

Investigation of Machine Learning Techniques to Aid in the Diagnosis of Neurodegenerative Diseases

Juliana Paula Félix¹, Hugo A. D. do Nascimento¹, Nilza Nascimento Guimarães²

¹Instituto de Informática, Universidade Federal de Goiás

²Departamento de Morfologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás

julianafelix@ufg.br, hadn@ufg.br, nilzang2@ufg.br

Abstract. *The thesis summarized in this document introduces alternative, rapid, low-cost, and effective solutions, aided by machine learning techniques, to support the diagnosis and differentiation of neurodegenerative diseases (NDDs) such as Parkinson's Disease, Huntington's Disease, and Amyotrophic Lateral Sclerosis. These diseases, characterized by the progressive loss of neurons, have no cure, and diagnosis is predominantly clinical. By leveraging novel features extracted from gait signals through dynamic fluctuation analysis and harmonic distortion, the thesis achieves highly accurate results with specificity and sensitivity ranging from 96% to 100% for automatic NDD classification, serving as a diagnostic aid system. Furthermore, it presents and discusses an innovative approach to NDD diagnosis focused on the patient's well-being, aiming to reduce examination duration and physical effort required for gait signal collection. These contributions represent innovations in the computational field with the potential to positively impact public health and enhance the quality of life of people with neurodegenerative diseases.*

Resumo. *A tese sumarizada neste documento apresenta soluções alternativas, rápidas, de baixo custo e eficazes, auxiliadas por técnicas de aprendizado de máquina, para apoiar o diagnóstico e diferenciação de doenças neurodegenerativas (NDDs) como a Doença de Parkinson, a Doença de Huntington e a Esclerose Lateral Amiotrófica. Essas doenças, caracterizadas pela perda progressiva de neurônios, não têm cura, e o diagnóstico é puramente clínico. Através do uso de características inovadoras extraídas de sinais da marcha, a partir da análise de flutuação dinâmica e da distorção harmônica, resultados com alta acurácia, especificidade e sensibilidade (96% – 100%) são alcançados para a classificação automática de NDDs, permitindo funcionar como um sistema de apoio ao diagnóstico. Além disso, esta tese apresenta e discute uma abordagem inovadora para o diagnóstico de NDDs focada no bem-estar do paciente, objetivando reduzir a duração do exame e o esforço físico necessário para a coleta de sinais de marcha. As contribuições apresentadas inovam no campo computacional, com potencial de impactar positivamente a saúde pública e a qualidade de vida de pessoas com doenças neurodegenerativas.*

1. Introdução

Doenças neurodegenerativas (NDDs), tais como a Doença de Parkinson (DP), Doença de Huntington (DH) e Esclerose Lateral Amiotrófica (ELA), causam, dentre outros sintomas, a aparição de movimentos involuntários, fraqueza dos músculos e dificuldade de

manter uma marcha estável [Maragakis and Rothstein 2006]. As NDDs não têm cura e, em níveis avançados, podem levar à morte [Heemels 2016]. O diagnóstico, que é uma tarefa difícil devido à falta de testes definitivos que permitam confirmar essas doenças, se torna ainda mais complexo em estágios iniciais, dada a sobreposição de sintomas com outras condições [Erkinen et al. 2018], o que retarda o tratamento e agravam os sintomas.

Um atrativo método alternativo para detecção precoce de NDDs consiste na análise da dinâmica da marcha, visto que alterações no movimento são uma das muitas alterações vivenciadas por pessoas afetadas com NDDs [Heemels 2016]. Formalmente, o ciclo da marcha é definido como o período entre o contato inicial e subsequente do mesmo pé com o solo [Abu-Faraj et al. 1999], e pode ser dividido ainda em intervalos de apoio (contato com o solo) e balanço (tempo no ar), ilustrados na Figura 1.

