

Análise de Imagens Faciais para Identificação do Parkinson em Contextos de Dados Limitados

**José Nazareno A. Rodrigues¹, Caio B. A. A. Pereira¹, Rodrigo de M. S. Veras¹,
Pedro H. X. R. Barros², Kelson J. Almeida², Kelson Rômulo T. Aires¹**

¹Universidade Federal do Piauí, Departamento de Computação,
Teresina – PI – Brasil

²Universidade Federal do Piauí, Departamento de Medicina Especializada,
Teresina – PI – Brasil

jose.nazareno@ifpi.edu.br,

{caioburton, rveras, pedrohx587, kelson.almeida, kelson}@ufpi.edu.br

Abstract. This study investigates the use of 2D facial images to identify Parkinson's Disease (PD) in data-limited scenarios. The exploration of facial expressions in 2D images remains scarce in the literature. To help bridge this gap, a complete pipeline was implemented. Predefined facial feature extraction approaches were compared with automatic extraction using pre-trained convolutional neural network models. Among the evaluated models, the combination of DenseNet-201 for feature extraction and SVM for classification achieved the best performance, reaching 94.75% accuracy. The results reinforce the potential of 2D facial image analysis for PD identification, despite challenges related to data scarcity.

Resumo. Este estudo investiga o uso de imagens faciais 2D para identificar a Doença de Parkinson (DP) em cenários com dados limitados. A exploração de expressões faciais em imagens 2D ainda é restrita na literatura. Para ajudar a suprir essa lacuna, foi implementado um pipeline completo. Compararam-se abordagens de extração de características faciais pré-definidas e automática via modelos de redes neurais convolucionais pré-treinadas. Entre os modelos avaliados, a combinação da DenseNet-201 para extração de características e SVM para classificação obteve o melhor desempenho, atingindo 94,75% de acurácia. Os resultados reforçam o potencial da análise de imagens faciais 2D para identificar a DP, apesar dos desafios relacionados à escassez de dados.

1. Introdução

A Doença de Parkinson (DP) é a segunda doença neurodegenerativa mais comum no mundo, afetando cerca de 2–3% da população global acima dos 65 anos (Skaramagkas et al. 2023). O diagnóstico precoce da DP permanece desafiador devido à ausência de testes definitivos, especialmente nos estágios iniciais. Geralmente realizado por especialistas, o diagnóstico clínico baseia-se na observação de sintomas como tremores em repouso, rigidez muscular e hipomimia — uma redução significativa da expressividade facial.

A hipomimia tem se destacado como um biomarcador promissor para a identificação precoce da DP, refletindo a rigidez muscular e a redução dos movimentos faciais (Rajnoha et al. 2018). Ela impacta diretamente a comunicação não verbal dos pacientes. Apesar de sua relevância, ela ainda é pouco explorada na literatura, especialmente em estudos baseados na análise de imagens faciais estáticas. Abordagens baseadas nesse tipo de imagem podem oferecer vantagens significativas, como acessibilidade, baixo custo e fácil integração em soluções computacionais, permitindo sua aplicação em regiões com menor poder econômico (Rajnoha et al. 2018; Rodrigues and Aires 2021). Entretanto, grande parte das pesquisas utiliza dispositivos especializados, como sensores de movimento e gravações de vídeo, que, embora precisos, apresentam alto custo e acesso restrito (Rajnoha et al. 2018; Su et al. 2021).

Em relação à análise de imagens, as redes neurais convolucionais (CNNs) têm demonstrado alta eficácia. No entanto, métodos tradicionais de aprendizado de máquina (AM) continuam sendo relevantes, especialmente em cenários com conjuntos de dados limitados (Rodrigues et al. 2025). Essa situação é comum na área da saúde. Estudos baseados em imagens faciais frequentemente enfrentam a escassez de dados. Este é um fator crítico para o treinamento de modelos de aprendizado profundo, que usualmente exigem milhares de amostras por classe para alcançar um bom desempenho (Rajnoha et al. 2018; Moustafa et al. 2023).

Por outro lado, apesar do sucesso relativo das técnicas tradicionais de AM, essas abordagens enfrentam desafios ao lidar com imagens 2D brutas, que possuem um grande volume de informações. Assim, é comum a aplicação de etapas de pré-processamento e extração de características antes da classificação (Rajnoha et al. 2018). Esse cenário ressalta a importância de se investigar diferentes estratégias de extração de características antes da aplicação de classificadores tradicionais de AM.

Diante desse contexto, este trabalho investiga a viabilidade da identificação acessível da DP por meio da análise de imagens faciais bidimensionais estáticas, utilizando AM tradicional e aprendizado profundo. A metodologia abrange a comparação entre técnicas de extração de características pré-definidas e a extração automática por modelos pré-treinados de aprendizado profundo. Classificadores como *Multi-Layer Perceptron (MLP)*, *Support Vector Machine (SVM)*, *Random Forest* e *Naive Bayes* foram empregados para avaliar o desempenho das características extraídas. A validação cruzada estratificada foi aplicada para garantir que múltiplas imagens do mesmo paciente fossem mantidas no mesmo conjunto, assegurando a robustez dos resultados (Bey et al. 2020).

A próxima seção apresenta os materiais e métodos utilizados na condução do presente trabalho. A Seção 3 apresenta e discute os resultados, comparando-os com o estado da arte. Por fim, a Seção 4 apresenta as conclusões, fazendo também as considerações sobre as limitações deste estudo e indicando possibilidades de trabalhos futuros.

2. Materiais e Métodos

A metodologia deste estudo foi estruturada em sete etapas. São elas: (1) obtenção das imagens faciais, (2) organização das imagens, tendo em vista viabilizar a utilização da validação cruzada estratificada, (3) pré-processamento, (4) extração de características, (5) treinamento dos modelos, (6) avaliação e análise comparativa dos desempenhos obtidos, e (7) otimização de hiperparâmetros. A **Figura 1** apresenta o *pipeline* para estas etapas.

