remoto em doenças com alta variabilidade sintomática ao longo do tempo.

No que se refere às estratégias de modelagem, observa-se coexistência entre métodos tradicionais de aprendizado supervisionado e arquiteturas baseadas em redes neurais profundas e modelos sequenciais. Modelos como Random Forest, SVM e KNN são frequentemente utilizados em cenários com amostras menores [Machado-Jaimes et al. 2022, Rykov et al. 2021, Tsai et al. 2022], enquanto arquiteturas CNN, LSTM e modelos híbridos são aplicados quando há maior volume e complexidade de dados [Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024, Price et al. 2024]. A dimensão temporal dos dados de wearables emerge como fator relevante: [Saylam and Durmaz İncel 2023, Saylam and Durmaz İncel 2024] demonstraram que abordagens que consideram a dependência temporal das séries fisiológicas produzem resultados superiores em comparação a modelos que tratam os dados de forma estática. Resultado convergente foi observado por [Rao et al. 2023], no qual o Hidden Markov Model, ao incorporar correlações entre semanas consecutivas de dados, superou significativamente o Random Forest estático em todas as tarefas de classificação de escores PRO.

Apesar dos resultados experimentais promissores, a maioria dos estudos permanece no estágio exploratório ou de prova de conceito. A acurácia elevada reportada em alguns trabalhos deve ser interpretada com cautela, considerando limitações como tamanhos amostrais reduzidos, ausência de validação externa e contextos de coleta controlados [Mullick et al. 2022, Tsai et al. 2022, Sun et al. 2023]. Uma exceção relevante é o trabalho de [Zhang et al. 2025], que se destaca pelo volume amostral expressivo ( $n = 10.129$  participantes de uma população geral do Reino Unido), representando avanço significativo em relação à tendência de amostras pequenas e contextos controlados dominante na área. Ainda assim, poucos estudos descrevem de forma detalhada os pipelines completos de pré-processamento, tratamento de dados faltantes e estratégias de reprodutibilidade, o que dificulta comparações diretas entre abordagens.

## 5. Limitação

Esta revisão apresenta limitações inerentes ao seu desenho metodológico. Por tratar-se de uma revisão rápida da literatura [Smela et al. 2023], foram adotadas simplificações como a restrição a três bases de dados, delimitação temporal entre 2020 e 2025 e triagem conduzida por um único revisor, o que pode ter resultado na exclusão de estudos relevantes. O número relativamente reduzido de trabalhos incluídos também limita a amplitude da síntese.

A heterogeneidade metodológica entre os estudos envolvendo diferentes dispositivos, instrumentos de avaliação, populações e métricas de desempenho dificultou comparações diretas e inviabilizou a condução de meta-análise quantitativa. Soma-se a isso a variabilidade na descrição dos pipelines de processamento, com muitos artigos omitindo etapas como pré-processamento, tratamento de dados faltantes e validação externa,

comprometendo a reprodutibilidade e a profundidade comparativa.

A predominância de estudos exploratórios com amostras reduzidas e contextos específicos restringe a generalização dos resultados, sendo poucos os trabalhos com validação clínica ampliada ou avaliação longitudinal. Além disso, alguns estudos foram conduzidos em ambientes laboratoriais controlados, limitando sua extrapolação para cenários clínicos reais. Por fim, a rápida evolução tecnológica da área pode fazer com que avanços recentes ainda não estejam contemplados nesta síntese, reforçando a necessidade de atualizações periódicas.

## 6. Conclusão

Esta revisão rápida da literatura analisou 13 estudos publicados entre 2020 e 2025 que investigam a integração entre questionários de saúde, dispositivos vestíveis e técnicas de aprendizado de máquina. Os resultados evidenciam avanço consistente na combinação de dados autorrelatados e biomarcadores digitais, com predominância de modelos supervisionados baseados em instrumentos clínicos padronizados, como PHQ-8, PHQ-9, entre outros instrumentos padronizados. Observa-se também o crescimento no uso de modelos aplicados a dados multimodais e séries temporais, incluindo abordagens híbridas e estratégias de aprendizado multitarefa. Os estudos abrangem diferentes contextos clínicos, como depressão, ansiedade, estresse, transtorno do pânico e condições físicas crônicas, reforçando o potencial de generalização dessas abordagens.