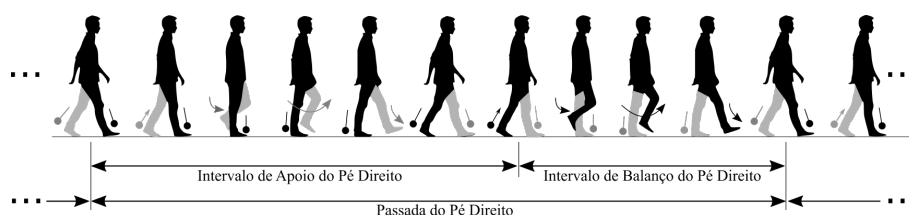


Figura 1. O ciclo da marcha e alguns parâmetros espaço-temporais.

Apesar da relevância de trabalhos anteriores que utilizam aprendizado de máquina para este fim, nossa revisão de literatura (Capítulo 2 da tese) identificou algumas lacunas, como a falta de estudos que relacionam a influência das diferentes fases da marcha (por ex., fases de apoio e balanço) para a discriminação de NDDs; a escassez de métodos automáticos para avaliar a severidade dessas doenças, e a concentração de estudos que realizam a análise de séries temporais de marcha a partir de registros públicos de cinco minutos de caminhada.

Na tese aqui apresentada, investigamos o uso de técnicas de aprendizado de máquina para auxiliar no diagnóstico e diferenciação de pacientes com DP, DH, ELA e indivíduos saudáveis a partir de sinais de marcha. Métodos computacionais alternativos, rápidos e de baixo custo são propostos para auxiliar nesta tarefa. Os métodos propostos inovam ao investigar, pela primeira vez na literatura, características obtidas a partir da distorção harmônica da série da marcha, e analisamos o efeito das fases da marcha (passada, apoio, balanço) na classificação dos sinais. Além disso, investigamos e propomos a redução do tempo necessário para a realização do exame de marcha, priorizando os pacientes. Os métodos propostos são validados em dois conjuntos de dados públicos. Os resultados confirmam a eficácia e eficiência dos métodos propostos, com potencial para auxiliar no diagnóstico de NDDs e serem explorados na prática.

2. Objetivos

Esta tese teve como objetivo desenvolver técnicas novas e não invasivas para auxiliar no diagnóstico de doenças neurodegenerativas (NDDs), especialmente as doenças de Parkinson, doença de Huntington, e a Esclerose Lateral Amiotrófica, utilizando informações da marcha e ferramentas de aprendizado de máquina. Mais precisamente, foram levantados os seguintes objetivos específicos:

- Investigar novas características que possam auxiliar na classificação das NDDs, bem como investigar como a fase da marcha (passada, balanço ou apoio) afeta sua classificação.
- Investigar se durações menores de caminhada fornecem informações suficientes para a discriminação automática de doenças neurodegenerativas.
- Investigar formas de avaliar a severidade de doenças neurodegenerativas a partir da análise da marcha e aprendizado de máquina.

3. Metodologia

A tese apresentada investiga diversos algoritmos e técnicas de aprendizado de máquina para alcançar os objetivos traçados. Em especial, características inovadoras, obtidas a partir dos sinais de marcha, são propostas. Diversas classificações binárias e multiclasse são apresentadas com o objetivo de classificar NDDs ou identificar sua severidade. Os métodos propostos são avaliados em dois conjuntos de dados públicos: GaitNDD¹ e GaitPDB². Ambos os conjuntos de dados utilizam sensores simples e de baixo custo (< R\$ 50 reais) alocados na sola dos pés dos participantes, e fornecem a força aplicada ao solo durante a caminhada. Estratégias de validação-cruzada são utilizadas. Os métodos propostos são avaliados em relação à sua acurácia, sensibilidade, e especificidade, dentre outras métricas, e comparadas com resultados da literatura.

4. Resultados

Os resultados alcançados podem ser divididos em três partes, alinhados com os objetivos específicos traçados. Esta seção descreve, de maneira sucinta, os resultados obtidos.

