

# Modeling complex human activities with inertial units, ambient sensors, and machine learning

Laura Sthefany Colombo<sup>1</sup>, André Luiz da Silva Conde<sup>1</sup>,  
Leonardo Tórtoro Pereira<sup>1</sup>, Caetano Mazzoni Ranieri<sup>1</sup>

<sup>1</sup>Instituto de Geociências e Ciências Exatas (IGCE)  
Universidade Estadual Paulista (UNESP)  
Rio Claro, SP, Brasil

{laura.colombo; andre.conde; leonardo.t.pereira; cm.ranieri}@unesp.br

**Abstract.** *Human activity recognition is key in applications such as healthcare, sports, and smart environments. Recent studies apply both classical algorithms and deep learning models. However, few address the impact of similar and complex activity patterns on model performance. This work compares Decision Trees, Random Forests, and a 1D Convolutional Neural Network (CNN-1D) using two public datasets: PAMAP2 and HWU-USP. On PAMAP2, the CNN-1D proved most robust ( $62.48\% \pm 10.31\%$  accuracy), while on HWU-USP the Random Forest was more stable ( $38.93\% \pm 3.94\%$  accuracy), highlighting the importance of hyperparameter tuning for complex tasks.*

**Resumo.** *O reconhecimento de atividades humanas é essencial para aplicações em saúde, esportes e ambientes inteligentes. Trabalhos recentes utilizam desde modelos clássicos até redes neurais profundas. Contudo, poucos exploram a influência da similaridade entre atividades complexas na acurácia dos modelos. Este trabalho compara árvores de decisão, florestas aleatórias e uma rede neural convolucional unidimensional (CNN-1D) em dois conjuntos de dados públicos: PAMAP2 e HWU-USP. No PAMAP2, a CNN-1D mostrou-se mais robusta (acurácia de  $62,48\% \pm 10,31\%$ ), enquanto no HWU-USP o modelo Floresta Aleatória foi mais estável (acurácia de  $38,93\% \pm 3,94\%$ ), destacando a importância do ajuste de hiperparâmetros para tarefas complexas.*

## 1. Introdução

Promover segurança e qualidade de vida por meio de soluções computacionais é o objetivo de sistemas para assistência à autonomia em domicílio (*Ambient Assisted Living* – AAL). Tais sistemas são especialmente relevantes para populações vulneráveis, como idosos e pessoas com mobilidade reduzida. Neste contexto, o reconhecimento de atividade humana (*Human Activity Recognition* – HAR) é uma tecnologia fundamental, uma vez que permite identificar o comportamento dos ocupantes de um ambiente inteligente (aquele com sistemas embarcados e tecnologias de informação e comunicação, para melhorar a experiência dos usuários) e adaptar o comportamento de seus componentes interativos e assistivos [Cicirelli et al. 2021, Jovanovic et al. 2022, Marques et al. 2021].

Para viabilizar a implementação de algoritmos para reconhecimento de atividades, pode-se recorrer à utilização de sensores, seja em dispositivos vestíveis (*wearables*) ou integrados ao ambiente via Internet das Coisas (IoT) [Marques et al. 2021]. A busca por

uma compreensão mais detalhada das atividades levou à adoção de sistemas multimodais, que consistem no uso de diferentes tipos de dados, provenientes de diferentes tipos de sensores (e.g., aceleração, sensores de contato). Embora essa abordagem enriqueça a análise, ela também introduz desafios em relação à complexidade do processamento de dados e à modelagem [Yadav et al. 2021].

Este trabalho apresenta uma análise comparativa do desempenho dos modelos de aprendizado de máquina: Árvore de Decisão (DT, do inglês *Decision Tree*), Floresta Aleatória (RF, do inglês *Random Forest*), e Rede Neural Convolutiva (CNN, do inglês: *Convolutional Neural Network*) unidimensional (1D), em duas bases de dados públicas de HAR com características distintas: (i) HWU-USP [Ranieri et al. 2021], focada em atividades domésticas complexas, e (ii) PAMAP2 [Reiss and Stricker 2012], centrada na execução de exercícios físicos. São comparadas a acurácia e a matriz de confusão dos modelos. A RF obteve melhor acurácia na base PAMAP2, após achatamento e balanceamento dos dados, enquanto a CNN-1D obteve o melhor resultado na base HWU-USP, após balanceamento de dados e otimização bayesiana.

Observa-se a limitação dos métodos tradicionais de aprendizado de máquina e ressalta-se a necessidade de *pipelines* mais avançados para o tratamento adequado do problema da modelagem de atividades humanas complexas com este conjunto de sensores. Sugere-se o uso de *feature engineering*, *data augmentation* e arquiteturas baseadas em atenção, e a fusão de dados de diferentes sensores.

