

# Combinando Tarefas de Escritas à Mão e *Ensemble Learning* para Identificação de Alzheimer

Reginaldo J. Silva <sup>1</sup>

Faculdade de Engenharia de Ilha Solteira/UNESP, Ilha Solteira, SP

Beatriz C. Debs <sup>2</sup>

Instituto de Ciências da Natureza/UNIFAL-MG, Alfenas, MG

Edvaldo J. R. Cardoso <sup>3</sup>

Faculdade de Medicina/UNIFAL-MG, Alfenas, MG

Angela L. Moreno <sup>4</sup>

Instituto de Ciências Exatas/UNIFAL-MG, Alfenas, MG

**Resumo.** A doença de Alzheimer é uma condição neurodegenerativa comum, afetando 8,5% da população mundial com mais de 60 anos, e a detecção precoce é crucial para a melhoria da qualidade de vida e a não progressão da doença. Entretanto, seu diagnóstico precoce depende de métodos de diagnóstico avançados, muitas vezes indisponível à maioria da população. Assim, esse estudo propôs a classificação de pacientes com Alzheimer e indivíduos saudáveis utilizando dados de escrita à mão do conjunto DARWIN, que inclui 25 tarefas categorizadas em memória/ditado, gráficas e cópia. Foram extraídos 18 atributos relacionados a tempo, pressão, velocidade e tremor, entre outros, para cada tarefa. Onze algoritmos de aprendizado de máquina, incluindo *Random Forest*, *Gradient Boosting* e *Decision Tree*, foram aplicados individualmente e combinados por *ensemble* (média das probabilidades e votação por maioria). Modelos baseados em árvores destacaram-se, com *Random Forest* alcançando acurácia de 0,8491 e AUC de 0,9095. Os métodos empregados constataram que tarefas gráficas e de cópia mostraram-se mais discriminativas, enquanto tarefas de memória/ditado foram mais desafiadoras. A combinação por *ensemble* melhorou a robustez do sistema.

**Palavras-chave.** Classificadores, *Machine Learning*, Algoritmos Baseados em Árvore, Diagnóstico Assistido por Máquina, Doença de Alzheimer, Conjunto de Dados DARWIN.

## 1 Introdução

A Doença de Alzheimer (DA), primeiramente descrita pelo médico alemão Alois Alzheimer em 1906, é uma condição neurodegenerativa caracterizada por um quadro clínico variável de progressiva perda cognitiva, especialmente da memória [3]. A doença, inicialmente, afeta o lobo temporal, responsável pela formação de novas memórias, e, em seus estágios finais, compromete todo o encéfalo. As principais características neuropatológicas incluem a presença de placas amiloides, descritas no trabalho original de Alzheimer, e atrofia cortical [9]. Atualmente, a DA é a causa mais comum de demência, considerada a doença degenerativa mais prevalente, cara e letal. Estima-se que sua prevalência triplique até 2050, representando um desafio significativo para os sistemas de saúde globais [14].

Diante desse cenário, diversas abordagens têm sido exploradas para o diagnóstico precoce, sendo que um dos maiores desafios é a diferenciação entre o declínio cognitivo (DC) normal associado

---

<sup>1</sup>reginaldo.silva@unesp.br

<sup>2</sup>beatriz.debs@sou.unifal-mg.edu.br

<sup>3</sup>edvaldo.cardoso@unifal-mg.edu.br

<sup>4</sup>angela.moreno@unifal-mg.edu.br

ao envelhecimento e os primeiros estágios da doença [5]. Enquanto o DC normal não responde às terapias medicamentosas, a DA é progressiva e irreversível, tornando o diagnóstico precoce crucial para o tratamento adequado e a implementação de intervenções que retardam a progressão da doença [6]. Nesse contexto, diversas abordagens têm sido exploradas para o diagnóstico precoce da DA, incluindo a associação com perda olfatória [13], síndrome metabólica [17] e marcadores plasmáticos [2], além de métodos avançados de imagem [16]. Entretanto, a evolução dos métodos computacionais combinados com aprendizado de máquina têm aberto novas possibilidades para a detecção precoce da doença, como a análise genética [10], a análise de neuroimagem [1] e a identificação de biomarcadores [15]. Entretanto, devido ao custo, o acesso de grande parte da população a esses métodos de diagnóstico avançados é limitado.

