

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**RECONHECIMENTO DE GESTOS DE
INGESTÃO DE LÍQUIDOS UTILIZANDO
REDES CONVOLUCIONAIS PROFUNDAS**

Manaus - AM
outubro de 2025

RECONHECIMENTO DE GESTOS DE INGESTÃO DE LÍQUIDOS UTILIZANDO REDES CONVOLUCIONAIS PROFUNDAS

Pedro Daniel da Silva Gohl

Orientador: *Prof. Dr. Eduardo James Pereira Souto*

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas - ICOMP-UFAM, como parte dos requisitos para obtenção do título de Mestre em Informática.

Manaus - AM
outubro /2025

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

G614r Gohl, Pedro Daniel da Silva
Reconhecimento de Gestos de Ingestão de Líquidos utilizando Redes
Convoluconais Profundas / Pedro Daniel da Silva Gohl. - 2025.
49 f. : il., color. ; 31 cm.

Orientador(a): Eduardo James Pereira Souto.
Dissertação (mestrado) - Universidade Federal do Amazonas, Programa
de Pós-Graduação em Informática, Manaus, 2025.

1. Reconhecimento de gestos. 2. Sensores inerciais. 3. Redes neurais
convolucionais. 4. Ingestão de líquidos . 5. Base de dados. I. Souto,
Eduardo James Pereira. II. Universidade Federal do Amazonas. Programa
de Pós-Graduação em Informática. III. Título



Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"RECONHECIMENTO DE GESTOS DE INGESTÃO DE LÍQUIDOS UTILIZANDO REDES CONVOLUCIONAIS PROFUNDAS"

PEDRO DANIEL DA SILVA GOHL

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Dr. Eduardo James Pereira Souto - **Presidente**

Profa. Dra. Eulanda Miranda dos Santos - **Membro Interno**

Dr. Hendrio Luis de Souza Bragança - **Membro Externo**

Manaus, 06 de outubro de 2025.



Documento assinado eletronicamente por **Eduardo James Pereira Souto, Professor do Magistério Superior**, em 06/10/2025, às 15:45, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Eulanda Miranda dos Santos, Professor do Magistério Superior**, em 06/10/2025, às 15:53, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2829071** e o código CRC **9DFFA67D**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio
Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

Referência: Processo nº 23105.043895/2025-61

SEI nº 2829071

Agradecimentos

Agradeço, em primeiro lugar, aos meus pais, por me concederem o privilégio da dedicação aos estudos e por serem a base de todas as minhas conquistas. A minha noiva, Andressa Diógenes, que me apoiou e acreditou em mim, sempre me incentivando a alcançar novos horizontes. Ao meu orientador, Professor Eduardo Souto, minha profunda gratidão pela orientação, pelo incentivo constante e pela amizade ao longo desta jornada.

Aos meus queridos amigos, Paulo Nellesen e Ayala Bernado, que não apenas abriram as portas de sua casa, mas também me acolheram com afeto nos momentos mais complicados. Serei eternamente grato.

Aos amigos que a academia me deu, Giovanni Zanardo e Leonardo Yuto, obrigado pelas inúmeras discussões e por tornarem o caminho mais leve.

As minhas colegas de laboratório Amanda Spellen e Laura Queiroz, que me ajudaram na jornada de tornar esse trabalho possível.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Código de Financiamento 001) e Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPEAM, que financiou parcialmente – por meio do projeto POSGRAD.

E por fim, mas não menos importante, agradeço a Rodrigo Tavares e Edwino Stein, pela amizade e por todo o suporte oferecido durante meu tempo em Manaus.

Resumo

A ingestão adequada de líquidos é essencial para a saúde, mas o acompanhamento contínuo ainda depende majoritariamente de autorrelato, sujeito a erros e pouco escalável. Reconhecer automaticamente o gesto de ingestão é o primeiro passo para monitorar a hidratação de forma passiva em ambiente real. Contudo, esse problema enfrenta dois gargalos: primeiro, a escassez de bases públicas de tamanho e diversidade suficientes para treinar modelos; segundo, a necessidade de modelos leves que generalizem entre indivíduos e possam rodar em dispositivos vestíveis.

Para minimizar esses desafios, este trabalho apresenta um conjunto de dados de sinais inerciais de punho (*InGesture*), destinado ao reconhecimento do gesto de ingestão de líquidos. O *InGesture* reúne registros de 50 participantes executando ingestão de líquidos e gestos análogos, totalizando 663 minutos de dados de acelerômetro e giroscópio rotulados em oito classes. Essa base supre uma lacuna crítica da literatura, a escassez de dados em escala e com diversidade suficiente para treinar e comparar modelos de forma justa e reprodutível, favorecendo a generalização entre indivíduos.

Com esse alicerce, a dissertação apresenta uma CNN 1D leve, projetada para capturar padrões temporais e preparada para execução em dispositivos vestíveis. A validação enfatiza a capacidade de generalização entre sujeitos e cobre tanto o cenário binário (ingestão vs. não ingestão) quanto o multiclasse, com comparação direta a *baselines* clássicos e arquiteturas profundas de referência.

Nos experimentos, o SVM obteve F1-score de 78,12% no cenário binário, enquanto a CNN 1D alcançou 87,30% (ingestão vs. não ingestão) e 87,50% no multiclasse (oito classes). Embora os dados tenham sido coletados em ambiente controlado e o foco atual seja a detecção de eventos (não a estimativa do volume ingerido), os resultados indicam a viabilidade do monitoramento contínuo e privado a partir de relógios inteligentes, estabelecendo base sólida para estudos em ambiente livre e para o acoplamento futuro entre detecção de gestos e estimativa de volume.

Palavras-chave: Reconhecimento de gestos, sensores inerciais, redes neurais convolucionais, ingestão de líquidos, base de dados

Abstract

Adequate fluid intake is essential for health, yet continuous monitoring still relies primarily on self-reporting, which is prone to errors and poorly scalable. Automatically recognizing drinking gestures is the first step toward passively monitoring hydration in real-world settings. However, this problem faces two main bottlenecks: first, the scarcity of public datasets with sufficient size and diversity to train models; second, the need for lightweight models that generalize across individuals and can run on wearable devices.

To address these challenges, this work presents a wrist-worn inertial signal dataset (*InGesture*) for liquid intake gesture recognition. *InGesture* comprises recordings from 50 participants performing drinking gestures and analogous actions, totaling 663 minutes of labeled accelerometer and gyroscope data across eight classes. This dataset fills a critical gap in the literature, the lack of large-scale, diverse data needed to train and fairly, reproducibly compare models, thereby promoting cross-subject generalization.

Building on this foundation, the dissertation proposes a lightweight 1D CNN designed to capture temporal patterns and optimized for execution on wearable devices. Validation emphasizes cross-subject generalization and covers both binary (drinking vs. non-drinking) and multiclass scenarios, with direct comparisons against classical *baselines* and state-of-the-art deep architectures.

In experiments, SVM achieved an F1-score of 78.12% in the binary scenario, while the 1D CNN reached 90.79% (drinking vs. non-drinking) and 87.50% in the multiclass setting (eight classes). Although data were collected in a controlled environment and the current focus is on event detection (not volume estimation), results indicate the feasibility of continuous, privacy-preserving monitoring using smartwatches, establishing a solid basis for future studies in free-living conditions and for coupling gesture detection with volume estimation.

Keywords: Gesture recognition, inertial sensors, convolutional neural networks, fluid intake, dataset

Lista de Figuras

Figura 2.1: Metodologia comum; aquisição; pré-processamento, extração de características e classificação. Adaptado (Hamatani et al., 2018) (Gomes e Sousa, 2019) (Huang et al., 2020)	6
Figura 2.2: Dados do Acelerômetro para o gesto de ingestão.	7
Figura 2.3: Dados do Giroscópio para o gesto de ingestão.	7
Figura 2.4: Dados do Acelerômetro para o gesto de ingestão com filtro de média-móvel aplicado.	8
Figura 2.5: Dados do Giroscópio para o gesto de ingestão com filtro de média-móvel aplicado.	8
Figura 4.1: Dispositivo de Monitoramento de Pulso Witmotion.	18
Figura 4.2: Exemplo dos sinais inerciais de gestos coletados durante a execução das coletas.	19
Figura 4.3: Participantes durante a coleta de dados.	19
Figura 4.4: Gestos análogos a ingestão.	19
Figura 5.1: Fluxo geral do processo de classificação utilizado nesta pesquisa. . . .	24
Figura 5.2: Esquema de validação LOSO.	25
Figura 5.3: Janelas deslizantes sobrepostas ($S < L$) e rotulação por <i>voto majoritário</i> . A barra inferior de cada janela resume a proporção temporal das classes no seu interior; o cartucho à direita explicita o rótulo atribuído (J1, J2, J3 e J4 no exemplo).	27
Figura 5.4: Arquitetura convolucional 1D utilizada para classificação de gestos. .	27

Lista de Tabelas

Tabela 4.1: Amostra de dados brutos anotados com sensores inerciais.	20
Tabela 4.2: Resumo dos gestos na base de dados, incluindo duração total, número de ocorrências e informações relevantes sobre o gesto de ingestão. . .	20
Tabela 4.3: Comparação entre <i>Base de dados</i> de ingestão de líquidos e gestos análogos.	21
Tabela 6.1: Hiperparâmetros avaliados nos experimentos.	34
Tabela 6.2: Resultados métodos clássicos no cenário binário	35
Tabela 6.3: Matriz de Confusão Média - SVM Binária (Huang et al. (2020))	35
Tabela 6.4: Matriz de Confusão Média — SVM Multiclasse (Huang et al. (2020)) . .	36
Tabela 6.5: Resultados modelos profundos no cenário binário	36
Tabela 6.6: Matriz de confusão Média método proposto	37
Tabela 6.7: Matriz de confusão Média Moccia et al. (2022) CNN	37
Tabela 6.8: Matriz de confusão Média Moccia et al. (2022) CNN_LSTM	37
Tabela 6.9: Matriz de Confusão Média — Senyurek et al. (2019) CNN_LSTM	38
Tabela 6.10Matriz de confusão Média Wang et al. (2024) TCN_MHA	38
Tabela 6.11Desempenho dos modelos profundos no cenário multiclasse	38
Tabela 6.12Matriz de Confusão Média — Método Proposto Multiclasse	39
Tabela 6.13Matriz de Confusão Média — Moccia et al. (2022) CNN_LSTM Multiclasse	39
Tabela 6.14Matriz de Confusão Média —Senyurek et al. (2019) CNN_LSTM Multi-classe	39
Tabela 6.15Matriz de Confusão Média — Wang et al. (2024) TCN_MHA Multiclasse	40

Sumário

Agradecimentos	4
Resumo	4
Abstract	6
Lista de Figuras	6
Lista de Tabelas	8
1 INTRODUÇÃO	1
1.1 Contextualização e Problema	1
1.2 Objetivo Geral	3
1.3 Objetivos Específicos	3
1.4 Contribuições	3
1.5 Organização do Texto	4
2 MONITORAMENTO DE GESTOS USANDO SENSORES VESTÍVEIS	5
2.1 Aquisição de Dados	7
2.2 Pré-processamento e Segmentação de Sinais	8
2.3 Extração de Características dos Sinais	9
2.4 Classificação	11
2.4.1 Modelos Rasos	11
2.4.2 Modelos Profundos	11
2.5 Considerações Finais	12
3 TRABALHOS RELACIONADOS	13
3.1 Abordagens para Detecção de Gestos de Ingestão	13
3.1.1 Sensores e Protocolos de Coleta	13
3.1.2 Pré-processamento e Segmentação de Sinais	14
3.1.3 Classificação: De Modelos Rasos a Profundos	15
3.2 Análise Comparativa e aspectos da Pesquisa Atual	16
3.3 Considerações Finais	16

4 InGesture: Um conjunto de dados de sensores inerciais para reconhecimento de gestos de ingestão de líquidos	17
4.1 Protocolo de Coleta	17
4.1.1 Participantes	17
4.1.2 Dispositivo de Coleta	17
4.1.3 Procedimento	18
4.2 Comparação com Outros Conjuntos de Dados	21
4.3 Aspectos Éticos e Legais	22
4.4 Considerações Finais	22
5 CLASSIFICAÇÃO DE GESTOS DE INGESTÃO DE LÍQUIDO	23
5.1 Fluxo geral do método proposto	23
5.2 Divisão dos Dados (Treino/Teste)	24
5.3 Pré-processamento dos Dados	25
5.4 Segmentação por Janela	25
5.5 Arquitetura de Classificação	27
5.6 Treinamento (LOSO) e Validação	28
5.7 Avaliação do Modelo e Métricas de Desempenho	28
5.8 Considerações Finais	29
6 EXPERIMENTOS E RESULTADOS	30
6.1 Protocolo Experimental	30
6.1.1 Base de Dados	30
6.1.2 Método de Validação	31
6.2 Pré-processamento	31
6.2.1 Dados Brutos	31
6.2.2 Re-amostragem	31
6.2.3 Normalização	32
6.2.4 Segmentação (Janelamento)	32
6.3 Modelagem e Classificação	33
6.3.1 Modelos de Classificação Clássicos (Baselines Rasos)	33
6.3.2 Configuração dos Experimentos	34
6.4 Resultados Aprendizado de máquina clássico	35
6.5 Resultados com Modelos de Aprendizado Profundo	36
6.5.1 Cenário Binário	36
6.5.2 Cenário Multiclasse	37
6.6 Discussão dos Resultados	41
6.7 Considerações Finais	41
7 CONCLUSÃO	43
7.1 Considerações Finais	43
7.2 Principais Contribuições	43

7.3 Dificuldades e Limitações Encontradas	44
7.4 Direções Futuras	44

INTRODUÇÃO

1.1 Contextualização e Problema

A água constitui o principal componente do corpo humano, correspondendo a aproximadamente entre 50% a 70% da massa corporal em adultos, e é indispensável para a manutenção de funções fisiológicas como transporte de nutrientes, termorregulação, homeostase celular e lubrificação articular ([EFSA Panel on Dietetic Products, Nutrition and Allergies \(NDA\), 2017](#); [Popkin et al., 2010](#); [Rosinger et al., 2023](#)). A insuficiência de ingestão hídrica pode levar à desidratação, condição caracterizada por balanço hídrico negativo, na qual a perda de líquidos excede a reposição. Evidências apontam que reduções moderadas (por exemplo, $\geq 2\%$ da massa corporal) já são suficientes para afetar desempenho cognitivo e psicomotor, incluindo atenção e memória ([Zhang et al., 2019](#)). Em quadros mais severos, observam-se impactos sistêmicos relevantes, com prejuízos ao funcionamento gastrointestinal, renal, cardiovascular e hepático ([Nishi et al., 2023](#)). Em populações vulneráveis, como idosos, a desidratação associa-se a desfechos adversos e declínio cognitivo em longo prazo ([Edmonds et al., 2021](#)). Apesar disso, a ingestão adequada de líquidos permanece frequentemente negligenciada em diferentes faixas etárias e contextos sociais ([Edmonds et al., 2021](#)).

