

# Uma Avaliação de Arquiteturas de Aprendizado Profundo para a Classificação de Úlceras do Pé Diabético

Francisco Santos<sup>1</sup>, Rodrigo Veras<sup>1</sup>, Elineide Santos<sup>1</sup>, Maila Lima Claro<sup>1</sup>,  
Luís Henrique Vogado<sup>1</sup>, Márcia Ito<sup>2</sup> e Andrea Bianchi<sup>3</sup>

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

<sup>2</sup>Faculdade de Tecnologia de São Paulo – São Paulo, Brasil

<sup>3</sup>Universidade Federal de Ouro Preto – Minas Gerais, Brasil

{fsantos, rveras}@ufpi.edu.br, andrea@ufop.edu.br,

{elineide.silva.inf, lhvogado, claromaila, marciaito2000}@gmail.com

**Abstract.** *A complication caused by diabetes mellitus is the appearance of wounds in the feet called diabetic foot ulcers. Late treatment can lead to the onset of infection or ischemia of the ulcer, which, in an advanced stage, can cause the amputation of the lower limbs. In this work, a comparison of the performance of several pre-trained deep learning architectures in the classification of images of diabetic foot ulcers was carried out. Our assessment considered four scenery, Three binaries - Healthy vs Ulcers; Healthy vs Ischemia; Healthy vs Infection - and one multiclass - Healthy vs Ulcer vs Ischemia vs Infection vs Infection and Ischemia. The results showed that our proposal could classify such images since the Kappa index reached values considered “Excellent” in the tests carried out. However, for the multiclass problem, it is still necessary to improve the use of these techniques.*

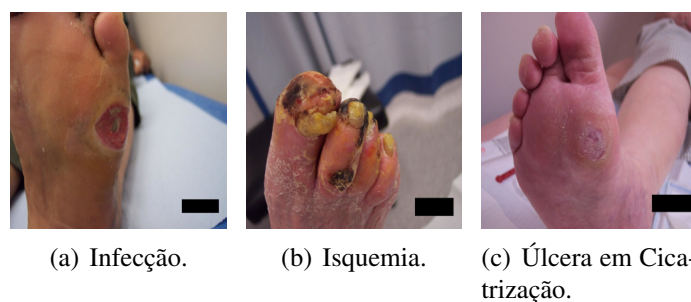
**Resumo.** *Uma complicação causada pelo diabetes mellitus é o aparecimento de feridas situadas na região dos pés denominadas úlceras do pé diabético. O tratamento tardio pode acarretar o surgimento de infecção ou isquemia da úlcera que, em estado avançado, pode ocasionar a amputação dos membros inferiores. Neste trabalho foi realizado um comparativo do desempenho de diversas arquiteturas de aprendizado profundo pré-treinadas na classificação de imagens de úlceras do pé diabético. Nossa avaliação levou em consideração quatro cenários, três binários - Saudável vs Úlceras; Saudável vs Isquemia; Saudável vs Infecção - e um multiclasse - Saudável vs Úlcera vs Isquemia vs Infecção vs Infecção e Isquemia. Os resultados alcançados demonstram que a nossa proposta consegue classificar tais imagens, visto que, nos testes realizados o índice Kappa atingiu valores considerados “Excelentes”. Contudo, para o problema multiclasse ainda é necessário aprimorar o uso dessas técnicas.*

## 1. Introdução

Uma complicação causada pelo diabetes *mellitus* é o aparecimento de feridas situadas na região dos pés denominadas úlceras do pé diabético. Este tipo de úlcera, normalmente, pode ser resultado de um trauma (pisar sobre objeto penetrante), estresse mecânico repetitivo (caminhar descalço ou com calçado inadequado) ou estresse mecânico continuamente aplicada (pressão não percebida estendida) [Cavanagh et al. 2005]. Em estágio

mais avançado as úlceras podem ocasionar a amputação dos membros infectados. Um estudo descrito por Goyal et al. [Goyal et al. 2020] aponta que cerca de 56% das úlceras do pé diabético são infectadas e 20% destas acarretam a amputação do membro atingido. Essa doença resulta em altos custos e elevada taxa de mortalidade, sendo, a identificação adequada das úlceras importante para determinar fatores que preveem o progresso de cura e o risco de amputação em cada paciente.