Nesse cenário, o uso de dados clínicos não convencionais, como a escrita à mão, emerge como uma abordagem promissora, pois a escrita é uma atividade complexa que envolve múltiplas funções cognitivas, motoras e perceptivas, sendo influenciada por áreas cerebrais afetadas precocemente pela DA, como o lobo temporal e o córtex pré-frontal. Estudos recentes têm demonstrado que alterações na escrita, como mudanças na pressão, velocidade, tamanho das letras e organização espacial, podem ser indicativos de comprometimento cognitivo [8]. Essas alterações, por sua vez, podem ser quantificadas e analisadas por meio de técnicas de processamento de sinais e aprendizado de máquina, oferecendo uma ferramenta não invasiva, de baixo custo e amplamente acessível para o diagnóstico precoce da DA. Além disso, a escrita é uma atividade cotidiana que pode ser facilmente monitorada ao longo do tempo, permitindo a identificação de mudanças sutis que podem preceder o aparecimento de sintomas clínicos mais evidentes [8]. A análise computacional da escrita à mão, portanto, não apenas complementa as abordagens diagnósticas tradicionais, mas também oferece uma abordagem para a detecção precoce e o monitoramento contínuo da doença.

Este trabalho propõe a utilização de dados de escrita à mão, processados por meio de técnicas de aprendizado de máquina, para a classificação de pacientes com DA e indivíduos saudáveis.

## 2 Metodologia

Este estudo utilizou o conjunto de dados DARWIN (*Diagnosis Alzheimer With haNdwriting*), que utiliza características de dados de escrita à mão para classificar pacientes entre aqueles com diagnóstico de DA e aqueles sem a doença. O DARWIN consiste em atributos derivados de 25 tarefas de escrita, categorizadas em gráficas, de cópia e de memória/ditado, conforme descrito no protocolo de aquisição de Cilia et al. [4]. O conjunto de dados inclui informações de 174 participantes, sendo 89 pacientes com DA e 85 indivíduos saudáveis. Para cada participante, foram extraídos 18 atributos relacionados ao tempo, pressão, velocidade, aceleração, tremor e dispersão, entre outros, para cada uma das 25 tarefas, conforme listadas na Tabela 1.

Para capturar diferentes aspectos cognitivos e motores, as tarefas foram agrupadas em três categorias principais: tarefas de memória/ditado (M), tarefas gráficas (G) e tarefas de cópia (C). Cada conjunto foi tratado individualmente, resultando em três matrizes de dados distintas. Por exemplo, o conjunto de tarefas de cópia, composto por 14 tarefas, gerou uma matriz com 174 linhas (uma para cada participante) e 252 colunas (14 tarefas  $\times$  18 atributos por tarefa). O mesmo procedimento foi aplicado aos conjuntos de memória/ditado e gráfico, resultando em matrizes com dimensões proporcionais ao número de tarefas em cada categoria.

Para a classificação, foram empregados 11 algoritmos de aprendizado de máquina, abrangendo diferentes paradigmas: métodos de *boosting* (*Ada Boosting*-AB) [7], modelos baseados em distância (*K-Nearest Neighbors*-KNN)[12], modelos Autoexpansíveis baseados em Teoria da Ressonância Adaptativa (*ART Fuzzy*-AF e *ART Euclidiana*-AE) [11], redes neurais (*Multilayer Perceptron*-MLP), modelos lineares ou de margem (*Support Vector Machine*-SVM e Regressão Logística-RL) e modelos baseados em árvores (*Random Forest*-RF, *Gradient Boosting*-GB, *Decision Tree*-DT e *Extra Trees*-ET) [12]. Cada algoritmo foi avaliado seguindo uma metodologia comum: os dados

Tabela 1: Descrição das tarefas e Categorias.

