

Avaliação de Classificadores Clássicos e Janelamento Temporal para Detecção de Congelamento da Marcha em Pacientes com Doença de Parkinson, Utilizando Dados de Sensores IMU

Clebson I. S. Silva, Ronaldo F. Zampolo e Antonio Pereira, Jr

Resumo— Este estudo investigou métodos clássicos de aprendizado de máquina para detectar congelamento da marcha na Doença de Parkinson. Avaliando diferentes tamanhos de janela temporal, o classificador Random Forest alcançou o melhor F1-score (0,92) com janela de 5 segundos. O desempenho estabiliza ou diminui em janelas maiores, indicando que há um ponto ótimo para equilíbrio entre sensibilidade e responsividade. Os resultados reforçam a viabilidade de modelos clássicos para monitoramento em tempo real em dispositivos com restrições computacionais.

Palavras-Chave— Doença de Parkinson, Congelamento da Marcha, Aprendizado de Máquina, Sensores Inerciais, Janelamento Temporal.

Abstract— This study investigated classical machine learning methods for detecting freezing of gait in Parkinson’s disease. By evaluating different temporal window sizes, the Random Forest classifier achieved the best F1-score (0.92) with a 5-second window. Performance stabilized or decreased with larger windows, indicating an optimal point for balancing sensitivity and responsiveness. The results highlight the feasibility of classical models for real-time monitoring in devices with computational constraints, offering accessible alternatives to deep learning approaches for the automatic detection of freezing of gait in practical applications.

Keywords— Parkinson’s Disease, Gait Freezing, Machine Learning, Inertial Sensors, Temporal Windowing.

I. INTRODUÇÃO

A Doença de Parkinson (DP) é uma desordem neurodegenerativa crônica e progressiva que afeta o movimento, impactando milhões de pessoas mundialmente [1]. Entre os diversos sintomas motores, o Congelamento da Marcha (FOG, do inglês *Freezing of Gait*) destaca-se como um dos mais comuns e incapacitantes, especialmente em estágios avançados da doença, impondo um fardo significativo na vida diária dos pacientes com DP e aumentando o risco de quedas e lesões associadas [2]. O FOG é definido como um episódio súbito e breve de incapacidade de marcha, caracterizado pela sensação

de que os pés estão “colados ao chão”, geralmente durante a iniciação da marcha e em situações que exigem adaptações rápidas do movimento, como atravessar corredores estreitos, evitar obstáculos e durante giros corporais. A detecção automática e assertiva do FOG é de grande importância para a avaliação clínica do paciente com DP, o monitoramento da progressão da doença e o desenvolvimento de estratégias de intervenção.

Nos últimos anos, a análise da marcha por meio de sensores vestíveis, particularmente Unidades de Medição Inercial (IMUs), emergiu como uma abordagem promissora para a detecção de FOG [3]. As IMUs, que geralmente possuem acelerômetros e giroscópios, são portáteis, de baixo custo, e capazes de capturar dados cinemáticos do movimento humano. Técnicas de aprendizado de máquina (ML) podem ser aplicadas aos dados provenientes destes sensores para classificar automaticamente os episódios de FOG e diversos estudos têm explorado o uso tanto de classificadores clássicos de ML quanto abordagens baseadas em aprendizado profundo, utilizando *datasets* proprietários ou públicos [3].

Este trabalho tem como objetivo investigar o desempenho de classificadores clássicos de aprendizado de máquina (*Support Vector Classifier - SVC*, *Random Forest - RForest*, *K-Nearest Neighbors - KNN* [4], e *XGBoost* [5]) e o impacto de diferentes tamanhos de janelas temporais na detecção de episódios de FOG em indivíduos com DP que usaram uma IMU localizada próximo ao calcanhar durante uma tarefa experimental com giro de 360 graus [6]. A tarefa de giro corporal representa um gatilho comum e clinicamente significativo para episódios de FOG, sendo uma das manobras mais eficientes para indução do fenômeno em ambiente controlado [7]. Embora os classificadores e o *dataset* utilizados já sejam bem estabelecidos na literatura, a principal contribuição deste trabalho reside na criação de um benchmark rigoroso e sistemático para a detecção de FOG neste *dataset* público. Investigou-se o impacto do tamanho da janela temporal, fornecendo um referencial de desempenho crucial com modelos clássicos. A demonstração de que modelos computacionalmente mais eficientes podem alcançar eficácia comparável à de arquiteturas de aprendizado profundo destaca um avanço importante no desenvolvimento de dispositivos vestíveis de baixo consumo energético para uso diário.

