



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

DFU-VGG, uma Nova e Aprimorada Rede VGG-19 para Classificação de Úlceras do Pé Diabético

Francisco das Chagas Torres dos Santos

Teresina-PI, Abril de 2022

Francisco das Chagas Torres dos Santos

DFU-VGG, uma Nova e Aprimorada Rede VGG-19 para Classificação de Úlceras do Pé Diabético

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Rodrigo de Melo Souza Veras

Teresina-PI

Abril de 2022

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Sistema de Bibliotecas da UFPI – SIBi/UFPI
Biblioteca Setorial do CCN

S231d Santos, Francisco das Chagas Torres dos.
DFU-VGG, uma nova e aprimorada rede VGG-19 para
classificação de úlceras do pé diabético / Francisco das
Chagas Torres dos Santos. – 2022.
52 f.: il.

Dissertação (Mestrado) – Universidade Federal do Piauí,
Centro de Ciências da Natureza, Pós-Graduação em Ciência
da Computação, Teresina, 2022.
“Orientador: Prof. Dr. Rodrigo de Melo Souza Veras”.

1. Sistemas Operacionais (Computação). 2. Arquitetura de
Redes. 3. Banco de Dados – Imagens – Úlceras do Pé
Diabético. 4. Redes VGGs. I. Veras, Rodrigo de Melo
Souza. II. Título.

CDD 005.43

“DFU-VGG uma nova e aprimorada rede VGG-19 para classificação de úlceras do pé diabético”

FRANCISCO DAS CHAGAS TORRES DOS SANTOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Aprovada por:

Prof. Rodrigo de Melo Souza Veras
(Presidente da banca examinadora)

Profa. Andrea Gomes Campos Bianchi
(Examinadora externa à instituição)

Prof. André Macedo Santana
(Examinador interno)

Prof. Vinicius Ponte Machado
(Examinador interno)

Teresina, 20 de abril de 2022.

*Aos meus pais, Gerônimo Correia dos Santos e Maria das Graças Torres de Andrade
(in memoriam), por sempre estarem ao meu lado me incentivando e me apoiando em
todas minhas decisões.*

Agradecimentos

Primeiramente, agradeço a Deus por ter me concedido este desafio e dado a mim forças suficientes para avançar mais esta etapa em minha vida. Agradeço aos meus queridos pais, Gerônimo Correia dos Santos e Maria das Graças Torres de Andrade (in memoriam), por terem me educado com muito carinho, paciência e amor, sendo eles exemplos de pais, heróis e guerreiros.

Aos meus irmãos, Maria da Paz Torres de Andrade, que contribuiu bastante com meus estudos e me inspirou a sempre ir mais longe, e José Domingo da Rocha Filho por nunca ter duvidado de mim. A minha esposa Rayane Edyene de Abreu Freitas Santos, que está ao meu lado, incentivando-me a nunca desistir e a toda sua família, que está sempre disposta a me ajudar.

Agradeço ao meu orientador, Rodrigo de Melo Souza Veras, por todos os conselhos, paciência e ajuda nesse período. Além do suporte de todos os meus amigos do laboratório LIMCI (*Laboratory of Image Processing and Computational Intelligence*), em especial Elineide Santos, Henrique e Maíla.

Aos meus amigos da Superintendência de Tecnologia da Informação - STI/UFPI pelo incentivo ao ingresso no mestrado e por todo apoio durante esta jornada. Em especial, ao professor Ricardo Lira que, com o desejo de qualificar ainda mais a equipe de trabalho, motivou os servidores da STI/UFPI a entrarem na seleção de mestrado.

Aos meus amigos da Comunidade Católica ORE, representada pelo seu fundador Padre Neto Rêgo, por todas as orações e palavras de fé. Por fim, agradeço a todos que contribuíram de forma direta ou indireta para que esta grande vitória fosse alcançada em minha vida.

*“Tudo quanto te vier à mão para fazer,
faze-o conforme as tuas forças,
(Eclesiastes 9:10)*

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 ou inadequado 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 das arquiteturas VGG-16, VGG-19, InceptionV3, ResNet50, InceptionResNetV2, DenseNet201, MobileNetV2 e EfficientNetB0 na classificação de imagens de Úlceras do Pé Diabético de duas bases de dados públicas com um total de 8.250 imagens. Para estas redes, foi aplicado um refinamento profundo com e sem alterações nas camadas finais (totalmente conectadas). Além disso, as redes VGG-16 e VGG-19 sofreram alterações na arquitetura interna, sendo adicionadas camadas de dropout e batch normalization. A avaliação proposta levou em consideração quatro classes: inexistente (que contém imagens de pele saudável, úlceras em processo de cicatrização e de úlceras sem isquemia ou infecção), isquemia, infecção e ambas (úlceras com isquemia e infecção). As melhores configuração das redes testadas foram as VGG-16 e VGG-19 com uma camada densa de 512 unidades e com camadas de batch normalization, que obtiveram índice Kappa acima dos 89,00% e uma acurácia média de 93,44% e 93,45%, respectivamente. Embora, as redes VGGs tenham obtidas métricas bem próximas, propomos para a classificação de úlceras do pé diabético a VGG-19 com batch normalization (denominada DFU-VGG), pois, esta configuração de rede classificou erroneamente como inexistente 7,65% das imagens de infecção, enquanto a VGG-16 com batch normalization fez o mesmo com 8,35% das imagens de infecção classificando-as como da classe inexistente. Os resultados alcançados demonstram que a DFU-VGG proposta consegue classificar corretamente as imagens, visto que, nos testes realizados o índice kappa atingiu valores considerados “Excelentes”.

Palavras-chaves: Úlceras do Pé Diabético, CNN, Refinamento, Dropout, Batch Normalization.

Abstract

A complication caused by diabetes mellitus is the appearance of wounds located in the region of the feet called Diabetic Foot Ulcers. Late or inadequate 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 the VGG-16, VGG-19, InceptionV3, ResNet50, InceptionResNetV2, DenseNet201, MobileNetV2 and EfficientNetB0 architectures in the classification of Diabetic Foot Ulcer images from two public databases with a total of 8,250 images was carried out. For these networks, a deep refinement with no changes in the final layers (fully connected) was applied. In addition, the VGG-16 and VGG-19 networks underwent changes in their internal architecture, with the addition of dropout and batch normalization layers. The proposed evaluation considered four classes: none (which contains images of healthy skin, ulcers in the process of healing and ulcers without ischaemia or infection), ischaemia, infection and both (ulcers with ischaemia and infection). The best configurations of the tested networks were the VGG-16 and VGG-19 with a dense layer of 512 neurons and with batch normalization layers, which obtained Kappa index larger than 89.00% and an average accuracy of 93.44% and 93.45%, respectively. Although VGGs networks have obtained very close metrics, we propose VGG-19 with batch normalization (called DFU-VGG) for the classification of diabetic foot ulcers, as it wrongly classified 7,65% of the infection images as none, while the VGG-16 with batch normalization did the same with 8,35% of the infection images. The results showed that DFU-VGG proposed is able to classify such images, since, in the tests performed, the kappa index reached values considered “Excelent”.

Keywords: Diabetic Foot Ulcers, CNN, Refinement, Dropout, Batch Normalization.

Lista de ilustrações

Figura 1 – Exemplos de imagens de pés com úlcera com isquemia, úlcera com infecção e úlcera em processo de cura.	2
Figura 2 – Exemplos de imagens de pele saudáveis e de úlceras da base DFU.	11
Figura 3 – Imagens de isquemia da base DFU 2020 com aumento de dados.	11
Figura 4 – Gráfico exibindo o percentual de imagens por classes após a união das bases DFUs.	12
Figura 5 – Fluxograma exemplificando a técnica de transferência de aprendizagem por ajuste fino.	16
Figura 6 – Exemplo comparativo de uma rede neural sem <i>dropout</i> e outra com aplicação de <i>dropout</i> . Fonte: (SRIVASTAVA et al., 2014)	17
Figura 7 – Exemplo comparativo de uma rede neural sem <i>batch normalization</i> e outra com aplicação de <i>batch normalization</i> . Fonte: (GARBIN; ZHU; MARQUES, 2020)	18
Figura 8 – Comparação das arquiteturas das redes VGG-16 e VGG-19.	20
Figura 9 – Arquitetura da rede VGG-19 com inclusão de camadas <i>batch normalization</i>	21
Figura 10 – Gráficos de treinamento das CNNs EfficientNetB0 (a) e DenseNet201 (b).	25
Figura 11 – Processo de ablação com uma camada com 256 (a), 512 (b) ou 1024 (c) unidades.	26
Figura 12 – Processo de ablação com duas camadas totalmente conectadas: uma de 512 e outra de 256 (a), uma de 1024 e outra de 256 (b) e uma de 1024 e outra de 512 (c) unidades.	26
Figura 13 – Gráficos de treinamento da CNNs EfficientNetB0 (a) e VGG-19 (b), ambas com uma camada densa de 512 unidades.	27
Figura 14 – Matriz de confusão visual da CNN DFU-VGG.	28
Figura 15 – Mapa de calor com as regiões de ativação para as imagens inexistente ((a) e (b)), infecção ((c) e (d)), isquemia ((e) e (f)) e ambas ((g) e (h)).	29
Figura 16 – Gráficos de treinamento da VGG-19 com uma camada densa de 512 unidades (a) e da DFU-VGG (b).	30
Figura 17 – Arquitetura da rede VGG-19 com inclusão de camadas <i>dropout</i>	31
Figura 18 – Gráficos de treinamento da CNN VGG-19 com uma camada densa com 512 unidades sem (a) e com (b) aplicação de <i>dropout</i>	32
Figura 19 – Arquitetura da rede VGG-19 com inclusão de camadas <i>dropout</i> e <i>batch normalization</i>	33

Figura 20 – Gráficos de treinamento da CNN VGG-19 sem (a) e com (b) aplicação de dropout e batch normalization. 34

Lista de tabelas

Tabela 1 – Sumário dos trabalhos encontrados na literatura.	9
Tabela 2 – Número de imagens por classes nas bases DFU.	12
Tabela 3 – Características das arquiteturas avaliadas.	14
Tabela 4 – Significado do Valor de K	19
Tabela 5 – Resultados da classificação com SFT e DFT.	23
Tabela 6 – Resultado das classificações obtidas com as CNNs originais (melhores valores em negrito).	24
Tabela 7 – Melhores resultados da classificação com as CNNs com alteração das camadas densas.	26
Tabela 8 – Resultados da classificação com Batch Normalization.	27
Tabela 9 – Resultados da classificação com adição de camada dropout.	32
Tabela 10 – Resultados da classificação com Batch Normalization e Dropout.	34

Sumário

1	INTRODUÇÃO	1
1.1	Motivação	2
1.2	Objetivo Geral	3
1.2.1	Objetivos Específicos	3
1.3	Organização da Defesa	3
1.4	Produção Científica e Contribuições	4
1.4.1	Trabalho Publicado	4
1.4.2	Trabalho Aceito	4
2	ESTADO DA ARTE	5
3	MATERIAIS E MÉTODOS	10
3.1	Base de Imagens	10
3.2	Aumento de dados	13
3.3	Arquiteturas de CNNs Avaliadas	13
3.4	Transferência de Aprendizagem e Ajuste Fino	15
3.5	Dropout e Batch Normalization	16
3.6	Métricas de Avaliação	18
3.7	Considerações Finais	19
4	MÉTODO PROPOSTO	20
4.1	DFU-VGG Proposta	21
4.2	Considerações Finais	22
5	RESULTADOS E DISCUSSÃO	23
5.1	Resultados com as Redes em suas Configurações Originais	24
5.2	Resultados com Alterações das Camadas Totalmente Conectadas	25
5.3	Resultados das VGGs com Batch Normalization	27
5.4	Resultados das VGGs com Dropout	30
5.5	Resultados das VGGs com dropout e batch normalization	32
6	CONCLUSÕES E TRABALHOS FUTUROS	35
6.1	Trabalhos Futuros	36
	REFERÊNCIAS	37

1 Introdução

Diabetes Mellitus é uma doença metabólica crônica, caracterizada por alto nível de glicose no sangue, que pode causar complicações como: cegueira, doenças cardiovasculares, falência renal e úlceras do pé diabético (*Diabetic Foot Ulcers - DFU*)(GOYAL; HASSANPOUR, 2020).

Estas úlceras consistem no aparecimento de feridas situadas na região dos pés, normalmente causadas por trauma (pisar sobre objetos penetrantes), estresse mecânico repetitivo (caminhar com calçado inadequado ou descalço) ou estresse mecânico continuamente aplicado (pressão não percebida)(CAVANAGH et al., 2005).

As DFUs necessitam de tratamento adequado, pois em estágio mais avançado podem ocasionar a amputação dos membros infectados. Um estudo descrito por 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 em lesões graves, tais lesões podem ser 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, conforme mostrado na Figura 1a. A isquemia consiste na distribuição inadequada de sangue pela região e, visualmente, pode ser indicada pela presença de má perfusão no pé ou nos dedos gangrenados (morte de tecidos afetados), conforme mostra a Figura 1b. Seu diagnóstico ocorre geralmente palpando os pulsos do pé e medindo a pressão sanguínea no pé e nos dedos (GOYAL et al., 2020). A Figura 1c exibe uma úlcera com isquemia e infecção. Algumas úlceras após o tratamento adequado alcançam um estado de cicatrização e ficam similares a pele saudável como mostra a Figura 1d.

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 neurais foram utilizadas em trabalhos de classificação de úlceras do pé diabético, por isso, com intuito de avaliar qual rede tem melhor desempenho em

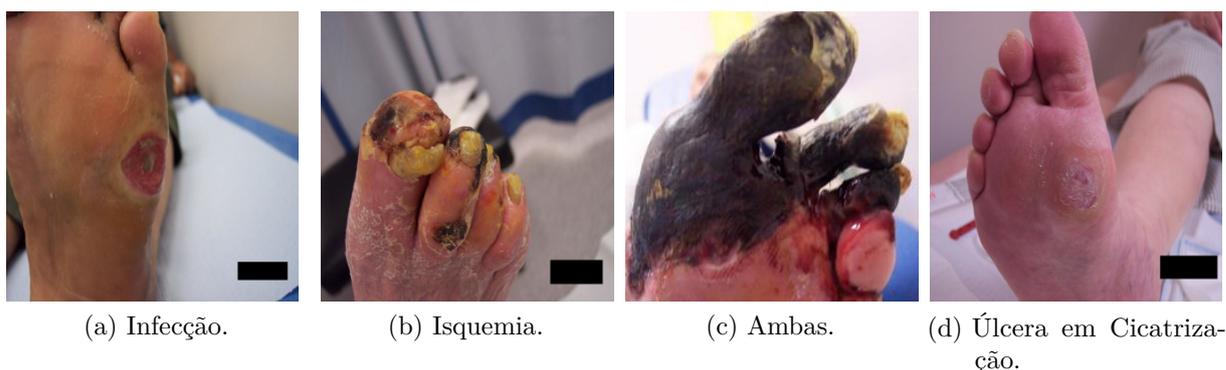


Figura 1 – Exemplos de imagens de pés com úlcera com isquemia, úlcera com infecção e úlcera em processo de cura.

acurácia, 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, InceptionresNetV2, DenseNet201, MobileNetV2 e EfficientNetB0.