Classificação de NDDs utilizando Análise de Flutuação Destendenciada

Nesta primeira abordagem proposta, investigamos novas formas de auxiliar no diagnóstico de NDDs, e investigamos o impacto das fases da marcha nesse processo. Propomos a utilização da Análise de Flutuação Destendenciada (DFA) [Peng et al. 1994], uma análise de passeio aleatório modificada, que permite a detecção de correlações de longo alcance em séries temporais, aplicada à classificação automática de séries de marcha. Adicionalmente, realizamos, pela primeira vez na literatura, uma análise abrangente do efeito das diferentes fases da marcha (passada, balanço, apoio) e do pé (direito/esquerdo) para a classificação automática de cada doença investigada, concluindo que certas fases são mais relevantes para determinadas doenças. Os resultados, que superam resultados da literatura, confirmam o potencial para auxiliar no diagnóstico de doenças neurodegenerativas. Essas investigações, incluindo resultados preliminares, resultaram em quatro publicações em veículos internacionais de qualis restrito [Félix et al. 2019, Felix et al. 2019a, Felix et al. 2019b, Felix et al. 2021], além de uma conferência local.

Investigando sessões de caminhada mais curtas para classificar NDDs

A fim de contribuir com a redução do tempo necessário para a realização do exame de marcha, investigamos o uso de apenas 1 minuto de dados de marcha, em vez dos tradicionais 5 minutos encontrados na literatura. Essa redução de tempo, combinada com um

¹<https://physionet.org/content/gaitnnd/1.0.0/>

²<https://physionet.org/content/gaitpdb/1.0.0/>

conjunto inovador de características derivadas da análise da distorção harmônica sobre as séries de marcha, resultaram em excelentes resultados. Além das classificações binárias, exploramos a multiclassificação para discriminar entre as quatro classes investigadas (DP, HD, ELA ou controle). O método proposto alcançou 100% de acurácia em todas as tarefas de classificação com validação cruzada leave-one-out, e 91,91% com validação cruzada de 10 dobras, superando muitos estudos anteriores. Além disso, os resultados para outras durações de experimentos (2 a 5 minutos) também corroboram com a proposta, indicando que mesmo que um paciente seja incapaz de andar durante todo o experimento ou tenha que interromper o exame de marcha inesperadamente, os dados parciais podem fornecer resultados significativos para a classificação. Algumas investigações e trabalhos preliminares foram publicados [Felix et al. 2020, Felix et al. 2022b, Chagas et al. 2024], e um prêmio de melhor trabalho foi recebido. Um artigo em periódico [Felix et al. 2023a] encontra-se atualmente em revisão por pares.

Avaliando a Severidade da Doença de Parkinson

Além da classificação de doenças neurodegenerativas, investigamos também a identificação e classificação da severidade da Doença de Parkinson utilizando sinais de marcha. A abordagem proposta consiste em um algoritmo de multiclassificação para distinguir pessoas saudáveis de pessoas com Parkinson em três diferentes níveis de severidade de acordo com a escala Hoehn & Yahr (HY). Para isso, características estatísticas e de dispersão foram extraídas e alimentadas em uma rede neural *feedforward*. Os resultados obtidos para os casos de teste mostram uma acurácia média de 98,79%, com sensibilidade média de 98,93% e especificidade de 99,34%. Além de superar resultados anteriores, os novos métodos relatados se destacam por utilizar uma rede neural simples, alimentadas por novas características harmônicas propostas nesta tese. As descobertas iniciais dessas abordagens foram publicadas em [Felix et al. 2022a], e resultados adicionais foram submetidos a uma revista [Felix et al. 2023b], que aguarda revisão por pares.

5. Discussão

Nesta tese, abordagens inovadoras para a análise de sinais de marcha foram propostas utilizando análise de flutuação destendenciada e características harmônicas em conjunto com a avaliação das diferentes fases do ciclo da marcha. Por meio de técnicas de aprendizado de máquina, os sistemas desenvolvidos conseguem identificar corretamente os sinais de marcha, contribuindo para o diagnóstico das doenças de Parkinson, Huntington e Esclerose Lateral Amiotrófica com alta acurácia, especificidade e sensibilidade (96% – 100%). Além do diagnóstico de doenças neurodegenerativas, também avaliamos a severidade da Doença de Parkinson. Identificar o estágio da doença possibilita adaptar o cuidado do paciente de forma específica, incluindo ajuste de medicamentos, terapias de reabilitação e suporte personalizado, melhorando a qualidade de vida do paciente e minimizando o impacto dos sintomas em sua rotina diária ao longo da progressão da doença.