#	Descrição da Tarefa	Cat.
1	Desenho da assinatura	M
2	Unir dois pontos com uma linha horizontal, continuamente por quatro vezes	G
3	Unir dois pontos com uma linha vertical, continuamente por quatro vezes	G
4	Retraçar um círculo (6 cm de diâmetro) continuamente por quatro vezes	G
5	Retraçar um círculo (3 cm de diâmetro) continuamente por quatro vezes	G
6	Copiar as letras “l”, “m” e “p”	C
7	Copiar as letras nas linhas adjacentes	C
8	Escrever cursivamente uma sequência de quatro letras “l” minúsculas, em um único movimento	C
9	Escrever cursivamente uma sequência de quatro bigramas “le” minúsculos, em um único movimento	C
10	Copiar a palavra “foglio”	C
11	Copiar a palavra “foglio” acima de uma linha	C
12	Copiar a palavra “mamma”	C
13	Copiar a palavra “mamma” acima de uma linha	C
14	Memorizar as palavras “telefono”, “cane”, e “negozio” e reescrevê-las	M
15	Copiar a palavra “bottiglia” de trás para frente	C
16	Copiar a palavra “casa” de trás para frente	C
17	Copiar seis palavras (regulares, irregulares, não palavras) nas caixas apropriadas	C
18	Escrever o nome do objeto mostrado em uma imagem (uma cadeira)	M
19	Copiar os campos de uma ordem postal	C
20	Escrever uma frase simples sob ditado	M
21	Retraçar uma forma complexa	G
22	Copiar um número de telefone	C
23	Escrever um número de telefone sob ditado	M
24	Desenhar um relógio, com todas as horas e posicionar os ponteiros às 11:05 ( <i>Clock Drawing Test</i> )	G
25	Copiar um parágrafo	C

foram divididos em conjuntos de treinamento (70%) e teste (30%), mantendo a proporção de pacientes com DA e indivíduos saudáveis em ambos os conjuntos, garantindo a representatividade, com otimização de hiperparâmetros realizada por meio de busca em grade e validação cruzada de 5-*folds* no conjunto de treinamento. A busca em grade permitiu testar combinações de hiperparâmetros para cada modelo, como o número de árvores e a profundidade máxima para RF, ou o número de vizinhos para KNN. Após a otimização, os modelos foram avaliados no conjunto de teste.

Para avaliar o desempenho dos modelos, foram utilizadas as métricas: acurácia (ACC) que mede a proporção de classificações corretas em relação ao total de amostras; sensibilidade (Se) e especificidade (Sp), avaliam a capacidade do modelo de identificar corretamente pacientes com DA e indivíduos saudáveis, respectivamente; precisão (Pre) reflete a proporção de verdadeiros positivos entre todas as previsões positivas; o *F<sub>1</sub>-score* combina precisão e sensibilidade em uma única métrica balanceada; e AUC-ROC (*Area Under the Curve - Receiver Operating Characteristic*), abreviadamente AUC, avalia a capacidade do modelo de distinguir entre as duas classes, independentemente do limiar de classificação.

Para cada um dos 11 modelos, após a avaliação em cada um dos três conjuntos de tarefas, as saídas foram combinadas utilizando duas estratégias de *ensemble*: votação por maioria e média das probabilidades. Na votação por maioria, a classe final foi determinada pela classe mais frequente entre as previsões das diferentes configurações do mesmo modelo. Já na média das probabilidades, a probabilidade média de pertencer à classe “DA” foi calculada a partir das probabilidades geradas por cada configuração do modelo, e a classe final foi determinada com base nessa média. Essa

abordagem permitiu explorar a diversidade das parametrizações de um mesmo modelo, reduzindo o viés individual e melhorando a robustez do sistema de classificação. Os resultados foram analisados tanto para os modelos individuais quanto para as combinações por *ensemble*, permitindo uma comparação abrangente do desempenho de cada abordagem.