Clebson I. S. Silva, Laboratório de Processamento de Sinais, Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Tecnologia, Universidade Federal do Pará, Belém-PA, Brasil (email: clebson.silva@itec.ufpa.br); Ronaldo F. Zampolo, Laboratório de Processamento de Sinais, Faculdade de Engenharia da Computação e Telecomunicações, Instituto de Tecnologia, Universidade Federal do Pará, Belém-PA, Brasil (email: zampolo@ufpa.br); Antonio Pereira, Jr, Laboratório de Processamento de Sinais, Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharias Elétrica e Biomédica, Instituto de Tecnologia, Universidade Federal do Pará, Belém-PA, Brasil (email: apereira@ufpa.br)

II. METODOLOGIA

Esta seção detalha os procedimentos adotados para a classificação de episódios de FOG, abrangendo a descrição do *dataset*, as etapas de pré-processamento dos sinais com foco no janelamento, a extração de características, os modelos de classificação empregados e o protocolo experimental.

A. Descrição do Dataset

Este estudo utilizou um *dataset* público [8] disponibilizado na plataforma Figshare (www.figshare.com). O *dataset* é composto pelos dados de 35 pacientes com DP idiopática e histórico de FOG (16 mulheres e 19 homens), com idades entre 44 e 84 anos. Todos os pacientes estavam sob efeito de medicação associada com o tratamento de DP.

1) *Tarefa Experimental*: Os participantes realizaram uma tarefa de "giro no lugar" (*turning-in-place*), caminhando em círculo alternadamente para a direita e para a esquerda, em ritmo auto-selecionado, durante 2 minutos. Cada participante realizou três sessões experimentais com intervalo de um mês entre elas.

2) *Dados de Sensores*: Durante a tarefa, os participantes utilizaram uma unidade de medição inercial (IMU) *Physilog 5* fixada próximo ao calcanhar do lado mais afetado pela doença. A IMU registrou acelerações lineares e velocidades angulares triaxiais a uma taxa de amostragem de 128 Hz.

3) *Anotação de FOG*: Os episódios de FOG foram identificados por dois especialistas em distúrbios do movimento, que revisaram os vídeos das sessões utilizando o *software* ELAN [9]. As anotações foram sincronizadas com os dados do IMU para rotulagem dos eventos.

B. Pré-processamento e Extração de Características

Foram analisadas janelas de 1 a 6 segundos, abrangendo e estendendo o intervalo de 2 a 4 segundos frequentemente citado como ideal na literatura [10], [11] para equilíbrio entre latência e detecção de padrões de FOG. O limite superior de 6 segundos evita uma redução excessiva do número de segmentos amostrais, o que comprometeria a robustez do treinamento. Cada janela segmentada recebeu rótulo binário (FOG/Não-FOG), conforme anotações do *dataset*. O critério de 75% para rotulagem FOG visou ampliar a quantidade de exemplos positivos sem diluir suas características, maximizando o aproveitamento das amostras e mantendo a representatividade dos padrões relevantes. Em cada janela, foram extraídas características temporais, espectrais e de correlação entre eixos dos sensores.

- **Características Temporais**: Para cada um dos seis eixos dos sensores, foram calculados: média, desvio padrão, valor máximo, valor mínimo, valor RMS (*Root Mean Square*), assimetria (*skewness*), curtose, contagem de cruzamentos por zero e SMA (*Signal Magnitude Area*).
- **Características Espectrais**: Utilizando a Transformada Rápida de Fourier (FFT), foram calculadas métricas como média e máximo da magnitude no espectro de frequência, e a potência espectral na banda de 3-8 Hz, que está associada ao tremor característico do FOG [12]. Adicionalmente, a razão da energia foi calculada como a razão

entre a potência na banda de tremor (3-8 Hz) e a potência no restante do espectro (obtida após a aplicação de um filtro rejeita-faixa de 3-8 Hz).

