



INSTITUTO FEDERAL DE ALAGOAS
CAMPUS MACEIÓ
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

IGOR MATHEUS BARROS MOREIRA,
JONATHAS RODRIGUES CORREIA

QUANTIFICAÇÃO DO SINAL DO TREMOR PARKINSONIANO EM EXAMES
MANUSCRITOS EM ESPIRAL UTILIZANDO MACHINE LEARNING

MACEIÓ, AL

2023

IGOR MATHEUS BARROS MOREIRA,
JONATHAS RODRIGUES CORREIA

QUANTIFICAÇÃO DO SINAL DO TREMOR PARKINSONIANO EM EXAMES
MANUSCRITOS EM ESPIRAL UTILIZANDO MACHINE LEARNING

Monografia apresentada ao curso de Graduação Sistemas de Informação, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação no Instituto Federal de Alagoas

Orientador: Dr. Leonardo Melo de Medeiros
Coorientador: Danilo Fernandes Costa

MACEIÓ, AL

2023



Dados Internacionais de Catalogação na Publicação
Instituto Federal de Alagoas
Campus Maceió
Biblioteca Benevides Monte

006.3

M838q

Moreira, Igor Matheus Barros.

Quantificação do sinal do tremor parkinsoniano em exames manuscritos em espiral utilizando *Machine Learning*/ Igor Matheus Barros Correia, Jonathas Rodrigues Correia. – Maceió, 2023.

47 f. : il., color.

Orientação: Prof. Dr. Leonardo Melo de Medeiros.

Coorientação: Prof. Danilo Fernandes Costa.

Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Instituto Federal de Alagoas, Campus Maceió. Maceió, 2023.

Arquivo no formato digital em PDF.,

1. Sistemas de Informação. 2. Doença de Parkinson – Micrografia. 3. *Machine Learning* – Diagnóstico da Doença de Parkinson. 4. Exames em espirais. I. Correia, Jonathas Rodrigues. II. Título.

IGOR MATHEUS BARROS MOREIRA,
JONATHAS RODRIGUES CORREIA

QUANTIFICAÇÃO DO SINAL DO TREMOR PARKINSONIANO EM EXAMES
MANUSCRITOS EM ESPIRAL UTILIZANDO MACHINE LEARNING

Monografia apresentada ao curso de Graduação
Sistemas de Informação, como requisito parcial
para obtenção do título de Bacharel em Siste-
mas de Informação no Instituto Federal de Ala-
goas

Aprovado em: Maceió, AL, 27 de Out. 2023

BANCA EXAMINADORA



Prof. Dr. Leonardo Melo de
Medeiros - Orientador
Instituto Federal de Alagoas - IFAL

Prof. Msc. Breno Jacinto Duarte da Costa
Instituto Federal de Alagoas - IFAL

Prof. Dr. Alex Emanuel Barros Costa
Instituto Federal de Alagoas - IFAL

*Este trabalho é dedicado às pessoas que,
sofrem do mal de Parkinson.*

AGRADECIMENTOS

Os agradecimentos principais são direcionados aos nossos parentes: Elenilda dos Santos Rodrigues, Danyele Lima Barros e José Adagnon Moreira Lima Junior, que deram apoio e educação a sua maneira.

Agradecimento especial para o Professor Orientador: Dr. Leonardo Melo de Medeiros por todo auxílio e conhecimento passado durante a construção deste trabalho.

Agradeço aos professores da coordenação do Curso de Sistemas de Informação que diretamente contribuíram para a nossa formação acadêmica.

*“Onde há um desejo, há um caminho.”
(Albert Einstein)*

RESUMO

Doença de Parkinson (DP) é uma enfermidade neurológica progressiva que se desenvolve gradualmente e provoca sintomas motores, como tremores, lentidão nos movimentos (bradicinesia) e desequilíbrio postural. A DP afeta milhões de pessoas em todo o mundo, especialmente a população idosa. Os dados da Organização Mundial de Saúde (OMS) mostram que aproximadamente 1% da população mundial com idade superior a 65 anos tem a doença. O diagnóstico da DP é difícil e o erro de diagnóstico é comum, principalmente nos estágios iniciais. A micrografia é uma técnica comumente empregada no diagnóstico da doença de Parkinson e consiste essencialmente em realizar exames manuscritos. Neste contexto, a Aprendizagem de Máquina (AM) surge como uma ferramenta poderosa para aplicar técnicas de classificação de dados que podem identificar padrões e auxiliar na detecção da DP. Este estudo utiliza dois modelos de Machine Learning, sendo eles, Support Vector Machine (SVM) e Random Forest Classifier (RFC), com o objetivo de classificar exames manuscritos e contribuir para melhorar a precisão e a eficácia no diagnóstico da DP. Os resultados evidenciam que ambos os modelos obtiveram um desempenho promissor, com altas taxas de acurácia, atingindo 86% para o RFC e 82% para o SVM. Esses resultados sugerem que a aplicação da Aprendizagem de Máquina na classificação de exames manuscritos pode ser uma ferramenta valiosa no diagnóstico da DP.

Palavras-chaves: Doença de Parkinson; Machine Learning; Micrografia; Exames em Espirais.

ABSTRACT

Parkinson's disease (PD) is a progressive neurological disease that develops gradually and causes motor symptoms, such as tremors, slow movement (bradykinesia) and postural imbalance. PD affects millions of people worldwide, especially the elderly population. Data from the World Health Organization (WHO) shows that approximately 1% of the world's population over the age of 65 has the disease. Diagnosing PD is difficult and misdiagnosis is common, especially in the early stages. Micrography is a technique commonly used in the diagnosis of Parkinson's disease and essentially consists of performing handwritten examinations. In this context, Machine Learning (ML) emerges as a powerful tool to apply data classification techniques that can identify patterns and assist in the detection of PD. This study uses two Machine Learning models, namely Support Vector Machine (SVM) and Random Forest Classifier (RFC), with the aim of classifying handwritten exams and contributing to improving the accuracy and effectiveness in diagnosing PD. The results show that both models achieved promising performance, with high accuracy rates, reaching 86% for the RFC and 82% for the SVM. These results suggest that the application of Machine Learning to classify handwritten exams can be a valuable tool in diagnosing PD.

Key-words: Parkinson's disease; Machine Learning; Micrography; Spiral Exams.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura	29
Figura 2 – Acurácia dos modelos utilizados	37
Figura 3 – Métricas de validação	37

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ANNs	<i>Artificial Neural Networks</i>
DBS	<i>Deep Brain Stimulation</i>
DST	<i>Dynamic Spiral Test</i>
DP	Doença de Parkinson
GB	<i>Gigabyte</i>
GS	<i>Grid Search</i>
HOG	<i>Histogram of Oriented Gradient</i>
HPO	<i>Hyper-parameter Optimization</i>
ISP	<i>Internet Service Provider</i>
K-NN	<i>K-Nearest Neighbors</i>
LXC	<i>Linux Containers</i>
ML	<i>Machine Learning</i>
MLP	<i>Multi-Layer Perceptron</i>
OMS	Organização Mundial de Saúde
OPF	<i>Optimum-Path Forest</i>
PD	<i>Parkinson Disease</i>
RBF	<i>Radial Basis Function</i>
RFC	<i>Random Forest Classifier</i>
RS	<i>Random Search</i>
SQL	<i>Structured Query Language</i>
SST	<i>Standard Spiral Test</i>
SVM	<i>Support Vector Machine</i>
UPDRS	<i>Unified Parkinson's Disease Rating Scale</i>
WHO	<i>World Health Organization</i>

SUMÁRIO

1	INTRODUÇÃO	13
	INTRODUÇÃO	13
1.1	MOTIVAÇÃO	15
1.2	OBJETIVO	16
1.2.1	Objetivo específico	16
1.3	TRABALHOS RELACIONADOS	16
1.4	ORGANIZAÇÃO DO TRABALHO	17
2	REVISÃO DE LITERATURA	18
2.1	DOENÇA DE PARKINSON (DP)	18
2.1.1	Fisiopatologia	18
2.1.2	Diagnóstico	19
2.1.3	Tratamento	19
2.2	<i>MACHINE LEARNING</i>	20
2.2.1	<i>Support Vector Machine (SVM)</i>	20
2.2.2	<i>Random forest classifier (RFC)</i>	21
2.2.3	<i>Histogram of Oriented Gradients (HOG)</i>	21
2.2.4	<i>Hyper-parameter Optimization (HPO)</i>	22
3	MATERIAS E MÉTODOS	24
3.1	NATUREZA DA PESQUISA	24
3.2	MATERIAIS	25
3.2.1	<i>Hardware</i>	25
3.2.2	<i>Software</i>	25
3.2.2.1	<i>Docker</i>	25
3.2.2.2	<i>Jupyter Notebook</i>	26
3.2.2.3	<i>Google Colab</i>	26
3.2.2.4	<i>Python</i>	27
3.2.2.5	<i>Scikit-Learn</i>	27
3.2.2.6	<i>NumPy</i>	27
3.2.2.7	<i>Pandas</i>	27
4	ARQUITETURA	29
4.1	PROCESSAR DADOS	29
4.1.1	Coletar Dados	29
4.1.2	Pré-processar Dados	30

4.1.2.1	Partição	30
4.1.2.2	Escala	31
4.1.2.3	<i>HOG Feature</i>	31
4.1.2.4	Viés	32
4.1.3	Engenharia de Características	33
4.1.3.1	Extração de Características	33
4.2	DESENVOLVER MODELO	34
4.2.1	Treinar, Sintonizar, Avaliar	34
4.2.1.1	Seleção do Algoritmo	34
4.2.1.2	Treinamento do Modelo	34
4.2.1.3	Ajuste de Hiperparâmetros	34
4.2.1.4	Métricas de Validação	35
4.3	RESULTADOS	35
4.3.1	<i>Support Vector Machine</i>	35
4.3.2	<i>Random Forest Classifier</i>	36
4.3.3	Comparação dos Resultados	36
5	CONCLUSÃO	38
5.1	TRABALHOS FUTUROS	39
	REFERÊNCIAS	40

1 INTRODUÇÃO

O parkinsonismo é definido como uma vasta categoria de doenças que apresentam diminuição da neurotransmissão dopaminérgica nos gânglios da base, estando estas classificadas em: parkinsonismo primário, secundário, plus e hereditárias degenerativas (SOUZA et al., 2001). Sendo a Doença de Parkinson (DP) a correspondente ao parkinsonismo primário descrita neste trabalho.

Estudos apontam que a prevalência aumenta com a idade, tendo estimativa de 5 a 26 casos a cada 100 mil pessoas por ano, sendo de aproximadamente 1% entre os indivíduos de 65 a 69 anos e, variando de 3% a 14,3% entre os idosos acima de 85 anos (WILLIS et al., 2010).

De acordo com a *The Parkinson's Foundation* (Parkinson's Foundation, 2018), mais de 10 milhões de pessoas em todo o mundo estão vivendo com DP. No Brasil, uma a cada quinhentas pessoas desenvolve o quadro da doença, estimando-se um total de 200 mil pessoas acometidas pelo mal de Parkinson.

Como o envelhecimento populacional está aumentando, as previsões é que a ocorrência da doença cresça em praticamente todas as maiores nações do mundo. Estudos indicam que até 2030, em países como a Indonésia o número de pessoas com a DP venha a dobrar (DORSEY et al., 2007).

A doença de Parkinson é o segundo distúrbio neurodegenerativo relacionado à idade mais comum após a doença de Alzheimer (BAGNOLI, 2021). A Doença ou Mal de Parkinson é uma desordem neurológica degenerativa progressiva de áreas específicas do sistema nervoso central (cérebro e medula espinhal) que acomete principalmente o sistema motor. O indivíduo portador desta patologia apresenta uma alteração da postura, um déficit progressivo de equilíbrio e coordenação, tremor, rigidez, uma marcha alterada, bradicinesia, e em estados mais avançados pode apresentar déficits de atenção e aprendizagem, entre outros (ZUCCO, 2009).

Como mencionado, a doença de Parkinson é uma doença neurodegenerativa, ou seja, acomete células nervosas e provoca sua morte. Mais especificamente, acomete células da substância negra, que fazem parte do sistema dopaminérgico dos núcleos da base, sistema que transmite sinais que controlam os movimentos do corpo (PEREIRA, 2000).

