

Uma Abordagem Multi Agentes de IA para Análise Multimodal e Gerenciamento Personalizado Da Doença De Parkinson

Eugenio Peixoto Junior
Department of Teleinformatics Engineering
Federal University of Ceará.
Fortaleza, Brazil
eugeniopeixotojr@alu.ufc.br

Felipe Cordeiro de Sousa
Department of Teleinformatics Engineering
Federal University of Ceará.
Fortaleza, Brazil
felipe.cordeiro@alu.ufc.br

Danilo Felipe Neto
Department of Teleinformatics Engineering
Federal University of Ceará
Fortaleza, Brazil
danilo.neto@alu.ufc.br

Victor Hugo C. de Albuquerque
Department of Teleinformatics Engineering.
Federal University of Ceará.
Fortaleza, Brazil
victor.albuquerque@ieee.org

Abstract—O trabalho tem como base em uma arquitetura colaborativa multimodal, composta por agentes autônomos especializados e impulsionados por Modelos de Linguagem de Larga ou Pequena Escala (LLMs/SLMs). A abordagem integra avanços recentes em DataOps, LLMops e VLMops, bem como estratégias de Recuperação Aumentada por Geração (RAG), incluindo variantes como Multimodal RAG, possibilitando a fusão eficiente de modalidades heterogêneas como texto clínico, exames de imagem, sinais sensoriais e gravações de voz, levando isso em conta, este estudo se diferencia por integrar, múltiplos modelos resultantes.

A metodologia inclui avaliações de custo computacional e eficiência operacional em ambientes simulados de produção, considerando métricas como uso de GPU, custo por chamadas de LLMs e escalabilidade da infraestrutura. Além disso, agentes inteligentes empregam técnicas de Aprendizado por Reforço (RL) em fase de pós-treinamento para adaptar estratégias de inferência com base em recompensas derivadas de métricas clínicas.

Os experimentos demonstraram desempenho competitivo entre diferentes classificadores, com destaque para o *Gradient Boosting* (acurácia de 83,85%, F1-score de 88,74%) e *HistGradient Boosting* (acurácia de 82,92%, F1-score de 88,12%, AUC-ROC de 0.909.), superando abordagens tradicionais de *Machine Learning* (ML) em robustez multimodal, porém em modelos convencionais, como SVM e KNN se mostraram, em alguns casos mais robustos em suas métricas de sensibilidade, identificando uma alta taxa de falsos positivos. A proposta se destaca pela capacidade de integração com fluxos clínicos reais, com potencial para se tornar um modelo de referência para plataformas neurointeligentes aplicadas à saúde digital.

Esses resultados reforçam a eficácia da abordagem proposta, demonstrando o potencial dos sistemas multiagentes colaborativos com RAG multimodal para a detecção precoce e precisa da Doença de Parkinson.

Os autores gostariam de agradecer ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) sob o número de concessão 15447/2023-0 e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código Financeiro 001 pelo apoio a esta pesquisa.

Index Terms—Doença de Parkinson, Multi-Agentes, Aprendizado de Máquina, Aprendizado Profundo, Redes Neurais Convolucionais, RAG multimodal, Diagnóstico médico, Fusão multimodal.

I. INTRODUÇÃO

A Doença de Parkinson (DP) é uma das doenças neurodegenerativas mais prevalentes no mundo, afetando cerca de 2 à 3% da população, principalmente nos idosos. Sua manifestação clínica abrange tanto sintomas motores, como, por exemplo, bradicinesia, rigidez e tremor em repouso, como também, não motores, incluindo alterações cognitivas, distúrbios do sono, sintomas depressivos e disfunções autonômicas [1]. Apesar dos sintomas não motores, que podem se manifestar muitos anos antes do diagnóstico clínico tradicional, sua natureza incerta e a falta de registros suficientes dificultam bastante a identificação precoce da doença.

Atualmente, o diagnóstico da DP permanece fundamentalmente clínico e, por consequência, pode depender da opinião de cada profissional, variar entre especialistas e ter dificuldades para realizar exames adicionais. Portanto, pesquisas por soluções tecnológicas possibilitam diagnósticos mais sensíveis, precoces e automatizados tem se intensificado [2]. A Inteligência Artificial (IA) já transformou a área da saúde e ainda tem muito potencial para novos avanços. Assim, a Engenharia Biomédica, especialmente na convergência com a IA, emerge como uma aliada no processo ao propor sistemas diagnósticos baseados em dados reais, processados de maneira automática e que podem ser atualizados constantemente.

