

# Identificação de Parkinson em Imagens Faciais Usando Modelos de Deep Learning Pré-Treinados

Caio B. A. A. Pereira<sup>1</sup>, Pedro H. X. R. Barros<sup>2</sup>, José N. A. Rodrigues<sup>3</sup>, Paulo H. R. Araújo<sup>1</sup>,  
Rodrigo N. Borges<sup>1</sup>, Kelson J. Almeida<sup>2</sup>, Rodrigo M. S. Veras<sup>1</sup>

<sup>1</sup>Departamento de Computação,  
Universidade Federal do Piauí - Teresina, Brasil

<sup>2</sup>Departamento de Medicina,  
Universidade Federal do Piauí - Teresina, Brasil

<sup>3</sup>Instituto Federal do Piauí - Teresina, Brasil

{caioburton, r\_borges, kelson.almeida, rveras}@ufpi.edu.br,  
{pedrohx587, phrod.devp}@gmail.com, jose.nazareno@ifpi.edu.br

**Abstract.** *This article evaluated the use of pre-trained deep learning models in the classification of facial images to differentiate between healthy individuals and patients with Parkinson's Disease (PD). We utilized a dataset consisting of 340 images of PD patients and 358 images of healthy individuals, applying techniques such as group cross-validation with 5 folds, data augmentation, and fine-tuning. The best performance was achieved with the DenseNet-201 model, which showed an average accuracy of 92.77%, F1-score of 92.48%, and a Kappa index of 85.33%. These results suggest that pre-trained CNNs are promising for the detection of Parkinson's Disease in facial images.*

**Resumo.** *Este artigo avaliou o uso de modelos de deep learning pré-treinados na classificação de imagens faciais para diferenciar entre indivíduos saudáveis e pacientes com Doença de Parkinson (DP). Utilizamos um dataset composto por 340 imagens de pacientes com DP e 358 imagens de indivíduos saudáveis, aplicando técnicas como validação cruzada por grupos com 5 folds, aumento de dados e ajuste fino. O melhor desempenho foi obtido com o modelo DenseNet-201, que apresentou uma acurácia média de 92,77%, F1-score de 92,48% e índice Kappa de 85,33%. Esses resultados sugerem que as CNNs pré-treinadas são promissoras para a detecção de Parkinson em imagens faciais.*

## 1. Introdução

A doença de Parkinson (DP) é a segunda doença neurodegenerativa mais prevalente, afetando aproximadamente 2% das pessoas com mais de 65 anos, com sua prevalência aumentando com o avançar da idade [Rajnoha et al. 2018]. Dados relatam que a doença de Parkinson atinge mais de 6 milhões de pessoas em todo o mundo, representando um aumento de 2,5 vezes ao longo das últimas três décadas [Huang et al. 2023]. Além dos sintomas clássicos como rigidez, instabilidade postural, bradicinesia e tremor, a DP também ocasiona alterações nas funções dopaminérgicas putaminais, orbitofrontais e amigdalares, resultando em hipomimia (conhecida como "rosto mascarado"), uma interferência gestual que limita a expressão emocional do paciente e impacta seu desenvolvimento social e qualidade de vida [Lee and Koh 2015].



(a) Com Doença de Parkinson. (b) Sem Doença de Parkinson.

**Figura 1. Exemplos de pessoas com e sem Doença de Parkinson.**

Embora seja amplamente reconhecido que ainda não existe uma cura definitiva para a doença de Parkinson, é consenso que o diagnóstico precoce e o tratamento imediato desempenham papéis essenciais na mitigação dos sintomas e na desaceleração da progressão da condição [Huang et al. 2023].

O diagnóstico da doença de Parkinson geralmente pode ser categorizado em dois tipos: *in-vivo* e *in-vitro*. O primeiro envolve o uso de aparelhos profissionais de diagnóstico por imagem. É altamente preciso, porém pode ser inconveniente devido à sua disponibilidade limitada em áreas com baixo desenvolvimento ou recursos financeiros limitados. Além disso, seu custo financeiro pode ser proibitivo para muitas famílias. Por outro lado, o segundo tipo de diagnóstico se baseia em três biomarcadores, sinal de fala, sinal de marcha e sinal de expressão facial, sendo esse último o mais promissor, pois não requer sensores profissionais como o sinal da marcha, tornando-o mais acessível e prático [Huang et al. 2023].