## 2. Revisão Bibliográfica

Os métodos para reconhecimento de atividade humana dependem diretamente da modalidade ou do conjunto de modalidades escolhidas para inferir com maior precisão qual atividade está sendo realizada. Isso ocorre porque diferentes modalidades de sensoramento capturam aspectos variados das atividades humanas, cada uma com vantagens e limitações específicas. Em geral, os sensores podem ser vestíveis ou instalados no ambiente, registrando o comportamento do indivíduo observado [Ranieri et al. 2021]. Por exemplo, sensores de movimento, como acelerômetros e giroscópios, são utilizados para captar movimentos físicos detalhados. Por sua vez, câmeras fornecem informações visuais ricas sobre posturas e gestos, enquanto microfones capturam dados auditivos úteis para atividades que envolvem sons característicos, como digitação ou passos.

Diversas abordagens se propuseram a analisar o desempenho de modelos de aprendizado de máquina para reconhecimento de atividades humanas. [Alagoz 2024] seguiu uma metodologia estruturada que envolveu coleta de dados de 30 voluntários realizando seis atividades distintas, capturadas por sensores de *smartphones* a 50 Hz. Após o pré-processamento dos sinais, que incluiu filtragem, segmentação e extração de características no domínio do tempo e frequência, os modelos XGBoost e MiniRocket [Dempster et al. 2021] foram treinados e avaliados utilizando validação cruzada de Monte Carlo. Os resultados evidenciaram o XGBoost com desempenho superior em tempo de treinamento e alta precisão, enquanto o MiniRocket destacou-se pela eficiência com dados brutos, sugerindo sua adequação para cenários com menos pré-processamento. Além disso, o estudo aponta que a fusão de sensores pode potencializar ainda mais o reconhecimento de atividades, contribuindo para inovações futuras na área.

[Guo et al. 2024] investigaram a fusão de sinais Wi-Fi e de sensores inerciais para

melhorar o HAR em residências inteligentes. Eles propuseram um sistema que combina informações de estado de canal Wi-Fi (CSI) com dados de acelerômetros e giroscópios de unidades inerciais, implementando seis algoritmos de aprendizado de máquina (SVM, MLP, árvore de decisão, floresta aleatória, regressão logística e k-NN) para reconhecer oito atividades diárias. Os experimentos demonstraram que a combinação CSI+IMU atingiu a maior acurácia, com o classificador SVM apresentando o melhor desempenho.

[Manouchehri and Bouguila 2023] propuseram um modelo generativo não supervisionado para HAR baseado em modelos ocultos de Markov (*Hidden Markov Models* – HMM). O método utiliza distribuições de Dirichlet escalonadas como funções de emissão no HMM, visando capturar padrões temporais sequenciais dos dados de atividades humanas. Os autores empregaram inferência variacional para treinar o modelo e avaliaram-no usando as bases de dados OPPORTUNITY e UCI HAR, demonstrando a viabilidade de reconhecer atividades diárias utilizando modelos estatísticos de séries temporais e destacando o potencial dos HMMs em aplicações de saúde e assistência.

[D’Arco et al. 2022] desenvolveram um sistema de HAR baseado em palmilhas inteligentes, avaliando o impacto de segmentação, tipos de sensores e seleção de atributos no desempenho. Nesse trabalho, um classificador SVM foi utilizado para reconhecer seis atividades de locomoção (e.g., subir escadas, sentar/levantar, caminhar), após otimização da janela deslizante (10 segundos com 50% de sobreposição) e redução de características de 272 para 227 atributos. Os resultados indicaram que sensores inerciais são mais confiáveis para detectar atividades dinâmicas, enquanto sensores de pressão são melhores para atividades estacionárias. Contudo, a combinação de ambos forneceu a maior acurácia global no reconhecimento das atividades analisadas, evidenciando os benefícios de abordagens multimodais com métodos tradicionais de aprendizado.

Observa-se que a maioria dos estudos em HAR com aprendizado de máquina tradicional ainda se concentra em modalidades isoladas de sensoriamento, sobretudo dados inerciais provenientes de dispositivos vestíveis. Abordagens multimodais, quando presentes, não exploram de forma sistemática a integração de sensores ambientais em cenários residenciais inteligentes. Este trabalho consiste em uma análise comparativa que avalie como diferentes técnicas se comportam diante da combinação de dados inerciais e sensores integrados ao ambiente, com vistas a identificar os modelos mais promissores para aplicações assistivas em ambientes inteligentes.

