



UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA LUSOFONIA AFRO-BRASILEIRA
INSTITUTO DE ENGENHARIAS E DESENVOLVIMENTO SUSTENTÁVEL
ENGENHARIA DE COMPUTAÇÃO

IVINA LORENA OLIVEIRA MOURA

PARKIN95: UMA PLATAFORMA DE APOIO AO DIAGNÓSTICO DE PARKINSON
INTEGRANDO MODELO PREDITIVO E INTERFACE CLÍNICA

REDEÇÃO – CEARÁ
2025

IVINA LORENA OLIVEIRA MOURA

**PARKIN95: UMA PLATAFORMA DE APOIO AO DIAGNÓSTICO DE PARKINSON
INTEGRANDO MODELO PREDITIVO E INTERFACE CLÍNICA**

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Engenharia de Computação, da Universidade da Integração Internacional da Lusofonia Afro-Brasileira, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Antonio Alisson Pessoa Guimarães

Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Sistema de Bibliotecas da UNILAB
Catalogação de Publicação na Fonte.

Moura, Ivina Lorena Oliveira.

M929p

Parkin95: uma plataforma de apoio ao diagnóstico de parkinson integrando modelo preditivo e interface clínica / Ivina Lorena Oliveira Moura. - Redenção, 2025.
65f: il.

Monografia - Curso de Engenharia de Computação, Instituto de Engenharias e Desenvolvimento Sustentável, Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, 2025.

Orientador: Prof. Dr. Antonio Alisson Pessoa Guimarães.

1. Doença de Parkinson. 2. Aprendizado de máquina. 3. Regressão logística regularizada. 4. Análise de voz. 5. Suporte à decisão Clínica. I. Título

CE/UF/BSCA

CDD 616.833

IVINA LORENA OLIVEIRA MOURA

**PARKIN95: UMA PLATAFORMA DE APOIO AO DIAGNÓSTICO DE PARKINSON
INTEGRANDO MODELO PREDITIVO E INTERFACE CLÍNICA**

Trabalho de Conclusão de Curso (TCC), na modalidade monografia, apresentado ao curso de Engenharia de Computação, da Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Campus das Auroras, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Data de aprovação: 16 de maio de 2025.

BANCA EXAMINADORA

Prof. Dr. Antonio Alisson Pessoa Guimarães (Orientador)

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

Profa. Dra. Marcia Roberta Falcão de Farias

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

Profa. Dra. Deborah Maria Vieira Magalhães

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

Prof. Dr. José Antônio Fernandes de Macêdo

Universidade Federal do Ceará (UFC)

Dedico este trabalho às pessoas que contribuíram para o meu crescimento pessoal, acadêmico e profissional. Em especial, agradeço aos meus pais, Nágila Moura e Júnior Moura, por me ensinarem o valor do estudo, e aos meus irmãos Virna Raquel, Eduardo Lincoln e Luis Neto, pelo apoio constante.

AGRADECIMENTOS

Meus sinceros agradecimentos aos meus pais, Júnior Moura e Nágila Moura, pelo esforço incansável e por me proporcionarem o privilégio de dedicar-me aos estudos e aproveitar as oportunidades que hoje tenho. Agradeço também aos meus irmãos, Virna Raquel, Eduardo Lincoln e Luis Neto, por estarem sempre ao meu lado, tornando a caminhada mais leve.

À tia Daniele, por estar sempre ao meu lado com conselhos, apoio e incentivo. Sou imensamente grata por acreditar no meu potencial, mesmo nos momentos em que eu mesma duvidei.

Aos meus amigos, que tive a sorte de encontrar ao longo do caminho e que tornaram a jornada acadêmica mais leve, Isaak Yuri, Artur Zico, Carlos Eduardo, Jorge Antônio, Alexandre, Thiago Queiroz, José Dembo, Gabriel Paixão e Misael Bessa. Em especial, àqueles que sempre contribuíram para o meu crescimento pessoal e acadêmico, ajudando-me a evoluir em cada etapa da jornada, Alcides Saraiva, Samuel Rodrigues, Álvaro Farias, Ithalo Luz e Lizandra Régia. Não há palavras que expressem plenamente minha gratidão por cada momento vivido ao lado de vocês.

Agradeço aos professores da instituição pelo conhecimento compartilhado, pela dedicação e pelo papel essencial na minha formação. Em especial, ao professor Alisson Guimarães, pela orientação, disponibilidade e por sempre me ensinar com paciência ao longo desses anos de trabalho em conjunto. Agradeço por sempre ter confiado em mim. Sua orientação fez toda a diferença.

Estendo também minha gratidão a todos os familiares que, mesmo não mencionados diretamente, sempre estiveram presentes com carinho, apoio e palavras de incentivo ao longo da minha caminhada.

*“Apesar de muito moço
me sinto são, e salvo, e forte”.*
(Belchior)

RESUMO

A Doença de Parkinson (DP) é uma enfermidade neurodegenerativa crônica que afeta o sistema nervoso central, comprometendo de forma significativa a função motora. Entre os principais sintomas destacam-se a acinesia (perda de movimentos voluntários), bradicinesia (lentidão motora), rigidez muscular e tremor de repouso. Além dessas manifestações motoras, é comum a presença de redução da expressão facial, instabilidade postural e alterações na fala, como disfonia e disartria. No âmbito emocional e cognitivo, os pacientes frequentemente apresentam quadros de depressão, ansiedade, sintomas psicóticos, prejuízos cognitivos e, em estágios mais avançados, demência. Com o avanço da tecnologia, técnicas de inteligência artificial — especialmente aquelas baseadas em aprendizado de máquina (Machine Learning) — têm sido amplamente estudadas como ferramentas auxiliares no diagnóstico da DP. Neste contexto, o presente trabalho contempla duas vertentes principais: a primeira consiste no desenvolvimento e validação de um modelo computacional baseado em Regressão Logística Regularizada (RLReg), treinado com amostras de voz para classificar indivíduos com ou sem a doença; a segunda refere-se à integração desse modelo à interface Parkin95 — uma aplicação desenvolvida para uso clínico por neurologistas, com o objetivo de apoiar a tomada de decisão diagnóstica. Em síntese, em fase de teste, o modelo alcançou 95% de acurácia, com recall de 100% para casos positivos, ou seja, identificou todos os pacientes com Parkinson sem falsos negativos. Na sequência, em decorrência dos resultados obtidos, desenvolveu-se a plataforma web interativa (Parkin95), implementada em Streamlit, a qual possibilita aos neurologistas inserir dados de pacientes, a partir de amostras de voz, e visualizar previsões em tempo real, otimizando o fluxo clínico. Como proposta futura, considera-se a expansão do sistema para integrar outros tipos de dados, como exames de imagem e sinais motores, além da migração do trabalho para um framework mais robusto, com o intuito de aprimorar a precisão do diagnóstico e o monitoramento da progressão da doença.

Palavras-chave: Doença de Parkinson, Aprendizado de Máquina, Regressão Logística Regularizada, Análise de Voz, Suporte à Decisão Clínica, Inteligência Artificial.

ABSTRACT

Parkinson's Disease (PD) is a chronic neurodegenerative disorder that affects the central nervous system, significantly impairing motor function. The main symptoms include akinesia (loss of voluntary movement), bradykinesia (slowness of movement), muscle rigidity, and resting tremor. In addition to these motor symptoms, it is common to observe reduced facial expression, postural instability, and speech impairments such as dysphonia and dysarthria. On the emotional and cognitive fronts, patients often experience depression, anxiety, psychotic symptoms, cognitive impairments, and, in more advanced stages, dementia. With technological advancements, artificial intelligence techniques—particularly those based on machine learning—have been widely studied as auxiliary tools in the diagnosis of PD. In this context, the present work encompasses two main aspects: the first involves the development and validation of a computational model based on Regularized Logistic Regression (RLReg), trained with voice samples to classify individuals as either affected or unaffected by the disease; the second refers to the integration of this model into Parkin95, an application developed for clinical use by neurologists, aiming to support diagnostic decision-making. In summary, during the testing phase, the model achieved 95% accuracy, with a 100% recall for positive cases, meaning it identified all patients with Parkinson's disease without any false negatives. Following these results, the Parkin95 interactive web platform was developed, implemented using Streamlit, allowing neurologists to input patient data from voice samples and visualize real-time predictions, thus optimizing clinical workflows. As a future proposal, the system is expected to be expanded to incorporate other types of data, such as imaging exams and motor signals, in addition to migrating the solution to a more robust framework, with the aim of enhancing diagnostic accuracy and monitoring disease progression.

Keywords: Parkinson's Disease, Machine Learning, Regularized Logistic Regression, Voice Analysis, Clinical Decision Support, Artificial Intelligence.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama da substância negra	19
Figura 2 – Classificação em árvore	28
Figura 3 – Classificação com Máquinas de Vetores de Suporte (SVM)	29
Figura 4 – Diagrama do K-NN	29
Figura 5 – Representação da Função Logística e sua Transformação Logit	30
Figura 6 – Diagrama do padrão de arquitetura MVC (Model-View-Controller)	37
Figura 7 – Fluxograma do processo de desenvolvimento de um modelo de Machine Learning, desde a definição do problema até a implantação	39
Figura 8 – Resultado gerado no console após a execução do algoritmo e classificação dos pacientes com base no banco de dados genérico.	43
Figura 9 – Colunas com o resultado pós sigmoide e sua classificação	44
Figura 10 – Exemplo do que era exibido no console	45
Figura 11 – Avaliações parciais obtidas durante o treinamento do modelo para cada vocábulo analisado.	48
Figura 12 – Comportamento do erro de diagnóstico por vocábulo	49
Figura 13 – Tela inicial do Parkin95	51
Figura 14 – Tela inicial com o sidebar	52
Figura 15 – Áreas de navegação	53
Figura 16 – Página Área de pacientes	54
Figura 17 – Tela com informações institucionais e acadêmicas dos responsáveis pelo projeto.	55
Figura 18 – Formulário para preenchimento dos dados do paciente	56
Figura 19 – Área dos pacientes já cadastrados	56
Figura 20 – Resultado exibido na tela para o usuário	57

LISTA DE TABELAS

Tabela 1 – Exemplo genérico de uma matriz confusão.	32
Tabela 2 – Matriz de confusão resultante após a remoção dos vocábulos mais associados a erros de diagnóstico.. . . .	49
Tabela 3 – Métricas de desempenho do modelo de Regressão Logística Regularizada	50

SUMÁRIO

1	INTRODUÇÃO	13
1.1	DIAGNÓSTICO	13
1.2	DIAGNÓSTICO DE DP POR FERRAMENTAS DE IA	15
1.3	OBJETIVOS	18
1.3.1	Objetivo geral	18
1.3.2	Objetivos específicos	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	A DOENÇA DE PARKINSON	19
2.1.1	Sintomas	20
2.1.2	Tratamentos e fármacos	21
2.1.3	Ferramentas sobre o monitoramento e diagnóstico da DP	21
2.2	ANÁLISE DE VOZ NA MEDICINA	22
2.3	APRENDIZADO DE MÁQUINA	24
2.3.1	Métodos de aprendizagem de máquina	25
2.3.2	Algoritmos comuns de classificação e de regressão	26
<i>2.3.2.1</i>	<i>Árvores de decisão</i>	<i>27</i>
2.4	REGRESSÃO LOGÍSTICA REGULARIZADA (RL-REG)	29
2.4.1	Detalhamento da transformação logística	30
2.4.2	Métricas de desempenho para classificação	32
2.5	TECNOLOGIAS PARA ANÁLISE DE DADOS E DESENVOLVIMENTO DE APLICAÇÕES	35
2.6	PADRÃO DE ARQUITETURA: MVC	36
3	METODOLOGIA	38
3.1	PROCESSO DE CONSTRUÇÃO DE UMA APLICAÇÃO DE APRENDIZADO DE MÁQUINA	38
3.1.1	Tratamento dos dados	39
3.1.2	Fase de treinamento/teste	42
3.1.3	Comitê avaliativo	42
3.1.4	Teste e validação	43
3.2	A INTERFACE GRÁFICA PARKIN95	45
3.2.1	Função e objetivo da aplicação	45
3.2.2	Tecnologias utilizadas	45
4	RESULTADOS E DISCUSSÕES	47
4.1	RESULTADOS DO TREINAMENTO DO MODELO	47
4.2	RESULTADOS DA INTERFACE PARKIN95	50

4.2.1	Funcionalidades	55
5	CONCLUSÕES E PERSPECTIVAS FUTURAS	58
	REFERÊNCIAS	60

1 INTRODUÇÃO

A Doença de Parkinson (DP) é uma condição neurológica crônica e progressiva que afeta o sistema nervoso central, comprometendo diretamente o controle motor. Seus sintomas surgem de forma gradual e incluem tremores involuntários, rigidez muscular, bradicinesia (lentidão dos movimentos) e instabilidade postural. Além disso, alterações na fala e na escrita são frequentemente observadas em pacientes. A causa principal está relacionada à degeneração de neurônios na substância negra, uma região cerebral responsável pela produção de dopamina — neurotransmissor essencial para a regulação dos movimentos voluntários (Sakar *et al.*, 2013).

Segundo a Organização Mundial da Saúde (Organização Mundial da Saúde, 2022), a Doença de Parkinson é a segunda condição neurodegenerativa mais prevalente globalmente, ficando atrás apenas do Alzheimer. Estima-se que aproximadamente 4 milhões de pessoas convivam com a doença em todo o mundo, sendo cerca de 200 mil no Brasil.

A doença recebeu esse nome em homenagem ao médico inglês James Parkinson, que descreveu seus sintomas pela primeira vez em 1817 (Hayes, 2019). Ela é considerada a segunda doença neurodegenerativa mais comum em pessoas idosas, com evolução crônica e progressiva (Chou, 2020). Segundo o estudo *Global Burden of Disease de 2016*, a Doença de Parkinson (DP) demonstra o crescimento mais rápido em prevalência, incapacidades e mortalidade entre os distúrbios neurológicos (Dorsey; Elbaz, 2018).

Embora a causa da Doença de Parkinson (DP) permaneça desconhecida, pesquisas indicam uma combinação de suscetibilidade genética e exposição a toxinas ambientais. A genética contribui com variantes que predispõem à doença, enquanto fatores ambientais, como a exposição contínua a pesticidas e herbicidas (comumente em áreas rurais), sugerem um vínculo com o estresse oxidativo induzido por químicos industriais (Cabreira; Massano, 2019; Balestrino; Schapira, 2019).

1.1 DIAGNÓSTICO

Os médicos diagnosticam a Doença de Parkinson conversando com o paciente e fazendo um exame físico completo. Não há exames de sangue, marcadores ou imagens que confirmem o diagnóstico de forma definitiva (Merritt, H., 2002; Moreira *et al.*, 2007).

A utilização de exames radiológicos cerebrais de rotina oferece pouca ajuda na identificação da Doença de Parkinson em relação a outras síndromes parkinsonianas

(Lang; Lozano, 1998). Segundo Rajput e Noyes (2024), o diagnóstico da Doença de Parkinson é predominantemente clínico, baseado na presença de sinais motores característicos, como tremor de repouso, bradicinesia e rigidez muscular. A suspeita geralmente surge quando se observa tremor unilateral em repouso, que tende a diminuir ou desaparecer durante a realização do teste dedo-nariz. Durante o exame neurológico, nota-se dificuldade na execução de movimentos alternados ou de rápida sucessão. A sensibilidade e a força muscular costumam permanecer preservadas, assim como os reflexos, embora estes possam ser de difícil elicitacão em função da rigidez ou tremor intensos.

Segundo Moreira *et al.* (2007), as manifestações motoras da Doença de Parkinson correspondem às seis características fundamentais tradicionalmente associadas ao parkinsonismo, são elas:

- Tremor em repouso a uma frequência de 4-5Hz.
- Rigidez, com maior resistência ao movimento passivo.
- Anormalidades posturais.
- Bradicinesia (lentidão dos movimentos, perda dos movimentos automáticos) e hipocinesia (diminuição da amplitude dos movimentos, especialmente os movimentos repetitivos).
- Perda dos reflexos posturais.
- Congelamento – Incapacidade transitória de executar movimentos ativos (bloqueio motor).

