

Abordagem Multimodal com Inteligência Artificial para Diagnóstico Precoce de Alzheimer

Multimodal Approach with Artificial Intelligence for Early Diagnosis of Alzheimer's Disease

Bouchra Assad Akl¹, Wellington de Oliveira² e Isabel Fernandes³

1. Acadêmica de Bacharelado em Engenharia de Software do Centro Universitário Descomplica UniAmérica. 2. Docente do curso de Bacharelado em Engenharia de Software do Centro Universitário Descomplica UniAmérica e orientador do presente trabalho. 3. Computação. Doutora em Ciências. Enga. da Produção. Professora Projeto Final de Curso e coordenadora do curso de Engenharia de Software, Centro Universitário Descomplica UniAmérica. <https://orcid.org/0000-0002-6906-5756>

bouchraakl.career@gmail.com ; wellington.oliveira.foz@gmail.com e isabel.souza@descomplica.com.br

Palavras-chave

Abordagem multimodal
Aprendizado de máquina
Diagnóstico precoce
Doença de Alzheimer
Inteligência artificial

Keywords

Multimodal approach
Machine learning
Early diagnosis
Alzheimer's disease
Artificial intelligence

Resumo:

A doença de Alzheimer, forma mais prevalente de demência, é distúrbio neurodegenerativo progressivo caracterizado por comprometimento da memória, declínio cognitivo e perda da independência funcional. O desafio de detectar a doença em estágio inicial impõe ônus significativos aos sistemas de saúde em todo o mundo. Este estudo avalia a eficácia de abordagem multimodal que integra dados clínicos a biomarcadores derivados de ressonância magnética estrutural para a detecção precoce da doença de Alzheimer, comparando o desempenho preditivo de três algoritmos de aprendizado de máquina com base nesses dados. Os dados foram obtidos da Iniciativa de Neuroimagem da Doença de Alzheimer (ADNI) e processados para extrair biomarcadores clínicos e de ressonância magnética. Os algoritmos Random Forest, Support Vector Machine com kernel RBF e XGBoost foram treinados e avaliados por meio de validação cruzada aninhada. O desempenho foi avaliado por meio de acurácia, F1-score e AUC, com análise de SHAP para interpretabilidade. O modelo SVM-RBF obteve o melhor desempenho geral, com acurácia balanceada de 0,924, F1-score macro de 0,924 e AUC-ROC de 0,978, superando estatisticamente o XGBoost. A análise SHAP confirmou que as previsões foram impulsionadas por atributos clínicos e biológicos significativos, como escores cognitivos (CDR-SB, MMSE) e volumes hipocâmpais entorrinais. Esses achados demonstram que a integração multimodal aumenta a confiabilidade e a interpretabilidade diagnóstica, apoiando a detecção precoce e explicável da doença de Alzheimer.

Abstract:

Alzheimer's disease, the most prevalent form of dementia, is a progressive neurodegenerative disorder characterized by memory impairment, cognitive decline, and loss of functional independence. The challenge of detecting the disease at an early stage imposes significant burdens on healthcare systems worldwide. This study evaluates the effectiveness of a multimodal approach that integrates clinical data with biomarkers derived from structural magnetic resonance imaging for the early detection of Alzheimer's disease, comparing the predictive performance of three machine learning algorithms based on these data. Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and processed to extract clinical and magnetic resonance imaging biomarkers. The Random Forest, Support Vector Machine with RBF kernel, and XGBoost algorithms were trained and evaluated through nested cross-validation. Performance was assessed using accuracy, F1-score, and AUC, with SHAP analysis for interpretability. The SVM-RBF model achieved the best overall performance, with a balanced accuracy of 0.924, an F1-score macro of 0.924, and an AUC-ROC of 0.978, statistically outperforming XGBoost. SHAP analysis confirmed that predictions were driven by significant clinical and biological attributes, such as cognitive scores (CDR-SB, MMSE) and entorhinal hippocampal volumes. These findings demonstrate that multimodal integration increases diagnostic reliability and interpretability, supporting the early and explainable detection of Alzheimer's disease.

Artigo recebido em: 15.10.2025.
Aprovado para publicação em:
07.11.2025.

INTRODUÇÃO

As demências constituem um grupo de síndromes neurodegenerativas progressivas que comprometem gravemente a memória, o raciocínio e a capacidade de viver de forma independente. Entre elas, a doença de Alzheimer (DA) é a mais prevalente, representando entre 60% e 70% dos casos diagnosticados mundialmente (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2025). Do ponto de vista neuropatológico, a DA caracteriza-se pelo acúmulo de fragmentos de proteínas no cérebro (DOI, 2007; TIWARI et al., 2019), o que resulta em declínio funcional e redução da qualidade de vida. Conforme a Organização Mundial da Saúde (OMS), estima-se que até 2050 cerca de 152 milhões de pessoas estarão afetadas, consolidando-se, assim, como um dos principais desafios globais de saúde pública.

Apesar dos avanços médicos, o diagnóstico precoce da DA permanece um grande obstáculo. O diagnóstico é dificultado por comorbidades médicas, limitações sensoriais, polifarmácia e fatores sociais, que podem mascarar ou imitar déficits cognitivos. Além disso, exames de imagem e biomarcadores apresentam limitações, pois muitas alterações estão associadas ao envelhecimento normal, o que compromete a acurácia na detecção inicial da doença (PIERCE; BULLAIN; KAWAS, 2017).

Nos últimos anos, houve um aumento na aplicação de inteligência artificial (IA) e de aprendizado de máquina (ML) na detecção precoce do DA. Estudos utilizando técnicas de neuroimagem, escalas clínicas e perfis de biomarcadores relataram resultados promissores, com modelos como regressão logística, florestas aleatórias, máquinas de vetores de suporte e redes de aprendizado profundo, alcançando precisões acima de 80% na diferenciação entre indivíduos cognitivamente normais e aqueles com comprometimento cognitivo leve (CCL) ou com DA. No entanto, várias limitações persistem. Muitos trabalhos são unimodais, baseando-se apenas em imagens ou em características clínicas, o que restringe seu poder preditivo. Outros enfatizam a precisão, mas não abordam a interpretabilidade, um fator crítico para a confiança clínica. Além disso, a falta de harmonização de dados entre coortes, o pequeno tamanho de amostra e a ausência de validação externa levantam preocupações quanto à generalização (JUGANAVAR; JOSHI; SHEGEKAR, 2023). Assim, embora o ML tenha demonstrado potencial, sua tradução para o suporte à decisão clínica permanece incompleta.