O monitoramento da ingestão de líquidos é tradicionalmente conduzido por métodos de autorrelato, como questionários e recordatórios dietéticos. Embora amplamente utilizados em estudos populacionais, tais métodos apresentam limitações conhecidas: vieses de sub/superestimação, erros de memória, dificuldade de quantificação de porções e omissão de ingestões de baixa saliência ao longo do dia ([Rogerson et al., 2023](#)). Além disso, a variabilidade intraindivíduo no padrão diário de consumo exige coletas repetidas para estimativas mais confiáveis ([Singh et al., 2025](#)), o que torna essas abordagens pouco adequadas para monitoramento fino e contínuo em nível individual ([Rogerson et al., 2023](#)).

Nesse contexto, a detecção automatizada de episódios de ingestão de líquidos a partir de sinais sensoriais surge como uma alternativa promissora. Em particular, o reconhecimento confiável do gesto de levar o recipiente à boca constitui um componente fundamental para sistemas de monitoramento automatizado, pois estabelece a base para inferir a ocorrência de eventos de consumo (Huang et al., 2020). Diferentes estratégias têm sido exploradas na literatura. Soluções instrumentadas no ambiente, como câmeras e recipientes inteligentes, podem produzir medições objetivas (por exemplo, por sensores de nível/peso) (Roy et al., 2022). Entretanto, tais soluções podem impor restrições práticas (dependência de recipientes específicos, limitações de usabilidade e variabilidade de precisão entre dispositivos comerciais) (Cohen et al., 2022). Em contraste, dispositivos vestíveis equipados com sensores inerciais (acelerômetros e giroscópios), especialmente posicionados no pulso, oferecem vantagens operacionais importantes: independência do tipo de recipiente, integração natural à rotina e preservação de privacidade por não dependerem de áudio/imagem. Resultados recentes indicam que modelos baseados em sinais inerciais podem identificar gestos de ingestão com desempenho elevado, inclusive em cenários mais próximos do mundo real (Wang et al., 2022).

Apesar desse avanço, persistem desafios que dificultam a translação dessas abordagens para aplicações cotidianas. O primeiro desafio é a escassez de bases públicas especificamente direcionadas à ingestão de líquidos. Modelos supervisionados para reconhecimento de gestos dependem de conjuntos de dados rotulados e representativos para (i) aprender padrões robustos frente à variabilidade intersujeitos e (ii) permitir avaliação reproduzível sob protocolos padronizados. Quando os dados permanecem privados, torna-se difícil reproduzir resultados, comparar abordagens de forma justa e estimar o real poder de generalização em condições diferentes das observadas durante a coleta (Huang et al., 2020; Wang et al., 2022). Esse cenário contrasta com áreas consolidadas de Reconhecimento de Atividades Humanas (HAR), onde bases abertas como PAMAP2 (Bleser et al., 2015) e *Opportunity* (Chavarriaga et al., 2013) foram determinantes para acelerar o desenvolvimento, a validação independente e a comparação entre métodos.

Um segundo aspecto diz respeito à implantação em dispositivos vestíveis. Embora arquiteturas profundas tenham alcançado alto desempenho na detecção de ingestão (Wang et al., 2022), o custo de inferência (tempo, memória e energia) pode ser relevante para uso embarcado. Nesse sentido, não se assume que *CNNs* sejam intrinsecamente “leves”; a complexidade varia conforme profundidade, número de filtros e resolução temporal dos sinais. Assim, além de buscar alto desempenho e generalização, é pertinente considerar arquiteturas mais enxutas e o compromisso entre acurácia e custo computacional.

Dessa forma, o problema de pesquisa deste trabalho pode ser sintetizado como: Como reconhecer automaticamente episódios de ingestão de líquidos a partir de sinais inerciais de pulso com alta acurácia e capacidade de generalização entre indivíduos e, adicionalmente, com um perfil de recursos que não inviabilize a implantação em dispositivos vestíveis?

Para endereçar essas lacunas, este trabalho desenvolve a base de dados *InGesture*, composta por coletas de 50 participantes, totalizando 1.325 ocorrências de ingestão e aproximadamente 39 mil segundos de sinais inerciais rotulados. Adicionalmente, propõe-se um modelo baseado em *rede convolucional unidimensional (1D-CNN)* para reconhecimento de eventos de ingestão, com ênfase em simplicidade arquitetural visando reduzir risco de sobreajuste e favorecer implantação em contexto embarcado. A avaliação é conduzida sob validação cruzada *Leave-One-Subject-Out (LOSO)*, quantificando a capacidade do modelo de generalizar para indivíduos não vistos durante o treinamento, cenário alinhado ao uso em aplicações reais.

1.2 Objetivo Geral

Desenvolver e avaliar um modelo de reconhecimento de episódios de ingestão de líquidos a partir de sinais inerciais de pulso, baseado em uma arquitetura 1D-CNN com custo computacional reduzido, visando generalização entre indivíduos e potencial viabilidade para execução embarcada em dispositivos vestíveis.

1.3 Objetivos Específicos

- Projetar e disponibilizar uma base de dados pública para reconhecimento de gestos de ingestão de líquidos, contemplando múltiplos participantes e classes de movimentos similares, de modo a apoiar reprodutibilidade e comparabilidade entre métodos.
- Propor uma arquitetura 1D-CNN orientada à eficiência, adequada a séries temporais multicanais de sensores inerciais.
- Analisar o compromisso entre desempenho e complexidade do modelo (p.ex., número de parâmetros e custo aproximado de inferência), discutindo implicações para implantação em dispositivos vestíveis.

1.4 Contribuições

Este trabalho apresenta duas contribuições principais:

1. Base de dados *InGesture*: um conjunto de dados dedicado ao reconhecimento de episódios de ingestão de líquidos a partir de sinais inerciais, com 50 participantes, 1.325 ocorrências de ingestão e aproximadamente 39 mil segundos de sinais rotulados, visando suprir a carência de dados públicos na área e apoiar avaliação reprodutível.

2. *Modelo 1D-CNN para ingestão*: proposta de uma arquitetura *Conv1D* para reconhecer eventos de ingestão e distingui-los de movimentos similares. A abordagem alcançou F1-score de 90,79% para a classe de ingestão na classificação binária e F1-score de 87,50% no cenário multiclasse, sob protocolo LOSO, evidenciando robustez e generalização.

1.5 Organização do Texto

O restante deste trabalho está organizado da seguinte maneira:

No Capítulo 2 são apresentados os conceitos fundamentais relacionados ao reconhecimento de gestos e séries temporais de sensores inerciais, bem como métricas e protocolos de avaliação empregados neste estudo.

No Capítulo 3 são discutidos os trabalhos relacionados ao reconhecimento de ingestão de líquidos com dispositivos vestíveis, destacando limitações associadas à generalização e à ausência de bases públicas.

No Capítulo 4 são descritos o protocolo de coleta, a rotulagem e o pré-processamento dos sinais, incluindo segmentação e normalização, culminando na construção da base InGesture.

No Capítulo 5 é apresentada a arquitetura proposta, o protocolo experimental (incluindo validação LOSO) e a análise dos resultados, com comparações e discussão de desempenho e complexidade.

No Capítulo 6 são apresentadas as conclusões, limitações e perspectivas de trabalhos futuros.

MONITORAMENTO DE GESTOS USANDO SENSORES VESTÍVEIS

O monitoramento de gestos utilizando sensores vestíveis tem se consolidado como uma área essencial para o desenvolvimento de soluções tecnológicas voltadas à captura e análise do comportamento humano em tempo real ([Gomes e Sousa, 2019](#)). Com o avanço das tecnologias de sensores e a popularização de dispositivos como relógios e pulseiras inteligentes, a coleta de dados sobre movimentos humanos tornou-se mais precisa e acessível ([Wang et al., 2022](#)). Equipados com sensores inerciais, como acelerômetros e giroscópios, esses dispositivos permitem a detecção detalhada de movimentos específicos do corpo, ampliando o escopo de aplicações, especialmente no campo da saúde, onde o monitoramento contínuo pode auxiliar na prevenção e no tratamento de condições relacionadas à mobilidade e à hidratação ([Wang et al., 2022](#); [Wellnitz et al., 2019](#)).

Sensores inerciais capturam informações contínuas sobre padrões de mobilidade e atividades diárias, como o ato de levar um copo à boca, essencial para sistemas de monitoramento de ingestão de líquidos ([Anderez et al., 2020](#)). Quando processados adequadamente, esses dados fornecem informações valiosas sobre os hábitos de consumo de líquidos, desempenhando um papel fundamental na prevenção da desidratação, particularmente em populações vulneráveis, como idosos, pacientes com condições crônicas e atletas ([Cohen et al., 2021](#)).

O desenvolvimento de modelos automáticos para o reconhecimento de gestos segue geralmente um processo bem definido, composto pelas etapas de aquisição de dados, pré-processamento, extração de características e classificação, conforme ilustrado na Figura 2.1 ([Gomes e Sousa, 2019](#); [Hamatani et al., 2018](#); [Huang et al., 2020](#)). Sensores inerciais desempenham um papel central na etapa de aquisição, fornecendo séries temporais contínuas que registram as variações dos movimentos ([Wang et al., 2022](#); [Wei et al., 2023](#)). Esses dados são submetidos ao pré-processamento, que envolve limpeza, organi-

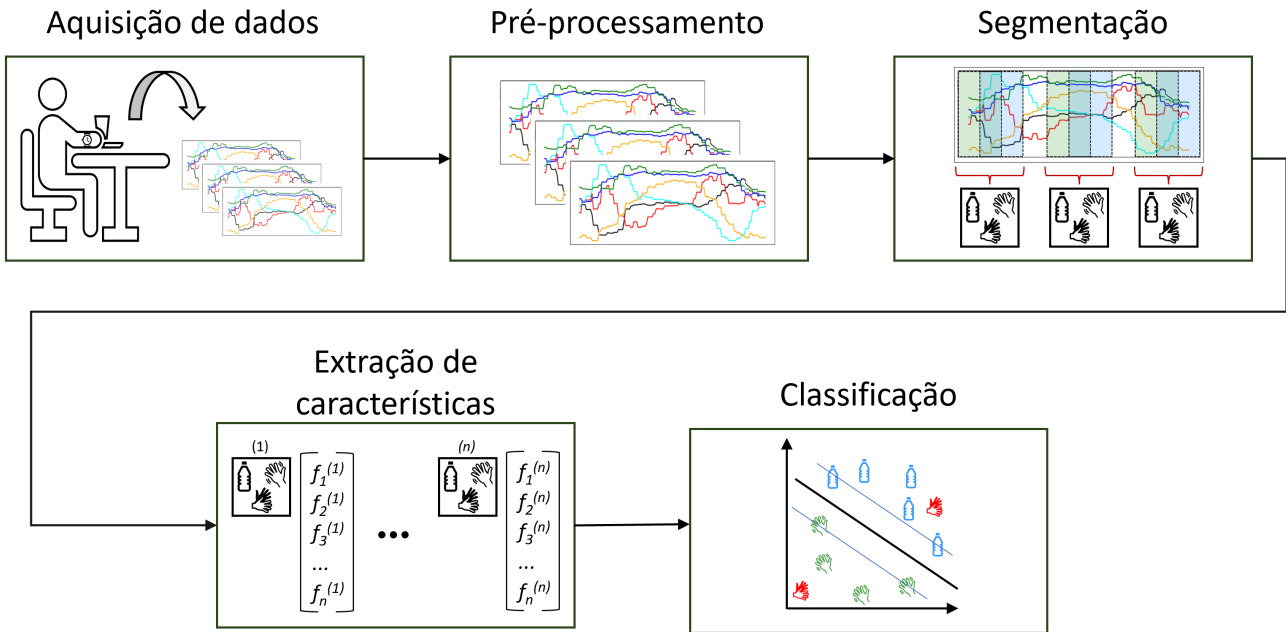


Figura 2.1: Metodologia comum; aquisição; pré-processamento, extração de características e classificação. Adaptado (Hamatani et al., 2018) (Gomes e Sousa, 2019) (Huang et al., 2020)

zação e segmentação dos gestos de interesse, preparando-os para as etapas subsequentes (Gomes e Sousa, 2019; Huang et al., 2020).

A extração de características é uma fase crucial nesse processo. Nessa etapa, os dados brutos dos sensores são transformados em representações que facilitam o aprendizado por algoritmos de classificação. Características bem projetadas têm o potencial de aumentar significativamente a precisão dos modelos (Bishop, 2006; Huang et al., 2020). Por fim, a etapa de classificação utiliza essas características para categorizar os gestos. A eficácia dessa fase depende da qualidade e quantidade dos dados disponíveis, bem como da escolha do modelo de classificação mais adequado. Tanto abordagens tradicionais quanto métodos de aprendizado profundo têm sido utilizados com sucesso em diferentes aplicações (Murphy, 2012). A seleção apropriada do modelo, aliada a um conjunto de dados robusto, é determinante para alcançar resultados precisos e confiáveis em tarefas de aprendizado de máquina.

As próximas seções detalham cada etapa desse processo, desde a aquisição de dados até a aplicação de algoritmos de aprendizado de máquina. Serão exploradas as principais técnicas e metodologias empregadas para garantir a precisão e a eficiência dos modelos de reconhecimento, com destaque para os avanços nas abordagens clássicas e profundas.

2.1 Aquisição de Dados

A etapa de aquisição de dados é fundamental para o desenvolvimento de sistemas de reconhecimento de gestos baseados em sensores vestíveis (Wang et al., 2022). Dispositivos como relógios e pulseiras inteligentes, equipados com sensores inerciais, como acelerômetros e giroscópios, são amplamente utilizados devido à sua capacidade de medir acelerações e rotações em múltiplos eixos. Essa funcionalidade permite a captura precisa de gestos, como o ato de ingestão de líquidos (Gomes e Sousa, 2019).

Segundo Hamilton (1994), os dados capturados por esses sensores são representados como uma sequência de pontos ordenados no tempo, formando uma série temporal. Essa estrutura temporal possibilita uma análise detalhada das variações nos movimentos ao longo do tempo. As Figuras 2.2 e 2.3 ilustram os sinais brutos coletados pelos acelerômetros e giroscópios durante o gesto de ingestão de líquidos, destacando os padrões de variação registrados nos diferentes eixos espaciais.

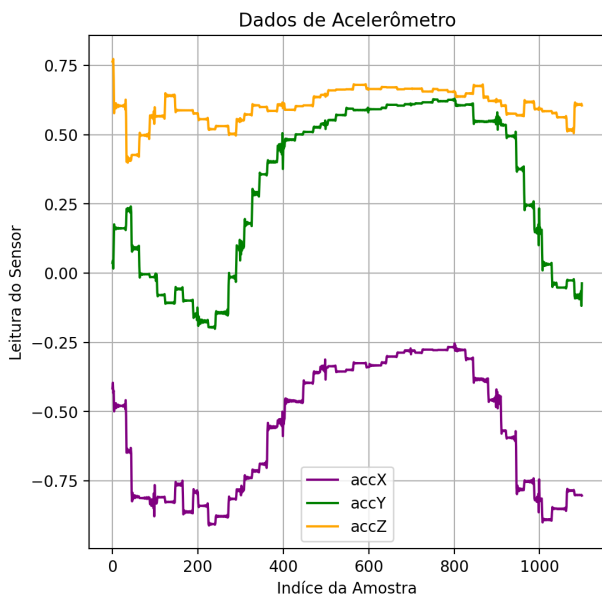


Figura 2.2: Dados do Acelerômetro para o gesto de ingestão.