Nesse contexto, a IA tem sido proposta como uma ferramenta promissora para aprimorar a acurácia, sensibilidade e precocidade dos diagnósticos médicos. Contudo, seu uso em contextos clínicos ainda demanda cautela: questões rela-

cionadas à explicabilidade dos modelos, à reprodutibilidade dos resultados e à confiabilidade frente à variabilidade dos dados reais precisam ser amplamente debatidas [3]; [4]. A aplicação indiscriminada da IA sem validação rigorosa pode introduzir vieses, gerar falsas promessas diagnósticas e impactar negativamente a tomada de decisão médica.

A Engenharia Biomédica, especialmente quando aliada à IA, pode contribuir de forma significativa para superar limitações dos métodos tradicionais. Porém, a simples substituição de métodos tradicionais por soluções automatizadas não garante melhorias significativas. Para que a IA realmente supere as limitações atuais, é essencial que os sistemas sejam transparentes, validados em ambientes clínicos reais e desenvolvidos com foco na segurança e aplicabilidade [5].

Este trabalho propõe uma abordagem inovadora que integra modelos de linguagem em pequena escala (SLMs), sistemas multiagentes especializados (SMA) e técnicas de Recuperação Aumentada por Geração Multimodal (Multimodal RAG) [6]; [7]. A proposta contempla agentes autônomos capazes de operar colaborativamente sobre dados textuais, imagéticos e sensoriais, com o objetivo de extrair sinais relevantes à caracterização da DP. Embora essas tecnologias estejam em estágio emergente, seu potencial reside na capacidade de combinar fontes heterogêneas de informação sem a necessidade de re-treinamento extensivo, o que pode aprimorar a precisão e a interpretabilidade dos sistemas de apoio ao diagnóstico [8]; [9]; [10].

Assim, este estudo busca não apenas propor um sistema computacional adaptável à prática clínica, mas também refletir criticamente sobre os benefícios e limitações dessas abordagens, contribuindo para a construção de soluções neurointeligentes mais robustas, seguras e aplicáveis à realidade da saúde digital.

II. METODOLOGIA

Este trabalho propõe um sistema de múltiplos agentes autônomos composto por múltiplos agentes autônomos, cada um impulsionado por diferentes modelos de linguagem de larga escala (LLMs), para previsão e prevenção da doença de Parkinson. Esses agentes operam em equipes especializadas em diferentes etapas do fluxo clínico, tomando decisões de forma colaborativa por meio de votações e discussões, o que aumenta a precisão diagnóstica.

A. Método proposto

A solução utiliza uma técnica avançada para recuperação de informação, o Multimodal RAG. Essa abordagem inovadora permite integrar dados multimodais (áudio, imagens, tabulares, textuais), sejam estruturados e não estruturados, oferecendo maior contexto e suporte à tomada de decisão preditiva. A arquitetura do nosso método é orientada por agentes com alto grau de autonomia e especialização.

B. Justificativa do Método Proposto

O diagnóstico precoce da DP ainda enfrenta limitações substanciais; variabilidade nos sintomas iniciais, subjetividade na

análise clínica e acesso restrito a exames complementares em larga escala. Métodos tradicionais são muitas vezes insensíveis a alterações sutis e precoces, o que justifica a investigação de alternativas baseadas em IA. A utilização de LLMs (Large Language Models) e RAG Multimodal busca superar essas barreiras ao permitir a integração e contextualização de múltiplas fontes de dados.

A escolha pelo Multimodal RAG fundamenta-se em sua capacidade de lidar com dados não estruturados (áudio, imagem, texto livre) e estruturados (sinais sensoriais, tabelas clínicas), promovendo recuperação contextual e geração de hipóteses mais robustas. Em comparação a abordagens unimodais, a RAG Multimodal apresenta maior flexibilidade e capacidade de generalização, apesar de desafios como sincronização temporal e compatibilização semântica entre modalidades.