Sua etiologia é tida como idiopática, mas estudos acreditam que a doença de Parkinson pode ser decorrente de um conjunto de fatores, sejam eles genéticos, toxinas ambientais, estresse oxidativo, anormalidades mitocondriais e/ou alterações do envelhecimento (PINHEIRO et al., 2006; PEREIRA; GARRETT, 2010). Como ainda o diagnóstico depende dos sintomas para detectar tal doença, geralmente, no momento da avaliação médica, o paciente já perdeu cerca de 70% dos neurônios dopaminérgicos (PEREIRA, 2017).

A DP não tem cura, mas vários tratamentos são conhecidos por aliviar os sintomas (MATHUR; PATHAK; BANDIL, 2019). Atualmente, seu tratamento é baseado na reposição de dopamina, embora abordagens alternativas, como a estimulação cerebral profunda (DBS), sejam adequadas para a doença em estágio avançado. Os tratamentos atualmente disponíveis oferecem um bom controle dos sintomas motores, mas não impedem a progressão da neurodegeneração, a evolução da doença e o aumento da incapacidade (BALESTRINO; SCHAPIRA, 2020).

Atualmente, o diagnóstico da DP é puramente clínico, não há nenhum exame, teste quantitativo ou biomarcador capaz de indicar o diagnóstico da doença. Em geral, a DP é avaliada por meio de escalas subjetivas, a mais utilizada é a Escala Unificada de Classificação da DP (Unified Parkinson's Disease Rating Scale – UPDRS), a qual é dividida em quatro partes de forma a servir como guia na avaliação motora e não motora do paciente. O tratamento da DP é majoritariamente sintomático e visa manter e melhorar a independência funcional e reduzir o desconforto dos pacientes (ANDRADE et al., 2017).

Portanto, sistemas criados com o intuito de facilitar parte do diagnóstico e avaliação médica tanto remotamente quanto localmente, indicando possíveis aspectos da doença e a melhora ou piora dos sintomas, trazem benefícios consideráveis para qualidade de vida desses pacientes idosos acometidos pela DP (SILVA, 2012; FONSECA, 2009).

Com o auxílio da Aprendizagem de Máquina (*Machine Learning* - ML) é possível utilizar técnicas de classificação de dados com o intuito de detectar padrões e auxiliar na resolução de problemas complexos do mundo real. Nos últimos anos o diagnóstico da Doença de Parkinson vem avançando graças aos investimentos realizados em pesquisas e a evolução dos aparatos computacionais. Através da computação, pesquisadores estão ampliando seus conhecimentos sobre a doença, influenciados por informações cada vez mais precisas, que vão desde o monitoramento de gestos, distúrbios vocais à identificação de alterações genéticas (SOUZA; MEDEIROS; OLIVEIRA, 2016).

No contexto da DP, trabalhos voltados a auxiliar o seu diagnóstico tem apresentado cada vez mais resultados positivos. Spadoto (SPADOTO et al., 2010; SPADOTO et al., 2011) propuseram uma aplicação baseada em técnicas evolutivas buscando maximizar a precisão e taxa de acerto no diagnóstico dessa enfermidade por meio do classificador Floresta de Caminhos Ótimos (*Optimum-Path Forest* - OPF) (PAPA; FALCÃO; SUZUKI, 2009; PAPA et al., 2012). (SOLEIMANIANGHAREHCHOPOGH; MOHAMMADI, 2013) utilizaram redes neurais artificiais (*Artificial Neural Networks* - ANNs) com perceptron multicamadas (*Multi-Layer Perceptron* - MLP) para diagnosticar os efeitos causados pelo mal de Parkinson. Já Pan et al. (PAN et al., 2012) analisaram o desempenho de máquinas de vetores de suporte (*Support Vector Machines* - SVMs), ANN-MLP e redes neurais de função de base radial para comparar o início do tremor em pacientes com mal de Parkinson. Estudos realizados por Watters et al. (WATTERS; PATEL, 1999) por meio de ANNs evidenciaram a degradação das vias dopaminérgicas do cére-

bro, indicando que o sistema semântico é o que sofre o maior prejuízo, ocasionando, assim, respostas mais lentas com uma quantidade maior de erros em tarefas de avaliação semântica em um paciente com mal de Parkinson.

A micrografia é outra abordagem amplamente utilizada para o diagnóstico de um paciente com doença de Parkinson, que é basicamente um exame de escrita (EICHHORN et al., 1996). Tal técnica é considerada uma medida objetiva para caracterização da doença, uma vez que um paciente com DP possivelmente apresenta a redução do tamanho da caligrafia e o tremor das mãos, que atualmente são apontados como os principais sintomas da doença de Parkinson. Hoje em dia, esse procedimento é frequentemente realizado por meio do preenchimento de formulários (PEREIRA et al., 2016).

A análise desses formulários é ainda amplamente utilizada por médicos e pesquisadores para um diagnóstico precoce do mal de Parkinson. Muito embora essa metodologia seja bastante eficaz, a mesma é dependente da interpretação de cada profissional, bem como o grande número de exames a serem analisados podem levar a diagnósticos cada vez menos precisos, principalmente, devido à fadiga causada pelo excesso de trabalho (PEREIRA, 2017).

Existe um grande número de exames de escrita que podem nos fornecer informações valiosas sobre o desenvolvimento da doença de Parkinson, pois é mais barato e mais fácil adquirir esse tipo de exame. Além disso, a maioria dos hospitais e clínicas tem exames de escrita apenas à mão, o que significa que eles precisam ser digitalizados antes da extração de informações. Normalmente, os pacientes são solicitados a desenhar espirais e meandros, que são então comparados com os templates (PEREIRA et al., 2016)

Nesse cenário, o presente trabalho tem como propósito realizar a classificação de exames manuscritos para o auxílio no diagnóstico da Doença de Parkinson, utilizando técnicas de extração de padrões de sinais motores finos. A abordagem central envolve a aplicação de métodos de aprendizado de máquina para identificar padrões e variações nas características dos sinais, culminando em uma detecção mais eficaz e precisa da doença.

1.1 MOTIVAÇÃO

A principal motivação deste trabalho de pesquisa é direcionada para o aprimoramento do diagnóstico da Doença de Parkinson. Especificamente, busca-se alcançar esse objetivo por meio da extração de padrões de exames manuscritos em espirais digitalizados. Além disso, almeja-se a classificação eficaz desses dados, com o propósito de aprimorar a avaliação médica associada a essa condição neurodegenerativa. O intuito fundamental é contribuir para um diagnóstico mais preciso e eficiente, beneficiando assim os pacientes e a prática clínica.

1.2 OBJETIVO

Realizar uma avaliação automática por meio da utilização de técnicas de reconhecimento de padrões e processamento de imagens, visando extrair informações através de formulários utilizados para o diagnóstico do mal de Parkinson.

Este trabalho tem como objetivo principal a aplicação de técnicas de aprendizado de máquina na classificação de exames manuscritos, com o propósito de contribuir para o diagnóstico eficaz da Doença de Parkinson. Para isso, busca-se utilizar modelos de *Machine Learning*, incluindo *Support Vector Machine* (SVM) e *Random Forest Classifier* (RFC), a fim de identificar padrões nos sinais motores finos presentes nos exames manuscritos em espirais.

1.2.1 Objetivo específico

- Propor um método de avaliação automatizada da Doença de Parkinson, utilizando técnicas de extração de características de sinais motores finos em exames manuscritos de espirais com o auxílio de modelos de *Machine Learning*;
- Apresentar técnicas de extração de características;
- Apresentar técnicas de otimização automática de hiperparâmetros.

1.3 TRABALHOS RELACIONADOS

Rosenblum et al. (ROSENBLUM et al., 2013) sugeriram que os exames de escrita podem ser usados para distinguir pacientes com DP de indivíduos saudáveis. Os autores empregaram a seguinte metodologia para apoiar sua suposição: 20 pacientes com DP e 20 indivíduos de controle foram solicitados a escrever seus nomes e endereços em um pedaço de papel anexado a uma mesa digital. Além disso, para cada toque, a pressão média e a velocidade foram medidas para computar informações espaciais e temporais. Os autores apresentaram taxas de reconhecimento muito boas, com 97,5% dos participantes classificados corretamente (100% dos controles e 95% dos pacientes com DP). Mais tarde, Drotár et al. (DROTAR et al., 2014) afirmou que o movimento durante a escrita de um texto consiste não apenas nos movimentos da mão na superfície, mas também nas trajetórias no ar realizadas quando a mão se move no ar de um golpe para o próximo. Os autores demonstraram que a avaliação dos movimentos da mão no ar durante a caligrafia da frase tem um impacto maior do que a avaliação pura dos movimentos na superfície, levando a acurácias de classificação de 84% e 78%, respectivamente.

Isenkul et al. (ISENKUL; SAKAR; KURSUN, 2014) propuseram um novo teste de caligrafia chamado *dynamic spiral test* (DST) para analisar o desempenho motor de pacientes com doença de Parkinson. Os autores coletaram as amostras de caligrafia com a ajuda de um tablet gráfico e um software especialmente desenvolvido. Dois testes, *dynamic spiral test* (DST)

e *standard spiral test* (SST) de 40 indivíduos, incluindo 25 pacientes com Parkinson, foram conduzidos, sendo analisadas a velocidade e aceleração.

Drotár et al. (DROTAR et al., 2016) mostraram como as características cinemáticas e de pressão extraídas das amostras de caligrafia podem ser usadas para classificar pacientes com Parkinson. Os autores também apresentaram o conjunto de dados PaHaW neste artigo. Amostras de caligrafia foram coletadas de 75 indivíduos com 37 pacientes com Parkinson, realizando 8 tarefas diferentes. Características cinemáticas e de pressão foram extraídas das amostras e seis funções básicas foram calculadas. Todos os recursos foram então normalizados para média zero e desvio padrão da unidade. Pelo teste Mann-Whitney U, apenas as 20 principais características foram selecionadas antes da classificação. Três classificadores AdaBoost, SVM e K-NN foram usados para classificação e a maior precisão de 81,3% foi obtida pelo classificador SVM.

Pereira et al. (PEREIRA et al., 2015) apresentaram alguns resultados em direção ao diagnóstico automatizado da DP através de técnicas baseada em visão computacional em um dataset composto por dezenas de amostras. Este dataset continha exames de escrita com imagens em espiral, sendo tais exames posteriormente digitalizados. Logo após, 9 características numéricas foram extraídas tanto do traço escrito, quanto do template espiral, ao medir as diferenças estatísticas entre eles. Os autores visavam descrever e comparar as distinções entre as duas imagens. Por fim, foi realizado experimentos usando três classificadores de padrões supervisionados, nomeados *Naive Bayes*, *Optimum-Path Forest* e *Support Vector Machines* com kernel RBF. Ao qual alcançaram, na sua melhor taxa de reconhecimento, uma acurácia em torno de 79%, o que é considerado bastante razoável.

1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho segue a seguinte estrutura: no Capítulo 2, abordamos a Revisão de Literatura, onde são apresentados os principais conceitos e estudos relacionados aos tópicos abordados nesta pesquisa. O Capítulo 3 detalha a seção de Materiais e Métodos, fornecendo informações sobre os recursos utilizados no estudo e a metodologia empregada na pesquisa. No Capítulo 4, foi concentrado na Arquitetura utilizada neste trabalho, explicando as etapas percorridas, os modelos de *Machine Learning* escolhidos, bem como os parâmetros e métricas de validação envolvidos. Por fim, o Capítulo 5 apresenta as Conclusões obtidas com base nos resultados, discutindo as contribuições do estudo, suas limitações e apontando possíveis direções para trabalhos futuros.

2 REVISÃO DE LITERATURA

Este capítulo introduz considerações fundamentais relacionadas aos tópicos centrais desta pesquisa. Através desta revisão, nossa intenção é construir uma base sólida de conhecimento, abordando tanto os principais aspectos relacionados à Doença de Parkinson quanto os conceitos essenciais ligados à área de *Machine Learning*.

2.1 DOENÇA DE PARKINSON (DP)

O médico inglês James Parkinson caracterizou pela primeira vez a doença de Parkinson em sua monografia de 1817 “*An Essay on the Shaking Paralysis*” (PARKINSON, 1817). Ele notou a natureza progressiva da doença e a grande incapacidade em que ela incorreu e a chamou de *paralisia agitante*. Charcot (CHARCOT, 1877), posteriormente expôs a doença, acrescentando bradicinesia e rigidez à constelação de sintomas, e renomeou a doença como *doença de Parkinson*.