As próximas seções detalham cada uma delas.

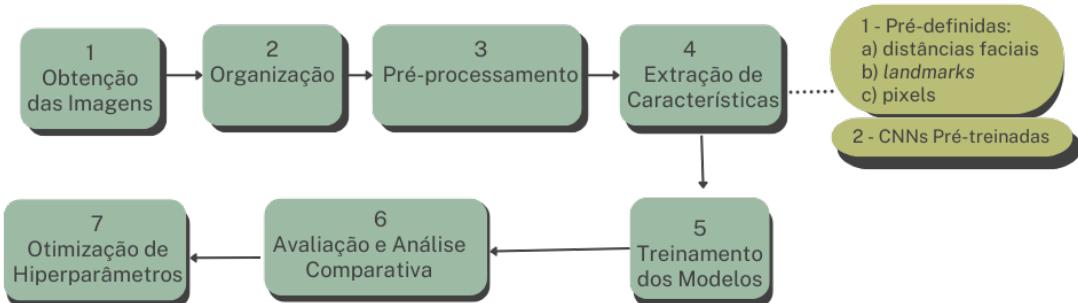


Figura 1. Pipeline implementado.

2.1. Obtenção dos Dados

Os dados foram obtidos após aprovação do Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal do Piauí. Foram recrutados 42 participantes (21 são portadores da DP). As imagens foram capturadas na Clínica de Neurologia da referida instituição.

Durante a captura das imagens solicitou-se aos participantes que expressassem emoções (felicidade, surpresa, medo, nojo, raiva e tristeza), resultando em cerca de três imagens por emoção. Foram coletadas 340 imagens de pacientes com DP e 358 de indivíduos sem Parkinson. Entre os participantes com DP, treze eram homens, com idade média de 61,4 anos e tempo médio de diagnóstico de 5,3 anos.

2.2. Organização dos Dados

Neste trabalho foram utilizadas apenas 346 imagens, sendo 169 provenientes de indivíduos com Parkinson e 177 de indivíduos sem Parkinson, resultando em um dataset relativamente balanceado. As imagens escolhidas representam expressões faciais associadas às emoções de felicidade, nojo e tristeza. A seleção dessas três emoções baseou-se em testes preliminares, que evidenciaram sua maior relevância para diferenciar as classes de indivíduos com e sem Parkinson.

Os dados foram organizados em uma estrutura de pastas que agrupa as imagens por classe e por participante. Essa organização foi essencial para garantir que as imagens do mesmo participante não estivessem simultaneamente nos conjuntos de treino e teste em cada iteração da validação cruzada. Dessa forma, buscou-se reduzir o viés e assegurar uma avaliação mais realista e robusta do modelo. Para viabilizar o uso da validação cruzada estratificada a biblioteca *Scikit-Learn* foi utilizada. Ela implementa essa técnica por meio da classe *StratifiedKFold*.

2.3. Pré-processamento

As imagens foram redimensionadas para um tamanho uniforme de 224 x 224 pixels. Os valores de pixel foram normalizados para o intervalo [0,1] a fim de melhorar a estabilidade do modelo durante o treinamento.

2.4. Extração de Características

Foram realizadas duas abordagens: (1) características pré-definidas e (2) extração com modelos CNN pré-treinados. Para a primeira abordagem três técnicas foram aplicadas: (1a) extração de distâncias faciais, (1b) extração de pontos faciais (*landmarks*) e (1c) extração de pixels. As bibliotecas OpenCV e DLib foram utilizadas para detecção de rostos e pontos faciais, com o preditor “shape_predictor_68_face_landmarks.dat” desta última biblioteca. Na abordagem 1a foram extraídas distâncias euclidianas específicas entre pontos faciais: entre os olhos, do nariz ao queixo, da largura da boca e da boca ao nariz. Na abordagem 1b, todas as coordenadas (X e Y) dos 68 pontos faciais foram extraídas, totalizando 136 características. Na abordagem 1c, todos os pixels da imagem foram utilizados como características.

Para a abordagem com modelos CNNs pré-treinadas foram utilizados os modelos *DenseNet-201*, *MobileNet-V1*, *MobileNet-V2*, *ResNet-50* e *VGG-16*, todos pré-treinados com a base de dados *imagenet*. Em cada caso, as camadas convolucionais do modelo foram congeladas. A saída de cada imagem foi transformada em um vetor de características, possibilitando o uso dos modelos pré-treinados como extratores de características.

As características extraídas de cada imagem foram organizadas em um arquivo do tipo *Comma Separated Values* (CSV), onde cada linha representa uma imagem específica. Esse arquivo inclui além das características faciais extraídas, um identificador único do participante (Id) e um rótulo de classe para classificação binária: 0 indicando “com Parkinson” e 1 indicando “sem Parkinson”. Esse processo é representado pela **Figura 2** em relação à abordagem 1 (extração de características pré-definidas). Os arquivos CSV utilizados neste estudo foram disponibilizados para acesso público¹.

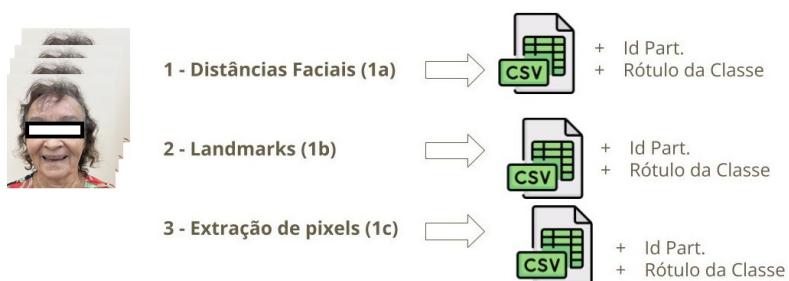


Figura 2. Processo: etapa extração de características pré-definidas.