C. Recuperação Aumentada por Geração (RAG)

A arquitetura RAG permite que LLMs acessem informações externas para gerar respostas mais precisas, dispensando re-treinamentos custosos [11]. É composta por dois componentes principais: um recuperador (baseado em DPR) e um gerador (encoder-decoder).

1) *Multimodal RAG*: Frente à variedade de dados médicos (imagens, sinais, áudios, textos), técnicas como o RAG Multimodal surgem para adaptar a recuperação a diferentes formatos de dados. Dentre elas, são utilizadas três abordagens principais:

- **Multimodal Embedding**: uso de modelos como CLIP e Wav2CLIP para gerar vetores a partir de imagens ou áudios [12], [13].
- **Summaries Search**: conversão de dados multimodais em texto via LLMs, gerando metadados descritivos que são embutidos por modelos textuais como BERT.
- **Hybrid Retrieval**: combinação de embeddings textuais e multimodais para melhorar a acurácia de recuperação, usando LLMs para gerar sumários e consultas híbridas.

D. Sistemas Multi-Agentes

Os *Multi-Agent Systems* (MAS) representam um avanço na inteligência artificial ao permitir que múltiplos agentes, nesse contexto baseado em pequenos modelos de linguagem (Small Language Models - SLM), colaborem para resolver tarefas complexas. Esses sistemas podem ser categorizados de acordo com as modalidades de dados processados e também a quantidade de diferentes modelos empregados em sua configuração:

- **Unimodal**: agentes atuam sobre uma única modalidade de dados, como texto.
 - *SMoA (Sparse Mixture-of-Agents)*: propõe agentes especializados colaborando para melhorar a qualidade das respostas geradas por LLMs.
 - *Multi-Agent Collaboration Mechanisms*: introduz um framework extensível para guiar pesquisas futuras sobre colaboração entre agentes baseados em LLMs.
- **Bimodal**: agentes operam com duas modalidades de dados, como texto e imagem.

- *Cooperative Sentiment Agents (Co-SA)*: utiliza agentes cooperativos para análise de sentimentos em dados multimodais, como imagens e textos.
- **Multimodal**: agentes integram múltiplas fontes de dados, como texto, imagem, áudio e vídeo.
 - *LMAgent*: apresenta agentes multimodais interagindo em ambientes simulados, com tarefas como compras online e transmissões ao vivo.
 - *SpeechAgents*: simula comunicação humana com base em sinais multimodais, como fala e texto.
 - *Beyond Text*: propõe o uso de uma plataforma sem código para facilitar a implementação de sistemas multimodais em ambientes corporativos.

1) *Responsabilidades dos agentes*: O sistema é composto por múltiplos agentes com papéis distintos, esta estrutura descentralizada permite simular a dinâmica de uma equipe clínica real, com agentes agindo de forma autônoma mas coordenada.

O sistema é composto por múltiplos agentes com papéis distintos:

Validador - O agente responsável por validar todo tipo de dado após realização das etapas de pré-processamento e processamento, desde o dado não estruturado ao dado estruturado, independente da modalidade. Ao final, gera um relatório para controle e monitoramento de dados classificados como ouro - aqueles que passaram por todas as etapas e estão prontos para serem usados nos modelos.

Recrutador - O agente responsável por recrutar a equipe de especialistas após análise prévia de disponibilidade de dados para análise e gerenciamento da doença de DP.

Especialista - Esses agentes são especialistas baseado no contexto no qual estão inseridos. Esses agentes especialistas trabalham em um grupo com outros especialistas, discutem entre si, chegam a um consenso e, por fim, tomam a decisão médica referente à consulta do paciente, complementando com o fornecimento de relatórios técnicos sobre possíveis anomalias e patologias relacionadas à consulta do paciente.

Moderador - Esse agente modera discussões, resume informações e incentiva opiniões entre os agentes especialistas em um ambiente de equipe multidisciplinar. O moderador é o principal responsável em casos de discordância entre os especialistas, mediando discussões divididas em rodadas e turnos de fala. O moderador registra o relatório da reunião da MDT, destacando todas as informações importantes abordadas durante a reunião.

Esta estrutura descentralizada permite simular a dinâmica de uma equipe clínica real, com agentes agindo de forma autônoma mas coordenada. Apesar de promissora, reconhece-se a ausência de evidências empíricas robustas de que tal divisão seja superior a arquiteturas monolíticas. Estudos comparativos futuros são necessários.