### **3. Metodologia**

De modo a observar a capacidade de predição de diferentes modelos tradicionais de aprendizado de máquina nas bases de dados de atividades físicas, o seguinte projeto de estudo foi idealizado: o pré-processamento dos dados, que inclui a segmentação em janelas das séries temporais presentes nos dados e o balanceamento das classes; implementação dos modelos e a definição de hiperparâmetros via testes preliminares ao estudo; e a execução da divisão em treino e teste para comparação da acurácia, *f1-score*, precisão e *recall*.

#### **3.1. Bases de Dados**

Existem diversas bases de dados públicas para avaliar os métodos de reconhecimento de atividades humanas, contemplando diferentes cenários e requisitos. Para os objetivos deste trabalho, foram adotadas as bases de dados HWU-USP Activities Dataset e PA-MAP2.

A primeira, HWU-USP [Ranieri et al. 2021], que contém dados de atividades complexas em um ambiente doméstico. O conjunto de dados original inclui múltiplos sensores: uma câmera RGB-D, duas Unidades de Medição Inercial (*Inertial Measurement Units* – IMUs) no pulso e cintura, e sensores ambientais (contato e movimento). Os dados foram coletados de 16 participantes realizando nove atividades de vida diária, como *fazendo uma xícara de chá e arrumando a cozinha*. Para este estudo, foram selecionados somente os dados dos sensores ambientais e das IMUs.

A segunda base foi a PAMAP2 (Physical Activity Monitoring) [Reiss and Stricker 2012], um benchmark para o reconhecimento de atividades físicas. A PAMAP2 utiliza exclusivamente sensores vestíveis: três IMUs (peito, pulso e tornozelo) e um monitor de frequência cardíaca, coletados de 9 participantes realizando 12 atividades principais, como *walking*, *running* e *ascending stairs*. A atividade com ID 0 (atividades transitórias) foi removida, conforme prática comum.

### 3.2. Pré-processamento de Dados

Para adequar os dados aos modelos, as séries temporais brutas foram segmentadas em janelas, segundo a abordagem de *sliding window*. Os parâmetros foram ajustados para cada base, conforme testes preliminares: para a HWU-USP, utilizou-se uma janela de 50 amostras com passo de 3, enquanto para a PAMAP2, uma janela de 80 amostras com passo de 20. Somente janelas contendo uma única atividade (rótulo puro) foram mantidas.

Para garantir uma avaliação adequada da capacidade de generalização dos modelos, a divisão entre treino e teste foi realizada com base no identificador dos participantes, evitando vazamento de dados.

- **HWU-USP:** 10 participantes para treino e 6 para teste.
- **PAMAP2:** 7 participantes para treino e 2 para teste.

Dado o desbalanceamento de classes em ambas as bases, aplicou-se uma estratégia de balanceamento durante o treinamento, utilizando pesos de classe (`class_weight='balanced'`). Essa técnica instrui os modelos a dar maior importância aos erros nas classes minoritárias.

### 3.3. Implementação dos Modelos

Para a classificação das atividades, foram implementados três modelos distintos: uma Árvore de Decisão (Decision Tree), uma Floresta Aleatória (Random Forest) e uma Rede Neural Convolutiva (CNN) unidimensional. Os modelos clássicos foram implementados com a biblioteca Scikit-learn (utilizando a semente aleatória 42 para os testes), enquanto a CNN foi construída com TensorFlow/Keras. Os códigos utilizados para implementação estão disponíveis publicamente<sup>1</sup>.

#### 3.3.1. Rede Neural Convolutiva (CNN 1D)

Para a classificação das séries temporais, foi implementada uma Rede Neural Convolutiva unidimensional (CNN 1D) utilizando a biblioteca TensorFlow/Keras. A

---

<sup>1</sup><https://drive.google.com/drive/folders/1kuUmsJY7cc6BZE1iKWF4mEzDH1E4Pomf?usp=sharing>

arquitetura da rede foi projetada para capturar padrões temporais locais nos dados dos sensores, com implementações específicas para cada base de dados.

Para a base HWU-USP, a abordagem foi dividida em duas etapas. Primeiramente, um *modelo base* foi construído com duas camadas convolucionais (Conv1D) e uma camada densa (Dense). Em seguida, para refinar o modelo, empregou-se uma *Otimização Bayesiana* com o Keras Tuner [O'Malley et al. 2019] para encontrar os melhores hiperparâmetros. Foram otimizados os seguintes parâmetros:

- O número de filtros nas camadas convolucionais (variando de 32 a 256).
- O número de neurônios na camada densa (de 128 a 512).
- A taxa de Dropout (de 0,2 a 0,5).
- A taxa de aprendizado do otimizador Adam ([1e-2, 1e-3, 1e-4]).