Em relação ao tratamento da doença de Parkinson vários medicamentos orais são comumente utilizados para aliviar os sintomas (Fahn *et al.*, 2004; Rajput; Noyes, 2024). Alguns tratamentos medicamentosos:

- Carbidopa/levodopa: É o tratamento padrão-ouro da DP. A levodopa é convertida em dopamina no cérebro, e a carbidopa evita que essa conversão ocorra fora do cérebro, reduzindo efeitos colaterais.
- Amantadina: Um antiviral com efeito antiparkinsoniano moderado. Pode ajudar a reduzir discinesias (movimentos involuntários causados por uso prolongado de levodopa).

- Inibidores de MAO-B: Impedem a degradação da dopamina no cérebro, prolongando seus efeitos.
- Anticolinérgicos: Utilizados principalmente para controlar tremores. Têm uso limitado devido a efeitos colaterais cognitivos, especialmente em idosos.
- Agonistas dopaminérgicos: Imitam a dopamina e são usados sozinhos ou com levodopa. Podem causar efeitos colaterais como sonolência ou distúrbios de controle de impulsos.
- Inibidores de COMT: Prolongam o efeito da levodopa ao impedir sua degradação periférica. Sempre usados em associação com levodopa, especialmente quando os sintomas flutuam.

E tratamentos não medicamentosos:

- Cirurgia: Indicada quando os medicamentos não controlam mais bem os sintomas ou causam efeitos adversos significativos.
- Exercícios e medidas adaptativas: Fisioterapia, terapia ocupacional e fonoaudiologia ajudam a manter a mobilidade, equilíbrio e fala. Importantes para qualidade de vida.

1.2 DIAGNÓSTICO DE DP POR FERRAMENTAS DE IA

Estamos vivendo uma nova revolução industrial, impulsionada pelo avanço de tecnologias emergentes, principalmente pela Inteligência Artificial (IA). As máquinas deixaram de executar apenas tarefas manuais e passaram a realizar também atividades que exigem raciocínio, assumindo funções tradicionalmente associadas à inteligência humana. Uma das áreas com maior crescimento na aplicação da automação é o diagnóstico médico assistido por computador. Atualmente, sistemas baseados em Inteligência Artificial têm alcançado altos níveis de acurácia, sendo em alguns casos comparáveis ou até superiores aos obtidos por profissionais de saúde humanos, especialmente em tarefas específicas como análise de imagens médicas (Ludermir, 2021). Um exemplo é o robô desenvolvido pela empresa *iFlytek*, que se destacou por ser aprovado no exame nacional para licenciamento de médicos da China. Conhecido como “Médico Assistente Robô”, esse sistema é capaz de registrar sintomas, analisar imagens de tomografia computadorizada e fornecer um diagnóstico inicial. Embora não tenha como objetivo substituir os médicos, ele foi projetado para auxiliá-los e aumentar a eficiência no atendimento (Saracco, 2017).

A aplicação da Inteligência Artificial (IA) no diagnóstico de doenças tem se mostrado um campo promissor, oferecendo avanços significativos, embora ainda enfrente desafios no desenvolvimento. Entre os principais benefícios destacam-se a capacidade de processar grandes volumes de dados de forma rápida, a padronização das análises e o potencial para diagnósticos mais precoces e precisos. Um exemplo é uma ferramenta de IA que demonstrou capacidade para detectar a Doença de Parkinson com até 15 anos de antecedência, alcançando 96% de acurácia na identificação dos casos (Lennon, 2023). Outra linha de pesquisa que demonstra o potencial da Inteligência Artificial para o diagnóstico precoce da Doença de Parkinson é a análise de sinais eletroencefalográficos (EEG), que permite identificar alterações neurológicas sutis ainda nas fases iniciais da doença (Silva *et al.*, 2023). Além disso, a IA pode contribuir para a redução da variabilidade entre especialistas humanos na interpretação de exames médicos. O software AIDP, por exemplo, obteve alta precisão na diferenciação de formas de parkinsonismo por meio da análise automatizada de imagens de ressonância magnética, demonstrando o potencial da IA para oferecer diagnósticos mais consistentes e confiáveis (Vaillancourt *et al.*, 2025). Outra vantagem importante diz respeito à não invasividade e acessibilidade dos métodos baseados em Inteligência Artificial, como apontam (Moraes *et al.*, 2023). Muitas dessas abordagens utilizam dados obtidos por meios não invasivos, como a análise de padrões vocais ou a captação de informações por dispositivos vestíveis. Isso permite que o processo diagnóstico ocorra de forma mais confortável ao paciente e, ao mesmo tempo, amplia seu alcance, já que não exige necessariamente procedimentos clínicos complexos ou equipamentos de alto custo.

Apesar do grande potencial da Inteligência Artificial (IA) na área da saúde, uma das principais desvantagens ainda reside na dificuldade de interpretabilidade dos modelos, especialmente os mais complexos, como as redes neurais profundas. Como apontado por Sadeghi *et al.* (2024), a falta de explicabilidade pode comprometer a confiança dos profissionais de saúde e dificultar a adoção clínica de sistemas automatizados. Essa limitação é especialmente crítica em contextos que exigem decisões transparentes e auditáveis, como o diagnóstico médico, onde compreender os fatores que levaram a determinada predição é tão importante quanto a própria acurácia do modelo. Outro aspecto relevante são as questões éticas e de privacidade. A utilização de dados de saúde em IA requer conformidade com regulamentações de privacidade, como o GDPR (Regulamento Geral de Proteção de Dados) e a HIPAA (Lei de Portabilidade e Responsabilidade de Seguro de Saúde), o que levanta preocupações éticas sobre consentimento informado e o uso secundário de dados. Como destacado por Marengo *et al.* (2023), o entusiasmo pela inovação tecnológica frequentemente entra em conflito com as questões regulatórias e éticas, especialmente no que se refere à proteção de dados, à privacidade, ao acesso a dispositivos móveis e às disparidades tecnológicas e

sociais que podem surgir. O artigo também enfatiza o risco de dependência tecnológica e os potenciais impactos dessa dependência na relação entre médicos e pacientes, além de questionar a autonomia do paciente ao depender de decisões automatizadas, o que pode comprometer a confiança e a qualidade do cuidado.

Diante desse cenário, no qual o monitoramento de sintomas tem avançado, mas ainda existem limitações quanto ao suporte direto ao diagnóstico da Doença de Parkinson, identificamos a necessidade de investigar métodos que possam auxiliar na tomada de decisão médica durante a avaliação clínica do paciente. Essa lacuna evidencia a importância de desenvolver ferramentas complementares que integrem dados objetivos e algoritmos preditivos, contribuindo para diagnósticos mais precisos e precoces.

Vários estudos têm abordado a aplicação de aprendizado de máquina no diagnóstico precoce da Doença de Parkinson, com foco na utilização de sinais de fala, dados motores e outros parâmetros clínicos. [Yadav e Ghosh \(2020\)](#) propõem uma estrutura de aprendizado de máquina inovadora para a previsão da Doença de Parkinson, utilizando dados de fala e movimento para o diagnóstico. [Coudray *et al.* \(2021\)](#) revisam modelos de aprendizado de máquina aplicados ao diagnóstico precoce, destacando diversas abordagens, incluindo o uso de sinais motores, imagens e outros dados clínicos. Apesar de os estudos mencionados abordarem uma gama mais ampla de dados, a proposta desenvolvida neste trabalho adota uma abordagem mais focada, utilizando exclusivamente dados de amostras de voz para a detecção da Doença de Parkinson. Essa abordagem permite explorar de maneira mais profunda as características vocais associadas à doença e implementar um classificador baseado em regressão logística regularizada, sem recorrer a outros tipos de dados. Essa especialização no uso de amostras de voz visa aprimorar a precisão do diagnóstico, ao mesmo tempo em que torna o processo mais simples e acessível.

O restante do trabalho está organizado da seguinte forma. O Capítulo 2 introduz os princípios teóricos acerca da Doença de Parkinson, abordando seus sintomas, tratamentos e a relevância da análise vocal no diagnóstico, além de estabelecer os métodos de aprendizado de máquina e as tecnologias empregadas. No terceiro capítulo, a metodologia empregada é detalhada, desde a análise dos dados até a elaboração do modelo e da interface Parkin95. O Capítulo 4 aborda os resultados alcançados, concentrando-se no desempenho do classificador e nas capacidades da plataforma. Finalmente, o Capítulo 5 apresenta as conclusões da pesquisa, suas restrições e propostas para futuras pesquisas, como a ampliação do sistema com novos dados e instrumentos.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Desenvolver e integrar um modelo de Regressão Logística Regularizada a uma interface gráfica interativa (Parkin95), com o intuito de apoiar neurologistas no diagnóstico da Doença de Parkinson a partir da análise de amostras de voz.

1.3.2 Objetivos específicos

- Realizar a classificação dos pacientes com base nos dados obtidos a partir das características presentes nos sinais de voz de cada paciente.
- Implementar um modelo de classificação utilizando Regressão Logística Regularizada baseado em dados vocais de pacientes com e sem Doença de Parkinson.
- Avaliar o desempenho do modelo por meio de métricas estatísticas como acurácia, precisão, sensibilidade e F1-Score.
- Desenvolver uma interface gráfica de fácil utilização para profissionais da saúde, utilizando a biblioteca Streamlit.
- Integrar o modelo preditivo à interface Parkin95, permitindo a inserção de dados e visualização dos resultados de forma intuitiva.
- Documentar o funcionamento do sistema e suas funcionalidades, visando sua aplicabilidade clínica.
- Propor melhorias futuras, como a inclusão de novos tipos de dados (exames de imagem e sinais motores) e migração para frameworks mais robustos.

2 FUNDAMENTAÇÃO TEÓRICA

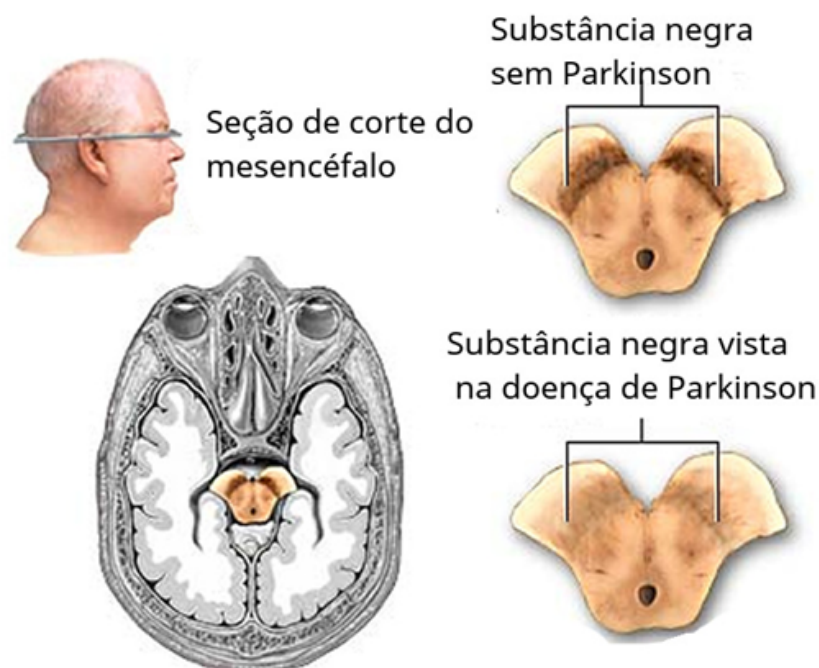
Neste tópico serão tratados os tópicos iniciais sobre doença de Parkinson (DP), aprendizado de máquina e o algoritmo de regressão logística regularizada.

2.1 A DOENÇA DE PARKINSON

A doença de Parkinson (DP) é uma condição neurológica crônica que afeta o sistema nervoso central, responsável pelo controle do movimento. Segundo [Silva e Carvalho \(2019\)](#), estima-se que a prevalência da Doença de Parkinson seja de 100 a 200 casos a cada 100 mil habitantes, afetando cerca de 10 milhões de pessoas em todo o mundo.

Seus sintomas incluem tremores involuntários, rigidez muscular, lentidão dos movimentos (bradicinesia), instabilidade postural e, em alguns casos, alterações na fala e na escrita. A DP está associada à degeneração progressiva de células nervosas na substância negra, uma região dos gânglios basais localizada no cérebro, responsável pela produção de dopamina. ([Costa; Paixão, 2018](#))

Figura 1 – Diagrama da substância negra



Fonte: [Costa e Paixão \(2018\)](#)

A dopamina é um neurotransmissor essencial para a modulação dos circuitos motores, permitindo movimentos suaves, coordenação postural e supressão de mo-

vimentos involuntários. Quando ocorre a degeneração dessas células, a produção de dopamina diminui significativamente, reduzindo a comunicação entre as células nervosas e comprometendo o controle dos movimentos musculares. Isso resulta em tremores, dificuldade de iniciar movimentos, hipocinesia (redução da amplitude dos movimentos), alterações posturais e perda de coordenação. Esses sintomas refletem a incapacidade dos gânglios basais de desempenhar suas funções regulatórias normais sobre o movimento voluntário e a postura (Sakar *et al.*, 2013).

2.1.1 Sintomas

O tremor em repouso em uma das mãos, geralmente é o primeiro sintoma. Geralmente, o tremor começa nas mãos ou nos pés, afetando um lado do corpo mais do que o outro. A mandíbula e a língua também podem ser afetadas, mas a voz normalmente não é diretamente atingida. No entanto, a fala pode ficar mais baixa e monótona, às vezes com gagueira. Com o avanço da doença, o tremor pode diminuir, enquanto a rigidez dos músculos se torna mais evidente. Em alguns casos, onde a rigidez e a lentidão dos movimentos são os principais sintomas, o tremor em repouso pode ser muito discreto ou até inexistente. Outro sintoma importante é a rigidez muscular, que pode ocorrer independentemente do tremor. Ao movimentar uma articulação afetada, o profissional de saúde pode observar reflexos irregulares, devido às variações na intensidade da rigidez (Monteiro *et al.*, 2017).

Um dos principais sinais da doença de Parkinson é a bradicinesia, caracterizada pela lentidão anormal dos movimentos, tanto voluntários quanto automáticos. A rigidez e a hipocinesia (redução da capacidade de se mover) contribuem para dores musculares e sensação de fadiga. Pacientes com a doença também podem apresentar micrografia, ou seja, escrevem com letras muito pequenas, devido à hipocinesia e ao comprometimento do controle muscular nas extremidades. A instabilidade postural tende a se desenvolver mais tardiamente. Caso esteja presente no início da doença, é necessário considerar diagnósticos alternativos (Rajput; Noyes, 2024).

A DP também pode levar a complicações respiratórias, sendo a principal causa de morte entre os pacientes. A pneumonia por aspiração é a complicação mais frequente nesse contexto (Onodera *et al.*, 2000).

Além dos sintomas motores, a doença de Parkinson (DP) pode causar sintomas não motores, como alterações de humor, distúrbios do sono, problemas de memória e dificuldades cognitivas. Também podem surgir complicações psiquiátricas, incluindo depressão e, em estágios avançados, demência (Gonçalves; Alvarez; Arruda, 2007).

A demência desenvolve-se em um terço dos pacientes diagnosticados, podendo ter a diminuição da fluência verbal. Menciona-se também transtornos do sono, a insônia e a privação do sono podem piorar a depressão e o comprometimento cognitivo e colaborar para a sonolência diurna excessiva (Monteiro *et al.*, 2017).

A DP é a segunda doença neurodegenerativa mais comum, ficando atrás apenas da doença de Alzheimer. Geralmente, manifesta-se entre os 50 e 79 anos, mas, em casos raros, pode ocorrer em crianças ou adolescentes (Rajput; Noyes, 2024).

Embora a doença de Parkinson tenha sido inicialmente vista como uma condição esporádica, evidências crescentes indicam que fatores genéticos desempenham um papel significativo. Estima-se que entre 10% e 25% dos casos possam estar relacionados a predisposições genéticas, especialmente entre pacientes diagnosticados em idades mais jovens (Singleton; Farrer; Bonifati, 2013).

2.1.2 Tratamentos e fármacos

A doença de Parkinson (DP) não tem cura, logo, os tratamentos disponíveis destinam-se para aliviar e controlar os sintomas, visando melhorar a qualidade de vida do paciente. Azevedo e Cardoso (2009) destacam que, embora a levodopa seja a principal medicação no tratamento da doença de Parkinson, outras opções terapêuticas também estão disponíveis. Estas incluem medicamentos neuroprotetores, cujo efeito na progressão da degeneração neuronal ainda não é comprovado, além de drogas que potencializam a ação da levodopa, como os agonistas dopaminérgicos e os inibidores da catecol-O-metiltransferase (COMT).