A DP está associada a um risco aumentado de mortalidade por todas as causas e uma redução na expectativa de vida (ISHIHARA et al., 2007; XU et al., 2014) e incapacidade grave. A doença de Parkinson é uma doença progressiva e incapacitante com fardos emocionais, econômicos e sociais tanto para os pacientes quanto para os cuidadores; portanto, pode ser útil encaminhá-los para associações sem fins lucrativos que oferecem suporte a pacientes com DP e seus familiares, fornecendo informações e iniciativas úteis (BALESTRINO; SCHAPIRA, 2020).

2.1.1 Fisiopatologia

A doença de Parkinson é caracterizada pela morte de neurônios dopaminérgicos na substância negra. A marca patológica da doença de Parkinson é o corpo de Lewy, uma inclusão neuronal que consiste principalmente em agregações de proteína α -sinucleína. O modelo mais amplamente citado para explicar a progressão neuropatológica da doença de Parkinson é a hipótese de Braak (BRAAK et al., 2003). Para Armstrong et al. (ARMSTRONG; OKUN, 2020), esse modelo sugere que a doença de Parkinson começa (estágios 1 e 2) na medula e no bulbo olfatório. Essa patologia precoce está associada a sintomas que ocorrem antes do início do distúrbio do movimento, como o distúrbio do comportamento do sono do movimento rápido dos olhos (no qual os indivíduos perdem a paralisia do sono normal do movimento rápido dos olhos e representam fisicamente seus sonhos durante o sono) e diminuição do olfato. Nos estágios 3 e 4, a patologia progride para a parte compacta da substância negra e outras estruturas do mesencéfalo e do prosencéfalo basal. A patologia nessas áreas está associada aos sintomas motores clássicos da doença de Parkinson. A doença de Parkinson é tipicamente diagnosticada

nesta fase. Na doença de Parkinson avançada, a patologia progride para os córtices cerebrais com início de comprometimento cognitivo e alucinações (BRAAK et al., 2003).

A Força-Tarefa para Distúrbios do Movimento reconheceu três estágios no início da doença de Parkinson (OBESO et al., 2017; BERG et al., 2013; STERN; LANG; POEWE, 2012): (1) a fase pré-clínica, na qual a neurodegeneração começa, mas os pacientes não apresentam sintomas clínicos; (2) a fase prodrômica, na qual os sintomas estão presentes, mas são insuficientes para fazer o diagnóstico da doença de Parkinson; e (3) a fase clínica, na qual os sintomas parkinsonianos são manifestos e reconhecíveis.

2.1.2 Diagnóstico

A doença de Parkinson continua sendo um diagnóstico clínico. Os sintomas assimétricos de tremor de repouso, bradicinesia e rigidez com resposta favorável à terapia dopaminérgica sugerem seu diagnóstico. As características exclusivas podem abranger disautonomia grave, alucinações precoces, demência precedendo os sintomas motores e instabilidade postural e congelamento nos primeiros 3 anos após o diagnóstico (GELB; OLIVER; GILMAN, 1999).

Para Berg et al. (BERG et al., 2015), o diagnóstico da doença de Parkinson baseia-se principalmente no histórico e no exame físico. O histórico deve avaliar os sintomas motores e não motores. O histórico familiar de um parente de primeiro grau com doença de Parkinson aumenta a probabilidade de um diagnóstico de doença de Parkinson.

Já Newman et al. e Schrag et al. (NEWMAN et al., 2009; SCHRAG; BEN-SHLOMO; QUINN, 2002), abordam que o diagnóstico da doença de Parkinson é difícil e o erro de diagnóstico é comum, principalmente nos estágios iniciais. Um médico que raramente diagnostica a doença de Parkinson deve considerar o encaminhamento de um paciente com suspeita de tê-la a um médico que tenha mais experiência com a doença para confirmar o diagnóstico (NETWORK, 2010; BRITAIN, 2006). Nenhuma regra de decisão clínica tem utilidade comprovada no diagnóstico da doença precoce (NETWORK, 2010), embora os critérios do Parkinson's UK Brain Bank melhorem a precisão do diagnóstico em pacientes com doença avançada (BRITAIN, 2006). Dada a incerteza inerente do diagnóstico da doença precoce e a crescente precisão diagnóstica com a progressão da doença, os médicos que cuidam de pacientes com doença de Parkinson devem reavaliar periodicamente o diagnóstico (NETWORK, 2010).

2.1.3 Tratamento

Os tratamentos atualmente disponíveis são sintomáticos e não param a neurodegeneração. Embora nos estágios iniciais da doença a terapia farmacológica proporcione um bom controle dos sintomas, nos estágios avançados da doença alguns sintomas resistentes à levodopa (como quedas e congelamento, disartria, disfagia e engasgos, demência, alucinações, sonolência diurna e incontinência urinária) surgem, causando o aumento da incapacidade na DP avançada (BALESTRINO; SCHAPIRA, 2020). Além disso, as complicações associadas ao tratamento

farmacológico aumentam as dificuldades no manejo dos estágios avançados da DP. A maioria dos pacientes perde a autonomia na fase avançada da doença; os principais motivos de internação em asilos e mortalidade são os sintomas de resistência à levodopa (AARSLAND et al., 2000; GOETZ; STEBBINS, 1993)

2.2 MACHINE LEARNING

De acordo com Arthur Samuel (SAMUEL, 1959), *Machine Learning* é definido como o campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados.

Machine Learning (ML) é usado para ensinar as máquinas a lidar com os dados com mais eficiência. Às vezes, depois de visualizar os dados, não podemos interpretar as informações extraídas dos dados. Nesse caso, aplicamos aprendizado de máquina (MAHESH, 2020). Técnicas baseadas em aprendizado de máquina foram aplicadas com sucesso em diversos campos, desde reconhecimento de padrões, visão computacional, engenharia espacial, finanças, entretenimento e biologia computacional até aplicações biomédicas e médicas (MAHESH, 2020).

Machine Learning possui três tipos de algoritmos classificadores, sendo eles: aprendizado supervisionado, não supervisionado e por reforço. O aprendizado supervisionado é usado quando os dados e seu rótulo são conhecidos. Por outro lado, não supervisionado é usado para prever dados quando os rótulos de classe são desconhecidos, por exemplo, agrupamento K-Means. O aprendizado por reforço é muito popular na teoria dos jogos, onde o agente executa as ações possíveis e recebe recompensas ou penalidades, que atuam como condição de aprendizado para os agentes (Department of Computer Science and Engineering, The Maharaja Sayajirao University of Baroda, India et al., 2019).

2.2.1 Support Vector Machine (SVM)

O *Support Vector Machine* (SVM) é um algoritmo de aprendizado de máquina que tem sido usado principalmente para problemas de classificação (SVEC; POPOLO; TITZE, 2003; LARSEN et al., 2000).

O SVM é um sistema de aprendizado de nova geração baseado em avanços recentes na teoria de aprendizado estatístico. É um algoritmo para dados lineares e não lineares. Ele transforma os dados originais em uma dimensão superior, de onde pode encontrar um hiperplano para separação dos dados usando tuplas de treinamento essenciais chamadas vetores de suporte. Uma *Support Vector Machine* é um classificador discriminativo formalmente definido por um hiperplano separador. Em outras palavras, dado treinamento rotulado, a máquina de vetores de suporte constrói um hiperplano ou um conjunto de hiperplanos em um espaço alto ou de dimensão infinita, que pode ser usado para classificação, regressão ou outras tarefas. Intuitivamente, uma boa separação é alcançada pelo hiperplano que possui maior distância até o ponto

de dados de treinamento mais próximo de qualquer classe denominada margem funcional, pois em geral quanto maior a margem menor o erro de generalização do classificador (BIND et al., 2015).

2.2.2 *Random forest classifier (RFC)*

O *Random forest classifier* (RFC) é uma das técnicas de aprendizado de conjunto mais bem-sucedidas, que se mostraram técnicas muito populares e poderosas no reconhecimento de padrões e aprendizado de máquina para classificação de alta dimensão e problemas distorcidos (BREIMAN, 2001).

RFC é um conjunto de árvores de decisão onde se constrói uma floresta (MAJHI; KASHYAP, 2022). Depois que a floresta é formada, as instâncias de teste são percoladas em cada árvore e as árvores fazem suas respectivas previsões de classe. A taxa de erro de uma floresta aleatória depende da força de cada árvore e da correlação entre quaisquer duas árvores. Ele pode ser usado para classificar a importância das variáveis em um problema de regressão ou classificação de maneira natural (BIND et al., 2015).

2.2.3 *Histogram of Oriented Gradients (HOG)*

Um dos métodos de extração de características simples e eficazes é o descritor de características HOG. É um descritor de características rápido e eficiente, que também foi demonstrado que os recursos do HOG são descritores bem-sucedidos para detecção. Principalmente, se usado para detecção de objetos em processamento de imagem e visão computacional. Usando HOG, a forma e a aparência da imagem podem ser descritas. Ele divide a imagem em pequenas células e calcula as direções das bordas. Para melhorar a precisão, os histogramas podem ser normalizados (GREESHMA; SREEKUMAR, 2019).

Em 2005, o descritor HOG foi proposto pela primeira vez por Dalal e Triggs (DALAL; TRIGGS, 2005). Teve um resultado quase perfeito na detecção de pedestres com fundo complexo. O descritor HOG tem os principais benefícios sobre outros descritores. Como pode operar em células locais, é invariante a transformações geométricas e fotométricas, exceto para orientação de objetos. Recentemente, o descritor HOG com classificador SVM é comumente usado em várias aplicações para detecção e reconhecimento de objetos robustos (SURASAK et al., 2018).

HOG é um descritor de características inicialmente proposto para detecção de pedestres (DALAL; TRIGGS, 2005). Ele conta o número de ocorrências de orientação de gradiente em uma janela de detecção. As principais etapas para calcular os recursos do HOG podem ser resumidas da seguinte forma (ZHOU et al., 2020). 1) Cálculo do gradiente: Nesta etapa, os gradientes espaciais nas direções horizontal e vertical são computados. Esses dois gradientes são então usados para calcular as magnitudes e os ângulos do gradiente; 2) Binning de orientação: Nesta etapa, a imagem é dividida em pequenas regiões conectadas chamadas células. A magni-

tude do gradiente de cada pixel em uma célula é votada em diferentes caixas de orientação de acordo com o ângulo do gradiente; 3) Descrição da característica: Nesta etapa, as células adjacentes são agrupadas em blocos. Cada bloco é normalizado por sua norma L2. Os histogramas de blocos normalizados em uma janela de detecção são concatenados para formar um descritor.

2.2.4 *Hyper-parameter Optimization (HPO)*

O ajuste de hiperparâmetros é considerado um componente chave na construção de um modelo de ML eficaz, especialmente para modelos de ML baseados em árvores e redes neurais profundas, que possuem muitos hiperparâmetros (HUTTER; KOTTHOFF; VANSCHOREN, 2019). O ajuste de hiperparâmetros é o processo de identificação do conjunto de valores de hiperparâmetros que se espera que produza o melhor modelo de predição de todos os conjuntos de valores de hiperparâmetros que estão sendo examinados (JIANG; XU, 2022).

O processo de ajuste de hiperparâmetros é diferente entre os diferentes algoritmos de ML devido aos seus diferentes tipos de hiperparâmetros, incluindo hiperparâmetros categóricos, discretos e contínuos (DECASTRO-GARCIA et al., 2019). A obtenção de uma configuração adequada para os hiperparâmetros de um algoritmo de ML requer conhecimento específico, intuição e, muitas vezes, tentativa e erro (MANTOVANI et al., 2015). Esses fatores têm inspirado o aumento da pesquisa em técnicas de otimização automática de hiperparâmetros; a chamada Otimização de Hiperparâmetros (HPO) (OLOF, 2018).