2.5. Treinamento

Para o treinamento dos modelos foram utilizados diversos classificadores de AM tradicional: *Multi-Layer Perceptron (MLP)*, *Support Vector Machine (SVM)*, *Decision Tree*, *Random Forest*, *Naive Bayes* e o *K-Nearest Neighbors (KNN)*. Esses algoritmos foram aplicados e avaliados a fim de se identificar o desempenho de cada um deles na classificação a partir das características extraídas das imagens faciais.

2.6. Avaliação

Os modelos foram treinados utilizando validação cruzada estratificada com 10 *folds*, garantindo que as imagens do mesmo participante não estivessem simultaneamente nos con-

¹Disponível em: https://bit.ly/parkinson_sbcas.

juntos de treino e teste. Segundo Bey et al. 2020, essa técnica aprimora a validação cruzada convencional ao organizar os registros em dobras estratificadas, agrupando pacientes semelhantes e minimizando vieses de superestimação da acurácia causados por duplicatas. Esse método é especialmente relevante em cenários onde a repetição de amostras pode levar a estimativas excessivamente otimistas. No presente estudo, o identificador do participante foi utilizado como critério de agrupamento, assegurando que múltiplas imagens do mesmo indivíduo fossem mantidas no mesmo conjunto, evitando a contaminação entre treino e teste. Com isso, a avaliação buscou ser mais consistente, evitando a superestimação do desempenho dos classificadores (Bey et al. 2020).

A avaliação do desempenho dos classificadores foi realizada por meio de métricas amplamente utilizadas na literatura: acurácia, precisão, revocação (*recall*), F1-score e coeficiente Kappa. Esses indicadores foram calculados para todos os modelos.

2.7. Otimização

A otimização de hiperparâmetros foi realizada nos classificadores com melhor desempenho (*Random Forest*, *SVM* e *MLPClassifier*), explorando diferentes combinações de parâmetros para aprimorar a performance dos modelos. Para o *Random Forest* foram ajustados os parâmetros *n_estimators* (número de árvores), *max_depth* (profundidade máxima), *min_samples_split* e *min_samples_leaf* (mínimo de amostras para dividir um nó e formar uma folha, respectivamente), além do *bootstrap*, que controla o reamostramento. Configurações específicas, como *min_samples_split* = 5 e *min_samples_leaf* = 1, foram adotadas para equilibrar complexidade e generalização. O parâmetro *max_depth* foi definido como *None*, permitindo o crescimento completo das árvores, embora com risco de sobreajuste em conjuntos menores (Hastie et al. 2009). O uso de *bootstrap*=*True* buscou aumentar a diversidade das árvores, favorecendo a generalização do modelo (Breiman 2001).

Para o *SVM* foram testados quatro *kernels* (*linear*, *RBF*, *polynomial* e *sigmoid*), variando os parâmetros *C* (controle da margem de erro), *gamma* (influência das amostras de treino) e *degree* (grau do kernel polinomial). O *C* foi testado com valores entre 0.1 e 100, enquanto *gamma* foi ajustado para *scale*, *auto* e valores numéricos (0.01, 0.1 e 1). O *degree*, aplicável ao kernel polinomial, foi avaliado para os valores 2, 3 e 4.

No *MLPClassifier*, foram exploradas cinco configurações para o tamanho das camadas ocultas: (50,), (100,), (50, 50), (100, 50) e (100, 100). As funções de ativação *tanh* e *relu* foram analisadas, assim como os otimizadores *adam* e *sgd*. A regularização L2 (*alpha*) foi ajustada para 0.0001, 0.001 e 0.01, e a taxa de aprendizado (*learning_rate*) foi testada nas modalidades *constant* e *adaptive*, permitindo ajustes dinâmicos ao longo do treinamento.

3. Resultados e Discussão

Esta seção apresenta e discute os resultados obtidos nas diferentes abordagens de extração de características e classificação. A análise é estruturada em quatro partes principais, organizadas para facilitar a compreensão e destacar os principais achados do estudo.

3.1. Extração de Características Pré-definidas

Nesta abordagem os arquivos CSV utilizados e gerados apresentaram pequenas discrepâncias referentes ao número de imagens processadas. Para a extração de distâncias

faciais (1a) e *landmarks* (1b) foram registrados 164 entradas para a classe 0 (com Parkinson), enquanto o esperado era 169. Já para a classe 1 (sem Parkinson), foram registradas 177 entradas, correspondendo ao número total de imagens disponíveis para essa classe. Tal discrepância pode indicar dificuldades das bibliotecas de visão computacional utilizadas, em detectar rostos em algumas imagens da classe 0. Na extração de *pixels* (1c), não foram observadas anomalias no arquivo CSV gerado, correspondendo com exatidão ao número de imagens para cada classe.

Quanto aos atributos, os arquivos CSV para a abordagem manual apresentaram, respectivamente, 6 (seis), 138 (cento e trinta e oito) e 50.178 (cinquenta mil, cento e setenta e oito) colunas para as abordagens (a), (b) e (c). Esses números correspondem à quantidade esperada de atributos, além dos campos adicionais para a classe e o Id do participante. Essa conformidade indica o êxito obtido nas etapas de pré-processamento e extração de características, assegurando que os dados estejam devidamente estruturados e adequados para serem utilizados pelos classificadores. Além disso, a etapa de organização dos dados foi fundamental para garantir que imagens do mesmo participante não estivessem simultaneamente nos conjuntos de treino e teste, utilizando o método de validação cruzada estratificada.

