



# Modelagem inteligente para detecção dos níveis de retinopatia diabética

Intelligent modeling for detection of diabetic retinopathy levels in medical images

Guilherme Fernandes da Silva<sup>1</sup>, Ricardo Menezes Salgado<sup>2</sup>

<sup>1</sup>Graduando em Ciência da Computação na Universidade Federal de Alfenas (Unifal-MG).

ORCID: [0000-0002-0965-3820](https://orcid.org/0000-0002-0965-3820)

**Email:** [guilherme.fernandes@sou.unifal-mg.edu.br](mailto:guilherme.fernandes@sou.unifal-mg.edu.br)

<sup>2</sup>Doutor em Engenharia Elétrica.

Departamento de Ciência da Computação

Universidade Federal de Alfenas.

ORCID: [0000-0002-0989-6259](https://orcid.org/0000-0002-0989-6259)

**Email:** [ricardo.salgado@unifal-mg.edu.br](mailto:ricardo.salgado@unifal-mg.edu.br)

**Correspondência:** Av. Jovino Fernandes Sales, 2600 - Prédio C - Sala C 301 J. Bairro Santa Clara - Alfenas/MG CEP: 37133-840

**Copyright:** Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

**Conflito de interesses:** os autores declaram que não há conflito de interesses.

## Como citar este artigo

Nomes dos autores no formato exigido (Exemplo - nome: Luiz Roberto de Oliveira / Formato exigido: Oliveira, LR de). Título do trabalho. Revista de Saúde Digital e Tecnologias Educacionais. [online], volume X, n. X. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, dezembro de 2020, p. XX-XX. Disponível em: <http://periodicos.ufc.br/resdite/index>. Acesso em "dia/mês/ano".

**Data de recebimento do artigo:** XX/XX/XXX

**Data de aprovação do artigo:** XX/XX/XXX

**Data de publicação:** XX/XX/XXX

## Resumo

A Retinopatia Diabética é uma das principais patologias que causam a perda parcial e total da visão quando não identificadas nos estágios iniciais. A literatura médica mostra que o tratamento precoce adequado evita progressão da doença e reduz as chances de perda de visão.

Os estágios desta patologia são identificados por médicos oftalmologistas, porém é comum em estágio inicial ocorrer divergência de opiniões entre eles para identificação do diagnóstico. Além disso, comumente há falta de especialistas para realizar o diagnóstico, já que muitas pessoas desenvolvem este acometimento, e a falta do diagnóstico precoce pode causar danos irreversíveis ao paciente.

O objetivo deste trabalho foi aplicar técnicas em diferentes modelos nas bases de dados propostas originadas das duas competições sobre Retinopatia Diabética no

Kaggle, utilizando *Deep Learning* e a técnica de transferência de aprendizagem combinando com métodos que auxiliam no desempenho do modelo na tarefa de classificação. Os resultados experimentais mostram que fatores como a base de dados e a arquitetura escolhida podem influenciar no desempenho do modelo na etapa de testes. Modelos treinados na maior base de dados foram mais generalistas e obtiveram um desempenho melhor que os modelos treinados nas outras bases propostas. E entre as quatro arquiteturas utilizadas (**InceptionResNetV2**, **NasNet**, **Xception** e **EfficientNetB0**), é possível concluir que o **EfficientNetB0** foi o menos generalista e que obteve o pior desempenho comparado às outras arquiteturas.

**Palavras-chave:** Retinopatia Diabética, Aprendizado Profundo, Redes Neurais Convolucionais, Transferência de Aprendizagem.

### **Abstract**

*Diabetic Retinopathy is one of the main pathologies that cause partial and total loss of vision when not identified in the early stages. Medical literature shows that adequate early treatment prevents disease progression and reduces the chances of vision loss.*

*The stages of this pathology are identified by ophthalmologists, but it is*

*common in the initial stage for there to be divergence of opinions between them to identify the diagnosis. In addition, there is usually a lack of specialists to diagnose this population, since many people develop this condition, and the lack of an early diagnosis can cause non-reversible damage to the patient.*

*The objective of this work was to apply techniques in different models in the proposed databases originating from the two competitions on Diabetic Retinopathy in Kaggle, using Deep Learning and the transfer of learning technique combined with methods that help in the performance of the model in the classification task. The experimental results show that factors such as the database and the chosen architecture can influence the performance of the model in the testing phase. Models trained on the largest database were more generalist and performed better than models trained on the other proposed databases. And among the four architectures used (**InceptionResNetV2**, **NasNet**, **Xception** and **EfficientNetB0**), it is possible to conclude that **EfficientNetB0** was the least generalist and that it obtained the worst performance compared to the other architectures.*

**Keywords:** Diabetic retinopathy, Deep Learning, Convolutional Neural Networks, Transfer Learning

## 1. Introdução

Diabetes mellitus (DM) é um dos principais problemas atuais para a população mundial. A Federação Internacional de Diabetes estimou em 2017 que 425 milhões de pessoas, ou seja 8,8% de toda população mundial entre 20 e 79 anos vivem com este problema, e que em 2045 seja superior a 629 milhões de pessoas afetadas<sup>1</sup>. A maioria dos casos tem predominância em países desenvolvidos, devido a fatores como transição epidemiológica, urbanização e estilo de vida mais sedentário quando comparado com países subdesenvolvidos<sup>2,3</sup>.