Algumas razões importantes para aplicar técnicas HPO a modelos de ML são as seguintes (HUTTER; KOTTHOFF; VANSCHOREN, 2019):

1. Reduz o esforço humano necessário, uma vez que muitos desenvolvedores de ML gastam um tempo considerável ajustando os hiperparâmetros, especialmente para grandes conjuntos de dados ou algoritmos de ML complexos com um grande número de hiperparâmetros.
2. Melhora o desempenho dos modelos de ML. Muitos hiperparâmetros de ML têm diferentes ideais para obter o melhor desempenho em diferentes conjuntos de dados ou problemas.
3. Torna os modelos e pesquisas mais reproduzíveis. Somente quando o mesmo nível de processo de ajuste de hiperparâmetros é implementado é que diferentes algoritmos de ML podem ser comparados de forma justa; portanto, usar o mesmo método HPO em diferentes algoritmos de ML também ajuda a determinar o modelo de ML mais adequado para um problema específico.

Os métodos teóricos de decisão, uma das técnicas de otimização utilizadas no HPO, baseiam-se no conceito de definir um espaço de busca de hiperparâmetros e, em seguida, detectar as combinações de hiperparâmetros no espaço de busca, selecionando, em última análise, a

combinação de hiperparâmetros de melhor desempenho (YANG; SHAMI, 2020). Grid search (GS) (BERGSTRA et al., 2011) é uma abordagem teórica de decisão que busca exaustivamente a configuração ideal em um domínio fixo de hiperparâmetros. O Random Search (RS) (BERGSTRA; YOSHUA, 2012) é outro método teórico de decisão que seleciona aleatoriamente combinações de hiperparâmetros no espaço de pesquisa, dados tempo de execução e recursos limitados. Em GS e RS, cada configuração de hiperparâmetro é tratada de forma independente.

3 MATERIAS E MÉTODOS

Este capítulo oferece uma visão detalhada dos recursos e abordagens que embasaram o desenvolvimento deste trabalho. Além disso, apresenta a metodologia empregada para alcançar os objetivos propostos. Nas seções subsequentes, abordaremos as etapas cruciais da pesquisa, que desempenham um papel essencial na compreensão dos resultados obtidos.

3.1 NATUREZA DA PESQUISA

Não existe uma única forma de realizar uma pesquisa científica. As metodologias podem possuir as mais variadas formas ou até mesmo o uso de mais de uma delas. Para Gerhardt e Silveira (GERHARDT et al., 2009), elas podem ser classificadas quanto a sua abordagem qualitativa e quantitativa; quanto a sua natureza em pesquisa básica ou aplicada; quanto a seus objetivos em pesquisa exploratória, descritiva ou explicativa; quanto a seus procedimentos sendo experimental; bibliográfica, documental, de campo, EX-POST-FACTO, levantamento, *SURVEY* e estudo de caso.

O presente trabalho adotou uma abordagem metodológica exploratória para auxiliar no diagnóstico da Doença de Parkinson por meio de exames manuscritos em espirais, os quais foram aplicados em modelos de *Machine Learning*.

Para Sampieri et al. (SAMPIERI; COLLADO; LUCIO, 1991) os estudos exploratórios servem para aumentar o grau de familiaridade com fenômenos relativamente desconhecidos, obter informações sobre a possibilidade de levar adiante uma investigação mais completa sobre um contexto particular da vida real e estabelecer prioridades para investigações posteriores, entre outras utilizações. Porém, os estudos exploratórios em poucas ocasiões constituem um fim em si mesmos. “Eles se caracterizam por serem mais flexíveis em sua metodologia em comparação com os estudos descritivos ou explicativos, e são mais amplos e dispersos que estes dois últimos tipos (por exemplo, buscam observar tantas manifestações do fenômeno estudado quanto for possível)” (SAMPIERI; COLLADO; LUCIO, 1991). Para Malhotra (MALHOTRA, 2001), a pesquisa exploratória “é um tipo de pesquisa que tem como principal objetivo o fornecimento de critérios sobre a situação-problema enfrentada pelo pesquisador e sua compreensão”.

Neste estudo de objetivo exploratório, optamos por uma abordagem predominantemente quantitativa na pesquisa. Conforme definido por Malhotra (MALHOTRA, 2001), a pesquisa quantitativa busca quantificar os dados e aplicar análises estatísticas. Isso nos permitiu extrair características específicas dos exames manuscritos relacionadas à identificação da Doença de Parkinson, através de modelos de ML e analisar métricas de desempenho dos resultados gerados por esses modelos de aprendizagem de máquina. A escolha por essa abordagem quantitativa foi motivada pela necessidade de uma análise mensurável dos dados, visando uma avaliação

rigorosa do desempenho dos modelos e a obtenção de resultados quantificáveis.

Para complementar a pesquisa exploratória, buscou-se embasar teoricamente o estudo por meio de uma pesquisa bibliográfica abrangente sobre a doença de Parkinson e suas relações com os exames manuscritos. A revisão bibliográfica proporcionou uma compreensão ampla de pesquisas anteriores, identificando lacunas na análise da doença de Parkinson e fornecendo uma base sólida para a investigação exploratória. Os estudos e artigos científicos revisados forneceram informações cruciais sobre as características clínicas da doença, métodos de diagnóstico existentes e avanços em tecnologias de *Machine Learning* aplicadas a exames médicos. Essa revisão de literatura contribuiu para o embasamento teórico do trabalho e permitiu contextualizar as motivações por trás do estudo, bem como estabelecer conexões entre os resultados obtidos e as descobertas científicas prévias.

3.2 MATERIAIS

Esta seção descreverá os principais recursos utilizados durante a construção deste trabalho.

3.2.1 *Hardware*

Realizou-se o processamento dos modelos de *Machine Learning* utilizando-se de um notebook com as seguintes especificações:

- Host: Inspiron 7580
- OS: Manjaro Linux x86_64
- Memory: 3947MiB / 15712MiB
- CPU: Intel i7-8565U (8) @ 4.600GHz
- GPU: NVIDIA GeForce MX150

3.2.2 *Software*

Os programas, bibliotecas e linguagem de programação utilizados são descritos a seguir.

3.2.2.1 *Docker*

A virtualização de sistemas operacionais e aplicativos, também conhecida como contêiner (por exemplo, Docker (MERKEL; others, 2014)) e LXC (HELSLEY, 2009)), tornou-se popular desde 2013 com o lançamento do projeto de código aberto Docker (docker.com) e com o crescente interesse dos provedores de nuvem (DUA; RAJA; KAKADIA, 2014; BERNSTEIN, 2014) e Provedores de Serviços de Internet (ISP) (NATARAJAN et al., 2016).

Docker é uma plataforma de código aberto que executa aplicativos e torna o processo mais fácil de desenvolver e distribuir. O Docker foi projetado para simplificar a criação, implantação e execução de aplicativos usando contêineres. A containerização permite que o usuário execute aplicativos em um ambiente virtual, empacotando todos os elementos necessários, como arquivos, bibliotecas e outros componentes essenciais juntos. Esses containers continuam rodando de forma isolada sobre o kernel do sistema operacional (BOETTIGER, 2015).

Para o desenvolvimento deste trabalho foi utilizado a imagem `jupyter/tensorflow-notebook:python-3.10`.

3.2.2.2 *Jupyter Notebook*

O Jupyter Notebook é o sistema mais amplamente utilizado para programação interativa (SHEN, 2014). Ele foi projetado para tornar a análise de dados mais fácil de documentar, compartilhar e reproduzir. O sistema foi lançado em 2013, e hoje existem mais de 1 milhão de notebooks no GitHub (RULE; TABARD; HOLLAN, 2018). O Jupyter Notebook originou-se do IPython (PEREZ; GRANGER, 2007) e, além do Python, oferece suporte a uma variedade de linguagens de programação, como Julia, R, Javascript e C.

O Jupyter Notebook é uma ferramenta baseada em navegador de código aberto que funciona como um notebook de laboratório virtual para oferecer suporte a fluxos de trabalho, código, dados e visualizações detalhando o processo de pesquisa. É legível por máquina e por humanos, o que facilita a interoperabilidade e a comunicação acadêmica. Esses notebooks podem residir em repositórios *on-line* e fornecer conexões para objetos de pesquisa, como conjuntos de dados, códigos, documentos de métodos, fluxos de trabalho e publicações que residem em outro lugar (RANGLES et al., 2017).

3.2.2.3 *Google Colab*

O Google criou o Google Colaboratory, amplamente conhecido como Google Colab, que é uma estrutura *on-line* onde se pode escrever, executar códigos de *Deep Learning* e *Machine Learning*. Nesta estrutura, diferentes versões do python e diferentes ambientes *runtime* estão disponíveis. Além disso, ele pode baixar *datasets* enormes diretamente dos servidores para o Google Drive em alta velocidade. Você pode montar sua unidade com o Google Colab e ele pode buscar o arquivo necessário após sua autenticação (KANANI; PADOLE, 2019).

O Google Colab (WELCOME...) fornece GPU para pesquisa para usuários que não têm recursos suficientes ou não podem pagar por um. O serviço Google Colab fornece 12,72 GB de RAM e 358,27 GB de espaço em disco rígido em um *runtime*. Cada *runtime* dura 12 horas, após as quais o tempo é desconectado e o usuário deve estabelecer uma nova conexão.

3.2.2.4 *Python*

Python é uma linguagem de programação interpretada mantida pela Python Foundation e por um comunidade de colaboradores. A linguagem funciona em praticamente todos as plataformas e sistemas existentes e é largamente usada em áreas como inteligência artificial, banco de dados, biotecnologia e *data science* (MENEZES, 2022).

Usou-se a versão 3.10 da linguagem já integrada na imagem docker do Jupyter.

3.2.2.5 *Scikit-Learn*

Scikit-learn (PEDREGOSA et al., 2011) é uma biblioteca de aprendizado de máquina de propósito geral escrita em Python. Ele fornece implementações eficientes de algoritmos de última geração, acessíveis a não especialistas em aprendizado de máquina e reutilizáveis em disciplinas científicas e campos de aplicação. Ele também aproveita a interatividade e a modularidade do Python para fornecer prototipagem rápida e fácil.

O Scikit-learn expõe uma ampla variedade de algoritmos de aprendizado de máquina, supervisionados e não supervisionados, usando uma interface consistente e orientada a tarefas, permitindo assim uma fácil comparação de métodos para um determinado aplicativo. Como ele depende do ecossistema científico do Python, ele pode ser facilmente integrado a aplicativos fora da faixa tradicional de análise de dados estatísticos. É importante ressaltar que os algoritmos, implementados em uma linguagem de alto nível, podem ser usados como blocos de construção para abordagens específicas para um caso de uso, por exemplo, em imagens médicas (MICHEL et al., 2012).

3.2.2.6 *NumPy*

O NumPy está disponível gratuitamente em <http://numpy.scipy.org> e é oferecido sob uma licença muito generosa. Ele surgiu de um módulo original chamado Numeric (e às vezes também chamado de numpy) escrito em 1995 (OLIPHANT, 2007).

NumPy é o pacote fundamental para computação científica em Python. É uma biblioteca Python que fornece um objeto array multidimensional, vários objetos derivados (como arrays e matrizes mascarados) e uma variedade de rotinas para operações rápidas em arrays (NumPy Developers,). Ele lida com persistência de matriz eficiente (entrada e saída) e fornece operações básicas, como produto escalar. A maioria das bibliotecas científicas do Python, incluindo o scikit-learn, usa matrizes NumPy como tipo de dados de entrada e saída (ABRAHAM et al., 2014)

3.2.2.7 *Pandas*

A biblioteca pandas, em desenvolvimento desde 2008, destina-se a preencher a lacuna na riqueza de ferramentas de análise de dados disponíveis entre o Python e as numerosas

plataformas de computação estatística e linguagens de banco de dados de domínio específico (MCKINNEY, 2011).

Este pacote fornece estruturas de dados complexas e funções especificamente projetadas para tornar o trabalho nelas fácil, rápido e eficaz. Este pacote é o núcleo da análise de dados em Python. O conceito fundamental deste pacote é o *DataFrame*, uma estrutura de dados tabular bidimensional com rótulos de linha e coluna (NELLI, 2018).

O Pandas aplica as propriedades de alto desempenho da biblioteca NumPy à manipulação de dados em planilhas ou em bancos de dados relacionais (bancos de dados SQL). De fato, usando indexação sofisticada, será fácil realizar muitas operações nesse tipo de estrutura de dados, como remodelagem, fatiamento, agregações e seleção de subconjuntos (NELLI, 2018).