2) *Decision Making*: A tomada de decisão nesses sistemas é elemento central para o desempenho do método, uma vez que envolve a interação entre múltiplos agentes com diferentes funções e níveis de especialização. Neste trabalho, adotou-se

uma abordagem híbrida, combinando mecanismos colaborativos, votação majoritária e mediação por agente moderador. Tal escolha se fundamenta na semelhança com a prática clínica multidisciplinar, onde o consenso entre especialistas é construído por meio de discussões orientadas.

Entretanto, reconhece-se que sistemas de votação podem ser sensíveis à presença de agentes enviesados ou à propagação de erros sistêmicos. Assim, destaca-se a importância de, em versões futuras, incorporar mecanismos de avaliação de confiabilidade interagente, de modo a mitigar tais riscos e fortalecer a robustez do processo decisório.

E. Métricas de Avaliação

1) *Avaliação do Classificador*: Para quantificar o desempenho dos classificadores empregados, utilizamos as seguintes métricas:

- **Precisão**:

$$\text{Precision} = \frac{TP}{TP + FP},$$

onde TP (True Positives) são os verdadeiros positivos e FP (False Positives) são os falsos positivos.

- **Sensibilidade**:

$$\text{Recall} = \frac{TP}{TP + FN},$$

onde FN (False Negatives) são os falsos negativos.

- **F_1 -score**:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

que corresponde à média harmônica entre precisão e sensibilidade.

- **AUC-ROC**: Área sob a curva ROC (Receiver Operating Characteristic), que plota a taxa de verdadeiros positivos $TPR = TP/(TP + FN)$ em função da taxa de falsos positivos $FPR = FP/(FP + TN)$. O AUC varia de 0,5 (classificador aleatório) a 1 (classificador perfeito).

2) *Avaliação da Geração*: A validação da geração representa a segunda etapa da avaliação de sistemas RAG, sendo responsável por verificar se a resposta gerada, com base nos dados recuperados, é clara, precisa e coerente. Diversas métricas são empregadas para essa avaliação:

- **BLEU (Bilingual Evaluation Understudy)**: avalia a precisão com base na sobreposição de n -grams entre a resposta gerada e a referência, penalizando sentenças demasiadamente curtas ou longas.
- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**: foca no *recall*, mensurando a proporção de n -grams, bigramas ou subsequências mais longas (ROUGE-1, ROUGE-2, ROUGE-L) presentes na referência e corretamente recuperadas pela resposta do modelo.
- **BERTScore**: baseada em modelos de linguagem pré-treinados como o BERT, essa métrica calcula a similaridade semântica entre a resposta gerada e a referência por meio de *embeddings* de palavras. Mede precisão, *recall* e F_1 .

- **Perplexity:** avalia a fluência da resposta gerada, mensurando o quão bem o modelo prevê a próxima palavra em uma sequência. Quanto menor a perplexidade, melhor a capacidade preditiva do modelo.

3) *Avaliação da recuperação:* Essa categoria de métrica é uma etapa crucial para identificar o potencial e a robustez do sistema durante a geração de respostas com base na recuperação de documentos armazenados no banco de dados. Atualmente, duas métricas importantes são empregadas para medir esses aspectos: relevância do contexto e recuperação do contexto.

- **Semantic Textual Similarity (STS)** Um modelo é usado para gerar representações vetoriais (embeddings) do texto, a partir das quais a similaridade semântica entre os documentos recuperados e o texto é calculada. A comparação entre o texto e os documentos recuperados é feita calculando-se a similaridade de cosseno [?]. A fórmula de similaridade é a seguinte:

$$\text{Similaridade de cosseno} = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \|\vec{v}_2\|} \quad (1)$$

F. Dataset

O conjunto de dados utilizado neste estudo resulta da integração de diversas bases sobre Doença de Parkinson, provenientes de diferentes fontes.