A análise facial tem avançado significativamente em diversas áreas, incluindo o uso de sistemas de Reconhecimento de Expressões Faciais (FER) para identificar padrões associados a condições médicas. Esses sistemas geralmente operam em três etapas principais: detecção de pontos de referência facial, extração de características e classificação. Algoritmos de aprendizado de máquina ou *deep learning* são frequentemente empregados, utilizando grandes conjuntos de dados de imagens de indivíduos saudáveis e de pacientes [Pegolo et al. 2022]. No contexto da doença de Parkinson, a análise de imagens faciais é relevante devido às alterações motoras sutis e à diminuição da expressividade facial, assim como observado na comparação feita na Figura 1. As imagens capturam essas mudanças, como a rigidez facial e a redução na amplitude dos movimentos faciais. Esses modelos são capazes de reconhecer e classificar essas diferenças, contribuindo para a detecção precoce da doença e fornecendo uma ferramenta complementar no monitoramento clínico.

A metodologia deste projeto propõe avaliar e comparar diferentes arquiteturas de Redes Neurais Convolucionais (CNNs) aplicadas na classificação de imagens faciais de indivíduos saudáveis e pacientes com Doença de Parkinson. Serão implementados e testados modelos de aprendizagem profunda, com ênfase nas CNNs, utilizando a técnica de validação cruzada por grupos (*GroupKFold*) para garantir que os mesmos indivíduos não sejam utilizados simultaneamente nos conjuntos de treinamento e teste, assegurando a ro-

bustez dos resultados. O foco está em identificar as arquiteturas mais eficazes na distinção entre as classes, com o objetivo de melhorar a acurácia e eficiência da classificação, contribuindo para a detecção precoce da doença e potencializando intervenções clínicas apropriadas.

O trabalho está organizado da seguinte maneira: a Seção 2 aborda a literatura existente sobre o tema. A Seção 3 descreve a metodologia aplicada no desenvolvimento do projeto, incluindo informações sobre as bases de imagens médicas selecionadas, além dos métodos e parâmetros de avaliação propostos. As Seções 4 e 5 apresentam os resultados obtidos e discutem o progresso do projeto, bem como as perspectivas futuras.

## 2. Trabalhos Relacionados

Ao longo dos últimos anos, diversos pesquisadores se empenharam em compreender e propor soluções para o reconhecimento da DP através da análise da expressão facial, fazendo uso de recursos computacionais. Esta seção visa elucidar, bem como resumir, alguns dos trabalhos que integram esse campo da ciência.

O estudo feito por Rajnoha et al. [Rajnoha et al. 2018] investiga a possibilidade de identificar hipomímia, um sintoma característico da DP, através da análise estática da face. Para isso, foram incluídos 50 pacientes com DP e 50 controles saudáveis pareados por idade e gênero. Utilizando métodos de reconhecimento facial e classificadores convencionais, como florestas aleatórias e *XGBoost*, os pesquisadores automatizaram a identificação da hipomímia. O algoritmo de árvore de decisão alcançou a maior acurácia, atingindo 67,33%. Embora os resultados indicassem que a análise estática da face poderia auxiliar no diagnóstico de hipomímia na DP, ela não se mostrou tão precisa quanto abordagens baseadas em vídeos.