Na avaliação das CNNs foram utilizadas duas bases de dados públicas, uma contendo 2.295 imagens e a outra 5.955 imagens pertencentes a quatro classes de diagnóstico, são elas: inexistente (que contém imagens de pele saudável, úlceras em processo de cicatrização e de úlceras sem isquemia ou infecção), isquemia (úlceras apenas com isquemia), infecção (úlceras apenas com infecção) e ambas (úlceras com isquemia e infecção). Após a classificação, é feita uma análise comparativa com intuito de identificar qual arquitetura de rede obteve melhor desempenho na classificação.

Além de avaliar as redes padrões, foi realizado um trabalho de configuração das camadas densas das redes, testando com uma camada densa e variando a quantidade de neurônios: 256, 512 e 1024. Depois foi testado com duas camadas densas seguindo as configurações: 512-256, 1024-256 e 1024-512 neurônios para as respectivas camadas densas. Após comparar os resultados das seis configurações de rede já citadas, foi realizada alterações nas arquiteturas internas das redes VGG-16 e VGG-19, adicionando as mesmas camadas de *dropout* (DP), em outra configuração, camadas *batch normalization* (BN) e uma configuração com estas duas camadas.

1.1 Motivação

Com base nas complicações causadas pelo diabetes, o que inclui as úlceras do pé diabético, devido ao fato de que elas precisam de um tratamento adequado, pois, quando não tratadas adequadamente podem evoluir para um estado mais grave, sendo necessário a amputação do membro do paciente e considerando que o tratamento para esta patologia é caro e muitas vezes de difícil acesso, visto que, a quantidade de médicos não

é suficiente para atender toda a demanda. Dessa forma, motivados a ajudar os médicos e pacientes no combate a esta patologia, acreditamos que a utilização de CNNs seja adequada para a classificação de imagens de DFU e que isso permitirá uma triagem dos pacientes, pois, a classificação prévia indicará quais pacientes precisam urgentemente de acompanhamento médico. Conseqüentemente, sabendo a situação atual da úlcera, mais pessoas serão acompanhadas e tratadas de forma mais rápida e este acompanhamento ajudará no processo de cicatrização da úlcera.

1.2 Objetivo Geral

Este trabalho tem como objetivo principal propor uma arquitetura de CNN para a classificação de imagens de úlceras do pé diabético.

1.2.1 Objetivos Específicos

Durante a pesquisa foram estabelecidos os objetivos específicos abaixo:

- Avaliar diferentes CNNs em suas configurações originais;
- Avaliar o comportamento das CNNs com alterações de suas camadas totalmente conectadas;
- Avaliar o comportamento das CNNs com a inserção de camadas de *dropout*;
- Avaliar o comportamento das CNNs com a inserção de camadas de *batch normalization*;
- Avaliar o comportamento das CNNs com a inserção de camadas de *dropout* e *batch normalization*.

1.3 Organização da Defesa

Este trabalho está organizado em: Capítulo 2 apresenta os trabalhos relacionados ao tema proposto; no Capítulo 3 são apresentadas as bases de imagens, o aumento natural de dados, o ajuste fino, as técnicas de *dropout* e *batch normalization* e as métricas de avaliação utilizadas. O Capítulo 4 apresenta o método proposto para solucionar o problema de classificação de imagens de úlceras do pé diabético. O Capítulo 5 contém a discussão dos resultados mediante sua comparação com aos trabalhos da literatura, já no Capítulo 6 é apresentada a conclusão.

1.4 Produção Científica e Contribuições

Ao longo deste trabalho, um artigo científico foi publicado e outro aceito no meio científico.

1.4.1 Trabalho Publicado

- **Francisco Santos**, Rodrigo Veras, Elineide Santos, Maila Lima Claro, Luís Henrique Vogado, Márcia Ito e Andrea Bianchi. **Uma Avaliação de Arquiteturas de Aprendizado Profundo para a Classificação de Úlceras do Pé Diabético**, Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), 2021, p. 323-334.

1.4.2 Trabalho Aceito

- **Francisco Santos**, Rodrigo Veras, Elineide Santos, Luís Henrique Vogado, Márcia Ito, Andrea G. C. Bianchi e João Manuel R. S. Tavares. **DFU-VGG a Novel and Improved VGG-19 Network for diabetic foot ulcer classification**, 26TH International Conference on Pattern Recognition (ICPR), 2022.

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. À luz desta afirmação, as principais metodologias avaliadas são explanadas a seguir.

No trabalho de [Veredas, Mesa e Morente \(2010\)](#), os autores propuseram uma metodologia de identificação automática de tecidos usando uma técnica que combina redes neurais e classificadores Bayesianos. Para obter cor e textura padrões, inicialmente foi realizada a segmentação das 113 imagens 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 de tecidos: pele, cicatrização, granulação, descamação e necrose. Esta abordagem obteve uma acurácia de 91,50% em seus resultados.

Em [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 do centro de Lesões da Escola de Medicina da Universidade de Massachusetts (UMass) 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; ZEMEL; CARREIRA-PERPINÁN, 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. \(2020\)](#) 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%, recall 93,60% e F1-score de 94,50% na classificação de 754 imagens de pés de pacientes com úlcera do pé diabético e pele saudável do Centro de Diabéticos do Nasiriyah Hospital, no sul do Iraque. Para fins de comparação, as redes pré-treinadas GoogleNet, VGG16 e AlexNet foram ajustadas e treinadas novamente para esta tarefa, no entanto, estas redes não superaram a rede proposta.

Goyal et al. (2017), utilizando a validação cruzada 10-fold em um conjunto com 1.679 imagens rotuladas, realizou a classificação binária: pele saudável e úlcera. Para esta classificação, foi proposta uma rede neural convolucional denominada DFUNet, que combina dois tipos de camadas convolucionais, ou seja, camadas de convolução tradicionais no início da rede que usam um único filtro convolucional seguido por camadas convolucionais paralelas, que usam várias camadas convolucionais para extração de vários recursos da mesma entrada. O DFUNet, resumido na Fig. 4, é dividido em três seções principais: as camadas de inicialização inspiradas no GoogLeNet, camadas de convolução paralela para discriminar o DFU de forma mais eficaz do que as camadas de rede anteriores e, por último, ambas as camadas totalmente conectadas e um classificador de saída baseado em softmax. A DFUNet proposta utilizou aumento de dados e obteve 0,934 de Sensibilidade, 0,939 de F-measure, 0,962 de AUC e 0,92 de Acurácia.

Goyal et al. (2020) realizou a classificação de imagens em isquemia x não isquemia e infecção x não infecção. 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 CIELAB) e textura (*Local binary pattern* - LBP, *Histogram of Oriented Gradient* - 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. Estas imagens são de pés de pacientes do Lancashire Teaching Hospitals. 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.

Amin et al. (2020) propôs uma arquitetura que classifica e localiza diferentes tipos de imagens de DFU como isquemia e infecção. Primeiramente, é realizada a classificação

usando uma rede neural com 16 camadas convolucionais em combinações com os classificadores: Naive Bayes, KNN, Softmax, Ensemble e Árvore de Decisão. Após esta classificação, as imagens de entrada são passadas para o modelo YOLOv2-DFU que é projetado pela combinação de YOLOv2 e uma rede aleatória de 172 camadas para a localização da região anormal. Foram utilizadas imagens de pacientes do Lancashire Teaching Hospitals. A base contém 15.762 imagens, incluindo as imagens resultantes do aumento natural de dados. Para a classificação isquemia x não isquemia, a melhor acurácia de 97,90% foi com o classificador Naive Bayes e para infecção x não infecção, foi de 99,60% com árvore de decisão.

Das, Roy e Mishra (2021a), para a classificação binária de úlceras do pé diabético e pele normal em 1.679 *patches* de imagens obtidas da Lancashire Teaching Hospitals, propôs uma rede convolucional denominada DFU-SPNet. Esta rede consiste em três camadas de convolução paralelas empilhadas de tamanho de kernel 1×1 , 3×3 e 5×5 . A camada anterior do primeiro bloco paralelo empilhado é uma única camada de convolução de 7×7 seguida por uma camada de transição. A camada de transição consiste em *batch normalization* e ativação de LeakyReLU e ajuda a padronizar as entradas para acelerar o treinamento. A saída concatenada de cada paralelo empilhado bloco de convolução é passado por uma sequência de 1×1 convolução, transição e camada MaxPooling 2×2 . Finalmente, a saída dos blocos de convolução paralelos empilhados passam por uma camada Flatten e por uma camada totalmente conectada de 32 unidades e em seguida, por uma camada *dropout*. A camada de saída consiste em uma função de ativação sigmóide para obter um valor predito binário da classe 0 (Normal) ou 1 (Anormal). O DFU-SPNet usando o otimizador SGD obteve acurácia de 96,40%, F1-Score de 95,4% e AUC de 0,974.

Rostami et al. (2021) realizaram classificação binária das imagens e classificação em três classes: úlceras diabéticas, úlceras venosas e cirúrgicas. Foram utilizados dois classificadores de imagens: um sendo uma arquitetura AlexNet pré-treinada com pesos ajustados usando o próprio conjunto de dados e o outro, um classificador que aplica a técnica de janela deslizante na imagem da ferida de entrada para extrair nove sub-regiões de igual tamanho junto com a etapa de classificação do patch para prever o tipo de ferida. Para cada imagem de entrada, cada um dos classificadores gera pontuações de classificação para todas as classes. O tipo de ferida para a imagem inteira é então previsto pela votação da maioria no rótulo de predição das sub-regiões que foram detectadas como feridas. Então, o vetor de características alimenta um classificador Perceptron Multicamadas (MLP) de quatro camadas com duas camadas ocultas que têm oito e sete neurônios, respectivamente. O número de nós nas camadas de entrada e saída é determinado com base no tipo de problema de classificação. A saída do classificador MLP é o tipo da ferida da imagem de entrada. Foi utilizado um conjunto de 400 imagens coletadas durante dois anos no Centro de Feridas e Vascular da Advancing the Zenith of Healthcare (AZH), Milwaukee, Wisconsin, Estados Unidos. Foi utilizada a validação cruzada 5-fold, sendo obtida acurácia

máxima de 91,90% e média de 87,70% para problemas de classificação de 3 classes.

Galdran, Carneiro e Ballester (2021) realizaram a classificação de imagens da base DFU do desafio 2021 composta com 15.683 imagens, sendo estas classificadas em quatro classes: sem infecção e sem isquemia, presença de isquemia, presença de infecção e presença de infecção e isquemia na mesma úlcera. Foram utilizados CNNs e Transformadores de Visão (*Vision Transformers - ViT*), uma arquitetura poderosa para aplicações de processamento de linguagem natural. De um modo geral, os ViT consideram as imagens como sequências de pequenos patches semelhantes a palavras ou tokens, dessa forma, não há noção de distância dentro de uma imagem. Quatro arquiteturas foram analisadas: Big Image Transfer (BiT) a ResNeXt50, EfficientNet, ViT e Transformadores de imagem com eficiência de dados (Data-efficient Image Transformers - DeIT), um refinamento do ViT com melhores estratégias de pré-treinamento. A melhor arquitetura foi a BiT-ResNeXt50, no entanto, a solução vencedora do desafio DFUC 2021 foi uma combinação linear das previsões extraídas do BiT-ResNeXt50 e EfficientNet B3 que obteve 62,16% de F1-score, 88,55% de AUC, 65,22% de Recall e precisão de 61,40%.

Al-Garaawi et al. (2022) realizaram a classificação binária de imagens DFU nos seguintes cenários: normal x anormal, isquemia x não isquemia e infecção x não infecção. Durante o estudo, foram investigados os benefícios do uso de códigos LBP (*Local Binary Patterns - LBP*) como entradas para modelos CNN na classificação DFU, sendo projetada uma arquitetura CNN com três entradas diferentes: DFU-RGB-Net usando as imagens RGB originais; DFU-TEX-Net usando imagens LBP mapeadas com codificação de textura e DFU-RGBTEX-Net usando imagens LBP mapeadas com codificação de textura e RGB. As etapas do processo são: (i) extrair os recursos de textura usando o método LBP básico, e então converter os códigos LBP extraídos para um espaço 3D para tornar os códigos LBP adequados como entrada da CNN, e (ii) treinar vários modelos CNN nas imagens RGB e nos códigos LBP mapeados separadamente e em recursos de textura treinados apenas na classificação DFU em comparação com os modelos CNN treinados em imagens RGB apenas. Foi proposta uma CNN com apenas quatro camadas de convolução e duas camadas totalmente conectadas sem preenchimento. Cada camada de convolução consiste em uma função de ativação de unidade linear retificadora (ReLU), seguida por uma camada de agrupamento. O satisfatório tamanho da arquitetura CNN proposta é encontrado empiricamente, onde o número de convolução e max-pooling é aumentado gradativamente. Logo após, o número de filtros é ajustado gradativamente e, em seguida, é escolhida a rede com melhor desempenho. Foi utilizada a base DFU 2021, sendo 16.790 imagens para classificação normal x anormal incluindo um aumento de 10 vezes e para classificação isquemia x não-isquemia 9.870 imagens e para infecção x não-infecção 5.892 imagens. Os resultados mostraram que a DFU-RGBTEX-Net proposta forneceu melhor desempenho do que os métodos baseados em CNN, com 94,10% de acurácia e 98,10% de AUC para a classificação normal x anormal, 99,00% de acurácia e 99,50% de AUC para a

classificação de isquemia e 74,20% de acurácia e 82,00% de AUC para a classificação de infecção.

Neste trabalho em questão, para a classificação de úlceras do pé diabético, propomos uma CNN com a rede VGG-19 em *backbone* e inclusão de camadas *batch normalization*. Utilizando 8.250 imagens das bases DFU de 2020 e 2021 e a validação cruzada 5-fold, foi realizada a classificação em quatro classes: inexistente, infecção, isquemia e ambas (úlceras com infecção e isquemia). A CNN proposta, denominada DFU-VGG obteve 93,45% de acurácia e 93,56% de precisão.