1) Modalidades:

- **Sensorial:** aceleração nos eixos X, Y e Z e giroscópio nos eixos X, Y e Z, coletados por caneta inteligente [14], em testes de HandDP com 20 indivíduos saudáveis (8 M e 12 F) e 20 com DP moderada (12 M e 8 F), cujos escores médios de UPDRS e Hoehn & Yahr foram $28,5 \pm 5,5$ e $2,2 \pm 0,2$, respectivamente;
- **Textual:** histórico clínico e familiar extraídos da base PPMI (Parkinson's Progression Markers Initiative);
- **Imagens:** tomografia computadorizada (TC) e ressonância magnética (RM) obtidas na PPMI;
- **Audio:** gravações de voz representam uma enunciação prolongada da vogal /a/ em formato WAV (16 kHz, 16 bits), realizadas no smartphone pessoal de cada participante [15].

2) *Dados Demográficos:* Como se trata de um dado sensível, decidimos resumir as informações de maneira que preserve ao máximo a identidade dos pacientes participantes. Então, apresentamos na tabela I os principais atributos demográficos dos dois grupos.

TABLE I
CARACTERÍSTICAS DEMOGRÁFICAS DOS PARTICIPANTES

Bases	Atributo	Controle	Parkinson
Sensorial - caneta	Idade (anos)	55 – 88	55 – 88
	Sexo (M/F)	8 M / 12 F	12 M / 8 F
Áudio	Idade (anos)	18 – 79	43 – 85
	Média de Idade (anos)	47.732	67.035
	Sexo (M/F)	25 M / 16 F	19 M / 21 F

3) *Pré-processamento:* O fluxo padrão, aplicado a todas as modalidades, consiste em:

limpeza → normalização → codificação

A aplicação da normalização *Z-score* nos sinais sensoriais assume uma distribuição aproximadamente normal, o que pode não ser representativo para todos os pacientes, especialmente em fases avançadas da doença. Já o uso de *SMOTE* para balanceamento das classes visa mitigar o viés para o grupo majoritário, mas pode induzir overfitting em conjuntos com baixa diversidade original, especialmente em dados de movimentos motores.

Em cada modalidade:

- **Sensorial:** Remoção do segmento inicial que não ocorre atividade, remoção da tendência linear (componente CC) do sinal, aplicação do filtro passa-banda Butterworth para isolar o movimento da escrita, normalização *Z-score*, codificação de dados categóricos usando OneHotEncoder;
- **Textual:** remoção de stopwords, emojis, caracteres especiais e conversão de letras para minúscula;
- **Imagens:** redimensionamento para 224×224 e equalização de histograma;
- **Áudio:** corte de silêncio (nível de energia curta-janela 12,5 ms); normalização em $[-1, 1]$; filtragem de ruído por thresholds de dB.

4) *Processamento:* Assim como na etapa anterior o fluxo padrão, consiste em:

engenharia de características → aumento de dados

Em cada modalidade:

- **Sensorial:**
 - 1) segmentação de tarefas do teste handDP, em: desenhar círculos no formulário, desenhar círculos no ar, desenhar espirais no formulário, desenhar meandros no formulário, desenhar diadococinesia com a mão direita e a esquerda;
 - 2) criação de timestamp unificando hora e milissegundos;
 - 3) criação de novas características para cada canal AcX, AcY, AcZ, GyX, GyY e GyZ;
 - 4) aumento de dados na classe minoritária (grupo saudável) usando SMOTE.
- **Textual:** vetorização e criação de embedding contextual;
- **Imagens:** aumento de dados usando transformações nas imagens originais. Transformações empregadas: flip, randomcut
- **Áudio:**
 - 1) características de fonação - Jitter, Shimmer, HNR, NHR;
 - 2) índices não lineares - RPDE, DFA, PPE.

III. RESULTADOS

Os resultados obtidos neste estudo foram gerados por três categorias de agentes:

- 1) **Agentes de Machine Learning Tradicional:** aplicam algoritmos clássicos (por exemplo, SVM, Random Forest e KNN) para classificar pacientes em “saúdável” ou “com Doença de Parkinson” com base em dados sensoriais e de áudio (os arquivos .wav são convertidos em tabelas .csv após o pré-processamento).
- 2) **Agentes de Aprendizado Profundo:** utilizam redes neurais convolucionais ajustadas para processamento de imagens médicas (ressonância magnética – RM, e tomografia computadorizada – TC), explorando arquiteturas como ResNet e VGG adaptadas ao nosso conjunto de dados.
- 3) **Agentes Baseados em Modelos de Linguagem (SLMs):** empregam modelos pré-treinados (LLama 3.2, Deepseek R1) para análise de dados textuais — histórico clínico, anotações médicas e questionários — e extração de características semânticas.