O modelo final utilizado nos experimentos foi então treinado com os melhores hiperparâmetros encontrados pela busca: 96 filtros nas camadas convolucionais, 352 neurônios na camada densa, taxa de *dropout* de 0.2 e taxa de aprendizado de 0.001.

Para a base PAMAP2, foi adotada uma arquitetura de CNN 1D com foco em regularização. A rede incluiu camadas de BatchNormalization após as convoluções para estabilizar o treinamento, e utilizou GlobalAveragePooling1D em vez de Flatten, uma técnica que reduz o número de parâmetros e ajuda a prevenir o sobreajuste. Além disso, foi aplicada regularização L2 (com fator 0,005) em todas as camadas convolucionais e densas.

Ambos os modelos finais de CNN foram compilados com o otimizador Adam e a função de perda `sparse_categorical_crossentropy`. O treinamento utilizou os pesos de classe para lidar com o desbalanceamento dos dados. Para prevenir o sobreajuste, foi empregado um *callback* de `EarlyStopping`, que monitorou a acurácia de validação (`val_accuracy`) para a HWU-USP e a perda de validação (`val_loss`) para a PAMAP2, com paciência de 3 e 5 épocas, respectivamente.

### 3.3.2. Floresta Aleatória (Random Forest)

Para a classificação com um modelo de aprendizado de máquina clássico, foi utilizado o algoritmo Floresta Aleatória. Uma etapa de pré-processamento foi necessária para adequar os dados de entrada. Como os dados foram janelados em um formato 3D (*amostras, passos de tempo e características*), eles foram remodelados (achatados) para o formato 2D (*amostras e características*) exigido pelo classificador do Scikit-learn. Isso foi feito concatenando todas as características de todos os passos de tempo de cada janela em um único vetor.

Para a base HWU-USP, a busca pelos melhores hiperparâmetros foi realizada utilizando uma busca aleatória com validação cruzada (`RandomizedSearchCV`). Os seguintes parâmetros foram explorados:

- `n_estimators`: O número de árvores na floresta [100, 200, 300].
- `max_depth`: A profundidade máxima das árvores [10, 20, 30, None].
- `min_samples_split`: O número mínimo de amostras para dividir um nó [2, 5, 10].
- `min_samples_leaf`: O número mínimo de amostras em um nó folha [1, 2, 4].

- `bootstrap`: Se amostras de bootstrap são usadas para construir as árvores [True, False].

A busca foi configurada para avaliar 100 combinações de parâmetros com 5 dobras de validação cruzada (*5-fold cross-validation*). O modelo final utilizado nos experimentos foi aquele com os melhores hiperparâmetros retornados por este processo: `n_estimators=300`, `min_samples_split=5`, `min_samples_leaf=1`, `max_depth=30` e `bootstrap=True`.

Para a base PAMAP2, foi treinado um modelo de Floresta Aleatória com `n_estimators=100`, que já constitui um *baseline*. Em ambos os casos, o parâmetro `class_weight='balanced'` foi utilizado durante o treinamento para mitigar o efeito do desbalanceamento de classes presente nos dois conjuntos de dados. O parâmetro `n_jobs=-1` também foi usado para paralelizar o treinamento e utilizar todos os núcleos de processamento disponíveis.

### 3.3.3. Árvore de Decisão (Decision Tree)

Como um modelo clássico e interpretável, foi implementada uma Árvore de Decisão (`DecisionTreeClassifier`) da biblioteca `Scikit-learn`. Modelos baseados em árvores são propensos a sobreajuste se não tiverem sua complexidade controlada, por isso a metodologia também incluiu uma etapa de otimização de hiperparâmetros. Da mesma forma que o Floresta Aleatória, os dados de entrada no formato 3D (*amostras, passos de tempo, características*) foram remodelados para o formato 2D (*amostras, características*), concatenando as informações de cada janela em um único vetor.

Para a base HWU-USP, um *baseline* foi inicialmente treinado sem restrições de profundidade. Em seguida, foi realizada uma busca em grade com validação cruzada (`GridSearchCV`) para encontrar a combinação de hiperparâmetros que melhor generaliza para dados não vistos. Os seguintes parâmetros foram avaliados:

- `criterion`: O critério para medir a qualidade da divisão ['gini', 'entropy'].
- `max_depth`: A profundidade máxima da árvore [10, 20, 30, None].
- `min_samples_split`: O número mínimo de amostras necessário para dividir um nó [2, 5, 10].
- `min_samples_leaf`: O número mínimo de amostras em um nó folha [1, 2, 4].

O modelo com os melhores parâmetros foi selecionado para a comparação, utilizando: `criterion='entropy'`, `max_depth=10`, `min_samples_leaf=4` e `min_samples_split=10`.