### 3 Resultados

A Tabela 2 apresenta os resultados individuais dos 11 modelos para cada um dos três conjuntos de tarefas: memória/ditado, gráficas e cópia. Ao se analisar o problema abordado, precisa-se focar não na acurácia, mas nas métricas sensibilidade e AUC-ROC. Para as tarefas de memória/ditado e cópia, pode-se perceber que as redes neurais autoexpansíveis AF e AE obtiveram o mesmo desempenho do algoritmo RF, com Acurácia de 66,04% e uma AUC-ROC de 0,8664, uma sensibilidade de 25% o que leva a classificar 75% dos pacientes com DA como saudáveis. Entretanto, esses modelos conseguem identificar corretamente os pacientes saudáveis devido ao alto valor de precisão e especificidade (100%). Com uma sensibilidade tão baixa esse modelo não é adequado para a identificação de pacientes com DA, pois pode levar a um atraso no tratamento. Não conseguindo capturar as diferenças entre os pacientes saudáveis e com DA, isso pode ser observado pelo seu MCC de 0,3928 e  $F_1$ -score de apenas 0,4000, o mesmo acontecendo com as tarefas gráficas para a rede autoexpansível AF e os algoritmos KNN e RF. Entretanto, a rede autoexpansível AE, ao classificar as tarefas gráficas, conseguiu uma sensibilidade de 70,83%, sendo o método mais sensível para esse tipo de tarefa e identificando 65,57% dos pacientes saudáveis, o que leva a um  $F_1$ -score mais balanceado, obtendo um *score* de 0,6667 e uma AUC de 0,7227.

Em todas as tarefas, os classificadores AB, MLP, SVM, RL, GB, DT e ET obtiveram uma acurácia de 73,88% e precisão de 85,71%, com sensibilidade de 50% e especificidade de 93,10%, com uma pontuação de  $F_1$ -score de 0,6316 e AUC de 0,7155, mostrando consistência no desempenho desses classificadores. O mesmo acontecendo para o KNN em relação à tarefa de memória/ditado.

Ao se observar os resultados da sensibilidade dos diversos modelos, nota-se que eles tiveram dificuldade em identificar pacientes com DA, possivelmente devido à complexidade das tarefas de memória/ditado que avaliam habilidades cognitivas mais afetadas pela doença.

Contudo, a situação de equilíbrio relativo se rompe quando analisamos a Tabela 3, que apresenta os resultados combinados utilizando as estratégias de *ensemble*: média das probabilidades e votação por maioria. Tais estratégias permitiram explorar a diversidade dos modelos, melhorando a robustez e a generalização do sistema de classificação. O RF combinado por média das probabilidades obteve a maior AUC (0,9095), com acurácia de 84,91%, sensibilidade de 84,62%, um MCC de 0,7077 e  $F_1$ -score de 0,8462, o que o destaca como o modelo mais eficaz. A votação por maioria também mostrou resultados consistentes, com acurácia de 81,13% e  $F_1$ -score de 0,8148 para o *Random Forest*. Entretanto, as redes neurais Autoexpansíveis AF e AE, que haviam tido um desempenho comparável ao classificador RF, em alguns casos superando-o, não conseguiram acompanhar seu desempenho, focando apenas em identificar pacientes saudáveis, com especificidade de 100% e 88,24%, com AUC de 0,8218 e 0,8118, respectivamente.

O GB também apresentou desempenho superior quando combinado por média das probabilidades, com acurácia de 0,8302 e  $F_1$ -score de 0,8163. Esses resultados reforçam a eficácia dos modelos baseados em árvores, que se beneficiaram da combinação de múltiplas previsões para melhorar a precisão e a robustez. Algoritmos mais simples, como DT, também se destacaram na combinação por *ensemble* tendo resultados consistentes para as duas técnicas. O DT alcançou uma acurácia de 83,02%, um  $F_1$ -score de 0,8235 e um MCC de 0,6652, demonstrando que mesmo modelos mais simples podem ser altamente eficazes quando combinados com outras abordagens.