O tratamento tardio pode acarretar lesões graves classificadas em infecção ou isquemia. A infecção se caracteriza pela presença de inflamação ou purulência, também pode ocorrer o aumento da vermelhidão dentro ou ao redor da úlcera. A Figura 1(a) apresenta um exemplo de úlcera com infecção. A isquemia consiste na distribuição inadequada de sangue pela região e, visualmente, pode ser indicada pela presença de má reperfunção no pé ou nos dedos pretos gangrenados<sup>1</sup>, conforme mostra a Figura 1(b). Seu diagnóstico ocorre geralmente palpando os pulsos do pé e medindo a pressão sanguínea no pé e nos dedos dos pés [Goyal et al. 2020]. Algumas úlceras após o tratamento adequado alcançam um estado de cicatrização e ficam similares a pele saudável como mostra a Figura 1(c). Neste trabalho, para efeito de avaliação, as úlceras com estas características são classificadas como pele saudável.



**Figura 1. Exemplos de imagens de pés com úlcera com isquemia, úlcera com infecção e úlcera em processo de cura.**

Atualmente, a avaliação do pé diabético é baseada nos sinais e sintomas da diabetes [Solís-Sánchez et al. 2016], o que inclui uma avaliação visual da úlcera. Esta avaliação é subjetiva, pois depende dos critérios adotados pelo médico, que pode não ter conhecimento de todas as variações visuais que as úlceras podem apresentar. Um sistema computacional pode auxiliar o médico nessa avaliação por meio da combinação de processamento digital de imagens e inteligência artificial. Dessa maneira, é possível classificar a ferida e acompanhar a evolução da área cicatrizada de forma automatizada.

Diferentes redes foram utilizadas em trabalhos anteriores, por isso, com intuito de avaliar qual rede tem melhor desempenho, este trabalho tem como objetivo realizar um estudo comparativo da classificação de imagens de úlceras do pé diabético utilizando as Redes Neurais Convolucionais (*Convolutional Neural Networks - CNNs*): VGG-16, VGG-19, InceptionV3, ResNet50 e DenseNet201. Na avaliação das CNNs foi utilizada uma base de dados pública contendo 2.295 imagens pertencentes a cinco classes, são elas: pele saudável (PS), úlcera (UC), úlcera com isquemia (UIS), úlcera com infecção (UIN) e úlcera com isquemia e infecção (UII). Na experimentação, foram avaliados quatro cenários considerando a classificação entre pele saudável e úlceras, pele saudável e

<sup>1</sup>Morte de tecidos afetados.

úlceras com isquemia, pele saudável e úlceras com infecção e por fim, entre as cinco classes. Após a classificação, é feita uma análise comparativa com intuito de identificar qual arquitetura de rede obteve melhor desempenho na classificação.

Este artigo está dividido da seguinte forma: na Seção 2 é apresentado o estado da arte, onde são citados os trabalhos mais relevantes sobre o tema. Na Seção 3 é apresentada a base de imagens, as técnicas aplicadas e as métricas de avaliação adotadas no trabalho. Os resultados e discussões são apresentados na Seção 4 e as conclusões e trabalhos futuros na Seção 5.

## 2. Estado da Arte

Diversas metodologias foram desenvolvidas nos últimos anos com a finalidade de oferecer soluções automatizadas e precisas para o diagnóstico de patologias de pé diabético. As principais metodologias propostas são explanadas a seguir.

No trabalho de Veredas et al. [Veredas et al. 2010], os autores propuseram uma metodologia de identificação automática de tecidos usando uma técnica que combina redes neurais e classificadores Bayesianos. As 113 imagens utilizadas foram inicialmente segmentadas usando uma técnica de crescimento de regiões. No entanto, esse procedimento não é trivial, uma vez que as imagens de úlceras apresentam grande variedade de forma, cor e textura. Em seguida as características foram extraídas e, então, realizado o treinamento das redes neurais supervisionadas para diferenciar tecidos necrosados dos demais tecidos. A saída desse sistema é usada pelos classificadores bayesianos para classificar os tecidos em cinco tipos diferentes de tecido: pele, cicatrização, granulação, descamação e necrose. Essa abordagem obteve uma acurácia de 91,50% em seus resultados.

Em Wang et al. [Wang et al. 2016], os autores propuseram uma metodologia que segmenta a região lesionada em imagens de úlceras de pé capturadas por uma caixa padronizada utilizando o classificador *Support Vector Machine* (SVM). Essa caixa possui iluminação e distância controladas. As 100 imagens capturadas são segmentadas usando a técnica de geração de *superpixels* denominada *Simple Linear Iterative Clustering* (SLIC) [Achanta et al. 2012]. Após a etapa de segmentação, são extraídas características de cor e textura que são utilizadas no treinamento do classificador. Diante disso, a metodologia proposta consiste em dois estágios para a classificação, sendo o primeiro estágio composto por um conjunto de SVM binários que executam diferentes testes nos conjuntos de imagens, onde os resultados são coletados e usados no próximo estágio. No segundo estágio, apenas os *superpixels* classificados como lesão são classificados usando um novo SVM binário. As imagens resultantes são processadas aplicando operações morfológicas seguida de detecção de regiões conectadas e um método de reclassificação baseado na abordagem *Conditional Random Field* (CRF) [He et al. 2004] para regiões não feridas marcadas incorretamente e preencher regiões de feridas não marcadas. Esse método obteve uma sensibilidade de 73,30% e uma especificidade de 94,60%.