Azevedo e Cardoso (2009) explicam que os agonistas dopaminérgicos são medicamentos que estimulam diretamente os receptores dopaminérgicos, sendo desenvolvidos para superar as limitações da levodopa. Com uma meia-vida mais longa que a levodopa, esses medicamentos proporcionam uma estimulação mais prolongada e contínua dos núcleos da base, ao contrário da levodopa, que gera uma estimulação pulsátil e pode levar a complicações devido a esse padrão intermitente de estimulação.

2.1.3 Ferramentas sobre o monitoramento e diagnóstico da DP

Algumas ferramentas que já são utilizadas para ajudar a identificar mudanças nos sintomas e a monitorar a eficácia dos tratamentos (Medscape Notícias Médicas, 2024). Exemplos de Dispositivos e Tecnologias:

- Smartwatches: Diversos smartwatches, como os das marcas Samsung e Apple,

são equipados com sensores capazes de acompanhar sinais vitais, padrões de sono e níveis de atividade física, podendo também ser integrados a aplicativos desenvolvidos especialmente para auxiliar no monitoramento da Doença de Parkinson ([Lone Star Neurology, 2023](#)).

- Anéis Inteligentes: Dispositivos como o Oura Ring utilizam sensores biométricos avançados para monitorar parâmetros como frequência cardíaca, variabilidade da frequência cardíaca (HRV), padrões de sono e temperatura corporal, oferecendo métricas relevantes para o acompanhamento contínuo de pacientes com Doença de Parkinson ([Anhesini, 2024](#)).
- Cinturões com sensores: Cinturões equipados com sensores inerciais podem ser utilizados para monitorar padrões de marcha e outros movimentos corporais, proporcionando uma análise mais abrangente e precisa dos sintomas motores da Doença de Parkinson ([Moreau et al., 2023](#)).

2.2 ANÁLISE DE VOZ NA MEDICINA

O objetivo desta seção é explorar como estudos anteriores utilizaram a voz como ferramenta para o diagnóstico de doenças neurológicas, mencionando outras doenças além do Parkinson. Através da análise de características vocais, como a modulação da frequência, a articulação e o ritmo, é possível identificar padrões indicativos de distúrbios motores e cognitivos.

Além disso, serão mencionadas outras condições neurológicas em que a voz tem sido utilizada como biomarcador, como a Doença de Alzheimer e a Esclerose Lateral Amiotrófica (ELA), destacando a versatilidade e o potencial da voz na detecção precoce e no monitoramento contínuo de diversas doenças neurodegenerativas.

A análise acústica da voz tem demonstrado grande potencial como ferramenta auxiliar no diagnóstico de doenças neurodegenerativas. Alterações na frequência fundamental, intensidade, *jitter*, *shimmer* e nas características dos formantes são frequentemente observadas em pacientes com Parkinson, Alzheimer e ELA. Essas alterações podem servir como biomarcadores sensíveis e específicos, permitindo a detecção precoce e o acompanhamento da progressão dessas doenças de forma não invasiva e objetiva.

A análise de voz na medicina utiliza avançadas técnicas de processamento de sinais e aprendizado de máquina para identificar padrões associados a diferentes patologias. Por meio da captura de gravações de áudio, extraem-se características acústicas detalhadas, como coeficientes MFCC, *jitter* e *shimmer*. Esses dados são

então processados por modelos estatísticos e redes neurais profundas, permitindo a detecção de anomalias vocais e contribuindo para diagnósticos mais precisos, eficientes e personalizados. Diversos estudos demonstraram a viabilidade do uso de sinais de voz no diagnóstico de doenças neurológicas. Alguns exemplos incluem:

- Esclerose Lateral Amiotrófica (ELA)

A progressão da Esclerose Lateral Amiotrófica (ELA) impacta diretamente a musculatura envolvida na fala, resultando no desenvolvimento de disartria. Diante disso, pesquisas têm utilizado gravações de voz de pacientes para identificar alterações na taxa de fala e na qualidade vocal, que podem servir como indicadores precoces da doença. Essas análises permitem detectar mudanças sutis, como variações na frequência, amplitude, *jitter* e *shimmer*, que estão associadas ao comprometimento neuromuscular característico da ELA. Um estudo relevante nesse contexto é o artigo “*Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis*” (Stegmann *et al.*, 2020). Publicado na revista *npj Digital Medicine*, o trabalho propõe um método inovador baseado na análise de fala para a detecção precoce e o monitoramento da progressão da ELA, com foco nas alterações bulbares — que afetam funções como fala e deglutição. A abordagem utiliza uma plataforma digital e remota, permitindo a coleta e análise frequente de dados de voz, o que facilita o acompanhamento contínuo dos pacientes. Esse método demonstra potencial para complementar os diagnósticos clínicos tradicionais, oferecendo uma ferramenta não invasiva e acessível para o manejo da doença.

- Doença de Alzheimer

A Doença de Alzheimer (DA) é um transtorno neurodegenerativo progressivo e irreversível que compromete principalmente a memória e diversas funções cognitivas. À medida que a doença evolui, indivíduos acometidos pelo Alzheimer enfrentam dificuldades para realizar tarefas cotidianas, apresentam alterações comportamentais e, com o tempo, podem perder a capacidade de comunicação, afetando significativamente sua qualidade de vida. A análise computacional da linguagem tem se mostrado uma ferramenta promissora na área da saúde, especialmente no contexto de doenças neurodegenerativas. Fraser, Meltzer e Rudzicz (2015) exploraram a aplicação dessa abordagem no estudo da Doença de Alzheimer (DA), demonstrando que características linguísticas presentes na fala narrativa podem servir como biomarcadores para a doença. Os resultados obtidos pelos autores sugerem que a análise computacional da linguagem pode complementar os métodos tradicionais de diagnóstico, oferecendo

informações valiosas sobre a progressão da doença e o impacto da DA nas funções cognitivas.

A voz é considerada um biomarcador relevante para o diagnóstico e monitoramento de doenças neurológicas, como a Doença de Parkinson (DP), por ser um indicador não invasivo e de baixo custo. Alterações na produção da fala, como variações no ritmo, fluência, intensidade e clareza vocal, podem refletir mudanças neurológicas e motoras associadas à progressão de distúrbios cognitivos e motores. Além disso, a voz é facilmente acessível e pode ser analisada com o auxílio de tecnologias de processamento de sinais, tornando-se uma ferramenta valiosa para a detecção precoce e acompanhamento da evolução de doenças neurodegenerativas, sem a necessidade de procedimentos invasivos ou caros.

2.3 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (AM), ou *Machine Learning* em inglês, é uma área da inteligência artificial (IA) que permite aos sistemas adquirir conhecimentos e melhorar seu desempenho de forma autônoma, sem intervenção direta no código por parte de humanos. Os modelos gerados através de AM conseguem reconhecer padrões nos dados aos quais são expostos e, a partir disso, são capazes de melhorar o seu próprio desempenho. Surgiu na década de 1960 como um subcampo da IA, com o propósito de identificar padrões a partir de dados. No início, seu foco estava voltado exclusivamente para aplicações computacionais, mas, a partir da década de 1990, suas possibilidades se ampliaram significativamente, alcançando diversas áreas (Izbicki; Santos, 2020).

Fagundes *et al.* (2024) afirmam que modelos de aprendizado de máquina (AM) são usados em uma ampla gama de áreas do conhecimento humano e têm impactado fortemente as ciências da saúde. O objetivo geral dos modelos de AM é fazer projeções sobre eventos futuros com base em dados históricos, descobrir padrões, relações e percepções ocultos nos dados e fornecer informações e recomendações para que humanos possam tomar decisões mais informadas (Howard, 2019). O AM faz uso de algoritmos complexos para criar modelos preditivos. Esses algoritmos podem ser categorizados em diferentes tipos, como aprendizado supervisionado, não supervisionado, semi-supervisionado e aprendizado por reforço, cada um com suas particularidades e aplicações específicas.

No livro, escrito por Géron (2019), intitulado "*Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems*", o *Machine Learning* é ótimo para:

- Problemas cujas soluções existentes exigem muitos ajustes ou longas listas de regras: um único algoritmo de Machine Learning pode, muitas vezes, simplificar o código e ter um desempenho melhor do que a abordagem tradicional.
- Problemas complexos para os quais uma abordagem tradicional não oferece uma boa solução: as melhores técnicas de Machine Learning podem, talvez, encontrar uma solução.
- Ambientes dinâmicos: um sistema de Machine Learning pode se adaptar a novos dados.
- Obter insights sobre problemas complexos e grandes volumes de dados.

2.3.1 Métodos de aprendizagem de máquina

- Aprendizado Supervisionado

Segundo [Delahoz e Labrador \(2014\)](#), algoritmos supervisionados utilizam dados previamente rotulados e validados para treinar um modelo. Após o treinamento, esse modelo é capaz de identificar padrões e fazer previsões com base em um conjunto de testes sem rótulos. O aprendizado supervisionado pode ser dividido em subseções de regressão e classificação.

Os objetivos do aprendizado supervisionado são compreender o relacionamento entre os atributos e a classe, identificando as características dos dados mais relevantes para prever a classe ou o valor de uma variável. Além disso, busca-se prever a classe de novos exemplos, utilizando o modelo treinado para classificar novos dados, ou seja, atribuindo a eles uma classe ou valor específico.

- Regressão: A regressão é uma tarefa de aprendizado supervisionado cujo objetivo é prever um valor contínuo ou numérico com base nas variáveis de entrada, gerando uma saída representada por uma variável contínua.
- Classificação: A classificação é uma tarefa de aprendizado supervisionado cujo objetivo é prever a classe ou o rótulo de uma entrada, resultando em uma saída representada por uma variável discreta.
- Aprendizado Não-supervisionado: O aprendizado não supervisionado, conforme descrito por [Hinton, Sejnowski e Poggio \(1999\)](#), não requer dados rotulados para treinamento. O algoritmo recebe dados não rotulados e, em seguida, o modelo

identifica a estrutura oculta dos dados sem intervenção humana, organizando-os em *clusters* ou classes.

Os objetivos do aprendizado não supervisionado são identificar representações úteis dos exemplos, como entender a organização dos dados e determinar as características mais relevantes. Além disso, busca-se encontrar *clusters*, ou seja, agrupar os dados em categorias de objetos semelhantes. O aprendizado não supervisionado também visa reduzir a dimensionalidade dos dados, descobrir as causas subjacentes aos dados, identificar as variáveis responsáveis pelos dados observados e estimar a distribuição de probabilidade dos mesmos.

- Aprendizado de máquina por reforço

O aprendizado por reforço (*Reinforcement Learning - RL*) é um paradigma de aprendizado voltado para o aprendizado de controle de um sistema com o objetivo de maximizar uma medida de desempenho numérica que expressa um objetivo de longo prazo. O que distingue o aprendizado por reforço do aprendizado supervisionado é que o aprendiz recebe apenas um *feedback* parcial sobre as previsões realizadas (Szepesvári, 2010). Os modelos de reforço aprendem os resultados corretos por meio de recompensas e penalidades, aplicando métodos de tentativa e erro semelhantes aos usados pelos seres humanos. O objetivo central do aprendizado por reforço (RL) é fazer com que o agente aprenda a tomar uma sequência de ações que maximize a soma total das recompensas ao longo do tempo, que pode variar de ganhar um jogo a realizar uma tarefa específica. Ao contrário de outras abordagens de aprendizado de máquina, que utilizam dados rotulados, o agente no RL aprende por meio da interação direta com o ambiente. Ele explora diferentes ações, observa as consequências e ajusta sua estratégia para obter melhores resultados. O RL é especialmente eficaz em problemas que envolvem decisões sequenciais, nas quais a escolha de uma ação em um determinado estado impacta os estados futuros e as recompensas subsequentes. Além disso, os agentes de RL têm a capacidade de se adaptar a ambientes dinâmicos, tornando-os resilientes a mudanças, incertezas e imprevistos ao longo do tempo (Geeks for Geeks, 2024).

2.3.2 Algoritmos comuns de classificação e de regressão

Os algoritmos de *Machine Learning* são essenciais para permitir que sistemas aprendam a partir de dados e tomem decisões sem serem explicitamente programados. Cada tipo de problema, seja ele de classificação, regressão ou agrupamento, exige a

aplicação de algoritmos específicos que melhor se adequam à tarefa e ao conjunto de dados disponível. Alguns algoritmos de classificação:

2.3.2.1 Árvores de decisão

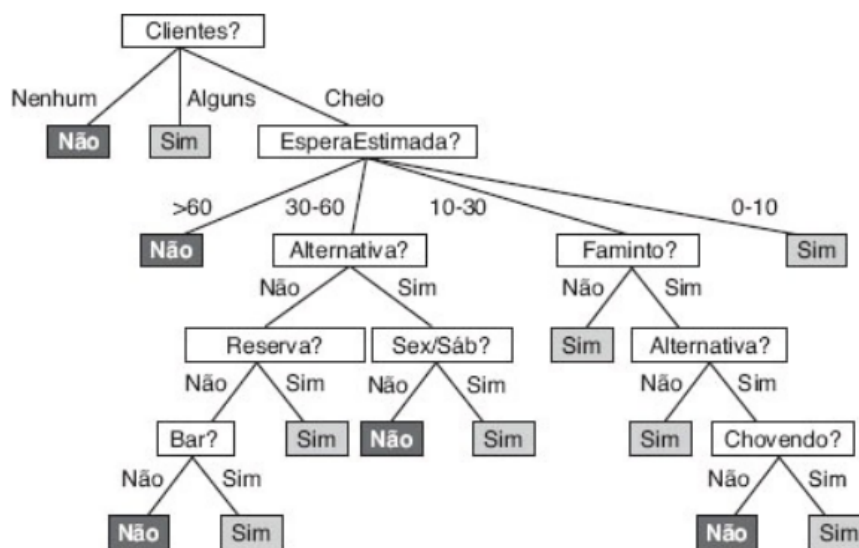
Uma árvore de decisão toma suas decisões por meio de uma sequência de testes estruturados. Cada nó interno da árvore corresponde a um teste sobre o valor de um dos atributos de entrada, A_i , enquanto as ramificações desse nó representam os possíveis valores que o atributo pode assumir, $A_i = v_{ik}$. Por sua vez, cada nó folha especifica o valor final a ser retornado pela função, ou seja, a decisão ou classificação resultante. A representação em forma de árvore de decisão é intuitiva e de fácil compreensão para os seres humanos, tanto que muitos manuais práticos, como guias de reparo automotivo, são organizados como extensas árvores de decisão, abrangendo centenas de páginas (Norvig; Russell, 2013).

A profundidade de uma árvore de decisão está diretamente relacionada à sua complexidade, que é controlada pelos nós de decisão. À medida que a árvore se torna mais complexa, o modelo tende a apresentar maior precisão no conjunto de dados de treinamento. No entanto, árvores com complexidade moderada (influenciada pelo tamanho do conjunto de dados) são geralmente preferidas, pois oferecem um equilíbrio entre ajuste (*fitting*) e generalização, priorizando um melhor desempenho preditivo em dados desconhecidos. Para alcançar esse equilíbrio, técnicas como pré-processamento ou poda (*pruning*) são frequentemente aplicadas. O pré-processamento é utilizado durante a construção da árvore para evitar que ela cresça excessivamente, enquanto a poda é uma técnica pós-processamento que remove nós e ramos desnecessários após a criação da árvore, simplificando sua estrutura e melhorando sua capacidade de generalização (Chen, 2024).

- Máquinas de Vetores de Suporte (SVM)

A Máquina de Vetores de Suporte (*Support Vector Machine - SVM*) é um algoritmo de aprendizado de máquina supervisionado amplamente reconhecido por sua eficácia em tarefas de classificação e regressão. Fundamentado na teoria de otimização convexa, o SVM busca identificar um hiperplano ótimo que maximize a margem de separação entre classes em um espaço de características, mesmo em cenários de alta dimensionalidade. Essa abordagem permite ao algoritmo generalizar bem para dados não vistos, minimizando o risco de sobreajuste (*overfitting*). Além disso, por meio do uso de funções de kernel, o SVM é capaz de mapear dados linearmente

Figura 2 – Classificação em árvore



Fonte: [Norvig e Russell \(2013\)](#)

inseparáveis para espaços de maior dimensionalidade, onde a separação linear se torna viável, ampliando sua aplicabilidade a problemas complexos e não lineares.