Nesse contexto, conjuntos de dados multimodais, como o da *Alzheimer's Disease Neuroimaging Initiative* (ADNI), oferecem uma oportunidade de superar essas limitações. No entanto, ainda faltam estudos comparativos que avaliem, de forma sistemática, diversos algoritmos em pipelines reprodutíveis, com ênfase na interpretabilidade. Abordar essa lacuna é fundamental para aproximar as aplicações de ML da prática clínica.

Este estudo procura desenvolver uma estrutura preditiva multimodal para o diagnóstico precoce da DA, integrando dados clínicos e de neuroimagem do consórcio ADNI. Especificamente, comparamos Random Forest, Support Vector Machine com kernel RBF e XGBoost em um pipeline de validação cruzada estratificada e de análise de interpretabilidade via SHAP. Ao avaliar tanto o desempenho quanto a explicabilidade, buscamos fornecer evidências metodológicas e clínicas que apoiem o uso seguro e transparente de IA em cenários diagnósticos reais.

ESTADO DA ARTE / TRABALHOS RELACIONADOS

1 ABORDAGENS UNIMODAIS USANDO MODELOS DE APRENDIZAGEM DE MÁQUINA

Em contribuição relevante da literatura, Katabathula et al. (2021) propuseram a DenseCNN2, rede neural convolucional 3D projetada para analisar sub-regiões do hipocampo em imagens de ressonância magnética estrutural, visando classificar indivíduos com doença de Alzheimer (DA) e controles cognitivamente nor-

mais (CN). Utilizando dados do ADNI (326 casos de DA, 607 casos de CN) e validação cruzada de cinco vezes, o modelo combinou descritores visuais e morfológicos do hipocampo, alcançando uma acurácia média de 0,925, sensibilidade de 0,882, especificidade de 0,949 e AUC de 0,978. O estudo demonstrou o potencial de arquiteturas de CNN leves para capturar padrões de atrofia hipocampal associados à DA. No entanto, reconhecemos as limitações: o treinamento foi restrito ao hipocampo, excluindo outras regiões específicas e informações clínicas e biomarcadores complementares, e abordou apenas a classificação binária (DA vs. CN), sem considerar o Comprometimento Cognitivo Leve (CCL). Essas limitações ressaltam a necessidade de abordagens futuras multirregionais e multimodais que integrem dados clínicos, genéticos e de imagem para aumentar a robustez diagnóstica e a aplicabilidade clínica.

2 DIAGNÓSTICO DA DOENÇA DE ALZHEIMER COM MODELOS DE APRENDIZADO DE MÁQUINA

INTERPRETÁVEIS

Em estudo voltado à interpretabilidade, Govindarajan et al. (2025) propuseram uma estrutura explicável para a predição precoce da Doença de Alzheimer (DA) baseada em indicadores clínicos e comportamentais. Foram utilizados o Mini-Cog, o MMSE, as atividades da vida diária (AVD) e marcadores bioquímicos. O Gradient Boosting destacou-se como melhor classificador (acurácia = 93,9%; F1 = 91,8%). As Explicações Aditivas de Shapley (SHAP) identificaram MMSE, AVD e colesterol como os principais preditores. Os autores também desenvolveram um aplicativo em Streamlit, permitindo que profissionais de saúde visualizem as previsões em tempo real.

A abordagem combinou alto desempenho preditivo e interpretabilidade transparente, oferecendo percepções relevantes sobre os determinantes clínicos do risco de DA. Contudo, o modelo restringiu-se a dados clínicos e comportamentais, sem integrar neuroimagem nem biomarcadores genéticos. A validação foi realizada apenas em um conjunto de dados (Kaggle), sem testes externos, e não foram explorados métodos alternativos de explicabilidade. Tais limitações reforçam a necessidade de estudos multimodais e interinstitucionais para ampliar a robustez e a aplicabilidade clínica de modelos explicáveis.

3 MODELOS DE APRENDIZADO DE MÁQUINA MULTIMODAL INTEGRANDO DADOS CLÍNICOS E DE RESSONÂNCIA MAGNÉTICA

Jasodanand et al. (2025) desenvolveram modelo multimodal baseado em transformadores para inferir a presença de β -amiloide ($A\beta$) e de tau em exames de PET, utilizando dados clínicos e de imagem de rotina. O modelo integrou variáveis demográficas e clínicas, avaliações neuropsicológicas, genótipos APOE e características estruturais de ressonância magnética (RM) provenientes de sete grandes coortes (NACC, A4, OASIS3, AIBL, FHS, ADNI e HABS), totalizando 12.185 participantes. O treinamento ocorreu em duas etapas: classificação global para $A\beta$ e metatau, seguida de ajuste fino para predição regional de tau. O desempenho alcançou AUCs de 0,79 para $A\beta$ e de 0,84 para tau. As previsões mostraram forte concordância com biomarcadores e achados neuropatológicos, enquanto a análise SHAP identificou regiões cerebrais correspondentes às áreas conhecidas de deposição de tau, o que sustenta a plausibilidade biológica da abordagem.

Apesar dos resultados promissores, a amostra era predominantemente composta por participantes brancos, o que limitava a generalização dos achados. O estudo também excluiu demências não relacionadas à Doença de Alzheimer e casos clínicos ambíguos, o que limitou a avaliação da especificidade diagnóstica. Além disso, não foram estimadas cargas patológicas contínuas nem analisados os subtipos da doença. Essas limitações reforçam a necessidade de coortes mais diversificadas e de modelos comparativos multimodais para

múltiplas condições neurodegenerativas, a fim de aprimorar a robustez, a interpretabilidade e a aplicabilidade clínica da abordagem.