De toda a população com DM, acredita-se que mais de um terço apresenta sinais iniciais de Retinopatia Diabética (RD), e um terço sofre de sintomas da patologia, considerada a principal causa da perda parcial ou total da visão entre pessoas entre 20 e 74 anos<sup>4</sup>.

A Retinopatia Diabética gera a perda parcial ou total da visão devido ao dano à retina por pequenas lesões nos vasos sanguíneos, que possuem a função do transporte de nutrientes para a retina. Com estes microaneurismas nos vasos, ocorre o extravasamento desses nutrientes na retina, causando a perda de visão total ou parcial<sup>5</sup>. A RD pode ser classificada em cinco estágios, sendo:

- Sem Retinopatia Diabética aparente: quando não ocorre alterações;
- Retinopatia diabética não proliferativa (leve): quando ocorrem pequenos inchaços nos vasos sanguíneos da retina, ou seja, micro aneurismas;
- Retinopatia diabética não proliferativa moderada: alguns vasos sanguíneos são bloqueados, com maior número de micro aneurismas que o estágio anterior.
- Retinopatia diabética não proliferativa grave: Ocorrem hemorragias, manchas e sangramento das veias, devido à falta de oxigênio pelo bloqueio dos vasos sanguíneos;
- Retinopatia diabética proliferativa: Ocorre a presença de novos vasos sanguíneos, mas com algum tipo de anomalia e mais frágeis, gerando extravasamento de resíduos que penetram na retina, geralmente causando a perda da visão.

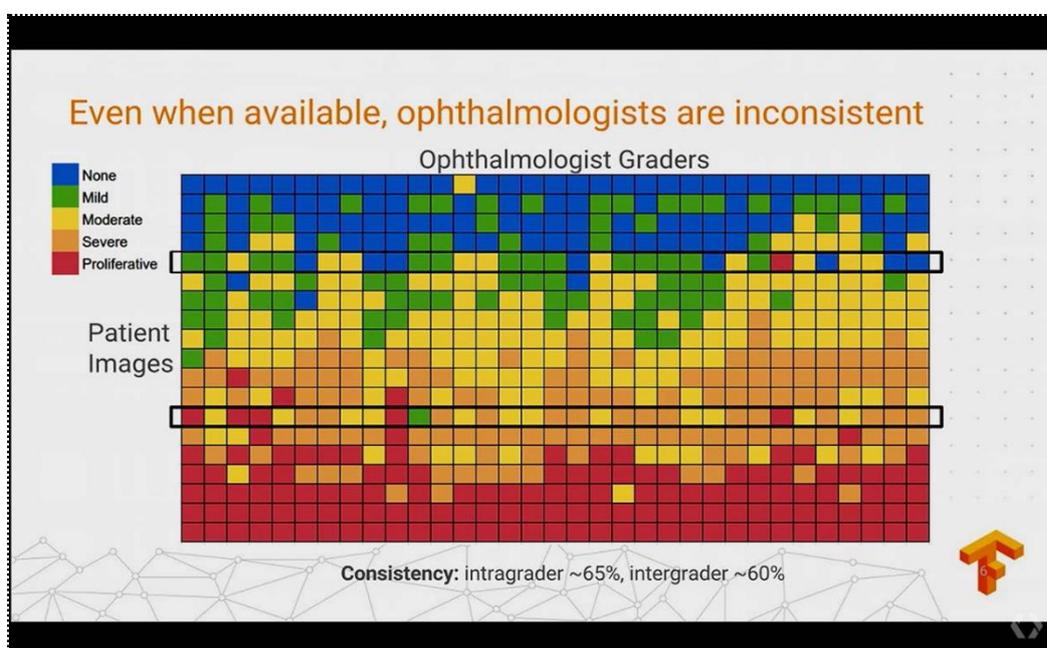
O estágio inicial da RD não apresenta sinais de alertas evidentes, sendo assim um desafio diagnosticá-lo através do olhar médico, sendo este essencial para redução das chances de o paciente ter sua visão afetada<sup>6</sup>.