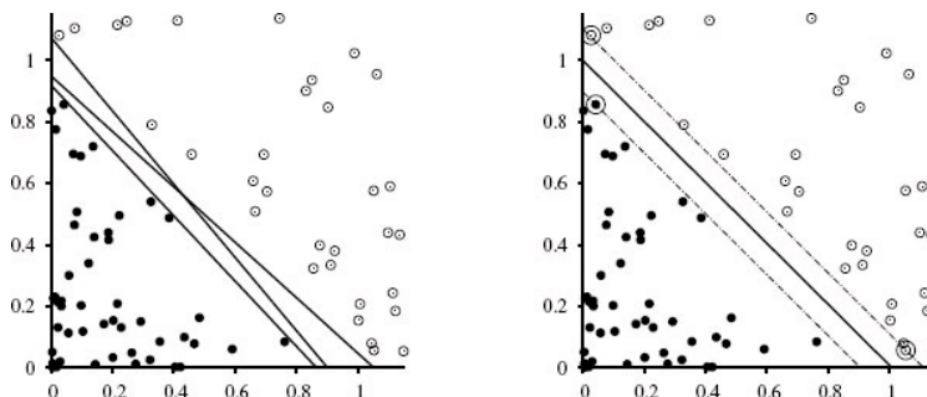
De acordo com [Norvig e Russell \(2013\)](#) os SVMs utilizam o conceito de kernel para mapear os dados para um espaço de maior dimensão, permitindo que sejam separáveis linearmente mesmo quando não são linearmente separáveis no espaço original. Esse truque de kernel possibilita uma separação eficiente de dados em problemas complexos.

Os SVMs são classificados como métodos não paramétricos, pois armazenam exemplos de treinamento, mas, na prática, mantêm apenas um subconjunto deles, conhecidos como vetores de suporte, que são críticos para a construção do modelo. Essa característica proporciona aos SVMs a flexibilidade de representar funções complexas, ao mesmo tempo que são robustos ao sobreajuste. A proposta dos SVMs é que determinados exemplos têm maior relevância do que outros, e focar neles pode resultar em uma melhor capacidade de generalização do modelo.

- K-Vizinhos Mais Próximos(K-NN)

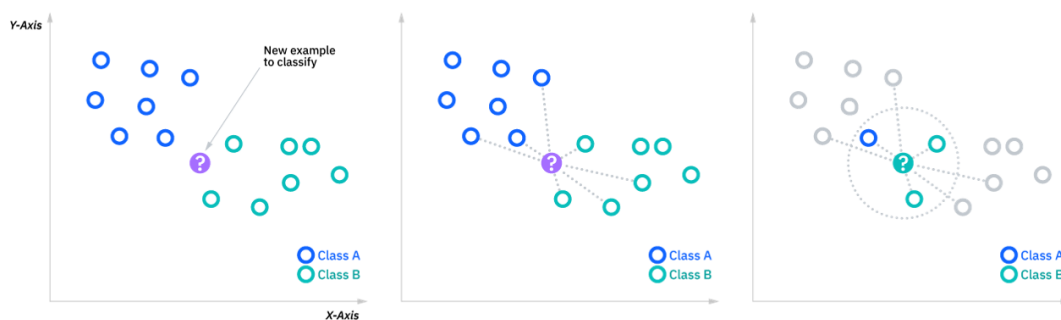
O método de *k-Nearest Neighbours(kNN)* é uma técnica de classificação não paramétrica conhecida por sua simplicidade e eficácia em vários cenários ([Guo et al., 2003](#)).

Segundo [Wang \(2002\)](#) para classificar um registro de dados e t , o algoritmo

Figura 3 – Classificação com Máquinas de Vetores de Suporte (SVM)

Fonte: [Norvig e Russell \(2013\)](#)

identifica os k vizinhos mais próximos a ele, formando uma “vizinhança”. A classe de t é definida por uma “votação” determinada pelos vizinhos, os mais próximos em distância atribuem um peso à “votação” maior que os vizinhos mais distantes. O valor de k é importantíssimo para o desempenho que o método irá ter. Para definir o k ideal, uma abordagem comum é testar o algoritmo com diferentes valores de k e selecionar aquele que oferece o melhor desempenho, como maior precisão ou menor taxa de erro. Essa estratégia ajuda a garantir que o modelo seja ajustado de forma adequada aos dados.

Figura 4 – Diagrama do K-NN

Fonte: [IBM \(2024\)](#)

Em problemas de regressão, o método kNN funciona de forma semelhante à classificação, mas em vez de prever classes, ele estima valores contínuos. Para isso, ele calcula a média dos valores dos k vizinhos mais próximos para um determinado ponto. A distância euclidiana é a métrica mais comum utilizada para medir a proximidade entre os dados. Ela calcula a distância "em linha reta" entre dois pontos ([IBM, 2024](#)).

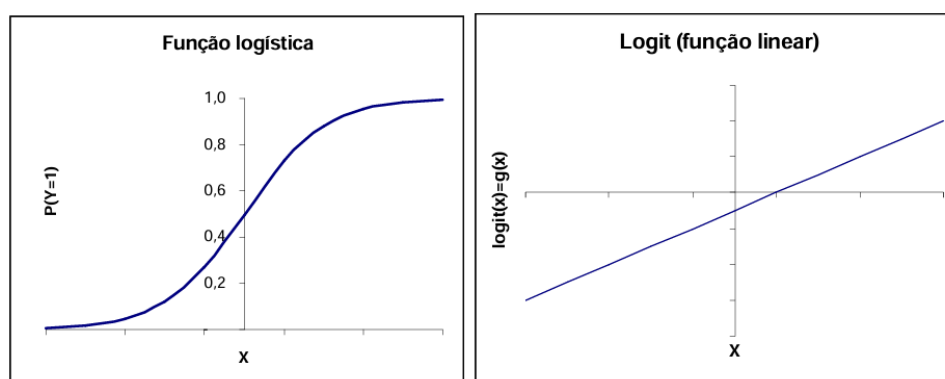
2.4 REGRESSÃO LOGÍSTICA REGULARIZADA (RL-REG)

[Hosmer e Lemeshow \(1989\)](#) destacam que, ao iniciar uma análise utilizando a regressão logística, é essencial ter em mente que o propósito desse método é

semelhante ao de outras abordagens estatísticas de modelagem. O principal objetivo é construir um modelo que represente de forma precisa e eficiente a relação entre uma variável dependente, também chamada de resposta, e um conjunto de variáveis independentes, frequentemente denominadas covariáveis. O modelo deve ser ao mesmo tempo simples e adequado à realidade do problema estudado, garantindo uma interpretação coerente dos resultados.

Conforme [Hosmer e Lemeshow \(1989\)](#), o modelo logístico é escolhido para analisar variáveis com resposta binária por dois motivos principais: sua flexibilidade e simplicidade matemática, além da clareza e riqueza na interpretação dos resultados que ele oferece. Segundo [Bittencourt \(2003\)](#), a Figura 5 representa a função logística, com seu formato característico em "S", demonstrando a relação linear entre a variável x e o logit $g(x)$.

Figura 5 – Representação da Função Logística e sua Transformação Logit



Fonte: [Bittencourt \(2003\)](#)

A função logística é definida como:

$$P(Y = 1 | X) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

2.4.1 Detalhamento da transformação logística

- Entrada: O Logito

A função logística recebe como entrada um valor numérico, denominado logito (z), que é calculado a partir de uma combinação linear das variáveis preditoras. Esse valor pode variar de $-\infty$ a $+\infty$ e é expresso pela seguinte equação:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.2)$$

A função logística aplica uma transformação não linear ao logito (z) por meio da seguinte fórmula:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2.3)$$

Essa equação é conhecida como **função sigmoide** devido ao seu formato em "S". A função possui as seguintes propriedades:

- Para $z \rightarrow +\infty$, $f(z) \rightarrow 1$;
- Para $z \rightarrow -\infty$, $f(z) \rightarrow 0$.

Essa transformação "comprime" o valor de z no intervalo $[0, 1]$, garantindo que a saída seja interpretável como uma probabilidade.

- Saída: Probabilidade estimada

O resultado da função logística é um valor no intervalo $[0, 1]$, que representa a probabilidade $P(y = 1 | x)$ de um evento ocorrer, dadas as características de entrada x . Por exemplo:

- Se $f(z) = 0.8$, há 80% de chance de o evento ocorrer;
- Se $f(z) = 0.2$, há 20% de chance de o evento ocorrer.

Essa saída é particularmente útil em problemas de classificação binária, onde o objetivo é prever a probabilidade de uma observação pertencer a uma das duas classes (por exemplo, comprar ou não comprar um produto).

- Interpretação e Aplicação

A função logística é amplamente utilizada em modelos de regressão logística, uma técnica estatística para problemas de classificação binária. A partir das probabilidades estimadas, é possível definir um limiar de decisão (geralmente 0,5) para classificar as observações:

- Se $f(z) \geq 0.5$, a observação é classificada na **classe positiva** (evento ocorre);
- Se $f(z) < 0.5$, a observação é classificada na **classe negativa** (evento não ocorre).

2.4.2 Métricas de desempenho para classificação

Para avaliar o desempenho do modelo de classificação desenvolvido neste trabalho, foram utilizadas as seguintes métricas: acurácia, matriz de confusão, *Recall* e *F-score*. Essas métricas permitem analisar não apenas o desempenho global do modelo, mas também sua capacidade de identificar corretamente os casos positivos e negativos, o que é essencial no contexto de diagnóstico médico.

- Matriz confusão

É uma tabela que resume os resultados das previsões do modelo, comparando os valores reais com os valores preditos. A matriz de confusão é estruturada com base em quatro componentes principais:

Tabela 1 – Exemplo genérico de uma matriz confusão.

	Classe Preditiva: Positiva	Classe Preditiva: Negativa
Classe Real: Positiva	VP	FN
Classe Real: Negativa	FP	VN

Fonte: Elaboração própria

- VP (Verdadeiros Positivos): Casos positivos corretamente classificados.
- VN (Verdadeiros Negativos): Casos negativos corretamente classificados.
- FP (Falsos Positivos): Casos negativos classificados incorretamente como positivos.
- FN (Falsos Negativos): Casos positivos classificados incorretamente como negativos.

Observa-se, na matriz de confusão, a presença de uma diagonal principal, composta pelos valores corretamente classificados, representados pelos Verdadeiros Positivos (VP) e Verdadeiros Negativos (VN). Além disso, é importante destacar que, em contextos médicos, os Falsos Negativos (FN) — situações em que o modelo não identifica corretamente um paciente que realmente possui a doença — representam um risco mais crítico. Por esse motivo, muitas vezes é desejável que o algoritmo atribua maior peso a esses erros, buscando minimizar o número de falsos negativos, uma vez que a não identificação da doença pode atrasar o início do tratamento adequado.

- Acurácia

A acurácia é uma métrica que mede a proporção de previsões corretas, tanto positivas quanto negativas, em relação ao total de amostras avaliadas. Em outras palavras, ela indica a taxa de acerto geral do modelo, calculada pela divisão do número de previsões corretas pelo número total de previsões realizadas.

A fórmula da acurácia é dada por:

$$\text{Acurácia} = \frac{\text{Previsões Corretas}}{\text{Total de Previsões}} \quad (2.4)$$

Em termos de matriz de confusão:

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} \quad (2.5)$$

- *Recall*

O *Recall* é uma métrica que mede a frequência com que um modelo de aprendizado de máquina identifica corretamente as instâncias positivas (verdadeiros positivos)

entre todas as amostras positivas reais no conjunto de dados. O *Recall* pode ser calculado dividindo o número de verdadeiros positivos pelo número total de instâncias positivas. É uma métrica cuja escala varia de 0 a 1, podendo ser expressa em forma percentual. Valores mais altos indicam melhor desempenho do modelo na identificação correta das instâncias positivas (AI, 2023).

Segundo (AI, 2023), o *Recall* responde à pergunta: um modelo de *machine learning* consegue encontrar todas as instâncias da classe positiva?

$$\text{Recall} = \frac{VP}{VP + FN} \quad (2.6)$$

- Precisão

A precisão é uma métrica utilizada para verificar a exatidão das previsões positivas feitas por um modelo. Ela indica a fração dos casos que foram corretamente classificados como positivos entre todos aqueles que o modelo identificou como positivos, incluindo os erros. Essa medida é particularmente relevante em cenários em que falsos positivos podem gerar consequências significativas (Scikit-learn, 2025).

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.7)$$

- *F-Score*

O *F-score*, também chamado de *F1-score*, é uma medida da precisão de um modelo em um conjunto de dados. Ele é utilizado para avaliar sistemas de classificação binária, que categorizam exemplos como "positivo" ou "negativo". É uma maneira de combinar a precisão e o *Recall* do modelo, sendo definido como a média harmônica entre eles (DeepAI, 2023).

$$\text{F-Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (2.8)$$

A seleção da métrica mais adequada depende diretamente do contexto e dos objetivos específicos do problema que está sendo analisado. Em cenários nos quais os falsos negativos geram impactos significativos — como no caso do diagnóstico de doenças — o *Recall* tende a ser priorizado, uma vez que é fundamental identificar

corretamente todas as instâncias positivas, mesmo que isso ocasione um aumento no número de falsos positivos.

2.5 TECNOLOGIAS PARA ANÁLISE DE DADOS E DESENVOLVIMENTO DE APLICAÇÕES

Diversas ferramentas auxiliam no desenvolvimento de projetos de aprendizado de máquina, sendo Python uma das mais populares nessa área (Silva, 2024). A linguagem Python é amplamente utilizada devido à sua capacidade de criar scripts e programas funcionais, além de oferecer diversas bibliotecas para resolver problemas matemáticos e computacionais (Guedes, 2022).

A evolução da linguagem Python foi impulsionada pelo desenvolvimento de novas bibliotecas de suporte, permitindo a análise e manipulação eficiente de diversos tipos de dados. Dentre as mais utilizadas pela comunidade científica, destacam-se (McKinney, 2012):

- NumPy: possibilita a criação, manipulação e visualização de arrays, além da execução de operações matemáticas avançadas entre eles.
- Pandas: aproveita a alta performance do NumPy e oferece funcionalidades para trabalhar com dados estruturados de forma eficiente.
- SciPy: reúne pacotes voltados para a computação científica, incluindo ferramentas para cálculos de integrais, derivadas e processamento de sinais.

Além das bibliotecas mencionadas, também vale destacar o Streamlit, uma ferramenta para a criação de aplicativos interativos em Python. Segundo (Awari, 2023), o Streamlit permite a construção de aplicações web baseadas em Python, onde cada seção do aplicativo é definida por uma função Python. Essas funções são executadas dinamicamente conforme a interação do usuário, possibilitando atualizações rápidas e em tempo real na interface da aplicação.

Outra ferramenta utilizada para auxiliar no armazenamento de dados é o *SQLite* que é sistema de gerenciamento de banco de dados relacional (SGBDR). Em comparação com SGBDRs tradicionais, como *MySQL* e *PostgreSQL*, o *SQLite* apresenta uma arquitetura simplificada e um menor *overhead*. Sua ausência de um servidor dedicado e a capacidade de armazenar o banco de dados em um único arquivo facilitam a implantação e o gerenciamento. O *SQLite* é ideal para aplicações que não exigem alta concorrência ou recursos avançados de gerenciamento de usuários, como aplicações embarcadas, dispositivos móveis e aplicações *desktop* de uso individual. Sua

leveza, portabilidade e facilidade de uso o tornam uma boa alternativa para projetos que buscam simplicidade e eficiência.

2.6 PADRÃO DE ARQUITETURA: MVC

O modelo MVC (*Model-View-Controller*), é um padrão de arquitetura de software que surgiu no começo da década de 80, mas se tornou muito popular na criação de aplicações WEB (Luciano; Alves, 2017).

A arquitetura Modelo-Visão-Controlador (MVC) apresenta uma dinâmica estruturada, na qual todas as requisições realizadas pela aplicação são inicialmente encaminhadas à camada Controller. Esta, por sua vez, é responsável por acionar a camada Model, encarregada do processamento da lógica da aplicação. Após o processamento, os dados são repassados à camada View, que tem a função de exibir o resultado ao usuário. O padrão MVC promove a separação entre as camadas de apresentação, lógica de negócio e controle do fluxo da aplicação, o que contribui significativamente para a modularidade, reutilização de componentes e maior facilidade na manutenção do sistema (Luciano; Alves, 2017).