A Tabela 1 resume os trabalhos encontrados na literatura em termos de ano de publicação, técnica(s) de classificação utilizada(s), número de imagens utilizadas, número de bases de imagens utilizadas e a métrica alcançada, que pode ser Acurácia (A), Sensibilidade (S), Especificidade (E) e Precisão (P). Em todos esses trabalhos, as imagens usadas para validação foram extraídas das mesmas bases de imagens usadas para ajustar seus parâmetros.

Tabela 1 – Sumário dos trabalhos encontrados na literatura.

Referência	Técnica(s)	N. de ima- gens	N. de classes	Performance(%)
(VEREDAS; MESA; MORENTE, 2010)	Neural Networks Bayesian Classifiers	113	5	A: 91,50
(WANG et al., 2016)	SVM	100	2	S: 73,30 E: 94,60
(GOYAL et al., 2017)	DFUNet	1,423	2	A: 92,50
(ALZUBAIDI et al., 2020)	DFU-QUTNet SVM KNN	754	2	P: 95,40
(GOYAL et al., 2020)	InceptionV3 ResNet50 InceptionResNetV2 SVM	1,459	2	A Isq: 90,00 A Inf: 73,00
(AMIN et al., 2020)	Neural Networks Naive bayes Neural Networks Decision tree	15,762	2	A Isq: 97,90 A Inf: 99,60
(DAS; ROY; MISHRA, 2021b)	DFU_SPNet	1,679	2	A: 96,40
(ROSTAMI et al., 2021)	AlexNet Sliding window MLP	400	3	A max: 91,90 A average: 87,70
(GALDRAN; CARNEIRO; BALLESTER, 2021)	BiT-ResNeXt50 EfficientNet B3	15,683	4	AUC: 88,55 P: 61,40
(AL-GARAAWI et al., 2022)	DFU-RGB-TEX-Net	32,552	2	A: 94,00 A Isq: 99,00 A Inf: 74,20
Santos, Francisco das C. T. dos	DFU-VGG: VGG-19 com Batch Normalization	8,250	4	A: 93,45 P: 93,56

3 Materiais e Métodos

Este trabalho propõe definir uma arquitetura de CNN para a classificação de imagens de úlceras do pé diabético. Neste capítulo, serão apresentadas as bases de imagens, o aumento natural de dados, as arquiteturas de CNNs avaliadas, as técnicas de aprendizagem, as técnicas de *dropout* e *batch normalization* e as métricas de avaliação utilizadas.

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 (Remoção de tecidos necróticos e desvitalizados), durante cinco anos no *Lancashire Teaching Hospitals*. Esta base de imagens denominada *Diabetic Foot Ulcer* (DFU)¹ é pública com *ground truth* produzido com apoio de dois profissionais de saúde (um médico consultor e um médico especialista em pé diabético) e em caso de desacordo entre eles, a decisão final era tomada pelo médico mais experiente.

Neste trabalho, foi utilizada a união das bases DFU de 2020 e a base atualizada em 2021. A base DFU 2020 que tivemos acesso é resultado do trabalho [Goyal et al. \(2020\)](#), que realizou duas classificações binárias: (1) úlceras com infecção e úlceras sem Infecção e (2) úlceras com isquemia e úlceras sem insquemia. A base DFU 2020, além de conter classes diferentes das classes da base DFU 2021, contém imagens oriundas do aumento natural de dados. A classe com úlceras que apresentam infecção contém a imagem original da úlcera e imagens com aproximação da mesma úlcera. A classe com úlceras que apresentam isquemia contém a imagem da úlcera e imagens de rotação e espelhamento. Por isso, foi necessário um trabalho na base DFU 2020 para adptá-la as classes existentes na base DFU 2021. A base DFU 2021 estava dividida em quatro classes: (1) inexistente, que contém imagens de pele saudável, úlceras em processo de cicatrização e úlceras sem infecção ou isquemia (Figura 2a); (2) infecção, que contém imagens de úlceras apenas com infecção (Figura 2b); (3) isquemia, que contém imagens de úlceras apenas com isquemia (Figura 2c) e (4) ambas, conjunto com imagens de úlceras contendo infecção e isquemia ao mesmo tempo (Figura 2d).

Primeiramente foi realizada a exclusão das imagens resultantes do aumento natural de dados, que apresentam o nome da imagem original mais um M indicando espelhamento e um R indicando rotação, conforme podemos ver na Figura 3, onde há a imagem 000505_30.jpg, o espelhamento 000505_30_M.jpg, as rotações 000505_30_R1.jpg,

¹ Disponível em: <http://www2.docm.mmu.ac.uk/STAFF/M.Yap/dataset.php>.

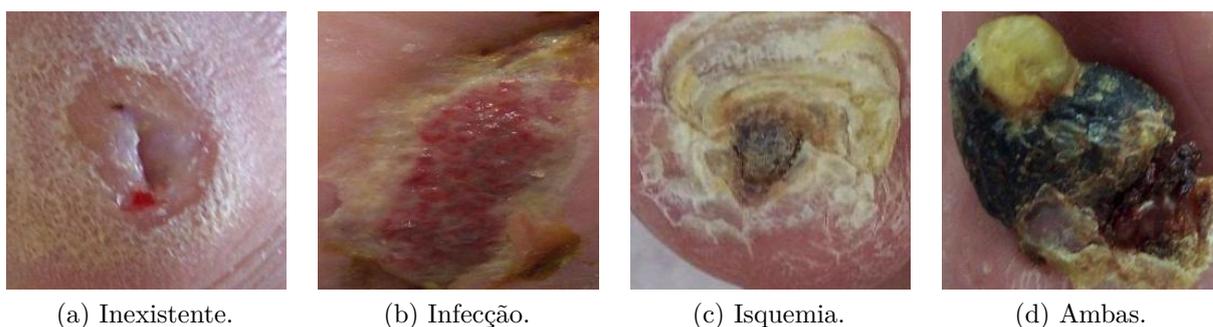


Figura 2 – Exemplos de imagens de pele saudáveis e de úlceras da base DFU.

000505_30_R2.jpg e 000505_30_R3 e também combinações de rotações e espelhamento. O mesmo ocorre com a imagem 000514_10.jpg, que tem espelhamento e rotações. Esta nomenclatura padronizada permitiu a criação de um *script* em Python para comparar os nomes dos arquivos e apagar as imagens indesejadas: Primeiramente, foram excluídas as imagens resultantes do aumento de dados. Optamos por não utilizar as imagens provenientes do aumento de dados com o intuito de diminuir a quantidade de imagens de cada classe para permitir uma comparação manual (de forma visual), onde foram comparadas as imagens em cada classe, sendo anotada a correspondência entre as mesmas, dessa forma constatou-se que as imagens comuns as classes compartilhavam da mesma nomenclatura.

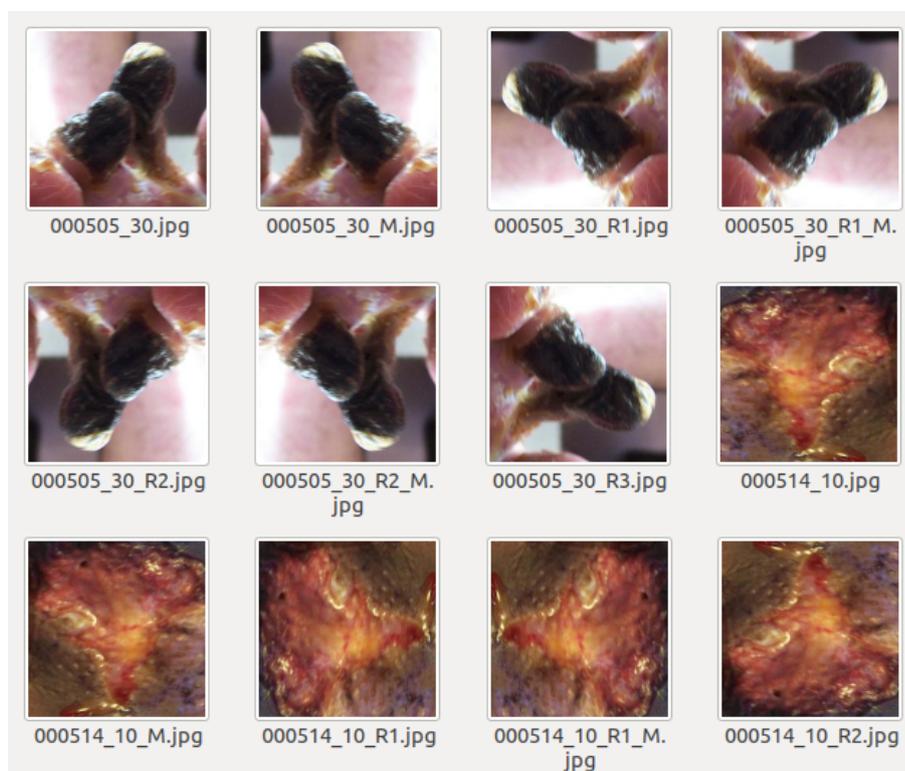


Figura 3 – Imagens de isquemia da base DFU 2020 com aumento de dados.

O segundo passo foi dividir a base DFU 2020 nas mesmas classes existentes na base DFU 2021, esta divisão foi realizada com um script em Python que compara o nome

de cada imagem de infecção com o nome de todas as imagens de isquemia e caso encontre imagens iguais nas duas classes significa que estas imagens deverão pertencer a classe ambas, ou seja, é uma imagem que apresenta infecção e isquemia ao mesmo tempo. A comparação por nome foi suficiente, pois a análise visual das imagens confirmou que imagens comuns as classes estavam com a mesma nomenclatura. Caso a imagem presente na classe infecção não esteja presente na classe isquemia significa que esta imagem é apenas da classe infecção e caso a imagem presente na classe isquemia não esteja presente na classe infecção significa que esta imagem é apenas de isquemia.

A Tabela 2 indica o número de imagens por classe das bases de imagens analisadas: base DFU 2020, base DFU 2021 e ambas as bases. Observa-se que as bases já estão com as mesmas classes e não há imagens provenientes do aumento natural de dados. 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. A Figura 4 mostra o percentual de imagens por classes considerando as duas bases. As classes estão desbalanceadas em relação a quantidade de imagens. A classe inexistente contém 46,46% das imagens e a classe isquemia contém apenas 3,07% das imagens. Não foi realizada alguma ação para balancear a quantidade de imagens nas classes.

Tabela 2 – Número de imagens por classes nas bases DFU.

Classes	DFU 2020	DFU 2021	União das Bases DFUs
Inexistente	1281	2552	3833
Isquemia	26	227	253
Infecção	779	2555	3334
Ambas	209	621	830
Total	2.295	5.955	8.250

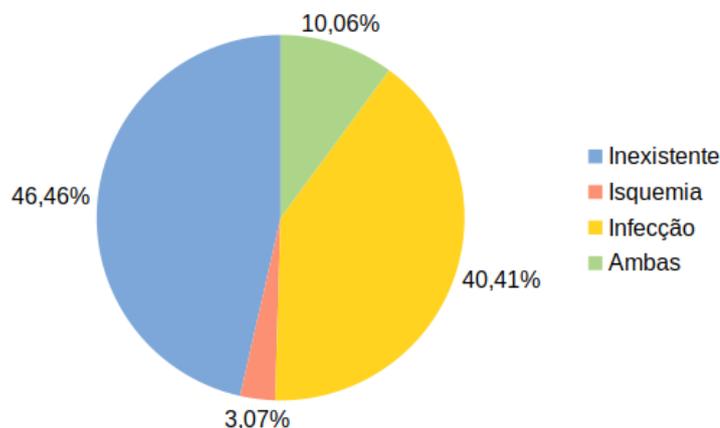


Figura 4 – Gráfico exibindo o percentual de imagens por classes após a união das bases DFUs.

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; conseqüentemente, o aumento na capacidade de generalização da arquitetura (WANG; PEREZ et al., 2017).

Neste trabalho, durante a etapa de refinamento da CNN, foi utilizado o aumento de dados proveniente do Keras, que é uma biblioteca de software de código aberto para redes neurais artificiais, escrita em Python e atua como uma interface para a biblioteca de aprendizado de máquina e inteligência artificial TensorFlow (CHOLLET et al., 2015). Esta é uma plataforma de código aberto para aprendizado de máquina. 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.

3.3 Arquiteturas de CNNs Avaliadas

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, Shlens e Le (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 arquiteturas avaliadas são apresentadas na Tabela 3, sendo comparadas em termos de: ano de publicação, profundidade topológica da rede, número de parâmetros e ano de publicação.

Os modelos VGG-16 e VGG-19 são versões da rede VGGNet proposta por Simonyan e Zisserman (2015). Estas CNN's utilizam vários filtros com menor dimensionalidade para substituir um filtro com maior dimensionalidade. A VGG-16 possui 16 camadas treináveis, sendo 13 camadas convolucionais divididas em cinco blocos e três camadas totalmente conectadas. Entre cada bloco convolucional existe uma camada MaxPooling. Por fim, duas camadas totalmente conectadas com 4.096 unidades e uma camada de saída com a

Tabela 3 – 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
InceptionResNetV2	572	55.873.736	2017
DenseNet201	201	20.242.984	2017
MobileNetV2	88	3.538.984	2018
EfficientNetB0	240	5.330.571	2019

função de ativação *softmax*. A VGG-19 possui 19 camadas treináveis, sendo 3 camadas convolucionais a mais que a VGG-16. Ao todo, as redes VGG-16 e VGG-19 possuem cerca de 138 e 143 milhões de parâmetros, respectivamente.

A Rede Neural Residual proposta por [He et al. \(2016\)](#) foi desenvolvida para solucionar o problema do desaparecimento do gradiente ou *vanishing gradient* que ocorre quando são adicionadas muitas camadas em um modelo sequencial. A arquitetura ResNet é formada por blocos residuais que pulam as conexões entre a entrada do próprio bloco e a saída. Os mapas residuais são mais fáceis de otimizar e, conseqüentemente, evitam a degradação causada pelo grande número de camadas. [He et al. \(2016\)](#) propuseram a ResNet em cinco arquiteturas com diferentes profundidades: 18, 34, 50, 101 e 152 camadas treináveis. Esses modelos alcançaram taxas de erro menores que as arquiteturas mais populares no estado da arte, como a VGG-16 e a GoogLenet. Neste trabalho foi avaliada a ResNet50, que possui uma topologia de 168 camadas, sendo apenas 50 delas treináveis.