- **Características de Wavelet**: Energia e desvio padrão dos coeficientes da transformada wavelet discreta (DWT) com wavelet 'db4' no nível 4. As características baseadas em wavelet foram incluídas por sua capacidade de analisar sinais não estacionários.
- **Características de Correlação**: Coeficientes de correlação de Pearson entre pares de eixos ortogonais para os dados de acelerômetro e giroscópio.
- **Características Derivadas de GRF**: A Força de Reação ao Solo (GRF - *Ground Reaction Force*) é um dos parâmetros cinemáticos da análise da marcha e pode ser inferida a partir dos dados de aceleração no eixo longitudinal do paciente, o peso do sujeito. Foram calculadas a média, desvio padrão e valor máximo da GRF estimada para cada janela.

Um detalhamento completo de cada uma destas características, seu domínio de origem e o número de componentes associados é apresentado na Tabela I.

As 99 características resultantes foram então normalizadas utilizando a técnica de padronização (subtração da média e divisão pelo desvio padrão), antes de serem utilizadas para o treinamento dos modelos de classificação.

C. Classificação e Comparação dos Classificadores

A classificação dos episódios de FOG foi realizada utilizando quatro algoritmos de aprendizado de máquina: *Support Vector Classifier* (SVC), *Random Forest* (RForest), *K-Nearest Neighbors* (KNN) e *XGBoost*. Os modelos foram implementados com o uso da biblioteca Scikit-learn [4] para SVC, RForest e KNN, e da biblioteca XGBoost [5].

Para assegurar a reprodutibilidade do estudo, os hiperparâmetros de cada classificador foram definidos conforme apresentado na Tabela II.

O desempenho de cada classificador foi rigorosamente avaliado através de um protocolo de validação cruzada *Leave-One-Subject-Out* (LOSO). Nesta abordagem, a cada iteração, os dados de um paciente são completamente reservados para o conjunto de teste, enquanto os dados de todos os outros pacientes são utilizados para o treinamento do modelo. Este método é considerado padrão-ouro em aplicações de saúde, pois oferece uma estimativa robusta e realista da capacidade de generalização do modelo para novos indivíduos não vistos durante o treinamento [13].

As métricas consideradas para a comparação de desempenho foram Acurácia, F1-score, Sensibilidade (para a classe FOG), Especificidade (para a classe Não-FOG) e a Área Sob a Curva ROC (AUC).

III. RESULTADOS

A Tabela III apresenta os resultados detalhados do desempenho dos classificadores para diferentes tamanhos de janela temporal (1 a 6 s). As métricas consideradas incluem Acurácia, F1-score (F1), Sensibilidade para FOG (Sens.), Especificidade para Não-FOG (Espec.) e AUC.

TABELA I
DETALHAMENTO DAS 99 CARACTERÍSTICAS EXTRAÍDAS PARA A CLASSIFICAÇÃO DE FOG

Característica	Domínio	Sensor(es) de Origem	Nº de Componentes
Média	Temporal	ACC / GYR	6
Desvio Padrão	Temporal	ACC / GYR	6
Valor Máximo	Temporal	ACC / GYR	6
Valor Mínimo	Temporal	ACC / GYR	6
Valor RMS	Temporal	ACC / GYR	6
Assimetria (Skewness)	Temporal	ACC / GYR	6
Curtose	Temporal	ACC / GYR	6
Cruzamentos por Zero	Temporal	ACC / GYR	6
Área da Magnitude do Sinal (SMA)	Temporal	ACC / GYR	6
Média da Magnitude (FFT)	Espectral	ACC / GYR	6
Máximo da Magnitude (FFT)	Espectral	ACC / GYR	6
Potência Espectral	Espectral	ACC / GYR	6
Razão de Energia	Espectral	ACC / GYR	6
Energia dos Coeficientes	Wavelet	ACC / GYR	6
Desvio Padrão dos Coeficientes	Wavelet	ACC / GYR	6
Média da GRF Estimada	Derivada	Apenas ACC (SI)	1
Desvio Padrão da GRF Estimada	Derivada	Apenas ACC (SI)	1
Máximo da GRF Estimada	Derivada	Apenas ACC (SI)	1
Correlação de Pearson	Correlação	Entre eixos do ACC	3
Correlação de Pearson	Correlação	Entre eixos do GYR	3
Total de Características			99