A idade média e o número de pessoas com diabetes devem aumentar, o que se torna um problema devido à insuficiência de oftalmologistas para diagnosticar a RD. O

diagnóstico tardio aumenta as chances de irreversibilidade e perda de visão do paciente<sup>7</sup>. A discordância entre médicos é um problema importante, pois a doença não apresenta sinais clássicos, resultando em tratamentos inadequados<sup>8</sup>.

A Figura 1 mostra a variância no diagnóstico entre médicos oftalmologistas na identificação dos estágios da RD, especialmente nos estágios leves, moderados e graves, onde cada cor representa uma categoria diferente da doença e cada linha uma imagem de fundo de olho de determinado paciente.

**Figura 1:** “Mesmo quando disponíveis, os oftalmologistas são inconsistentes”



**Fonte:** SHISHI, Duan. **First TensorFlow Developer Conference: TensorFlow's development over the years.** ProgrammerSought. Disponível em:

<https://www.programmersought.com/article/3066123972/>. Acesso em: 22 fev. 2023.

A Inteligência Artificial (IA) tem se mostrado uma ferramenta valiosa na área da saúde, especialmente na detecção de doenças. O uso de modelos de *Deep Learning* para análises de imagens médicas, como radiografias e tomografias, tem se mostrado eficiente na detecção de doenças como câncer ou problemas cardíacos<sup>9</sup>.

Na área oftalmológica, a IA tem se mostrado promissora na detecção de problemas como a Retinopatia Diabética. Estudos têm mostrado que modelos de *Deep Learning* podem ser treinados para reconhecer padrões de imagens de fundo de olho, e assim ajudar no diagnóstico preciso e precoce<sup>10,11</sup>.

Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) são uma subclasse de redes neurais profundas que foram especialmente projetadas para lidar com dados que possuem uma estrutura especial, como as imagens. As CNNs têm mostrado excelentes resultados na resolução de diversos problemas de Visão Computacional, incluindo a detecção de doenças.

Na detecção de doenças, as CNNs são amplamente utilizadas para análise de imagens médicas, como radiografias e tomografias, e para a classificação de diagnósticos. As CNNs aprendem a extrair características relevantes dos dados, e a combinar essas características para realizar a tarefa de classificação. Essas características são aprendidas por meio de filtros e camadas que agem como uma forma de seleção dessas características.

É importante destacar que apesar das CNNs terem se mostrado uma ferramenta poderosa no auxílio de doenças e com tendência de crescer no campo da Medicina, é preciso destacar que essas redes neurais ainda precisam ser avaliadas e validadas rigorosamente antes de serem utilizadas em ambientes clínicos, a fim de garantir a segurança no diagnóstico do paciente.

A técnica de transferência de aprendizagem é relevante na Inteligência Artificial, pois permite que modelos apliquem conhecimentos adquiridos em uma tarefa para solucionar problemas semelhantes, economizando tempo e recursos computacionais<sup>12</sup>. A transferência de aprendizado também pode levar a resultados mais precisos e generalizáveis.

Diante disso, o objetivo deste artigo foi desenvolver uma solução que classifique as imagens com os cinco níveis da Retinopatia Diabética, que busca inicialmente avaliar o desempenho de diferentes modelos em diferentes bases de dados e determinar qual obteve o melhor desempenho, aplicando técnicas de pré-processamento, divisão de dados e transferência de aprendizagem. Serão testadas as técnicas de transferência de aprendizagem nas arquiteturas **Xception**, **InceptionResNetV2**, **NasNet** e **EfficientNetB0** da biblioteca Keras, treinadas nas bases de dados das competições Diabetic Retinopathy Detection que ocorreu em 2015 sendo patrocinada pelo California Healthcare Foundation e da competição APTOS 2019 Blindness Detection que ocorreu em 2019 sendo patrocinada pelo Asia Pacific Tele-Ophthalmology Society (APTOS) disponíveis na plataforma Kaggle<sup>13,14</sup>.

O presente artigo encontra-se organizado da seguinte forma: Na seção 1 contém a introdução e motivação do desenvolvimento deste trabalho. A seção 2 contextualiza o

conceito de arquitetura de uma rede neural convolucional e a técnica de transferência de aprendizagem. A seção 3 aborda a metodologia para a realização dos experimentos. Na seção 4 é realizada uma discussão sobre os resultados obtidos por cada modelo em suas respectivas bases de dados bem como apresenta-se os resultados obtidos a partir da aplicação dos modelos treinados na base de dados de teste. A seção 5 finaliza o trabalho apresentando as discussões finais e as considerações.