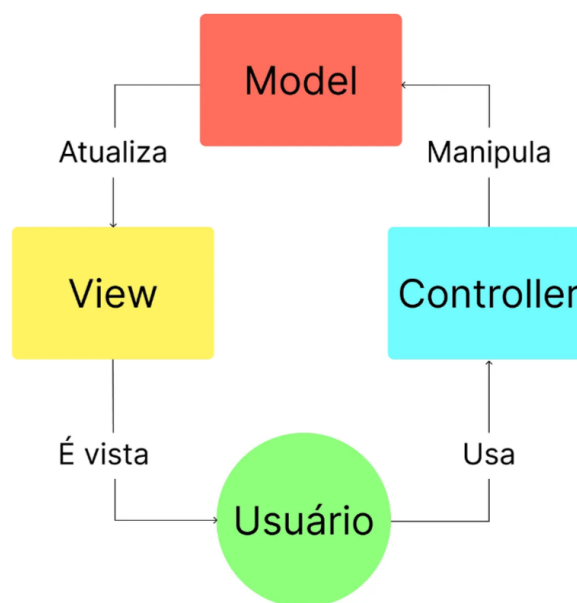
A camada *Model* é responsável por representar e gerenciar os dados e as regras de negócio da aplicação. Ela define as estruturas de dados, bem como os métodos e operações associadas ao seu tratamento, como criação, leitura, atualização e exclusão de informações. Além disso, é a partir das ações do usuário, interpretadas pela camada *Controller*, que o *Model* é atualizado, garantindo a consistência e integridade dos dados. Essa separação permite uma maior modularidade e facilita a manutenção e a escalabilidade do sistema.

A camada *View* é responsável pela apresentação visual das informações ao usuário, atuando como a interface gráfica da aplicação. Sua função principal consiste em exibir os dados processados pela camada *Model*, organizando-os de forma compreensível e esteticamente adequada, conforme os modelos previamente definidos. Além disso, é na *View* que ocorre a interação direta com o usuário, possibilitando a realização de ações por meio de elementos visuais como botões, formulários, menus e outros componentes gráficos. Essa camada não possui responsabilidade sobre a lógica de negócio, limitando-se à representação e captura de eventos da interface, que são posteriormente tratados pela camada *Controller*.

A camada *Controller* é responsável por intermediar a comunicação entre a camada *View* e a camada *Model*, sendo encarregada da implementação da lógica de controle da aplicação. Sua principal função consiste em interpretar as ações rea-

lizadas pelo usuário na interface (*View*) e, a partir disso, executar os procedimentos necessários para processar essas solicitações, acionando os métodos apropriados da camada *Model*. Dessa forma, o *Controller* atua como um mecanismo de coordenação, assegurando que as alterações no estado da aplicação ocorram de maneira estruturada, coerente e conforme as regras de negócio estabelecidas. Essa separação contribui para a modularidade do sistema, favorecendo a manutenção, a escalabilidade e a reutilização de componentes (Papst, 2021).

Figura 6 – Diagrama do padrão de arquitetura MVC (Model-View-Controller)



Fonte: Papst (2021)

3 METODOLOGIA

Nesta seção, será apresentada a metodologia adotada para o desenvolvimento deste trabalho, estruturada em duas etapas: a construção do classificador preditivo e a arquitetura e implementação da interface Parkin95. A primeira parte aborda o processo de desenvolvimento do modelo de classificação utilizado para auxiliar no diagnóstico da Doença de Parkinson. Em seguida, são descritas as ferramentas e etapas envolvidas na criação da interface gráfica.

3.1 PROCESSO DE CONSTRUÇÃO DE UMA APLICAÇÃO DE APRENDIZADO DE MÁQUINA

O processo de construção de uma aplicação de *Machine Learning* envolve várias etapas. Géron (2019) apresenta uma abordagem prática que combina conceitos teóricos com aplicações reais. Seguindo essa metodologia, o primeiro passo é definir claramente o problema e entender os dados disponíveis. Para tal, é necessário identificar se o problema envolve classificação, regressão ou clusterização, além de analisar a estrutura dos dados, verificando a presença de valores faltantes, distribuição das variáveis e possíveis anomalias. A visualização dos dados também é essencial para identificar padrões que possam influenciar o sistema.

Após essa análise inicial, os dados precisam ser preparados para a aplicação no algoritmo. Esse processo inclui limpeza dos dados, tratando valores ausentes e removendo *outliers*, além da transformação dos dados, como normalização e codificação de variáveis categóricas. Também é fundamental dividir os dados em conjuntos de treinamento, validação e teste, garantindo que o modelo seja avaliado corretamente.

Com os dados devidamente ajustados, inicia-se a etapa de seleção e treinamento do algoritmo. A escolha da técnica mais adequada é um passo crucial e, uma vez definida, o treinamento é realizado utilizando o conjunto de dados apropriado. Para aprimorar os resultados, são feitos

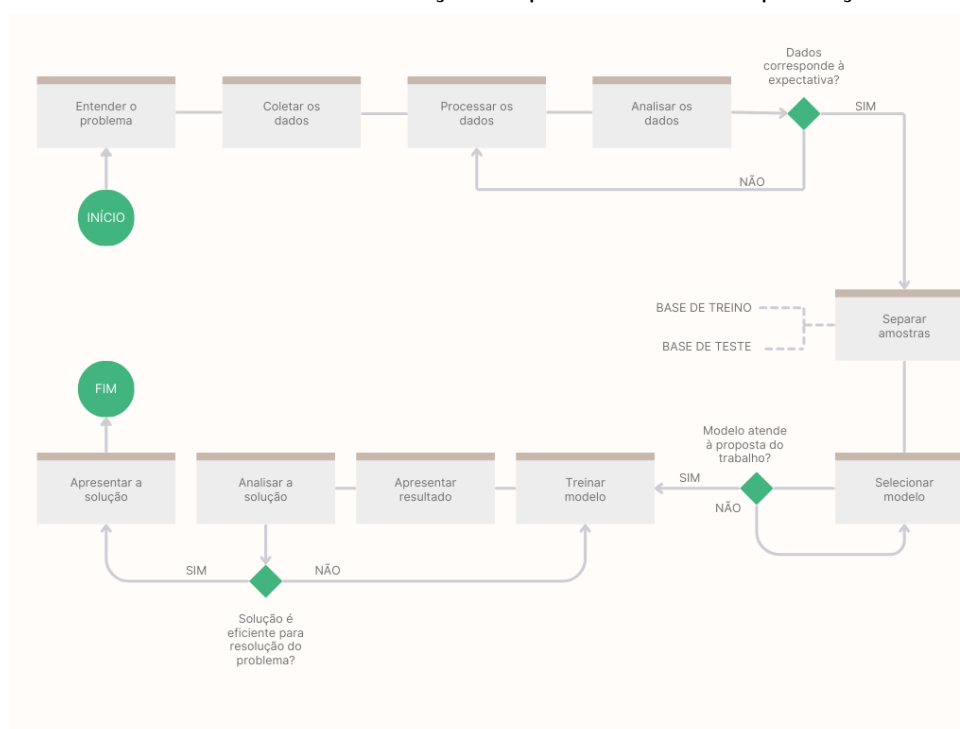
A avaliação do desempenho é feita utilizando métricas apropriadas para cada tipo de problema. No caso de problemas de classificação, são utilizadas métricas como acurácia, precisão, *recall* e *F1-score*. Além disso, a validação cruzada é uma técnica importante para garantir que o modelo consiga generalizar bem para novos dados.

Após percorridas essas etapas, a abordagem está pronta para ser implantado em produção, um processo conhecido como *deploy*. Essa fase garante que o sistema preditivo possa ser utilizado no mundo real, respondendo a novas entradas de dados e

gerando previsões. Por fim, para assegurar a eficácia contínua do algoritmo, torna-se fundamental realizar o monitoramento e a manutenção de forma sistemática. Essas atividades são essenciais para garantir que o sistema permaneça funcional e adaptável, ajustando-se adequadamente às possíveis variações nos dados ao longo do tempo.

A Figura 7 resume de maneira visual e organizada todo o processo discutido anteriormente, destacando os principais passos envolvidos no desenvolvimento de um modelo de *Machine Learning*. Ela serve como um guia prático para compreender a sequência lógica das etapas, desde a definição do problema até a implantação e monitoramento contínuo.

Figura 7 – Fluxograma do processo de desenvolvimento de um modelo de Machine Learning, desde a definição do problema até a implantação



Fonte: Elaboração própria.

3.1.1 Tratamento dos dados

Para o desenvolvimento de um sistema auxiliar na detecção de distúrbios do Parkinson por meio da análise vocal dos pacientes, a metodologia adotada envolveu a seleção e o processamento das características vocais mais relevantes, seguidos pela fase de treinamento e avaliação do desempenho do modelo gerado.

De acordo com [AMAZON \(2018\)](#), durante a análise dos dados é fundamental refletir sobre algumas questões importantes:

- Os dados estão de acordo com as expectativas?
- Houve algum problema no processo de coleta dos dados?
- Quais classes apresentam maior frequência no conjunto de dados?
- A quantidade de valores ausentes ou inválidos está acima do esperado?

Nesse contexto, a partir das diretrizes mencionadas, foram analisados os dados públicos disponibilizados no *UCI Machine Learning Repository: Parkinson Speech Dataset with Multiple Types of Sound Recordings Data Set*, os quais foram devidamente pré-processados para garantir a qualidade da análise. Este *dataset* contém gravações de voz de 20 pacientes diagnosticados com Parkinson (6 mulheres e 14 homens) e 20 pessoas saudáveis (10 homens e 10 mulheres), que recorreram ao Departamento de Neurologia da Faculdade de Medicina de Cerrahpasa, na Universidade de Istambul ([Kursun Olcay; Gurgen, 2013](#)).

Após a coleta do conjunto de dados de treinamento, que consiste em múltiplos tipos de gravações de som, e a realização dos experimentos, em linha com os achados obtidos, foi coletado um conjunto de teste independente de pacientes com Parkinson, sob o mesmo processo de exame médico, nas mesmas condições.

Durante a investigação dos sinais de voz, a equipe médica, todos os pacientes, com Parkinson ou não, foram submetidos a testes fonoaudiólogos, sendo necessário pronunciar alguns vocábulos e frases curtas, dos quais, foram extraídas 26 características das amostras de voz deste novo conjunto de dados.

As amostras de voz no arquivo de dados de treinamento estão organizadas da seguinte maneira:

- 1: vogal sustentada (aaa...)
- 2: vogal sustentada (ooo...)
- 3: vogal sustentada (uuu...)
- 4-13: números de 1 a 10
- 14-17: frases curtas
- 18-26: palavras

De acordo com [AMAZON \(2018\)](#), também destaca que é uma boa prática analisar a correlação entre cada variável e a variável alvo. De modo geral, recomenda-se a inclusão das variáveis com alta correlação, uma vez que tendem a apresentar maior poder preditivo, enquanto variáveis com baixa correlação costumam ser descartadas por, provavelmente, serem irrelevantes.

Com o intuito de seguir essa boa prática e orientar a identificação de distúrbios relacionados ao Parkinson, o processo inicial foi cuidadosamente estruturado com a seleção das características vocais mais relevantes. A base de dados contempla 26 características de voz. No entanto, para efeitos de alimentar o algoritmo proposto nesse trabalho tomou-se as 10 características mais relevantes em termos de correlação ao diagnóstico de Parkinson.

As variáveis mais correlacionadas com o diagnóstico foram organizadas em ordem decrescente de correlação, da mais relevante para a menos relevante. São elas, nesta ordem: *Jitter (local, absolute)*, *Shimmer (apq11)*, *Maximum pitch*, *Degree of voice breaks*, *Fraction of locally unvoiced frames*, *Standard deviation*, *Jitter (rap)*, *Jitter (ddp)*, *Jitter (ppq5)* e *Mean pitch*.

Segundo [Little et al. \(2009\)](#), *jitter*: refere-se à variação na frequência fundamental (F0) da voz, ou seja, pequenas oscilações no tom de um ciclo vocal para o outro. Já o *shimmer*: refere-se à variação na amplitude da voz, dada as pequenas mudanças na intensidade do som de um ciclo vocal para o outro. Ambos os parâmetros são indicativos de instabilidade na produção vocal, uma característica frequentemente observada em indivíduos com DP, decorrente da rigidez muscular e das alterações no controle motor fino da laringe.

Além disso, métricas como o *Maximum pitch* (frequência máxima atingida), *Standard deviation* (desvio padrão da frequência fundamental) e *Degree of voice breaks* (grau de quebras vocais) reforçam a presença de dificuldades no controle da prosódia e da estabilidade vocal, aspectos frequentemente comprometidos em estágios iniciais da Doença de Parkinson. Tais alterações refletem oscilações involuntárias na emissão vocal, muitas vezes associadas à rigidez muscular e à bradicinesia laríngea. Outrossim, a variável *Fraction of locally unvoiced frames* — que representa a proporção de segmentos da fala sem emissão sonora — revela pausas anormais e interrupções na fonação, aspectos que podem indicar falhas na coordenação neuromuscular do trato vocal. Esses padrões são amplamente reconhecidos como biomarcadores acústicos sensíveis para a detecção precoce da doença ([Tsanas et al., 2010](#); [Little et al., 2009](#)).

3.1.2 Fase de treinamento/teste

Na fase de treinamento, o algoritmo de RL-Reg foi capacitado com um conjunto de dados de 40 pacientes, sendo eles, 80% para treinamento e 20% para teste, a divisão dos dados em conjuntos de treinamento e teste *foi realizada de forma aleatória por meio de um algoritmo em Python*. O conjunto de treinamento incluiu 16 pacientes saudáveis e 16 pacientes com Doença de Parkinson (DP), enquanto o conjunto de teste consistiu em 4 pacientes saudáveis e 4 pacientes com DP, considerando a análise fonética dos vocábulos pronunciado por cada um deles. Durante esse processo, foram geradas avaliações parciais de cada vocábulo, que foram submetidas a uma análise detalhada das métricas de desempenho do modelo, incluindo precisão, *recall*, *F1-score*, além da elaboração de uma matriz de confusão. A avaliação parcial de cada vocábulo consistiu na classificação binária dos pacientes, atribuindo a cada um o diagnóstico de Parkinson (1) ou Não-Parkinson (0).

Para efeitos de melhor entendimento da base de dados, os vocábulos foram rotulados por uma letra do alfabeto. Coincidentemente, como havia 26 vocábulos disponíveis cada um foi rotulado com uma letra. Por exemplo, a vogal sustentada (aaa. . .) foi rotulada pela letra A, a vogal sustentada (ooo. . .) foi rotulada pela letra B, e assim sucessivamente.

3.1.3 Comitê avaliativo

Após a conclusão da fase de treinamento, foi instituído um critério no diagnóstico que foi posteriormente incorporado a um banco de dados genérico, a fim de aprimorar nossa compreensão da estrutura que desejava-se construir e organizar. O banco de dados genérico compreende as classificações parciais dos 26 vocábulos referentes aos 40 pacientes.

Após a análise inicial, cada paciente foi representado por uma matriz de 26x10, onde as linhas correspondiam às amostras de voz e as colunas às variáveis mais relevantes. A seleção das variáveis mais relevantes para o modelo é realizada por meio de um critério de correlação. Essas matrizes foram processadas pela função sigmoide, resultando em um vetor binário (0s e 1s) para cada paciente.

Ao término de cada avaliação por vocábulo, ocorria uma soma dos valores zero e um, onde a prevalência determinava o diagnóstico definitivo do paciente. Nessa etapa, foi realizada uma análise detalhada dos padrões de erro do modelo, identificando os vocábulos com maior frequência de classificações incorretas em relação ao diagnóstico esperado. Além disso, considerando a presença de um conjunto de 26 letras, existia

a possibilidade de uma situação em que a classificação resultasse em uma divisão equitativa de 13/13, caracterizando um cenário de indeterminação.