No estudo de Hou et al. [Hou et al. 2021], os pesquisadores desenvolveram um modelo de inteligência artificial (*IA*) para identificar características faciais associadas à doença de Parkinson em vídeos 2D, sem a necessidade de marcações. Eles coletaram 140 vídeos de expressões faciais de pacientes com DP e controles, e treinaram o modelo para reconhecer características específicas da doença. O modelo alcançou uma alta precisão na identificação das características da DP, com taxas de acerto de até 83% usando características geométricas e 86% com características de textura. Esses resultados indicaram que a *IA* pode ser uma ferramenta promissora para auxiliar no diagnóstico precoce da DP, complementando a avaliação de neurologistas. Além disso, o uso de vídeos para análise facial ofereceu a possibilidade de monitoramento remoto dos pacientes, o que foi particularmente útil durante a pandemia de COVID-19, quando visitas presenciais ao médico eram limitadas.

Pegolo et al. [Pegolo et al. 2022] objetivam desenvolver um índice quantitativo para hipomímia, capaz de distinguir indivíduos saudáveis de patológicos e de ser aplicado na classificação de emoções. Para isso, foi implementado um algoritmo de rastreamento facial baseado no *Facial Action Coding System*, introduzindo uma nova métrica chamada Índice de Mobilidade Facial (IMF), que considera distâncias entre pares de feições geométricas. O estudo inclui uma comparação entre controles saudáveis e pacientes com DP. Os resultados indicam que o IMF pode quantificar o comprometimento na DP e ser útil na classificação de emoções. Observou-se diferenças estatisticamente significativas para todas as emoções quando as distâncias foram consideradas e para fe-

licidade e raiva quando o IMF foi levado em conta. As melhores classificações foram obtidas com os algoritmos *Random Forest* e *kNN*, conforme avaliado pela métrica *AUC*.

A obra de Jiang et al. [Jiang et al. 2022] explora a utilização da fenotipagem facial para o pré-diagnóstico médico, uma técnica recentemente bem-sucedida para avaliar pré-clinicamente diversas doenças genéticas raras. A biometria facial está ligada a problemas médicos ou genéticos importantes. O estudo visa estender essa abordagem para a DP, propondo uma estrutura de pré-diagnóstico facial direcionada à privacidade, especialmente voltada para a *Inteligência Artificial das Coisas (AIoT)*, para analisar o tratamento de Estimulação Cerebral Profunda (ECP) em pacientes com DP. A estrutura proposta inclui uma nova abordagem de preservação da privacidade baseada em borda para realizar diagnósticos faciais profundos de forma privada, integrando-se a um esquema de comunicação multipartidária seguro. Os experimentos realizados com dados faciais de pacientes com DP demonstram pela primeira vez a viabilidade de usar padrões faciais para avaliar a diferença facial em pacientes tratados com ECP. Além disso, a implementação de uma estrutura segura de pré-diagnóstico facial confirma sua precisão, revelando seu potencial como um serviço confiável para classificar a gravidade da DP em pacientes.

Huang et al. [Huang et al. 2023] abordam a importância do diagnóstico precoce da DP devido ao seu impacto significativo na qualidade de vida dos pacientes e seus familiares. O diagnóstico *in vitro* baseado em expressões faciais tem recebido crescente atenção por sua capacidade de distinguir características específicas, como o "rosto mascarado" associado à DP, e sua acessibilidade. No entanto, as abordagens atuais enfrentam limitações devido à escassez de dados de treinamento e à fraca capacidade preditiva dos modelos. Para superar esses desafios, é proposto um novo método guiado por expressão facial que se baseia no aumento de dados de treinamento e na utilização de redes neurais profundas. Este método envolve a síntese de imagens de expressões faciais virtuais e a seleção de imagens de alta qualidade para treinamento do modelo. Extensivos experimentos foram realizados, incluindo a colaboração com um hospital para criar um conjunto de dados faciais da DP, demonstrando a eficácia do método proposto para o diagnóstico de DP.