A InceptionV3 foi proposta por [Szegedy et al. \(2016\)](#) como sucessora das arquiteturas GoogLenet e InceptionV2. Com profundidade topológica de 159 camadas e 24 milhões de parâmetro, a InceptionV3 possui blocos simétricos e assimétricos, camadas convolucionais, *average* e *MaxPolling*, concatenação de características, dropouts e camadas densas. A capacidade de fatorizar convoluções é uma das principais características da InceptionV3 e tem como principal objetivo a redução no número de parâmetros e conseqüentemente a redução do custo atrelado a essas operações.

[Szegedy, Ioffe e Vanhoucke \(2016\)](#) apresentaram um novo mecanismo para ser incorporado nas arquiteturas de redes neurais convolucionais. Nesse trabalho, eles propuseram a adição da normalização em lote durante o treinamento dos modelos. Eles observaram que a adição desse método em modelos de classificação de imagens do estado da arte resultou em um aumento substancial na velocidade de treinamento da rede. Para avaliação dos resultados, eles modificaram a arquitetura Inception ([SZEGEDY et al., 2014](#)) adicionando sua proposta de normalização. Essa variante conhecida com Inception-v2 ([IOFFE; SZEGEDY, 2015](#)) atingiu os melhores resultados nos benchmarks do desafio de

classificação ImageNet.

A rede DenseNet foi proposta por [Huang et al. \(2017\)](#). Neste modelo, cada camada recebe todas as camadas anteriores como sua entrada, enquanto sua saída é alimentada em todas as camadas posteriores. As conexões densas no DenseNet aliviam os problemas de desaparecimento e explosão de gradiente e facilitam a reutilização de recursos. A DenseNet tem diferentes versões, conforme a quantidade de camadas na rede neural. Neste trabalho foi avaliada a DenseNet201, que possui profundidade topológica de 201 camadas e cerca de 20 milhões de parâmetros.

A arquitetura MobileNetV2 ([SANDLER et al., 2018](#)) foi projetada para uso em dispositivos móveis, diminuindo significativamente o número de operações e memória necessárias, mantendo a mesma precisão. A principal contribuição é um módulo de camada residual invertido com gargalo linear. Este módulo recebe como entrada uma representação compactada de baixa dimensão que é primeiro expandida para alta dimensão e filtrada com uma convolução leve em profundidade. Os recursos são posteriormente projetados de volta para um representação de baixa dimensão com uma convolução linear.

[Tan e Le \(2019\)](#) estudaram sistematicamente o dimensionamento das CNN e identificaram que o equilíbrio cuidadoso da profundidade, largura e resolução da rede podem levar a um melhor desempenho. Com base nesta observação, proporam um método que dimensiona uniformemente todas as dimensões de profundidade, largura e resolução usando um coeficiente de dimensionamento simples, mas altamente eficaz. Foi obtida uma família de modelos, chamada EfficientNets, que alcança precisão e eficiência muito melhores do que as CNNs anteriores. Neste trabalho foi utilizada a rede EfficientNetB0.

3.4 Transferência de Aprendizagem e Ajuste Fino

Técnicas de transferência de aprendizagem 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 5, é apresentado um exemplo da técnica de transferência de aprendizado por ajuste fino.

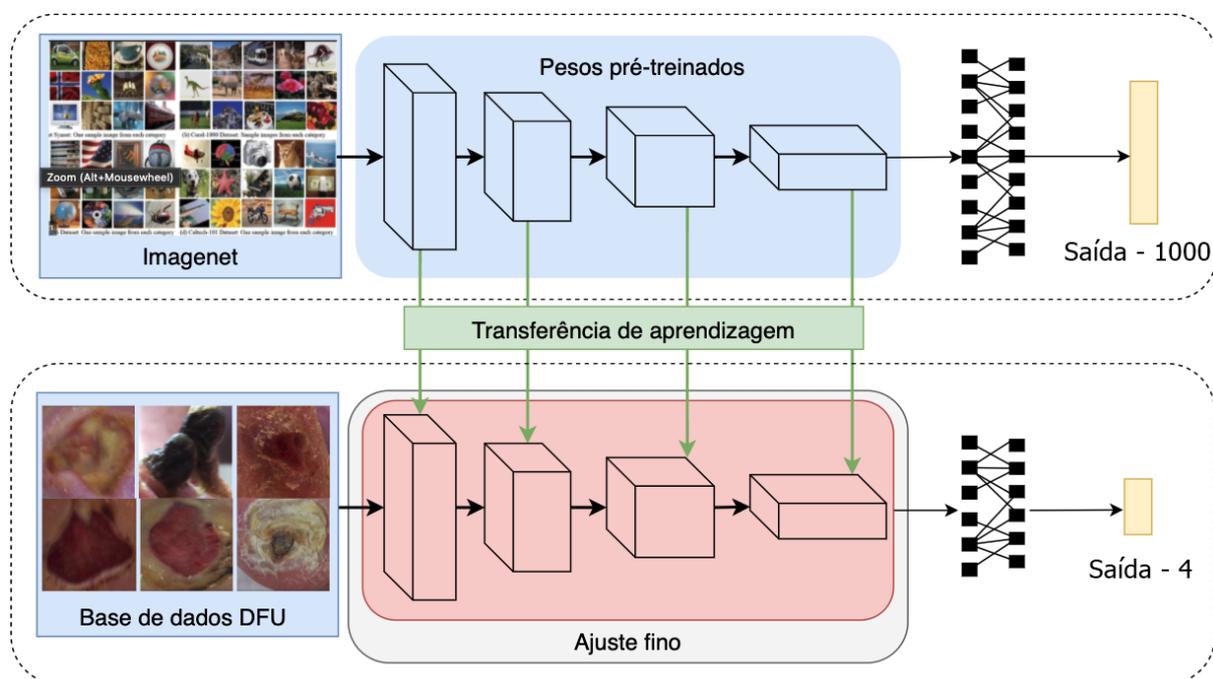


Figura 5 – Fluxograma exemplificando a técnica de transferência de aprendizagem por ajuste fino.

De acordo com [Izadyazdanabadi 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, 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.

Neste trabalho, no processo de ablação, primeiramente realizamos os refinamentos das redes avaliadas utilizando os dois ajustes finos, SFT e DFT, e constatamos que o DFT foi mais adequado ao problema proposto, visto que com este ajuste, obtivemos métricas com valores maiores.

3.5 Dropout e Batch Normalization

Apresentada em 2012 por [Hinton et al. \(2012\)](#), a técnica dropout objetiva evitar o problema de *overfitting* no pequeno conjunto de dados de treinamento ao treinar uma rede profunda de várias camadas com muitas camadas ocultas de nós de unidade não

linear (LEE; SONG, 2019). O *dropout* é uma técnica que evita o *overfitting* e fornece uma maneira de combinar muitas arquiteturas de rede neural diferentes de forma exponencial e eficiente (SRIVASTAVA et al., 2014). O *dropout* consiste em descartar unidades em uma rede neural, ou seja remover temporariamente uma unidade e todas conexões de entrada e saída, conforme mostra a Figura 6.

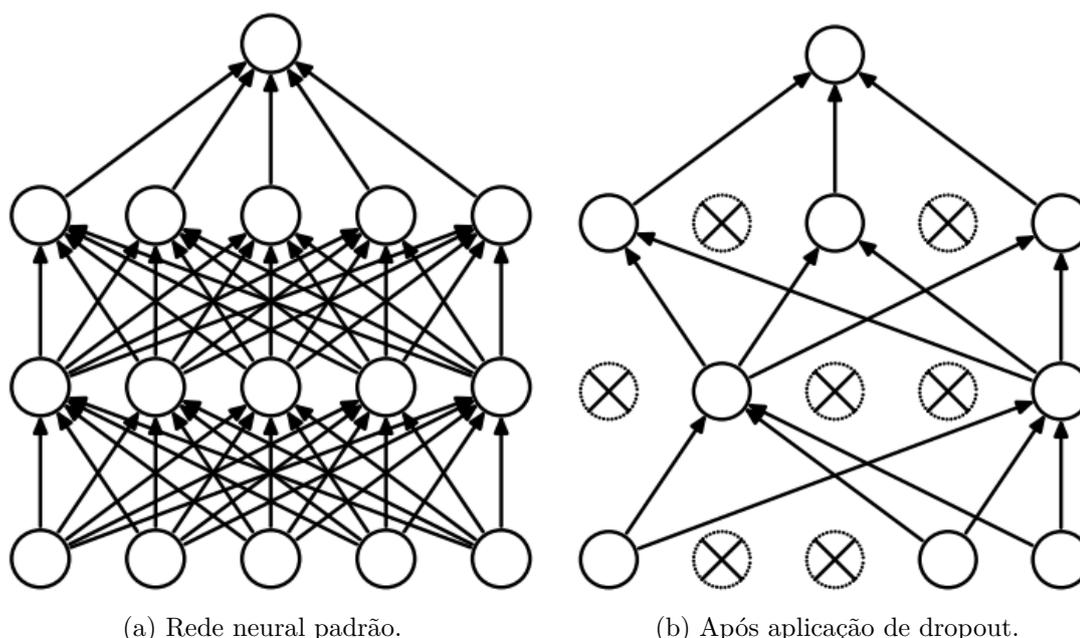
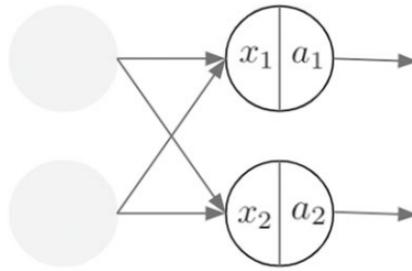


Figura 6 – Exemplo comparativo de uma rede neural sem *dropout* e outra com aplicação de *dropout*. Fonte: (SRIVASTAVA et al., 2014)

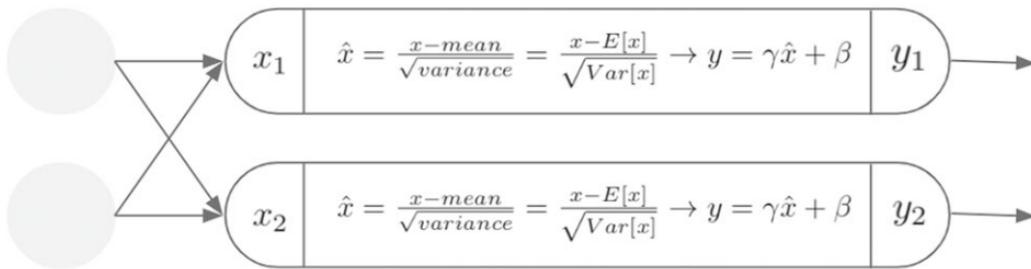
No treinamento da rede, embora o otimizador *Stochastic Gradient Descent* (SGD) com minilote seja simples e eficaz, ele necessita de um ajuste cuidadoso dos hiperparâmetros e dos valores iniciais dos parâmetros do modelo. Caso contrário, a convergência do treinamento será dificultada, pois a rede neural começará com a maioria das ativações saturadas. Além disso, as entradas de cada camada são afetadas pelos parâmetros de todas as camadas anteriores, amplificando pequenas alterações nos parâmetros da rede à medida que a mesma se torna mais profunda.

Dessa forma, o SGD retarda o treinamento, exigindo taxas de aprendizado mais baixas e inicialização cuidadosa dos parâmetros, e torna notoriamente difícil treinar modelos com não linearidades de saturação, o que afeta a robustez das redes e desafia suas aplicações extensas. Segundo Wu et al. (2018), estes problemas são resolvidos normalizando as entradas das camadas, técnica conhecida como *batch normalization*, que realiza a normalização em cada minilote de treinamento.

Segundo Garbin, Zhu e Marques (2020), o tempo de treinamento é reduzido significativamente com a normalização em lote, normalizando a entrada de cada camada na rede, não apenas a camada de entrada. Esta abordagem permite o uso de taxas de



(a) Rede neural padrão.



(b) Após aplicação de batch normalization.

Figura 7 – Exemplo comparativo de uma rede neural sem *batch normalization* e outra com aplicação de *batch normalization*. Fonte: (GARBIN; ZHU; MARQUES, 2020)

aprendizagem mais altas, o que por sua vez reduz o número de etapas de treinamento que a rede precisa para convergir. A Figura 7a exibe uma rede que utiliza os valores da camada anterior inalterados e a Figura 7b exibe uma camada com *batch normalization*, onde os valores de entrada são normalizados para ter uma média de zero e variância de um.

3.6 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 3.1, 3.2, 3.3, 3.4 e 3.5) da classificação:

$$A = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.1)$$

$$P = \frac{VP}{VP + FP} \quad (3.2)$$

$$R = \frac{VP}{VP + FN} \quad (3.3)$$

$$F1 - score = 2 * \frac{P * R}{P + R} \quad (3.4)$$

$$K = \frac{observado - esperado}{1 - esperado} \quad (3.5)$$

O índice kappa (K) considera todos os elementos da matriz de confusão e funciona como uma medida de associação usada para descrever e testar o grau de concordância (confiabilidade e precisão) na classificação. De acordo com [Landis e Koch \(1977\)](#), este índice assume valores entre 0 e 1, sendo o resultado qualificado de acordo com a Tabela 4.

Tabela 4 – Significado do Valor de K

Valor de K	Significado
$K \leq 0,2$	Ruim
$0,2 < K \leq 0,4$	Moderado
$0,4 < K \leq 0,6$	Bom
$0,6 < K \leq 0,8$	Muito bom
$K > 0,8$	Excelente

3.7 Considerações Finais

Este capítulo expôs as bases de imagens utilizadas, as características das redes CNNs avaliadas e as técnicas de treinamento utilizadas. Apresentou-se também a metodologia de avaliação de desempenho das CNNs.

4 Método Proposto

Objetivando melhorar as métricas obtidas pelas redes sequenciais na classificação das imagens do pé diabético e considerando que a normalização em lote permite uma aprendizagem mais alta, reduzindo o número de etapas para a rede convergir, este trabalho propõe um processo de ablação nas redes VGG-16 e VGG-19 com inclusão de camadas *batch normalization*.

As redes VGGS foram escolhidas para este processo de ablação por apresentarem bons resultados nos testes realizados, serem simples de alterar e não apresentarem camadas de normalização em suas arquiteturas originais. Conforme mostra a Figura 8, as redes VGG-16 e VGG-19 tem arquiteturas semelhantes, ambas possuem cinco blocos com camadas convolucionais, no entanto, nos blocos três, quatro e cinco a VGG-19 apresenta uma camada a mais.