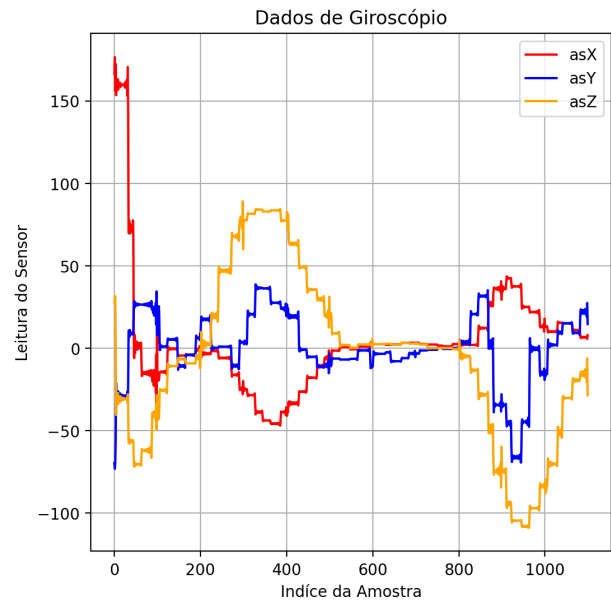


Figura 2.3: Dados do Giroscópio para o gesto de ingestão.

Esses dados constituem a base para as etapas subsequentes, como o pré-processamento e a extração de características, que são essenciais para o desenvolvimento de modelos de aprendizado de máquina. Conforme destacado por Bishop (2006), a qualidade das características extraídas dos sinais brutos tem um impacto direto no desempenho dos sistemas de classificação. Portanto, garantir a precisão e a integridade dos dados coletados é crucial para o sucesso de sistemas que visam o reconhecimento confiável de gestos de ingestão.

2.2 Pré-processamento e Segmentação de Sinais

Após a aquisição dos dados, os sinais brutos passam por um processo de pré-processamento para remover ruídos e artefatos, garantindo a qualidade e a integridade dos dados para as etapas subsequentes. Essa etapa é essencial para otimizar os sinais e envolve técnicas como filtragem e normalização.

A filtragem é aplicada para eliminar ruídos indesejados dos sinais. Filtros passa-baixa são frequentemente utilizados, permitindo a passagem de componentes de baixa frequência e atenuando as de alta frequência, que geralmente correspondem a ruídos. Filtros como o Butterworth e o de média móvel são amplamente empregados devido à sua simplicidade e eficácia na atenuação de sinais ruidosos. As Figuras 2.4 e 2.5 mostram os sinais do acelerômetro e giroscópio após a aplicação do filtro de média móvel. Esse filtro substitui cada ponto do sinal pela média de um conjunto de pontos vizinhos, resultando na atenuação eficaz das componentes de alta frequência (Bendat e Piersol, 2010; Harris, 1978). A escolha do filtro e de seus parâmetros depende das características do sinal e do tipo de ruído presente.

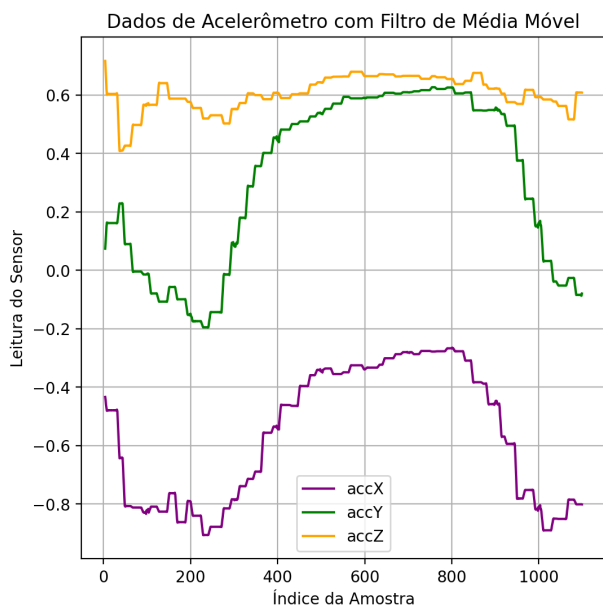


Figura 2.4: Dados do Acelerômetro para o gesto de ingestão com filtro de média-móvel aplicado.

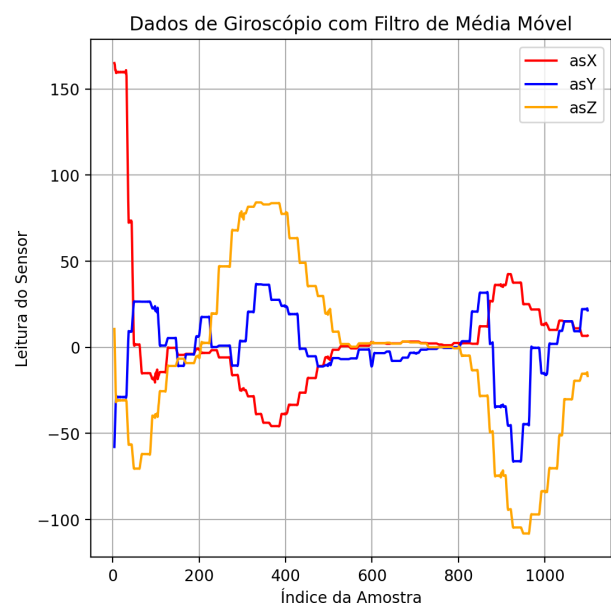


Figura 2.5: Dados do Giroscópio para o gesto de ingestão com filtro de média-móvel aplicado.

A normalização é outra etapa importante do pré-processamento. Ela ajusta os dados para um intervalo padrão, facilitando a comparação entre diferentes segmentos e contribuindo para a estabilidade e eficiência dos algoritmos de aprendizado de máquina. Métodos comuns incluem a escala mínima-máxima, que transforma os dados para o intervalo entre

0 e 1, e a padronização, que ajusta os valores para média zero e desvio padrão unitário (Bishop, 2006).

A segmentação de sinais é uma etapa crítica, pois divide os dados contínuos em segmentos menores que podem ser analisados individualmente. Esses segmentos são geralmente definidos por janelas de tempo fixas, comumente utilizando janelas do tamanho médio do gesto, com sobreposição de 50%, para garantir a captura eficaz dos eventos de interesse (Wang et al., 2022). A segmentação precisa impacta diretamente a extração de características subsequente, influenciando o desempenho dos modelos de classificação (Bishop, 2006).

2.3 Extração de Características dos Sinais

A extração de características é uma etapa essencial para métodos de aprendizado supervisionado, particularmente para modelos rasos. O objetivo desse processo é transformar os dados segmentados em representações adequadas para os algoritmos de aprendizado de máquina, extraindo informações relevantes que permitam a classificação eficiente dos gestos. A seguir, são apresentadas as principais características utilizadas no processamento de dados obtidos por acelerômetros e giroscópios.

- Média: Representa o valor médio dos pontos de dados e é calculada para cada eixo dos sensores inerciais (Altun e Barshan, 2010). É calculada como:

$$\text{Média} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.1)$$

onde, N é o número total de pontos de dados no segmento e x_i é o valor do i -ésimo ponto de dados.

- Desvio Padrão: Mede a dispersão dos pontos de dados em torno da média, refletindo a variabilidade do sinal (Maurer et al., 2006). Um desvio padrão elevado indica que os dados estão mais dispersos em relação à média. A fórmula é:

$$\text{Desvio Padrão} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \text{Média})^2} \quad (2.2)$$

onde, Média é o valor médio dos pontos de dados.

- Variância: Fornece uma medida da dispersão dos pontos de dados e é calculada como o quadrado do desvio padrão (Kwapisz et al., 2011). A variância quantifica o quanto os dados variam em relação à média.

$$\text{Variância} = \frac{1}{N} \sum_{i=1}^N (x_i - \text{Média})^2 \quad (2.3)$$

- Máximo e Mínimo: Representam os valores extremos dos pontos de dados, utilizados para identificar os limites superiores e inferiores dos sinais (Banos et al., 2014).

$$\text{Máximo} = \max_{1 \leq i \leq N} x_i, \quad \text{Mínimo} = \min_{1 \leq i \leq N} x_i \quad (2.4)$$

- Amplitude: Calculada como a diferença entre os valores máximo e mínimo, fornecendo uma medida do intervalo dos dados (Bao e Intille, 2004).

$$\text{Amplitude} = \text{Máximo} - \text{Mínimo} \quad (2.5)$$

- Curtose: Avalia o grau de concentração dos dados nas caudas ou no pico da distribuição, indicando se a distribuição é mais achatada ou mais pontuda que a distribuição normal (Westfall, 2014). A curtose é útil para identificar a presença de outliers e a "pesadez" das caudas da distribuição.

$$\text{Curtose} = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^N \left(\frac{x_i - \text{Média}}{\text{Desvio Padrão}} \right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \quad (2.6)$$

- Assimetria (*Skewness*): Mede a assimetria da distribuição dos dados, indicando a tendência dos dados em se desviarem para a direita (assimetria positiva) ou para a esquerda (assimetria negativa) em relação à média (Lara e Labrador, 2013).

$$\text{Assimetria} = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{x_i - \text{Média}}{\text{Desvio Padrão}} \right)^3 \quad (2.7)$$

As características extraídas são utilizadas como entrada para diversos algoritmos de aprendizado de máquina. Entre os modelos frequentemente empregados estão AdaBoost, Árvore de Decisão, Floresta Aleatória, Naïve Bayes, k-Nearest Neighbor e Máquina de Vetores de Suporte. Esses modelos são treinados para classificar atividades de ingestão, reconhecer gestos e, em alguns casos, estimar a quantidade de ingestão de líquidos (Altun e Barshan, 2010; Kwapisz et al., 2011; Lara e Labrador, 2013). A escolha do conjunto de características e do modelo de aprendizado desempenha um papel crucial no desempenho do sistema de classificação.

2.4 Classificação

Após a extração de características dos sinais, a etapa final do processo consiste na classificação, onde algoritmos de aprendizado de máquina são empregados para reconhecer gestos de ingestão de líquidos com base nos dados coletados pelos sensores. As abordagens de classificação podem ser agrupadas em métodos tradicionais de aprendizado supervisionado, conhecidos como modelos rasos, e métodos avançados baseados em aprendizado profundo.

2.4.1 Modelos Rasos

Modelos tradicionais como Árvores de Decisão, Florestas Aleatórias, Máquinas de Vetores de Suporte (SVMs) e k-Nearest Neighbors (k-NN), têm sido amplamente utilizados no reconhecimento de gestos. Esses métodos dependem fortemente da extração manual de características dos dados dos sensores, o que exige intervenção especializada para garantir a eficácia. Por exemplo, [Gomes e Sousa \(2019\)](#) aplicaram uma Floresta Aleatória para detectar gestos de ingestão de líquidos, alcançando precisão de 97% em condições controladas e 85% em ambientes de vida livre. Apesar do bom desempenho, esses métodos apresentam limitações quando aplicados a ambientes mais dinâmicos ou diante da variabilidade entre usuários, dificultando a generalização do modelo.

Soluções mais avançadas dentro dessa abordagem foram propostas, como o sistema Fluid-Meter de [Hamatani et al. \(2018\)](#), que utiliza Campos Aleatórios Condicionais (CRF) para, além de detectar eventos de ingestão, quantificar o volume ingerido. Esse modelo atingiu uma precisão média de 91,4% na detecção de ingestão e um erro médio de 15,3% na estimativa do volume ingerido. Tais avanços ilustram como modelos rasos podem ser adaptados para cenários específicos, embora ainda dependam de características pré-definidas e enfrentem dificuldades em capturar padrões mais complexos nos dados.

2.4.2 Modelos Profundos

O advento do aprendizado profundo trouxe novas possibilidades para o reconhecimento de gestos, permitindo que características sejam extraídas automaticamente diretamente dos dados brutos dos sensores, eliminando a necessidade de etapas manuais de extração. Redes Neurais Convolucionais (CNNs) têm demonstrado grande potencial em tarefas de reconhecimento de atividades humanas (HAR), como demonstrado por [Ronao e Cho \(2016\)](#). Nesse estudo, uma arquitetura CNN alcançou uma precisão de 94,79% na classificação de atividades básicas utilizando sensores de smartphones, superando significativamente os métodos tradicionais.

Avançando nessa linha, [Wang et al. \(2022\)](#) propuseram uma Rede Convolucional Temporal Multiestágio (MS-TCN) para monitorar gestos de ingestão de líquidos em condições de vida livre. O modelo obteve um F1-score de 94% em ambientes semi-controlados e 90% em condições de vida livre. A MS-TCN se destaca por sua capacidade de capturar dependências temporais de longo alcance, o que é crucial para a detecção precisa de gestos que variam em duração e estilo entre diferentes indivíduos. No entanto, desafios permanecem, como a necessidade de generalização para diferentes perfis de usuários e a complexidade computacional, que pode limitar sua aplicação em dispositivos com restrições de recursos.

2.5 Considerações Finais

Neste capítulo foram explorados os conceitos fundamentais e as etapas principais envolvidas no monitoramento de gestos utilizando sensores vestíveis, com destaque para o reconhecimento de atividades relacionadas à ingestão de líquidos. O avanço dos dispositivos vestíveis equipados com sensores inerciais, como acelerômetros e giroscópios, viabilizou a coleta precisa e contínua de dados de movimento, abrindo novas possibilidades para aplicações na área da saúde ([Gomes e Sousa, 2019](#); [Wang et al., 2022](#)).

A importância de um processo estruturado, desde a aquisição de dados até a classificação, foi apresentada como fundamental para o desenvolvimento de sistemas robustos de reconhecimento de gestos ([Hamatani et al., 2018](#)). As etapas de pré-processamento e segmentação dos sinais são essenciais, pois afetam diretamente a qualidade das características extraídas e o desempenho dos modelos de aprendizado ([Bishop, 2006](#)).

Além disso, foram apresentados exemplos das principais características utilizadas no desenvolvimento de modelos de reconhecimento de gestos baseados em algoritmos rasos e métodos avançados de aprendizado profundo. Esses conceitos servirão de base para a compreensão e avaliação do método proposto, permitindo a comparação entre abordagens tradicionais e profundas no contexto do reconhecimento de gestos de ingestão.

TRABALHOS RELACIONADOS

O monitoramento de gestos de ingestão de líquidos por meio de sensores vestíveis se consolidou como uma área de pesquisa promissora, impulsionada pela necessidade de soluções automáticas e objetivas para a gestão da saúde. A análise de trabalhos correlatos é fundamental para contextualizar o cenário atual, identificar as abordagens mais eficazes e compreender as lacunas que justificam novas investigações. Este capítulo apresenta uma análise de estudos seminais que serviram de base para a concepção desta dissertação, abordando desde os protocolos de coleta de dados até as arquiteturas de aprendizado de máquina empregadas. A seleção destes trabalhos foi guiada pela sua relevância na detecção de gestos de ingestão, no uso de sensores inerciais e na aplicação de técnicas de classificação que informaram diretamente as decisões metodológicas aqui adotadas.