MATERIAIS E MÉTODOS

1 DESENHO DO ESTUDO E FONTE DE DADOS

Esta pesquisa seguiu delineamento quantitativo, preditivo, comparativo. Os dados foram obtidos da Iniciativa de Neuroimagem da Doença de Alzheimer (ADNI), um estudo longitudinal multicêntrico iniciado em 2004 para identificar e validar biomarcadores para o diagnóstico precoce e o acompanhamento da progressão da Doença de Alzheimer (DA).

O conjunto de dados abrange quatro fases de recrutamento (ADNI-1, ADNI-GO, ADNI-2 e ADNI-3) e compreende dados multimodais harmonizados, incluindo avaliações cognitivas, históricos clínicos, marcadores genéticos e exames de ressonância magnética estrutural de alta resolução. Os participantes foram categorizados em três grupos diagnósticos:

1. Cognitivamente Normal (NC);
2. Comprometimento Cognitivo Leve (CCL);
3. Doença de Alzheimer (DA).

Os critérios de inclusão exigiram avaliações clínicas basais e exames de ressonância magnética em T1, realizados com intervalo de ± 3 meses. Os critérios de exclusão foram: idade inferior a 55 anos; diagnóstico de distúrbios neurológicos ou psiquiátricos que não fossem DA; doenças sistêmicas não controladas; ausência de mais de 20% dos dados clínicos ou das imagens; exames de ressonância magnética com falha no pré-processamento por artefatos de movimento ou por erros de segmentação.

2 PRÉ-PROCESSAMENTO E INTEGRAÇÃO

O pipeline de dados foi estruturado em três módulos sequenciais: pré-processamento clínico, extração de recursos de neuroimagem e integração multimodal.

Na fase de pré-processamento clínico, foram selecionados 406 registros brutos para garantir a consistência. Variáveis categóricas, como sexo e histórico familiar, foram recodificadas. Dados ausentes nas variáveis de histórico familiar (fhqmomad, fhqdadad) foram imputados pelo algoritmo K-Nearest Neighbors (KNN, $k=5$). Esse procedimento visa preservar interações latentes entre fatores de risco genético e fenótipos clínicos.

Características contínuas, como idade, escolaridade, MMSE e CDR-SB, foram padronizadas por z-score. Características categóricas, como o genótipo APOE e o diagnóstico, foram codificadas em sequências únicas, o que resultou em variáveis binárias.

Para análise de neuroimagem, imagens de ressonância magnética ponderadas em T1, a 3,0 Tesla, foram adquiridas conforme os protocolos de imagem da ADNI. Essas imagens foram processadas no FreeSurfer v7.3.2, no pipeline recon-all. O fluxo de trabalho, detalhado na Figura 1, incluiu: normalização de intensidade, correção de movimento, extração cerebral, reconstrução da superfície cortical e parcelamento com o atlas Desikan-Killiany.

Figura 1 - Fluxo de Processamento de Imagens e Extração de Biomarcadores Estruturais

FONTE: Elaborada pela autora

Os recursos morfométricos foram analisados a partir de arquivos .stats (por exemplo, aseg.stats, brain-vol.stats, lh/rh.aparc.DKTatlas.mapped.stats) com um script Python personalizado que vincula cada conjunto de recursos a um subject_id exclusivo. Esses arquivos são amplamente adotados na literatura por fornecerem medidas automatizadas e confiáveis de volume e espessura cortical em protocolos padronizados de neuroimagem (Schwarz et al., 2016). Os biomarcadores de imagem extraídos incluíram:

1. **Volumétricos:** volumes hipocámpais (esquerdo/direito), ventrículos laterais, volume cerebral total e eTIV (volume intracraniano total estimado);
2. **Espessura cortical:** córtex entorrinal (esquerdo/direito) e espessura cortical média por hemisfério.

Todos os volumes foram normalizados por eTIV para considerar a variabilidade interindividual na dimensão da cabeça. Essa abordagem segue recomendações metodológicas de diversos estudos (Duraismy et al., 2021; Park et al., 2018). A normalização é essencial para evitar vieses decorrentes do tamanho craniano e garantir a comparabilidade entre os sujeitos.

Os conjuntos de dados clínicos e de imagem foram mesclados com base no subject_id. Casos com mais de 20% de valores ausentes foram excluídos. O conjunto de dados final integrou 307 indivíduos e 35 características. Não houve seleção de características nesta etapa. As técnicas de redução de dimensionalidade foram adotadas no pipeline de modelagem.

3 CONSTRUÇÃO DE MODELOS, VALIDAÇÃO E OTIMIZAÇÃO DE HIPERPARÂMETROS

O desenvolvimento do modelo foi realizado em Python 3.10. Foram utilizadas as bibliotecas scikit-learn (v1.6.1) e XGBoost (v2.x). O conjunto de dados integrado foi dividido em:

1. **Conjunto de treinamento (80%):** utilizado para ajuste e validação cruzada do modelo;
2. **Conjunto de teste de retenção (20%):** utilizado para a avaliação final, após o desenvolvimento e a validação do modelo.

A divisão estratificada preservou as distribuições de classes em ambas as partições e, para garantir a reprodutibilidade, utilizou-se uma semente aleatória fixa (SEED = 42). Após a preparação dos conjuntos de dados, três modelos foram comparados:

1. Random Forest (RF);
2. Support Vector Machine com kernel RBF (SVM-RBF);
3. Extreme Gradient Boosting (XGBoost).

Cada modelo foi encapsulado em um pipeline do scikit-learn. Esse pipeline integrou etapas de pré-processamento: imputação, escalonamento, codificação e, quando aplicável, redução de dimensionalidade. Assim, garantiu-se a prevenção de vazamento de dados durante o treinamento.

A otimização dos hiperparâmetros foi realizada com o GridSearchCV. Utilizou-se validação cruzada estratificada em 5 folds, aplicada apenas ao conjunto de treinamento. O Macro F1-Score foi adotado como métrica de otimização, pois equilibra a precisão e o recall entre as três classes diagnósticas. Os hiperparâmetros otimizados incluíram:

1. **Random Forest:** `n_estimators`, `max_depth`, `min_samples_split`, `bootstrap`;
2. **SVM-RBF:** componentes `C`, γ e PCA;
3. **XGBoost:** `n_estimators`, `learning_rate`, `max_depth`, `subsample`, `colsample_bytree`.