4 ARQUITETURA

Neste capítulo, é explorado a arquitetura central empregada no estudo. Através dele, é possível compreender todos os componentes que compõem essa arquitetura e de como eles se interconectam para alcançar os resultados discutidos no capítulo seguinte.

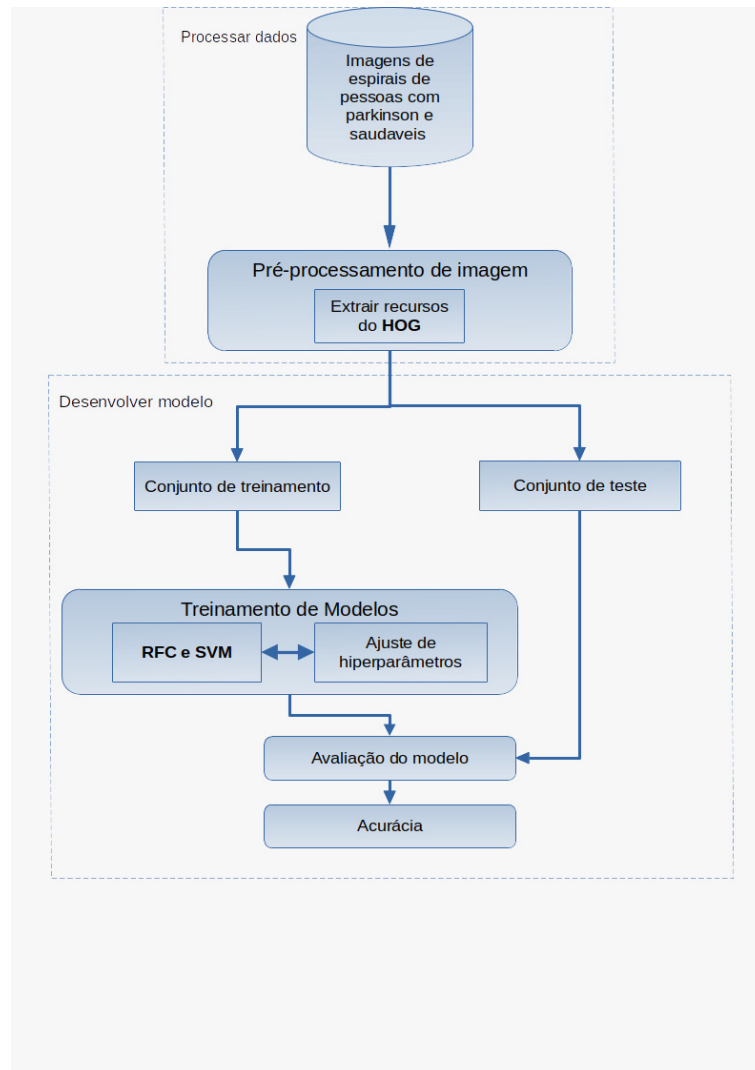


Figura 1 – Arquitetura

4.1 PROCESSAR DADOS

4.1.1 Coletar Dados

Para este trabalho, utilizou-se o dataset disponibilizado no paper de Zham et al. (ZHAM et al., 2017), que pode ser encontrado no Kaggle por meio do link: <https://www.kaggle.com/datasets/kmader/parkinsons-drawings>. Esse dataset é composto por imagens de exames manuscritos em espirais (*spiral*) e ondas (*waves*).

O agrupamento de imagens *spiral* está dividido em dois subgrupos distintos: os conjuntos *testing* e *training*. O grupo *testing* é composto por 15 imagens para cada classe, sendo elas *Healthy* (saudáveis) e *Parkinson*. Já o grupo *training* também possui duas classes, *Healthy* e *Parkinson*, e conta com 36 imagens para cada uma dessas classes.

Cabe ressaltar que, para a elaboração deste trabalho, concentrou-se apenas no agrupamento de imagens *spiral*. As imagens contidas nesse agrupamento foram utilizadas para treinar os modelos de *Machine Learning* com o objetivo de realizar a classificação entre indivíduos saudáveis e aqueles com a doença de *Parkinson*.

4.1.2 Pré-processar Dados

Antes de aplicar os modelos de *Machine Learning*, foram realizados procedimentos de pré-processamento nos dados. Isso incluiu o particionamento do dataset, a padronização das imagens, normalização dos valores dos pixels e a adequação das dimensões para garantir a consistência nos dados utilizados. A partir desses dados preparados, foram aplicados os modelos de SVM e *Random Forest*, permitindo assim a avaliação e comparação do desempenho de cada abordagem na classificação dos exames manuscritos em *spiral*.

4.1.2.1 Partição

Como mencionado, o Dataset utilizado neste trabalho foi importado do Kaggle. No entanto, esse dataset apresentava uma estrutura na qual as quantidades de dados para treinamento e teste estavam separadas de forma fixa, o que exigia uma intervenção manual para modificar essas proporções. Diante da necessidade de avaliar o desempenho do modelo com diferentes proporções de treinamento e teste, optou-se por simplificar a estrutura do dataset. Utilizando a biblioteca *Pandas*, procedeu-se com a junção das imagens, eliminando as subdivisões anteriores. Como resultado, o dataset passou a conter apenas os grupos *Healthy* e *Parkinson*, proporcionando maior flexibilidade na definição das proporções de treinamento e teste, sem a necessidade de intervenção manual.

A fim de dividir os dados em conjuntos de treinamento e teste, utilizou-se o método *train_test_split*. Essa função divide o dataset em duas partes, atribuindo a primeira ao conjunto de treinamento e a segunda ao conjunto de teste. Definiu-se um *test_size* de 0.21, indicando que 21% dos dados foram alocados para o conjunto de teste e 79% para o conjunto de treinamento.

Além disso, foi empregado o parâmetro *random_state* para controlar a aleatoriedade da divisão dos dados. Ao definir um valor específico para o *random_state*, no caso, *random_state* = 10, garantiu-se que a divisão seja consistente e replicável em diferentes execuções do código. Isso significa que, ao utilizar o mesmo valor para o *random_state*, a divisão dos dados será sempre a mesma, o que possibilita a reprodutibilidade dos resultados e facilita a comparação entre diferentes modelos ou experimentos.

4.1.2.2 Escala

Para garantir consistência e preparar as imagens para a análise, todas as imagens de exames manuscritos em espirais foram redimensionadas para a mesma dimensão de 256x256 pixels. Esse redimensionamento ajuda a criar um padrão uniforme para as imagens, facilitando o processamento subsequente.

Além disso, as imagens originais foram convertidas para a escala de cinza. Essa conversão atendeu a dois propósitos: primeiro, reduziu a complexidade das imagens, mantendo apenas informações de intensidade luminosa, o que é crucial para capturar texturas e detalhes; segundo, a conversão para a escala de cinza foi um pré-requisito para a aplicação da técnica *Histogram of Oriented Gradients* (HOG). O HOG é sensível a variações de intensidade nas imagens e, portanto, a conversão para a escala de cinza era necessária para utilizar essa técnica eficazmente.

Essas etapas de pré-processamento são fundamentais para garantir que todas as imagens estejam em um formato padronizado, adequado para a análise posterior nos modelos de *Machine Learning*.

4.1.2.3 HOG Feature

Para melhorar a extração de informações relevantes das imagens, uma técnica amplamente utilizada no pré-processamento de imagens foi empregada, especialmente em problemas de visão computacional e reconhecimento de padrões.

Conforme explicado na documentação do Scikit-image, acessível através do link: https://scikit-image.org/docs/stable/auto_examples/features_detection/plot_hog.html, o processo do HOG é composto por diversas etapas, sendo que cada uma delas contribui para a criação de um vetor de características destinado à representação das imagens. Estas etapas seguem a sequência a seguir:

1. Equalização de Normalização de Imagem Global: Essa etapa opcional utiliza a compressão gama (lei de potência) para reduzir os efeitos de iluminação nas imagens. A intensidade da textura da imagem é normalmente proporcional à iluminação da superfície local, e a compressão gama ajuda a reduzir os efeitos de sombreamento e variações de iluminação.
2. Cálculo de Gradientes de Imagem de Primeira Ordem: Esta etapa captura informações de contorno, silhueta e texturas, resistindo melhor às variações de iluminação. Os gradientes são calculados a partir do canal de cor localmente dominante para proporcionar invariância de cor.
3. Produzir uma codificação que seja sensível ao conteúdo local da imagem: Aqui, a janela da imagem é dividida em pequenas regiões espaciais, denominadas “células”. Para cada

célula, acumulamos um histograma 1-D local de orientações de gradiente ou borda sobre todos os pixels na célula. Este histograma 1-D combinado de nível de célula forma a representação básica de “histograma de orientação”. Cada histograma de orientação divide o intervalo do ângulo do gradiente em um número fixo de compartimentos predeterminados. As magnitudes de gradiente dos pixels na célula são usadas para votar no histograma de orientação.

4. Cálculo da normalização: A normalização apresenta melhor invariância para iluminação, sombreamento e contraste de borda. Esse processo é realizado por meio da acumulação de uma medida de energia do histograma local em grupos de células chamados de blocos. Normalmente, cada célula individual é compartilhada entre vários blocos, mas suas normalizações dependem do bloco e, portanto, são diferentes. A célula, portanto, aparece várias vezes no vetor de saída final com diferentes normalizações. Embora possa parecer redundante, essa abordagem melhora o desempenho. Os descritores resultantes dessa normalização são chamados de *Histograma de Gradiente Orientado* (HOG).
5. Combinação de Descritores HOG: Todos os blocos de uma densa grade sobreposta são coletados para formar um vetor de características combinadas. Este vetor de características é utilizado no classificador de janela, que é um componente essencial para a detecção de objetos.

Esses vetores de características foram então utilizados como entrada para os modelos SVM e RFC, permitindo que esses modelos aprendam padrões e relacionamentos relevantes entre os grupos *Healthy* e *Parkinson* com base nas características identificadas pelo HOG.

4.1.2.4 Viés

O *dataset* utilizado neste trabalho apresenta uma característica positiva, sendo classificado como balanceado, pois é composto por um total de 102 imagens, distribuídas de forma equilibrada entre os grupos *Healthy* (saudável) e *Parkinson*. Essa distribuição igualitária é de extrema importância no treinamento de modelos de *Machine Learning*, pois previne que o modelo desenvolva viés e favoreça uma classe em detrimento da outra. Em *datasets* desbalanceados, onde uma das classes possui um número muito maior ou menor de amostras do que a outra, o modelo pode aprender a ser mais preciso para a classe majoritária e menos preciso para a classe minoritária.

Essa abordagem proporciona uma maior confiabilidade nas previsões do modelo, uma vez que ele não será influenciado pelo desequilíbrio na quantidade de amostras de cada classe. Além disso, o balanceamento contribui significativamente para a interpretabilidade do modelo, pois sua performance não é distorcida pelas diferenças na proporção de amostras entre as classes.

Entretanto, é importante considerar outro fator que pode introduzir enviesamento no *dataset*: o fato de o modelo também ser utilizado por pessoas de idade avançada, as quais podem apresentar tremor nas mãos sem necessariamente ter a doença de Parkinson. Essa circunstância pode influenciar as características das imagens em espiral no dataset, uma vez que tremores relacionados à idade podem ser erroneamente associados ao diagnóstico de Parkinson.

A presença de tremor relacionado à idade nas amostras pode levar o modelo a aprender características que não são exclusivas da Doença de Parkinson, mas sim comuns em pessoas mais velhas. Isso pode resultar em previsões enviesadas, onde o modelo tende a classificar erroneamente amostras de pessoas idosas como pertencentes ao grupo Parkinson.

4.1.3 Engenharia de Características

Uma vez que os dados tenham sido selecionados, torna-se necessário extrair características adequadas para o objetivo em questão, o que é conhecido como engenharia de características. A engenharia de características representa o processo de coletar características a partir de dados brutos e aplicar transformações, possibilitando, posteriormente, aplicação de algoritmos de aprendizado de máquina. Esse processo desempenha um papel de extrema importância em qualquer modelo de aprendizado de máquina, pois, em algumas situações, pode até mesmo proporcionar um melhor desempenho ao modelo utilizado.