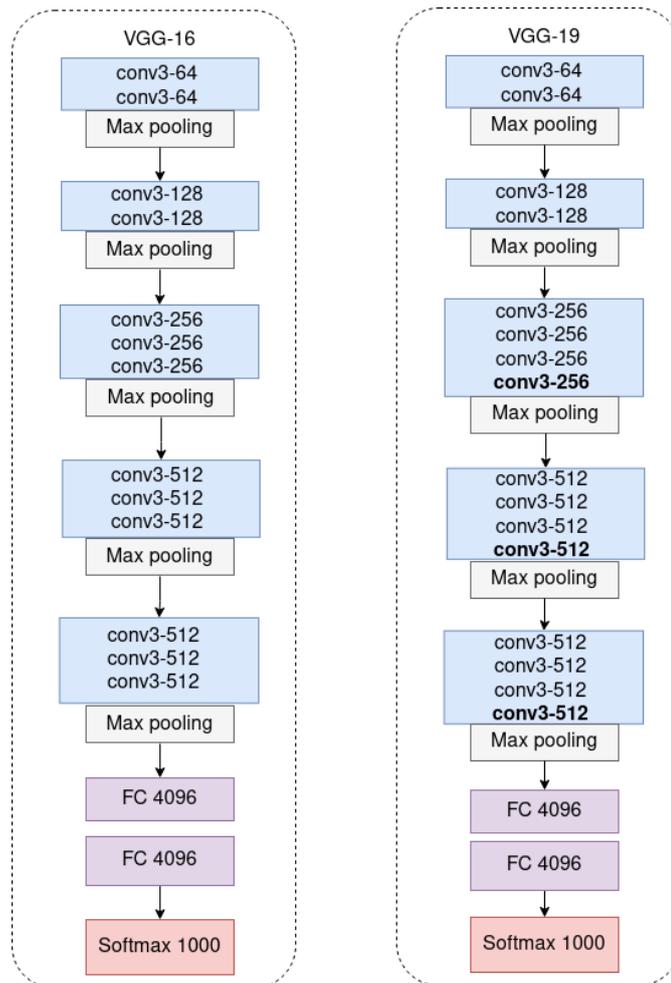


Figura 8 – Comparação das arquiteturas das redes VGG-16 e VGG-19.

4.1 DFU-VGG Proposta

A DFU-VGG proposta é uma rede neural convolucional com a rede VGG-19 em *backbone* e com operações de normalização em lote introduzidas após os blocos convolucionais e camadas densas com menor dimensionalidade, conforme mostra a Figura 9. Observamos que foram adicionadas cinco camadas *batch normalization*, uma em cada bloco de camadas convolucionais, antes da camada *max pooling* e as camadas totalmente conectadas foram substituídas por uma camada *global average pooling* e uma camada totalmente conectada com 512 unidades e a camada *softmax* foi ajustada para 4 unidades, que correspondem a quantidade de classes trabalhadas.

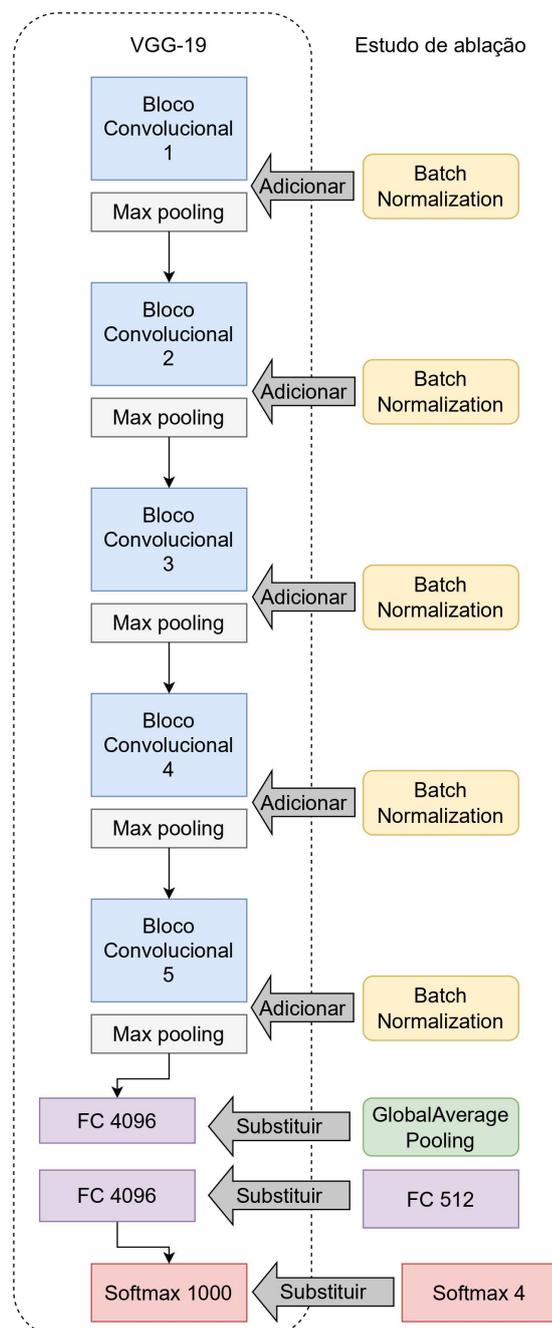


Figura 9 – Arquitetura da rede VGG-19 com inclusão de camadas batch normalization.

A camada *Global Average Pooling* é uma operação de pooling que obtém a média de cada mapa de recursos e o vetor resultante, foi projetada para substituir camadas totalmente conectadas em CNNs clássicas. Uma vantagem desta camada sobre as camadas totalmente conectadas é que ele é mais nativa a estrutura de convolução, impondo correspondências entre mapas de recursos e categorias. Esta configuração de rede obteve os melhores resultados na classificação das imagens de úlceras do pé diabético, por isso, esta é a nossa DFU-VGG proposta. A rede foi refinada com 500 épocas e 64 passos por época.

4.2 Considerações Finais

Este capítulo expôs os estudos de ablação realizados nas redes VGG-16 e VGG-19 com a inclusão das camadas de normalização. Dentre as configurações testadas, os melhores resultados foram com a rede VGG-19, sendo esta a nossa DFU-VGG proposta. Os resultados dos testes realizados com esta proposta são apresentados na Subseção 5.3.

5 Resultados e Discussão

Nos testes iniciais, as arquiteturas de CNN's avaliadas foram treinadas com 100 épocas e 64 passos por época. Observou-se que esta quantidade de épocas não foi suficiente para a aprendizagem, por isso, os demais testes foram realizados com 500 épocas. Foram realizados refinamentos SFT e DFT nas redes com alteração das camadas totalmente conectadas, onde, constatou-se que os resultados obtidos com DFT foram melhores que os resultados SFT. Por isso, os demais testes foram realizados utilizando apenas o DFT. A Tabela 5 mostra as métricas obtidas com SFT e DFT para as redes com uma camada de 512 unidades.

Tabela 5 – Resultados da classificação com SFT e DFT.

SFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K(%)
VGG-16 512	64,11	62,22	64,11	62,92	38,80
VGG-19 512	63,93	61,72	63,93	62,49	38,84
ResNet50 512	54,03	55,48	54,03	51,46	20,23
InceptionV3 512	68,97	68,50	68,97	68,53	48,21
InceptionResNetV2 512	73,64	73,35	73,64	73,14	55,95
DenseNet201 512	74,98	75,13	74,98	74,87	58,34
MobileNetV2 512	73,52	73,07	73,52	73,02	55,84
EfficientNetB0 512	46,50	21,63	46,50	29,52	0,0
DFT					
Arquitetura	A(%)	P(%)	R(%)	F1-Score(%)	K(%)
VGG-16 512	84,21	84,32	84,21	84,22	74,11
VGG-19 512	76,74	76,98	76,74	76,58	61,40
ResNet50 512	80,14	80,06	80,14	80,00	67,19
InceptionV3 512	84,76	84,87	84,76	84,76	75,17
InceptionResNetV2 512	76,07	76,56	76,07	75,85	60,18
DenseNet201 512	85,48	85,51	85,48	85,44	76,06
MobileNetV2 512	76,44	76,44	76,44	76,16	60,71
EfficientNetB0 512	56,28	48,85	56,28	52,30	22,30

Provavelmente uma das técnicas mais populares para estimar o erro de um classificador é a validação cruzada k-fold (k-cv), onde o conjunto de dados é dividido em k subconjuntos, um classificador é aprendido usando um subconjunto e o valor de erro é calculado testando o classificador no subconjunto restante. Finalmente, a estimativa k-cv do erro é a valor médio dos erros cometidos em cada subconjunto. Assim, o k-cv estimador de erro depende de dois fatores: o conjunto de treinamento e o partição em subconjuntos (RODRIGUEZ; PEREZ; LOZANO, 2010). Neste trabalho, as duas bases de imagens DFU foram unificadas e depois divididas em 5 subconjuntos, ou seja, 5-fold,

cada um dos subconjuntos foi dividido numa proporção de 80% para treinamento e 20% para testes. Foram realizados o refinamento e a validação em cada um dos subconjuntos e depois extraídas as métricas. Portanto, os resultados exibidos nas tabelas das próximas subseções correspondem a média aritmética simples das métricas obtidas na classificação de cada um dos cinco subconjuntos. Para facilitar o entendimento das tabelas, os valores do índice kappa foram multiplicados por 100.

5.1 Resultados com as Redes em suas Configurações Originais

Para atender ao objetivo de encontrar uma CNN adequada a classificação das Úlceras do Pé Diabético, inicialmente, foi realizado o refinamento das redes VGG-16, VGG-19, InceptionV3, ResNet50, InceptionResNetV2, DenseNet201, MobileNetV2 e EfficientNetB0 em suas configurações originais. Durante o refinamento das redes VGG-16 e VGG-19, foi necessário redimensionar as imagens de entrada para 224x224 pixels, uma vez que as imagens de entrada com tamanho 112x112 pixels gerava erro na camada Flatten, camada presente na arquitetura original das redes VGGs e que realiza um “achatamento”, ou seja, transforma a entrada em uma matriz de uma dimensão. As demais redes, como não tem a camada Flatten em sua arquitetura original, não apresentaram esta restrição.

Todos as redes exibidas na Tabela 6 foram refinadas com o tamanho da imagens de entrada em 224x224 pixels. Os resultados obtidos com as redes em suas configurações originais em 500 épocas são exibidos na Tabela 6, que tem um índice kappa entre 40% e 60%, ou seja, considerado bom. A InceptionV3 obteve índice kappa moderado. A DenseNet201 obteve melhor acurácia em relação as demais e o maior desvio padrão de acurácia foi de 2,51% na classificação com a VGG-16.

Tabela 6 – Resultado das classificações obtidas com as CNNs originais (melhores valores em negrito).

CNN	A(%)	P(%)	R(%)	F(%)	K(%)
VGG-16	71,95±2,51	63,44±2,68	71,95±2,51	67,23±2,51	50,42±4,60
VGG-19	73,17±0,92	65,05±0,64	73,17±0,92	68,52±0,82	52,71±1,58
ResNet50	74,23±0,59	65,39±0,56	74,23±0,59	69,36±0,54	54,47±1,03
InceptionV3	57,66±6,19	60,95±7,20	57,66±6,19	50,10±8,96	23,27±12,58
InceptionResNetV2	51,14±5,14	44,94±6,54	51,14±5,14	40,51±8,00	9,74±10,02
DenseNet201	75,06±0,71	66,58±0,26	75,06±0,71	70,24±0,58	56,03±1,10
MobileNetV2	61,53±0,07	53,87±0,06	61,53±0,07	55,19±0,11	30,87±14,90
EfficientNetB0	46,56±0,10	21,68±0,10	46,56±0,10	29,59±0,10	0±0

Os gráficos da Figura 10 permitem comparar o treinamento da rede que apresentou pior resultado na classificação, EfficientNetB0 (Figura 10a) e da rede que apresentou o melhor desempenho, DenseNet201 (Figura 10b). Observamos que a EfficientNetB0

apresentou um gráfico linear para acurácia de validação, não ocorrendo aumento do valor da mesma com o aumento do número de épocas. A DenseNet201, a medida que aumentava o número de épocas, o treinamento apresentou um aumento da acurácia de validação e uma diminuição da perda.

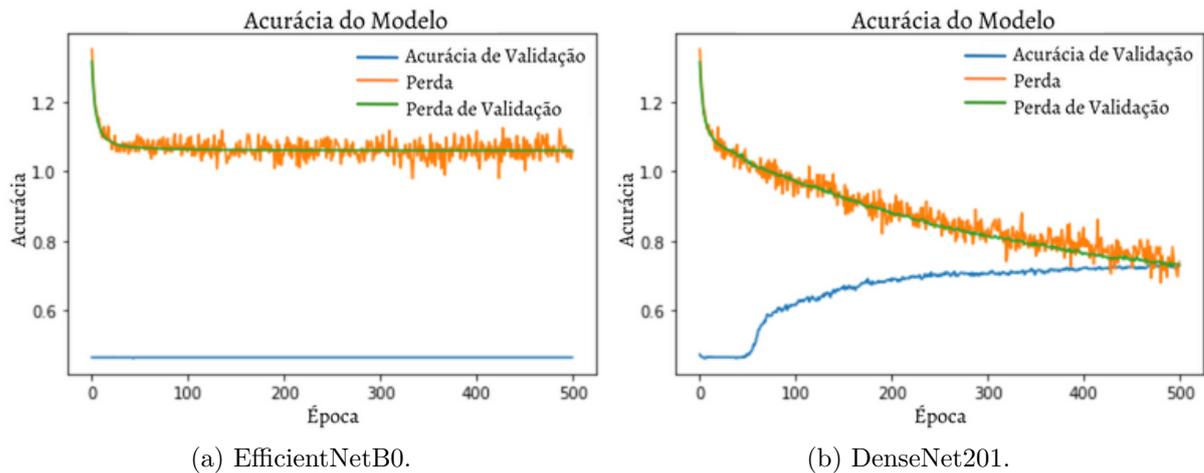


Figura 10 – Gráficos de treinamento das CNNs EfficientNetB0 (a) e DenseNet201 (b).

5.2 Resultados com Alterações das Camadas Totalmente Conectadas

Mantendo as camadas convolucionais das redes, foi inserida uma camada de *Global Average Pooling* e em seguida as camadas totalmente conectadas em dois cenários: (1) uma camada conectada com a quantidade de neurônios assumindo os seguintes valores: 256, 512 e 1024, como mostra a Figura 11 e (2) duas camadas totalmente conectadas com 512-256, 1024-256 e 1024-512 neurônios para as respectivas camadas densas, conforme exibido na Figura 12.