## 2. Fundamentação Teórica

Redes Neurais Convolucionais (CNN) surgiram no final dos anos 80 como variante das redes neurais feedforward, mas tornaram-se amplamente usadas a partir de 2010 devido ao aumento de dados e poder de processamento<sup>15</sup>. As CNNs consistem em camadas de neurônios interconectados com a capacidade de realizar operações de convolução, essa operação é responsável por extrair recursos ou características importantes das imagens de entrada, basicamente é aplicada uma operação em cada parte da imagem de entrada usando um conjunto de filtros. Cada filtro é uma pequena matriz de pesos que é deslizada sobre a imagem de entrada, produzindo um único valor de saída. Esse processo é repetido para cada posição de entrada, produzindo um mapa de características da imagem de entrada que são importantes para o problema<sup>16</sup>.

As CNNs podem ser treinadas para realizar tarefas de classificação de imagens, analisando as imagens através de várias camadas de processamento, cada uma aprendendo características cada vez mais complexas<sup>17</sup>. Uma das vantagens é a habilidade de aprender características diretamente dos dados, sem a necessidade de pré-processamento manual<sup>18</sup>.

### 2.1 Transferência de Aprendizagem

A técnica de transferência de aprendizagem é usada para acelerar o treinamento dos modelos. Isso consiste em usar um modelo pré-treinado em uma tarefa similar como base para outro problema relacionado, transferindo o conhecimento adquirido para a nova tarefa, economizando tempo e recursos computacionais. Para problemas na área de oftalmologia, as arquiteturas utilizadas geralmente fazem parte da biblioteca KERAS e incluem pesos pré-treinados para a previsão e extração de recursos.

A vantagem de se utilizar as arquiteturas disponíveis no KERAS é que já foram testadas e avaliadas em grandes conjuntos de dados mostrando-se eficazes para a tarefa de classificação de imagem, sendo desenvolvidas por especialistas em Inteligência Artificial. Uma outra vantagem é que muitas dessas arquiteturas são pré-treinadas em grandes conjuntos de dados, o que significa que a rede pode ser usada como ponto de partida para uma tarefa de classificação específica.

## 2.2 Ensemble

Redes neurais são métodos não lineares que podem ser adaptadas com base na quantidade de dados disponíveis. Isso é devido às redes neurais aprenderem através de algoritmos de treinamento estocástico, produzindo um conjunto de pesos diferentes para cada vez que são treinadas. A técnica de Ensemble é utilizada para melhorar a taxa de acerto e normalizar os pesos, combinando previsões de múltiplos modelos treinados pela técnica de validação cruzada, sendo a previsão final a classe que obtém mais votos. No Ensemble, múltiplos modelos são treinados com diferentes conjuntos de dados ou com diferentes configurações e, em seguida, são combinados para produzir uma previsão final. O objetivo dessa combinação de múltiplos modelos é que o resultado final seja mais preciso e robusto do que o resultado produzido por cada modelo individual.

## 3. Métodos

Para resolver o problema de Retinopatia Diabética, é utilizada uma modelagem computacional que contempla as etapas de pré-processamento das imagens, seleção de bases para treinamento, aplicação da técnica de Data Augmentation, validação cruzada, duas etapas de treinamento e a aplicação da técnica de Ensemble nos modelos.

Para o pré-processamento das imagens, foi utilizado a linguagem Python e a biblioteca OpenCV para cortar as imagens e isolar a imagem colorida, eliminando ao máximo o fundo preto. As bases de dados de treinamento foram selecionadas a partir de duas competições referentes a Retinopatia Diabética na plataforma Kaggle, surgindo três bases com tamanhos e distribuição de classes diferentes. A etapa de Data Augmentation foi utilizada para aumentar o tamanho dos conjuntos de dados criados, utilizando a técnica de distorção dos dados. Já a técnica de validação cruzada (K-fold) foi utilizada para avaliar

a performance do modelo em diferentes conjuntos de dados. E a etapa de treinamento foi dividida em duas fases, pré-treinamento e fine tuning e após o treinamento foi aplicado a técnica de ensemble nos modelos treinados em cada fold.

### 3.1 Pré-Processamento

O treinamento dos modelos foi feito com as imagens pré-processadas, pelo fato das imagens originais terem variações na posição do olho na imagem, então foi feito o corte para isolar a imagem colorida e remover o máximo possível do fundo preto. A Figuras 2 mostra um antes e depois da aplicação do método de corte, visando isolar a imagem do globo ocular e reduzir ao mínimo os ruídos externos.