3.1 Abordagens para Detecção de Gestos de Ingestão

A detecção automática de gestos de ingestão segue um pipeline bem estabelecido, que compreende a aquisição de dados, pré-processamento, segmentação e classificação. Os trabalhos analisados, embora compartilhem essa estrutura, apresentam variações significativas em suas abordagens, fornecendo um panorama rico das técnicas disponíveis.

3.1.1 Sensores e Protocolos de Coleta

A escolha do sensor e do protocolo de coleta de dados define a qualidade e a natureza das informações disponíveis para o treinamento dos modelos. O uso de Unidades de Medição Inercial (IMU), que integram acelerômetros e giroscópios, é predominante na área devido à sua capacidade de capturar detalhadamente a dinâmica dos movimentos.

O posicionamento do sensor no pulso é uma constante nos trabalhos analisados, sendo considerado o local ideal por ser não intrusivo e altamente eficaz na captura de gestos de mão-em-boca. Os ambientes de coleta, no entanto, variam, influenciando diretamente a complexidade e a generalização dos dados. Estudos como os de [Moccia et al. \(2022\)](#) e [Huang et al. \(2020\)](#) foram conduzidos em ambientes de laboratório controlados, onde os participantes realizaram um conjunto pré-definido de gestos. Essa abordagem permite a criação de um dataset limpo e bem anotado, ideal para a validação inicial de algoritmos. [Moccia et al. \(2022\)](#), por exemplo, envolveram 22 participantes em sessões de 10 minutos, enquanto [Huang et al. \(2020\)](#) contaram com 20 participantes realizando uma sequência de 9 atividades distintas.

Em contrapartida, [Wang et al. \(2024\)](#) avançam em direção a cenários mais realistas, propondo as bases de dados FD-I e FD-II, coletados em ambientes que descrevem como "quase-livres" (*near-free-living*). Nesses estudos, os participantes foram acompanhados por pesquisadores durante suas rotinas diárias, garantindo a captura de dados em contextos variados, como em casa, no trabalho ou em restaurantes. Essa metodologia, embora mais complexa, resulta em dados mais representativos da variabilidade encontrada no mundo real. Finalmente, [Senyurek et al. \(2019\)](#) optaram por utilizar uma base de dados pública que, apesar de coletado em ambiente controlado, incluía atividades confundidoras críticas como comer e fumar, oferecendo um desafio relevante para a distinção de gestos.

3.1.2 Pré-processamento e Segmentação de Sinais

Após a coleta, os sinais brutos dos sensores requerem um tratamento para remover ruídos e estruturá-los para a classificação. A filtragem é um passo comum, com [Moccia et al. \(2022\)](#) aplicando um filtro passa-baixa *Butterworth* de quarta ordem, enquanto [Huang et al. \(2020\)](#) utilizaram um filtro de média móvel para suavizar os sinais. A normalização dos dados também é uma prática padrão, comumente realizada através da padronização (média zero e desvio padrão unitário) para garantir que diferentes canais de sensores contribuam de forma equilibrada para o modelo.

Para a segmentação, a técnica de janelas deslizantes é a abordagem predominante. A definição do tamanho da janela e da sobreposição é um parâmetro crucial que impacta o desempenho. [Moccia et al. \(2022\)](#) testaram janelas de 2s, 3s e 6s, selecionando a de 3s (150 amostras a 50Hz) com 75% de sobreposição como a ideal. [Huang et al. \(2020\)](#), por sua vez, realizaram uma extensa avaliação de janelas entre 64 e 256 amostras com diferentes percentuais de sobreposição. Essa análise detalhada da literatura sobre o impacto do janelamento foi fundamental para a definição do protocolo experimental desta dissertação.

3.1.3 Classificação: De Modelos Rasos a Profundos

A etapa de classificação é onde os dados processados são utilizados para identificar os gestos de interesse. Os trabalhos analisados exploraram desde modelos clássicos de aprendizado de máquina até arquiteturas de aprendizado profundo.

Modelos Rasos

Modelos rasos, que dependem da extração manual de características, servem como um importante ponto de partida e base de comparação. [Huang et al. \(2020\)](#) forneceram uma análise comparativa robusta ao avaliar seis classificadores distintos: Adaptive Boosting (AdaBoost), Árvore de Decisão (DT), Floresta Aleatória (RF), Naïve Bayes (NB), k-vizinho próximo (k-NN) e Máquina de vetor de suporte (SVM). Para alimentar esses modelos, eles extraíram um conjunto de 88 características estatísticas e de domínio da frequência. Em seus resultados, o modelo AdaBoost alcançou o melhor desempenho na detecção de atividades de ingestão, com uma acurácia de 94,42%. Esta análise aprofundada de classificadores rasos foi a principal inspiração para a etapa de validação comparativa realizada neste trabalho.

Modelos Profundos

O aprendizado profundo elimina a necessidade de extração manual de características, permitindo que o modelo aprenda representações hierárquicas diretamente dos dados brutos. Uma arquitetura recorrente na literatura é a combinação de Redes Neurais Convolucionais (CNN) com Memória de Longo Prazo (LSTM).

Tanto [Senyurek et al. \(2019\)](#) quanto [Moccia et al. \(2022\)](#) implementaram uma arquitetura CNN-LSTM. Essa abordagem híbrida é particularmente eficaz, pois as camadas de CNN atuam como extratoras de características locais dos sinais, enquanto as camadas LSTM modelam as dependências temporais de longo prazo entre essas características. [Moccia et al. \(2022\)](#) demonstraram que a CNN-LSTM superou todos os outros modelos testados, incluindo SVM e RF, alcançando um F1-score médio de 90,5% na classificação de múltiplos gestos.

Indo além, [Wang et al. \(2024\)](#) propuseram uma arquitetura mais avançada, que combina uma Rede Convolucional Temporal (TCN) com um módulo de Atenção Multi-Cabeças (*Multi-Head Attention*). A TCN é projetada para capturar dependências temporais em múltiplas escalas de forma mais eficiente que as Redes Neurais recorrentes (RNN) tradicionais, enquanto o mecanismo de atenção permite que o modelo foque dinamicamente nas partes mais informativas da sequência de dados.

3.2 Análise Comparativa e aspectos da Pesquisa Atual

A análise conjunta desses trabalhos revela uma clara progressão na complexidade e eficácia das abordagens para o reconhecimento de gestos de ingestão. Uma conclusão consistente é a superioridade dos modelos de aprendizado profundo em relação aos modelos rasos, especialmente em tarefas que envolvem alta variabilidade de gestos. A capacidade das CNN de aprender características relevantes automaticamente supera a extração manual, que pode não capturar toda a complexidade dos sinais.

A eficácia da arquitetura CNN-LSTM, demonstrada por [Moccia et al. \(2022\)](#) e [Senyurek et al. \(2019\)](#), validou a importância de modelar tanto os padrões locais quanto as sequências temporais, inspirando a exploração de modelos híbridos nesta dissertação. Adicionalmente, a análise de múltiplos classificadores rasos, conduzida por [Huang et al. \(2020\)](#), forneceu uma base sólida e um protocolo de comparação para avaliar o desempenho do modelo aqui proposto em relação a abordagens mais tradicionais. Por fim, a necessidade de lidar com a variabilidade dos gestos em ambientes menos controlados, como abordado por [Wang et al. \(2024\)](#), justificou o foco em desenvolver uma arquitetura robusta e generalizável, capaz de operar para além das condições estritas de laboratório.

3.3 Considerações Finais

Este capítulo revisou os principais trabalhos que fundamentam a pesquisa em reconhecimento de gestos de ingestão de líquidos com sensores vestíveis. Foi observado um consenso no uso de sensores inerciais de pulso e uma clara tendência em direção a modelos de aprendizado profundo, que demonstram maior poder de generalização e acurácia. As metodologias de pré-processamento, segmentação por janelas e as arquiteturas de classificação (profunda e "rasa") descritas nesses estudos foram essenciais para definir o pipeline experimental e os modelos de base para comparação nesta dissertação. A partir desta base sólida, o capítulo seguinte detalhará a metodologia proposta, que busca avançar em relação ao estado da arte ao abordar as lacunas identificadas.

InGesture: Um conjunto de dados de sensores inerciais para reconhecimento de gestos de ingestão de líquidos

Neste capítulo é apresentado o protocolo de coleta de dados desenvolvido para capturar gestos de ingestão de líquidos e gestos análogos utilizando sensores inerciais. São detalhadas as características dos dados coletados, os procedimentos de processamento aplicados e as limitações enfrentadas ao longo do processo. Adicionalmente, é realizada uma análise comparativa com outros conjuntos de dados relevantes disponíveis na literatura, destacando as contribuições e singularidades do conjunto de dados proposto.

4.1 Protocolo de Coleta

4.1.1 Participantes

Foram recrutados 50 participantes para este estudo (34 homens, 16 mulheres, $28,5 \pm 10,5$ anos, altura = 170 ± 8 cm, peso = $68,3 \pm 13,2$ kg). Os participantes variaram em idade de 18 a 67 anos, peso de 47,6 a 111,1 kg e altura de 150 a 188 cm, representando uma amostra diversificada da população. Alguns participantes contribuíram com múltiplas sessões, totalizando 65 sessões de coleta.

4.1.2 Dispositivo de Coleta

Para a captura dos dados foi utilizado o dispositivo **WT901BLECL5** (Figura 4.1), um sensor inercial multi-eixo que integra acelerômetro, giroscópio e magnetômetro. O dispositivo foi



Figura 4.1: Dispositivo de Monitoramento de Pulso Witmotion.

fixado no pulso dominante de cada participante, permitindo a captura dos movimentos associados aos gestos de interesse. O dispositivo foi calibrado antes de cada sessão de coleta para garantir a precisão dos dados.

Os dados coletados consistem em sinais contínuos de acelerômetro e giroscópio, registrados a uma taxa de amostragem de 200 Hz. Essa alta resolução temporal permite capturar detalhes sutis dos movimentos, oferecendo uma representação rica dos gestos realizados. Os sinais foram coletados nos três eixos espaciais do acelerômetro (Acc_x , Acc_y , Acc_z) e giroscópio (Gyr_x , Gyr_y , Gyr_z). A Figura 4.2 ilustra exemplos dos sinais inerciais coletados, separados por tipo de gesto.

4.1.3 Procedimento

As atividades foram realizadas em ambiente controlado, com os participantes posicionados em pé ou sentados, como ilustrado na Figura 4.3. Cada sessão consistiu em uma sequência pré-definida de gestos, intercalando ações de ingestão de líquidos e gestos análogos. Para avaliar a influência do tipo de recipiente nos sinais coletados, os gestos de ingestão foram realizados utilizando diferentes recipientes como garrafas, copos e xícaras.

Cada participante executou os gestos na seguinte sequência de atividades: Condição Livre (0) → Ingestão de Líquidos (1) → Atender o Celular (2) → Coçar a Cabeça (3) → Ajustar os Óculos (4) → Passar a Mão no Rosto (5) → Segurar o Queixo (6) → Esticar as Mãos Atrás do Pescoço (7). Essa sequência foi repetida 4 vezes por cada participante. A Figura 4.4 ilustra os gestos investigados e a Tabela 4.2 faz um resumo sobre a duração e ocorrências dos gestos.

Os gestos foram rotulados em tempo real utilizando um aplicativo desenvolvido especificamente para a coleta, que sincronizava os eventos com os sinais inerciais. A Tabela 4.1 exemplifica os dados brutos anotados, indicando os pontos de início e fim de cada gesto.

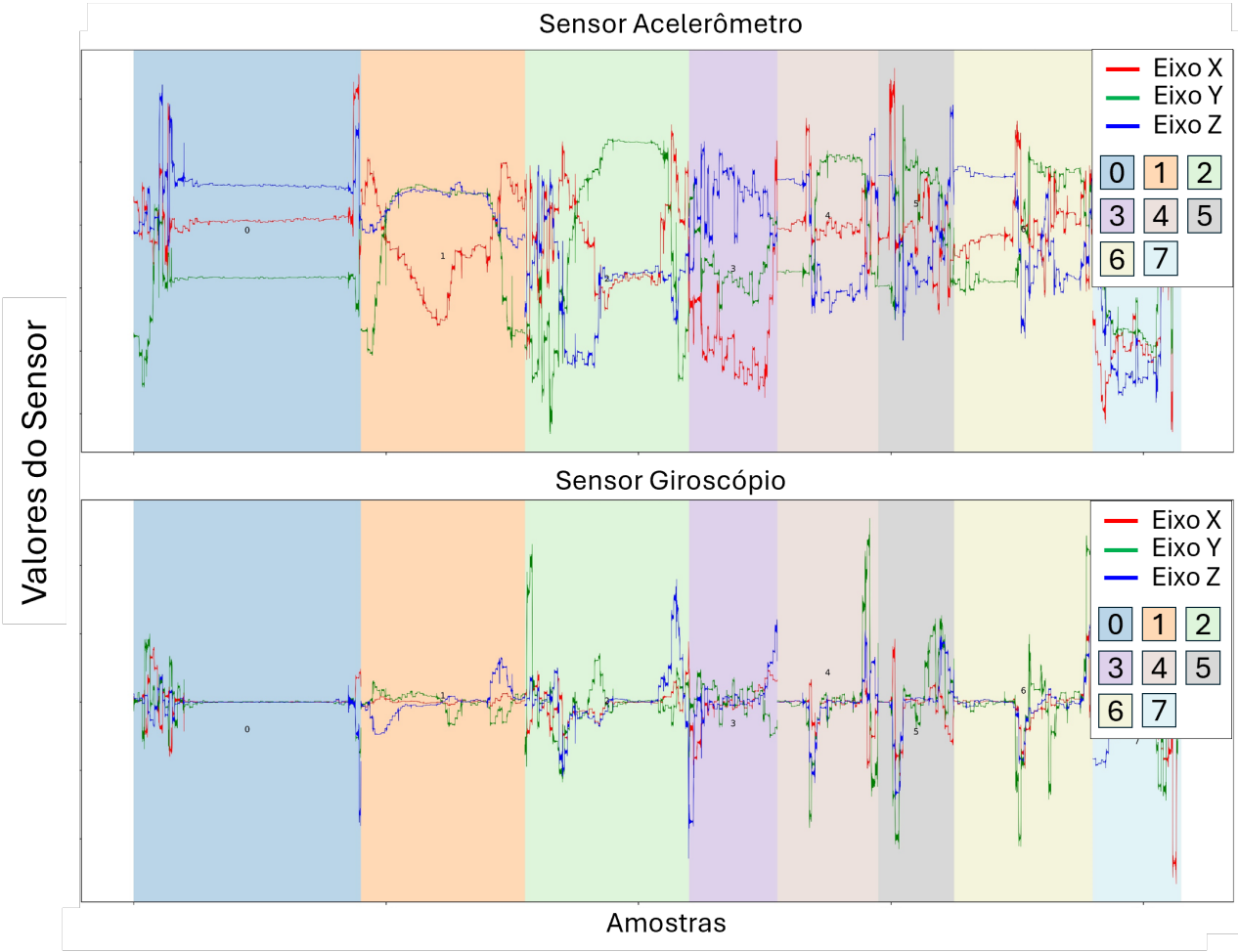


Figura 4.2: Exemplo dos sinais inerciais de gestos coletados durante a execução das coletas.



Figura 4.3: Participantes durante a coleta de dados.