Outro grande diferencial do projeto desenvolvido é a preocupação com o bem-estar dos pacientes durante os exames de marcha. Para isso, investigamos a adoção de intervalos mais curtos de caminhada (1 minuto), visando reduzir o esforço físico e aumentar a adesão dos pacientes aos exames, que podem apresentar dificuldades em caminhar por longos períodos, mesmo em fases iniciais da NDD. A redução da duração do exame de marcha também poderia facilitar a coleta de novos dados de marcha, visando

a disponibilização de novos conjuntos de dados públicos. Em particular, a existência de novos dados de marcha de pessoas com NDDs seriam extremamente benéficos para o campo do diagnóstico assistido por computador, que sofre com a falta deste tipo de dados para melhor generalizar os algoritmos de aprendizagem de máquina.

Enquanto os métodos tradicionais de coleta de dados de marcha, como plataformas de força ou sistemas complexos de câmeras, exigem ambientes controlados e estão associados a um alto custo financeiro, as abordagens propostas e investigadas nesta tese oferecem alternativas economicamente viáveis e de baixo custo computacional, o que permite uma fácil implementação em unidades básicas de saúde e no Sistema Único de Saúde (SUS). Por este motivo, algoritmos clássicos de aprendizado de máquina e redes neurais rasas foram exploradas, em contraste com algoritmos de aprendizado profundo, que geralmente requerem tanto a existência de mais dados, quanto recursos computacionais mais complexos. Assim, as soluções alternativas, rápidas, de baixo custo e eficazes para auxiliar no diagnóstico, diferenciação e mensuração de severidade de doenças neurodegenerativas aqui propostas, representam uma contribuição relevante para o campo do diagnóstico assistido por computador, com potencial para impactar positivamente a saúde pública e a qualidade de vida dessas pessoas.

6. Contribuições Científicas

Diversos artigos científicos diretamente relacionados à tese foram publicados em meios nacionais e internacionais nas áreas de computação (IEEE COMPSAC, Qualis A2; IEEE CCECE, Qualis A3; ISVC/LNCS, Qualis A4), computação aplicada à saúde (SBCAS, Qualis A4), inteligência artificial (IEEE ICTAI, Qualis A3), bioinformática (IEEE BIBM, Qualis A2), engenharia biomédica (CLAIB & CBEB, Qualis B4) e automação (CBA, Qualis B4), além de algumas conferências locais. Um dos artigos recebeu um prêmio de melhor trabalho, e três periódicos aguardam revisão por pares.

Além de publicações, o projeto recebeu apoio financeiro da CAPES, incluindo bolsas de pós-doutorado e de professor visitante (Edital nº 30/2022). Apesar dos desafios da pandemia do COVID-19, estabelecemos uma parceria com uma unidade da Secretaria de Estado da Saúde de Goiás, visando implementar e estender o projeto para outras unidades de saúde, além de coletar novos dados de marcha que serão disponibilizados publicamente para a comunidade científica. Finalmente, este projeto tem contribuído para a formação acadêmica, com alunos de mestrado e graduação (iniciação científica, tecnológica e trabalhos de conclusão de curso) desenvolvendo estudos relacionados, sob supervisão da autora. Outras contribuições científicas são listadas no Apêndice A da tese.