Por outro lado, modelos como KNN e SVM tiveram desempenho inferior em comparação com os modelos baseados em árvores. O KNN combinado por média das probabilidades obteve uma

Tabela 2: Resultados dos onze métodos de classificação para diferentes tipos de tarefas.

		AB	KNN	AF	AE	MLP	SVM	RL	GB	RF	DT	ET
Memória/Ditado	ACC	<b>0,7358</b>	<b>0,7358</b>	0,6604	0,6604	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	0,6604	<b>0,7358</b>	<b>0,7358</b>
	Pre	0,8571	0,9167	<b>1,0000</b>	<b>1,0000</b>	0,8571	0,8571	0,8571	0,8571	<b>1,0000</b>	0,8571	0,8571
	Se	0,5000	0,4583	0,2500	0,2500	<b>0,5000</b>	<b>0,5000</b>	<b>0,5000</b>	<b>0,5000</b>	0,2500	0,5000	<b>0,5000</b>
	Sp	0,9310	0,9655	<b>1,0000</b>	<b>1,0000</b>	0,9310	0,9310	0,9310	0,9310	<b>1,0000</b>	0,9310	0,9310
	F <sub>1</sub>	<b>0,6316</b>	0,6111	0,4000	0,4000	<b>0,6316</b>	<b>0,6316</b>	<b>0,6316</b>	<b>0,6316</b>	0,4000	<b>0,6316</b>	<b>0,6316</b>
	MCC	0,4867	<b>0,5041</b>	0,3928	0,3928	0,4867	0,4867	0,4867	0,4867	0,3928	0,4867	0,4867
	AUC	0,7155	0,8563	<b>0,8664</b>	<b>0,8664</b>	0,7155	0,7155	0,7155	0,7155	<b>0,8664</b>	0,7155	0,7155
Gráfica	ACC	<b>0,7358</b>	0,7170	0,6604	0,6792	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	0,7358	0,6604	<b>0,7358</b>	<b>0,7358</b>
	Pre	0,8571	0,7143	<b>1,0000</b>	0,6296	0,8571	0,8571	0,8571	0,8571	<b>1,0000</b>	0,8571	0,8571
	Se	0,5000	0,6250	0,2500	<b>0,7083</b>	0,5000	0,5000	0,5000	0,5000	0,2500	0,5000	0,5000
	Sp	0,9310	0,7931	<b>1,0000</b>	0,6552	0,9310	0,9310	0,9310	0,9310	<b>1,0000</b>	0,9310	0,9310
	F <sub>1</sub>	0,6316	0,6667	0,4000	<b>0,6667</b>	0,6316	0,6316	0,6316	0,6316	0,4000	0,6316	0,6316
	MCC	<b>0,4867</b>	0,4255	0,3928	0,3619	<b>0,4867</b>	<b>0,4867</b>	<b>0,4867</b>	<b>0,4867</b>	0,3928	<b>0,4867</b>	<b>0,4867</b>
	AUC	0,7155	0,7759	<b>0,8664</b>	0,7227	0,7155	0,7155	0,7155	0,7155	<b>0,8664</b>	0,7155	0,7155
Cópia	ACC	<b>0,7358</b>	0,6604	0,6604	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	<b>0,7358</b>	0,6604	<b>0,7358</b>	<b>0,7358</b>
	Pre	0,8571	<b>1,0000</b>	<b>1,0000</b>	0,8571	0,8571	0,8571	0,8571	0,8571	<b>1,0000</b>	0,8571	0,8571
	Se	<b>0,5000</b>	0,2500	0,2500	<b>0,5000</b>	<b>0,5000</b>	<b>0,5000</b>	<b>0,5000</b>	<b>0,5000</b>	0,2500	<b>0,5000</b>	<b>0,5000</b>
	Sp	0,9310	<b>1,0000</b>	<b>1,0000</b>	0,9310	0,9310	0,9310	0,9310	0,9310	<b>1,0000</b>	0,9310	0,9310
	F <sub>1</sub>	<b>0,6316</b>	0,4000	0,4000	0,6316	<b>0,6316</b>	<b>0,6316</b>	<b>0,6316</b>	<b>0,6316</b>	0,4000	<b>0,6316</b>	<b>0,6316</b>
	MCC	<b>0,4867</b>	0,3928	0,3928	0,4867	<b>0,4867</b>	<b>0,4867</b>	<b>0,4867</b>	<b>0,4867</b>	0,3928	<b>0,4867</b>	<b>0,4867</b>
	AUC	0,7155	<b>0,8664</b>	<b>0,8664</b>	0,7155	0,7155	0,7155	0,7155	0,7155	<b>0,8664</b>	0,7155	0,7155