A **Tabela 1** apresenta o desempenho dos classificadores para as abordagens de extração de características pré-definidas, evidenciando a influência do método de extração nos resultados. O *MLPClassifier* obteve melhor desempenho com distâncias faciais, enquanto o *SVM* se destacou na extração de *landmarks* e o *Random Forest* obteve maior acurácia com a abordagem baseada em *pixels*. Esses achados reforçam a importância de alinhar a escolha do classificador ao tipo de característica extraída, considerando tanto as particularidades do conjunto de dados quanto as exigências computacionais de cada método.

Tabela 1. Desempenho dos classificadores utilizando extração de características pré-definidas.

Classificador	Acur. (%)	Prec. (%)	Rec. (%)	F1 (%)	Kappa (%)
Método de extração: 1a - distâncias faciais					
MLPClassifier	72,65 ± 12,30	75,64 ± 17,51	80,64 ± 10,82	76,21 ± 9,06	44,49 ± 23,69
SVM	70,40 ± 14,65	71,11 ± 14,13	74,86 ± 15,65	72,42 ± 13,67	39,41 ± 29,01
DecisionTree	67,04 ± 10,29	75,39 ± 16,51	64,20 ± 14,44	66,89 ± 17,87	33,44 ± 18,93
RandomForest	69,08 ± 9,67	76,63 ± 15,63	67,87 ± 19,04	69,02 ± 9,19	37,87 ± 17,16
KNN	70,10 ± 11,33	70,06 ± 15,00	77,60 ± 16,26	72,45 ± 12,59	39,27 ± 21,97
NaiveBayes	65,71 ± 8,49	66,01 ± 13,81	84,04 ± 19,92	71,21 ± 8,09	30,01 ± 17,75
Método de extração: 1b - Landmarks					
MLPClassifier	61,13 ± 12,61	65,92 ± 15,51	60,35 ± 15,67	61,63 ± 12,11	21,19 ± 26,24
SVM	84,35 ± 10,35	87,38 ± 8,74	81,83 ± 22,11	82,55 ± 14,87	67,62 ± 21,31
DecisionTree	55,76 ± 14,97	62,77 ± 17,32	54,27 ± 18,13	55,51 ± 14,25	13,98 ± 28,37
RandomForest	63,65 ± 21,85	64,84 ± 20,17	67,36 ± 25,72	64,88 ± 21,86	27,61 ± 43,15
KNN	61,32 ± 15,82	62,72 ± 16,90	70,73 ± 17,62	65,31 ± 14,47	22,38 ± 30,69
NaiveBayes	54,53 ± 12,75	57,09 ± 15,01	77,64 ± 19,01	63,67 ± 11,25	8,09 ± 22,25
Método de extração: 1c - pixels					
MLPClassifier	75,90 ± 12,82	83,10 ± 14,45	74,63 ± 26,96	74,08 ± 16,07	50,39 ± 26,25
SVM	75,70 ± 12,87	82,22 ± 16,58	74,47 ± 24,53	73,32 ± 19,27	48,06 ± 28,95
DecisionTree	65,36 ± 13,24	69,26 ± 12,81	62,76 ± 22,18	63,64 ± 15,45	30,56 ± 25,84
RandomForest	79,61 ± 12,32	85,90 ± 12,13	77,57 ± 23,95	75,81 ± 23,29	56,92 ± 27,03
KNN	77,62 ± 13,13	81,46 ± 21,87	72,57 ± 29,35	73,02 ± 23,18	54,49 ± 28,10
NaiveBayes	64,17 ± 19,13	66,64 ± 20,51	76,01 ± 24,07	68,07 ± 16,90	27,28 ± 37,10

3.2. Extração com Modelos CNN Pré-treinados

A **Tabela 2** apresenta os resultados dos classificadores combinados com diferentes modelos pré-treinados de CNNs como extratores de características. O *SVM* associado à *DenseNet-201* obteve o melhor desempenho geral, alcançando uma acurácia de $93,27\% \pm 6,83$, precisão de $92,93\% \pm 8,45$, *recall* de $94,33\% \pm 12,48$ e F1-score de $92,83\% \pm 8,06$. O coeficiente Kappa ($86,22\% \pm 14,06$) indica uma alta concordância estatística além do acaso, reforçando a confiabilidade dessa abordagem.

A *MobileNet-V1* apresentou o desempenho mais modesto, com $83,59\% \pm 16,04$ de acurácia usando o *MLPClassifier*. Já a *MobileNet-V2* mostrou uma melhor generalização, atingindo $90,60\% \pm 8,91$ de acurácia com o *SVM*. Isso a torna uma alternativa viável para dispositivos com menor capacidade de processamento. O *ResNet-50* também apresentou bons resultados, com $90,98\% \pm 8,95$ de acurácia, mas não superou a *DenseNet-201* em nenhuma métrica. No caso do *VGG-16*, o *Random Forest* foi o classificador mais eficiente, atingindo $90,15\% \pm 8,20$ de acurácia e uma alta precisão de $93,74\% \pm 7,35$, indicando que essa arquitetura ainda é competitiva como extratora de características.

Esses achados destacam a relevância da escolha da arquitetura da CNN e do classificador para otimizar o desempenho em aplicações específicas. O *SVM* foi o classificador com melhor desempenho na maioria dos casos, especialmente quando combinado com a *DenseNet-201* como extratora de características.

Tabela 2. Desempenho dos classificadores utilizando modelos pré-treinados.