Para cada categoria, avaliamos desempenho usando as métricas de acurácia, sensibilidade, precisão e F1-score, apresentadas na Tabela II. Além disso, foi computado o gráfico que mostra a curva ROC AUC que relaciona a taxa de verdadeiros positivos e falsos positivos, conforme imagem 1.

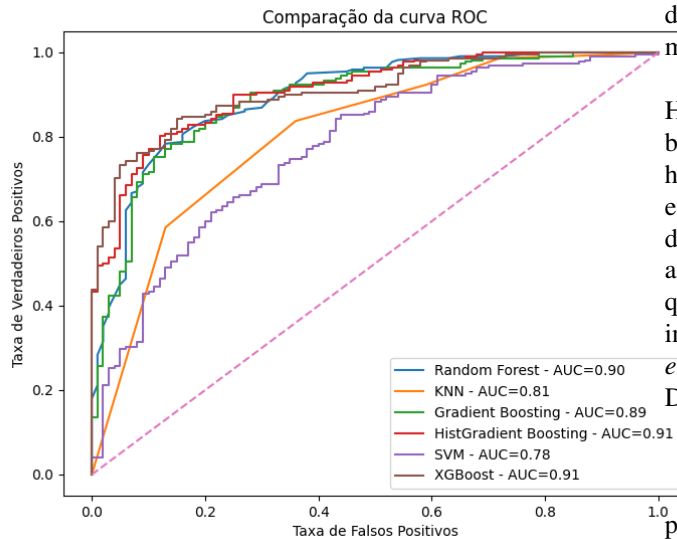


Fig. 1. Curva ROC AUC

TABLE II
RESULTADOS OBTIDOS A PARTIR DA BASE DE DADOS SENSORIAL USANDO AGENTES BASEADOS EM ML TRADICIONAL

Modelo	Acurácia	Precisão	Sensibilidade	F1-Score	ROC AUC
Random Forest	0.8136	0.8491	0.8874	0.8678	0.8916
KNN	0.7640	0.7765	0.9234	0.8436	0.8064
Gradient Boosting	0.8385	0.8541	0.9234	0.8874	0.8880
HistGradient Boosting	0.8292	0.8465	0.9189	0.8812	0.9093
SVM	0.7578	0.7535	0.9640	0.8459	0.7756
XGBoost	0.8261	0.8578	0.8964	0.8767	0.9082

A Figura 2 apresenta as matrizes de confusão, nas quais são comparados os resultados reais com as previsões feitas pelos modelos.

IV. DISCUSSÃO

A Tabela II evidencia que os modelos de *ensemble* apresentaram o melhor desempenho geral na tarefa de classificação de pacientes com DP e controles saudáveis. Em particular, o **Gradient Boosting** alcançou a acurácia mais alta (83,85 %) e o maior F₁-score (88,74 %), indicando um bom equilíbrio entre precisão e sensibilidade. Já o **HistGradient Boosting** obteve a maior área sob a curva ROC (AUC = 0,909), sugerindo forte capacidade de discriminação entre as classes.

Por outro lado, técnicas baseadas em distância (KNN) e margem (SVM) privilegiaram a *recall*, atingindo 92,34 % e 96,40 %, respectivamente, mas sacrificaram a precisão (77,65 % no KNN e 75,35 % no SVM). Esse comportamento é esperado quando se busca minimizar falsos negativos, porém pode não ser o mais adequado em cenários clínicos onde o custo de um falso positivo também é relevante.

O **Random Forest** apresentou desempenho intermediário, com acurácia de 81,37 % e AUC de 0,892, confirmando sua robustez mesmo sem afinar tanto hiperparâmetros. Já o **XGBoost** (acurácia de 82,61 % e AUC de 0,908) mostrou-se competitivo com o HistGradient Boosting, reforçando o valor de métodos baseados em gradiente para conjuntos de dados multimodais.