Alzubaidi et al. [Alzubaidi et al. 2020] propuseram uma rede convolucional profunda denominada DFU-QUTNet, que classifica as imagens em pele normal ou anormal com baixo custo computacional. Esta arquitetura é capaz de aumentar a largura da rede sem a necessidade de ampliar sua profundidade, dessa maneira, ocorre o aumento da aprendizagem dos dados para realizar a classificação. Os autores avaliaram dois cenários,

o primeiro consiste na utilização da DFU-QUTNet como classificador, já o segundo é composto pela extração das características utilizando a DFU-QUTNet pré-treinada e os classificadores SVM e *K-Nearest Neighbors* (KNN) na etapa de predição. Utilizando a DFU-QUTNet+SVM foi obtida precisão de 95,40% na classificação de 754 imagens de pés de pacientes com úlcera do pé diabético e pele saudável.

Goyal et al. [Goyal et al. 2020] propuseram uma metodologia de classificação de imagens. Inicialmente, a metodologia adotada realiza um aumento natural de dados com a finalidade de melhorar o desempenho do algoritmo de identificação das lesões, visto que as imagens de pé diabéticos ocupam uma região bem reduzida em relação ao total. Após a definição da região de interesse, foram extraídas informações de cor (RGB e CI-ELAB) e textura (*Local binary pattern* - LBP, *Histogram of Oriented Gradien* - HOG) das 1.459 imagens adotadas. Também, foram gerados *superpixels* aplicando a técnica SLIC [Achanta et al. 2012] para segmentar as imagens e facilitar a extração de características. Outro detalhe empregado foi o modelo de rede *Ensemble*, que combina as redes neurais InceptionV3, ResNet50 e InceptionResNetV2 com o classificador SVM [Wang et al. 2016] para realizar as predições. Os resultados obtidos por essa abordagem foram acurácia de 90,00% na classificação de imagens em isquemia e de 73,00% na classificação em imagens de infecção para a base de imagens adotada.

### 3. Materiais e Métodos

Objetivando realizar uma análise comparativa das classificações de imagens de úlceras do pé diabético foram avaliadas as Redes Neurais Convolucionais pré-treinadas: VGG16, VGG19, InceptionV3, ResNet50 e DenseNet201. Para implementação das CNNs foi utilizada linguagem de programação Python e a biblioteca de rede neural de código aberto Keras.

#### 3.1. Base de Imagens

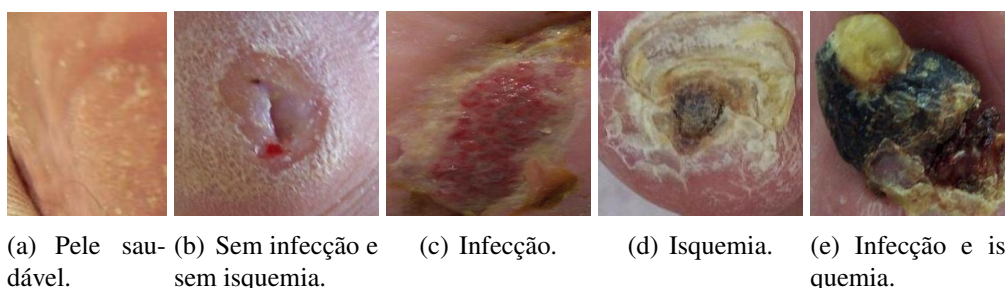
O conjunto de dados utilizado neste estudo contém imagens de úlceras do pé diabético capturadas com três tipos de câmeras (Kodak DX4530, Nikon D3300 e Nikon COOLPIX P100) após o desbridamento<sup>2</sup> durante cinco anos no *Lancashire Teaching Hospitals*. Goyal et al. [Goyal et al. 2020] realizaram a identificação de casos de isquemia e de infecção nestas imagens produzindo a base de dados pública *Diabetic Foot Ulcer* (DFU)<sup>3</sup>, contando com apoio de dois profissionais de saúde (consultor médico e médico especialista em pé diabético) para a produção do *ground truth*. A decisão final foi tomada pelo médico mais experiente quando houve desacordo entre os profissionais.