4.1.3.1 Extração de Características

A extração de características desempenha um papel fundamental no treinamento de modelos de *Machine Learning*, especialmente ao lidar com dados complexos. Os modelos de *Machine Learning* dependem de características significativas para aprender padrões e relacionamentos relevantes nos dados. A extração cuidadosa dessas características permite que os modelos sejam alimentados com informações relevantes e discriminativas, facilitando a diferenciação entre os grupos *Healthy* e Parkinson nos exames em espiral.

Uma vez extraídas as características, cada modelo de *Machine Learning* utiliza abordagens diferentes para incorporar essas informações. O modelo SVM, por exemplo, mapeia as características para um espaço de alta dimensionalidade. O SVM busca encontrar um hiperplano de separação ótimo entre as classes nesse novo espaço. Foi aplicado um kernel apropriado, como o kernel *radial basis function* (RBF), que permite capturar relacionamentos não lineares entre as características extraídas. Essa abordagem permite que o modelo SVM aprenda a distinguir padrões complexos nos exames em espiral.

No modelo *Random Forest Classifier*, a diversidade das árvores é aproveitada para extrair informações relevantes dos exames em espiral. Embora cada árvore de decisão individual não realize explicitamente a extração de características, o *Random Forest Classifier* treina cada árvore com um subconjunto aleatório das características previamente extraídas dos exames em

espiral. Ao considerar várias árvores de decisão independentes, o RFC é capaz de capturar a variabilidade das características extraídas e explorar diferentes combinações delas.

4.2 DESENVOLVER MODELO

4.2.1 Treinar, Sintonizar, Avaliar

4.2.1.1 Seleção do Algoritmo

A seleção cuidadosa dos algoritmos é fundamental para garantir que o modelo seja capaz de aprender padrões complexos presentes nos dados e realizar uma classificação precisa.

Nesse contexto, consideramos dois algoritmos principais: *Support Vector Machine* (SVM) e *Random Forest Classifier* (RFC). Cada um desses algoritmos possui características distintas que podem se alinhar com os requisitos da tarefa.

A escolha dos algoritmos foi baseada nas características dos dados e, principalmente, na capacidade de cada algoritmo em lidar com a tarefa de classificação dos dados. Ao optar por uma abordagem diversificada, buscamos explorar as forças e fraquezas de cada algoritmo, determinando assim qual deles apresentaria o melhor desempenho para a tarefa em questão, que é a classificação dos exames manuscritos em espirais entre as classes *Healthy* e *Parkinson*.

4.2.1.2 Treinamento do Modelo

O processo de treinamento consiste em fornecer ao algoritmo um conjunto de dados previamente rotulados. O algoritmo utiliza esses exemplos rotulados para ajustar seus parâmetros internos, buscando minimizar a diferença entre as previsões feitas e os rótulos reais.

É importante destacar que a busca do melhor resultado é uma etapa iterativa e repetida. Os modelos são alimentados com os dados de treinamento várias vezes, ajustando-se gradualmente para se tornarem mais eficazes na detecção da doença de Parkinson. A eficácia do treinamento é monitorada por meio de métricas de desempenho, que avaliam a capacidade do modelo de fazer previsões precisas para novas entradas de dados.

4.2.1.3 Ajuste de Hiperparâmetros

Um aspecto crítico durante o desenvolvimento dos modelos é o ajuste nos hiperparâmetros, uma etapa que visa encontrar a combinação ideal de parâmetros que otimize o desempenho do modelo. Os hiperparâmetros são valores configuráveis que afetam diretamente o processo de treinamento e a capacidade de generalização dos modelos.

Nesse contexto, empregamos a técnica de *Hyper-parameter Optimization* (HPO), que envolve a busca sistemática e automática das melhores combinações de hiperparâmetros para os modelos de *Machine Learning*.

O conjunto de hiperparâmetros que resulta no melhor desempenho é selecionado como a configuração final do modelo. Essa abordagem de sintonia de hiperparâmetros assegura que os modelos estejam bem ajustados, maximizando sua capacidade de fazer previsões precisas e confiáveis.

4.2.1.4 Métricas de Validação

Para avaliar a capacidade e o desempenho dos modelos, é crucial empregar métricas de validação que proporcionem informações quantitativas sobre sua eficácia na tarefa de detecção da Doença de Parkinson. A busca por uma acurácia acima de 80% é um indicador importante para garantir a confiabilidade dos resultados.

Nesse contexto, utilizamos diversas métricas de avaliação para medir a performance dos modelos. A acurácia é uma métrica fundamental que representa a proporção de previsões corretas em relação ao total de previsões feitas.

Além da acurácia, também selecionamos outras métricas para validar estatisticamente os modelos, tais como: *precision* (ou taxa de verdadeiros positivos), que mede a proporção de exemplos classificados como Parkinson que realmente pertencem a essa classe; *recall* (ou sensibilidade), que avalia a proporção de exemplos verdadeiramente pertencentes à classe Parkinson que foram corretamente identificados pelo modelo; e *F1-score*, uma métrica que combina a *precision* e o *recall*, oferecendo uma medida abrangente do desempenho do modelo, equilibrando a importância de ambas as métricas.

Ao avaliar as métricas em conjunto, buscamos garantir que o modelo tenha uma performance satisfatória, evitando viés em direção à classe majoritária e atingindo o objetivo de uma acurácia acima de 80%.

4.3 RESULTADOS

Nesta seção, apresentamos os resultados obtidos durante a avaliação dos modelos de aprendizado de máquina para a detecção da Doença de Parkinson a partir de exames manuscritos em espirais. Cada modelo foi submetido a um processo de treinamento e validação rigoroso, utilizando diferentes conjuntos de dados e técnicas de ajuste de hiperparâmetros.

4.3.1 *Support Vector Machine*

Como mencionado nas seções anteriores, para alcançar as métricas de validação, foi conduzido ajustes nos hiperparâmetros do modelo SVM por meio da aplicação de técnicas de Otimização de Hiperparâmetros (HPO). A busca pelos parâmetros ideais foi conduzida ao explorar o espaço de valores definido no primeiro *output* de código abaixo. Após a conclusão bem-sucedida da fase de otimização, o modelo SVM foi treinado usando os parâmetros finais, os quais estão apresentados no segundo *output* de código. Essa abordagem permitiu um atri-

moramento na capacidade do SVM em discernir com maior precisão os exames manuscritos comprometidos pela Doença de Parkinson dos exames saudáveis.

```
{RandomizedSearchCV(cv=5, estimator=SVC(), n_iter=100, n_jobs=-1,
param_distributions={'C': [0.1, 1, 10, 100, 1000],
'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
'kernel': ['rbf', 'linear']},
random_state=10, return_train_score=True, scoring='accuracy',
verbose=1)}
```

```
{'kernel': 'rbf', 'gamma': 0.1, 'C': 10}
```

4.3.2 *Random Forest Classifier*

Da mesma forma que abordado para o modelo SVM, a avaliação do desempenho do Random Forest Classifier também passou por um processo de otimização de hiperparâmetros (HPO). Esse procedimento visou encontrar a combinação mais eficaz considerando o espaço de valores definido no primeiro *output* de código definido abaixo:

```
{'n_estimators': [200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000],
'max_features': ['sqrt'],
'max_depth': [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None],
'min_samples_split': array([ 2,  3,  4,  5,  6,  7,  8,  9, 10, 11]),
'min_samples_leaf': array([1, 2, 3, 4]),
'bootstrap': [True, False]}
}
```

```
{'n_estimators': 200,
'min_samples_split': 10,
'min_samples_leaf': 2,
'max_features': 'sqrt',
'max_depth': 60,
'bootstrap': False}
```

4.3.3 *Comparação dos Resultados*

Os resultados do treinamento dos modelos revelaram que o classificador Random Forest se destacou em termos de desempenho. Ele demonstrou uma notável capacidade em distinguir os padrões relevantes para a identificação da doença de Parkinson nos exames de espirais. Esse destaque é evidenciado pela sua acurácia, que atingiu o valor de 86%, como mostrado no gráfico abaixo:

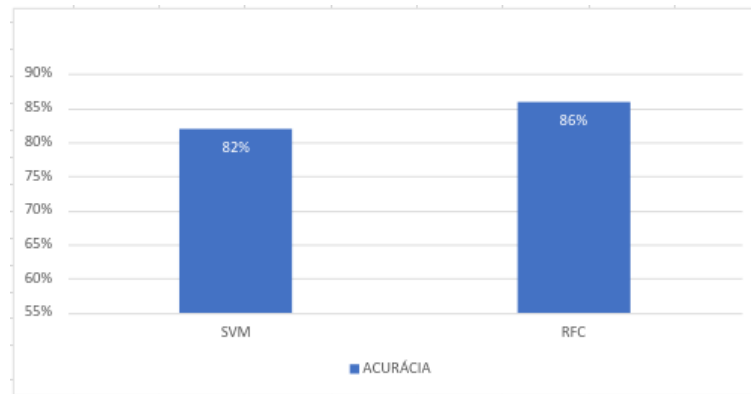


Figura 2 – Acurácia dos modelos utilizados

Além da acurácia, utilizamos outras métricas de validação para uma compreensão mais abrangente do desempenho dos modelos. Estas métricas incluem: Precision, Recall e F1-Score.

As Figura 3 apresentam o desempenho dessas métricas de validação para cada um dos classificadores testados na predição da doença de Parkinson.

		PRECISION	RECALL	F1-SCORE
SVM	Healthy	73%	100%	85%
	Parkinson	100%	64%	78%
RFC	Healthy	79%	100%	88%
	Parkinson	100%	73%	84%

Figura 3 – Métricas de validação

Em resumo, o gráfico e a tabela fornecem uma visão das diferenças no desempenho dos classificadores, destacando os valores do *Random Forest Classifier* como as melhores métricas encontradas.

5 CONCLUSÃO

A condução deste estudo permitiu explorar de forma abrangente a aplicação de técnicas de aprendizado de máquina na quantificação do sinal do Tremor Parkinsoniano por meio de exames manuscritos em espirais. Ao longo deste trabalho, foram abordados diversos aspectos, desde a revisão bibliográfica sobre a doença até a arquitetura empregada nos modelos, assim como a avaliação dos classificadores selecionados.

Foi possível vislumbrar a complexidade da doença de Parkinson, a problemática do seu diagnóstico e da sua avaliação subjetiva. Ficou evidente os ganhos que podem ser proporcionados pela utilização de ferramentas que permitem classificar os exames, especialmente para a população idosa afetada pela doença de Parkinson.

Ao explorar a arquitetura dos modelos, observamos que a combinação de técnicas de pré-processamento, como a padronização das imagens para uma escala uniforme, e a aplicação do Histograma de Gradiente Orientado (HOG) permitiu a extração eficaz de características relevantes dos exames manuscritos. Essas características, por sua vez, se mostraram cruciais para o desempenho dos classificadores.

Os resultados dos experimentos de validação e teste revelaram que o classificador *Random Forest* obteve a melhor performance entre os três algoritmos, alcançando uma acurácia de 86%. Esse resultado é promissor, pois indica que o modelo foi capaz de aprender padrões e relações relevantes nos exames manuscritos em espirais, contribuindo para uma identificação mais precisa de indícios da doença de Parkinson.

No entanto, é importante ressaltar que este estudo não está isento de limitações. A quantidade limitada de dados disponíveis para treinamento pode influenciar a generalização dos modelos. Além disso, considerando a natureza da doença de Parkinson, a utilização de outros tipos de dados e características poderia enriquecer ainda mais a análise. Desta forma, é fundamental destacar que este estudo não se destina a ser um método definitivo de diagnóstico da doença de Parkinson, mas sim uma ferramenta para uma compreensão mais aprofundada da doença.

Em suma, este estudo representa um passo importante na aplicação de técnicas de *Machine Learning* no auxílio da detecção da doença de Parkinson a partir de exames manuscritos em espirais. Este trabalho abre caminho para investigações futuras, visando a possibilidade de desenvolver um mecanismo de quantificação do sinal motor da DP. Tal avanço permitiria uma avaliação mais precisa e objetiva do progresso da doença, contribuindo assim para uma melhoria significativa na qualidade de vida dos pacientes afetados.