Classificador	Acur. (%)	Prec. (%)	Rec. (%)	F1 (%)	Kappa (%)
Modelo Pré-treinado: DenseNet-201					
MLPClassifier	$85,22 \pm 17,09$	$88,37 \pm 16,77$	$85,98 \pm 17,82$	$85,95 \pm 14,78$	$69,61 \pm 35,67$
SVM	$93,27 \pm 6,83$	$92,93 \pm 8,45$	$94,33 \pm 12,48$	$92,83 \pm 8,06$	$86,22 \pm 14,06$
DecisionTree	$69,61 \pm 11,52$	$71,43 \pm 7,75$	$67,37 \pm 22,35$	$67,71 \pm 14,62$	$40,01 \pm 21,75$
RandomForest	$86,98 \pm 14,06$	$90,06 \pm 12,57$	$85,67 \pm 20,80$	$86,31 \pm 14,43$	$73,72 \pm 27,86$
KNN	$69,27 \pm 15,31$	$65,09 \pm 14,42$	$96,82 \pm 6,61$	$76,97 \pm 10,14$	$35,28 \pm 33,20$
NaiveBayes	$70,49 \pm 11,17$	$87,47 \pm 12,25$	$51,30 \pm 19,50$	$62,36 \pm 16,74$	$42,02 \pm 20,95$
Modelo Pré-treinado: MobileNet-V1					
MLPClassifier	$83,59 \pm 16,04$	$87,90 \pm 15,87$	$84,93 \pm 16,82$	$84,86 \pm 13,68$	$65,86 \pm 33,78$
SVM	$83,24 \pm 14,39$	$85,71 \pm 15,27$	$88,49 \pm 17,09$	$85,02 \pm 11,98$	$64,72 \pm 30,86$
DecisionTree	$63,59 \pm 18,73$	$66,79 \pm 24,12$	$62,66 \pm 27,03$	$61,88 \pm 21,97$	$26,78 \pm 38,07$
RandomForest	$80,67 \pm 18,95$	$87,35 \pm 18,14$	$80,49 \pm 23,99$	$81,09 \pm 17,17$	$59,85 \pm 39,92$
KNN	$51,39 \pm 6,83$	$51,39 \pm 6,83$	$100,00 \pm 0,00$	$67,62 \pm 6,09$	$0,00 \pm 0,00$
NaiveBayes	$68,51 \pm 10,83$	$96,26 \pm 6,42$	$40,63 \pm 22,40$	$53,28 \pm 23,68$	$37,42 \pm 22,24$
Modelo Pré-treinado: MobileNet-V2					
MLPClassifier	$80,83 \pm 16,42$	$85,70 \pm 16,23$	$79,87 \pm 21,29$	$80,72 \pm 15,18$	$61,35 \pm 33,25$
SVM	$90,60 \pm 8,91$	$89,13 \pm 13,00$	$95,56 \pm 7,87$	$91,40 \pm 8,26$	$80,94 \pm 17,30$
DecisionTree	$63,43 \pm 20,44$	$66,85 \pm 21,09$	$68,98 \pm 23,70$	$65,80 \pm 17,98$	$26,84 \pm 41,45$
RandomForest	$82,47 \pm 17,98$	$85,38 \pm 18,43$	$83,78 \pm 18,71$	$83,30 \pm 15,96$	$64,59 \pm 36,45$
KNN	$54,54 \pm 8,00$	$53,23 \pm 5,93$	$100,00 \pm 0,00$	$69,27 \pm 5,20$	$4,68 \pm 14,04$
NaiveBayes	$70,66 \pm 14,14$	$90,00 \pm 30,00$	$42,30 \pm 29,66$	$53,58 \pm 28,90$	$41,05 \pm 29,73$
Modelo Pré-treinado: ResNet-50					
MLPClassifier	$87,07 \pm 13,63$	$91,56 \pm 12,88$	$86,12 \pm 18,70$	$87,19 \pm 13,27$	$73,67 \pm 27,52$
SVM	$90,98 \pm 8,95$	$91,27 \pm 11,70$	$93,89 \pm 12,92$	$91,34 \pm 8,56$	$81,09 \pm 18,87$
DecisionTree	$85,65 \pm 12,87$	$91,09 \pm 10,06$	$80,46 \pm 26,59$	$81,82 \pm 21,93$	$69,95 \pm 27,47$
RandomForest	$82,87 \pm 17,96$	$86,13 \pm 17,49$	$84,35 \pm 18,82$	$83,71 \pm 14,51$	$64,18 \pm 38,16$
KNN	$70,32 \pm 15,84$	$67,96 \pm 15,19$	$95,67 \pm 8,95$	$77,98 \pm 8,92$	$36,84 \pm 34,22$
NaiveBayes	$71,58 \pm 13,97$	$94,42 \pm 10,27$	$46,22 \pm 27,26$	$57,77 \pm 26,64$	$43,19 \pm 27,89$
Modelo Pré-treinado: VGG-16					
MLPClassifier	$85,76 \pm 13,21$	$84,88 \pm 13,96$	$93,03 \pm 12,01$	$87,86 \pm 10,99$	$69,30 \pm 28,85$
SVM	$87,03 \pm 10,81$	$84,96 \pm 13,61$	$96,00 \pm 12,00$	$88,77 \pm 8,86$	$71,99 \pm 24,13$
DecisionTree	$66,20 \pm 17,03$	$72,63 \pm 19,19$	$63,51 \pm 24,28$	$64,50 \pm 16,88$	$31,71 \pm 35,53$
RandomForest	$90,15 \pm 8,20$	$93,74 \pm 7,35$	$88,45 \pm 17,43$	$89,54 \pm 9,99$	$79,80 \pm 16,46$
KNN	$51,39 \pm 6,83$	$51,39 \pm 6,83$	$100,00 \pm 0,00$	$67,62 \pm 6,09$	$0,00 \pm 0,00$
NaiveBayes	$68,23 \pm 11,81$	$98,00 \pm 4,00$	$40,24 \pm 21,38$	$53,89 \pm 20,51$	$38,03 \pm 22,10$

3.3. Impacto da otimização de hiperparâmetros

Em vista dos resultados alcançados, infere-se que a atividade de otimização ajudou a maximizar os resultados, cumprindo com o seu propósito de melhorar o desempenho dos classificadores. Esses ajustes permitiram que cada método de extração fosse combinado com o classificador e a configuração de hiperparâmetros mais adequados, maximizando o desempenho da classificação. A **Tabela 3** proporciona uma análise detalhada em relação às métricas utilizadas nas abordagens baseadas em extração manual e apoiada em modelos CNNs pré-treinados.