Neste trabalho, para ficar mais próximo da realidade do serviço do médico, as imagens da base DFU foram separadas em cinco classes: (1) imagens de pele saudável (Figura 2a); (2) imagens de úlceras sem infecção e sem isquemia (Figura 2b); (3) imagens de úlceras com infecção (Figura 2c); (4) imagens de úlceras com isquemia (Figura 2d) e (5) imagens de úlceras com infecção e isquemia (Figura 2e).

A Tabela 1 indica o número de imagens por classe do conjunto de 2.295 imagens analisadas. Para realizar esta separação, as imagens positivas de infecção foram comparadas com as imagens positivas de isquemia, possibilitando a identificação de imagens

<sup>2</sup>Remoção de tecidos necróticos e desvitalizados.

<sup>3</sup>Disponível em: <http://www2.docm.mmu.ac.uk/STAFF/M.Yap/dataset.php>.



**Figura 2. Exemplos de imagens de pele saudáveis e de úlceras de pé diabético da base DFU.**

comuns aos dois grupos<sup>4</sup> e de imagens pertencentes apenas a infecção ou a isquemia. A obtenção de imagens de úlceras sem infecção e sem isquemia foram obtidas com a comparação das imagens negativas de isquemia com as imagens negativas de infecção.

**Tabela 1. Número de imagens por classes do banco de dados DFU.**

Classes	Número de Imagens
Pele Saudável (PS)	641
Úlcera sem Infecção e sem Isquemia (UC)	640
Úlcera com Isquemia (UIS)	26
Úlcera com Infecção (UIN)	779
Úlcera com Infecção e Isquemia (UII)	209
<b>Total</b>	<b>2.295</b>

A base DFU original contém imagens oriundas do aumento natural de dados. Para uma úlcera com infecção existem três imagens e para uma úlcera com isquemia existem sete imagens. Nos nossos experimentos, não utilizamos as imagens extras provenientes do aumento de dados original. Contudo, na etapa de refinamento da rede foi aplicado um conjunto de operações de aumento de dados conforme descrito no final da Subseção 3.2. Para a avaliação (teste) utilizamos apenas as imagens originais.

### 3.2. Aumento de dados

Diante dos milhares de parâmetros existentes nas arquiteturas de CNNs, o custo computacional para o treinamento é alto, necessitando de uma grande quantidade de dados. Diante disso, trabalhos encontrados na literatura utilizaram a técnica de aumento de dados com o intuito de solucionar o problema de classificação de imagens do pé diabético. Esta técnica consiste na criação de um novo conjunto de imagens utilizando transformações em imagens, como: rotação, translação e inversão. O aumento de dados tem como principais objetivos a redução no *overfitting* da rede e; consequentemente, o aumento na capacidade de generalização da arquitetura [Perez and Wang 2017].

Neste trabalho, durante a etapa de refinamento da CNN, foi utilizado o aumento de dados proveniente do Keras. Este aumento consiste na geração de novas amostras durante o treinamento, aplicando operações aleatórias em *batches*. As técnicas utilizadas foram: rotação, translação vertical e horizontal, cisalhamento e zoom. Também foi aplicada a inversão horizontal e vertical, assim como o preenchimento por reflexão foi aplicada para substituir *pixels* em preto resultantes das técnicas de rotação e de translação. Por fim, foi realizada a normalização dos *pixels* da imagem de entrada para 0 e 1.

<sup>4</sup>Úlceras com infecção e isquemia.

### 3.3. Arquiteturas de CNNs

Foram exploradas e avaliadas as arquiteturas de CNNs projetadas para o *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) [Russakovsky et al. 2015]. De acordo com Kornblith et al. [Kornblith et al. 2019], quanto melhor o desempenho da arquitetura no conjunto de dados ImageNet, melhor será a transferência para outros conjuntos de dados de imagens naturais. Além disso, outro fator determinante para a seleção de arquiteturas no estado da arte foi o desempenho obtido em trabalhos da literatura, como em Vogado et al. [Vogado et al. 2021].

As redes sequenciais como VGG-16 e VGG-19 [Simonyan and Zisserman 2014] facilitam mudanças na estrutura da arquitetura, enquanto as redes Residuais e Inception apresentaram melhores resultados no ILSVRC. A ResNet50 [He et al. 2016], a InceptionV3 [Szegedy et al. 2016] e a DenseNet201 [Huang et al. 2017] possuem menos parâmetros que os VGGNets, mas apresentam maior profundidade, conforme observado na Tabela 2. As arquiteturas avaliadas são comparadas em termos de: ano de publicação, profundidade topológica da rede (incluindo normalização de lote, camadas de ativação, etc.), número de parâmetros, ano de publicação.