Para abordar essa probabilidade, foi desenvolvida uma função destinada a minimizar tais indeterminações. Essa função opera excluindo a variável com a menor correlação em relação ao diagnóstico final, a partir do conjunto das 10 variáveis mais fortemente correlacionadas com o diagnóstico. Em outros termos, a última coluna da base de dados, identificada como a de menor correlação, era eliminada. A Figura 8 ilustra a estrutura utilizada para a análise dos pacientes. Nesse esquema, cada linha representa um paciente (0,1,2, ... 10), enquanto as colunas correspondem às variáveis do modelo, representadas pelos vocábulos (A, B, C, ... Z). A coluna "ALVO" indica o diagnóstico real do paciente, enquanto a coluna "Resultado"exibe a classificação atribuída pelo algoritmo com base na contagem dos valores 0 e 1. Por exemplo, no caso do paciente 0, cujo diagnóstico real é Parkinson (1), o algoritmo também classificou como Parkinson (1). Adicionalmente, a coluna "Acertou?" verifica se a predição do algoritmo está em conformidade com o diagnóstico real.

Figura 8 – Resultado gerado no console após a execução do algoritmo e classificação dos pacientes com base no banco de dados genérico.

	A	B	D	E	F	G	H	I	J	K	L	N	O	P	Q	S	T	U	V	W	X	Y	Z	ALVO	quantidade_1	quantidade_0	resultado	Acertou?	
0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	20	6	1	sim
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	26	0	1	sim
2	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	1	1	1	1	21	5	1	sim
3	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	1	1	1	18	8	1	sim
4	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	0	1	0	1	1	19	7	1	sim
5	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	24	2	1	sim
6	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	23	3	1	sim
7	0	1	1	0	0	1	0	1	1	1	1	1	1	0	0	1	1	1	0	0	1	0	1	1	1	16	10	1	sim
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	24	2	1	sim
9	0	1	0	0	1	0	1	1	1	0	1	1	0	0	0	1	1	1	0	1	0	0	1	1	1	15	11	1	sim
10	0	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	19	7	1	sim

Fonte: Autoria própria.

3.1.4 Teste e validação

Como mencionado anteriormente, na fase inicial, abrangendo o treinamento do algoritmo, as ações foram conduzidas por meio do Octave. Posteriormente, para a aplicação e execução dos resultados obtidos no treinamento, a abordagem se baseou na linguagem Python com a utilização da biblioteca *Pandas* e a biblioteca *Numpy*. Nesse ponto, foi onde pôde ser visto a parte prática após o treinamento.

Durante esse processo, os dados individuais de cada um dos 40 pacientes foram extraídos a partir de arquivos Excel. No dataset original, todos os pacientes estavam agrupados em um único conjunto de dados; portanto, nesta fase, os dados foram separados por paciente. Em seguida, os parâmetros obtidos durante a fase de treinamento foram importados do Octave, realizando as operações matriciais. Os resultados dessas operações foram armazenados na variável "resultados", servindo

como base para a análise e avaliação do modelo.

Após a segmentação dos dados de cada paciente, e a exportação dos dados do Octave. A função de ativação sigmoide foi aplicada para calcular a probabilidade de diagnóstico, processo que foi otimizado com o uso da biblioteca SciPy. Os dados processados foram então submetidos a uma regra de decisão: se o valor resultante fosse superior a 0,5, o paciente era classificado como portador de Parkinson (1); caso contrário, era considerado saudável (0).

A Figura 9 mostra uma maneira mais clara de como foram os resultados da função sigmoide ao longo dos dados de um paciente. Este novo *dataframe* compreendia duas colunas: "Sigmoide" e "Classificação". A função de classificação atribuía "1" para "Parkinson", caso contrário, "Não-Parkinson" de acordo com a situação.

Figura 9 – Colunas com o resultado pós sigmoide e sua classificação

	Sigmoide	Classificação
0	0.000000e+00	não-Parkinson
1	0.000000e+00	não-Parkinson
2	0.000000e+00	não-Parkinson
3	1.000000e+00	Parkinson
4	1.000000e+00	Parkinson
5	4.117883e-140	não-Parkinson
6	1.000000e+00	Parkinson
7	1.000000e+00	Parkinson
8	0.000000e+00	não-Parkinson
9	1.000000e+00	Parkinson
10	0.000000e+00	não-Parkinson
11	1.956965e-239	não-Parkinson
12	1.000000e+00	Parkinson
13	1.000000e+00	Parkinson
14	1.000000e+00	Parkinson
15	1.000000e+00	Parkinson
16	1.000000e+00	Parkinson
17	1.000000e+00	Parkinson
18	1.000000e+00	Parkinson
19	1.000000e+00	Parkinson
20	1.000000e+00	Parkinson
21	0.000000e+00	não-Parkinson
22	0.000000e+00	não-Parkinson
23	1.000000e+00	Parkinson
24	1.000000e+00	Parkinson
25	1.083701e-62	não-Parkinson

Fonte: Elaboração própria.

Conforme anteriormente mencionado, quando ocorria uma equitativa classificação de 13/13, um cenário de indeterminação era identificado. Nesse contexto, a última coluna, a menos correlacionada, era removida e, posteriormente, o cálculo original do código era novamente iniciado. Subsequentemente, uma decisão era tomada pelo algoritmo com base na contagem dos valores zero e um. O diagnóstico definitivo era, então, exibido no console como exibido na Figura 10.

Figura 10 – Exemplo do que era exibido no console

```
quantidade de 1's: 16
quantidade de 0s: 10

*****DIAGNÓSTICO*****

Paciente tem Parkinson
```

Fonte: Elaboração própria.

3.2 A INTERFACE GRÁFICA PARKIN95

Esta seção descreve a camada de interação com o usuário, intitulada Parkin95, abordando as tecnologias utilizadas, o objetivo e o fluxo de interação. As funcionalidades serão abordados nos resultados e discussões.

3.2.1 Função e objetivo da aplicação

O ambiente visual desenvolvido, é um ponto principal neste trabalho, atuando como ponte intuitiva e eficiente entre o usuário (profissional de saúde) e o modelo preditivo da Doença de Parkinson. Sua funcionalidade abrange o cadastro simplificado de pacientes, a execução direta de predições e a visualização objetiva dos resultados.

A Parkin95 oferece uma interface gráfica intuitiva que permite a neurologistas utilizarem, de forma acessível, o modelo preditivo da Doença de Parkinson desenvolvido e treinado nas etapas anteriores deste trabalho, com base em Regressão Logística Regularizada. A aplicação foi projetada para simplificar a entrada dos dados clínicos, realizar a predição com o modelo previamente construído e exibir os resultados de forma clara, contribuindo para a agilidade e precisão na análise do risco diagnóstico.

A criação do sistema de interface, visou tornar o modelo preditivo acessível a usuários sem conhecimento técnico em programação ou ciência de dados. Ao eliminar a manipulação direta de código, a interface facilita o uso prático em ambientes clínicos, democratizando a inteligência artificial na saúde para profissionais da medicina de maneira segura, intuitiva e eficiente.

3.2.2 Tecnologias utilizadas

Para o desenvolvimento, utilizou-se a biblioteca streamlit por ser uma biblioteca Python de código aberto voltada à criação rápida de interfaces web interativas, especialmente adequada para projetos que envolvem ciência de dados e aprendizado de máquina. Sua sintaxe simples e integração nativa com bibliotecas como *Pandas*,

NumPy e *Scikit-learn* possibilitaram a construção de uma interface funcional sem a necessidade de conhecimentos avançados em desenvolvimento web. Além disso, o *Streamlit* permite a prototipagem ágil e a visualização clara dos resultados, o que se alinha ao objetivo do projeto de oferecer uma ferramenta acessível e de uso intuitivo para profissionais da saúde.

O *SQLite* foi adotado como sistema de gerenciamento de banco de dados (SGBD) por ser leve, autônomo e de fácil integração com aplicações Python. Por não exigir um servidor dedicado, o *SQLite* se mostrou ideal para o escopo do projeto *Parkin95*, que visa oferecer uma solução simples e portátil. A manipulação eficiente de dados em aplicações de pequeno a médio porte, como o armazenamento de informações de usuários e pacientes, é facilitada pela estrutura em arquivos do SGBD. Essa escolha também simplificou a implementação, preservando a confiabilidade e a segurança no acesso aos dados.

A estruturação do projeto adotou o padrão arquitetural *Model-View-Controller* (MVC), visando uma clara separação de responsabilidades entre a lógica de negócios (*Model*), a interface com o usuário (*View*) e o controle do fluxo de dados (*Controller*). Essa abordagem promove uma melhor organização do código, facilitando a manutenção, o desenvolvimento em equipe e a escalabilidade da aplicação (Buschmann *et al.*, 1996).

4 RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos com base na metodologia aplicada, estruturados em duas partes, de forma análoga à organização metodológica. A primeira parte refere-se à etapa de treinamento e avaliação do modelo preditivo, onde são apresentados os dados relacionados ao desempenho do algoritmo de Regressão Logística Regularizada, com foco em métricas como acurácia, precisão, *recall* e *F1-score*. A segunda parte descreve os resultados relacionados à interface gráfica Parkin95, desenvolvida com o objetivo de integrar os dados analisados e fornecer uma ferramenta prática de apoio ao diagnóstico.

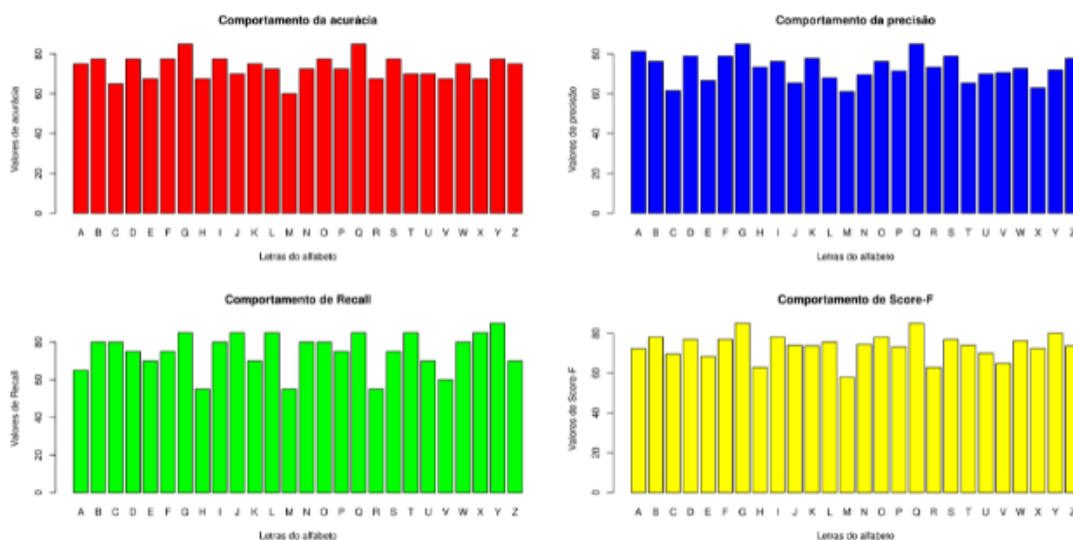
4.1 RESULTADOS DO TREINAMENTO DO MODELO

Nesta subseção, são apresentados os resultados obtidos durante o processo de treinamento do modelo preditivo, incluindo métricas de desempenho, validação e análises relevantes que demonstram a eficácia da abordagem utilizada.

Durante a etapa de treinamento do modelo, utilizou-se o algoritmo de Regressão Logística Regularizada, o qual foi alimentado com um conjunto de dados composto por registros vocais de 40 pacientes. O foco da análise esteve nas características fonéticas dos vocábulos pronunciados por cada participante, permitindo que o modelo identificasse padrões associados à presença ou ausência da Doença de Parkinson. Cada vocábulo pronunciado foi analisado individualmente, gerando avaliações parciais que possibilitaram observar o desempenho do modelo em diferentes contextos fonéticos. Essas avaliações foram posteriormente submetidas à análise de métricas de desempenho amplamente utilizadas em aprendizado de máquina, como precisão, *recall* e *F1-score*, a fim de quantificar a eficácia do modelo.

O objetivo dessa análise parcial era realizar uma classificação binária para cada paciente com base nos dados vocais: atribuir 1 para casos identificados como Parkinson e 0 para os que não apresentavam sinais da doença. Isso possibilitou um diagnóstico automatizado inicial a partir de informações fonéticas específicas. Como resultado da etapa de treinamento, foram obtidas avaliações parciais correspondentes a cada vocábulo, conforme ilustrado na Figura 11.

Figura 11 – Avaliações parciais obtidas durante o treinamento do modelo para cada vocábulo analisado.



Fonte: Elaboração Própria.

Após essa análise inicial, foi elaborado um gráfico que destaca quais vocábulos (letras) apresentaram maior taxa de erro na classificação dos pacientes. O objetivo foi identificar quais sons vocais geravam mais incertezas ou confusões para o modelo durante a tarefa de diagnóstico. Para essa avaliação comparativa, a métrica utilizada foi a precisão, que indica a proporção de pacientes corretamente identificados como portadores da Doença de Parkinson em relação ao total de diagnósticos positivos realizados pelo modelo.

Ao observar os vocábulos com menor precisão, foi possível inferir quais sinais vocais estavam mais associados a erros de diagnóstico, apontando para possíveis limitações na representação fonética desses sons ou maior variabilidade entre os pacientes. A Figura 12 foi elaborada para ilustrar de forma mais clara o comportamento do erro de diagnóstico por vocábulo, destacando aquelas que apresentaram maior taxa de erro.

Figura 12 – Comportamento do erro de diagnóstico por vocábulo

Fonte: Elaboração própria.

Com base na análise anterior, os vocábulos que apresentaram maior correlação com erros de diagnóstico — ou seja, aqueles que contribuíram significativamente para classificações incorretas — foram identificados e posteriormente removidos do conjunto de dados. Essa estratégia teve como objetivo avaliar se a exclusão de sinais vocais menos confiáveis poderia impactar positivamente o desempenho geral do modelo.

Após a retirada desses vocábulos, o algoritmo de Regressão Logística Regularizada foi novamente executado, desta vez com um conjunto de dados mais enxuto e teoricamente mais robusto. A nova rodada de treinamento e avaliação possibilitou uma análise comparativa entre o desempenho anterior e o atual, destacando possíveis melhorias na capacidade preditiva do modelo.

Os resultados dessa nova abordagem estão representados por meio da matriz de confusão Tabela 2, que evidencia a relação entre as classificações reais e previstas, permitindo avaliar métricas como acurácia, precisão, *recall* e *F1-score* em um cenário livre dos vocábulos mais propensos a erro.

Tabela 2 – Matriz de confusão resultante após a remoção dos vocábulos mais associados a erros de diagnóstico..

	Classe Preditiva: Positiva	Classe Preditiva: Negativa
Classe Real: Positiva	20	0
Classe Real: Negativa	2	18

Fonte: Elaboração própria

Com base na matriz de confusão gerada após a exclusão dos vocábulos mais

propensos a erro, foi possível calcular as principais métricas de desempenho do modelo. A acurácia refletiu a proporção total de classificações corretas, considerando tanto pacientes diagnosticados corretamente com Parkinson quanto aqueles corretamente identificados como saudáveis.

A precisão indicou a capacidade do modelo de acertar quando previu a presença da doença, ou seja, a proporção de verdadeiros positivos em relação ao total de diagnósticos positivos realizados. Já o *recall* representou a proporção de pacientes com Parkinson corretamente identificados entre todos os que realmente possuíam a doença.

Por fim, o *F1-score*, métrica harmônica entre precisão e *recall*, forneceu uma visão equilibrada do desempenho do modelo. Esses resultados permitiram uma avaliação mais precisa da eficácia do modelo após a etapa de refinamento do conjunto de dados.

Tabela 3 – Métricas de desempenho do modelo de Regressão Logística Regularizada

Métrica	Valor (%)
Acurácia	95,00
Precisão	90,90
Recall	100,00
Score-F1	95,23

Fonte: E laboração própria.

Os resultados indicaram que o algoritmo estudado, Regressão Logística Regularizada, apresentou-se como uma abordagem promissora para a detecção da Doença de Parkinson em estágios iniciais. A integração entre análise de dados e prática clínica revela um potencial significativo para aprimorar o diagnóstico e o tratamento de doenças neurológicas.