**Figura 2:** Comparação Antes e Depois do pré-processamento.



**Fonte:** Elaboração dos autores.

### 3.2 Base de dados de treinamento

Esse artigo aborda o problema da classificação do nível da retinopatia diabética em duas competições na plataforma Kaggle. O objetivo foi comparar qual arquitetura obteve um melhor desempenho, e analisar qual base de dados teve um melhor desempenho no treinamento dessas arquiteturas, com o objetivo de justificar a hipótese se a arquitetura e a base de dados podem influenciar no desempenho de um modelo.

Diante disso, surgiram três bases de dados a partir das duas competições propostas no Kaggle. A primeira base de dados refere-se à junção das duas bases originais, totalizando 18.697 imagens para o treinamento. A segunda base de dados é referente ao

balanceamento das duas bases de dados originais, ou seja, o objetivo desta base é manter o número de imagens de todos os diagnósticos próximos, sendo distribuída da forma que exista por volta de 1.000 imagens para cada classe, totalizando 5.010 imagens de treinamento. Por fim, a última base de dados foi somente com os dados de treinamento da competição que ocorreu no ano de 2019, totalizando 3.662 imagens.

Durante o processo de treinamento dos modelos, 70% das imagens de cada base foram utilizadas para a etapa de treinamento e os outros 30% das imagens foram utilizadas para a etapa de validação de cada modelo, essa divisão é adotada como um padrão pelo fato de manter uma quantidade suficiente de dados para treinamento do modelo e ainda ter dados suficientes para avaliar o desempenho na etapa de treinamento.

A Tabela 1 mostra a distribuição da quantidade de imagens de cada classe referente às bases de dados criadas da origem das duas competições sobre retinopatia diabética presente no Kaggle.

**Tabela 1:** Distribuição das bases de dados

Base de dados	Número total de imagens
Competição 2015 e Competição 2019	18.698 imagens
Balanceada	5.010 imagens
Competição 2019	3.662 imagens

Fonte: Elaboração dos autores.

### 3.3 Base de dados de teste

Este artigo possui uma base de dados para teste que é referente a Private Label da competição realizada no ano de 2015. Basicamente, um conjunto Private Label é referente ao conjunto oficial que foi usado como benchmark de testes para pontuação em uma competição. Portanto, temos um total de 42.670 imagens disponibilizadas como Private Label na competição Diabetic Retinopathy Detection.

### 3.4 Data Augmentation

A técnica de aumento de dados (Data Augmentation) foi utilizada para expandir o tamanho do conjunto de treinamento com base nas próprias imagens da base de dados.

Com uma quantidade significativa de imagens para a etapa de treinamento, os modelos tendem a resultar em uma melhor capacidade de ajustar aos dados e uma maior taxa de generalização. Essa técnica aumenta artificialmente o tamanho do conjunto de dados de treinamento por distorção ou superamostragem. A distorção de dados transforma as imagens existentes de forma que seu rótulo seja preservado. Isso abrange ampliações como transformações geométricas e de cores, apagamento aleatório e transferência de estilo neural<sup>19</sup>. Para este problema, foi aplicado os seguintes filtros em 20% das imagens de treinamento:

- *vertical\_flip* = Gira a imagem verticalmente
- *horizontal\_flip* = Gira a imagem horizontalmente

### 3.5 Validação Cruzada

A técnica de Validação Cruzada (K-fold) é uma maneira de avaliar a performance de um modelo em um conjunto de dados, basicamente é realizada a divisão do conjunto de dados em “k” partes, onde o modelo é treinado em k-1 folds e testado no fold restante, repetindo o processo em k vezes. A performance é avaliada como a média das k iterações. Essa técnica é útil para avaliar a performance do modelo em grandes conjuntos de dados e identificar problemas de overfitting<sup>20</sup>.

Neste trabalho, foi aplicada a técnica com k=5 nos dados de treinamento utilizando a biblioteca de aprendizagem de máquina Scikit-learn, onde é fornecido um método para realizar esse procedimento. Já a escolha do k é devido ser amplamente aceito na literatura, pois existe uma boa relação entre a quantidade de dados usados para treinar e validar o modelo, e demonstra ser uma boa escolha devido a capacidade de fornecer uma estimativa da performance do modelo com relativamente pouco viés<sup>21</sup>.

### 3.6 Modelos

Para garantir uma comparação justa entre os modelos treinados em diferentes bases de dados, foram utilizados os mesmos parâmetros e metodologias de treinamento em todas elas. O processo de treinamento foi dividido nas etapas de pré-treinamento e Fine

Tuning. Dessa forma, pode-se comparar os resultados de maneira justa e identificar qual modelo e em qual base de dados obteve o melhor desempenho.