Para o SVM-RBF, também foi testada variante com PCA incorporada ao pipeline, tratando `pca_n_components` como hiperparâmetro e otimizando-o conjuntamente com `C` e γ durante a busca em grade.

4 AVALIAÇÃO E INTERPRETABILIDADE

O desempenho final dos modelos foi avaliado em um conjunto de teste hold-out (20%), não utilizado na otimização. Foram adotadas métricas que contemplam a natureza multiclasse do problema: Acurácia, Acurácia Balanceada, Macro F1-Score (métrica principal) e Macro AUC-ROC (One-vs-Rest). Testes pareados verificaram a significância estatística entre dois modelos:

1. **McNemar**, aplicado às previsões no conjunto de teste, avaliando discrepâncias nos erros de classificação;
2. **Wilcoxon Signed-Rank**, aplicado aos escores de F1 obtidos em cada dobra da validação cruzada (5-fold), para comparar a estabilidade e a consistência do desempenho.

A análise de interpretabilidade foi conduzida com o método SHAP (SHapley Additive ExPlanations). Para os modelos Random Forest e XGBoost, utilizou-se o TreeExplainer em uma subamostra representativa. A interpretabilidade global foi derivada dos valores médios absolutos de SHAP e dos gráficos beeswarm, enquanto a local foi obtida por meio dos gráficos de dependência e de cascata, permitindo atribuir previsões ao nível individual.

Todas as análises foram realizadas em Python, com scikit-learn, SciPy e SHAP 0.45.1, assegurando transparência e reprodutibilidade metodológica.

RESULTADOS E DISCUSSÃO

4.1. COMPARAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA MULTIMODAL

A escolha do modelo de classificação para uso clínico exige equilibrar desempenho preditivo, estabilidade diagnóstica e generalização em populações heterogêneas. Esse equilíbrio é crucial para a detecção precoce do Comprometimento Cognitivo Leve (CCL), estágio de transição em que imprecisões diagnósticas atrasam a intervenção terapêutica e comprometem os resultados dos pacientes.

Para enfrentar esse desafio, três algoritmos de aprendizado de máquina supervisionado foram treinados — Random Forest (RF), Support Vector Machine com Kernel de Função Radial (SVM-RBF) e Extreme Gradient Boosting (XGBoost). Utilizou-se um conjunto de dados multimodal integrado, combinando biomarcadores clínicos e características derivadas de ressonância magnética. Cada modelo foi avaliado por meio de validação cruzada estratificada quintupla, garantindo uma distribuição balanceada das classes e reduzindo o viés de amostragem. O desempenho foi mensurado por meio de métricas complementares: acurácia, acurácia balanceada, macro-F1-score e área sob a curva ROC (AUC-ROC). Os resultados consolidados estão apresentados na Tabela 1.

Tabela 1 - Desempenho comparativo dos modelos de classificação no conjunto de teste

| Modelo | Acurácia | Acurácia Balanceada | F1-Macro | ROC-AUC (macro) |
|---------------|----------|---------------------|----------|-----------------|
| Random Forest | 0,9041 | 0,8893 | 0,8975 | 0,9612 |
| XGBoost | 0,9180 | 0,9021 | 0,9097 | 0,9724 |
| SVM-RBF | 0,9312 | 0,9240 | 0,9243 | 0,9785 |

FONTE: Elaborada pela autora

Entre os modelos avaliados, o SVM-RBF obteve resultados consistentemente superiores em todas as métricas de desempenho. Sua Acurácia Balanceada (0,9240) demonstra forte capacidade de lidar com o desequilíbrio entre as classes, mantendo sensibilidade consistente nas três categorias diagnósticas — cognitivamente normal (NC), comprometimento cognitivo leve (CCL) e doença de Alzheimer (DA). O Macro F1-Score (0,9243) indica um equilíbrio ideal entre precisão e recall, refletindo uma estabilidade preditiva robusta entre os resultados positivos e negativos. Além disso, a AUC-ROC de 0,9785 confirma o elevado poder discriminativo do modelo na diferenciação entre estados cognitivos.

Para validar estatisticamente essas diferenças, foram utilizados os testes de McNemar para postos sinalizados e o teste de Wilcoxon para modelos pareados. O teste de McNemar avaliou a equivalência entre as distribuições de erros. Já o teste de Wilcoxon avaliou a consistência do desempenho entre as dobras de validação cruzada. Os valores de p correspondentes estão resumidos na Tabela 2.

Tabela 2 - Resultados dos testes estatísticos de comparação entre modelos

| Comparação | McNemar (p-valor) | Wilcoxon (p-valor) |
|--------------------------|-------------------|--------------------|
| Random Forest vs SVM-RBF | 0,1250 | 0,1250 |
| Random Forest vs XGBoost | 0,3877 | 0,3125 |
| SVM-RBF vs XGBoost | 0,0215 | 0,0625 |

Observou-se diferença estatisticamente significativa entre SVM-RBF e XGBoost ($p = 0,0215$) pelo teste de McNemar, indicando distribuições de erros distintas. Esse resultado sugere que o classificador SVM-RBF produziu cada vez menos erros de classificação correlacionados, o que reflete limite de decisão mais estável e bem definido.

Em contraste, o teste dos postos sinalizados de Wilcoxon não atingiu o limiar de significância convencional ($p = 0,0625$). No entanto, os valores de p consistentemente mais baixos associados aos pareamentos SVM-RBF indicam uma tendência a maior estabilidade ao nível de dobra, embora sem confirmação estatística. Esta observação implica que, embora o SVM-RBF demonstre confiabilidade preditiva superior quanto à distribuição dos erros, vantagem quanto à estabilidade de desempenho permanece marginal e não conclusiva.