Foram realizados refinamentos com 50 épocas para todas estas configurações de redes, onde observou-se que era necessário mais épocas para que as redes pudessem obter melhores resultados nas classificações. No entanto, algumas configurações de rede foram melhores, dessa forma, optou-se por testar estas configurações com 500 épocas. Os resultados do treinamento com 500 épocas são apresentados na Tabela 7. A VGG-19 com uma camada densa de 512 unidades obteve melhor acurácia e um desvio padrão pequeno, o que indica que esta rede obteve acurácias bem próximas da acurácia média para todos os cinco subconjuntos.

Os gráficos da Figura 13 permitem a comparação entre o treinamento da rede que apresentou pior resultado na classificação, EfficientNetB0 com uma camada de 512 unidades (Figura 13a) e a rede que apresentou o melhor desempenho, VGG-19 com uma

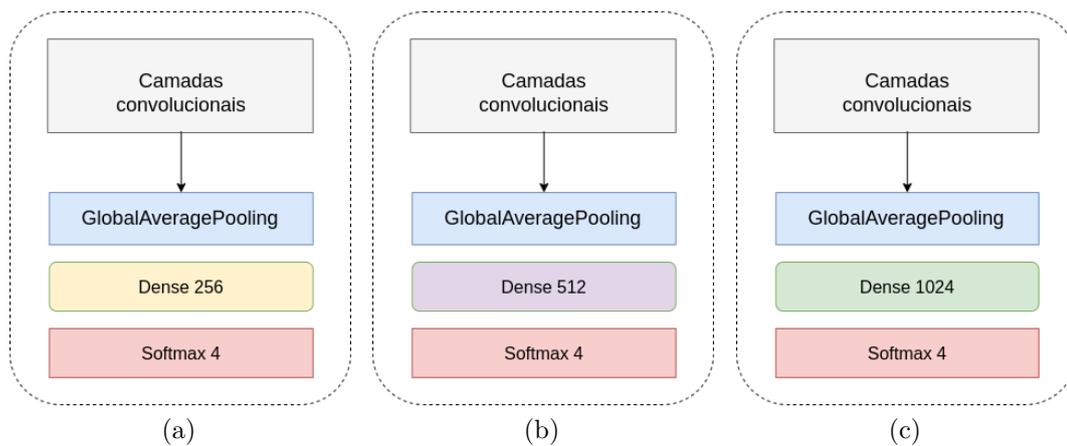


Figura 11 – Processo de ablação com uma camada com 256 (a), 512 (b) ou 1024 (c) unidades.

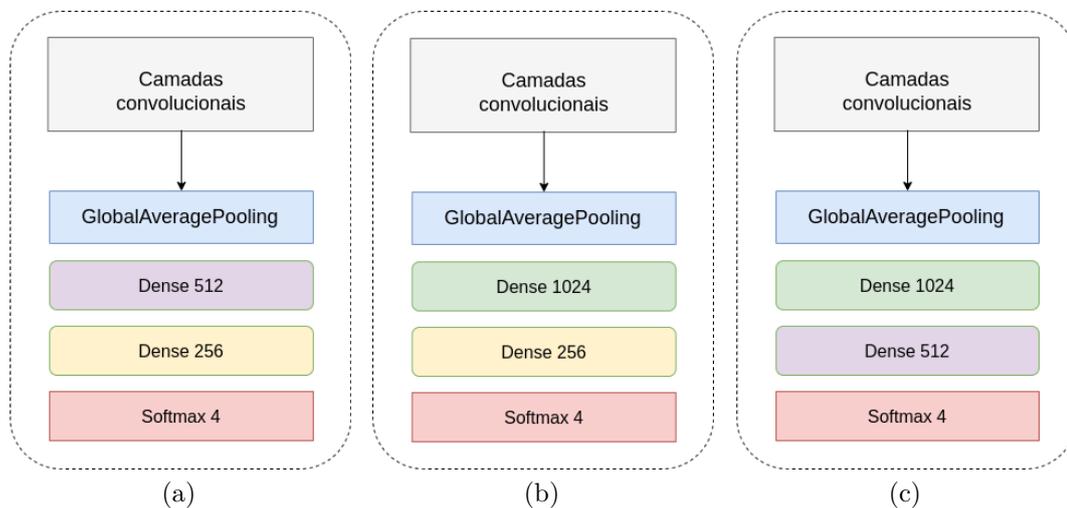


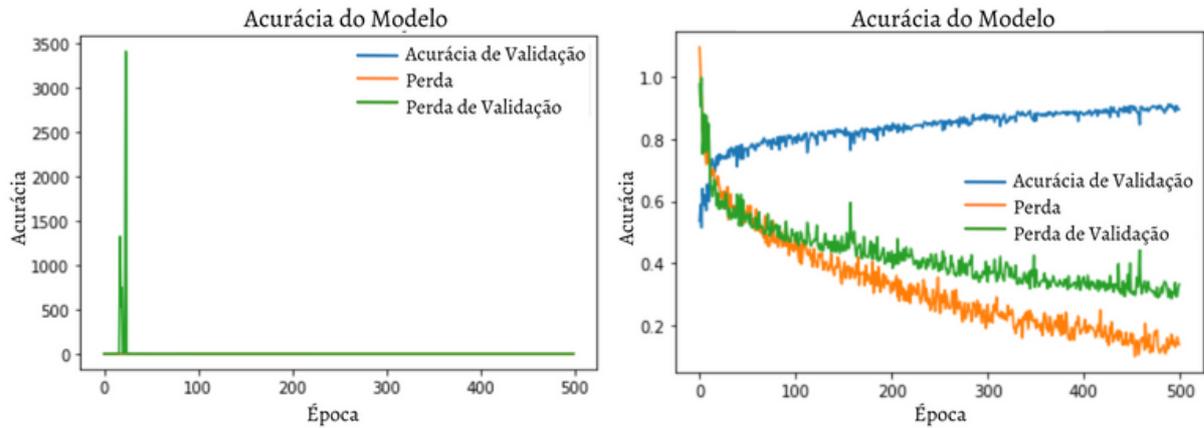
Figura 12 – Processo de ablação com duas camadas totalmente conectadas: uma de 512 e outra de 256 (a), uma de 1024 e outra de 256 (b) e uma de 1024 e outra de 512 (c) unidades.

Tabela 7 – Melhores resultados da classificação com as CNNs com alteração das camadas densas.

CNN	$A(\%)$	$P(\%)$	$R(\%)$	$F(\%)$	$K(\%)$
VGG-16 512	90,55±0,44	90,60±0,42	90,55±0,44	90,54±0,45	84,47±0,72
VGG-19 512	91,40±0,35	91,45±0,37	91,40 ±0,35	91,40±0,37	85,88±0,59
ResNet50 512	90,24±0,75	90,37±0,70	90,24±0,75	90,25±0,75	83,98±1,22
InceptionV3 512	81,60±1,03	81,97±0,88	81,60±1,03	81,54±1,04	69,56±1,72
InceptionResNetV2 256	85,03±1,14	85,12±1,21	85,03±1,14	84,98±1,15	75,28±1,90
DenseNet201 1024-512	90,04±0,64	90,24±0,70	90,04±0,64	90,04±0,65	83,63±1,04
MobileNetV2 512-256	81,43±0,01	81,51±0,01	81,43±0,01	81,33±0,01	69,34±1,25
EfficientNetB0 512	43,12±3,40	22,56±10,00	43,12±3,40	28,79±8,00	16,71±3,34

camada de 512 unidades (Figura 13b). Para a EfficientNetB0, o valor da acurácia de validação se manteve constante do início ao fim do treinamento, conforme observamos no

gráfico da Figura 13a e o valor da perda, inicialmente apresentou uma variação, mas logo este também se manteve constante. Já a VGG-19, a medida que aumentava o número de épocas, apresentou um aumento da acurácia de validação e uma diminuição da perda.



(a) EfficientNetB0 com uma densa de 512 unidades. (b) VGG-19 com uma densa de 512 unidades.

Figura 13 – Gráficos de treinamento da CNNs EfficientNetB0 (a) e VGG-19 (b), ambas com uma camada densa de 512 unidades.

5.3 Resultados das VGGs com Batch Normalization

As camadas de batch normalization foram adicionadas nas redes sequenciais VGG-16 e VGG-19 conforme descrito na Subseção 4.1, onde definimos a rede DFU-VGG. Não houve inserção de camadas de normalização nas demais redes testadas neste trabalho, pois estas possuem um grande número de camadas e apresentam camadas de normalização em suas arquiteturas originais.

A Tabela 8 mostra que as redes VGG-16 com adição da camada de normalização (VGG-16 + BN) e VGG-19 com adição da camada de normalização (DFU-VGG) em um treinamento com 500 épocas obtiveram acurácia, precisão, recall e f1-score acima dos 93% e índice kappa superior a 89%, que segundo a Tabela 4 é considerado excelente. Ambas as redes obtiveram desvio padrão menor que 1%.

Tabela 8 – Resultados da classificação com Batch Normalization.

CNN	$A(\%)$	$P(\%)$	$R(\%)$	$F(\%)$	$K(\%)$
VGG-16 512	90,55±0,44	90,60±0,42	90,55±0,44	90,54±0,45	84,47±0,72
VGG-16 + BN	93,44±0,26	93,46±0,25	93,44±0,26	93,43±0,26	89,21±0,43
VGG-19 512	91,40±0,35	91,45±0,37	91,40 ±0,35	91,40±0,37	85,88±0,59
DFU-VGG	93,45±0,34	93,56±0,30	93,45±0,34	93,46±0,34	89,24±0,58

Dentre os trabalhos do estado da arte (Seção 2), apenas Galdran, Carneiro e Ballester (2021) realizou a classificação em quatro classes, que são as mesmas classificadas

pela DFU-VGG, e obteve precisão de 60,53% utilizando Bit-ResNeXt50. Conforme mostra a Tabela 8, a DFU-VGG obteve precisão de 93,56%.

A DFU-VGG obteve acurácia de 93,45% na classificação das quatro classes e o trabalho de [Veredas, Mesa e Morente \(2010\)](#) obteve acurácia de 91,50% na classificação de tecidos em cinco classes: pele, cicatrização, granulação, descamação e necrose. No entanto, não é possível comparar estes dois trabalhos, visto que ambos utilizaram classes bem diferentes. Outro trabalho que também não permite uma comparação direta é o de [Rostami et al. \(2021\)](#) que utiliza as classes úlceras diabéticas, úlceras venosas e cirúrgicas. Também não é possível uma comparação trivial dos trabalhos que realizam a classificação binária, no entanto, analisando a matriz de confusão (Ver Figura 14) da DFU-VGG vemos a possibilidade de comparar o percentual de acerto da classificação com uma das métricas da classe correspondente obtida nos demais trabalho.

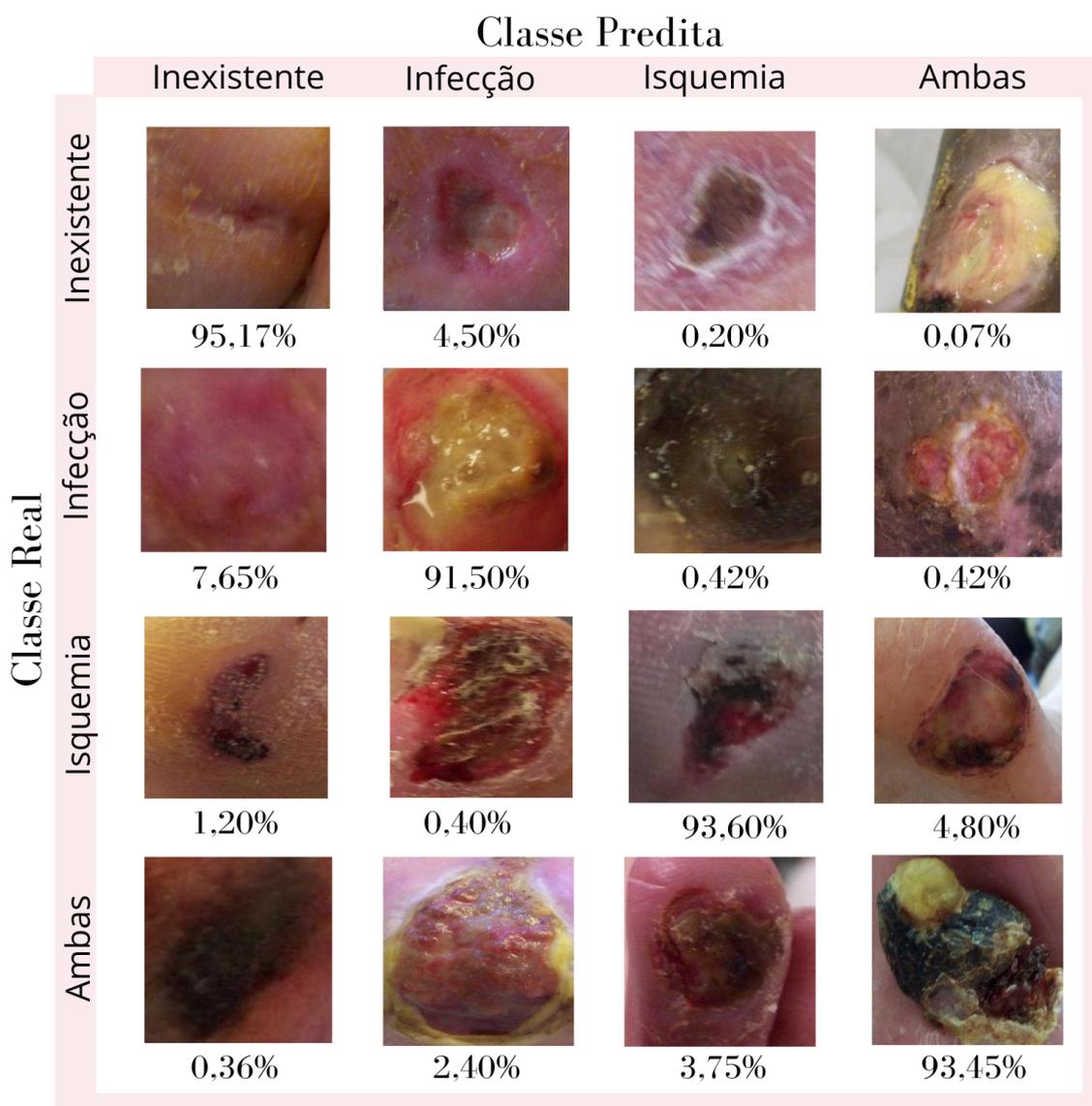


Figura 14 – Matriz de confusão visual da CNN DFU-VGG.