O treinamento de uma rede neural começa geralmente com a inicialização dos pesos com valores aleatórios, mas neste caso, os pesos foram inicializados com valores copiados de um modelo já treinado na base de dados ImageNet com mais de 14 milhões de imagens distribuídas em 1.000 classes. Isso permite iniciar o treinamento de um ponto melhor e reduz o tempo de treinamento, garantindo que o modelo já tenha aprendido recursos relevantes para o problema<sup>22</sup>. A Tabela 2 descreve os parâmetros utilizados no pré-treinamento, onde os modelos foram treinados por 5 épocas, com uma taxa de aprendizagem de 0,001, utilizando a entropia cruzada para atribuição dos valores entre as classes e o otimizador Adam para medir o desempenho de acerto da predição.

Fine Tuning é um processo de treinamento que ocorre após a etapa de pré-treinamento que permite utilizar redes neurais pré-treinadas para reconhecer novas classes. Isso é feito removendo a camada final da rede e substituindo-a por uma nova camada conectada relevante para o problema. É esperado que a rede pré-treinada tenha aprendido alguns filtros úteis para a classificação, enquanto a nova camada possui pesos aleatórios<sup>23</sup>, com isso a taxa de aprendizagem é 10 vezes menor do que a etapa de pré-treinamento pelo fato de que na etapa de pré-treinamento, o modelo é treinado para reconhecer recursos gerais, enquanto na etapa de Fine-Tuning o modelo deve ser ajustado para reconhecer recursos específicos do problema. Normalmente a quantidade de épocas na etapa de Fine-Tuning é maior que a etapa de Pré-Treinamento, permitindo que a rede refine ainda mais os recursos aprendidos na etapa anterior para a tarefa específica, aumentando assim a eficácia da rede. Os parâmetros descritos na Tabela 2, indicam que o modelo foi novamente treinado por 10 épocas, com uma taxa de aprendizagem 10 vezes menor que a etapa de pré-treinamento, ou seja 0,0001, utilizando a entropia cruzada para atribuição dos valores entre as classes e o otimizador Adam para medir o desempenho de acerto da predição.

**Tabela 2:** Parâmetros para treinamento

Parâmetros	Pré-Treinamento	Fine-Tuning
Dimensão imagem	512 x 512	512 x 512
Epochs	5	10

<b>Learning Rate</b>	0,001	0,0001
<b>Loss</b>	Categorical crossentropy	Categorical crossentropy
<b>Optimizer</b>	Adam	Adam

Fonte: Elaboração dos autores.

## 4. Resultados

Todas as arquiteturas foram treinadas e testadas na plataforma Kaggle, utilizando a GPU disponível gratuitamente e a biblioteca KERAS, que permitiu acessar as arquiteturas e aplicar a técnica de transferência de aprendizagem.

### 4.1 Experimento com o conjunto de Private Label

Para este experimento, todos os modelos foram treinados nas bases de treinamento propostas, com o objeto de realizar uma comparação justa sobre qual modelo e qual base de dados obteve um melhor resultado.

Os resultados dos modelos de classificação do nível de retinopatia diabética são apresentados na Tabela 3, sendo divididos pelos modelos treinados nas arquiteturas propostas na maior base de dados, base de dados balanceada e menor base de dados.

Para cada modelo são fornecidos dois valores de acurácia: "Acc Validação" representa a acurácia do modelo na etapa de validação e "Acc Private Label" representa a acurácia do modelo em imagens de rótulos privados, na qual o modelo não teve contato na etapa de treinamento.

Os modelos treinados na base de dados com maior número de imagens que apresentaram melhores resultados na etapa de validação são das arquiteturas **InceptionResNetV2** e **NasNet**, enquanto a arquitetura **EfficientNetB0** teve o pior desempenho. Analisando a etapa de treinamento, consta que além da arquitetura **EfficientNetB0** ter uma acurácia menor em relação às outras, foi a menos generalista. Os modelos treinados na arquitetura **EfficientNetB0** não aprenderam as características discriminativas da classe 1 (Leve) e 3 (Severa).

Já os modelos treinados na base de dados balanceada, é possível observar que novamente as arquiteturas **InceptionResNetV2** e **NasNet** obtiveram um melhor resultado, enquanto a arquitetura **Xception** teve o pior desempenho. Analisando a etapa de treinamento dessas arquiteturas, todos os modelos conseguiram generalizar bem, ou seja, todos os modelos conseguiram aprender algumas características referentes às classes. Já na etapa de testes com os dados não vistos, os modelos tiveram bastante dificuldade na generalização do modelo, obtendo um desempenho bem abaixo da etapa de validação. Esse tipo de resultado é considerado como um overfitting, que ocorre quando o modelo aprende demais com os dados de treinamento e não consegue generalizar para novos dados.