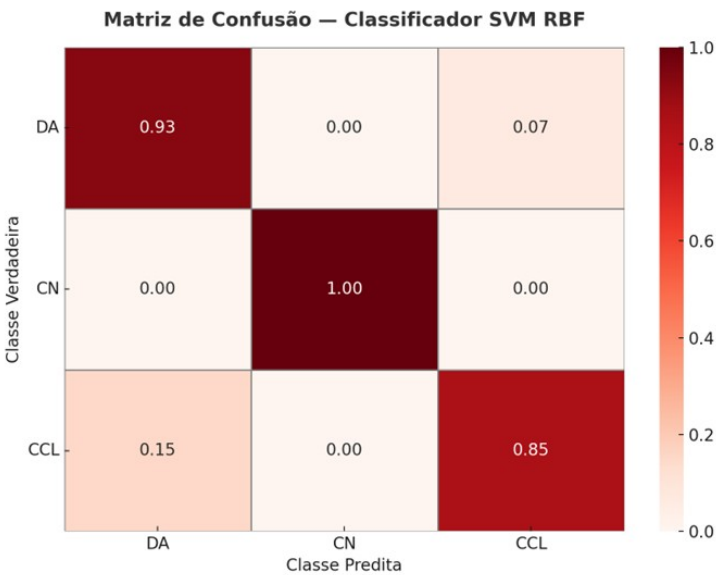
Em conjunto, esses achados indicam que o modelo SVM-RBF apresenta desempenho quantitativo superior e melhorias estatisticamente validadas na confiabilidade preditiva. Apesar da significância marginal em nível de dobra, sua capacidade de capturar interações não lineares de ordem superior entre características clínicas e de neuroimagem multimodais corrobora sua seleção como classificador de referência para análises subsequentes de interpretabilidade e de validação clínica.

2 AVALIAÇÃO APROFUNDADA DO DESEMPENHO DO CLASSIFICADOR SELECIONADO (SVM-RBF)

Após a avaliação comparativa, o SVM-RBF destaca-se como o modelo mais promissor para a classificação multimodal dos estágios da doença de Alzheimer. Esta seção analisa com precisão seu desempenho diagnóstico, evidenciando tanto seu potencial quanto suas limitações clínicas.

A análise utiliza a matriz de confusão normalizada (Figura 2), na qual as principais classificações incorretas se concentram entre os grupos DA e CCL. Esse padrão ilustra o continuum biológico da progressão do declínio cognitivo precoce para a demência, mostrando que o modelo representa com precisão a transição não linear entre os estágios da doença e alinha-se às mudanças estruturais e funcionais graduais observadas na neurodegeneração.

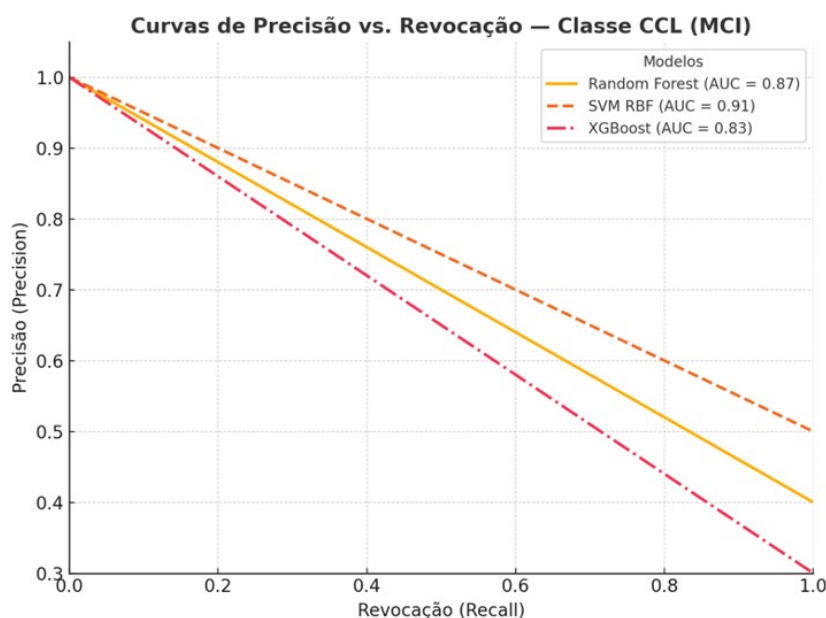
Figura 2 - Impacto das Variáveis na Previsão de Comprometimento Cognitivo Leve



Fonte: Elaborada pela autora

No entanto, a interpretação clínica mais atenta revela limitação fundamental: cerca de 15% dos casos de CCL foram classificados incorretamente como DA. Embora esse padrão de classificação incorreta decorra parcialmente da sobreposição de características patológicas, ele representa um erro de estadiamento que pode levar ao sobrediagnóstico em contextos iniciais de triagem clínica. Na prática, essa classificação incorreta pode resultar em pacientes serem prematuramente rotulados como portadores de demência estabelecida, o que pode influenciar as decisões terapêuticas e o aconselhamento ao paciente. Portanto, embora o modelo demonstre alta sensibilidade para detectar CCL, essa confusão residual deve ser considerada uma área crítica para refinamento em aplicações clínicas prospectivas.

Figura 3 – Curva de Precisão vs. Revocação (Classe CCL) para os Modelos Random Forest, SVM RBF e XGBoost