O trabalho de [Goyal et al. \(2020\)](#) obteve acurácia de 73,00% na classificação binária infecção x não infecção (pele normal, úlcera em processo de cicatrização e úlcera apenas com isquemia) e 90,00% na classificação binária isquemia x não isquemia (pele normal, úlcera em processo de cicatrização e úlcera apenas com infecção). [Al-Garaawi et al. \(2022\)](#) obteve em seu trabalho 74,20% de acurácia para a classificação infecção x não infecção e 99,00% para isquemia x não isquemia. [Amin et al. \(2020\)](#) obteve 99,60% de acurácia para infecção x não infecção com árvore de decisão e 97,90% para isquemia x não isquemia com o classificador Naive Bayes. Conforme mostra a Tabela 8, a DFU-VGG obteve 93,45% de acurácia na classificação em quatro classes e a matriz de confusão apresenta que 91,50% das imagens apenas com infecção e 93,60% das imagens apenas isquemia foram classificadas corretamente.

Dentre os resultados apresentados na matriz de confusão (Figura 14), o mais preocupante é a classificação de 7,65% das imagens de infecção como inexistente, o que significa que 7,65% das imagens de úlceras com infecção foram classificadas como úlceras sem infecção e sem isquemia, o que na prática, faria com que esse percentual de pacientes ficassem sem tratamento adequado, acreditando que estas úlceras estavam em processo de cicatrização. Esta taxa de erro justifica a escolha da VGG-19 com *batch normalization* como nossa rede proposta, pois, embora a VGG-16 + BN e DFU-VGG tenham obtido métricas bem próximas, ao analisar a matriz de confusão da VGG-16 + BN (matriz de confusão não mostrada aqui), observamos que a VGG-16 + BN classificou erroneamente como inexistente 8,35% das imagens de infecção e este como já explicado precisa ser minimizado.

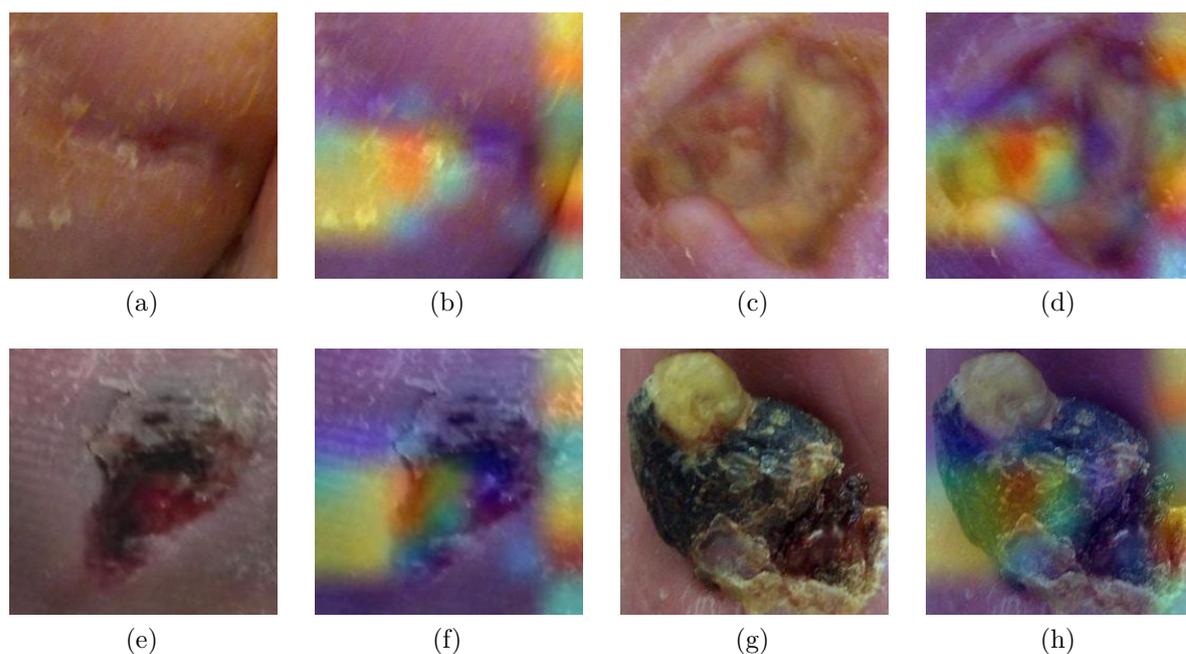
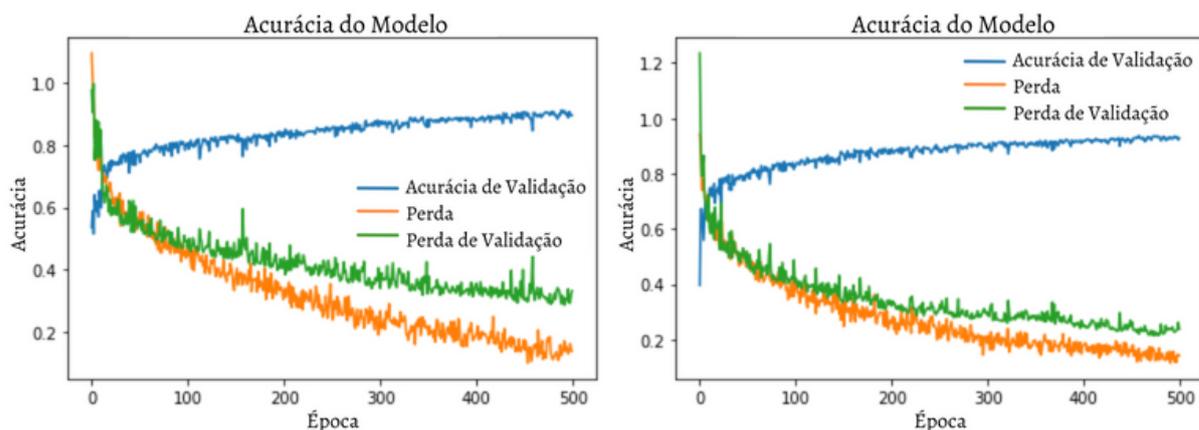


Figura 15 – Mapa de calor com as regiões de ativação para as imagens inexistente ((a) e (b)), infecção ((c) e (d)), isquemia ((e) e (f)) e ambas ((g) e (h)).

Na Figura 15 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 rede DFU-VGG. Como observamos, as regiões de ativação correspondem a uma parcela da úlcera e não a úlcera como um todo e também não há uma diferença visual acentuada entre os mapas de calor, o que sugere que as mesmas regiões foram ativadas para todas as classes.

Os gráficos da Figura 16 permitem a comparação entre o treinamento da rede VGG-19 com uma camada de 512 unidades (Figura 16a) e a DFU-VGG (Figura 16b). Observamos que com o aumento do número de épocas, ambas as redes apresentaram um aumento da acurácia de validação e uma diminuição da perda, no entanto, a rede proposta logo nas primeiras 100 épocas apresenta uma parábola com uma curva maior se comparada a parábola da mesma sem batch normalization.



(a) VGG-19 com uma densa de 512 unidades. (b) DFU-VGG: VGG-19 com uma densa de 512 unidades com batch normalization.

Figura 16 – Gráficos de treinamento da VGG-19 com uma camada densa de 512 unidades (a) e da DFU-VGG (b).

5.4 Resultados das VGGs com Dropout

Considerando que a camada de *dropout* evita o *overfitting*, achamos interessante testar as redes VGGs com a inserção desta camada. Portanto, foi realizada a inserção de camadas dropout nas redes VGG-16 e VGG-19 e estas configurações são denominadas de VGG-16 + DP e VGG-19 + DP, respectivamente. A Figura 17 apresenta a arquitetura da rede VGG-19 após este processo de ablação, onde foi inserida uma camada de *dropout* após cada camada *max pooling*, as camadas totalmente conectadas foram substituídas por uma camada *global average pooling* e uma camada totalmente conectada com 512 unidades e a camada *softmax* foi ajustada para 4 unidades, que correspondem a quantidade de classes trabalhadas.

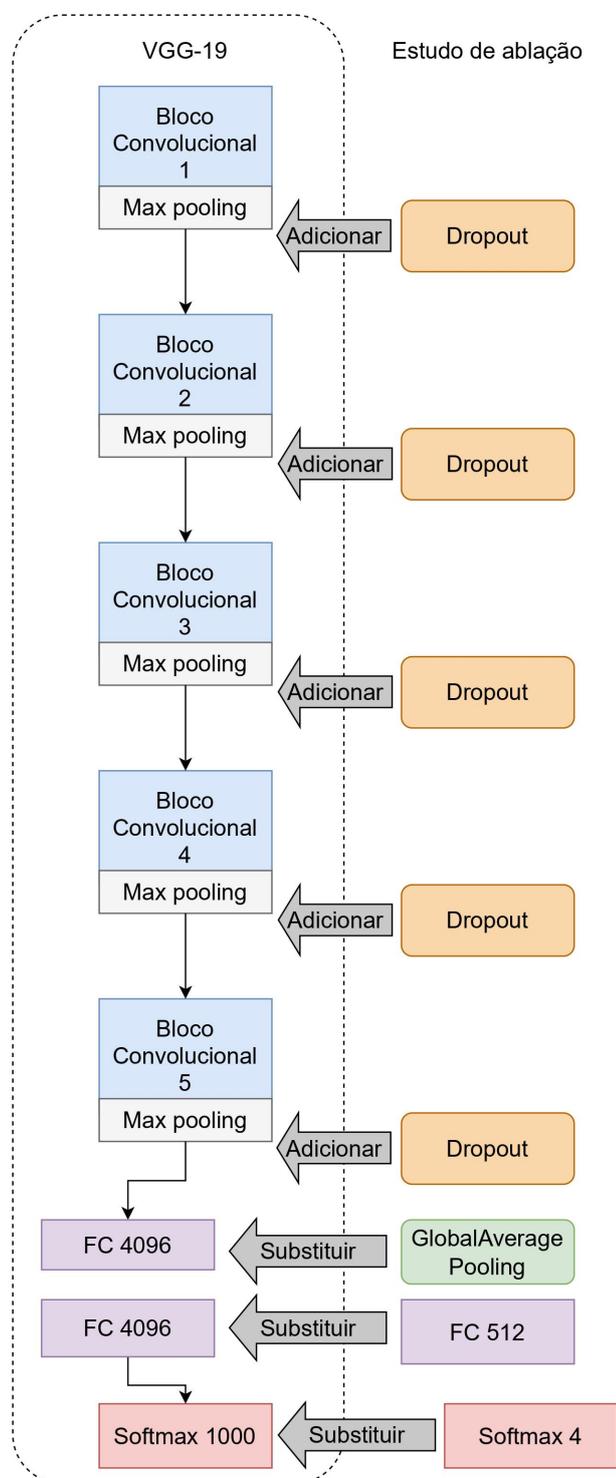


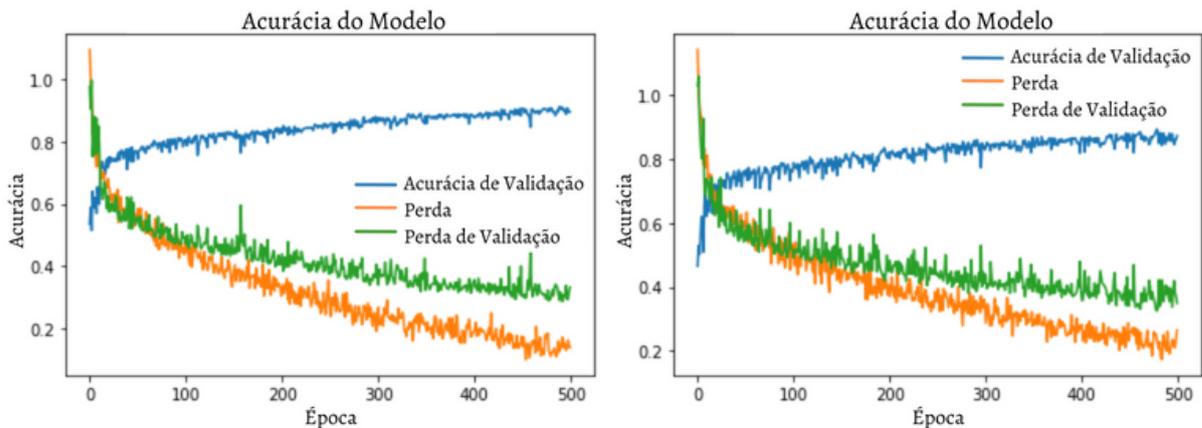
Figura 17 – Arquitetura da rede VGG-19 com inclusão de camadas dropout.

Embora esta abordagem seja uma maneira para melhorar os resultados, nos testes realizados, os resultados obtidos não superaram a rede proposta. A Tabela 9 mostra que os resultados obtidos foram inferiores inclusive aos resultados das redes sem esta camada.

Tabela 9 – Resultados da classificação com adição de camada dropout.

CNN	A(%)	P(%)	R(%)	F(%)	K(%)
VGG-16 512	90,55±0,44	90,60±0,42	90,55±0,44	90,54±0,45	84,47±0,72
VGG-16 + DP	87,89±0,66	87,98±0,60	87,89±0,66	87,89±0,66	80,09±1,15
VGG-19 512	91,40±0,35	91,45±0,37	91,40 ±0,35	91,40±0,37	85,88±0,59
VGG-19 + DP	88,86±0,40	89,00±0,39	88,86±0,40	88,87±0,41	81,70±0,65
DFU-VGG	93,45±0,34	93,56±0,30	93,45±0,34	93,46±0,34	89,24±0,58

Os gráficos da Figura 18 permitem a comparação entre o treinamento da rede VGG-19 com uma camada de 512 unidades (Figura 18a) e desta com adição das camadas de dropout (Figura 18b). Observamos que com o aumento do número de épocas, ambas as redes apresentaram um aumento da acurácia de validação e uma diminuição da perda, no entanto, a rede com dropout não apresentou um melhor desempenho.



(a) VGG-19 com uma densa de 512 unidades. (b) VGG-19 com uma densa de 512 unidades com dropout.

Figura 18 – Gráficos de treinamento da CNN VGG-19 com uma camada densa com 512 unidades sem (a) e com (b) aplicação de dropout.