4.2 RESULTADOS DA INTERFACE PARKIN95

Esta seção descreve os principais resultados obtidos no processo de desenvolvimento e funcionamento da interface gráfica Parkin95, voltada ao apoio diagnóstico da Doença de Parkinson. A interface foi registrada como propriedade intelectual junto ao Instituto Nacional da Propriedade Industrial (INPI), sob o número BR512025002207-7. São abordados aspectos relacionados à usabilidade da aplicação, à integração com o modelo preditivo treinado e às funcionalidades implementadas com o objetivo de fornecer suporte eficiente e acessível a profissionais da área da saúde. Destacam-se sua usabilidade, a integração com o modelo preditivo treinado e as funcionalidades

implementadas para auxiliar no diagnóstico da Doença de Parkinson. Ao acessar a aplicação via web, o usuário é recebido pela Página Inicial, que apresenta brevemente as informações e o objetivo da Parkin95, contextualizando o propósito da ferramenta no apoio ao diagnóstico da Doença de Parkinson. A Figura 13 apresenta a tela inicial do classificador, conforme desenvolvida no projeto.

Figura 13 – Tela inicial do Parkin95

NAVEGAÇÃO >



PARKIN95 - Predição do diagnóstico da doença de Parkinson

Informações

O aplicativo foi desenvolvido para prever se os pacientes têm Parkinson com base em informações específicas coletadas. Ele utiliza algoritmo de Regressão Logística Regularizada para analisar dados fornecidos pelos usuários, que são os tremores nos sinais de voz dos pacientes. Ao processar essas informações, o aplicativo gera uma avaliação que indica a probabilidade de o usuário ter a doença de Parkinson. O objetivo da ferramenta é auxiliar e melhorar a precisão diagnóstica na identificação precoce da condição. Vale ressaltar que os resultados do aplicativo não substituem uma avaliação médica profissional, sendo uma ferramenta complementar para a detecção precoce da doença.

Pesquisa desenvolvida por Antônio Alisson, Ivina Lorena 2023-2024

Fonte: Elaboração Própria.

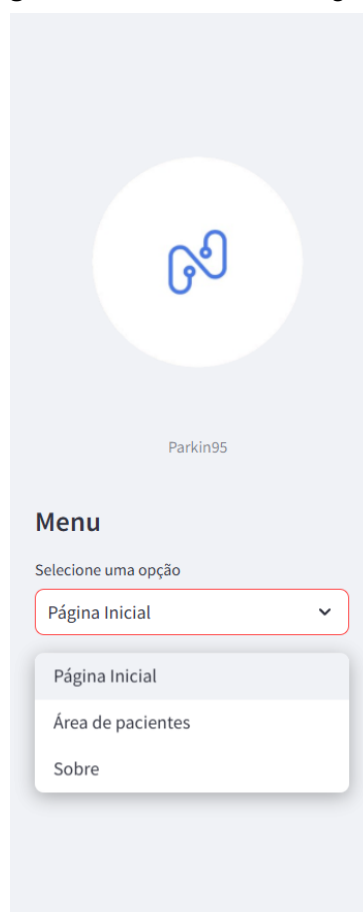
No canto superior esquerdo, encontra-se a barra de navegação, que permite o acesso às funcionalidades da interface, os quais estão representados na Figura 14.

Figura 14 – Tela inicial com o sidebar



Fonte: Elaboração Própria.

A Figura 15 apresenta a barra lateral (*sidebar*) e as áreas disponíveis para navegação e seleção pelo usuário.

Figura 15 – Áreas de navegação

Fonte: Elaboração própria.

Na lateral esquerda da tela, encontra-se o menu lateral (*sidebar*), onde o usuário pode navegar até a seção "Área de pacientes". Ao acessá-la, é exibido um espaço centralizado contendo um modelo de planilha disponível para download, o qual serve como guia para o preenchimento correto dos dados clínicos necessários à análise. Esse arquivo orienta o usuário quanto aos vocábulos e métricas extraídas dos pacientes que serão cadastrados (Figura 16).

Logo abaixo, a interface apresenta duas seções expansíveis (*expanders*):

A primeira, intitulada "Cadastro de Pacientes", contém os campos necessários para o registro de um novo paciente. Ao preencher e submeter essas informações, os dados são automaticamente armazenados no banco de dados da aplicação.

A segunda, chamada "Lista de Pacientes", exibe uma tabela com todos os registros cadastrados. Ao lado de cada paciente, há dois botões de ação: "Resultado", responsável por realizar a comunicação direta com o modelo preditivo treinado, retornando a classificação diagnóstica para o paciente selecionado e "Excluir", que

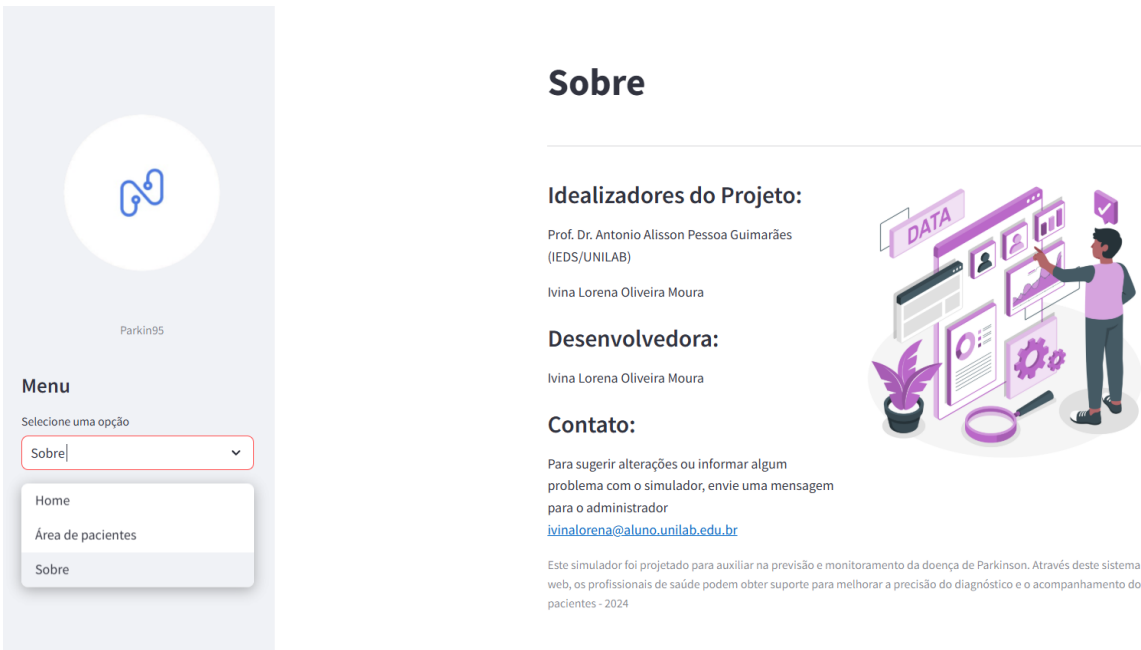
remove permanentemente o paciente da tabela, eliminando também seu diagnóstico e informações pessoais associadas.

Figura 16 – Página Área de pacientes



Fonte: Elaboração própria.

A tela "Sobre"(Figura 17), acessível através da barra de navegação lateral, essa seção também disponibiliza dados institucionais e pessoais da desenvolvedora, como nome, curso, universidade e orientador, permitindo que o usuário conheça o contexto acadêmico em que o projeto foi desenvolvido.

Figura 17 – Tela com informações institucionais e acadêmicas dos responsáveis pelo projeto.


Sobre

Idealizadores do Projeto:
 Prof. Dr. Antonio Alisson Pessoa Guimarães (IEDS/UNILAB)
 Ivina Lorena Oliveira Moura

Desenvolvedora:
 Ivina Lorena Oliveira Moura

Contato:
 Para sugerir alterações ou informar algum problema com o simulador, envie uma mensagem para o administrador
ivinalorena@aluno.unilab.edu.br

Este simulador foi projetado para auxiliar na previsão e monitoramento da doença de Parkinson. Através deste sistema web, os profissionais de saúde podem obter suporte para melhorar a precisão do diagnóstico e o acompanhamento dos pacientes - 2024

Fonte: Elaboração própria.

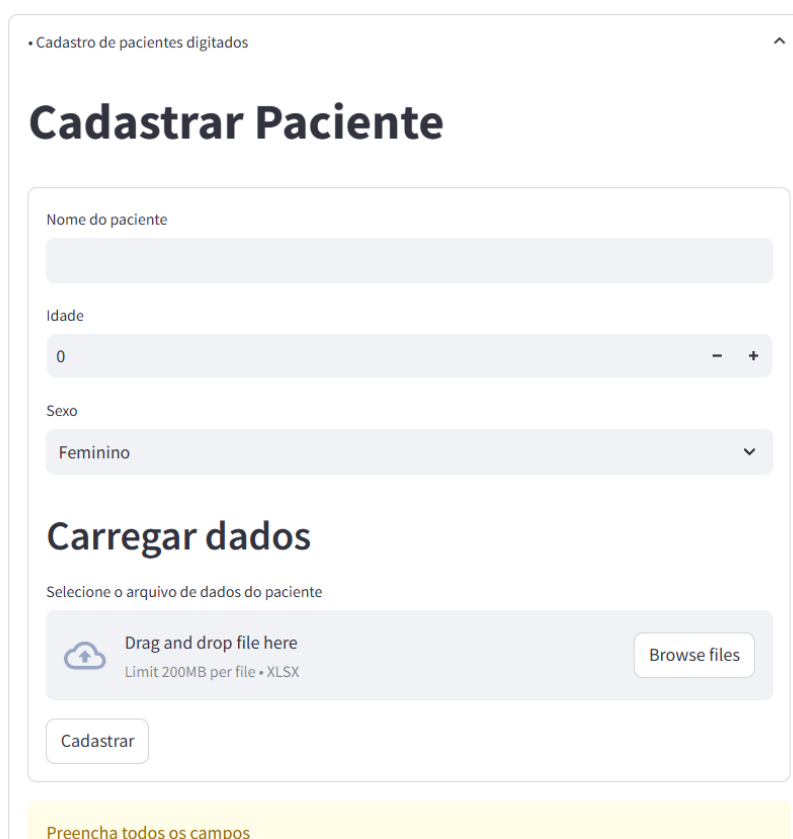
4.2.1 Funcionalidades

O Parkin95 foi desenvolvido para ser simples e direto. As funcionalidades implementadas visam atender às necessidades principais do processo de diagnóstico, garantindo praticidade na operação. As funcionalidades:

- Cadastro e visualização de pacientes

A interface permite o cadastro de novos pacientes por meio da inserção de informações clínicas relevantes para a predição, como medidas obtidas por meio de análises de voz.

Na Figura 18, é apresentada a interface de cadastro de pacientes. Para registrar um novo paciente, é necessário preencher os campos obrigatórios: nome, idade, sexo e enviar o arquivo contendo os dados vocais correspondentes. Após o preenchimento, as informações são armazenadas no banco de dados. O gerenciamento dos dados e a execução das consultas (*queries*) foram realizados por meio do sistema SQLite, com o apoio da ferramenta *DB Browser for SQLite*. Essas informações serão exibidas na seção "Lista de Pacientes", uma aba dedicada a apresentar todos os registros previamente cadastrados no banco de dados. Nessa área, o usuário poderá visualizar, consultar e gerenciar os dados dos pacientes de forma organizada e eficiente.

Figura 18 – Formulário para preenchimento dos dados do paciente

• Cadastro de pacientes digitados

Cadastrar Paciente

Nome do paciente

Idade

 - +

Sexo

 ▾

Carregar dados

Selecione o arquivo de dados do paciente

Drag and drop file here
Limit 200MB per file • XLSX

Preencha todos os campos

Fonte: Elaboração própria.

Na seção "Lista de Pacientes", serão exibidos os pacientes já cadastrados (Figura 19), acompanhados de um botão denominado 'Resultado'. Ao clicar nesse botão, o sistema processará os dados e apresentará a predição do diagnóstico para o paciente selecionado (Figura 20).

Figura 19 – Área dos pacientes já cadastrados

ID	Nome	Idade	Excluir	Predição
1	Paciente 12	77	<input type="button" value="Excluir"/>	<input type="button" value="Resultado"/>

Este simulador foi projetado para auxiliar na previsão e monitoramento da doença de Parkinson. Através deste sistema web, os profissionais de saúde podem obter suporte para melhorar a precisão do diagnóstico e o acompanhamento dos pacientes - 2024

Fonte: Elaboração própria.

- Predição da Doença de Parkinson

Com base nos dados inseridos, a aplicação realiza a comunicação direta com

o modelo preditivo treinado anteriormente. A predição é feita de forma automática e imediata, classificando se o paciente possui ou não indícios compatíveis com a Doença de Parkinson, segundo os padrões aprendidos pelo algoritmo de Regressão Logística Regularizada.

Figura 20 – Resultado exibido na tela para o usuário

A interface apresenta uma seção intitulada "Lista de Pacientes" com uma tabela de dados. Abaixo da tabela, há um cabeçalho "Resultado:" seguido de uma caixa de texto destacada em rosa que contém o resultado da predição.

ID	Nome	Idade	Excluir	Predição
1	Paciente 12	77	Excluir	Resultado

Resultado:

O paciente Paciente 12 apresenta uma **predisposição** para a doença de Parkinson.

Este simulador foi projetado para auxiliar na previsão e monitoramento da doença de Parkinson. Através deste sistema web, os profissionais de saúde podem obter suporte para melhorar a precisão do diagnóstico e o acompanhamento dos pacientes - 2024

Fonte: Elaboração própria.

5 CONCLUSÕES E PERSPECTIVAS FUTURAS

O desenvolvimento deste trabalho resultou em uma aplicação funcional, voltada para o apoio ao diagnóstico da Doença de Parkinson com base em características vocais dos pacientes. Com a utilização da Regressão Logística Regularizada como algoritmo preditivo, foi possível alcançar métricas de desempenho consistentes, incluindo alta acurácia, elevado *recall* e boa precisão. Esses resultados evidenciam o potencial do modelo na detecção precoce da doença, contribuindo para o aprimoramento de ferramentas de suporte à decisão clínica.

Além disso, a construção de uma interface gráfica com o uso da biblioteca Streamlit permitiu uma visualização clara dos resultados e facilitou a interação por parte dos usuários finais, especialmente profissionais da área da saúde. A organização modular do sistema, aliada ao uso de banco de dados com SQLite, contribuiu para a escalabilidade e manutenção do projeto.

Como aprimoramento futuro, propõe-se a substituição do ambiente de desenvolvimento atual (Streamlit) por um *framework* mais robusto e flexível, como o Flask. Essa migração permitirá um controle mais detalhado das rotas, integração facilitada com *front-ends* personalizados (HTML/CSS/JavaScript) e maior liberdade para implementar recursos mais complexos, como autenticação avançada, *dashboards* interativos e integrações com APIs externas.

Outras possibilidades de expansão incluem:

- Implementação de novos algoritmos de aprendizado de máquina para comparação de desempenho.
- Validação do modelo com bases de dados mais amplas e diversificadas.
- Inclusão de novos tipos de dados biomédicos (como sinais de movimento ou imagens médicas) para enriquecer a base de análise.
- Implantação com autenticação segura para uso por profissionais da saúde.
- Gravações de voz automática.
- Comparação com outros algoritmos de aprendizado de máquina: Implementar e avaliar modelos alternativos, como Random Forest, XGBoost, SVM ou redes neurais profundas, a fim de comparar seu desempenho com o modelo atual.

- Análise comparativa com estudos relacionados: Investigar artigos científicos que utilizam a mesma base de dados e realizar uma comparação sistemática de desempenho, considerando métricas como acurácia, sensibilidade, especificidade e F1-score.
- Exploração de diferentes técnicas de pré-processamento e seleção de atributos: Avaliar o impacto de diferentes métodos de normalização, imputação e redução de dimensionalidade sobre os resultados do modelo.

Conclui-se, portanto, que os resultados alcançados até aqui representam um avanço significativo na aplicação de técnicas de aprendizado de máquina voltadas à saúde, abrindo caminho para futuras melhorias e validações clínicas.