**Tabela 2. Características das arquiteturas avaliadas.**

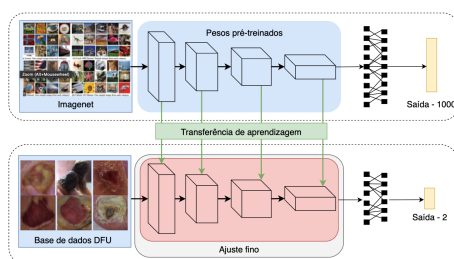
Arquitetura	Profundidade Topológica	Número de parâmetros	Ano
VGG-16	23	138.357.544	2014
VGG-19	26	143.667.240	2014
ResNet50	168	25.636.712	2015
InceptionV3	159	23.851.784	2016
DenseNet201	201	20.242.984	2017

### 3.4. Transferência de Aprendizagem e Ajuste Fino

Técnicas de transferência de aprendizado frequentemente empregadas em aprendizado profundo utilizam pesos que são pré-treinados em grandes conjuntos de dados, como o conjunto de dados da competição da *ImageNet*, apresentado em [Russakovsky et al. 2015]. Esse procedimento diminui a necessidade de retrainar todos os parâmetros da CNN do início [Yosinski et al. 2014]. Duas abordagens são frequentemente empregadas quando os pesos pré-treinados são usados para transferência de aprendizado.

A primeira é a extração de características utilizando as camadas totalmente conectadas das CNNs como entrada para classificadores do estado da arte. A segunda é o ajuste fino ou *fine-tuning*. Nessa abordagem, os pesos das camadas convolucionais são congelados e um retraining é realizado apenas nas camadas totalmente conectadas com uma taxa de aprendizagem baixa. Comparada com a primeira abordagem, ainda exige grande poder computacional, uma vez que é necessário retomar o treinamento da CNN com o conjunto de dados de destino, adaptando o modelo para o domínio desejado. Na Figura 3, é apresentado um exemplo da técnica de transferência de aprendizado por ajuste fino.

De acordo com Izadyyazdanabadi et al. 2018, existem dois tipos de ajuste fino: *Shallow Fine Tuning* (SFT) e *Deeply Fine-Tuning* (DFT). O ajuste por meio do SFT consiste no congelamento das camadas iniciais da CNN, geralmente os conjuntos de camadas convolucionais. Essas são consideradas mais gerais e conseguem representar características de forma, textura e cor. As camadas superiores costumam ser específicas de domínio,



**Figura 3. Fluxograma exemplificando a técnica de transferência de aprendizagem por ajuste fino.**

carregando conteúdo semântico dos rótulos das instâncias. A abordagem DFT permite treinar toda a rede, adaptando até as primeiras camadas. Embora exija um custo computacional mais alto, assim como maior quantidade de dados, geralmente apresenta resultados superiores ao SFT em determinados contextos.

### 3.5. Métricas de Avaliação

A avaliação das CNNs foi calculada com base nos valores obtidos pela matriz de confusão. Com base na matriz, obtemos quatro valores, denominados Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN). Nesse contexto, são calculados as métricas de Acurácia ( $A$ ), Precisão ( $P$ ),  $Recall$  ( $R$ ),  $F1-score$  e o  $Kappa$  ( $K$ ) (Equações 1 e 2) da classificação:

$$F1 - score = 2 * \frac{P * R}{P + R} \quad (1)$$

$$K = \frac{observado - esperado}{1 - esperado} \quad (2)$$

O índice  $Kappa$  ( $K$ ) considera todos os elementos da matriz de confusão. De acordo com Landis e Koch [Landis and Koch 1977], este índice assume valores entre 0 e 1, sendo o resultado qualificado de acordo com o  $K$  da seguinte maneira: Ruim:  $K \leq 0,2$ ; Moderado:  $0,2 < K \leq 0,4$ ; Bom:  $0,4 < K \leq 0,6$ ; Muito bom:  $0,6 < K \leq 0,8$ ; Excelente:  $K > 0,8$ .

## 4. Resultados e Discussões

As arquiteturas de CNN's avaliadas foram treinadas em 50 épocas com 32 steps per epoch, sendo realizado para cada arquitetura e para classificação o SFT e o DFT. Além disso, a função de custo aplicada durante o treinamento das CNNs foi a *binary cross-entropy* para as classificações binárias e *categorical cross-entropy* para a classificação multiclasse, o otimizador selecionado foi o SGD. Os resultados obtidos não mostraram necessidade de aumentar a quantidade de épocas para a classificação binária.