Para a base PAMAP2, foi utilizado um modelo com o hiperparâmetro `max_depth=20` para controlar a complexidade. Em ambos os cenários, o parâmetro `class_weight='balanced'` foi empregado para mitigar o viés causado pelo desbalanceamento de classes nos conjuntos de dados.

## 3.4. Validação em Múltiplas Amostras

Para atender à solicitação de uma validação mais confiável e avaliar a generalização dos modelos, um segundo conjunto de experimentos foi conduzido. A escolha da estratégia

de amostragem foi adaptada às características de cada conjunto de dados a fim de garantir a melhor estimativa de performance.

Para a base HWU-USP, que conta com 16 participantes, optou-se pela metodologia de Bootstrap para gerar 25 amostras. Em cada iteração, os participantes do conjunto de treino foram sorteados com reposição, e os participantes não selecionados (Out-of-Bag) compuseram o conjunto de teste. Já para a base PAMAP2, que possui um número menor de participantes (9), foi adotada uma abordagem sistemática de validação cruzada por participante. Este método garante que todos os indivíduos sejam utilizados no conjunto de teste ao longo das iterações, fornecendo uma avaliação mais estável para conjuntos de dados com um número limitado de sujeitos.

Diferentemente do experimento inicial, os modelos de Árvore de Decisão, Floresta Aleatória e a CNN-1D foram treinados com seus hiperparâmetros padrão, sem a aplicação de otimização bayesiana ou busca em grade. O objetivo foi avaliar o desempenho base e a estabilidade de cada arquitetura frente à variabilidade dos dados. O pré-processamento com janelas deslizantes e normalização seguiu a mesma metodologia do estudo inicial.

## 4. Resultados

### 4.1. Resultados do Experimento Inicial

Nesta seção, são apresentados os resultados obtidos a partir da avaliação de diferentes modelos de aprendizado de máquina aplicados aos conjuntos de dados HWU-USP e PAMAP2. Os experimentos consideraram diversas configurações de pré-processamento e ajuste, incluindo tratamentos como achatamento de dados, balanceamento de classes, *Grid Search* e *Keras Tuner*, como mencionado na Seção 3.

As Tabelas 1 e 2 apresentam os resultados encontrados para cada método e tratamento aplicado, nos conjuntos de dados PAMAP2 e HWU-USP respectivamente, e utilizando as divisões de treino e teste mencionadas na Seção 3.

**Tabela 1. Acurácia obtida para cada modelo e os tratamentos aplicados, no conjunto de dados PAMAP2.**

Modelo	Tratamentos Aplicados	Acurácia	Precision	Recall	F1-Score
Árvore de Decisão	Achatamento + Balanceamento dos Dados	76,40%	71,59%	72,69%	69,47%
Floresta Aleatória	Achatamento + Balanceamento dos Dados	79,38%	79,85%	78,00%	74,12%
CNN-1D	Balanceamento dos Dados	71,19%	67,75%	70,92%	66,00%

Os resultados obtidos indicam que, no conjunto de dados PAMAP2 (Tabela 1), todos os modelos avaliados atingiram acurácias superiores a 70%. O melhor desempenho foi obtido pelo modelo Floresta Aleatória (79,38%), seguido pelo *Decision Tree* (76,40%) e pela *CNN-1D* (71,19%). Não obstante, o *Recall* também ficou acima de 70% em todos

**Tabela 2. Acurácia obtida para cada modelo e os tratamentos aplicados, no conjunto de dados HWU-USP.**

Modelo	Tratamentos Aplicados	Acurácia	Precision	Recall	F1-Score
Decision-Tree	Achatamento + Balanceamento dos Dados	33,83%	33,11%	32,11%	32,22%
Decision-Tree	Achatamento + Balanceamento dos Dados + Grid Search	34,51%	33,56%	33,67%	32,22%
Floresta Aleatória	Achatamento + Balanceamento dos Dados	36,98%	37,11%	36,89%	36,44%
Floresta Aleatória	Achatamento + Balanceamento dos Dados + GridSearch	37,12%	37,33%	37,11%	36,55%
CNN-1D	Balanceamento dos Dados	40,90%	39,22%	40,22%	39,56%
CNN-1D	Balanceamento dos Dados + Keras Tuner	44,93%	42,89%	44,44%	43,00%

os casos, e também maior para o modelo Floresta Aleatória (79,85%), indicando boa identificação de instâncias relevantes (verdadeiros positivos) no conjunto de dados. No caso da precisão, apenas o modelo CNN-1D ficou abaixo de 70% (67,75%), mas ainda próximo, indicando também que a taxa de falsos positivos é baixa. Por fim, o *F1-Score* é acima de 70% apenas para o Floresta Aleatória (74,12%), mas próximo para os outros, indicando bom equilíbrio entre precisão e *recall*. Esses resultados indicam que o PAMAP2 possui padrões bem definidos entre as classes, especialmente após os tratamentos de achatamento e balanceamento dos dados.