TABELA II
HIPERPARÂMETROS DEFINIDOS PARA CADA CLASSIFICADOR

Modelo	Hiperparâmetro	Valor
SVC	kernel	'rbf'
	C	1.0
	gamma	'scale'
	probability	True
	random_state	42
RForest	n_estimators	100
	min_samples_split	2
	min_samples_leaf	1
	random_state	42
	n_neighbors	5
KNN	weights	'uniform'
	algorithm	'auto'
	leaf_size	30
	p	2
XGBoost	n_estimators	100
	max_depth	6
	learning_rate	0.3
	random_state	42

A análise dos resultados, detalhados na Tabela III, revela uma clara tendência de melhoria no desempenho com o aumento do tamanho da janela, com o ponto ótimo de desempenho geral sendo alcançado em 5 segundos. O comportamento dos classificadores variou significativamente conforme a duração da janela. O classificador SVC destacou-se em janelas mais curtas, apresentando o melhor F1-score em 1s e 2s, e manteve uma alta especificidade em todas as configurações.

A partir de 3 segundos, os modelos de *ensemble* assumiram a liderança. O Random Forest (RForest) emergiu como o modelo mais robusto, alcançando o melhor F1-score global do estudo (0.9233) com uma janela de 5s. Em um contraste notável, o XGBoost, que apresentou o desempenho mais fraco na janela de 1s, demonstrou uma forte recuperação com janelas maiores, tornando-se altamente competitivo e chegando a liderar em F1-score e sensibilidade em outras configurações. O modelo KNN, por sua vez, embora competitivo, obteve seu principal destaque ao alcançar a maior especificidade na janela de 6s

A Figura 1 mostra a sensibilidade (FOG) dos classificadores em função do tamanho da janela temporal, complementando os valores apresentados na Tabela III.

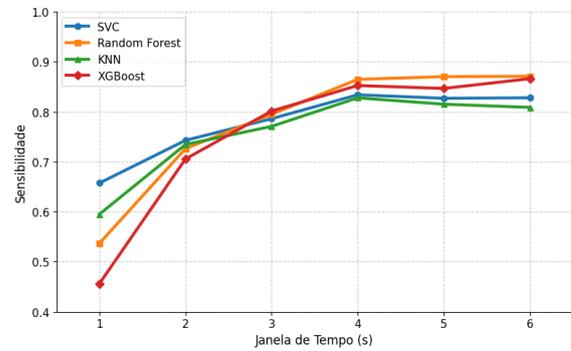


Fig. 1. Sensibilidade dos classificadores em função do tamanho da janela temporal.

Além do desempenho dos modelos, foi realizada uma análise de importância das características extraídas, utilizando o critério de Informação Mútua. A Figura 2 apresenta o ranque das características de acordo com sua relevância para a tarefa de classificação.

As barras laranjas representam características derivadas do giroscópio, enquanto as barras azuis referem-se ao acelerômetro. Nota-se que a maior parte das variáveis mais relevantes está relacionada ao giroscópio, especialmente no eixo SI (Superior-Inferior).

IV. DISCUSSÃO

Os resultados demonstram o impacto significativo do tamanho da janela temporal na eficácia da detecção de FOG utilizando classificadores clássicos. A melhoria de desempenho com janelas maiores (4 s) sugere que segmentos de maior duração podem capturar de forma mais completa as características dinâmicas dos episódios de FOG, que são eventos complexos e podem variar em sua manifestação temporal [6]. Janelas mais curtas, como a de 1 segundo, podem fragmentar excessivamente o evento, dificultando a distinção entre FOG e não-FOG, o que é refletido nos menores valores de *F1-score* e sensibilidade, especialmente para o *XGBoost* nessa configuração.

TABELA III

DESEMPENHO DOS MODELOS POR JANELA DE TEMPO. MELHOR VALOR POR MÉTRICA EM AZUL; PIOR VALOR EM VERMELHO.