### 4.1. Classificações Binárias

Os experimentos envolvendo a classificação binária das classes são subdivididos em três sub-problemas: a classificação entre pele saudável e com a presença de úlcera; classificação entre pele saudável e úlceras com isquemia e entre pele saudável e úlceras com infecção. Para cada sub-problema, os experimentos foram realizados com SFT e DFT.

#### 4.1.1. Saudável vs Úlcera

Na Tabela 3 são apresentados os resultados da classificação com SFT e DFT, observa-se que os resultados obtidos pela arquitetura DenseNet201 com SFT foram superiores as demais com SFT, atingindo 96,83% de acurácia e Kappa de 0,93. Ao comparar os resultados obtidos por ambos os métodos, observa-se que o refinamento por meio do DFT apresentou melhores resultados em todas as arquiteturas, uma vez que as camadas convolucionais conseguiram aprender as características mais básicas da base de dados. Três arquiteturas obtiveram o mesmo resultado, são elas a VGG-19, ResNet50 e InceptionV3 com 99,60% de acurácia e Kappa de 0,99 e a DenseNet201 obteve 100,00% nas métricas e valor 1,00 para o índice Kappa.

**Tabela 3. Resultados da Classificação Binária: Saudável vs Úlcera com SFT e DFT.**

SFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	90,51	90,52	90,51	90,51	0,81
VGG-19	88,54	88,66	88,54	88,53	0,77
ResNet50	82,61	89,67	82,61	82,59	0,65
InceptionV3	93,28	93,28	93,20	93,28	0,86
DenseNet201	<b>96,83</b>	<b>96,83</b>	<b>96,83</b>	<b>96,83</b>	<b>0,93</b>

DFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	99,21	99,22	99,21	99,21	0,98
VGG-19	99,60	99,60	99,60	99,60	0,99
ResNet50	99,60	99,60	99,60	99,60	0,99
InceptionV3	99,60	99,60	99,60	99,60	0,99
DenseNet201	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>

#### 4.1.2. Saudável vs Isquemia

Para a classificação Saudável vs Isquemia, todas as arquiteturas testadas mantendo os pesos pré-treinados da ImageNet conseguiram acurácia acima dos 95,00%. Conforme mostra a Tabela 4, a ResNet50 com SFT obteve a menor acurácia, com 95,52% e valor 0,00 para o índice Kappa, o que significa que foi classificada corretamente apenas a classe com mais imagens, saudável. Nesta classificação, a InceptionV3 e a DenseNet201 obtiveram melhores métricas, divergindo apenas no valor de *F1-score* e de K.

No geral, para a classificação Saudável vs Isquemia, as CNNs com DFT obtiveram um aumento no valor das métricas em relação às CNNs com SFT. Conforme mostra a Tabela 4, a ResNet50 melhorou o valor do índice Kappa de 0,00 para 1,00 e, assim como a VGG-19 e a DenseNet201, obteve 100% nas demais métricas. Apenas a InceptionV3 manteve os mesmos valores para acurácia, precisão e *recall* com SFT e DFT, mas com DFT obteve melhorias nas outras métricas.

#### 4.1.3. Saudável vs Infecção

Para a classificação Saudável vs Infecção as redes se comportaram de forma parecida com o problema Saudável vs Isquemia, com a ResNet50 obtendo novamente as menores



**Tabela 4. Resultados da Classificação Saudável vs Isquemia com SFT e DFT.**

SFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	96,26	96,40	96,26	94,97	0,27
VGG-19	97,01	97,10	97,01	96,29	0,48
ResNet50	95,52	91,24	95,52	93,33	0,00
InceptionV3	<b>99,25</b>	<b>98,53</b>	<b>99,25</b>	98,36	0,79
DenseNet201	<b>99,25</b>	<b>98,53</b>	<b>99,25</b>	<b>99,22</b>	<b>0,90</b>
DFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	99,25	99,25	99,25	99,22	0,90
VGG-19	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>
ResNet50	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>
InceptionV3	99,25	99,25	99,25	99,22	0,90
DenseNet201	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>

métricas em relação as demais arquiteturas com SFT, neste caso, com valores abaixo dos 89,00% e a DenseNet201 com SFT, conforme detalhado na Tabela 5, obteve métricas maiores que as demais redes. Nestes testes, todas as arquiteturas obtiveram o valor do índice Kappa acima de 0,8, exceto a ResNet50 que obteve valor 0,75 que é considerado muito bom.