Em contraste, no conjunto HWU-USP (Tabela 2), os modelos apresentaram desempenho significativamente inferior, com acurácias abaixo de 45%. O melhor resultado foi alcançado pela *CNN-1D*, ajustada com *Keras Tuner*, com 44,93% de acurácia. A inclusão de otimização de hiperparâmetros contribuiu para o ganho em desempenho, ainda que modesto. Modelos baseados em árvores, como o Floresta Aleatória, apresentaram desempenho inferior, mesmo após a aplicação de *Grid Search* para ajuste de hiperparâmetros. Os resultados se mantiveram também para a precisão, *recall* e *F1-score*, com o CNN-1D tratado com o *Keras Tuner* possuindo os melhores resultados. Mesmo assim, seus valores foram, respectivamente, 42,89%, 44,44% e 43,00%, bem abaixo do outro conjunto de dados. Não obstante, estes dados evidenciam que a classificação foi equilibrada, mantendo uma boa proporção entre *recall* e precisão.

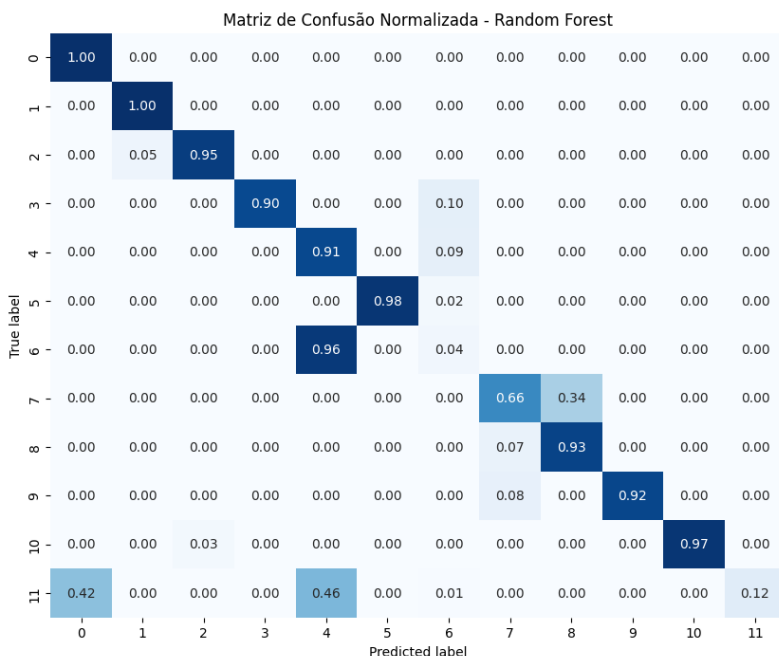
As discrepâncias entre os dois conjuntos de dados indicam a necessidade de técnicas de extração de características aprimorada e modelos capazes de capturar dependências



temporais e contextuais mais sutis. No caso do HWU-USP, o desafio adicional reside no fato de que as atividades presentes — como *preparando uma xícara de chá, fazendo um sanduíche, lavando as louças*, entre outras — ocorrem em um mesmo ambiente e envolvem gestos similares. Essa homogeneidade contextual e motora pode gerar sobreposição nos padrões de sensores, dificultando a distinção entre classes. Por outro lado, o PAMAP2 apresenta maior diversidade de atividades físicas, contribuindo para padrões mais distintivos e separáveis.

A Figura 1 apresenta a matriz de confusão da aplicação do algoritmo Floresta Aleatória no conjunto de dados PAMAP2, ou seja, o modelo que obteve melhor desempenho. Nela, podemos observar que o modelo tem bom acerto para a maioria das categorias (valores próximos de 1,0 na diagonal principal). Porém, para a categoria 7, os acertos são mais próximos de 0,66 e, para as categorias 6 e 11, ele é próximo de 0. Logo, tais categorias são aquelas que trazem maior dificuldade para o aprendizado do modelo, possivelmente por serem atividades similares.