Janela	Modelo	Acurácia	F1	Sens.	Espec.	AUC
1s	SVC	0.8766	0.8225	0.6573	0.9485	0.8029
	RForest	0.8543	0.7767	0.5367	0.9583	0.7475
	KNN	0.8324	0.7639	0.5950	0.9102	0.7526
	XGBoost	0.8155	0.7168	0.4561	0.9332	0.6946
2s	SVC	0.9117	0.8721	0.7427	0.9648	0.8538
	RForest	0.9064	0.8638	0.7261	0.9631	0.8446
	KNN	0.8935	0.8492	0.7344	0.9435	0.8390
	XGBoost	0.8968	0.8498	0.7054	0.9570	0.8312
3s	SVC	0.9230	0.8886	0.7860	0.9647	0.8754
	RForest	0.9245	0.8912	0.7948	0.9640	0.8794
	KNN	0.9087	0.8694	0.7707	0.9507	0.8607
	XGBoost	0.9214	0.8879	0.8013	0.9581	0.8797
4s	SVC	0.9407	0.9132	0.8338	0.9720	0.9029
	RForest	0.9428	0.9179	0.8646	0.9657	0.9152
	KNN	0.9351	0.9055	0.8277	0.9666	0.8972
	XGBoost	0.9393	0.9126	0.8523	0.9648	0.9085
5s	SVC	0.9374	0.9077	0.8268	0.9693	0.8981
	RForest	0.9471	0.9233	0.8701	0.9693	0.9197
	KNN	0.9330	0.9011	0.8150	0.9670	0.8910
	XGBoost	0.9392	0.9114	0.8465	0.9659	0.9062
6s	SVC	0.9390	0.9098	0.8278	0.9710	0.8994
	RForest	0.9443	0.9196	0.8708	0.9655	0.9182
	KNN	0.9379	0.9071	0.8086	0.9752	0.8919
	XGBoost	0.9454	0.9207	0.8660	0.9683	0.9171

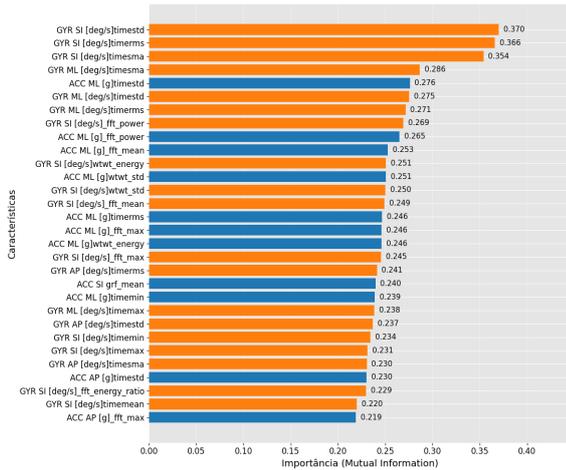


Fig. 2. Ranque de importância das 30 melhores características para detecção de congelamento de movimento, calculado por meio de *Mutual Information*. As métricas derivadas do giroscópio (GYR) e acelerômetro (ACC) nos eixos SI (superior-inferior) e ML (medio-lateral) e AP (Antero-Posterior).

Além do impacto do tamanho da janela, a análise de importância das características (ver Fig. 2) mostra outra informação relevante. A predominância das características do giroscópio, especialmente no eixo Superior-Inferior (SI), está alinhada com a fenomenologia do FOG, que se manifesta como uma interrupção do movimento para a frente acompanhada por tremores de alta frequência nas pernas [14]. A velocidade angular (medida pelo giroscópio) é mais sensível a esses tremores do que a aceleração linear. Esta constatação reforça a importância de incluir sensores giroscópicos em dispositivos vestíveis para a detecção eficaz do FOG, um ponto que é corroborado pelos resultados de alto desempenho dos modelos alimentados com essas características.

O classificador *Random Forest* (*RForest*), particularmente

com a janela de 5 segundos, emergiu como o modelo mais promissor, alcançando os melhores valores globais para Acurácia, *F1-score* e AUC. A versão do *Random Forest* (*RForest*) com janela de 6 s atingiu maior sensibilidade global, enquanto *XGboost* atingiu a maior especificidade global para o mesmo tamanho de janela. A arquitetura do *RForest*, que combina múltiplas árvores de decisão, provavelmente contribuiu para sua robustez e capacidade de generalização, lidando bem com a complexidade e variabilidade dos dados de FOG. O SVC também apresentou um desempenho consistentemente bom, especialmente com janelas menores (1 s e 2 s) e alcançando a maior especificidade global com 4 s, indicando sua forte capacidade de identificar corretamente os períodos de não-FOG. O KNN atingiu melhor especificidade com janelas de 6 s. O *XGBoost* mostrou um desempenho variável, sendo o mais fraco com janelas de 1 s, mas melhorando consideravelmente e até liderando em sensibilidade com janelas de 6 s.