acurácia de 71,70% e uma AUC de 0,8908, enquanto o SVM alcançou uma acurácia de 77,36% e uma AUC de 0,8865. Esses resultados sugerem que, embora esses modelos obtiveram desempenho satisfatório em termos de AUC, sua capacidade de classificação foi inferior à dos modelos baseados em árvores. Entretanto, quase todos os classificadores obtiveram resultados melhores utilizando a técnica de combinação por *ensemble*.

## 4 Discussão

Os resultados deste estudo destacam a eficácia de modelos baseados em árvores, como RF, GB e DT, na classificação de pacientes com DA utilizando dados de escrita à mão. Esses modelos compartilham a capacidade de capturar relações não lineares e interações complexas entre as características, sendo eficazes para dados que envolvam múltiplos atributos derivados de tarefas de escrita. Além disso, a robustez desses modelos a ruídos e dados desbalanceados contribuiu para seu desempenho superior.

A importância das tarefas individuais no diagnóstico da DA também foi evidenciada pelos resultados. Tarefas de memória/ditado, que avaliam a capacidade de recordação e processamento de informações, demonstraram ser mais desafiadoras para os modelos, refletindo os déficits cognitivos característicos da doença. Por outro lado, tarefas gráficas e de cópia, que avaliam habilidades motoras finas e coordenação, foram mais discriminativas, sugerindo que alterações motoras podem ser um indicador confiável da DA.

A combinação dos resultados por *ensemble*, utilizando média das probabilidades e votação por maioria, demonstrou ser uma estratégia interessante para melhorar a robustez e a generalização dos modelos. A média das probabilidades foi particularmente útil para suavizar as previsões individuais, enquanto a votação por maioria garantiu decisões mais consistentes.

Tabela 3: Resultados das métricas para diferentes métodos de classificação.