A partir das obras analisadas, fica evidente que o desenvolvimento e o uso de técnicas computacionais avançadas são essenciais para aprimorar o reconhecimento da Doença de Parkinson (DP) por meio da análise facial. Estudos anteriores demonstram a viabilidade de identificar sintomas como a hipomímia e destacam a eficácia de modelos de inteligência artificial para detectar características faciais da DP. No entanto, este trabalho se diferencia ao utilizar modelos de deep learning pré-treinados e validação cruzada por grupos, garantindo maior robustez e precisão na classificação. Além disso, enquanto outros estudos exploram análises dinâmicas ou métodos tradicionais, esta abordagem foca na detecção a partir de imagens estáticas, simplificando a coleta de dados e ampliando a aplicabilidade clínica. Essas inovações proporcionam uma base sólida para a proposta atual, contribuindo significativamente para avanços na detecção precoce e no monitoramento da DP.

### **3. Materiais e Métodos**

Nesta seção, são detalhados os recursos e o método empregados no desenvolvimento deste trabalho. O objetivo é classificar imagens faciais para diferenciar indivíduos saudáveis de

pacientes com Doença de Parkinson (DP).

Para isso, foram selecionadas, ajustadas e avaliadas oito Redes Neurais Convolucionais. A implementação das *CNNs* foi realizada utilizando a linguagem de programação *Python*, juntamente com as bibliotecas de aprendizado profundo *Keras* e *TensorFlow*.

Nos tópicos seguintes, abordamos detalhadamente o método proposto, a base de imagens selecionada e suas características, técnicas de aumento de dados, as *CNNs* selecionadas e as métricas de avaliação utilizadas para desempenho.

### 3.1. Base de Imagens

O dataset é composto por um conjunto de imagens capturadas na Clínica de Neurologia do Hospital Universitário da Universidade Federal do Piauí. A coleta das imagens foi realizada após a devida autorização do comitê de ética<sup>1</sup>. Foram utilizadas fotografias de 21 pacientes com a doença e 21 sem a doença. Durante a coleta foi pedido a esses pacientes que expressassem as seguintes emoções: felicidade, surpresa, medo, nojo, raiva e tristeza, sendo cerca de 3 imagens por emoção, totalizando, portanto, 340 imagens de pacientes com Doença de Parkinson (DP) e 358 imagens de indivíduos saudáveis. Entre os pacientes com DP, 13 eram homens, com média de idade de 61,4 anos e tempo médio de diagnóstico de 5,3 anos.

Inicialmente, o dataset estava organizado por emoções, mas neste trabalho foi decidido utilizar uma classificação binária, unindo as imagens de todas as emoções. Essa abordagem visa simplificar a análise e focar na diferenciação entre indivíduos com e sem a doença, independentemente das expressões emocionais capturadas nas imagens.

### 3.2. Aumento de Dados

O aumento de dados consiste em um conjunto de técnicas destinadas a melhorar a qualidade e a diversidade dos dados, com o objetivo de aprimorar o desempenho dos modelos de aprendizado por meio da criação de novas amostras a partir das já existentes. De acordo com Mumuni e Mumuni [Mumuni and Mumuni 2022], conjuntos de dados maiores, mais diversificados e representativos aumentam a eficácia dos modelos de aprendizado profundo em dados não observados.

Para evitar problemas como o *overfitting*, as técnicas de aumento de dados foram aplicadas exclusivamente ao subconjunto de treino. As imagens de entrada passaram por um processo de pré-processamento, onde os valores de pixel foram normalizados, e em seguida, as imagens foram submetidas a várias transformações, incluindo rotação de até 20 graus, deslocamento horizontal e vertical de até 20% da largura e altura, cisalhamento de até 20%, variação de zoom de até 20%, e inversão horizontal para espelhamento das imagens. Em casos onde as transformações criaram regiões vazias, as áreas resultantes foram preenchidas com os valores dos pixels mais próximos (*nearest neighbor interpolation*), garantindo a integridade das imagens transformadas.

### 3.3. Redes Neurais Convolucionais

Redes Neurais Convolucionais (*CNNs*) são uma classe de algoritmos de aprendizado profundo projetadas para processar e analisar dados. Elas operam utilizando camadas con-

---

<sup>1</sup>Certificado de Apresentação de Apreciação Ética: 68623423.6.0000.8050.

volucionais que aplicam filtros sobre a entrada, extraindo características importantes enquanto preservam relações espaciais. Em problemas de classificação, as *CNNs* aprendem automaticamente a identificar padrões e características distintas nas imagens, permitindo a categorização precisa dos objetos nelas contidos [Tajbakhsh et al. 2016].