Apesar dos avanços recentes em aprendizado profundo, classificadores clássicos de aprendizado de máquina ainda representam uma alternativa viável para a detecção de FOG, especialmente em cenários com restrições computacionais. No entanto, abordagens mais antigas e simplificadas — como métodos baseados em limiares ou modelos lineares — historicamente apresentaram desempenho relativamente modesto. Por exemplo, Bächlin *et al.* [15] relataram um sistema vestível *online* de detecção de FOG com sensibilidade de 73.1% e especificidade de 81.6% usando apenas características espectrais. Mesmo com janelas similares (3 s), nossos modelos clássicos superaram esse patamar, alcançando *F1-score* de 0.89 (3 s) e 0.92 (4 s). Mazilu *et al.* alcançaram sensibilidade/especificidade superiores a 95% com um *smartphone* e algoritmos de ML [10]; contudo, tais valores foram obtidos em cenários sem validação *leave-one-subject-out* [10]. Quando se exige generalização entre sujeitos, o desempenho tende a

cair, tornando os resultados aqui reportados ($F1 = 0.92$ com um único sensor no calcanhar) competitivos. Estudos recentes empregando *deep learning* – Sigcha *et al.* com CNN-LSTM em uma IMU de cintura [11] e Borzì *et al.* com CNN usando sensor único de tornozelo [16] – reportam *AUC-scores* na faixa de 0.93, mas à custa de maior complexidade computacional e, muitas vezes, uso de múltiplos sensores ou sobreposição de janelas para compensar latência. A meta-análise de Pardoel *et al.* [17] indica que mais de 50% dos trabalhos baseados em IA superam 90% de acurácia, mas enfatiza a queda de desempenho em ambientes não controlados. Assim, nossos resultados mostram que modelos de *ensemble* clássicos, quando alimentados por um conjunto abrangente de atributos e por janelas ≥ 3 s, podem igualar a eficácia de arquiteturas profundas, com menor custo operacional – aspecto crucial para dispositivos de baixo consumo energético em uso diário.

Quanto ao tamanho de janela, a literatura converge para durações de 2 a 4 s como compromisso entre latência e captura de padrões [10], [11]. Nossos achados corroboram essa faixa, mas sugerem ganho adicional ao estender para 5 s, pelo menos em tarefas de giro de 360° . Isso pode refletir a natureza prolongada dos episódios de FOG durante curvas, em contraste ao FOG de início de marcha observado em outros protocolos, que é mais breve.

É notável que, embora a tendência geral seja de melhoria, o desempenho ótimo para a métrica *F1-score* foi alcançado com a janela de 5 s, e não na de 6 s, sugerindo que janelas excessivamente longas podem começar a incorporar dinâmicas não relacionadas ao evento, diluindo as características do FOG.

Limitações e perspectivas

As principais limitações deste estudo, que definem as direções para trabalhos futuros, incluem a necessidade de validação dos resultados em atividades de vida diária, visto que os dados foram coletados em ambiente controlado. A metodologia de rotulagem, baseada em um limiar de 75%, também pode ser aprimorada com estratégias contínuas para capturar episódios de FOG mais curtos. Como perspectivas, sugere-se a exploração de uma granularidade de janela mais fina (com durações > 1 s), a condução de um estudo de ablação para isolar o impacto de cada subconjunto de sensores, e a aplicação de técnicas de seleção automática de características para otimizar o modelo e reduzir sua dimensionalidade, visando à aplicação em dispositivos vestíveis de baixo consumo.

V. CONCLUSÕES

Este estudo avaliou o impacto do tamanho da janela temporal e o desempenho de classificadores clássicos na detecção de FOG. A contribuição fundamental deste trabalho foi o estabelecimento de um benchmark sistemático que demonstrou como a seleção adequada de parâmetros pode otimizar a detecção em um dataset público. Os resultados comprovaram que a combinação de janelas temporais de 4 a 5 segundos com modelos baseados em *ensemble*, como o Random Forest, oferece um pipeline robusto para a detecção de FOG, alcançando um *F1-score* superior a 0.92. Mais significativamente, o estudo demonstrou que estes modelos clássicos, de menor

custo computacional, podem igualar a eficácia de arquiteturas de aprendizado profundo mais complexas, reforçando sua viabilidade para implementação em dispositivos vestíveis de uso contínuo, onde a eficiência energética é crucial.