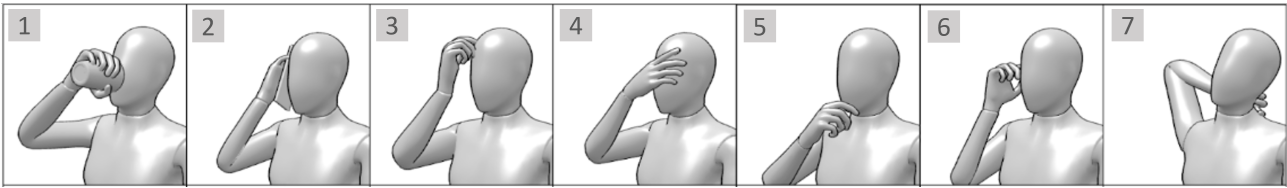


Figura 4.4: Gestos análogos a ingestão.

time	Acc_x	Acc_y	Acc_z	Gyr_x	Gyr_y	Gyr_z	label
42929521.0	-0.957	0.066	0.205	-2.136	-0.916	0.305	0
42929525.0	-0.957	0.071	0.204	-2.764	-0.007	-0.231	0
42945653.0	-0.788	0.274	0.715	-4.761	55.176	-15.991	1
42945658.0	-0.821	0.223	0.734	-9.163	55.562	-28.487	1
42962814.0	-0.011	-0.302	1.591	-102.844	-20.203	-68.604	2
42962818.0	0.303	0.045	1.292	-55.634	-64.136	-63.101	2
42990536.0	-0.003	0.826	0.118	27.771	85.999	282.837	3
42990540.0	-0.379	0.864	0.085	33.711	107.965	322.249	3
43049928.0	-0.884	-0.157	0.257	-0.122	0.793	-0.122	4
43049933.0	-0.943	-0.185	0.267	-1.953	-4.861	-14.530	4
43111394.0	-0.475	-0.063	0.815	0.671	-2.197	0.793	5
43111399.0	-0.468	-0.079	0.795	-11.744	-18.539	-13.916	5
43170315.0	0.052	1.114	0.360	1.038	1.587	3.784	6
43170319.0	0.056	1.120	0.363	1.142	1.786	3.922	6
43231276.0	-1.170	-0.152	0.730	11.169	-8.057	44.067	7
43231280.0	-1.309	-0.218	0.777	17.134	-1.390	34.651	7

Tabela 4.1: Amostra de dados brutos anotados com sensores inerciais.

Tabela 4.2: Resumo dos gestos na base de dados, incluindo duração total, número de ocorrências e informações relevantes sobre o gesto de ingestão.

Código	Gesto	Duração Total (s)	Ocorrências
0	Condição Livre	28708.50	2694
1	Ingestão de Líquidos	5670.50	1325
2	Atender celular	2921.00	666
3	Coçar a cabeça	462.00	132
4	Ajustar óculos	486.00	132
5	Passar a mão no rosto	476.00	132
6	Segurar Queixo	506.50	132
7	Esticar as mãos atrás da cabeça	561.50	132
-	Total (Todos os Gestos Combinados)	39792.00	-
-	Duração média do gesto de ingestão	4.28 segundos	-

4.2 Comparação com Outros Conjuntos de Dados

Para contextualizar a relevância e singularidade do nosso conjunto de dados, nomeado InGesture, foi realizada uma comparação detalhada com outros *datasets* proeminentes na área de reconhecimento de gestos relacionados à ingestão de líquidos e atividades análogas. A Tabela 4.3 sintetiza as principais características desses conjuntos de dados.

Bases	Dispositivos	Participantes	Freq. Amostragem	Gestos	Ambiente
InGesture (Proposto)	WT901BLECL5	50	200 Hz	Ingestão de líquidos, atender celular, coçar a cabeça, ajustar óculos, passar a mão no rosto, segurar o queixo, esticar as mãos atrás do pescoço	Controlado
FD-I (Wang et al., 2024)	Shimmer3 IMU	34	64 Hz	Comer e beber	Livre
Clemson (Shen et al., 2016)	IMU no pulso	271	15 Hz	Ingestão de alimentos e bebidas	Controlado
FreeFIC (Kyritsis et al., 2021)	Acelerômetro, Giroscópio	12	100 Hz	Comer e beber	Livre

Tabela 4.3: Comparação entre *Base de dados* de ingestão de líquidos e gestos análogos.

O InGesture distingue-se como o único conjunto de dados dedicado à captura e análise de gestos de ingestão de líquidos, com uma abordagem estruturada para incluir tanto os gestos principais quanto os análogos. Essa especificidade contrasta com outros conjuntos de dados, que abrangem gestos relacionados a ingestão de alimentos e líquidos, mas não apresentam o mesmo foco detalhado e sistemático.

A variabilidade de gestos no InGesture é um dos seus grandes diferenciais. Ele inclui uma gama diversificada de ações, desafiando os modelos a reconhecer padrões em gestos semelhantes, o que melhora a generalização. Essa complexidade é essencial para o desenvolvimento de modelos mais robustos e aplicáveis em cenários práticos.

Outro destaque é a amostra de 50 participantes, significativamente maior e mais representativa em termos de idade, gênero e variação de características físicas quando comparada as outras bases de dados. Essa abrangência reduz vieses e aumenta a aplicabilidade dos modelos treinados com o InGesture.

As outras bases de dados, como o FD-I e o FreeFIC, embora utilizem dados coletados em ambientes livres, têm um foco mais genérico em gestos associados a alimentação, o que pode limitar sua aplicação específica ao reconhecimento de ingestão de líquidos. Além disso, a menor variedade de gestos e amostra reduzida de participantes comprometem a capacidade de generalização dos modelos desenvolvidos a partir deles. Já o Clemson Dataset, embora ofereça dados coletados em ambientes parcialmente controlados, também

apresenta menor amplitude de gestos e participantes, restringindo sua aplicabilidade em cenários mais complexos.

4.3 Aspectos Éticos e Legais

Este estudo foi conduzido em conformidade com as diretrizes éticas aplicáveis a pesquisas envolvendo seres humanos. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Amazonas, sob o número de protocolo **70951823.0.0000.5020**. Todos os participantes foram devidamente informados sobre os objetivos do estudo, procedimentos, possíveis riscos e benefícios, antes de assinarem o Termo de Consentimento Livre e Esclarecido. Esse termo garantiu a confidencialidade das informações e o uso exclusivo dos dados para fins acadêmicos e científicos.

Além disso, foram implementadas medidas rigorosas para assegurar a privacidade e a proteção dos dados coletados. Todos os dados foram anonimizados, impossibilitando a identificação dos participantes, e armazenados em ambientes seguros.

4.4 Considerações Finais

Neste capítulo detalhamos os protocolos de coleta de dados desenvolvidos, enfatizando as metodologias aplicadas e as contribuições do conjunto de dados InGesture para a área de reconhecimento de gestos. A análise comparativa com outras bases de dados evidencia as vantagens da nossa abordagem, como a diversidade de gestos e a amplitude da amostra de participantes, que aumentam a representatividade e robustez dos modelos desenvolvidos com base nesses dados.

Os desafios enfrentados durante a coleta, assim como as limitações inerentes ao uso de um ambiente controlado, abrem caminho para futuras pesquisas voltadas à expansão do conjunto de dados, incluindo novos contextos e cenários de coleta em ambiente livre.

CLASSIFICAÇÃO DE GESTOS DE INGESTÃO DE LÍQUIDO

Este capítulo descreve em detalhe a metodologia adotada para o reconhecimento de gestos de ingestão de líquidos a partir de sinais inerciais obtidos com sensores de pulso. O foco é apresentar o pipeline de classificação de forma clara e reproduzível, desde os sinais brutos até a etapa final de classificação. A coleta dos dados e sua caracterização foram descritas no Capítulo 4, enquanto a avaliação experimental e os resultados quantitativos são discutidos no Capítulo 6.

5.1 Fluxo geral do método proposto

O processo mostrado na Figura 5.1 inicia-se com os *dados brutos* de acelerômetro e giroscópio e, em cada iteração, procede-se à divisão LOSO, que é crucial para medir generalização entre sujeitos e, sobretudo, para evitar vazamento de informação entre treino e teste. Em seguida, o pré-processamento cumpre dois papéis essenciais: a reamostragem reduz redundância temporal e custo computacional mantendo a dinâmica relevante; a normalização Z-score equaliza escalas entre eixos/sujeitos, estabilizando o treinamento e impedindo que amplitudes maiores dominem o ajuste do modelo. A segmentação por janela transforma séries contínuas em amostras de comprimento fixo (condição necessária para redes neurais) e o voto majoritário na janela mitiga incertezas de fronteira e ruído de anotação. O treinamento sob LOSO garante que estatísticas (p. ex., da normalização) sejam estimadas apenas no treino e aplicadas ao teste do mesmo *fold*, preservando a validade da avaliação. Por fim, a classificação produz as probabilidades por classe para cada janela e as métricas agregadas por *fold* (ex: F1-score) sintetizam o desempenho de forma comparável e reproduzível. Cada decisão segue práticas consolidadas em HAR com IMU e aplicações de ingestão (Moccia et al., 2022; Wang et al., 2024).

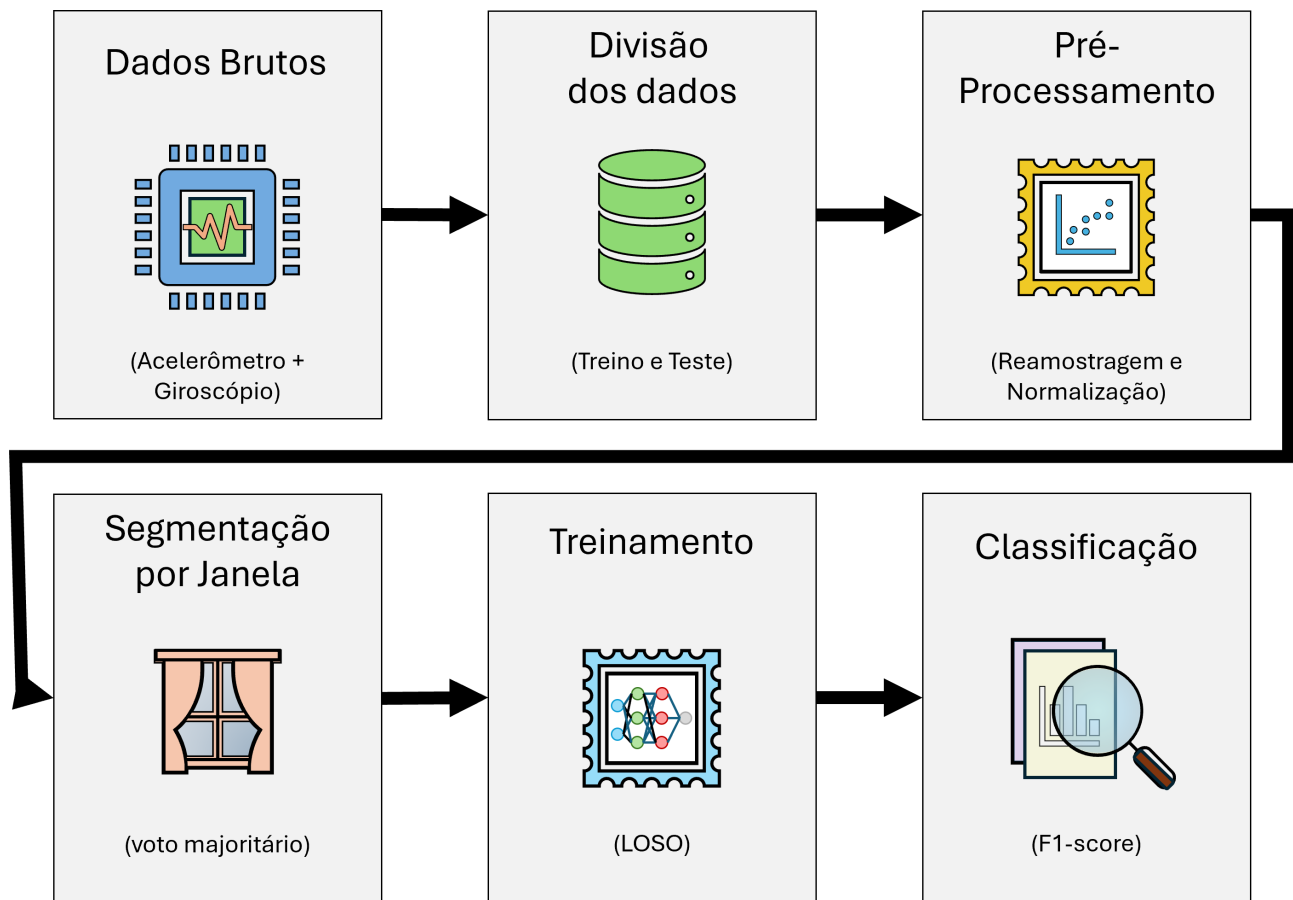


Figura 5.1: Fluxo geral do processo de classificação utilizado nesta pesquisa.

5.2 Divisão dos Dados (Treino/Teste)

A divisão dos dados é realizada **antes** do pré-processamento para evitar vazamento de informação. O particionamento segue o esquema *Leave-One-Subject-Out* (LOSO): em cada iteração, todos os registros de um participante compõem o conjunto de teste, enquanto os dados dos demais participantes compõem o treino. Essa configuração mede generalização intersujeitos, prática comum em HAR com sensores vestíveis (Moccia et al., 2022; Wang et al., 2024). Todas as estatísticas calculadas em etapas posteriores (por exemplo, normalização) são estimadas exclusivamente no conjunto de treino do respectivo *fold* LOSO e aplicadas ao teste.

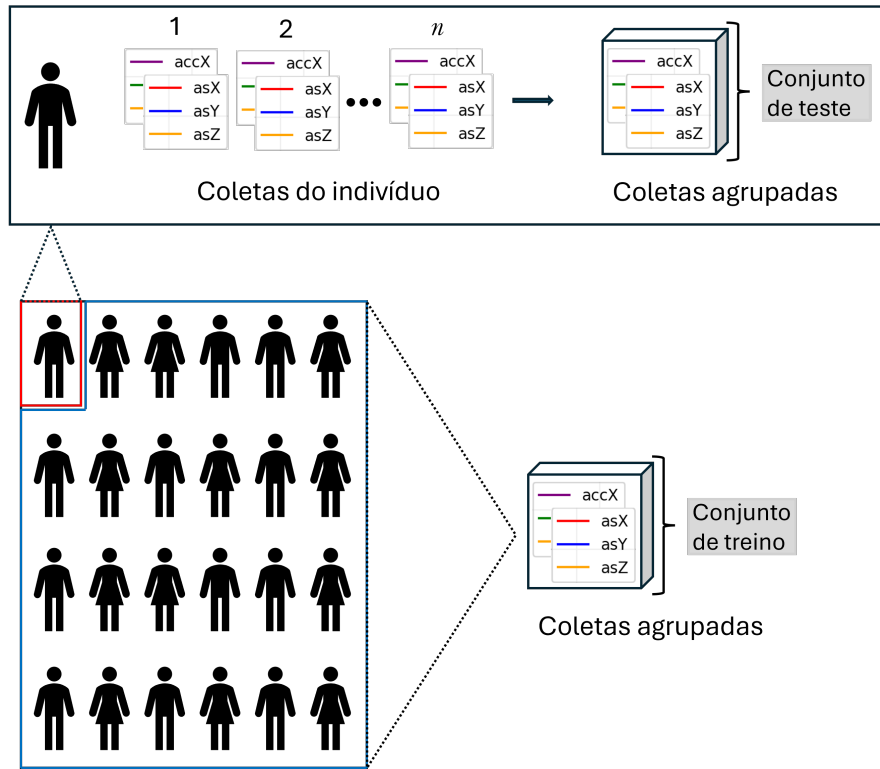


Figura 5.2: Esquema de validação LOSO.