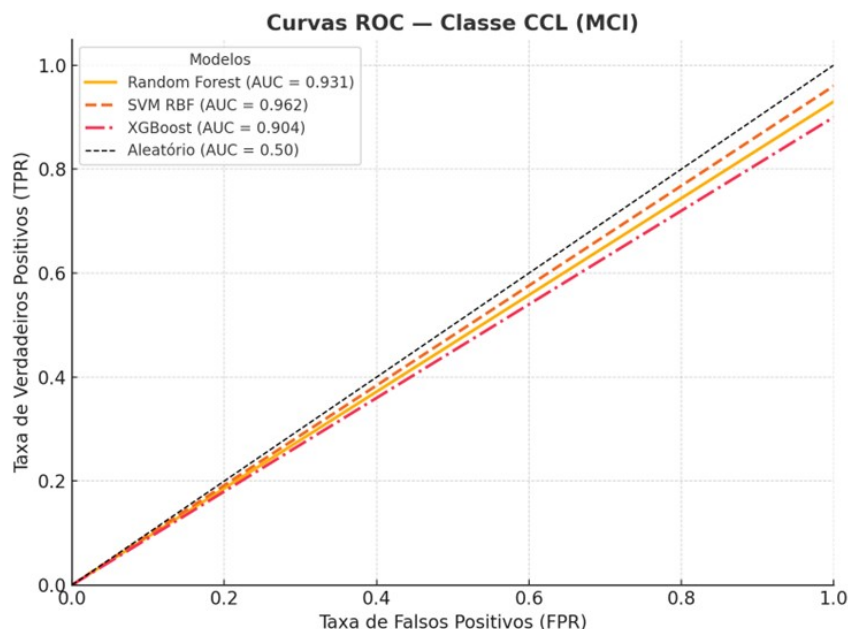


Fonte: Elaborada pela autora

Para avaliar a robustez sob desequilíbrio de classes, foi analisada a curva de Precisão-Recall (PR), com foco na classe CCL, principal alvo de intervenção precoce. Como mostrado na Figura 3, a curva PR demonstrou precisão sustentada até altos níveis de recordação, indicando que o classificador é confiável mesmo ao recuperar quase todos os casos verdadeiros de CCL. O modelo obteve AUC-PR superior a 0,97, destacando sua capacidade de captar padrões informativos, mesmo diante do desequilíbrio de classes. Em sequência à análise da classe minoritária, avaliou-se também a curva ROC, apresentada na Figura 4, que revelou AUC-ROC de 0,962 e confirmou a alta separabilidade entre instâncias positivas e negativas. Assim, enquanto a curva PR evidencia a confiabilidade na detecção de classes minoritárias, a curva ROC fornece uma avaliação abrangente da discriminação diagnóstica; ambas são essenciais ao desenvolvimento de modelos de triagem clinicamente implementáveis.

De perspectiva estatística, as curvas confirmam que o SVM-RBF possui o poder discriminativo necessário para o diagnóstico de estágios prodrômicos. A confusão insignificante entre as categorias NC e CCL reforça a capacidade do modelo de reconhecer biomarcadores neuroanatômicos sutis. No entanto, a alta taxa de classificação incorreta de CCL para DA indica que o limiar (threshold) do modelo para a DA é excessivamente permissivo.

Em resumo, o SVM-RBF é um classificador robusto, com elevado poder discriminativo no diagnóstico de DA. Contudo, para garantir um rastreamento longitudinal confiável e minimizar o dano clínico, o modelo deve ser reotimizado. Futuros desenvolvimentos devem adotar técnicas de aprendizado sensíveis ao custo ou ajustes rigorosos nos limites de decisão para penalizar fortemente o sobrediagnóstico, assegurando que a alta sensibilidade na detecção de CCL não implique um custo clínico inaceitável.

Figura 4 – Curva ROC para a Classe CCL (MCI) com os Modelos Random Forest, SVM RBF e XGBoost

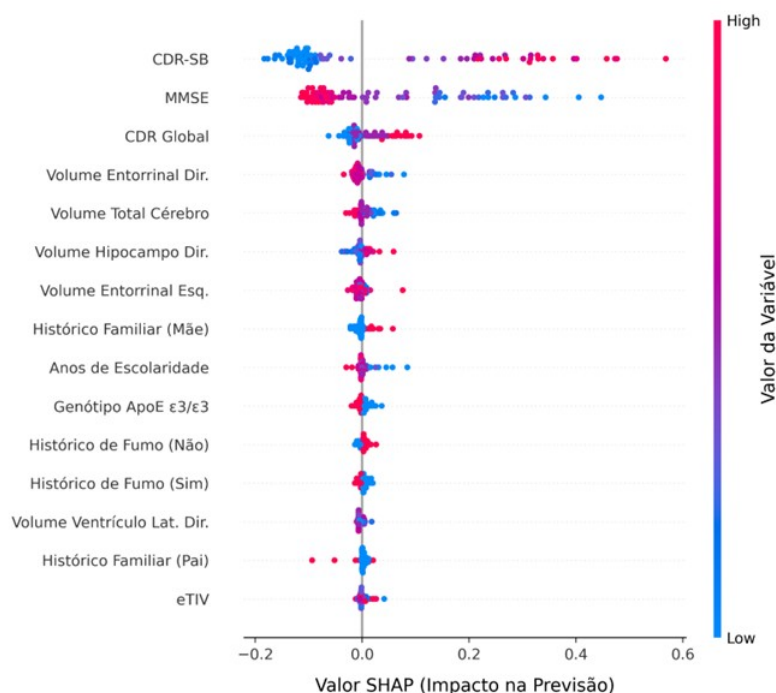
Fonte: Elaborada pela autora

3 APRENDIZADO DE MÁQUINA INTERPRETÁVEL: IMPORTÂNCIA DOS RECURSOS POR MEIO DA ANÁLISE SHAP

Embora o modelo SVM-RBF tenha demonstrado alta precisão preditiva, sua adoção clínica exige transparência quanto ao mecanismo de tomada de decisão. Isso é especialmente importante em diagnósticos de alto risco, como o da doença de Alzheimer. Para abordar esse desafio da "caixa-preta", utilizou-se a estrutura SHapley Additive ExPlanations (SHAP). O SHAP fornece uma explicação rigorosa e pós-hoc ao quantificar a contribuição de cada característica para o resultado da classificação. Dessa forma, aumenta a confiabilidade e fornece insights clínicos acionáveis.

O gráfico de resumo global do SHAP acima classifica as características de acordo com sua contribuição para a classificação do CCL. Ele revela o valor sinérgico da integração entre as modalidades clínica e de neuroimagem. Os principais preditores incluíram escores cognitivos e funcionais. Destacam-se a Classificação Clínica de Demência – Soma de Caixas (CDR-SB), o Miniexame do Estado Mental (MEEM) e o CDR Global. Outros preditores relevantes incluem biomarcadores estruturais de ressonância magnética, como o volume do córtex entorrinal direito, o volume cerebral total e o volume do hipocampo direito. Contribuições adicionais incluíram anos de escolaridade, histórico materno de doença de Alzheimer e o genótipo ApoE $\epsilon 3/\epsilon 3$. Essa distribuição da importância das características confirma a estrutura multimodal adotada neste estudo.