REFERÊNCIAS

- [1] Yoav Ben-Shlomo, Sirwan Darweesh, Jorge Llibre-Guerra, Connie Marras, Marta San Luciano, and Caroline Tanner, “The epidemiology of parkinson’s disease,” *The Lancet*, vol. 403, no. 10423, pp. 283–292, 2024.
- [2] Wei-Shan Zhang, Chao Gao, Yu-Yan Tan, and Sheng-Di Chen, “Prevalence of freezing of gait in parkinson’s disease: a systematic review and meta-analysis,” *Journal of neurology*, vol. 268, no. 11, pp. 4138–4150, 2021.
- [3] Wei Zhang, Hong Sun, Debin Huang, Zixuan Zhang, Jinyu Li, Chan Wu, Yingying Sun, Mengyi Gong, Zhi Wang, Chao Sun, et al., “Detection and prediction of freezing of gait with wearable sensors in parkinson’s disease,” *Neurological Sciences*, vol. 45, no. 2, pp. 431–453, 2024.
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [5] Tianqi Chen and Carlos Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [6] Yifan Guo, Jian Yang, Yifan Liu, Xian Chen, and Guang-Zhong Yang, “Detection and assessment of parkinson’s disease based on gait analysis: A survey,” *Frontiers in Aging Neuroscience*, vol. 14, pp. 916971, 2022.
- [7] Carolina I Conde, Charlotte Lang, Christian R Baumann, Chris A Easthope, William R Taylor, and Deepak K Ravi, “Triggers for freezing of gait in individuals with parkinson’s disease: a systematic review,” *Frontiers in Neurology*, vol. 14, pp. 1326300, 2023.
- [8] C. Ribeiro de Souza, R. Miao, I. Ávila de Oliveira, A. C. De Lima-Pardini, D. F. De Campos, C. Silva-Batista, and et al., “A public dataset of video, acceleration, and angular velocity in individuals with parkinson’s disease during the turning-in-place task,” 2022, Dataset.
- [9] M. Gilat, “How to annotate freezing of gait from video: a standardized method using open-source software,” *Journal of Parkinson’s Disease*, vol. 9, pp. 821–824, 2019.
- [10] S. Mazilu, M. Hardegger, G. Troster, E. Gazit, and J. M. Hausdorff, “Online detection of freezing of gait with smartphones and machine learning techniques,” in *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*, 2012, pp. 123–130.
- [11] Luis Sigcha, Nuria Costa, Iván Pavón, Ricardo Costa, Pedro Arezes, and Jorge López, “Deep learning approaches for detecting freezing of gait in parkinson’s disease patients through on-body acceleration sensors,” *Sensors*, vol. 20, no. 7, pp. 1895, 2020.
- [12] Martina Mancini, Bastiaan R. Bloem, Fay B. Horak, Simon J. G. Lewis, Alice Nieuwboer, and Jurgen Nonnekes, “Clinical and methodological challenges for assessing freezing of gait: Future perspectives,” *Movement Disorders*, vol. 34, no. 6, pp. 783–790, 2019.
- [13] Sajeev Kunjan, Tyler S Grummett, Kenneth J Pope, David MW Powers, Sean P Fitzgibbon, T Bastiampillai, M Battersby, and Trent W Lewis, “The necessity of leave one subject out (loso) cross validation for eeg disease diagnosis,” in *International conference on brain informatics*. Springer, 2021, pp. 558–567.
- [14] Marika Falla, Giovanni Cossu, and Alessio Di Fonzo, “Freezing of gait: overview on etiology, treatment, and future directions,” *Neurological Sciences*, vol. 43, no. 3, pp. 1627–1639, 2022.
- [15] M. Bächlin, M. Plotnik, D. Roggen, N. Giladi, J. M. Hausdorff, and G. Tröster, “Wearable assistant for parkinson’s disease patients with the freezing of gait symptom,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2010.
- [16] Laura Borzì, Luis Sigcha, Daniel Rodríguez-Martín, and et al., “Real-time detection of freezing of gait in parkinson’s disease using multi-head convolutional neural networks and a single inertial sensor,” *Artificial Intelligence in Medicine*, vol. 135, pp. 102459, 2023.
- [17] Mieke Pardoel, Gurleen Shalin, Julie Nantel, and Edward D. Lemaire, “Insights into parkinson’s disease-related freezing of gait detection and prediction: A comprehensive review,” *Sensors*, vol. 24, no. 12, pp. 3959, 2024.