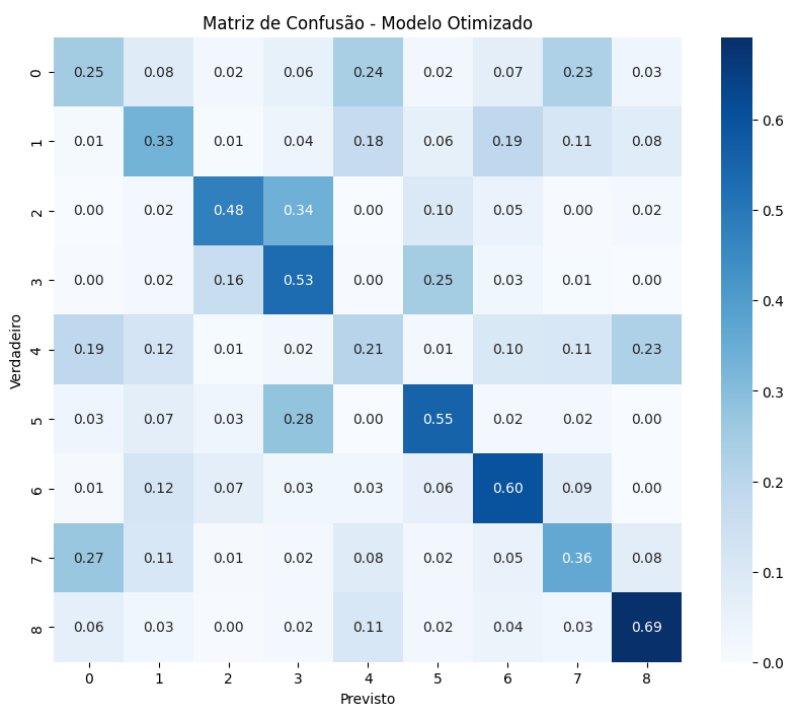
Por fim, a Figura 2 apresenta a matriz de confusão da aplicação do algoritmo *CNN-ID* no conjunto de dados HWU-USP, ou seja, o modelo que obteve melhor desempenho. Nela, observamos que, apesar da diagonal principal ter, no geral, os maiores valores, eles são mais próximos ao intervalo de 0,5 a 0,6, indicando precisão mediana. Em especial, as categorias 2, 3, 5, 6 e 8 obtiveram melhor desempenho, sendo aquelas mais fáceis para o modelo distinguir. Porém, as categorias 0, 1, 4 e 7 obtiveram valores entre 0,2 a 0,4, indicando maior dificuldade de predição. As matrizes de confusão dos outros modelos podem ser verificadas no projeto Open Science Framework (OSF) desta pesquisa<sup>2</sup>



**Figura 1. Matriz de Confusão: Radom Forest no conjunto de dados PAMAP2.**

Em síntese, o pré-processamento aplicado ao PAMAP2 demonstrou-se eficaz, mas revelou-se insuficiente para lidar com a complexidade das tarefas do HWU-USP, exigindo

<sup>2</sup>[https://osf.io/bmuwn/?view\\_only=1f30a1016ce04dec89bbb70e2fa7d5b5](https://osf.io/bmuwn/?view_only=1f30a1016ce04dec89bbb70e2fa7d5b5)



**Figura 2. Matriz de Confusão do modelo CNN-1D com otimização no conjunto de dados HWU-USP.**

abordagens mais especializadas.

#### 4.2. Análise de Generalização

Para avaliar a capacidade de generalização dos modelos, os resultados do experimento de validação em múltiplas amostras foram consolidados. A Tabela 3 apresenta a acurácia média e o desvio padrão para cada modelo nos dois conjunto de dados.

**Tabela 3. Resultados da validação em múltiplas amostras (Acurácia Média  $\pm$  Desvio Padrão).**

Conjunto de dados	Modelo	Acurácia
PAMAP2	Árvore de Decisão	37,95% $\pm$ 5,66%
PAMAP2	Floresta Aleatória	43,73% $\pm$ 8,13%
PAMAP2	CNN-1D	62,48% $\pm$ 10,31%
HWU-USP	Árvore de Decisão	34,42% $\pm$ 3,06%
HWU-USP	Floresta Aleatória	38,93% $\pm$ 3,94%
HWU-USP	CNN-1D	37,47% $\pm$ 3,52%

A análise dos novos resultados revelou importantes nuances sobre a robustez dos modelos. No conjunto PAMAP2, a CNN-1D mostrou-se resiliente à ausência de otimização, mantendo o melhor desempenho médio (62,48%). Em contrapartida, a performance dos modelos clássicos foi acentuadamente inferior à do experimento inicial (Tabelas 1 e 2), evidenciando sua forte dependência do ajuste fino.