De forma geral, identificam-se três principais estratégias de integração: o uso de questionários como variável-alvo em modelos supervisionados; a combinação multimodal de dados fisiológicos e autorrelatos; e a predição contínua de escores clínicos a partir de dados passivos provenientes de dispositivos vestíveis. Apesar dos avanços, a maioria das investigações ainda se encontra em estágio exploratório, com amostras reduzidas e limitada validação externa. Embora estudos com maior escala indiquem potencial ampliado de generalização, permanecem desafios relacionados à padronização metodológica, reprodutibilidade e aplicabilidade clínica.

Em síntese, a literatura aponta um potencial significativo para o uso integrado de questionários digitais e dados de dispositivos vestíveis como suporte à avaliação clínica contínua e personalizada. No entanto, são necessários estudos longitudinais, com amostras mais amplas e maior rigor metodológico, para consolidar a maturidade científica e viabilizar a aplicação prática dessas soluções em contextos reais.

## 7. Agradecimentos

Os autores agradecem à Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo apoio financeiro a este trabalho.

## Referências

- Jacobson, N. C. and Feng, B. (2022). Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. *Translational Psychiatry*, 12:336.
- Machado-Jaimes, L., Bustamante-Bello, M. R., and Argüelles-Cruz, A. (2022). Development of an intelligent system for the monitoring and diagnosis of the well-being. *Sensors*, 22(24):9719.

- Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., and Pulkki-Råback, L. (2021). Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry*, 12:625247. PMID: 33584388.
- Mullick, T., Radovic, A., Shaaban, S., and Doryab, A. (2022). Predicting depression in adolescents using mobile and wearable sensors: Multimodal machine learning-based exploratory study. *JMIR Formative Research*, 6(6):e35807.
- Price, G. D., Heinz, M. V., Collins, A. C., and Jacobson, N. C. (2024). Detecting major depressive disorder presence using passively-collected wearable movement data in a nationally-representative sample. *Psychiatry Research*, 332:115693.
- Price, G. D., Heinz, M. V., Song, S. H., Nemesure, M. D., and Jacobson, N. C. (2023). Using digital phenotyping to capture depression symptom variability: detecting naturalistic variability in depression symptoms across one year using passively collected wearable movement and sleep data. *Translational Psychiatry*, 13(1):381.
- Rao, K., Speier, W., Meng, Y., Wang, J., Ramesh, N., Xie, F., Su, Y., Nowell, W. B., Curtis, J. R., and Arnold, C. (2023). Machine learning approaches to classify self-reported rheumatoid arthritis health scores using activity tracker data: Longitudinal observational study. *JMIR Formative Research*, 7:e43107. PMID: 37017471.
- Rykov, Y., Thach, T.-Q., Bojic, I., Christopoulos, G., and Car, J. (2021). Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, 9(10):e24872.
- Saylam, B. and Durmaz İncel, Ö. (2023). Quantifying digital biomarkers for well-being: Stress, anxiety, positive and negative affect via wearable devices and their time-based predictions. *Sensors*, 23(21):8987.
- Saylam, B. and Durmaz İncel, Ö. (2024). Multitask learning for mental health: Depression, anxiety, stress (DAS) using wearables. *Diagnostics*, 14(5):501.
- Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E., and Boyer, L. (2023). Rapid literature review: definition and methodology. *Journal of Market Access & Health Policy*, 11(1):2241234.
- Sun, S., Folarin, A. A., Zhang, Y., Cummins, N., Garcia-Dias, R., Stewart, C., Ranjan, Y., Rashid, Z., Conde, P., Laiou, P., et al. (2023). Challenges in using mhealth data from smartphones and wearable devices to predict depression symptom severity: retrospective analysis. *Journal of medical Internet research*, 25:e45233.
- Tsai, C.-H., Chen, P.-C., Liu, D.-S., Kuo, Y.-Y., Hsieh, T.-T., Chiang, D.-L., Lai, F., and Wu, C.-T. (2022). Panic attack prediction using wearable devices and machine learning: Development and cohort study. *JMIR Medical Informatics*, 10(2):e33063. PMID: 35166679.
- Zhang, Y., Stewart, C., Ranjan, Y., Conde, P., Sankesara, H., Rashid, Z., Sun, S., Dobson, R. J., and Folarin, A. A. (2025). Large-scale digital phenotyping: Identifying depression and anxiety indicators in a general uk population with over 10,000 participants. *Journal of Affective Disorders*, 375:412–422.