**Tabela 5. Resultados da Classificação Saudável vs Infecção com SFT e DFT.**

SFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	93,63	93,65	93,63	93,64	0,87
VGG-19	92,22	92,23	92,22	92,21	0,84
ResNet50	87,98	88,25	87,98	88,01	0,75
InceptionV3	95,05	95,05	95,05	95,04	0,90
DenseNet201	<b>97,17</b>	<b>97,17</b>	<b>97,20</b>	<b>97,16</b>	<b>0,94</b>
DFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>
VGG-19	98,93	98,94	98,93	98,94	0,97
ResNet50	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>
InceptionV3	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>
DenseNet201	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>1,00</b>

Ao treinar todas as camadas das arquiteturas para a classificação Saudável vs infecção, as arquiteturas obtiveram um aumento no valor de todas as métricas conforme mostra a Tabela 5. As redes VGG-16, ResNet50, InceptionV3 e DenseNet201 com DFT obtiveram 1,00 para o índice Kappa e 100,00% para as demais métricas. A arquitetura com menor desempenho foi a VGG-19 que atingiu métricas acima dos 97,00% e 0,97 para o índice Kappa, o que é excelente.

## 4.2. Classificação Multiclasse

A classificação multiclasse consistiu em classificar as imagens em cinco classes: (1) imagens de pele saudável; (2) imagens de úlceras sem infecção e sem isquemia; (3) imagens de úlceras com isquemia; (4) imagens de úlceras com infecção e (5) imagens de úlceras com infecção e isquemia.

Ao utilizar o SFT, conforme mostra a Tabela 6, os resultados foram bem abaixo se comparados aos resultados das classificações binárias. Neste caso, a rede que apresentou

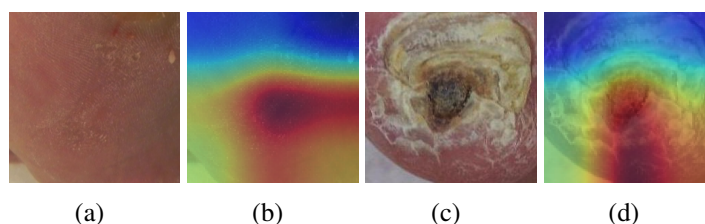
melhor desempenho foi a DenseNet201 que obteve acurácia de 85,62% e índice Kappa de 0,79.

**Tabela 6. Resultados da classificação multiclases com SFT e DFT.**

SFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	74,94	74,94	73,72	74,09	0,64
VGG-19	74,72	74,72	74,38	74,19	0,64
ResNet50	55,12	55,12	49,86	52,31	0,35
InceptionV3	80,61	80,61	80,27	80,16	0,72
DenseNet201	<b>85,62</b>	<b>85,62</b>	<b>85,14</b>	<b>85,25</b>	<b>0,79</b>
DFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K
VGG-16	91,28	91,28	91,28	91,19	0,87
VGG-19	<b>92,37</b>	<b>92,37</b>	91,39	91,78	<b>0,89</b>
ResNet50	90,41	90,41	90,20	90,24	0,86
InceptionV3	90,41	90,41	90,65	90,40	0,86
DenseNet201	92,15	92,15	<b>92,04</b>	<b>92,00</b>	<b>0,89</b>

A Tabela 6 mostra que ao treinar todos os pesos das redes, as mesmas atingiram métricas acima dos 90,00% e um índice Kappa com valor acima de 0,80 que é considerado excelente. Os resultados destas métricas mostram que o diagnóstico se assemelha bastante ao cenário real, mesmo com valores inferiores as métricas das classificações binárias.

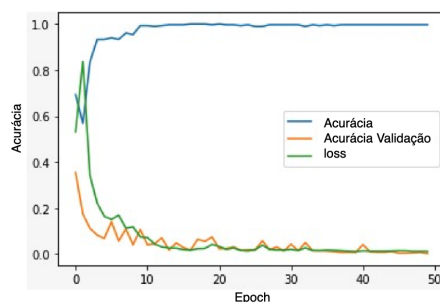
Na Figura 4 são apresentados os mapas de calor com as regiões de ativação consideradas durante a extração de características e conseqüentemente a predição da arquitetura. Na obtenção desses exemplos, foi selecionada a melhor CNN no contexto multiclases. Observam-se nas classes de isquemia e infecção que as regiões com cores mais intensas estão concentradas nas lesões provocadas pela úlcera. A Figura 5 exibe os valores de acurácia de treino e validação e *loss* durante as épocas do treinamento das CNNs. O valor de *loss* diminui no decorrer das épocas, um dos indícios de que não há *overfitting*. Outros indícios que tendem a reduzir a possibilidade de *overfitting* e foram empregados neste trabalho são: utilização do aumento de dados durante o treinamento da arquitetura, o uso da transferência de aprendizagem e ajuste-fino que tendem a reduzir a quantidade de parâmetros treináveis.