Em síntese, os modelos de gradiente (Gradient Boosting, HistGradient Boosting e XGBoost) combinaram alta acurácia, bom F₁-score e excelente AUC, destacando-se como as melhores opções para a classificação aqui proposta. As decisões entre eles podem se basear em fatores adicionais, como tempo de treinamento, interpretabilidade e sensibilidade a desbalanceamentos. Por fim, a consistência dos resultados sugere que a fusão de dados sensoriais, de áudio e de imagem fornece informação complementar e reforça a eficácia de abordagens *ensemble* em aplicações clínicas de apoio ao diagnóstico de DP.

V. CONCLUSÃO

Neste trabalho, apresentamos uma abordagem multimodal para o diagnóstico de DP, integrando dados sensoriais, de áudio, imagem e texto. A expectativa inicial era avaliar três categorias de agentes, baseados em métodos clássicos de aprendizado de máquina, em redes profundas para imagens médicas e em modelos de linguagem pré-treinados para dados textuais, demonstrando que os métodos de ensemble por gradiente alcançam as melhores combinações de acurácia, F₁-score e AUC. Além disso, mostramos que agentes voltados para maximizar a sensibilidade podem ser úteis em cenários clínicos de triagem. Vale lembrar que esses resultados serão bem mais assertivos ao finalizar e fundir a análise textual que os agentes baseados em SLMs ofereceriam.

Os resultados reforçam a viabilidade de explorar múltiplas fontes de informação de forma complementar, indicando que a fusão adequada de características sensoriais, acústicas, visuais

Matriz de confusão

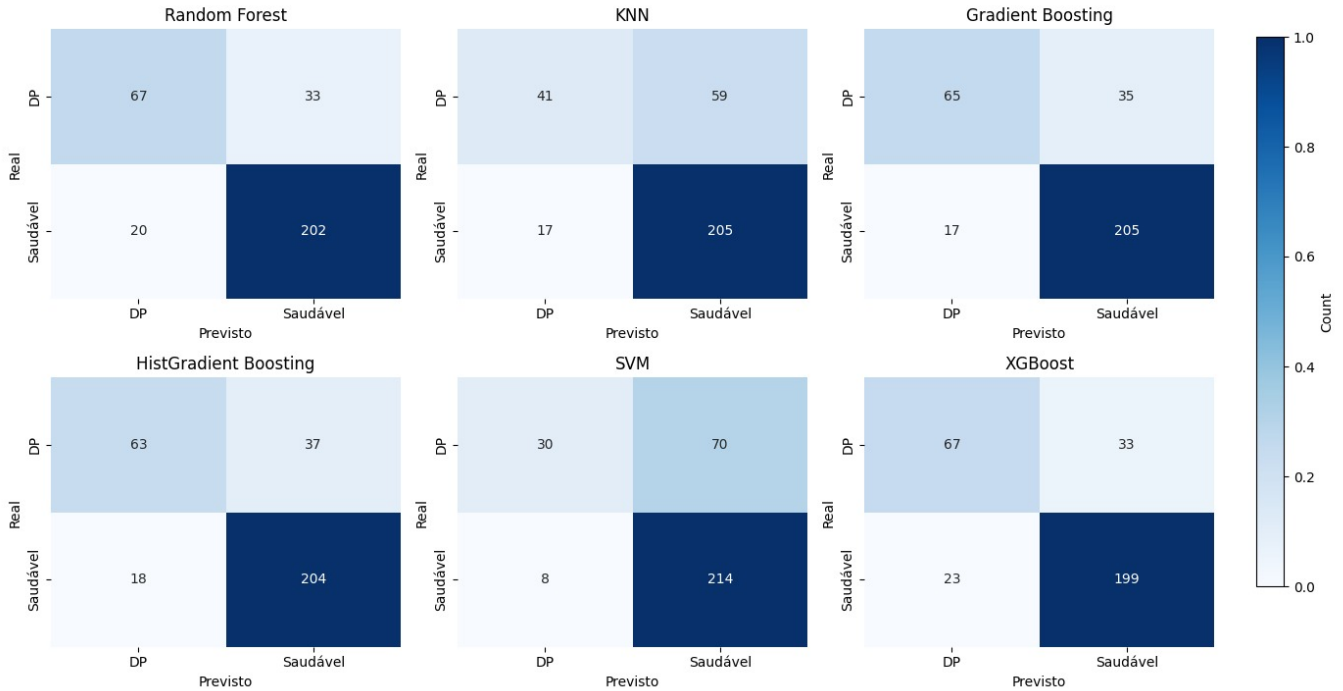


Fig. 2. Matriz de confusão

e clínicas potencializa a capacidade de discriminação entre pacientes com DP e controles saudáveis. A robustez dos modelos de gradiente e a consistência das métricas suportam seu uso em sistemas de apoio à decisão clínica.