Já no cenário mais complexo do HWU-USP, o Floresta Aleatória apresentou o resultado mais estável (acurácia de 38,93%). Este resultado sugere que a natureza de

ensemble do Floresta Aleatória oferece uma linha de base robusta e estável para atividades domésticas com padrões similares, mesmo sem otimização. A performance da CNN-1D, por sua vez, foi inferior à de sua versão otimizada no experimento inicial (44,93%), o que reforça a importância do ajuste de hiperparâmetros para sua arquitetura neste cenário. O baixo desvio padrão em todos os modelos sugere que a dificuldade de classificação é uma característica intrínseca dos dados.

## 5. Conclusão

Este estudo demonstrou que, embora modelos clássicos - Árvore de Decisão e Floresta Aleatória - e uma CNN-1D atinjam acurácias acima de 70% em atividades físicas bem definidas (PAMAP2), seu desempenho cai para menos de 45% em tarefas domésticas similares (HWU-USP), como lavar a louça, preparar uma tigela de cereais e fazer um sanduíche, cujos padrões de movimentação apresentam semelhanças. A otimização por *Keras Tuner* elevou a CNN-1D ao melhor resultado nesse cenário, mas ainda evidencia a limitação de abordagens tradicionais diante de padrões sobrepostos. Uma análise subsequente em múltiplas amostras aprofundou esses achados: com modelos sem ajuste de hiperparâmetros, a CNN-1D confirmou sua superioridade no PAMAP2 (acurácia de 62,48%), enquanto o Floresta Aleatória se mostrou mais estável para o HWU-USP (acurácia de 38,93%). Uma possível causa para esta baixa capacidade de predição dá-se pela similaridade das atividades realizadas, que podem exigir modelos mais complexos para identificar pequenas alterações nos padrões, ou o uso de sensores diversos, que potencializem as diferenças nesses padrões. Além disso, a relativa pequena taxa amostral dos bancos de dados pode influenciar nestes resultados, demandando abordagens que consigam bons resultados em bases menores, ou o uso de estratégias de *data augmentation*.

Esses achados reforçam a necessidade de *pipelines* mais avançados de *feature engineering*, capturando a magnitude vetorial, velocidade e aceleração dos sinais, e análises na frequência como a Transformada de Fourier para identificação das frequências dominantes, aplicação de *data augmentation* para séries temporais, devido as poucas amostras disponíveis e a variação destas, o aumento de dados pode beneficiar o aprendizado das atividades, e adoção de arquiteturas baseadas em atenção (como *Transformers*) que capturem dependências temporais sutis, permitindo a captura de informações contextuais, para abaixar um braço antes este deve ter sido levantado, como quando deseja-se alcançar uma prateleira. Estudos futuros também devem explorar a fusão de múltiplos tipos de sensores — ambientais e inerciais — para enriquecer a representação das atividades, com o acesso aos dados dos sensores em itens permite-se a identificação de objetos sendo utilizados na execução da atividade, contribuindo para melhorar o desempenho das classificações.

## Referências

- Alagoz, C. (2024). Comparative analysis of xgboost and minirocket algorithms for human activity recognition. *arXiv preprint arXiv:2402.18296*.
- Cicirelli, G., Marani, R., Petitti, A., Milella, A., and D’Orazio, T. (2021). Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population. *Sensors*, 21(10).

- Dempster, A., Schmidt, D. F., and Webb, G. I. (2021). Minirocket: A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 248–257.
- D'Arco, L., Wang, H., and Zheng, H. (2022). Assessing impact of sensors and feature selection in smart-insole-based human activity recognition. *Methods and Protocols*, 5(3):45.
- Guo, W., Yamagishi, S., and Jing, L. (2024). Human activity recognition via wi-fi and inertial sensors with machine learning. *IEEE Access*, 12:18821–18836.
- Jovanovic, M., Mitrov, G., Zdravevski, E., Lameski, P., Colantonio, S., Kampel, M., Tellioglu, H., and Florez-Revuelta, F. (2022). Ambient assisted living: Scoping review of artificial intelligence models, domains, technology, and concerns. *J Med Internet Res*, 24(11):e36553.
- Manouchehri, N. and Bouguila, N. (2023). Human activity recognition with an hmm-based generative model. *Sensors*, 23(3):1390.
- Marques, G., Bhoi, A. K., Albuquerque, V., and S., H. (2021). *IoT in Healthcare and Ambient Assisted Living*. Springer.
- O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Keras Tuner. <https://github.com/keras-team/keras-tuner>.
- Ranieri, C. M., MacLeod, S., Dragone, M., Vargas, P. A., and Romero, R. A. F. (2021). Activity recognition for ambient assisted living with videos, inertial units and ambient sensors. *Sensors*, 21(3):768.
- Reiss, A. and Stricker, D. (2012). Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE.
- Yadav, S. K., Tiwari, K., Pandey, H. M., and Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems*, 223:106970.