**Figura 4. Mapas de calor com as regiões de ativação para imagens saudáveis((a) e (b)) e com úlcera((c) e (d)).**

## 5. Conclusão

Neste trabalho, foram treinadas cinco arquiteturas CNNs (VGG-16, VGG-19, ResNet50, InceptionV3 e DenseNet201) para classificação de úlceras do pé diabético: (1) Saudável vs Úlceras; (2) Saudável vs Isquemia; (3) Saudável vs Infecção e (4) Saudável vs Úlcera vs Isquemia vs Infecção vs Infecção e Isquemia. Os experimentos mostraram que ao treinar todos os pesos das redes, os resultados foram melhores que os obtidos com o SFT.



**Figura 5. Comportamento da acurácia e *loss* durante o treinamento e validação da arquitetura.**

Na classificação Saudável vs Úlceras, as redes VGG-19, ResNet50 e InceptionV3 alcançaram 99,60% de precisão e 0,99 para o índice Kappa. Na classificação Saudável vs Isquemia, a VGG-19 e a ResNet50 atingiram 100,00% de acurácia e índice Kappa igual a 1,00. Na classificação Saudável vs Infecção, as redes VGG-16, ResNet50 e InceptionV3 alcançaram 100,00% de precisão e índice Kappa igual a 1,00. Na classificação multi-classe, a rede que obteve melhor desempenho foi a VGG-19 com acurácia de 92,37% e 0,89 para o índice Kappa.

Dentre as arquiteturas avaliadas, observa-se que as redes obtiveram excelentes resultados nas classificações binárias, destacando-se a DenseNet201 que alcançou melhores resultados nos três problemas de classificação. Para a classificação multiclasse, que é um cenário mais próximo a realidade do serviço médico, ainda é necessário aprimorar ou aplicar novas técnicas com a finalidade de ampliar a diversidade de classes.

## Referências

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282.
- Alzubaidi, L., Fadhel, M., Oleiwi, S., Al-Shamma, O., and Zhang, J. (2020). Dfu\_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network. *Multimedia Tools and Applications*, 79(21-22):15655–15677.
- Cavanagh, P. R., Lipsky, B. A., Bradbury, A. W., and Botek, G. (2005). Treatment for diabetic foot ulcers. *The Lancet*, 366(9498):1725–1735.
- Goyal, M., Reeves, N. D., Rajbhandari, S., Ahmad, N., Wang, C., and Yap, M. H. (2020). Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. *Computers in Biology and Medicine*, 117:103616–103616.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.
- He, X., Zemel, R. S., and Carreira-Perpinán, M. A. (2004). Multiscale conditional random fields for image labeling. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004.*, volume 2, pages II–II. IEEE.

- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708.
- Izadyyazdanabadi, M., Belykh, E., Mooney, M., Martirosyan, N., Eschbacher, J., Nakaji, P., Preul, M. C., and Yang, Y. (2018). Convolutional neural networks: ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical cle images. *Journal of Visual Communication and Image Representation*, 54:10–20.
- Kornblith, S., Shlens, J., and Le, Q. V. (2019). Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2661–2671.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Solís-Sánchez, L. O., Ortiz-Rodríguez, J., Castañeda-Miranda, R., Martínez-Blanco, M., Ornelas-Vargas, G., Galván-Tejada, J. I., Galván-Tejada, C. E., Celaya-Padilla, J. M., and Castañeda-Miranda, C. L. (2016). Identification and evaluation on diabetic foot injury by computer vision. *IEEE International Conference on Industrial Technology (ICIT)*, pages 758–762.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826.
- Veredas, F., Mesa, H., and Morente, L. (2010). Binary tissue classification on wound images with neural networks and bayesian classifiers. *IEEE Transactions on Medical Imaging*, 2(29):410–427.
- Vogado, L., Veras, R., Aires, K., Araújo, F., Silva, R., Ponti, M., and Tavares, J. M. R. (2021). Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. *Sensors*, 21(9):2989–2989.
- Wang, L., Pedersen, P. C., Agu, E., Strong, D. M., and Tulu, B. (2016). Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification. *IEEE Transactions on Biomedical Engineering*, 64(9):2098–2109.
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, volume 2, pages 3320–3328, Cambridge, MA, USA.