Esses modelos podem ser utilizados na classificação de imagens médicas e auxiliar no diagnóstico mais preciso de doenças, como a Doença de Parkinson (*DP*). Nesse projeto, foram avaliadas oito *CNNs* pré-treinadas utilizando a base de dados ImageNet, que contém milhões de imagens rotuladas distribuídas em milhares de categorias. As *CNNs* selecionadas foram *MobileNet*, *ResNet50*, *DenseNet121*, *MobileNetV2*, *ResNet101*, *VGG19*, *DenseNet201* e *EfficientNetB3*.

### 3.4. Transferência de Aprendizagem

A transferência de aprendizado é uma abordagem de aprendizado de máquina que envolve a adaptação de um modelo pré-treinado em uma grande base de dados, como o ImageNet, para uma nova tarefa, aproveitando o conhecimento previamente adquirido.

O modelo proposto utiliza a técnica de *Shallow Fine-Tuning (SFT)*, na qual, conforme mencionado por [Tajbakhsh et al. 2016], apenas as camadas finais — geralmente as camadas totalmente conectadas — são ajustadas, enquanto as camadas convolucionais iniciais, que realizam a extração de características genéricas, permanecem intactas, assim como mostra a Figura 2. Esse método requer menos dados e recursos computacionais, permitindo que o novo modelo alcance um desempenho superior em um tempo de treinamento reduzido. O ajuste fino não só acelera a convergência e melhora a capacidade de generalização do modelo, mas também diminui a probabilidade de *overfitting*.

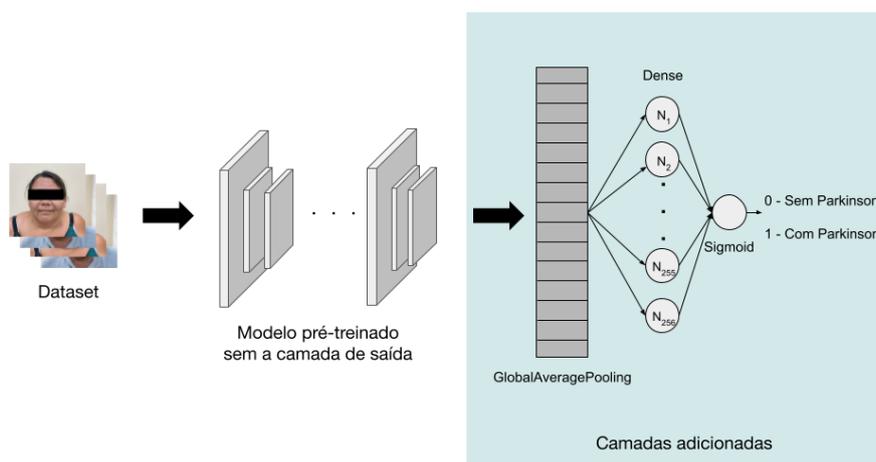


Figura 2. Arquitetura de Rede Neural com Ajuste Fino de Modelos Pré-Treinados.

### 3.5. Validação Cruzada

O projeto implementou a validação cruzada por grupos (*GroupKFold*) para avaliar o desempenho dos modelos de aprendizado de máquina, segmentando o conjunto de dados em 5 partes ou "folds". Em cada iteração, um *fold* era utilizado como conjunto de teste, enquanto os outros eram empregados para treinamento, repetindo o processo cinco vezes para garantir que todos os *folds* fossem testados. A técnica *GroupKFold* foi utilizada

para garantir que imagens do mesmo paciente não fossem incluídas simultaneamente nos conjuntos de treino, teste e validação, evitando assim o vazamento de dados e garantindo uma avaliação mais robusta. Essa metodologia é crucial para manter a integridade dos resultados e fornecer uma estimativa confiável da capacidade de generalização do modelo, especialmente em conjuntos de dados com possível desequilíbrio entre as classes, mitigando vieses e identificando possíveis casos de *overfitting*.