REFERÊNCIAS

- AI, Evidently. **Recall: A Metric for Classification**. 2023. Acesso em: 18 mar. 2025. Disponível em: <https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall#:~:text=Recall%20is%20a%20metric%20that,the%20number%20of%20positive%20instances>.
- AMAZON. **Amazon Machine Learning Developer Guide**. [S.l.: s.n.]: Amazon Web Services, Inc. and/or its affiliates, 2018.
- ANHESINI, Victória. **Relógios e anéis: conheça 5 wearables para monitorar a saúde e o bem-estar**. 2024. Publicado em 23 de novembro de 2024. Atualizado 5 meses atrás. Disponível em: <https://www.infomoney.com.br/saude/relogios-e-aneis-conheca-5-wearables-para-monitorar-a-saude-e-bem-estar/>.
- AWARI. **Streamlit: A Biblioteca Python que Revoluciona o Desenvolvimento de Aplicações Web**. 2023. Disponível em: <https://awari.com.br/streamlit-python/>. Acesso em: 28 fev. 2023.
- AZEVEDO, Luciana Lemos de; CARDOSO, Francisco. Ação da levodopa e sua influência na voz e na fala de indivíduos com doença de parkinson. **Revista da Sociedade Brasileira de Fonoaudiologia**, v. 14, n. 1, p. 21, 2009. Acesso em: 30 nov. 2024. Disponível em: <https://doi.org/10.1590/S1516-80342009000100021>.
- BALESTRINO, R.; SCHAPIRA, A. H. V. Parkinson disease. **European Journal of Neurology**, v. 27, n. 1, p. 27–42, 2019.
- BITTENCOURT, Hélio Radke. **Regressão logística politômica: revisão teórica e aplicações**. 2003. Trabalho de Conclusão de Curso.
- BUSCHMANN, Frank; MEUNIER, Regine; ROHNERT, Hans; SOMMERLAD, Peter; STAL, Michael. **Pattern-Oriented Software Architecture: A System of Patterns**. Chichester, England: John Wiley & Sons, 1996.
- CABREIRA, V.; MASSANO, J. Doença de parkinson: Revisão clínica e atualização. **Acta Médica Portuguesa**, v. 32, n. 10, p. 661–670, 2019.
- CHEN, L. Heart disease prediction utilizing machine learning techniques. **Transactions on Materials, Biotechnology and Life Sciences**, v. 3, p. 35–50, 2024.
- CHOU, Kelvin L. **Clinical manifestations of Parkinson disease**. [S.l.: s.n.]: UpToDate, 2020. https://www.uptodate.com/contents/clinical-manifestations-of-parkinson-disease?search=doenca%20de%20parkinson&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Acesso em: 30 abril 2025.
- COSTA, Jônntas Lennon Lima; PAIXÃO, Flávio Costa. Trabalho de Conclusão de Curso (Engenharia de Software), **Sistema para Auxiliar a Detecção da Doença de Parkinson Interpretando Sinais de Tremor de Repouso Parkinsoniano Utilizando Aprendizado de Máquinas**. Brasília, DF: [S.l.: s.n.], 2018. Faculdade UnB Gama (FGA). Orientadora: Dra. Lourdes Mattos Brasil. Coorientador: Bel. Ithallo Junior Alves Guimarães.

- COUDRAY, P. M. *et al.* Machine learning models in early diagnosis of parkinson's disease: A review. **IEEE Reviews in Biomedical Engineering**, v. 14, p. 213–225, 2021.
- DEEPAI. **F-score - Machine Learning Glossary and Terms**. 2023. Acessado em: 10 out. 2023. Disponível em: <https://deepai.org/machine-learning-glossary-and-terms/f-score>.
- DELAHOZ, Y. S.; LABRADOR, M. A. Survey on fall detection and fall prevention using wearable and external sensors. **Sensors**, v. 14, p. 19806–19842, 2014.
- DORSEY, E. R.; ELBAZ, A. Global, regional, and national burden of parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. **The Lancet Neurology**, Elsevier, v. 17, n. 11, p. 939–953, 2018.
- FAGUNDES, Thales Pardini; WICHMANN, Roberta Moreira; SILVA, Maria Eduarda de Lima e; OLIVEIRA, Tiago Almeida de. Big data em saúde do trabalhador: o quão distantes estamos? **Revista de Saúde Pública**, v. 49, 2024. Edição especial. Disponível em: <https://doi.org/10.1590/2317-6369/40522pt2024v49edcinq11>.
- FAHN, S.; OAKES, D.; SHOULSON, I.; AL. *et.* Levodopa and the progression of parkinson's disease. **The New England Journal of Medicine**, v. 351, n. 24, p. 2498–2508, 2004.
- FRASER, Kathleen C.; MELTZER, Jed A.; RUDZICZ, Frank. Linguistic features identify alzheimer's disease in narrative speech. **Journal of Alzheimer's Disease**, v. 49, n. 2, p. –, 2015.
- Geeks for Geeks. **Reinforcement Learning**. 2024. Acesso em: 14 dez. 2024. Disponível em: <https://www.geeksforgeeks.org/what-is-reinforcement-learning/>.
- GONÇALVES, L. H. T.; ALVAREZ, A. M.; ARRUDA, M. C. Pacientes portadores da doença de parkinson: significado de suas vivências. **Acta Paulista de Enfermagem**, v. 20, n. 1, p. 62–68, 2007.
- GUEDES, Bruna Campos. **Métodos de Correlação Aplicados a Imagens de Ressonância Magnética Funcional**. Uberlândia, Brasil, 2022.
- GUO, G.; WANG, H.; BELL, D.; BI, Y.; GREER, K. Knn model-based approach in classification. *In*: MEERSMAN, R.; TARI, Z.; SCHMIDT, D.C. (ed.). **On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE**. Springer, Berlin, Heidelberg, 2003, (Lecture Notes in Computer Science, v. 2888). p. 986–996. Disponível em: https://doi.org/10.1007/978-3-540-39964-3_62.
- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2nd. ed. [S.l.: s.n.]: O'Reilly Media, 2019. ISBN 978-1492032649.
- HAYES, M. T. Parkinson's disease and parkinsonism. **The American Journal of Medicine**, Elsevier, v. 132, n. 7, p. 802–807, 2019.
- HINTON, G.E.; SEJNOWSKI, T.J.; POGGIO, T.A. **Unsupervised Learning: Foundations of Neural Computation**. Cambridge, MA, USA: MIT Press, 1999.

HOSMER, D.; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley & Sons, 1989.

HOWARD, J. Artificial intelligence: implications for the future of work. **American Journal of Industrial Medicine**, v. 62, p. 917–926, 2019.

IBM. **O que é k-Nearest Neighbors (kNN)?** 2024. Acessado em: 21 fev. 2025. Disponível em: <https://www.ibm.com/br-pt/topics/knn>.

IZBICKI, Rafael; SANTOS, Tiago Mendonça dos. **Aprendizado de máquina: uma abordagem estatística**. 1. ed. S.l.: s.n., 2020. 272 p.

KURSUN OLCAY, Sakar Betul Isenkul M. Sakar C. Sertbas Ahmet; GURGEN, Fikret. **Parkinson's Speech with Multiple Types of Sound Recordings**. 2013. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5NC8M>.

LANG, A. E.; LOZANO, A. M. Parkinson's disease. first of two parts. **The New England Journal of Medicine**, v. 339, n. 16, p. 1044–1053, 1998.

LENNON, Annie. **AI tool detects Parkinson's up to 15 years early with 96% accuracy**. 2023. Fact checked by Hannah Flynn. Disponível em: https://www-medicalnewstoday-com.translate.google.com/articles/ai-tool-detects-parkinsons-up-to-15-years-early-with-96-accuracy?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc.

LITTLE, Max; MCSHARRY, Patrick; HUNTER, Eric; SPIELMAN, Jennifer; RAMIG, Lorraine. Suitability of dysphonia measurements for telemonitoring of parkinson's disease. **IEEE transactions on bio-medical engineering**, v. 56, p. 1015, 04 2009.

Lone Star Neurology. **Wearable Devices for Parkinson's Disease: Technological Solution**. 2023. Publicado em 17 de outubro de 2023. Atualizado em 10 de julho de 2024. Revisão médica por Ramin Ansari. Autor não especificado. Disponível em: https://lonestarneurology-net.translate.google.com/blog/wearable-devices-for-parkinsons-disease/?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=sge#:~:text=Pacientes%20de%20Parkinson%20podem%20gerenciar,resultados%20entre%20os%20dois%20dispositivos.

LUCIANO, Josué; ALVES, Wallison Joel Barberá. Padrão de arquitetura mvc: Model-view-controller. **EPeQ Fafibe**, v. 1, n. 3a, p. 102–107, 2017.

LUDERMIR, Teresa Bernarda. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, Instituto de Estudos Avançados da Universidade de São Paulo, v. 35, n. 101, p. 85–94, jan. 2021.

MARENGO, Lívia Luize; KOZYREFF, Alan Martinez; MORAES, Fabio da Silva; MARICATO, Laura Inês Gomes; BARBERATO-FILHO, Silvio. Tecnologias móveis em saúde: reflexões sobre desenvolvimento, aplicações, legislação e ética. **Revista panamericana de salud pública**, SciELO Public Health, v. 46, p. e37, 2023.

MCKINNEY, Wes. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython**. [S.l.]: O'Reilly Media, Inc., 2012.

- Medscape Notícias Médicas. **Dispositivos vestíveis para a doença de Parkinson: o futuro está aqui**. 2024. Texto publicado em 5 de dezembro de 2024. Autor individual não especificado; fonte: WebMD, LLC. Disponível em: <https://portugues.medscape.com/verartigo/6512052#:~:text=Notavelmente%2C%20os%20dispositivos%20Kinesia%20360,remoto%20da%20doen%C3%A7a%20de%20Parkinson>.
- Merritt, H. **Merrit Tratado de Neurologia**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2002.
- MONTEIRO, E. P.; WILD, L. B.; MARTINEZ, F. G.; PAGNUSSAT, A. de Souza; PEYRÉ-TARTARUGA, L. A. Aspectos biomecânicos da locomoção de pessoas com doença de parkinson: revisão narrativa. **Revista Brasileira de Ciências do Esporte**, v. 39, n. 4, p. 450–457, 2017. Disponível em: <https://doi.org/10.1016/j.rbce.2016.07.003>.
- MORAES, Joel Junior de; SILVA, Jéssica Oliveira; SANTOS, Marcos Vinícius; LIMA, Pedro Henrique. Impacto da tecnologia de inteligência artificial na medicina diagnóstica. **REASE - Revista Eletrônica de Administração e Saúde**, v. 10, n. 11, p. 1301–1307, 2023. Acesso em: 04 de maio de 2025. Disponível em: <https://periodicorease.pro.br/rease/article/view/10699>.
- MOREAU, C.; ROUAUD, T.; GRABLI, D.; AL. et. Visão geral sobre sensores vestíveis para o tratamento da doença de parkinson. **npj Parkinson's Disease**, v. 9, n. 153, 2023.
- MOREIRA, Camilla Silveira; MARTINS, Kamilly Farah Cardoso; NERI, Vanderson Carvalho; ARAÚJO, Paulo Gustavo. Doença de parkinson: como diagnosticar e tratar. **Revista Científica da Faculdade de Medicina de Campos**, v. 2, n. 2, 2007. Artigo de Revisão. Faculdade de Medicina de Campos, Campos dos Goytacazes, RJ, Brasil.
- NORVIG, Peter; RUSSELL, Stuart. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.
- ONODERA, H.; OKABE, S.; KIKUCHI, Y.; TSUDA, T.; ITOYAMA, Y. Impaired chemosensitivity and perception of dyspnoea in parkinson's disease. **The Lancet**, v. 356, n. 9231, p. 739–740, 2000. Comment in: *Lancet*. 2000;356(9247):2099, *Lancet*. 2000;356(9247):2100.
- Organização Mundial da Saúde. **Doença de Parkinson: uma abordagem de saúde pública: resumo técnico**. [S.l.], 2022. 32 p. Acessado em 19 de maio de 2025. Disponível em: https://www-who-int.translate.google.com/publications/i/item/9789240050983?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc.
- PAPST, Nathalia. **A Arquitetura de Código MVC e Como Implementá-la**. 2021. Acessado em: 12 mar. 2025. Publicado na Apple Developer Academy | Mackenzie. Disponível em: <https://medium.com/mackmobile/a-arquitetura-de-c%C3%B3digo-mvc-e-como-implement%C3%A1-la-eb58ca5f1006>.
- RAJPUT, Alex; NOYES, Eric. Web Page, **Doença de Parkinson**. Manual MSD, 2024. Disponível em: <https://www.msdmanuals.com/pt/profissional/dist%C3%B2rbios-neurol%C3%B3gicos/transtornos-de-movimento-e-cerebelares/doenca-de-parkinson>.

SADEGHI, Zahra; ALIZADEHSANI, Roohallah; CIFCI, Mehmet Akif; KAUSAR, Samina; REHMAN, Rizwan; MAHANTA, Priyakshi; BORA, Pranjali Kumar; ALMASRI, Ammar; ALKHAWALDEH, Rami S.; HUSSAIN, Sadiq; ALATAS, Bilal; SHOEIBI, Afshin; MOOSAEI, Hossein; HLADÍK, Milan; NAHAVANDI, Saeid; PARDALOS, Panos M. A review of explainable artificial intelligence in healthcare. **Computers and Electrical Engineering**, v. 118, p. 109370, 2024. ISSN 0045-7906. Disponível em: <https://doi.org/10.1016/j.compeleceng.2024.109370>.

SAKAR, Betül Erdogdu; ISENKUL, M. Erdem; SAKAR, C. Okan; SERTBAS, Ahmet; GURGEN, Fikret; DELIL, Sakir; APAYDIN, Hulya; KURSUN, Olcay. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. **IEEE Journal of Biomedical and Health Informatics**, Institute of Electrical and Electronics Engineers Inc., v. 17, n. 4, p. 828–834, 2013.

SARACCO, Roberto. **Congrats Xiaoyi. You are now a medical doctor**. 2017. IEEE Future Directions.

Scikit-learn. **Classification metrics**. 2025. https://scikit-learn.org/stable/modules/model_evaluation.html#precision-recall-f-measure-metrics. Acesso em: 1 maio 2025.

SILVA, Jorge Antônio Félix da. **Análise Comparativa de Algoritmos de Aprendizagem de Máquina para Diagnóstico da Doença de Parkinson com Base em Características Vocais**. 2024. Trabalho de Conclusão de Curso, Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Acarape.

SILVA, Thaianne Pereira da; CARVALHO, Claudia Reinoso Araujo de. Doença de parkinson: o tratamento terapêutico ocupacional na perspectiva dos profissionais e dos idosos. **Cad. Bras. Ter. Ocup.**, v. 27 (2), p. 331–344, 2019.

SILVA, Vitória M. B. da; CAVALCANTI, Raquel S.; VIANA, Welainny F.; SOUZA, Juliana G. de; GOMES, Bianca de L. X.; PEREIRA, Jessiane M. S.; SANTOS, Wellington P. dos. Diagnóstico precoce da doença de parkinson a partir de sinais eletroencefalográficos e inteligência artificial. Recife, Brasil, 2023.

SINGLETON, A. B.; FARRER, M. J.; BONIFATI, V. The genetics of parkinson's disease: Progress and therapeutic implications. **Movement Disorders**, v. 28, n. 1, p. 14–23, 2013.

STEGMANN, G.M.; HAHN, S.; LISS, J. *et al.* Early detection and tracking of bulbar changes in als via frequent and remote speech analysis. **npj Digital Medicine**, v. 3, n. 132, 2020. Disponível em: <https://doi.org/10.1038/s41746-020-00335-x>.

SZEPESVÁRI, Csaba. **Algorithms for reinforcement learning**. San Rafael, CA, USA: Morgan & Claypool, 2010. v. 4. 1–103 p. (Synthesis Lectures on Artificial Intelligence and Machine Learning, v. 4).

TSANAS, Athanasios; LITTLE, Max A; MCSHARRY, Patrick E; RAMIG, Lorraine O. Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 57, n. 4, p. 884–893, 2010.

VAILLANCOURT, David E.; BARMPOUTIS, Angelos; WU, Shengdi S. *et al.* Automated imaging differentiation for parkinsonism. **JAMA Neurology**, 2025. Published online March 17, 2025.

WANG, H. **Nearest Neighbours without k: A Classification Formalism based on Probability**. [S./], 2002.

YADAV, A.; GHOSH, S. A novel machine learning framework for the prediction of parkinson's disease using speech and motor data. **Journal of Medical Systems**, v. 44, n. 7, p. 1–9, 2020.