5.1 TRABALHOS FUTUROS

Diante dos bons resultados, o auxílio na detecção da doença de Parkinson por meio de exames manuscritos em espirais revela-se uma abordagem promissora. No entanto, é importante destacar que este estudo não se encerra aqui. Há espaço para expandir e refinar ainda mais essas técnicas. Abaixo são listados trabalhos que são vislumbrados como futuros:

- **Ampliação do Conjunto de Dados:** Até o momento, o treinamento dos modelos foi realizado com um conjunto de dados composto por 102 imagens. Para garantir a consistência do desempenho dos modelos, é essencial considerar a inclusão de um número de imagens significativamente maior e mais diversificado em futuras iterações deste estudo.
- **Utilização de Data Augmentation:** Uma estratégia a ser explorada é a implementação de técnicas de Data Augmentation. Isso envolve a criação de novas imagens a partir das existentes, aplicando operações como rotação, espelhamento e ajustes de brilho e contraste. A aplicação dessas técnicas pode aumentar consideravelmente o tamanho e a diversidade do conjunto de dados de treinamento. Como resultado, aprimora a capacidade dos modelos de aprendizado de máquina em lidar com uma ampla gama de variações nas imagens.
- **Treinamento dos Algoritmos com Formulários de Exames nos Padrões 'Waves' e 'Meander':** Cada modelo de *Machine Learning* será treinado separadamente utilizando conjuntos de dados específicos para os padrões 'Waves' e 'Meander'. Essa estratégia possibilitará uma avaliação mais abrangente do desempenho de cada modelo ao lidar com diferentes tipos de exames.

REFERÊNCIAS

- AARSLAND, D. et al. Predictors of Nursing Home Placement in Parkinson's Disease: A Population-Based, Prospective Study. *Journal of the American Geriatrics Society*, v. 48, n. 8, p. 938–942, ago. 2000. ISSN 0002-8614, 1532-5415. Disponível em: <https://agsjournals.onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.2000.tb06891.x>. Acesso em: 16 set. 2023. Citado na página 18.
- ABRAHAM, A. et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, v. 8, p. 14, 2014. Citado na página 25.
- ANDRADE, A. O. et al. Sinais e sintomas motores da doença de Parkinson: caracterização, tratamento e quantificação. *Novas tecnologias aplicadas à saúde: integração de áreas transformando a sociedade*, p. 282, 2017. Citado na página 13.
- ARMSTRONG, M. J.; OKUN, M. S. Diagnosis and Treatment of Parkinson Disease: A Review. *JAMA*, v. 323, n. 6, p. 548–560, fev. 2020. ISSN 0098-7484. Disponível em: <https://doi.org/10.1001/jama.2019.22360>. Acesso em: 12 ago. 2023. Citado na página 17.
- BAGNOLI, S. *Protein aggregation, ageing and neurodegeneration in the emerging model *Nothobranchius furzeri**. Tese (Doutorado) — Scuola Normale Superiore, set. 2021. Citado na página 12.
- BALESTRINO, R.; SCHAPIRA, A. Parkinson disease. *European Journal of Neurology*, v. 27, n. 1, p. 27–42, jan. 2020. ISSN 1351-5101, 1468-1331. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/ene.14108>. Acesso em: 09 out. 2023. Citado 3 vezes nas páginas 13, 17 e 18.
- BERG, D. et al. Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. *The Lancet Neurology*, v. 12, n. 5, p. 514–524, maio 2013. ISSN 14744422. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1474442213700474>. Acesso em: 30 set. 2023. Citado na página 18.
- BERG, D. et al. MDS research criteria for prodromal Parkinson's disease: MDS Criteria for Prodromal PD. *Movement Disorders*, v. 30, n. 12, p. 1600–1611, out. 2015. ISSN 08853185. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/mds.26431>. Acesso em: 19 ago. 2023. Citado na página 18.
- BERGSTR, J. et al. Algorithms for Hyper-Parameter Optimization. In: SHAW, TAYLOR, J. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2011. v. 24. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf. Acesso em: 19 ago. 2023. Citado na página 21.
- BERGSTR, J.; YOSHUA, B. Random search for hyper-parameter optimization. *Journal of machine learning research*, v. 13, n. 2, p. 281–305, fev. 2012. Citado na página 21.
- BERNSTEIN, D. Containers and Cloud: From LXC to Docker to Kubernetes. *IEEE Cloud Computing*, v. 1, n. 3, p. 81–84, 2014. Citado na página 23.

BIND, S. et al. A survey of machine learning based approaches for Parkinson disease prediction. *Int. J. Comput. Sci. Inf. Technol*, v. 6, n. 2, p. 1648–1655, 2015. Citado 2 vezes nas páginas 19 e 20.

BOETTIGER, C. An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, v. 49, n. 1, p. 71–79, jan. 2015. ISSN 0163-5980. Disponível em: <https://dl.acm.org/doi/10.1145/2723872.2723882>. Acesso em: 08 out. 2023. Citado na página 24.

BRAAK, H. et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, v. 24, n. 2, p. 197–211, mar. 2003. ISSN 01974580. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0197458002000659>. Acesso em: 26 ago. 2023. Citado na página 17.

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. ISSN 08856125. Disponível em: <http://link.springer.com/10.1023/A:1010933404324>. Acesso em: 30 set. 2023. Citado na página 20.

BRITAIN, N. C. C. for C. C. G. *Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care*. [S.l.]: Royal College of Physicians, 2006. ISBN 1860162835. Citado na página 18.

CHARCOT, J.-M. Lectures on the diseases of the nervous system, delivered at La Salpetriere. *London: The New Sydenham Society*, 1877. Citado na página 17.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. [S.l.: s.n.], 2005. v. 1, p. 886–893. Citado na página 20.

DECASTRO-GARCIA, N. et al. Effect of the Sampling of a Dataset in the Hyperparameter Optimization Phase over the Efficiency of a Machine Learning Algorithm. *Complexity*, v. 2019, p. 1–16, fev. 2019. ISSN 1076-2787, 1099-0526. Disponível em: <https://www.hindawi.com/journals/complexity/2019/6278908/>. Acesso em: 23 set. 2023. Citado na página 21.

Department of Computer Science and Engineering, The Maharaja Sayajirao University of Baroda, India et al. Deep Learning to Detect Skin Cancer using Google Colab. *International Journal of Engineering and Advanced Technology*, v. 8, n. 6, p. 2176–2183, ago. 2019. ISSN 22498958. Disponível em: <https://www.ijeat.org/portfolio-item/F8587088619/>. Acesso em: 12 ago. 2023. Citado na página 19.

DORSEY, E. R. et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, v. 68, n. 5, p. 384–386, jan. 2007. ISSN 0028-3878, 1526-632X. Disponível em: <https://www.neurology.org/lookup/doi/10.1212/01.wnl.0000247740.47667.03>. Acesso em: 08 out. 2023. Citado na página 12.

DROTAR, P. et al. Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. *Computer Methods and Programs in Biomedicine*, v. 117, n. 3, p. 405–411, 2014. ISSN 0169-2607. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169260714003204>. Acesso em: 30 set. 2023. Citado na página 15.

DROTAR, P. et al. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial Intelligence in Medicine*, v. 67, p. 39–46, fev. 2016. ISSN 09333657. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0933365716000063>. Acesso em: 30 set. 2023. Citado na página 15.

DUA, R.; RAJA, A. R.; KAKADIA, D. Virtualization vs Containerization to Support PaaS. In: *2014 IEEE International Conference on Cloud Engineering*. [S.l.: s.n.], 2014. p. 610–614. Citado na página 23.

EICHHORN, T. E. et al. Computational analysis of open loop handwriting movements in Parkinson's disease: A rapid method to detect dopamimetic effects. *Movement Disorders*, v. 11, n. 3, p. 289–297, maio 1996. ISSN 0885-3185, 1531-8257. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/mds.870110313>. Acesso em: 08 out. 2023. Citado na página 13.

FONSECA, G. Inacio Marques da. masterThesis, *Desenvolvimento de uma plataforma para monitoramento remoto de pacientes idosos*. 2009. Disponível em: <https://repositorio.ufpe.br/handle/123456789/2368>. Acesso em: 10 out. 2023. Citado na página 13.

GELB, D. J.; OLIVER, E.; GILMAN, S. Diagnostic Criteria for Parkinson Disease. *Archives of Neurology*, v. 56, n. 1, p. 33, jan. 1999. ISSN 0003-9942. Disponível em: <http://archneur.jamanetwork.com/article.aspx?doi=10.1001/archneur.56.1.33>. Acesso em: 12 ago. 2023. Citado na página 18.

GERHARDT, T. E. et al. *Métodos de pesquisa*. Ed. da UFRGS, 2009. ISBN 9788538600718. Disponível em: <https://lume.ufrgs.br/handle/10183/52806>. Acesso em: 16 set. 2023. Citado na página 22.

GOETZ, C. G.; STEBBINS, G. T. Risk factors for nursing home placement in advanced Parkinson's disease. *Neurology*, v. 43, n. 11, p. 2222–2222, nov. 1993. ISSN 0028-3878, 1526-632X. Disponível em: <https://www.neurology.org/lookup/doi/10.1212/WNL.43.11.2222>. Acesso em: 20 ago. 2023. Citado na página 18.

GREESHMA, K. V.; SREEKUMAR, K. K. Fashion-MNIST classification based on HOG feature descriptor using SVM. *International Journal of Innovative Technology and Exploring Engineering*, v. 8, n. 5, p. 960–962, 2019. Citado na página 20.

HELSLEY, M. LXC: Linux container tools. *IBM developerWorks Technical Library*, v. 11, 2009. Citado na página 23.

HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. (Ed.). *Automated Machine Learning: Methods, Systems, Challenges*. Cham: Springer International Publishing, 2019. (The Springer Series on Challenges in Machine Learning). ISBN 9783030053178 9783030053185. Disponível em: <http://link.springer.com/10.1007/978-3-030-05318-5>. Acesso em: 16 set. 2023. Citado na página 21.

ISENKUL, M.; SAKAR, B.; KURSUN, O. Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. *The 2nd international conference on e-health and telemedicine (ICEHTM-2014)*, v. 5, 2014. Citado na página 15.

ISHIHARA, L. S. et al. Estimated life expectancy of Parkinson's patients compared with the UK population. *Journal of Neurology, Neurosurgery & Psychiatry*, v. 78, n. 12, p. 1304–1309, dez. 2007. ISSN 0022-3050, 1468-330X. Disponível em: <https://jnnp.bmj.com/content/78/12/1304>. Acesso em: 23 set. 2023. Citado na página 17.

JIANG, X.; XU, C. Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data. *Journal of Clinical Medicine*, v. 11, n. 19, p. 5772, set. 2022. ISSN 2077-0383. Disponível em: <https://www.mdpi.com/2077-0383/11/19/5772>. Acesso em: 09 out. 2023. Citado na página 21.

KANANI, P.; PADOLE, M. Deep learning to detect skin cancer using google colab. *International Journal of Engineering and Advanced Technology Regular Issue*, v. 8, n. 6, p. 2176–2183, 2019. Citado na página 24.

LARSEN, K. et al. Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics*, v. 56, n. 3, p. 909–914, set. 2000. ISSN 0006341X. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.0006-341X.2000.00909.x>. Acesso em: 30 set. 2023. Citado na página 19.

MAHESH, B. Machine learning algorithms - a review. *International Journal of Science and Research (IJSR)*, v. 9, n. 1, p. 381–386, 2020. Citado na página 19.

MAJHI, B.; KASHYAP, A. Early Prediction of Parkinson's Disease Using Motor, Non-Motor Features and Machine Learning Techniques. In: *Deep Learning, Machine Learning and IoT in Biomedical and Health Informatics*. 1. ed. Boca Raton: CRC Press, 2022. p. 139–155. ISBN 9780367548445. Disponível em: <https://www.taylorfrancis.com/books/9780367548445/chapters/10.1201/9780367548445-10>. Acesso em: 12 ago. 2023. Citado na página 20.

MALHOTRA, N. K. Pesquisa de marketing: uma orientação aplicada. 3 .ed. Porto Alegre: Bookman, 2001. Citado na página 22.

MANTOVANI, R. G. et al. Effectiveness of Random Search in SVM hyper-parameter tuning. In: *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2015. p. 1–8. Citado na página 21.