Tabela 3. Desempenho dos melhores classificadores antes e depois da otimização, utilizando diferentes métodos de extração de características.

Classificador	Acur. (%)	Prec. (%)	Rec. (%)	F1 (%)	Kappa (%)
Método: Extração de distâncias faciais					
MLPClassifier	$72,65 \pm 12,30$	$75,64 \pm 17,51$	$80,64 \pm 10,82$	$76,21 \pm 9,06$	$44,49 \pm 23,69$
MLPClassifier*	$73,86 \pm 12,80$	$78,53 \pm 17,81$	$79,53 \pm 15,28$	$76,43 \pm 10,00$	$46,97 \pm 24,44$
Método: Extração de landmarks					
SVM	$84,35 \pm 10,35$	$87,38 \pm 8,74$	$81,83 \pm 22,11$	$82,55 \pm 14,87$	$67,62 \pm 21,31$
SVM*	$84,38 \pm 12,72$	$88,51 \pm 11,85$	$83,28 \pm 18,60$	$84,32 \pm 12,32$	$68,04 \pm 25,59$
Método: Extração de pixels					
RandomForest	$79,61 \pm 12,32$	$85,90 \pm 12,13$	$77,57 \pm 29,53$	$75,81 \pm 23,29$	$56,92 \pm 27,03$
RandomForest*	$81,20 \pm 12,18$	$88,70 \pm 12,69$	$77,12 \pm 23,66$	$79,37 \pm 14,45$	$61,55 \pm 24,43$
Modelo Pré-treinado: DenseNet-201					
SVM	$93,27 \pm 6,83$	$92,93 \pm 8,45$	$94,33 \pm 12,48$	$92,83 \pm 8,06$	$86,22 \pm 14,06$
SVM*	$94,75 \pm 5,50$	$93,50 \pm 7,80$	$95,70 \pm 6,90$	$93,88 \pm 6,82$	$87,50 \pm 12,30$
Modelo Pré-treinado: MobileNet-V1					
MLPClassifier	$83,59 \pm 16,04$	$87,90 \pm 15,87$	$84,93 \pm 16,82$	$84,86 \pm 13,68$	$65,86 \pm 33,78$
MLPClassifier*	$85,23 \pm 14,45$	$88,15 \pm 12,10$	$86,97 \pm 13,00$	$85,92 \pm 11,25$	$68,34 \pm 31,54$
Modelo Pré-treinado: MobileNet-V2					
SVM	$90,60 \pm 8,91$	$89,13 \pm 13,00$	$95,56 \pm 7,87$	$91,40 \pm 8,26$	$80,94 \pm 17,30$
SVM*	$91,90 \pm 7,50$	$90,85 \pm 11,20$	$96,45 \pm 6,58$	$92,70 \pm 7,90$	$83,45 \pm 15,90$
Modelo Pré-treinado: ResNet-50					
SVM	$90,98 \pm 8,95$	$91,27 \pm 11,70$	$93,89 \pm 12,92$	$91,34 \pm 8,56$	$81,09 \pm 18,87$
SVM*	$92,50 \pm 7,25$	$91,90 \pm 10,35$	$95,70 \pm 8,10$	$92,40 \pm 7,89$	$84,20 \pm 16,90$
Modelo Pré-treinado: VGG-16					
RandomForest	$90,15 \pm 8,20$	$93,74 \pm 7,35$	$88,45 \pm 17,43$	$89,54 \pm 9,99$	$79,80 \pm 16,46$
RandomForest*	$91,80 \pm 7,00$	$94,05 \pm 6,30$	$89,30 \pm 15,10$	$90,25 \pm 9,45$	$82,10 \pm 14,78$

* indica que o classificador de AM foi otimizado.

3.4. Considerações finais sobre os resultados

A **Figura 3** compara a acurácia dos métodos de extração de características pré-definidas com os de modelos CNNs pré-treinados, evidenciando a superioridade destas últimas. As CNNs apresentaram acurácia mais alta e menor variabilidade, com valores concentrados próximos a 90%, enquanto os métodos de extração de características pré-definidas exibiram maior dispersão (55% a 85%). Esses resultados indicam que as CNNs extraem características mais discriminativas e são mais robustas a variações nos dados.

A **Figura 4** compara os três melhores resultados de acurácia para cada tipo de extração, reforçando que as CNNs não apenas atingiram as maiores acurácia, como também apresentaram menor dispersão. A extração de características pré-definidas mostrou maior variabilidade, especialmente nos métodos baseados em *landmarks*, sugerindo maior sensibilidade às variações dos dados e dos classificadores utilizados.

Esses achados sugerem que a extração de características por CNNs pré-treinadas é a abordagem mais eficaz e estável para o reconhecimento de padrões faciais. Em

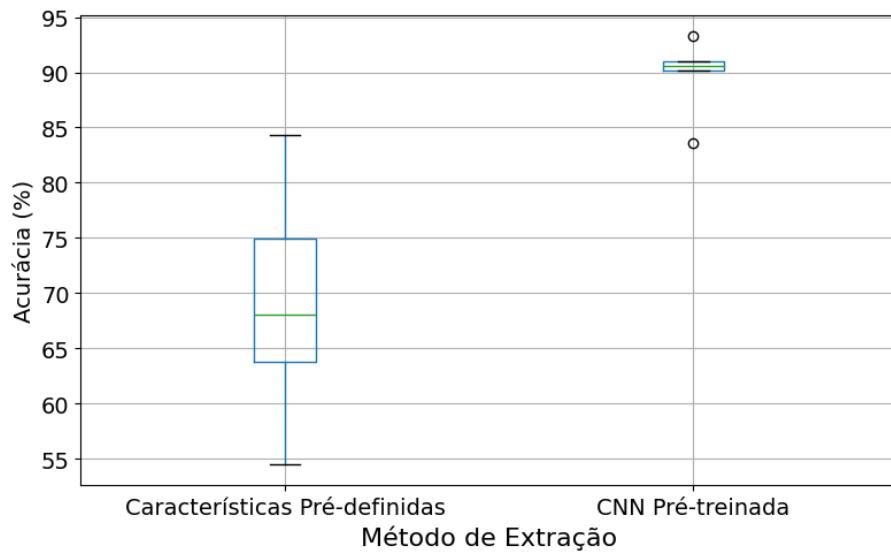


Figura 3. Boxplot: Acurácia por tipo de extração.