### 3.6. Métricas de Avaliação

As métricas empregadas para avaliar o desempenho do modelo incluíram acurácia, precisão, *recall*, *F1-Score* e o índice *Kappa* [Powers 2020]. A acurácia refere-se à proporção de previsões corretas em relação ao total de previsões realizadas, fornecendo uma visão geral da eficácia do modelo. A precisão mede a fração de verdadeiros positivos em relação a todas as previsões positivas, refletindo a exatidão das previsões positivas do modelo. O *recall* (ou sensibilidade) representa a proporção de verdadeiros positivos em relação ao total de casos realmente positivos, demonstrando a capacidade do modelo de identificar corretamente os casos positivos. O *F1-Score* é a média harmônica entre precisão e *recall*, proporcionando um equilíbrio entre ambas e sendo particularmente útil em situações de desequilíbrio entre as classes.

O índice kappa é recomendado como uma medida apropriada de exatidão, pois pode representar adequadamente a matriz de confusão. Este índice leva em consideração todos os elementos da matriz de confusão, não apenas aqueles na diagonal principal, ao contrário da precisão da classificação global, que considera apenas os elementos diagonais principais. O coeficiente kappa é calculado pela Equação 1:

$$K = \frac{(\text{observado} - \text{esperado})}{1 - \text{esperado}} \times 100 \quad (1)$$

De acordo com [Landis and Koch 1977], o valor de K pode ser interpretado da seguinte maneira:  $K \leq 20\%$ : Ruim;  $20\% < K \leq 40\%$ : Regular;  $40\% < K \leq 60\%$ : Bom;  $60\% < K \leq 80\%$ : Muito Bom; e  $K > 80\%$ : Excelente.

## 4. Resultados e Discussão

Fazendo uso de técnicas como aumento de dados, que incluem zoom, deslocamento e cisalhamento, além de transferência de aprendizado com redes pré-treinadas e *Shallow Fine Tuning (SFT)*, foram aplicadas técnicas de validação cruzada por grupos, como o k-fold Cross Validation com  $k = 5$  folds.

Na validação cruzada por grupos (*GroupKFold*) aplicada durante os experimentos, os dados foram divididos de forma a garantir que imagens do mesmo paciente não fossem incluídas simultaneamente nos conjuntos de treino, teste e validação. O conjunto de dados foi dividido em cinco partes ( $k = 5$ ), e em cada iteração, quatro dessas partes foram usadas para o treinamento e a quinta para teste. Dentro do conjunto de treinamento, uma divisão adicional foi feita para criar subconjuntos de treino e validação interna, utilizando uma proporção de 80% para treino e 20% para validação. Essa abordagem assegurou uma avaliação mais robusta e imparcial do desempenho dos modelos.

Durante os experimentos, cada modelo foi treinado por um total de 100 épocas, utilizando um *batch size* de 32. O otimizador escolhido foi o *Adam*, com uma taxa de

aprendizado inicial de 0,001. Esta taxa foi selecionada com base em testes preliminares, onde se observou uma convergência estável e eficiente. Para evitar o sobreajuste, foi implementado o *early stopping* com 10 épocas de paciência, monitorando a acurácia de validação e restaurando os melhores pesos do modelo. Oito *CNNs* foram treinadas, e seus resultados foram avaliados para determinar o desempenho de cada modelo.

Os experimentos e resultados apresentados nesta seção foram obtidos utilizando um computador equipado com um processador Intel(R) Core(TM) i7 de 9ª geração, com clock de 3,00 GHz, 16 GB de RAM e uma placa gráfica Nvidia RTX 2060 de 6 GB, composta por 1.920 núcleos.