MATHUR, R.; PATHAK, V.; BANDIL, D. Parkinson Disease Prediction Using Machine Learning Algorithm. In: RATHORE, V. S. et al. (Ed.). *Emerging Trends in Expert Applications and Security*. Singapore: Springer Singapore, 2019. v. 841, p. 357–363. ISBN 9789811322846 9789811322853. Disponível em: http://link.springer.com/10.1007/978-981-13-2285-3_42. Acesso em: 23 set. 2023. Citado na página 12.

MCKINNEY, W. pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing*, v. 14, n. 9, p. 1–9, 2011. Citado na página 25.

MENEZES, N. N. C. *Introdução à Programação com Python*. [S.l.]: Novatec Editora, 2022. ISBN 9788575222508. Citado na página 24.

MERKEL, D.; others. Docker: lightweight linux containers for consistent development and deployment. *Linux j*, v. 239, n. 2, p. 2, 2014. Citado na página 23.

MICHEL, V. et al. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, v. 45, n. 6, p. 2041–2049, jun. 2012. ISSN 00313203. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0031320311001439>. Acesso em: 12 ago. 2023. Citado na página 25.

NATARAJAN, S. et al. An analysis of container-based platforms for nfv. *IETF draft, Apr*, 2016. Citado na página 23.

NELLI, F. *Python Data Analytics: With Pandas, NumPy, and Matplotlib*. Berkeley, CA: Apress, 2018. ISBN 9781484239124 9781484239131. Disponível em: <http://link.springer.com/10.1007/978-1-4842-3913-1>. Acesso em: 10 out. 2023. Citado 2 vezes nas páginas 25 e 26.

NETWORK, S. I. G. *Diagnosis and Pharmacological Management of Parkinson's disease: A National Clinical Guideline*. [S.l.]: Edinburgh: Scottish Intercollegiate Guidelines Network, 2010. Citado na página 18.

NEWMAN, E. J. et al. Accuracy of Parkinson's disease diagnosis in 610 general practice patients in the West of Scotland. *Movement Disorders*, p. NA–NA, 2009. ISSN 08853185, 15318257. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/mds.22829>. Acesso em: 05 ago. 2023. Citado na página 18.

NumPy Developers. *NumPy documentation*. Disponível em: <https://numpy.org/doc/1.26/>. Acesso em: 30 set. 2023. Citado na página 25.

OBESO, J. et al. Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy. *Movement Disorders*, v. 32, n. 9, p. 1264–1310, set. 2017. ISSN 0885-3185, 1531-8257. Disponível em: <https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mds.27115>. Acesso em: 27 ago. 2023. Citado na página 18.

OLIPHANT, T. E. Python for Scientific Computing. *Computing in Science & Engineering*, v. 9, n. 3, p. 10–20, 2007. Citado na página 25.

OLOF, S. S. A Comparative Study of Black-box Optimization Algorithms for Tuning of Hyper-parameters in Deep Neural Networks. *Dissertation*, 2018. Citado na página 21.

PAN, S. et al. Parkinson's Disease tremor classification – A comparison between Support Vector Machines and neural networks. *Expert Systems with Applications*, v. 39, n. 12, p. 10764–10771, set. 2012. ISSN 09574174. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0957417412004629>. Acesso em: 26 ago. 2023. Citado na página 13.

PAPA, J. P. et al. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, v. 45, n. 1, p. 512–520, jan. 2012. ISSN 00313203. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0031320311003013>. Acesso em: 12 ago. 2023. Citado na página 13.

PAPA, J. P.; FALCÃO, A. X.; SUZUKI, C. T. N. Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, v. 19, n. 2, p. 120–131, jun. 2009. ISSN 0899-9457, 1098-1098. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/ima.20188>. Acesso em: 12 ago. 2023. Citado na página 13.

PARKINSON, J. An essay on the shaking palsy. *London: Whittingham and Rowland Sherwood, Neely and Jones*, 1817. Citado na página 17.

Parkinson's Foundation. *Parkinson's Foundation - Statistics*. 2018. Disponível em: <https://www.parkinson.org/understanding-parkinsons/statistics>. Acesso em: 17 ago. 2023. Citado na página 12.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, v. 12, p. 2825–2830, 2011. Citado na página 25.

PEREIRA, C. R. Aprendizado de máquina aplicado ao auxílio do diagnóstico da doença de Parkinson. jul. 2017. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/9299>. Acesso em: 23 set. 2023. Citado 2 vezes nas páginas 12 e 14.

PEREIRA, C. R. et al. A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, v. 136, p. 79–88, nov. 2016. ISSN 01692607. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0169260716301894>. Acesso em: 30 set. 2023. Citado na página 14.

PEREIRA, C. R. et al. A Step Towards the Automated Diagnosis of Parkinson's Disease: Analyzing Handwriting Movements. In: *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*. [S.l.: s.n.], 2015. p. 171–176. Acesso em: 23 set. 2023. Citado na página 16.

PEREIRA, D.; GARRETT, C. Risk factors for Parkinson disease: an epidemiologic study. *Acta Médica Portuguesa*, v. 23, n. 1, p. 15–24, fev. 2010. Disponível em: <https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/599>. Acesso em: 30 set. 2023. Citado na página 12.

PEREIRA, J. S. Distúrbio respiratório na doença de Parkinson. *Fisioterapia Brasil*, v. 1, n. 1, p. 23–26, 2000. Citado na página 12.

PEREZ, F.; GRANGER, B. E. IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, v. 9, n. 3, p. 21–29, 2007. Citado na página 24.

PINHEIRO, J. E. S. et al. *Tratado de geriatria e gerontologia*. Rio de Janeiro: [s.n.], 2006. Citado na página 12.

RANDLES, B. M. et al. Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. In: *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. [S.l.: s.n.], 2017. p. 1–2. Citado na página 24.

ROSENBLUM, S. et al. Handwriting as an objective tool for Parkinson's disease diagnosis. *Journal of Neurology*, v. 260, n. 9, p. 2357–2361, set. 2013. ISSN 0340-5354, 1432-1459. Disponível em: <http://link.springer.com/10.1007/s00415-013-6996-x>. Acesso em: 23 set. 2023. Citado na página 15.

RULE, A.; TABARD, A.; HOLLAN, J. D. Exploration and Explanation in Computational Notebooks. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, 2018. p. 1–12. ISBN 9781450356206. Disponível em: <https://dl.acm.org/doi/10.1145/3173574.3173606>. Acesso em: 09 out. 2023. Citado na página 24.

- SAMPIERI, R. H.; COLLADO, R. C. F.; LUCIO, P. B. *Metodología de la Investigación*. México: Ed. Me Graw-Hill Interamericana Editores, 1991. Citado na página 22.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM: J Res Dev.*, v. 3, p. 210–29, 1959. Citado na página 19.
- SCHRAG, A.; BEN-SHLOMO, Y.; QUINN, N. How valid is the clinical diagnosis of Parkinson’s disease in the community? *Journal of Neurology, Neurosurgery & Psychiatry*, v. 73, n. 5, p. 529–534, nov. 2002. ISSN 0022-3050, 1468-330X. Disponível em: <https://jnnp.bmj.com/content/73/5/529>. Acesso em: 24 set. 2023. Citado na página 18.
- SHEN, H. Interactive notebooks: Sharing the code. *Nature*, v. 515, n. 7525, p. 152–152, nov. 2014. ISSN 0028-0836, 1476-4687. Disponível em: <https://www.nature.com/articles/515151ax>. Acesso em: 23 set. 2023. Citado na página 24.
- SILVA, K. C. N. D. *Monitoramento da saúde humana através de sensores: análise de incertezas contextuais através da teoria da evidência de Dempster-Shafer*. Tese (Doutorado em Sistemas Digitais) — Universidade de São Paulo, São Paulo, nov. 2012. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-01082013-160349/>. Acesso em: 23 set. 2023. Citado na página 13.
- SOLEIMANIANGHAREHCHOPOGH, F.; MOHAMMADI, P. A Case Study of Parkinson’s Disease Diagnosis using Artificial Neural Networks. *International Journal of Computer Applications*, v. 73, n. 19, p. 1–6, jul. 2013. ISSN 09758887. Disponível em: <http://research.ijcaonline.org/volume73/number19/pxc3889206.pdf>. Acesso em: 17 ago. 2023. Citado na página 13.
- SOUZA, C. F. M. et al. A Doença de Parkinson e o Processo de Envelhecimento Motor. *Revista Neurociências*, v. 19, n. 4, p. 718–723, mar. 2001. ISSN 1984-4905, 0104-3579. Disponível em: <https://periodicos.unifesp.br/index.php/neurociencias/article/view/8330>. Acesso em: 02 set. 2023. Citado na página 12.
- SOUZA, H.; MEDEIROS, L. M. de; OLIVEIRA, M. C. Métodos de Aprendizagem de Máquina no auxílio à classificação da marcha em parkinsonianos. *XXV Congresso Brasileiro de Engenharia Biomédica*, p. 1725–1728, 2016. Citado na página 13.
- SPADOTO, A. A. et al. Improving Parkinson’s disease identification through evolutionary-based feature selection. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.: s.n.], 2011. p. 7857–7860. Citado na página 13.
- SPADOTO, A. A. et al. Parkinson’s disease identification through optimum-path forest. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. [S.l.: s.n.], 2010. p. 6087–6090. Citado na página 13.
- STERN, M. B.; LANG, A.; POEWE, W. Toward a redefinition of Parkinson’s disease. *Movement Disorders*, v. 27, n. 1, p. 54–60, jan. 2012. ISSN 08853185. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/mds.24051>. Acesso em: 09 out. 2023. Citado na página 18.
- SURASAK, T. et al. Histogram of oriented gradients for human detection in video. In: *2018 5th International Conference on Business and Industrial Research (ICBIR)*. [S.l.: s.n.], 2018. p. 172–176. Citado na página 20.

SVEC, J. G.; POPOLO, P. S.; TITZE, I. R. Measurement of vocal doses in speech: experimental procedure and signal processing. *Logopedics Phoniatrics Vocology*, v. 28, n. 4, p. 181–192, out. 2003. ISSN 1401-5439, 1651-2022. Disponível em: <http://www.tandfonline.com/doi/full/10.1080/14015430310018892>. Acesso em: 09 out. 2023. Citado na página 19.

WATTERS, P. A.; PATEL, M. [No title found]. *Neural Processing Letters*, v. 9, n. 2, p. 189–199, 1999. ISSN 13704621. Disponível em: <http://link.springer.com/10.1023/A:1018637711183>. Acesso em: 08 out. 2023. Citado na página 13.

WELCOME to Colaboratory! Disponível em: <https://colab.research.google.com/notebooks/welcome.ipynb#scrollTo=-Rh3-Vt9Nev9>. Acesso em: 10 out. 2023. Citado na página 24.

WILLIS, A. W. et al. Geographic and Ethnic Variation in Parkinson Disease: A Population-Based Study of US Medicare Beneficiaries. *Neuroepidemiology*, v. 34, n. 3, p. 143–151, 2010. ISSN 0251-5350, 1423-0208. Disponível em: <https://www.karger.com/Article/FullText/275491>. Acesso em: 08 out. 2023. Citado na página 12.

XU, J. et al. Parkinson's disease and risk of mortality: meta-analysis and systematic review. *Acta Neurologica Scandinavica*, v. 129, n. 2, p. 71–79, fev. 2014. ISSN 00016314. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/ane.12201>. Acesso em: 09 out. 2023. Citado na página 17.

YANG, L.; SHAMI, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, v. 415, p. 295–316, nov. 2020. ISSN 09252312. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0925231220311693>. Acesso em: 09 out. 2023. Citado na página 21.

ZHAM, P. et al. Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral. *Frontiers in Neurology*, v. 8, p. 435, set. 2017. ISSN 1664-2295. Disponível em: <http://journal.frontiersin.org/article/10.3389/fneur.2017.00435/full>. Acesso em: 17 out. 2023. Citado na página 27.

ZHOU, W. et al. Histogram of Oriented Gradients Feature Extraction From Raw Bayer Pattern Images. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 67, n. 5, p. 946–950, 2020. Citado na página 20.

ZUCCO, F. *FisioWeb - A atuação do Fisioterapeuta na Patologia de Parkinson*. 2009. Disponível em: http://www.wgate.com.br/conteudo/medicinaesaudefisioterapia/neuro/parkinson_fabiola.htm. Acesso em: 10 ago. 2023. Citado na página 12.