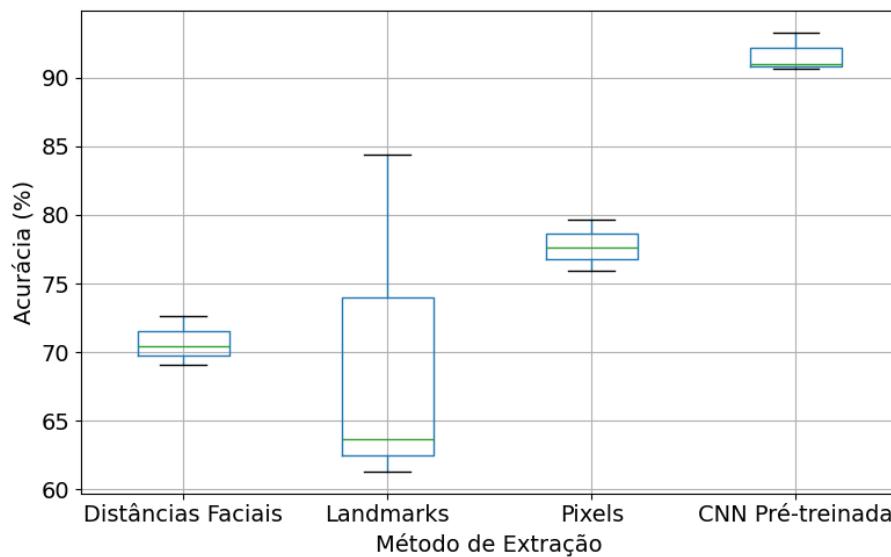


Figura 4. Boxplot: comparação dos 3 melhores resultados de acurácia por tipo de extração.

comparação com abordagens baseadas na extração de características pré-definidas, como distâncias faciais e *landmarks*, os métodos propostos neste estudo ofereceram melhores resultados, inclusive quando combinados com classificadores otimizados.

Em relação ao estado da arte, Rodrigues et al. 2025 apresenta uma revisão sistemática que consolida os principais estudos voltados à identificação da DP por meio de imagens faciais 2D. Dentre as abordagens identificadas, destaca-se a proposta de Rajnoha et al. 2018. Nela, a CNN é treinada para extrair 128 medidas faciais, que depois são analisados por algoritmos tradicionais de AM. O referido trabalho alcançou 67,33% de acurácia utilizando árvores de decisão para fazer a classificação. Desempenho um pouco melhor foi mencionado em Skibińska and Burget 2020, que atingiram 69% de acurácia também em contexto de dados limitados. Jin et al. 2020 alcançaram uma acurácia de 86% e F1-score de 75% a partir de expressões faciais relacionadas a sorrisos. Outro trabalho semelhante é o de Guan 2021. Nele, a partir de expressões faciais, analisadas com modelos de Regressão Logística e SVM foi reportado um resultado de 90,06% de acurácia.

No âmbito das abordagens de extração de características pré-definidas deste trabalho, observa-se que o classificador SVM, utilizando *landmarks*, atingiu uma acurácia de 84,35%. Este desempenho supera o que foi reportado em Rajnoha et al. 2018 e Skibińska and Burget 2020. Já a abordagem de extração automática de características a partir do modelo CNN pré-treinado *DenseNet-201*, em combinação com o classificador SVM, obteve desempenho superior a todos os trabalhos mencionados no parágrafo anterior, atingindo 94,85% de acurácia. Isto demonstra que os resultados desta pesquisa estão alinhados com o estado da arte.

Apesar dessa comparação, o uso extensivo de *datasets* proprietários identificado na literatura dificulta a realização de análises comparativas diretas entre os estudos (Rodrigues et al. 2025). Isso ocorre porque a natureza dos dados está diretamente relacionada ao desempenho dos classificadores. Nesse sentido, o único trabalho relacionado identificado na literatura que utilizou *dataset* semelhante foi o de Pereira et al. 2024. No referido trabalho, além do desempenho inferior com o *DenseNet-201* (92,77% de acurácia), foram utilizadas 698 imagens abrangendo todas as seis expressões faciais analisadas. Por outro lado, no presente trabalho foram utilizadas apenas 346 imagens, correspondentes a apenas três emoções: felicidade, nojo e tristeza. Além disso, o trabalho de Pereira et al. 2024 utilizou a técnica de validação cruzada convencional, que pode superestimar o desempenho dos modelos, já que imagens de um mesmo paciente podem estar simultaneamente nos conjuntos de treino e teste (Bey et al. 2020).

Adicionalmente, é válido destacar o trabalho de Warnakulasuriya et al. 2023, que, além de outras abordagens, como imagens de ressonância magnética, desenhos à mão e análise de voz, também utilizou imagens faciais. Essa abordagem multimodal reforça a relevância das expressões faciais também como um componente significativo na detecção da DP. No trabalho de Warnakulasuriya et al. 2023, foram consideradas três expressões faciais específicas: sorriso, nojo e surpresa, alcançando uma acurácia de 95%, semelhante à do presente trabalho. A **Tabela 4** apresenta uma síntese comparativa destes trabalhos relacionados.