**Tabela 1. Resultados dos modelos avaliados.**

Modelo	Acur. (%)	Prec. (%)	Rec. (%)	F1 (%)	Kappa (%)
<b>DenseNet-201</b>	<b>92,77 ± 7,09</b>	<b>93,93 ± 8,21</b>	<b>91,16 ± 8,30</b>	<b>92,48 ± 8,03</b>	<b>85,33 ± 14,40</b>
VGG-19	90,87 ± 6,00	97,69 ± 4,62	84,07 ± 9,91	90,13 ± 6,72	81,76 ± 12,03
MobileNet	89,18 ± 7,91	86,50 ± 10,09	94,97 ± 7,62	90,15 ± 7,16	77,54 ± 16,91
EfficientNet-B3	88,90 ± 4,82	94,89 ± 10,22	85,27 ± 9,91	88,82 ± 4,25	77,89 ± 9,50
ResNet101	88,10 ± 4,82	94,64 ± 6,59	82,41 ± 10,99	87,37 ± 5,48	76,43 ± 9,21
MobileNetV2	86,81 ± 9,42	97,66 ± 2,20	76,06 ± 19,52	84,14 ± 12,20	73,64 ± 18,98
ResNet50	86,98 ± 11,02	97,98 ± 2,75	77,00 ± 21,95	84,17 ± 14,14	74,39 ± 21,05
DenseNet-121	86,10 ± 7,63	91,14 ± 9,28	80,17 ± 13,48	84,80 ± 10,00	72,06 ± 15,80

Os melhores resultados obtidos durante o treinamento estão em negrito.

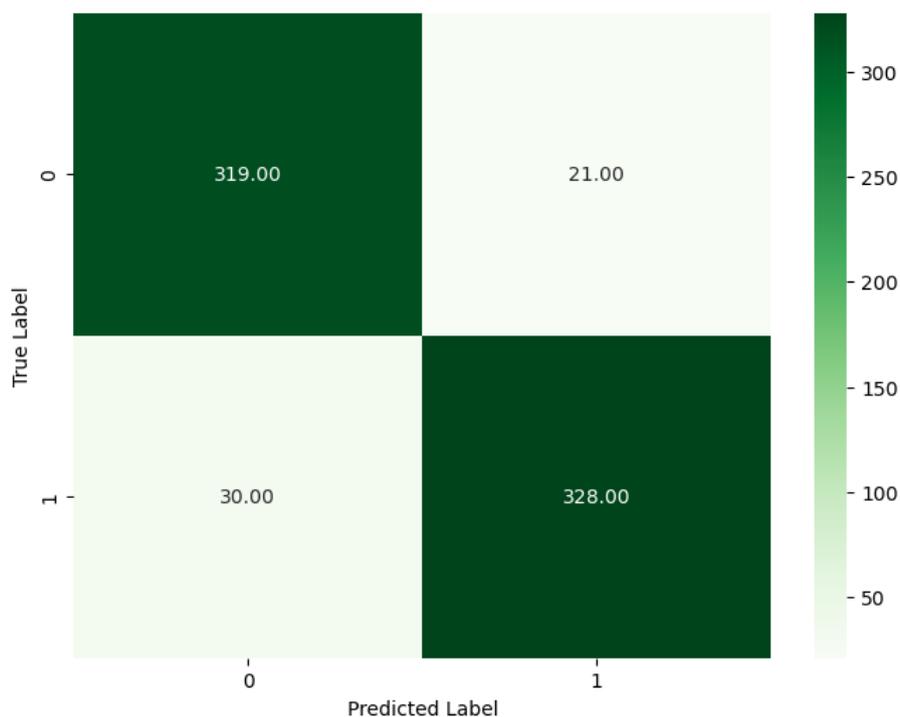
Entre os modelos avaliados na Tabela 1, o *DenseNet-201* se destaca como o líder em desempenho geral. Com uma acurácia média de 92,77% e um desvio padrão de  $\pm 7,09\%$ , o *DenseNet-201* demonstra um equilíbrio consistente entre as métricas avaliadas. Além disso, apresenta uma precisão de 93,93%, um recall de 91,16%, e um F1 score de 92,48%, mostrando sua habilidade em identificar corretamente tanto as instâncias positivas quanto as negativas. O coeficiente Kappa de 85,33% reforça a robustez do modelo na classificação das imagens faciais. Esses resultados indicam que o *DenseNet-201* é a escolha mais adequada entre os modelos testados para aplicações que exigem alta precisão e confiabilidade na identificação da Doença de Parkinson.