Por fim, os modelos treinados na base de dados com menor número de imagens, as arquiteturas **Xception** e **NasNet** apresentaram um melhor desempenho, no entanto todas as arquiteturas desempenharam bem na etapa de validação. Analisando profundamente a etapa de treinamento, a única arquitetura que realmente generalizou e conseguiu apreender características discriminativas de todas as classes foi a **Xception**, já as outras arquiteturas tiveram bastante dificuldade em aprender as características das classes 1 (Leve), 3 (Severa) e 4 (Proliferativa). Na etapa de testes com os dados não-vistos, ocorreu o mesmo problema de overfitting da base anterior.

**Tabela 3:** Resultado dos modelos treinados nas bases de dados propostas.

Bases de Dados		InceptionResNetV2	Xception	NasNet	EfficientNetB0
Maior Base de Dados	Acc Validação	84,25%	84,62%	84,79%	79,12%
	Acc Private Label	82,04%	75,68%	82,25%	73,59%
Base de Dados Balanceada	Acc Validação	80,20%	80,50%	68,80%	73,20%
	Acc Private Label	39,90%	54,56%	53,13%	27,56%
Menor Base de Dados	Acc Validação	86,02%	89,04%	86,48%	86,04%
	Acc Private Label	52,01%	63,40%	64,27%	51,17%

**Fonte:** Elaboração dos autores.

A arquitetura **NasNet** se mostrou superior em relação às outras arquiteturas, comparando os resultados sobre os dados não-vistos na etapa de testes, os modelos

treinados sobre a arquitetura **NasNet** na maior base de dados aprenderam mais características e foram mais generalistas do que os outros modelos da arquitetura **NasNet** treinados nas outras bases de dados propostas.

## 4.2 Discussão

Os resultados mostram que houve uma diferença significativa na acurácia entre os modelos treinados nas bases de dados propostas. A arquitetura **NasNet** treinada na maior base de dados foi a que obteve a maior acurácia na etapa de testes, com 82,25% de acertos, enquanto a arquitetura **EfficientNetB0** treinada na base de dados balanceada teve o menor percentual de acertos na etapa de testes, com 27,56%. O comportamento e a variação do desempenho na etapa de testes podem ser justificados como a base de dados utilizada para o treinamento dos modelos. Analisando as bases de dados de treinamento propostas, percebe-se que os modelos treinados na base de dados com maior número de imagens tiveram um desempenho superior aos modelos treinados nas outras bases de dados.

Comparando as arquiteturas utilizadas neste trabalho, a arquitetura **NasNet** foi a arquitetura que obteve melhores resultados e a que possui maior profundidade (786 camadas) e a arquitetura **EfficientNetB0** foi a arquitetura que obteve os piores desempenhos e a que possui menor profundidade (25 camadas), a profundidade da rede afeta diretamente a capacidade do modelo em aprender representações hierárquicas e detalhes das imagens, quanto mais profunda a rede, mais capacidade ela terá para aprender recursos complexos, e quanto menos profunda a rede, menos capacidade ela terá para aprender recursos complexos.

Isso sugere que a escolha da arquitetura e da base de dados podem influenciar os resultados, já que os modelos treinados na arquitetura **EfficientNetB0** tiveram o pior desempenho em todas as bases de dados. Além disso, a técnica de Ensemble não apresentou uma melhora significativa em relação ao desempenho individual de cada modelo, em alguns casos aumentando apenas 0,1% a acurácia, portanto essa técnica não foi eficaz neste trabalho. A análise dos resultados também indica que as arquiteturas treinadas com a base de dados maior têm um desempenho superior quando comparadas às treinadas nas outras bases. Pelo fato de ser uma base de dados mais ampla, os modelos treinados com esses dados são mais generalistas e são capazes de aprender e reconhecer

padrões e características mais sutis e complexas devido à grande quantidade de imagens disponíveis no treinamento.

## 5. Conclusão

Este artigo apresentou uma abordagem de modelagem de aprendizado profundo para identificação dos níveis de retinopatia diabética em imagens médicas. A retinopatia diabética é uma complicação ocular decorrente da diabetes que pode levar a perda parcial ou total da visão se não for detectada e tratada precocemente. A modelagem utilizou técnicas de pré-processamento, aplicação da transferência de aprendizagem nas arquiteturas **InceptionResNetV2**, **Xception**, **NasNet** e **EfficientNetB0** e o uso da técnica de Ensemble, visando avaliar o impacto das técnicas, das arquiteturas utilizadas e da base de dados para treinamento no desempenho dos modelos.