5.3 Pré-processamento dos Dados

O pré-processamento tem como objetivo padronizar a escala dos sinais e preparar entradas com comprimento fixo para os modelos (Wang et al., 2024).

Os sinais foram adquiridos a 200 Hz e re-amostrados para 50 Hz, reduzindo redundância temporal e custo computacional, preservando a dinâmica dos gestos de ingestão conforme práticas em HAR com IMUs (Moccia et al., 2022).

Para reduzir variabilidade entre sujeitos e sessões, aplica-se normalização Z-score por eixo (aceleração e giroscópio):

$$x' = \frac{x - \mu}{\sigma}.$$

As estatísticas μ e σ são estimadas no treino do fold LOSO e reutilizadas no teste do mesmo fold, evitando vazamento de informação (Wang et al., 2024).

5.4 Segmentação por Janela

Os sinais contínuos de IMU são transformados em amostras de comprimento fixo por meio de **janelas deslizantes** com **sobreposição**. Seja $x(t) \in \mathbb{R}^d$ o sinal multieixo e $a(t) \in \mathcal{C}$ a

anotação de classe no tempo. Definimos a i -ésima janela como

$$W_i = [t_0 + iS, t_0 + iS + L),$$

onde L é a duração da janela e S é o passo (*stride*) entre inícios de janelas consecutivas, com $S < L$ para garantir sobreposição.¹ Os valores específicos de L e S são apresentados no capítulo de Experimentos.

Cada janela W_i recebe um rótulo por voto majoritário com base na anotação temporal contida no seu interior. Para cada classe $c \in \mathcal{C}$, computa-se a duração (ou contagem de amostras) dessa classe na janela,

$$\text{dur}_c(W_i) = \int_{t \in W_i} \mathbf{1}[a(t) = c] dt \quad (\text{ou equivalente discreto em amostras}),$$

e define-se o rótulo da janela por

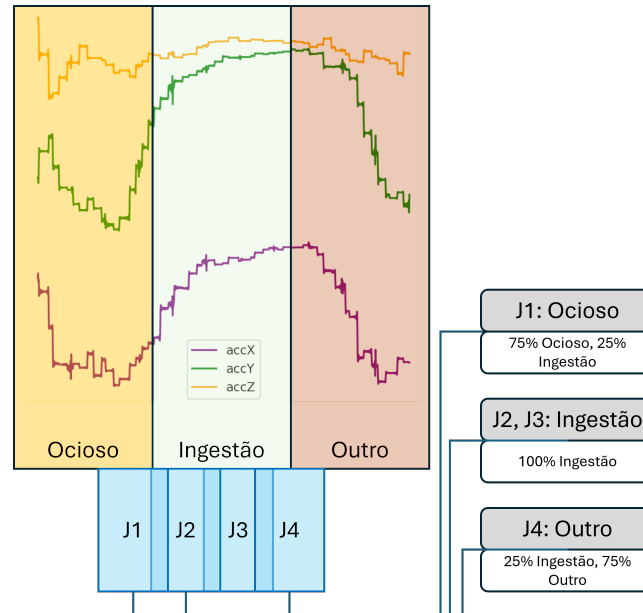
$$\hat{y}(W_i) = \arg \max_{c \in \mathcal{C}} \text{dur}_c(W_i).$$

Para fins descritivos, a proporção por classe na janela é $p_c(W_i) = \text{dur}_c(W_i)/L$ (fração da janela ocupada pela classe c).

Janelas que cruzam fronteiras entre classes podem conter múltiplos rótulos de *ground truth*. Nesses casos, o voto majoritário atribui à janela a classe com maior duração temporal interna. Isso está ilustrado na Fig. 5.3: a janela J1 cobre majoritariamente *Ocioso* (75%) e minoritariamente *Ingestão* (25%), logo $\hat{y}(W_1)=\text{Ocioso}$; a J2 e J3 recai integralmente sobre *Ingestão* ($p_{\text{Ingestão}}=100\%$), portanto $\hat{y}(W_2)=\text{Ingestão}$; e a J4 cruza *Ingestão* e *Outro* com predominância desta última (25% vs. 75%), resultando em $\hat{y}(W_4)=\text{Outro}$. Essa estratégia é amplamente utilizada em reconhecimento de gestos com IMU por estabilizar rótulos diante de variações de duração e incertezas de fronteira (Moccia et al., 2022; Wang et al., 2024).

¹Adota-se intervalo *fechado à esquerda e aberto à direita* $[\cdot, \cdot)$ para evitar contagem dupla em fronteiras.

Figura 5.3: Janelas deslizantes sobrepostas ($S < L$) e rotulação por *voto majoritário*. A barra inferior de cada janela resume a proporção temporal das classes no seu interior; o cartucho à direita explicita o rótulo atribuído (J1, J2, J3 e J4 no exemplo).



5.5 Arquitetura de Classificação

A arquitetura é uma CNN 1D (Figura 5.4) para séries temporais multivariadas de IMU, abordagem consolidada em HAR (Moccia et al., 2022; Wang et al., 2024). A entrada é a janela segmentada (após reamostragem e normalização). O processamento é feito por três ramificações convolucionais paralelas (Conv1D), cada uma com 128 filtros e tamanhos de kernel 3, 5 e 7. As saídas são concatenadas e passam por uma Conv1D adicional com 256 filtros (*refinamento*). Em seguida, aplica-se GlobalAveragePooling1D e uma camada Dense com softmax para a predição das classes.

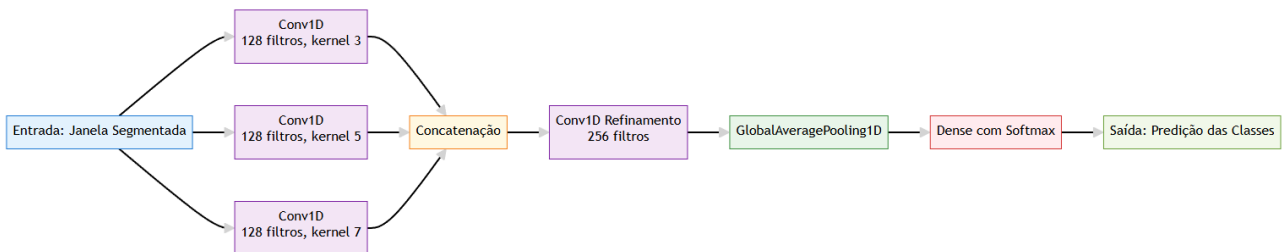


Figura 5.4: Arquitetura convolucional 1D utilizada para classificação de gestos.

5.6 Treinamento (LOSO) e Validação

O treinamento é executado por fold LOSO: em cada iteração, estima-se a normalização no treino, treina-se a rede somente com os sujeitos do treino e avalia-se no sujeito deixado de fora. Não há compartilhamento de amostras ou estatísticas entre treino e teste. Esse protocolo avalia a robustez intersujeitos, requisito em cenários de uso livre (Wang et al., 2024).

5.7 Avaliação do Modelo e Métricas de Desempenho

A avaliação do desempenho do modelo é realizada em cada *fold* do processo de validação cruzada LOSO. Para cada janela de dados de teste, a camada de saída `softmax` da rede neural calcula um vetor de probabilidades, indicando a confiança do modelo para cada uma das classes de ingestão. A classe com a maior probabilidade é então selecionada como a predição final para aquela janela.

Para quantificar a performance, utilizamos um conjunto de métricas padrão em tarefas de classificação, que oferecem uma visão completa da capacidade do modelo:

- Acurácia: Proporção de janelas classificadas corretamente. Serve como uma medida geral de desempenho.
- Precisão: Mede a proporção de predições positivas que estavam corretas. É um indicador da confiabilidade do modelo quando ele afirma que um gesto ocorreu.
- Revocação: Mede a proporção de gestos reais que foram corretamente identificados pelo modelo. Indica a capacidade do modelo de não omitir eventos.
- F1-Score: Média harmônica entre precisão e revocação, sendo uma métrica robusta que equilibra ambas, especialmente útil em cenários com classes desbalanceadas.

Ao final do treinamento de todos os *folds*, os resultados das métricas são agregados. A performance final do modelo é reportada através da média e do desvio padrão de cada métrica ao longo dos *folds*. Essa abordagem, alinhada às práticas de avaliação em HAR (Moccia et al., 2022; Wang et al., 2024), não apenas resume o desempenho geral, mas também indica a estabilidade e a capacidade de generalização do modelo entre diferentes sujeitos.

5.8 Considerações Finais

Este capítulo apresentou a metodologia de classificação desenvolvida, composta pelas etapas de pré-processamento (re-amostragem, normalização e segmentação por janelas), definição da arquitetura convolucional 1D e estratégia de treinamento com validação LOSO. Essas etapas estabelecem um fluxo coerente entre sinais brutos e classificação final, permitindo que o processo seja replicado em futuros estudos. A análise quantitativa dos resultados será apresentada no próximo capítulo.

EXPERIMENTOS E RESULTADOS

Este capítulo apresenta a avaliação empírica das arquiteturas de aprendizado de máquina desenvolvidas e investigadas para o reconhecimento de gestos de ingestão de líquidos. O objetivo central é conduzir uma análise quantitativa rigorosa, validando a eficácia do método proposto em relação a uma linha de base de métodos clássicos e a reimplementações de arquiteturas de ponta da literatura.

Todos os experimentos foram executados sobre a base de dados *InGesture*, detalhado no Capítulo 4, o que permitiu um controle metodológico preciso e uma avaliação justa e consistente entre todos os modelos. A análise foi estruturada em dois cenários de classificação (binário e multiclasse) para avaliar os modelos em tarefas de complexidade variada.

6.1 Protocolo Experimental

Para garantir a validade científica, a robustez e a reprodutibilidade dos achados, foi estabelecido um protocolo experimental detalhado, cujas decisões foram diretamente informadas pelos desafios e práticas discutidos nos trabalhos correlatos.

6.1.1 Base de Dados

A base de dados *InGesture* representa a base fundamental sobre a qual se assenta toda a validação empírica realizada. Conforme descrito no Capítulo 3, este conjunto de dados foi especificamente construído para esta pesquisa e possui características que o tornam ideal para os objetivos propostos. Com registros de 50 participantes, a base de dados oferece uma diversidade demográfica e de padrões de movimento que é crucial para o treinamento e teste de modelos generalizáveis. Ele contém oito classes de gestos, sendo uma de ingestão de líquidos e sete de movimentos análogos (ex: coçar a cabeça, atender o celular), que

servem como "distratores" realistas e desafiam a capacidade dos modelos de discriminar ações sutilmente diferentes. O *InGesture* apresenta desbalanceamento natural de classes (por exemplo, maior frequência de "ingestão" e "condição livre" em relação às demais). Para tornar a avaliação transparente e comparável, reportamos Precisão, Revocação e F1 por classe; além disso, destacamos explicitamente as métricas da classe de maior interesse (Ingestão)

6.1.2 Método de Validação

Um dos maiores desafios no reconhecimento de atividades baseado em sensores vestíveis é a variabilidade inter-sujeito. Um modelo pode apresentar alta acurácia para os indivíduos nos quais foi treinado, mas falhar ao ser aplicado a um novo usuário. Para endereçar diretamente este desafio e avaliar a capacidade de generalização dos modelos, adotou-se a estratégia de validação cruzada *Leave-One-Subject-Out (LOSO)*. Neste método, em cada iteração, os dados de um participante são completamente separados para o conjunto de teste, enquanto os dados de todos os outros 49 participantes são usados para o treinamento. O processo é repetido 50 vezes, até que cada participante tenha servido como teste uma vez. O resultado final é a média do desempenho em todas as iterações, fornecendo uma estimativa robusta e realista de como o modelo se comportaria em uma aplicação prática com novos usuários.

6.2 Pré-processamento

6.2.1 Dados Brutos

Após os dados brutos serem divididos em conjuntos de treino e teste, são submetidos a uma sequência de etapas de pré-processamento antes de serem utilizados pelos modelos de classificação. Diferentemente de outras abordagens da literatura, optou-se por não aplicar uma etapa de filtragem de sinal, a fim de avaliar a capacidade dos modelos de aprendizado profundo de aprenderem a lidar com os sinais em sua forma mais próxima da original, contendo ruídos naturais.

6.2.2 Re-amostragem

A primeira etapa consistiu na redução da frequência de amostragem original de 200Hz para as frequências de trabalho de 50Hz, que serão avaliadas nos experimentos. O processo de subamostragem foi realizado através de dizimação, conforme descrito no Algoritmo 1. Para o sinal do sensor, uma a cada N amostras era mantida, onde N é o fator de redução. Para

os rótulos, o rótulo da nova amostra foi definido pelo valor majoritário dentro da janela de N amostras correspondente, garantindo a consistência da anotação.

Algorithm 1 Reamostragem (Downsampling) dos Sinais

```

1: procedimento REAMOSTRAR(dados_x, dados_y, taxa_original, nova_taxa)
2:   fator_reamostragem  $\leftarrow$  int(taxa_original/nova_taxa)
3:   x_reamostrado  $\leftarrow$  dados_x[:,fator_reamostragem]
4:   y_reamostrado  $\leftarrow$  []
5:   para i  $\leftarrow$  0 to len(dados_y) step fator_reamostragem faça
6:     segmento_y  $\leftarrow$  dados_y[i : i + fator_reamostragem]
7:     se len(segmento_y) > 0 então
8:       rotulo_majoritario  $\leftarrow$  argmax(contagem_ocorrencias(segmento_y))
9:       Adicionar rotulo_majoritario a y_reamostrado
10:  retornar x_reamostrado, np.array(y_reamostrado)

```

6.2.3 Normalização

Após a reamostragem, os dados dos sensores foram normalizados utilizando a padronização Z-score, implementada através do `StandardScaler` da biblioteca *Scikit-learn*. Este processo transforma os dados para que tenham média zero e desvio padrão unitário. Para evitar o vazamento de dados, os parâmetros do `StandardScaler` (média e desvio padrão) foram ajustados exclusivamente com os dados de treinamento (usando o método `fit_transform`) em cada um dos *folds* da validação cruzada LOSO.

6.2.4 Segmentação (Janelamento)

Os sinais contínuos foram segmentados em janelas de tamanho fixo com sobreposição, utilizando a técnica de janelas deslizantes, conforme detalhado no Algoritmo 2. Para cada janela de dados gerada, um único rótulo foi atribuído usando voto majoritário naquele segmento temporal. Esse processo transforma a série temporal contínua em um conjunto de amostras de tamanho fixo, adequado para a entrada nos modelos de classificação.