Na Figura 3, a matriz de confusão do modelo *DenseNet-201* mostra um desempenho sólido na classificação das imagens. A baixa quantidade de falsos positivos e negativos sugere que o *DenseNet-201* está generalizando bem e realizando previsões precisas sobre os dados de teste.

## 5. Conclusão

Os experimentos realizados neste estudo demonstraram a eficácia de técnicas avançadas de aprendizado de máquina na classificação de imagens faciais de indivíduos com ou sem Doença de Parkinson (DP), destacando o impacto positivo do aumento de dados e da transferência de aprendizado. A utilização de validação cruzada por grupos (*GroupKFold*) garantiu uma avaliação robusta dos modelos, resultando em uma melhor generalização e desempenho.

Entre os modelos analisados, o *DenseNet-201* se destacou como o mais eficiente, alcançando uma acurácia média de 92,77% e apresentando resultados sólidos em métricas como precisão, *recall* e *F1-score*. Esses resultados confirmam a capacidade do modelo em realizar classificações precisas e o posicionam como uma opção preferencial para



**Figura 3. Matriz de Confusão Final Combinada das 5 rodadas do K-fold do modelo *DenseNet-201*.**

aplicações que exigem alta performance. A matriz de confusão associada ao modelo reforça a consistência dos resultados, evidenciando uma baixa taxa de falsos positivos e negativos.

É possível afirmar que resultados encorajadores foram obtidos, e que o aprimoramento futuro de algumas técnicas pode torná-los ainda melhores. Entre as estratégias para isso, destaca-se a ampliação da base de imagens utilizada, o que tornaria o treinamento mais abrangente. Além disso, o uso de técnicas de refinamento, como o *Deep Fine Tuning* (DFT), pode ser vantajoso, pois permite ajustar todas as camadas das CNNs à base de imagens de entrada, e não apenas as camadas finais. A exploração de outras arquiteturas de redes neurais também pode contribuir significativamente para avanços futuros.

Portanto, conclui-se que a combinação das técnicas adotadas, a exploração de outras redes, e a perspectiva de um aumento nos datasets estabelecem um caminho promissor para futuras investigações na área de reconhecimento de padrões e classificação de imagens faciais de pessoas com a Doença de Parkinson.

## Referências

- Hou, X., Zhang, Y., Wang, Y., Wang, X., Zhao, J., Zhu, X., and Su, J. (2021). A markerless 2d video, facial feature recognition-based, artificial intelligence model to assist with screening for parkinson disease: Development and usability study. *Journal of Medical Internet Research*, 23(11):e29554.
- Huang, W., Zhou, Y., Cheung, Y.-m., Zhang, P., Zha, Y., and Pang, M. (2023). Facial expression guided diagnosis of parkinson's disease via high-quality data augmentation. *IEEE Transactions on Multimedia*, 25:7037–7050.

- Jiang, R., Chazot, P., Pavese, N., Crookes, D., Bouridane, A., and Celebi, M. (2022). Private facial prediagnosis as an edge service for parkinson's dbs treatment valuation. *IEEE Journal of Biomedical and Health Informatics*, 26(6):2703–2713.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Lee, H. M. and Koh, S. B. (2015). Many faces of parkinson's disease: Non-motor symptoms of parkinson's disease. *Journal of Movement Disorders*, 8(2):92–97.
- Mumuni, A. and Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16:100258.
- Pegolo, E., Volpe, D., Cucca, A., Ricciardi, L., and Sawacha, Z. (2022). Quantitative evaluation of hypomimia in parkinson's disease: A face tracking approach. *Sensors*, 22(4).
- Powers, D. M. W. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.
- 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.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312.