Os resultados indicaram uma diferença significativa no desempenho geral dos modelos, com os modelos treinados com base na arquitetura **NasNet** alcançando o melhor desempenho em todas as bases de dados. Além disso, foi observado uma variação no desempenho dos modelos treinados nas diferentes bases de dados propostas, com os modelos treinados na maior base de dados superando os demais. Já a técnica Ensemble não apresentou nenhuma melhora ou impacto positivo no desempenho dos modelos.

Em geral, os resultados obtidos neste estudo sugerem que a utilização de arquiteturas pré-treinadas pode ser uma opção viável e eficiente para a classificação do nível de retinopatia diabética, e que ainda há espaço para aprimoramento desses modelos através da incorporação de novas técnicas.

Conclui-se que a escolha da arquitetura e da base de dados de treinamento podem influenciar significativamente nos resultados da análise e classificação de imagens médicas. A discriminação de sutis diferenças entre imagens semelhantes ainda é um desafio, bem como a variabilidade da qualidade das imagens, a falta de dados rotulados e a necessidade de modelos capazes de lidar com imagens de alta resolução. Estratégias para melhorar o desempenho dos modelos incluem o uso de técnicas de transferência de aprendizagem e a incorporação de conhecimento médico. No entanto, é importante lembrar que a tecnologia deve ser usada com cuidado e como parte de um processo de diagnóstico mais amplo que inclui avaliação clínica e julgamento médico.

## 6. Referências

1. CHO, Nam H. et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. **Diabetes research and clinical practice**, v. 138, p. 271-281, 2018.
2. BOYLE, James P. et al. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. **Population health metrics**, v. 8, n. 1, p. 1-12, 2010.
3. SAEEDI, Pouya et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. **Diabetes research and clinical practice**, v. 157, p. 107843, 2019.
4. LEE, Ryan; WONG, Tien Y.; SABANAYAGAM, Charumathi. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. **Eye and vision**, v. 2, n. 1, p. 1-25, 2015.
5. GUPTA, Ankita; CHHIKARA, Rita. Diabetic retinopathy: Present and past. **Procedia computer science**, v. 132, p. 1432-1440, 2018.
6. MOOKIAH, Muthu Rama Krishnan et al. Computer-aided diagnosis of diabetic retinopathy: A review. **Computers in biology and medicine**, v. 43, n. 12, p. 2136-2155, 2013.
7. ZHENG, Yingfeng; HE, Mingguang; CONGDON, Nathan. The worldwide epidemic of diabetic retinopathy. **Indian journal of ophthalmology**, v. 60, n. 5, p. 428, 2012.
8. ABRÀMOFF, Michael D.; GARVIN, Mona K.; SONKA, Milan. Retinal imaging and image analysis. **IEEE reviews in biomedical engineering**, v. 3, p. 169-208, 2010.
9. ESTEVA, Andre et al. Dermatologist-level classification of skin cancer with deep neural networks. **nature**, v. 542, n. 7639, p. 115-118, 2017.
10. KERMANY, Daniel S. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. **cell**, v. 172, n. 5, p. 1122-1131. e9, 2018.
11. GARGEYA, Rishab; LENG, Theodore. Automated identification of diabetic retinopathy using deep learning. **Ophthalmology**, v. 124, n. 7, p. 962-969, 2017.
12. XUE, Dan et al. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. **IEEE Access**, v. 8, p. 104603-104618, 2020.
13. Kaggle. Diabetic Retinopathy Detection [Competição online]. Disponível em: <https://www.kaggle.com/competitions/diabetic-retinopathy-detection>.
14. Kaggle. Aptos 2019 Blindness Detection [Competição online]. Disponível em: <https://www.kaggle.com/competitions/aptos2019-blindness-detection>.
15. HINTON, Geoffrey E. et al. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.
16. LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.
17. SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
18. HE, Kaiming et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 770-778.
19. SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. **Journal of big data**, v. 6, n. 1, p. 1-48, 2019.
20. NEMATZADEH, Zahra; IBRAHIM, Roliana; SELAMAT, Ali. Comparative studies on breast cancer classifications with k-fold cross validations using machine learning techniques. In: **2015 10th Asian control conference (ASCC)**. IEEE, 2015. p. 1-6.

21. CUNHA, João Paulo Zanola. Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos. 2019. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019. doi:10.11606/D.45.2019.tde-26082019-220647.
22. DA COSTA, Nator Junior Carvalho et al. Classification of X-ray images for detection of childhood pneumonia using pre-trained neural networks. **Revista Brasileira de Computação Aplicada**, v. 12, n. 3, p. 132-141, 2020.
23. SWATI, Zar Nawab Khan et al. Brain tumor classification for MR images using transfer learning and fine-tuning. **Computerized Medical Imaging and Graphics**, v. 75, p. 34-46, 2019.