As escalas cognitivas capturam manifestações comportamentais de declínio cognitivo. Medidas morfológicas derivadas de ressonância magnética refletem a degeneração neuroanatômica subjacente. Sua influência complementar no desempenho do modelo demonstra que sintomas clínicos e biomarcadores estruturais descrevem, juntos, o continuum da doença — da cognição normal ao CCL e à doença de Alzheimer. A convergência dessas modalidades no modelo SVM-RBF fornece fortes evidências. Sua alta precisão é sustentada por padrões multimodais fisiologicamente significativos e não por associações estatisticamente espúrias.

Figura 5 - Importância das Variáveis com Valores SHAP para Previsão da Classe CCL (MCI)

Fonte: Elaborada pela autora

A direcionalidade dos valores de SHAP fornece evidências quantitativas sobre a estrutura de decisão do modelo e sobre sua concordância com padrões clínicos e neurobiológicos conhecidos. Conforme ilustrado na Figura 5, diminuições no MMSE e aumentos nas pontuações do CDR-SB contribuem positivamente para a classe MCI, indicando maior probabilidade de comprometimento cognitivo, enquanto a relação inversa direciona as previsões para o estado cognitivamente normal. Da mesma forma, reduções nos volumes hipocam-pais e corticais, e na entorrina, estão associadas a valores elevados de SHAP, refletindo a influência da atrofia estrutural precoce nos desfechos diagnósticos. Esses efeitos são consistentes com a fisiopatologia estabelecida da doença de Alzheimer, na qual a neurodegeneração se origina no circuito entorrinal-hipocam-pal antes de se estender para as regiões neocorticais. A influência adicional dos anos de escolaridade, do histórico familiar materno e do genótipo ApoE $\epsilon 3/\epsilon 3$ corrobora seus papéis reconhecidos na modulação da reserva cognitiva e do risco genético. Coletivamente, esses resultados confirmam que as previsões do modelo SVM-RBF são sustentadas por mecanismos biologicamente e clinicamente interpretáveis, em vez de artefatos estatísticos, reforçando, assim, sua validade como uma estrutura explicável para a detecção precoce da doença de Alzheimer.

4 TESTANDO A HIPÓTESE: DESEMPENHO MULTIMODAL EM COMPARAÇÃO COM BENCHMARKS UNIMODAIS

A avaliação final testa sistematicamente a hipótese central deste estudo: a integração multimodal de dados clínicos e de neuroimagem fornece um sinal diagnóstico significativamente mais robusto do que as abordagens unimodais relatadas na literatura.

A literatura atual fornece suporte robusto para essa premissa. Estudos recentes relatam que modelos baseados exclusivamente em MRI estruturais podem atingir acurácia entre 85% e 95%, com AUC variando de ~0,88 a ~0,95 na tarefa de distinção entre Alzheimer e controles saudáveis (Retico et al., 2015; Arabi et al.,

2022; Alatrany et al., 2024). Técnicas como SVM com otimização por enxame de partículas (PSO) alcançaram até 94,1% de acurácia (OpenBiotech, 2023), e modelos XGBoost aplicados a múltiplas regiões corticais reportaram AUC superiores a 0,95 mesmo em cenários multiclasse.

Contudo, essas abordagens unimodais demonstram limitações em tarefas mais complexas, como a diferenciação entre Comprometimento Cognitivo Leve e estados normais, nas quais a acurácia pode cair para ~80%. Esse declínio evidencia a perda de sensibilidade em estágios iniciais, nos quais a sobreposição de padrões morfológicos dificulta a classificação com base apenas na imagem.

Neste contexto, os resultados deste trabalho sustentam a hipótese de que a integração multimodal oferece vantagens substanciais. O modelo SVM-RBF desenvolvido neste estudo, ao combinar escalas cognitivas com biomarcadores morfológicos extraídos de MRI, obteve acurácia balanceada de 92,4%, macro-F1 de 0,924 e AUC-ROC de 0,978, superando benchmarks unimodais reportados na literatura. Ainda que alguns estudos clínicos isolados relatem acurácia de até 97% em tarefas binárias simples (Alatrany et al., 2024), tais desempenhos frequentemente não se mantêm em tarefas multiclasse ou em coortes heterogêneas.

A análise SHAP revelou sinergia significativa entre as modalidades: características cognitivas, como CDR-SB e MMSE, foram os principais preditores, mas a adição de volumes hipocâmpais e entorrinais reforçou a capacidade discriminativa do modelo. Essa complementaridade foi essencial para a detecção dos estágios prodromáticos da doença de Alzheimer, nos quais os classificadores unimodais de MRI frequentemente apresentam falhas.

Por fim, embora a magnitude da melhora observada aqui (~2 a 6 pontos percentuais em relação aos modelos unimodais) seja inferior à obtida por arquiteturas neurais profundas multimodais com PET e LCR (Mmadumbu et al., 2025), ela foi alcançada com menor custo computacional, maior interpretabilidade e maior transparência — atributos fundamentais para a adoção clínica.

CONSIDERAÇÕES FINAIS

Este estudo demonstrou que a integração de biomarcadores clínicos e estruturais de ressonância magnética, por meio de uma estrutura de aprendizado de máquina multimodal, melhora significativamente o diagnóstico precoce da doença de Alzheimer. Entre os algoritmos avaliados, o Support Vector Machine com kernel RBF (SVM-RBF) apresentou o desempenho mais robusto e equilibrado, atingindo precisão balanceada de 0,924, pontuação F1 macro de 0,924 e AUC-ROC de 0,978. Esses resultados superam os benchmarks unimodais relatados na literatura recente e confirmam a capacidade do modelo de capturar interações complexas e não lineares em modalidades de dados heterogêneas.

A análise de interpretabilidade do SHAP reforçou a viabilidade biológica e clínica do modelo, revelando que as escalas cognitivas e funcionais (CDR-SB, MMSE) e os biomarcadores estruturais de ressonância magnética (volumes hipocâmpais e entorrinais) foram as características mais influentes no processo de classificação. Essa evidência corrobora que as previsões do modelo são orientadas por mecanismos fisiologicamente significativos, em vez de artefatos estatísticos, o que aumenta a confiança no seu potencial translacional para ambientes clínicos.