Para etapas seguintes, pretende-se finalizar as análises de fusão multimodal, incorporando estratégias avançadas de aprendizado profundo para combinar de forma otimizada todas as modalidades. Ademais, será desenvolvido um framework integrado para diagnóstico e gerenciamento da DP, visando uma ferramenta de apoio clínico com maior assertividade e aplicabilidade em ambientes reais.

REFERENCES

- [1] V. Skaramagkas, A. Pentari, Z. Kefalopoulou, and M. Tsiknakis, "Multi-modal deep learning diagnosis of parkinson's disease-a systematic review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.
- [2] M. Shen, P. Mortezaagha, and A. Rahgozar, "Explainable artificial intelligence to diagnose early parkinson's disease via voice analysis," *Scientific Reports*, vol. 15, no. 1, p. 11687, 2025.
- [3] P. Xia, J. Wang, and ..., "Mmedagent-rl: Optimizing multi-agent collaboration for multimodal medical reasoning," 2025. *arXiv preprint*.
- [4] P. Liu, X. Liu, R. Yao, and ..., "Hm-rag: Hierarchical multi-agent multimodal retrieval augmented generation," 2025. *arXiv preprint*.
- [5] L. Gao, K. Gao, W. Sun, Y. Zhang, R. Zhang, and L. Zhang, "Effect simulation and local adaptation of multi-agent collaborative governance in marine eco-economic systems: Evidence from china," *Ecological Economics*, vol. 228, p. 108451, 2025.
- [6] M. M. Abootorabi, A. Zobeiri, M. Dehghani, M. Mohammadkhani, B. Mohammadi, O. Ghahroodi, M. S. Baghshah, and E. Asgari, "Ask in any modality: A comprehensive survey on multimodal retrieval-augmented generation," *arXiv preprint arXiv:2502.08826*, 2025.
- [7] P. Liu, X. Liu, R. Yao, J. Liu, S. Meng, D. Wang, and J. Ma, "Hm-rag: Hierarchical multi-agent multimodal retrieval augmented generation," *arXiv preprint arXiv:2504.12330*, 2025.
- [8] M. Garg, S. Raza, S. Rayana, X. Liu, and S. Sohn, "The rise of small language models in healthcare: A comprehensive survey," *arXiv preprint arXiv:2504.17119*, 2025.
- [9] Y.-J. Chen, A. Albarqawi, and C.-S. Chen, "Reinforcing clinical decision support through multi-agent systems and ethical ai governance," *arXiv preprint arXiv:2504.03699*, 2025.
- [10] P. Xia, K. Zhu, H. Li, T. Wang, W. Shi, S. Wang, L. Zhang, J. Zou, and H. Yao, "Mmed-rag: Versatile multimodal rag system for medical vision language models," *arXiv preprint arXiv:2410.13085*, 2024.
- [11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020. NeurIPS 2020.
- [12] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, Proceedings of Machine Learning Research (PMLR), pp. 8748–8763, PMLR, 2021.
- [13] H.-H. Wu, P. Seetharaman, K. Kumar, and J. P. Bello, "Wav2clip: Learning robust audio representations from clip," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Singapore), pp. 4563–4567, IEEE, 2022.
- [14] E. P. Júnior, I. L. Delmiro, N. Magaia, F. M. Maia, M. M. Hassan, V. H. C. Albuquerque, and G. Fortino, "Intelligent sensory pen for aiding in the diagnosis of parkinson's disease from dynamic handwriting analysis," *Sensors*, vol. 20, no. 20, p. 5840, 2020.
- [15] F. Prior, T. Virmani, A. Iyer, L. Larson-Prior, A. Kemp, Y. Rahmatallah, *et al.*, "Voice samples for patients with parkinson's disease and healthy controls," 2023.