Algorithm 2 Segmentação por Janela Deslizante

```

1: procedimento JANELADESLIZANTE( $x, y, tamanho\_janela, passo$ )
2:    $dados \leftarrow [], rotulos \leftarrow []$ 
3:    $inicio \leftarrow 0$ 
4:   enquanto  $inicio + tamanho\_janela \leq \text{len}(x)$  faça
5:      $fim \leftarrow inicio + tamanho\_janela$ 
6:      $segmento\_x \leftarrow x[inicio : fim]$ 
7:      $segmento\_y \leftarrow y[inicio : fim]$ 
8:      $rotulos\_unicos, contagens \leftarrow \text{unicos}(segmento\_y, \text{return\_counts}=\text{True})$ 
9:      $rotulo\_majoritario \leftarrow rotulos\_unicos[\text{argmax}(contagens)]$ 
10:    Adicionar  $segmento\_x$  a  $dados$ 
11:    Adicionar  $rotulo\_majoritario$  a  $rotulos$ 
12:     $inicio \leftarrow inicio + passo$ 
13:  retornar  $\text{np.array}(dados), \text{np.array}(rotulos)$ 

```

6.3 Modelagem e Classificação

Foram investigadas duas famílias de modelos para a tarefa de classificação: os de aprendizado clássico, que exigem extração manual de características e servem como uma linha de base, e os de aprendizado profundo, que aprendem características de forma automática e representam o foco principal desta dissertação.

6.3.1 Modelos de Classificação Clássicos (Baselines Rasos)

Para estabelecer uma performance de referência, foram implementados modelos clássicos inspirados na análise de [Huang et al. \(2020\)](#). Estes métodos representam o paradigma tradicional de classificação, que depende de uma etapa explícita de extração de características. Para alimentar os classificadores, foi extraído um vetor de 48 atributos estatísticos para cada janela de dados, conforme detalhado no Algoritmo 3.

Algorithm 3 Extração de Características Estatísticas

```

1: procedimento EXTRAIRCARACTERISTICAS(segmentos)
2:   caracteristicas  $\leftarrow$  []
3:   para todo segmento in segmentos faça
4:     caracteristicas_segmento  $\leftarrow$  []
5:     para todo i in range(num_eixos) faça
6:       sinal  $\leftarrow$  segmento[:, i]
7:       media  $\leftarrow$  np.mean(sinal)
8:       desvio_pad  $\leftarrow$  np.std(sinal)
9:       var  $\leftarrow$  np.var(sinal)
10:      max_val  $\leftarrow$  np.max(sinal)
11:      min_val  $\leftarrow$  np.min(sinal)
12:      amplitude  $\leftarrow$  max_val - min_val
13:      curtose  $\leftarrow$  kurtosis(sinal)
14:      assimetria  $\leftarrow$  skew(sinal)
15:      Estender caracteristicas_segmento com [media, desvio_pad, var, ...]
16:      Adicionar caracteristicas_segmento a caracteristicas
17:   retornar np.array(caracteristicas)

```

6.3.2 Configuração dos Experimentos

Foram conduzidos experimentos comparando os modelos propostos com *baselines* da literatura, incluindo abordagens clássicas e de aprendizado profundo. Foram utilizados os hiperparâmetros listados na Tabela 6.1. A taxa de aprendizagem foi fixada em 1×10^{-3} , janelas em 3.0 segundos e re-amostragem em 50hz, assim como otimizador Adam em todos os experimentos, valores padrões que demonstraram boa convergência em testes preliminares.

Tabela 6.1: Hiperparâmetros avaliados nos experimentos.

Hiperparâmetro	Valores Testados
Otimizador	Adam
Tamanho da Janela	3.0s
Sobreposição	50%
Frequência de Amostragem	50Hz
Taxa de Aprendizagem	1×10^{-3} (fixa)

6.4 Resultados Aprendizado de máquina clássico

Para estabelecer uma linha de base robusta, foram avaliados seis modelos de aprendizado de máquina clássico, seguindo a abordagem comparativa de [Huang et al. \(2020\)](#). Esses modelos, conhecidos como "rasos", dependem de um processo de extração de características. Para esta análise, os sinais foram segmentados utilizando uma janela de 3 segundos e uma frequência de 50Hz, a partir da qual 48 atributos estatísticos e de domínio da frequência foram extraídos para alimentar os classificadores. Os resultados, obtidos no cenário de detecção de ingestão (binário), estão sumarizados na Tabela 6.2.

Tabela 6.2: Resultados métodos clássicos no cenário binário

Modelo	Precisão (%)	Revocação (%)	F1 Ingestão (%)
AdaBoost	61.99	47.00	53.47
Decision Tree	55.27	53.60	54.43
Random Forest	83.08	51.12	63.29
k-NN	46.25	36.74	40.95
Naïve Bayes	25.52	93.86	40.13
SVM	75.38	81.06	78.12

A análise da Tabela 6.2 revela que a máquina de vetor de suporte (SVM) apresentou o melhor F1-score (78,12%). Esse resultado o estabelece como o *baseline* mais generalista dentre os modelos clássicos. Outros modelos, como Florestas Aleatórias, também mostraram boa precisão (83,08%), mas com uma revocação inferior. Um caso notável é o *Naïve Bayes*, que, apesar de obter a maior revocação (93,86%), teve uma precisão extremamente baixa (25,52%), indicando uma alta taxa de falsos positivos que o torna inadequado para uma aplicação prática confiável. Com base nesta análise, o desempenho do SVM será o principal ponto de referência para a avaliação dos modelos de aprendizado profundo.

Tabela 6.3: Matriz de Confusão Média - SVM Binária ([Huang et al. \(2020\)](#))

Real \ Predita	0	1
0	291	8
1	11	35

Tabela 6.4: Matriz de Confusão Média — SVM Multiclasse (Huang et al. (2020))

Real \ Predita	0	1	2	3	4	5	6	7
0	245	5	2	0	0	0	0	0
1	16	27	2	0	0	0	0	0
2	9	5	8	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	1	1	1	0	0	0	0	0
5	1	0	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0
7	1	0	0	0	0	0	0	2

6.5 Resultados com Modelos de Aprendizado Profundo

Após a avaliação dos modelos clássicos, investigaram-se arquiteturas profundas que aprendem representações diretamente dos dados. Foram considerados o método proposto e reimplementações de modelos da literatura: CNN e CNN-LSTM Moccia et al. (2022), CNN-LSTM Senyurek et al. (2019) e TCN-MHA Wang et al. (2024).

6.5.1 Cenário Binário

A Tabela 6.5 indica que o TCN-MHA apresenta o melhor F1 para a classe 1 (93,38%), com precisão 94,18% e revocação 93,16%, caracterizando o perfil mais equilibrado entre evitar falsos positivos e reter verdadeiros positivos. A CNN Moccia et al. (2022) obtém a maior precisão de ingestão (96,98%), porém com revocação menor (87,82%), o que sugere maior conservadorismo na decisão positiva. O método proposto atinge precisão 89,69% e revocação 85,97% (F1 = 87,30%), ficando atrás do TCN-MHA no equilíbrio geral e da CNN

Tabela 6.5: Resultados modelos profundos no cenário binário

Modelo	Precisão Ingestão Média	Revocação Ingestão Média	F1 Médio Classe 1
Método Proposto	89.69	85.97	87.30
CNN (Moccia et al., 2022)	96.98	87.82	91.68
CNN-LSTM (Moccia et al., 2022)	89.61	80.41	83.04
CNN-LSTM (Senyurek et al., 2019)	72.08	66.43	66.23
TCN-MHA (Wang et al., 2024)	94.18	93.16	93.38

em precisão, mas superando o CNN-LSTM [Senyurek et al. \(2019\)](#) em todas as métricas reportadas.

No método proposto (Tabela 6.6), observam-se poucos falsos positivos (6,76) e falsos negativos moderados (9,50), com 56,58 verdadeiros positivos e 448,38 verdadeiros negativos, compatíveis com uma revocação ligeiramente inferior à precisão. Na CNN [Moccia et al. \(2022\)](#) (Tabela 6.7), os falsos positivos caem para 1,32 e os verdadeiros positivos são 43,04, com 5,48 falsos negativos; esse arranjo explica a precisão máxima, à custa de uma revocação menor que a do TCN-MHA. O CNN-LSTM reimplementado de [Moccia et al. \(2022\)](#) (Tabela 6.8) apresenta aumento simultâneo de falsos positivos (4,18) e de falsos negativos (9,30), refletindo em quedas de precisão (89,61%) e revocação (80,41%). Já o TCN-MHA (Tabela 6.10) combina falsos positivos muito baixos (1,86) com falsos negativos também reduzidos (1,74), o que se traduz na melhor F1 do cenário e em diagonal de maior pureza entre todos os modelos binários.

Tabela 6.6: Matriz de confusão Média método proposto

Real \ Predita	0	1
0	448	6
1	9	56

Tabela 6.7: Matriz de confusão Média [Moccia et al. \(2022\)](#) CNN

Real \ Predita	0	1
0	418	1
1	5	43

Tabela 6.8: Matriz de confusão Média [Moccia et al. \(2022\)](#) CNN_LSTM

Real \ Predita	0	1
0	415	4
1	9	39

6.5.2 Cenário Multiclasse

Na Tabela 6.11, a CNN [Moccia et al. \(2022\)](#) alcança a maior precisão média de ingestão (93,96%), enquanto o TCN-MHA obtém a maior revocação média (94,96%) e o maior F1 (92,89%), mantendo o melhor compromisso entre acurácia por classe e cobertura dos eventos de interesse. O método proposto registra F1 de 87,36%, com precisão 88,73% e revocação 87,11%, situando-se abaixo dos dois referenciais anteriores, porém à frente do CNN-LSTM [Senyurek et al. \(2019\)](#) em F1. Este último apresenta combinação atípica de precisão baixa (59,46%) e revocação alta (84,98%), sugerindo excesso de positivos preditos e, portanto, maior incidência de falsos positivos.

Tabela 6.9: Matriz de Confusão Média — [Senyurek et al. \(2019\)](#) CNN_LSTM

Real \ Predita	0	1
0	409	11
1	15	32

Tabela 6.10: Matriz de confusão Média [Wang et al. \(2024\)](#) TCN_MHA

Real \ Predita	0	1
0	341	1
1	1	33

As matrizes de confusão médias detalham onde ocorrem os erros. Para o método proposto (Tabela 6.12), a classe 0 domina o volume amostral e mantém diagonal elevada (378), com erros residuais dispersos principalmente para as classes 1 e 2; a classe 1 retém boa parte dos acertos (57), mas ainda sofre confusão para a classe 0, consistente com as métricas agregadas. No CNN-LSTM de [Moccia et al. \(2022\)](#) (Tabela 6.13), a diagonal permanece razoavelmente alta nas classes majoritárias (por exemplo, 319 para a classe 0 e 44 para a classe 1), mas há dispersões adicionais em classes menos frequentes, o que reduz o F1 frente à CNN pura. No CNN-LSTM de [Senyurek et al. \(2019\)](#) (Tabela 6.14), observa-se confusão acentuada da classe 2 para a classe 1 (15 fora da diagonal), com verdadeiro positivo reduzido na própria classe 2 (4), padrão compatível com a precisão baixa e a revocação alta da Tabela 6.11. Por fim, o TCN-MHA (Tabela 6.15) exibe as menores confusões fora da diagonal de modo amplo, com valores residuais baixos entre pares de classes e diagonais mais limpas, justificando o F1 máximo entre os modelos multiclasse.

Tabela 6.11: Desempenho dos modelos profundos no cenário multiclasse

Modelo	Precisão Ingestão Média	Revocação Ingestão Média	F1 Médio Classe 1
Método proposto	88.73	87.11	87.36
CNN (Moccia et al., 2022)	93.96	92.19	92.88
CNN-LSTM (Moccia et al., 2022)	90.68	90.24	89.82
CNN-LSTM (Senyurek et al., 2019)	59.46	84.98	69.22
TCN-MHA Wang et al. (2024)	91.45	94.96	92.89

Tabela 6.12: Matriz de Confusão Média — Método Proposto Multiclasse

Real \ Predita	0	1	2	3	4	5	6	7
0	378	5	4	0	1	1	0	1
1	6	57	1	0	0	0	0	0
2	5	1	24	0	0	0	0	0
3	1	0	0	2	0	0	0	0
4	2	0	0	0	1	0	0	0
5	1	0	1	0	0	1	0	0
6	1	0	1	0	0	0	0	0
7	1	0	0	0	0	0	0	4

Tabela 6.13: Matriz de Confusão Média — [Moccia et al. \(2022\)](#) CNN_LSTM Multiclasse

Real \ Predita	0	1	2	3	4	5	6	7
0	319	1	1	0	0	0	0	0
1	2	44	1	0	0	0	0	0
2	2	2	20	0	0	0	0	0
3	0	0	1	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	1	0	0	1	0	0
6	0	0	1	0	0	0	0	0
7	1	0	0	0	0	0	0	2

Tabela 6.14: Matriz de Confusão Média — [Senyurek et al. \(2019\)](#) CNN_LSTM Multiclasse

Real \ Predita	0	1	2	3	4	5	6	7
0	315	6	1	0	0	0	0	0
1	6	41	0	0	0	0	0	0
2	5	15	4	0	0	0	0	0
3	1	0	1	0	0	0	0	0
4	1	0	1	0	0	0	0	0
5	1	1	1	0	0	0	0	0
6	1	1	1	0	0	0	0	0
7	0	2	0	0	0	0	0	0

Tabela 6.15: Matriz de Confusão Média — Wang et al. (2024) TCN_MHA Multiclasse

Real \ Predita	0	1	2	3	4	5	6	7
0	320	1	0	0	0	0	0	0
1	0	46	1	0	0	0	0	0
2	1	2	19	0	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	1	0	0	1	0	0
6	0	0	1	0	0	0	0	0
7	0	0	0	0	0	0	0	2

6.6 Discussão dos Resultados

Três achados principais emergem das Tabelas 6.5 e 6.11, em conjunto com as matrizes de confusão binárias (Tabelas 6.6, 6.7, 6.8, 6.9, 6.10) e multiclasse (Tabelas 6.12, 6.13, 6.14, 6.15).

(i) Em relação ao patamar de referência de métodos clássicos, os modelos profundos superam o F1 binário da SVM (78,12%) com margens observáveis: no cenário binário, TCN-MHA atinge F1 de 93,38% e a CNN reimplementada de Moccia et al. (2022) alcança F1 de 91,68% (Tabela 6.5). No multiclasse, TCN-MHA registra F1 de 92,89% e a CNN 92,88% (Tabela 6.11).

(ii) Entre redes profundas, há diferenças consistentes entre precisão e revocação e seus reflexos nas matrizes de confusão. No binário, a CNN de Moccia et al. (2022) maximiza a precisão de Ingestão (96,98%) com revocação de 87,82% (Tabela 6.5); a Tabela 6.7 mostra falsos positivos de 1,32 e falsos negativos de 5,48. Já o TCN-MHA combina precisão 94,18% e revocação 93,16% com falsos positivos de 1,86 e falsos negativos de 1,74 (Tabela 6.10), o que explica o F1 mais alto no binário. No multiclasse, a matriz do TCN-MHA (Tabela 6.15) apresenta confusões fora da diagonal menores e mais distribuídas em valores baixos, enquanto a CNN-LSTM de Senyurek et al. (2019) (Tabela 6.14) concentra confusões específicas (por exemplo, da classe 2 para a classe 1), o que se reflete na combinação de precisão 59,46% e revocação 84,98% (Tabela 6.11).