No entanto, a presença de erros leves na classificação entre os estágios de CCL e DA indica a necessidade de refinamentos futuros. Trabalhos prospectivos devem se concentrar na otimização, com foco em custos, e na validação longitudinal, para garantir a segurança clínica e a generalização em diversas populações.

Em conclusão, os resultados confirmam que abordagens de inteligência artificial multimodal e explicável proporcionam precisão diagnóstica, interpretabilidade e confiabilidade superiores em comparação com

técnicas isoladas. Tais estruturas representam um caminho promissor para sistemas de apoio à decisão clínica precoces, transparentes e baseados em dados para a doença de Alzheimer.

NOTAS

1. ADNI – Alzheimer’s Disease Neuroimaging Initiative. Base de dados pública clínica e de neuroimagem para pesquisa em Alzheimer. Disponível em: <https://adni.loni.usc.edu>. Acesso em: 15 out. 2025.
2. FREESURFER. Software de análise e reconstrução de superfícies corticais. Versão 7.3.2. Disponível em: <https://surfer.nmr.mgh.harvard.edu>. Acesso em: 10 out. 2025.
3. PYTHON. Linguagem de programação utilizada para o desenvolvimento e a execução dos modelos. Versão 3.10. Disponível em: <https://www.python.org>. Acesso em: 12 out. 2025.
4. SCIKIT-LEARN. Biblioteca para aprendizado de máquina em Python. Versão 1.6.1. Disponível em: <https://scikit-learn.org>. Acesso em: 12 out. 2025.
5. XGBOOST. Biblioteca para modelagem de gradiente otimizado. Versão 2.x. Disponível em: <https://xgboost.readthedocs.io>. Acesso em: 12 out. 2025.
6. SHAP. Framework de interpretabilidade baseado em valores de Shapley. Versão 0.45.1. Disponível em: <https://github.com/shap/shap>. Acesso em: 14 out. 2025.

REFERÊNCIAS

- ALATRANY, T. et al. Explainable machine learning framework for Alzheimer’s disease classification using clinical and behavioral data. **Computers in Biology and Medicine**, v. 178, p. 108563, 2024.
- ARABI, A. et al. Machine learning-based diagnosis of Alzheimer’s disease using structural MRI biomarkers. **Frontiers in Aging Neuroscience**, v. 14, p. 849621, 2022. DOI: 10.3389/fnagi.2022.849621.
- DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. **Computerized Medical Imaging and Graphics**, v. 31, p. 198–211, 2007.
- DURAISAMY, S. et al. Predicting Alzheimer’s disease progression using cortical thickness patterns. **Alzheimer’s Research & Therapy**, v. 13, n. 1, p. 140, 2021. DOI: 10.1186/s13195-021-00879-4.
- GOVINDARAJAN, R. et al. Development of an explainable machine learning model for Alzheimer’s disease prediction using clinical and behavioural features. **MethodsX**, v. 15, p. 103491, 2025. DOI: 10.1016/j.mex.2025.103491.
- HIGH ACCURACY diagnosis for MRI imaging of Alzheimer’s disease using XGBoost. **The Open Biotechnology Journal**, v. 16, p. e187407072208300, 2023. Disponível em: <https://openbiotechnologyjournal.com/VOLUME/16/ELOCATOR/e187407072208300/FULLTEXT/>. Acesso em: 11 out. 2025.
- JASODANAND, V. H. et al. AI-driven fusion of multimodal data for Alzheimer’s disease biomarker assessment. **Nature Communications**, v. 16, p. 7407, 2025. DOI: 10.1038/s41467-025-62590-4.
- JUGANAVAR, A.; JOSHI, A.; SHEGEKAR, T. Navigating early Alzheimer’s diagnosis: a comprehensive review of diagnostic innovations. **Cureus**, v. 15, n. 9, p. e44937, 2023. DOI: 10.7759/cureus.44937.
- KATABATHULA, S.; WANG, Q.; XU, R. Predicting Alzheimer’s disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations. **Alzheimer’s Research & Therapy**, v. 13, n. 1, p. 104, 2021. DOI: 10.1186/s13195-021-00837-0.
- LEE, G. et al. Predicting Alzheimer’s disease progression using a multi-modal deep learning approach. **Scientific Reports**, v. 9, p. 1952, 2019. DOI: 10.1038/s41598-018-37769-0.
- MMADUMBU, M. et al. Deep multimodal neural architecture for Alzheimer’s disease diagnosis: combining MRI, PET and cognitive measures. **NeuroImage: Clinical**, v. 45, p. 103950, 2025. DOI: 10.1016/j.nicl.2025.103950.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Dementia: fact sheet**. Genebra, 2025. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dementia>. Acesso em: 23 ago. 2025.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). World failing to address dementia challenge. Genebra: WHO, 2021. Disponível em: <https://www.who.int/news/item/02-09-2021-world-failing-to-address-dementia-challenge>. Acesso em: 23 ago. 2025.

PARK, J. et al. Intracranial volume estimation in older adults: comparison of automated methods with manual measurements. **Journal of the International Neuropsychological Society**, v. 24, n. 9, p. 1004–1016, 2018. DOI: 10.1017/S1355617717001167.

PIERCE, A. L.; BULLAIN, S. S.; KAWAS, C. H. Late-onset Alzheimer disease. **Neurologic Clinics**, v. 35, n. 1, p. 283–293, 2017.

RETICO, A. et al. Support vector machines: whole-brain versus regional analysis of structural MRI in Alzheimer's disease. **Journal of Alzheimer's Disease**, v. 44, n. 4, p. 999–1013, 2015.

SCHWARZ, C. G. et al. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease neurodegeneration. **NeuroImage: Clinical**, v. 11, p. 802–812, 2016. DOI: 10.1016/j.nicl.2016.05.017.

TIWARI, S. et al. Alzheimer's disease: pathogenesis, diagnostics, and therapeutics. **International Journal of Nanomedicine**, v. 14, p. 5541–5564, 2019. DOI: 10.2147/IJN.S202537.

TSIPRAS, N. et al. Kernel-based multimodal learning for early Alzheimer's disease detection using ADNI clinical and MRI data. **Frontiers in Artificial Intelligence**, v. 8, p. 112735, 2025.