Referências

- Abu-Faraj, Z. O., Harris, G. F., Smith, P. A., and Hassani, S. (1999). Human gait and clinical movement analysis. *Wiley Encycl. of Electrical and Electro. Eng.*, pages 1–34.
- Chagas, A. L., Bucci, G., Felix, J., Fonseca, A., Nascimento, H., and Soares, F. (2024). Avaliando a sobreamostragem de dados temporais de marcha no diagnóstico automático de doenças neurodegenerativas. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), Goiânia, GO, Brazil, June 25–28, 2024*, pages 1–12. SBC.
- Erkkinen, M. G., Kim, M.-O., and Geschwind, M. D. (2018). Clinical neurology and epidemiology of the major neurodegenerative diseases. *Cold Spring Harbor Perspectives in Biology*, 10(4):a033118.

- Felix, J., Fonseca, A. U., Araújo, R., Nascimento, H., and Guimarães, N. (2022a). Classificação de Severidade da Doença de Parkinson Utilizando Sinais de Marcha e Aprendizado de Máquina. In *IX Congresso Latino-Americano de Engenharia Biomédica (CLAIB 2022) e o XXVIII Congresso Brasileiro de Engenharia Biomédica (CBEB 2022)*, Florianópolis, SC, Brazil, October 24–28, 2022, pages 601–606. SBEB.
- Felix, J., Fonseca, A. U., Nascimento, H., and Guimarães, N. (2022b). Rede Neural Multicamadas para Classificação de Doenças Neurodegenerativas a partir de Sinais de Marcha. In *XXIV Congresso Brasileiro de Automática*, pages 1354–1361. SBA.
- Felix, J. P., do Nascimento, H. A. D., Guimarães, N. N., Pires, E. D. O., da Silva Vieira, G., and de Souza Alencar, W. (2020). An Effective and Automatic Method to Aid the Diagnosis of Amyotrophic Lateral Sclerosis Using One Minute of Gait Signal. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2745–2751. IEEE.
- Felix, J. P., Fonseca, A. U., Araújo, R., Nascimento, H. A. D. d., and Guimarães, N. N. (2023a). A Patient-Focused Approach for Neurodegenerative Diseases Diagnosis using One-Minute Walking Data and Neural Networks. *Biomedical Signal Processing*. Submitted on April 2023.
- Felix, J. P., Fonseca, A. U., Araújo, R., Nascimento, H. A. D. d., and Guimarães, N. N. (2023b). Assessing Parkinson’s Disease Severity Using Gait Signals and a Feed-forward Neural Network. *Biomedical Signal Processing*. Submitted on February 2023.
- Felix, J. P., Nascimento, H. A. D. d., Guimarães, N. N., Pires, E. D. O., Da Fonseca, A. U., and Vieira, G. D. S. (2021). Automatic Classification of Amyotrophic Lateral Sclerosis through Gait Dynamics. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1942–1947. IEEE.
- Félix, J. P., Vieira, F. H. T., Cardoso, Á. A., Ferreira, M. V. G., Franco, R. A. P., Ribeiro, M. A., Araújo, S. G., Corrêa, H. P., and Carneiro, M. L. (2019). A Parkinson’s Disease Classification Method: An Approach Using Gait Dynamics and Detrended Fluctuation Analysis. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering, CCECE 2019, Edmonton, AB, Canada, May 5-8, 2019*, pages 1–4. IEEE.
- Felix, J. P., Vieira, F. H. T., da Silva Vieira, G., Franco, R. A. P., da Costa, R. M., and Salvini, R. L. (2019a). An Automatic Method for Identifying Huntington’s Disease using Gait Dynamics. In *31st IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Portland, USA, November 4-6, 2019*, pages 1659–1663. IEEE.
- Felix, J. P., Vieira, F. H. T., Franco, R. A. P., da Costa, R. M., and Salvini, R. L. (2019b). Diagnosing Huntington’s Disease Through Gait Dynamics. In *14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7-9, 2019, Proceedings, Part II*, Lecture Notes in Computer Science, pages 504–515. Springer.
- Heemels, M.-T. (2016). Neurodegenerative diseases. *Nature*, 539(7628):179–180.
- Maragakis, N. J. and Rothstein, J. D. (2006). Mechanisms of disease: astrocytes in neurodegenerative disease. *Nature Clinical Practice Neurology*, 2(12):679–689.
- Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical review e*, 49(2):1685.