Tabela 4. Comparação entre estudos relacionados.

Estudo	Abordagem	Dados	Resultados (Acur.)
Rajnoha et al. 2018	AM e aprendizado profundo a partir de imagens faciais extraídas de vídeos	Proprietário (100 participantes, 50 DP)	67,33%
Skibińska and Burget 2020	AM a partir de imagens faciais extraídas de vídeos (análise de mudanças emocionais durante a pronúncia de fala)	Proprietário (70 DP e 45 sem Parkinson) + FER2013	69,00%
Jin et al. 2020	AM e aprendizado profundo a partir de imagens extraídas de vídeos (expressões faciais)	Proprietário (64 participantes, 33 DP)	86,00%
Guan 2021	AM a partir de imagens faciais extraídas de vídeos (expressões faciais)	Proprietário	90,06%
Warnakulasu et al. 2023	AM e aprendizado profundo a partir de imagens faciais simples (expressões faciais: sorriso, nojo e surpresa)	Proprietário: 35 pacientes com DP e 70 indivíduos saudáveis	95,00%
Pereira et al. 2024	Aprendizado profundo a partir de imagens simples (expressões faciais: sorriso, tristeza, nojo, surpresa, raiva e medo)	Proprietário (42 participantes, 21 DP)	92,77%
Este estudo (2025)	AM e Aprendizado profundo a partir de imagens simples (expressões faciais: sorriso, nojo e tristeza)	Proprietário (42 participantes, 21 DP)	94,75%

4. Conclusões

Os resultados sugerem que a abordagem utilizada no presente estudo pode apoiar soluções computacionais acessíveis e de baixo custo por meio da exploração de imagens faciais bidimensionais estáticas. Os achados indicam que, embora as abordagens de extração de características pré-definidas tenham apresentado resultados significativos, as CNNs demonstraram maior desempenho, destacando-se na captura de padrões faciais. O problema do número limitado de participantes foi atenuado pelo uso de validação cruzada estratificada, que reduziu vieses decorrentes do tamanho reduzido da amostra e viabilizou modelos mais robustos. Apesar disso, o número restrito de participantes pode impactar a generalização dos modelos para diferentes populações. Trabalhos futuros devem incluir conjuntos de dados mais amplos e ainda, explorar outras possibilidades de extração de características, como outras distâncias faciais, visando aumentar a precisão dos modelos.

Referências

- [Bey et al. 2020] Bey, R., Goussault, R., Grolleau, F., Benchoufi, M., and Porcher, R. (2020). Fold-stratified cross-validation for unbiased and privacy-preserving federated learning. *Journal of the American Medical Informatics Association*, 27(8):1244–1251.
- [Breiman 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- [Guan 2021] Guan, Y. (2021). Application of logistic regression algorithm in the diagnosis of expression disorder in parkinson’s disease. In *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 2, pages 1117–1120. IEEE.
- [Hastie et al. 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

- [Jin et al. 2020] Jin, B., Qu, Y., Zhang, L., and Gao, Z. (2020). Diagnosing parkinson disease through facial expression recognition: video analysis. *Journal of Medical Internet Research*, 22(7):e18697.
- [Moustafa et al. 2023] Moustafa, K., Metawie, H., Saadoun, S., Sameh, N., Ibrahim, R., and Abdelsayed, A. (2023). Easy park: Mobile application for parkinson’s disease detection and severity level. In *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 1–8. IEEE.
- [Pereira et al. 2024] Pereira, C., Barros, P., Rodrigues, J., Araújo, P., Borges, R., Almeida, K., and Veras, R. (2024). Identificação de parkinson em imagens faciais usando modelos de deep learning pré-treinados. In *Anais da XII Escola Regional de Computação do Ceará, Maranhão e Piauí*, pages 169–178, Porto Alegre, RS, Brasil. SBC.
- [Rajnoha et al. 2018] Rajnoha, M., Mekyska, J., Burget, R., Eliasova, I., Kostalova, M., and Rektorova, I. (2018). Towards identification of hypomimia in parkinson’s disease based on face recognition methods. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 1–4. IEEE.
- [Rodrigues and Aires 2021] Rodrigues, J. N. A. and Aires, K. R. T. (2021). Uma pesquisa exploratória sobre a utilização de soluções apoiadas em inteligência artificial e tecnologias móveis com portadores de transtorno do espectro autista. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 410–415. SBC.
- [Rodrigues et al. 2025] Rodrigues, J. N. A., Veras, R. d. M. S., Neto, L. d. S. B., Barros, P. H. X. R., Moura, W. d. S., De Almeida, K. J. S., and Aires, K. R. T. (2025). Identification of parkinson’s disease through facial image classification: A systematic review. *IEEE Access*, pages 46720–46731.
- [Skaramagkas et al. 2023] Skaramagkas, V., Pentari, A., Kefalopoulou, Z., and Tsiknakis, M. (2023). Multi-modal deep learning diagnosis of parkinson’s disease-a systematic review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.
- [Skibińska and Burget 2020] Skibińska, J. and Burget, R. (2020). Parkinson’s disease detection based on changes of emotions during speech. In *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 124–130. IEEE.
- [Su et al. 2021] Su, G., Lin, B., Yin, J., Luo, W., Xu, R., Xu, J., and Dong, K. (2021). Detection of hypomimia in patients with parkinson’s disease via smile videos. *Annals of Translational Medicine*, 9(16).
- [Warnakulasuriya et al. 2023] Warnakulasuriya, N., De Silva, S., Madushika, J., Gamage, S., Jayawardena, S., and Karunasena, A. (2023). Multimodal fusion for enhanced parkinson’s disease screening: Integrating brain mri, hand drawing, facial expressions, and voice analysis. In *2023 5th International Conference on Advancements in Computing (ICAC)*, pages 531–536. IEEE.