Modelo	Método	ACC	Pre	Se	Sp	F <sub>1</sub>	MCC	AUC
AB	Média das Probabilidades	0,7736	0,7500	0,7500	0,7931	0,7500	0,5431	0,8908
	Votação por Maioria	0,7547	0,7200	0,7500	0,7586	0,7347	0,5072	-
KNN	Média das Probabilidades	0,7170	<b>1,0000</b>	0,3750	<b>1,0000</b>	0,5455	0,4972	0,8908
	Votação por Maioria	0,7170	<b>1,0000</b>	0,3750	<b>1,0000</b>	0,5455	0,4972	-
AF	Média das Probabilidades	0,7547	<b>1,0000</b>	0,4583	<b>1,0000</b>	0,6286	0,5626	0,8218
	Votação por Maioria	0,7547	<b>1,0000</b>	0,4583	<b>1,0000</b>	0,6286	0,5626	-
AE	Média das Probabilidades	0,7925	0,8824	0,6250	0,9310	0,7317	0,5930	0,8118
	Votação por Maioria	0,7736	0,8750	0,5833	0,9310	0,7000	0,5577	-
MLP	Média das Probabilidades	0,7736	0,7143	0,8333	0,7241	0,7692	0,5559	0,8405
	Votação por Maioria	0,7736	0,7143	0,8333	0,7241	0,7692	0,5559	-
SVM	Média das Probabilidades	0,7736	0,7500	0,7500	0,7931	0,7500	0,5431	0,8865
	Votação por Maioria	0,7736	0,7308	0,7917	0,7586	0,7600	0,5479	-
RL	Média das Probabilidades	0,7547	0,7200	0,7500	0,7586	0,7347	0,5072	0,8578
	Votação por Maioria	0,7736	0,7308	0,7917	0,7586	0,7600	0,5479	-
RF	Média das Probabilidades	<b>0,8491</b>	0,7857	<b>0,9167</b>	0,7931	<b>0,8462</b>	<b>0,7077</b>	<b>0,9095</b>
	Votação por Maioria	0,8113	0,7333	<b>0,9167</b>	0,7241	0,8148	0,6436	-
GB	Média das Probabilidades	0,8302	0,8000	0,8333	0,8276	0,8163	0,6590	0,8951
	Votação por Maioria	0,8113	0,7917	0,7917	0,8276	0,7917	0,6193	-
DT	Média das Probabilidades	0,8302	0,7778	0,8750	0,7931	0,8235	0,6652	0,8513
	Votação por Maioria	<b>0,8302</b>	0,7778	0,8750	0,7931	<b>0,8235</b>	<b>0,6652</b>	-
ET	Média das Probabilidades	0,7925	0,7241	0,8750	0,7241	0,7925	0,5991	0,8973
	Votação por Maioria	0,8113	0,8182	0,7500	0,8621	0,7826	0,6183	-

Valores em negrito indicam os melhores resultados para cada métrica.

## 5 Considerações Finais

Este trabalho teve como objetivo classificar pacientes entre aqueles com diagnóstico de DA e indivíduos saudáveis, utilizando dados de escrita à mão extraídos do conjunto de dados DARWIN. A metodologia envolveu a divisão das tarefas em três categorias principais (memória/ditado, gráficas e cópia), a aplicação de 11 algoritmos de aprendizado de máquina e a combinação dos resultados por meio de estratégias de *ensemble*, como média das probabilidades e votação por maioria. Os resultados demonstraram que modelos baseados em árvores, como *Random Forest*, *Gradient Boosting* e *Decision Tree*, foram os mais eficazes, destacando-se pela alta acurácia e AUC. Além disso, a combinação dos modelos por *ensemble* permitiu melhorar a robustez e a generalização do sistema de classificação.

Os resultados também evidenciaram a importância de incluir múltiplas tarefas no protocolo de avaliação, uma vez que diferentes categorias de tarefas capturam aspectos distintos da doença. Tarefas de memória/ditado, que avaliam habilidades cognitivas, mostraram-se mais desafiadoras para os modelos, refletindo os déficits característicos de DA. Por outro lado, tarefas gráficas e de cópia, que avaliam habilidades motoras finas, foram mais discriminativas, sugerindo que alterações motoras podem ser um indicador precoce e confiável da doença.

## Referências

- [1] M. Ahmadzadeh et al. "Neuroimaging and machine learning for studying the pathways from mild cognitive impairment to alzheimer's disease: a systematic review". Em: **BMC neurology** 23.1 (2023), pp. 309. DOI: [10.1186/s12883-023-03323-2](https://doi.org/10.1186/s12883-023-03323-2).