(iii) Para a classe Ingestão (classe 1), a diferença entre o método proposto e o melhor modelo é de 6,08 pontos de F1 no binário (87,30% vs. 93,38%) e de 5,53 pontos no multiclasse (87,36% vs. 92,89%) — ver Tabelas 6.5 e 6.11. Nas matrizes de confusão, o método proposto apresenta, no binário, 6,76 falsos positivos e 9,50 falsos negativos (Tabela 6.6), enquanto o TCN-MHA reduz ambos (1 FP; 1 FN na Tabela 6.10). No multiclasse, a Tabela 6.12 indica dispersão moderada da classe 1 para a classe 0 no método proposto, ao passo que a Tabela 6.15 mostra valores residuais menores fora da diagonal para o TCN-MHA. Esses padrões quantificam a diferença de desempenho na classe de interesse.

6.7 Considerações Finais

Com o protocolo LOSO, reamostragem a 50,Hz, normalização por eixo no treino e janelas de 3,s com voto majoritário, os resultados indicam: (a) no binário, TCN-MHA com F1 de 93,38% e CNN de Moccia et al. (2022) com precisão de 96,98% (Tabela 6.5); (b) no multiclasse, TCN-MHA com F1 de 92,89% e CNN com 92,88% (Tabela 6.11). As matrizes de confusão associadas (Tabelas 6.6–6.10 e 6.12–6.15) mostram menor confusão fora da diagonal para o TCN-MHA em ambos os cenários. A CNN 1D proposta apresenta F1 de 87,30% (binário) e 87,36% (multiclasse), com relação precisão–revocação mais próxima do

equilíbrio no binário (89,69% vs. 85,97%) e padrão de confusões coerente nas Tabelas 6.6 e 6.12. Esses resultados, em conjunto com a disponibilização do conjunto de dados e do protocolo de avaliação, permitem reproduzir e comparar melhorias futuras mantendo as mesmas métricas e referências de erro por classe.

CONCLUSÃO

7.1 Considerações Finais

Este trabalho teve dois objetivos centrais: (i) projetar, coletar e documentar um conjunto de dados específico para reconhecimento de ingestão por IMU de punho, com anotações temporais, metadados completos e protocolo de avaliação LOSO por sujeito, de modo a viabilizar comparações justas e reprodutíveis nos cenários binário e multiclasse; (ii) propor e avaliar um classificador leve baseado em CNN 1D operando sobre janelas fixas (após re-amostragem e normalização) e comparar seu desempenho, sob o mesmo fluxo, com modelos de referência da literatura (CNN, CNN-LSTM e TCN com atenção). Esses objetivos respondem às lacunas de escassez de bases específicas e de métodos compatíveis com vestíveis, e foram aferidos por métricas por classe (com ênfase em F1 de Ingestão) e pela análise de generalização intersujeitos em LOSO.

7.2 Principais Contribuições

A primeira contribuição é a consolidação do InGesture¹, uma base de dados para ingestão com sensores inerciais de pulso, cobrindo gestos de ingestão e gestos análogos, além de segmentos ociosos. A coleta foi realizada com um IMU no punho dominante, com amostragem original de 200 Hz, e anotações temporais em nível de evento feitas durante as sessões. O conjunto contempla 50 participantes e 8 classes (detalhadas no capítulo de dados), permitindo análises nos cenários binário (ingestão vs. não ingestão) e multiclasse. O desenho privilegia privacidade (ausência de áudio/vídeo) e portabilidade para aplicações em vestíveis. Cada sessão possui arquivos de sinais e de anotações sincronizados, com

¹<https://data.mendeley.com/datasets/fdxst56tcj/5>

identificação de participante e metadados essenciais (sensor, posição, frequência), o que viabiliza auditoria, reuso e extensões futuras.

A segunda contribuição é o desenvolvimento e a avaliação de uma arquitetura convolucional unidimensional (CNN 1D) otimizada para reconhecimento de ingestão em dados inerciais. O modelo foi projetado para operar sobre janelas fixas de 3 s reamostradas a 50 Hz, priorizando baixa complexidade computacional e viabilidade de execução em dispositivos de pulso. O classificador utiliza camadas convolucionais de pequena profundidade para extração de padrões locais de movimento e uma camada totalmente conectada para decisão final, mantendo compatibilidade com o fluxo de pré-processamento do protocolo LOSO. Os resultados obtidos nos cenários binário e multiclasse demonstram desempenho competitivo em relação a modelos mais complexos da literatura (CNN-LSTM e TCN-MHA), com balanço estável entre precisão e revocação e baixo custo de processamento, caracterizando uma alternativa eficiente para reconhecimento embarcado de gestos de ingestão.

7.3 Dificuldades e Limitações Encontradas

A principal dificuldade veio da variabilidade inter-sujeitos: diferenças de amplitude, ritmo e micro-padrões de movimento impactam a estabilidade da decisão em LOSO, especialmente no multiclasse. A rotulagem por janelamento introduz casos de fronteira com mistura de classes dentro da mesma janela, elevando a taxa de erros quando a evidência do gesto não ocupa a maior parte do intervalo; embora o voto majoritário reduza oscilações, ele pode penalizar eventos curtos ou deslocados. A opção por janelas de comprimento fixo e campo receptivo limitado na CNN também restringe a captura de dependências temporais mais longas, o que favorece modelos com dilatações ou atenção em situações de transição lenta entre fases do gesto. Do ponto de vista dos dados, mesmo com número expressivo de participantes, permanece a lacuna para contexto de vida livre: o protocolo controlado mitiga ruído de fundo e confusores ambientais; a generalização para situações não controladas (variação de utensílios, posturas, deslocamentos) exigirá validação adicional. Além disso, o estudo não explorou filtragem temporal (p.ex., passa-baixas, denoising). Por fim, utilizou-se apenas IMU de punho; sem sinais complementares (PPG, EMG ou múltiplos posicionamentos), algumas ambiguidades entre ingestão e gestos análogos permanecem difíceis de resolver.

7.4 Direções Futuras

Os resultados apontam uma trilha incremental e prática para reduzir terreno em relação ao estado da arte sem comprometer a leveza do método. No nível de modelagem, duas extensões são imediatas: (i) introduzir convoluções dilatadas em poucas camadas para

ampliar o campo receptivo efetivo; e (ii) em pré-processamento conduzir uma avaliação sistemática de filtros (passa-baixas, denoising por wavelets) e aumentações temporais (jitter, time-warp, scaling) para robustez a ruído e variabilidade inter-sujeitos. Em anotação e rotulagem, investigar critérios alternativos ao voto majoritário (como rótulo pelo centro da janela, esquemas ponderados pelo tempo de permanência ou janelas de tamanho variável sincronizadas com eventos) pode reduzir o viés em fronteiras. Em avaliação, além de métricas por janela, incorporar medidas por evento tornará a análise mais alinhada ao uso real. Em generalização, explorar adaptação por sujeito, auto/semi-supervisão para pré-treino com maior volume de dados não rotulados e testes em condição de vida livre fortalecerão a validade externa. Por fim, em implantação, medir latência, consumo de energia e memória em relógios comerciais permitirá orientar escolhas de arquitetura e parametrização sob restrições reais, mantendo a proposta dentro do escopo de uso em vestíveis.

Referências Bibliográficas

- Altun, K. e Barshan, B. (2010). Human activity recognition using inertial/magnetic sensor units. Em *Human Behavior Understanding*, páginas 38–51. Springer. [9](#), [10](#)
- Anderez, D. O., Lotfi, A., e Pourabdollah, A. (2020). Eating and drinking gesture spotting and recognition using a novel adaptive segmentation technique and a gesture discrepancy measure. *Expert Systems with Applications*, 140. [5](#)
- Banos, O., Galvez, J. M., Damas, M., Pomares, H., e Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499. [10](#)
- Bao, L. e Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. Em *Pervasive computing*, páginas 1–17. Springer. [10](#)
- Bendat, J. S. e Piersol, A. G. (2010). *Random Data: Analysis and Measurement Procedures*. John Wiley & Sons, 4th^a edição. [8](#)
- Bishop, C. M. (2006). Pattern recognition and machine learning. *Springer google schola*, 2:1122–1128. [6](#), [7](#), [9](#), [12](#)
- Bleser, G., Steffen, D., Reiss, A., Weber, M., Hendeby, G., e Fradet, L. (2015). Personalized physical activity monitoring using wearable sensors. Em *Smart health: Open problems and future challenges*, páginas 99–124. Springer. [2](#)
- Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. d. R., e Roggen, D. (2013). The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042. [2](#)
- Cohen, R., Fernie, G., e Fekr, A. R. (2021). Fluid intake monitoring systems for the elderly: A review of the literature. [5](#)
- Cohen, R., Fernie, G., e Roshan Fekr, A. (2022). Monitoring fluid intake by commercially available smart water bottles. *Scientific Reports*, 12(1):4402. [2](#)

- Edmonds, C. J., Gandy, J., e Hooper, L. (2021). Dehydration in older people: A systematic review of the effects of dehydration on health outcomes, cognitive performance and daily living. *Maturitas*, 151:40–48. [1](#)
- EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) (2017). Scientific opinion on dietary reference values for water. Relatório Técnico e15121, European Food Safety Authority. [1](#)
- Gomes, D. e Sousa, I. (2019). Real-time drink trigger detection in free-living conditions using inertial sensors. *Sensors (Switzerland)*, 19. [7](#), [5](#), [6](#), [11](#), [12](#)
- Hamatani, T., Elhamshary, M., Uchiyama, A., e Higashino, T. (2018). Fluidmeter: Gauging the human daily fluid intake using smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2. [7](#), [5](#), [6](#), [11](#), [12](#)
- Hamilton, J. D. J. D. (1994). *Time series analysis*. Princeton University Press. [7](#)
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83. Section on smoothing techniques, pages 65–70. [8](#)
- Huang, H. Y., Hsieh, C. Y., Liu, K. C., Hsu, S. J. P., e Chan, C. T. (2020). Fluid intake monitoring system using a wearable inertial sensor for fluid intake management. *Sensors (Switzerland)*, 20:1–17. [7](#), [8](#), [2](#), [5](#), [6](#), [14](#), [15](#), [16](#), [33](#), [35](#), [36](#)
- Kwapisz, J. R., Weiss, G. M., e Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2):74–82. [9](#), [10](#)
- Kyritsis, K., Diou, C., e Delopoulos, A. (2021). Wrist-mounted IMU data towards the investigation of free-living human eating behavior - the Free- living Food Intake Cycle (FreeFIC) dataset. [21](#)
- Lara, O. D. e Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials*, 15(3):1192–1209. [10](#)
- Maurer, U., Smailagic, A., Siewiorek, D. P., e Deisher, M. (2006). Activity recognition and monitoring using multiple sensors on different body positions. Em *International Workshop on Wearable and Implantable Body Sensor Networks*, páginas 113–116. IEEE. [9](#)
- Moccia, S., Solbiati, S., Khornehah, M., Bossi, F. F., e Caiani, E. G. (2022). Automated classification of hand gestures using a wristband and machine learning for possible application in pill intake monitoring. *Computer Methods and Programs in Biomedicine*, 219:106753. [8](#), [14](#), [15](#), [16](#), [23](#), [24](#), [25](#), [26](#), [27](#), [28](#), [36](#), [37](#), [38](#), [39](#), [41](#)
- Murphy, K. P. (2012). *Machine learning : a probabilistic perspective*. [6](#)

- Nishi, N., Otsuka, R., Tange, C., Nishita, Y., Kato, Y., Ando, F., e Shimokata, H. (2023). Water intake, hydration status and 2-year changes in cognitive function among community-dwelling older adults. *BMC Medicine*, 21:262. [1](#)
- Popkin, B. M., D'Anci, K. E., e Rosenberg, I. H. (2010). Water, hydration, and health. *Nutrition reviews*, 68(8):439–458. [1](#)
- Rogerson, G., Hewitt, S., Oliver, S., Maughan, R. J., e Betts, J. A. (2023). A systematic review of measurement methods used to assess fluid intake in free-living adults. *Journal of Human Nutrition and Dietetics*, 36(5):951–965. [1](#)
- Ronao, C. A. e Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, 59:235–244. [11](#)
- Rosinger, A. Y., Young, S. L., Dickinson, F., e Brewis, A. (2023). Water needs, water insecurity, and human biology. *Annual Review of Anthropology*, 52(1):359–375. [1](#)
- Roy, S., Paul, S., e Lee, S. (2022). Smart water bottle for monitoring daily water intake: Sensors, challenges, and future directions. *Sensors*, 22(7):2514. [2](#)
- Senyurek, V., Imtiaz, M., e Hassan, N. (2019). Detection of drinking via inertial sensor. Em *Presented at the 6th International Electronic Conference on Sensors and Applications*, volume 15, página 30. [8](#), [14](#), [15](#), [16](#), [36](#), [37](#), [38](#), [39](#), [41](#)
- Shen, Y., Salley, J., Muth, E., e Hoover, A. (2016). Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables. *IEEE journal of biomedical and health informatics*, 21(3):599–606. [21](#)
- Singh, R., Patel, M., Wong, K., Martinez, L., e Kim, J.-H. (2025). Minimum days estimation for reliable dietary intake: Implications for water and fluid consumption studies. *European Journal of Clinical Nutrition*, 79(2):301–310. [1](#)
- Wang, C., Kumar, T. S., De Raedt, W., Camps, G., Hallez, H., e Vanrumste, B. (2024). Eating speed measurement using wrist-worn imu sensors towards free-living environments. *IEEE Journal of Biomedical and Health Informatics*, 28(10):5816–5828. [8](#), [14](#), [15](#), [16](#), [21](#), [23](#), [24](#), [25](#), [26](#), [27](#), [28](#), [36](#), [38](#), [40](#)
- Wang, C., Kumar, T. S., Raedt, W. D., Camps, G., Hallez, H., e Vanrumste, B. (2022). Drinking gesture detection using wrist-worn imu sensors with multi-stage temporal convolutional network in free-living environments. volume 2022-July, páginas 1778–1782. Institute of Electrical and Electronics Engineers Inc. [2](#), [5](#), [7](#), [9](#), [12](#)
- Wei, B., Zhang, S., Diao, X., Xu, Q., Gao, Y., e Alshurafa, N. (2023). An end-to-end energy-efficient approach for intake detection with low inference time using wrist-worn sensor. *IEEE Journal of Biomedical and Health Informatics*, 27:3878–3888. [5](#)

- Wellnitz, A., Wolff, J.-P., e Kirste, C. H. T. (2019). Fluid intake recognition using inertial sensors. [5](#)
- Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. r.i.p. *The American Statistician*, 68(3):191–195. [10](#)
- Zhang, N., Du, S., Tang, Z., Zheng, M., Yan, R., Zhu, Y., e Ma, G. (2019). Effects of dehydration and rehydration on cognitive performance and mood among male college students in cangzhou, china: A self-controlled trial. *International Journal of Environmental Research and Public Health*, 16(11):1891. [1](#)