- [2] C. Bermudez et al. “Plasma biomarkers for prediction of Alzheimer’s disease neuropathologic change”. Em: **Acta neuropathologica** 146.1 (2023), pp. 13–29. DOI: [10.1007/s00401-023-02594-w](https://doi.org/10.1007/s00401-023-02594-w).
- [3] Z. Breijyeh e R. Karaman. “Comprehensive review on Alzheimer’s disease: causes and treatment”. Em: **Molecules** 25.24 (2020), pp. 5789. DOI: [10.3390/molecules25245789](https://doi.org/10.3390/molecules25245789).
- [4] N. D. Cilia et al. “An Experimental Protocol to Support Cognitive Impairment Diagnosis by using Handwriting Analysis”. Em: **Procedia Computer Science** 141 (2018), pp. 466–471. DOI: [10.1016/j.procs.2018.10.141](https://doi.org/10.1016/j.procs.2018.10.141).
- [5] A. Espinosa et al. “A longitudinal follow-up of 550 mild cognitive impairment patients: evidence for large conversion to dementia rates and detection of major risk factors involved”. Em: **Journal of Alzheimer’s Disease** 34.3 (2013), pp. 769–780. DOI: [10.3233/JAD-122002](https://doi.org/10.3233/JAD-122002).
- [6] C. Fonte et al. “Comparison between physical and cognitive treatment in patients with MIC and Alzheimer’s disease”. Em: **Aging (Albany NY)** 11.10 (2019), pp. 3138. DOI: [10.18632/aging.101970](https://doi.org/10.18632/aging.101970).
- [7] Y. Freund e R. E. Schapire. “Experiments with a new boosting algorithm”. Em: **Proceedings of the Thirteenth International Conference on International Conference on Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156.
- [8] J. Garre-Olmo et al. “Handwriting analysis for early detection of Alzheimer’s disease”. Em: **Journal of Alzheimer’s Disease** 68.4 (2019), pp. 1471–1480. DOI: [10.1002/9781119544487.ch18](https://doi.org/10.1002/9781119544487.ch18).
- [9] P. Gowda, P. H. Reddy e S. Kumar. “Deregulated mitochondrial microRNAs in Alzheimer’s disease: focus on synapse and mitochondria”. Em: **Ageing research reviews** 73 (2022), pp. 101529. DOI: [10.1016/j.arr.2021.101529](https://doi.org/10.1016/j.arr.2021.101529).
- [10] J. Li et al. “Identification of diagnostic genes for both Alzheimer’s disease and Metabolic syndrome by the machine learning algorithm”. Em: **Frontiers in immunology** 13 (2022), pp. 1037318. DOI: [10.3389/fimmu.2022.1037318](https://doi.org/10.3389/fimmu.2022.1037318).
- [11] A. L. Moreno. **Redes Neurais ART e ARTMAP com Treinamento Continuado**. Saarbrücken: Novas Edições Acadêmicas, 2016. ISBN: 978-3330740051.
- [12] A. C. Müller e S. Guido. **Introduction to Machine Learning with Python: A Guide for Data Scientists**. O’Reilly Media, Incorporated, 2018. ISBN: 9789352134571.
- [13] R. O. Roberts et al. “Association between olfactory dysfunction and amnesic mild cognitive impairment and Alzheimer disease dementia”. Em: **JAMA neurology** 73.1 (2016), pp. 93–101. DOI: [10.1001/jamaneurol.2015.2952](https://doi.org/10.1001/jamaneurol.2015.2952).
- [14] P. Scheltens et al. “Alzheimer’s disease”. Em: **The Lancet** 397.10284 (2021), pp. 1577–1590. DOI: [10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4).
- [15] A. Xu, V. L. Kouznetsova e I. F. Tsigelny. “Alzheimer’s disease diagnostics using mirna biomarkers and machine learning”. Em: **Journal of Alzheimer’s Disease** 86.2 (2022), pp. 841–859. DOI: [10.3233/JAD-215502](https://doi.org/10.3233/JAD-215502).
- [16] Y. Zhou et al. “Prediction of Alzheimer’s disease progression based on magnetic resonance imaging”. Em: **ACS Chemical Neuroscience** 12.22 (2021), pp. 4209–4223. DOI: [10.1021/acscchemneuro.1c00472](https://doi.org/10.1021/acscchemneuro.1c00472).
- [17] M. Zuin et al. “Metabolic syndrome and the risk of late onset Alzheimer’s disease: An updated review and meta-analysis”. Em: **Nutrition, Metabolism and Cardiovascular Diseases** 31.8 (2021), pp. 2244–2252. DOI: [10.1016/j.numecd.2021.03.020](https://doi.org/10.1016/j.numecd.2021.03.020).