5.5 Resultados das VGGs com dropout e batch normalization

Considerando que as redes VGGs com *batch normalization* obtiveram um bom desempenho, foram realizados testes com essas redes inserindo nelas camadas de *dropout*, obtendo as redes VGG-16 + BN + DP e VGG-19 + BN + DP. A Figura 19 mostra a arquitetura da VGG-19 com a inserção de camadas *batch normalization* antes da *max pooling* e após esta, inserção de *dropout*. As camadas totalmente conectadas foram substituídas por

uma camada *global average pooling* e uma camada totalmente conectada com 512 unidades e a camada *softmax* foi ajustada para 4 unidades, que correspondem a quantidade de classes trabalhadas.

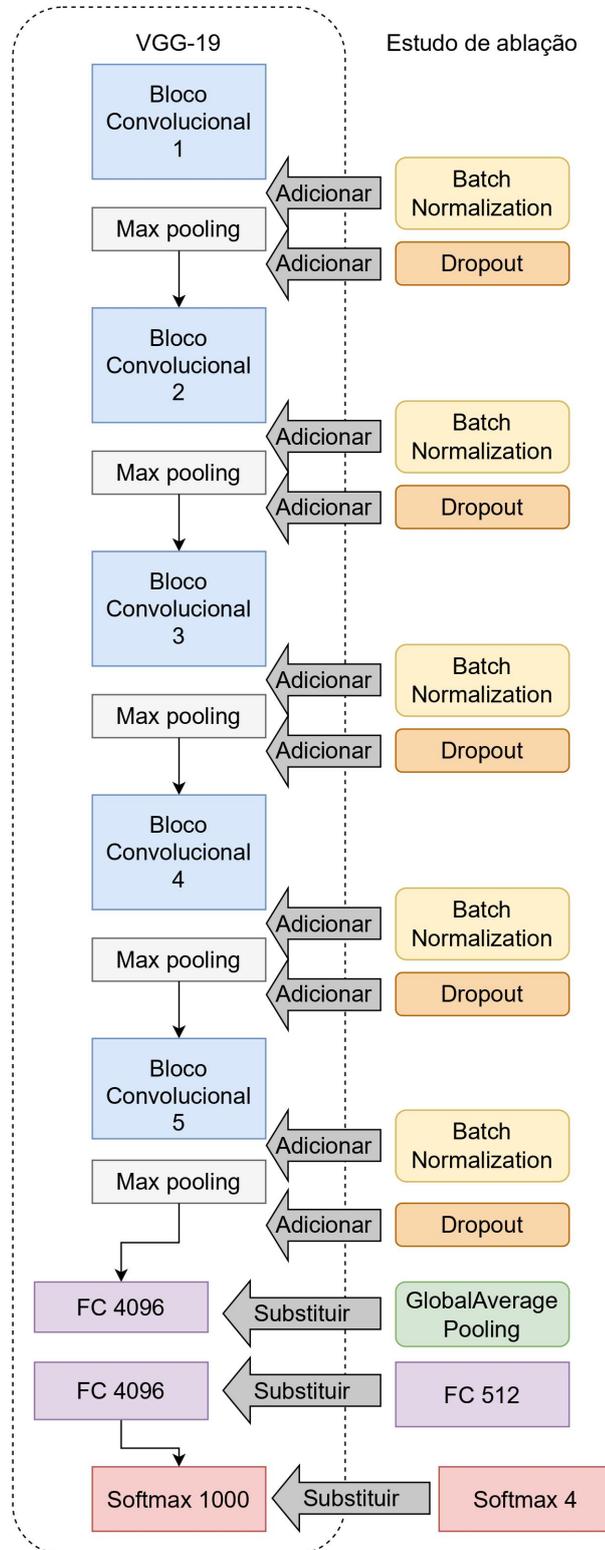


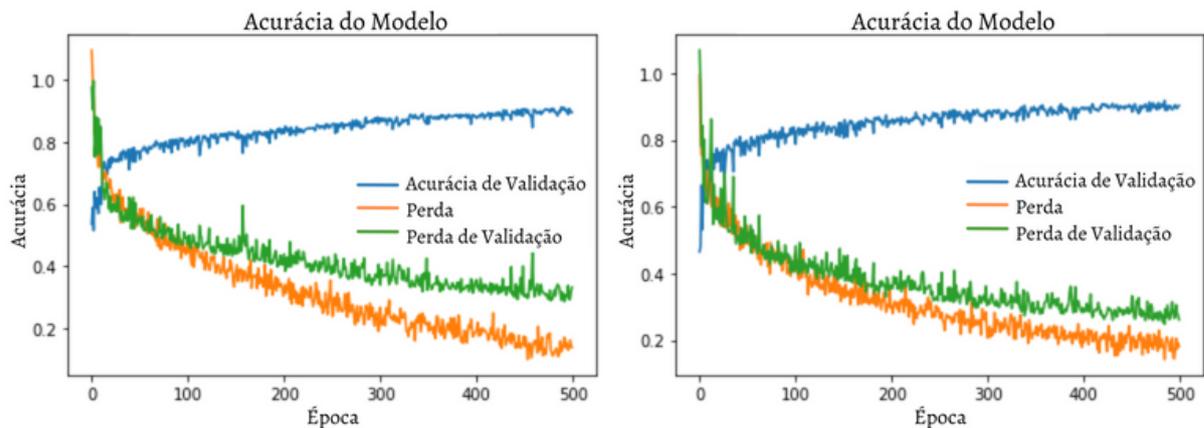
Figura 19 – Arquitetura da rede VGG-19 com inclusão de camadas dropout e batch normalization.

Ao utilizar as duas técnicas, dropout e batch normalization, observou-se que os resultados mostrados na Tabela 10 foram superiores aos obtidos pelas redes com alterações das camadas totalmente conectadas (Ver Tabela 7), no entanto, não superaram os resultados obtidos pela DFU-VGG.

Tabela 10 – Resultados da classificação com Batch Normalization e Dropout.

CNN	A(%)	P(%)	R(%)	F(%)	K(%)
VGG-16 512	90,55±0,44	90,60±0,42	90,55±0,44	90,54±0,45	84,47±0,72
VGG-16 + BN + DP	92,30±0,37	92,40±0,33	92,30±0,37	92,30±0,36	87,34±0,61
VGG-19 512	91,40±0,35	91,45±0,37	91,40 ±0,35	91,40±0,37	85,88±0,59
VGG-19 + BN + DP	92,36±0,56	92,46±0,54	92,36±0,56	92,36±0,56	87,43±0,93
DFU-VGG	93,45±0,34	93,56±0,30	93,45±0,34	93,46±0,34	89,24±0,58

Os gráficos da Figura 20 permitem a comparação entre o treinamento da rede VGG-19 com uma camada de 512 unidades (Figura 20a) e desta com adição das camadas de dropout e batch normalization (Figura 20b). Observamos que com o aumento do número de épocas, ambas as redes apresentaram um aumento da acurácia de validação e uma diminuição da perda, no entanto, a rede com batch normalization logo nas primeiras 100 épocas apresenta uma parábola com uma curva maior se comparada a parábola da mesma sem batch normalization.



(a) VGG-19 com uma densa de 512 unidades. (b) VGG-19 com uma densa de 512 unidades com dropout e batch normalization.

Figura 20 – Gráficos de treinamento da CNN VGG-19 sem (a) e com (b) aplicação de dropout e batch normalization.

6 Conclusões e Trabalhos Futuros

Neste trabalho, foram refinadas oito arquiteturas de CNNs (VGG-16, VGG-19, ResNet50, InceptionV3, InceptionResNetV2, DenseNet201, MobileNetV2 e EfficientNetB0) para classificação de úlceras do pé diabético, considerando as quatro classes: inexistente (que contém imagens de pele saudável, úlceras em processo de cicatrização e de úlceras sem isquemia ou infecção), isquemia, infecção e ambas (úlceras com isquemia e infecção).

Os experimentos com as CNNs em suas configurações originais apresentaram um baixo valor para o índice kappa, sendo que a DenseNet201 apresentou um kappa de 56,03% que é considerado bom, enquanto as redes InceptionResNetV2 e EfficientNetB0 apresentaram um valor kappa inferior a 20%, o que é considerado ruim. Em relação as demais métricas, as redes testadas obtiveram valores baixo, destacando-se a DenseNet201 com a maior acurácia de 75,06%.

Com estes dados, conclui-se que as redes em suas configurações originais não são adequadas para o problema de classificação de imagens DFU. Por outro lado, após o processo de ablação, estas redes apresentaram melhores resultados em relação a todas as métricas, com destaque para a VGG-19 com uma camada de 512, que obteve acurácia de 91,40% e um kappa de 85,88%, com exceção da EfficientNetB0 que continuou com um kappa considerado ruim. Neste processo de ablação, as redes foram refinadas com os ajustes finos SFT e DFT, onde, constatou-se que o DFT foi mais eficiente que o SFT.

O processo de ablação para as redes VGGs foi além das alterações das camadas totalmente conectadas, sendo realizadas alterações na arquitetura interna destas redes. Com a adição de camadas de *dropout* as redes VGGs apresentaram melhores resultados se comparados ao desempenho obtido anteriormente por estas redes em suas configurações originais, no entanto, tais resultados foram inferiores aos obtidos pelas redes VGGs com uma camada densa com 512 unidades.

A adição de camadas *batch normalization* melhorou os resultados obtidos pelas redes VGGs, sendo que a VGG-19 com normalização obteve métricas melhores que todas as obtidas pelas demais configurações de CNNs testadas, por isso, propomos esta configuração como nossa CNN, denominada DFU-VGG. A comparação dos resultados da DFU-VGG com os trabalhos do estado da arte não é trivial, pois a maioria destes utiliza número de classes diferentes. Apenas um trabalho realizou a classificação em quatro classe e obteve uma precisão de 61,40%, enquanto a DFU-VGG obteve 93,56% de precisão e índice kappa de 89,24%, que é considerado excelente.

6.1 Trabalhos Futuros

Diante dos resultados obtidos, acredita-se que é possível definir uma CNN específica para a classificação de imagens DFU que apresente um desempenho melhor que a rede DFU-VGG proposta. Sendo assim, como trabalhos futuros, investigaremos a combinação de redes para a classificação de imagens DFU, realizando primeiramente uma classificação binária, normal x anormal, e posteriormente a classificação em quatro classes. Com esta abordagem, pretende-se diminuir a quantidade de imagens com infecção que são classificadas como inexistentes.

Referências

- ACHANTA, R. et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 34, n. 11, p. 2274–2282, 2012. Citado 2 vezes nas páginas 5 e 6.
- AL-GARAAWI, N. et al. Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks. *Computers in Biology and Medicine*, v. 140, p. 105055, 2022. ISSN 0010-4825. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010482521008490>>. Citado 3 vezes nas páginas 8, 9 e 29.
- ALZUBAIDI, L. et al. Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network. *Multimedia Tools and Applications*, Springer Netherlands, v. 79, n. 21-22, p. 15655–15677, 2020. Citado 2 vezes nas páginas 5 e 9.
- AMIN, J. et al. An integrated design for classification and localization of diabetic foot ulcer based on cnn and yolov2-dfu models. *IEEE Access*, IEEE, v. 8, p. 228586–228597, 2020. Citado 3 vezes nas páginas 6, 9 e 29.
- CAVANAGH, P. R. et al. Treatment for diabetic foot ulcers. *The Lancet*, Elsevier, v. 366, n. 9498, p. 1725–1735, 2005. Citado na página 1.
- CHOLLET, F. et al. *Keras*. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>. Citado na página 13.
- DAS, S. K.; ROY, P.; MISHRA, A. K. Dfu_spnet: A stacked parallel convolution layers based cnn to improve diabetic foot ulcer classification. *ICT Express*, Elsevier, 2021. Citado na página 7.
- DAS, S. K.; ROY, P.; MISHRA, A. K. Dfu_spnet: A stacked parallel convolution layers based cnn to improve diabetic foot ulcer classification. *ICT Express*, Elsevier, 2021. Citado na página 9.
- GALDRAN, A.; CARNEIRO, G.; BALLESTER, M. Á. G. Convolutional nets versus vision transformers for diabetic foot ulcer classification. *Lecture Notes in Computer Science*, Springer, Cham, v. 13183, 2021. Disponível em: <https://doi.org/10.1007/978-3-030-94907-5_2>. Citado 3 vezes nas páginas 8, 9 e 27.
- GARBIN, C.; ZHU, X.; MARQUES, O. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, v. 79, p. 1–39, 05 2020. Citado 3 vezes nas páginas 9, 17 e 18.
- GOYAL, M.; HASSANPOUR, S. A refined deep learning architecture for diabetic foot ulcers detection. *CoRR*, abs/2007.07922, 2020. Disponível em: <<https://arxiv.org/abs/2007.07922>>. Citado na página 1.
- GOYAL, M. et al. Dfunet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, v. 4, n. 5, p. 728–739, 2017. Citado 2 vezes nas páginas 6 e 9.

- GOYAL, M. et al. Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. *Computers in Biology and Medicine*, Elsevier, v. 117, p. 103616–103616, 2020. Citado 5 vezes nas páginas 1, 6, 9, 10 e 29.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 14.
- HE, X.; ZEMEL, R. S.; CARREIRA-PERPINÁN, M. A. Multiscale conditional random fields for image labeling. In: IEEE. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004*. [S.l.], 2004. v. 2, p. II–II. Citado na página 5.
- HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012. Disponível em: <<http://arxiv.org/abs/1207.0580>>. Citado na página 16.
- HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 15.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015. Disponível em: <<http://arxiv.org/abs/1502.03167>>. Citado na página 14.
- IZADYYAZDANABADI, M. et al. Convolutional neural networks: ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical cle images. *Journal of Visual Communication and Image Representation*, Elsevier, v. 54, p. 10–20, 2018. Citado na página 16.
- KORNBLITH, S.; SHLENS, J.; LE, Q. V. Do better imagenet models transfer better? In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 2661–2671. Citado na página 13.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, JSTOR, v. 33, n. 1, p. 159–174, 1977. Citado na página 19.
- LEE, H.; SONG, J. Introduction to convolutional neural network using keras; an understanding from a statistician. *Communications for Statistical Applications and Methods*, v. 26, p. 591–610, 11 2019. Citado na página 17.
- RODRIGUEZ, J. D.; PEREZ, A.; LOZANO, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 3, p. 569–575, 2010. Citado na página 23.
- ROSTAMI, B. et al. Multiclass wound image classification using an ensemble deep cnn-based classifier. *Computers in Biology and Medicine*, Elsevier, p. 104536, 2021. Citado 3 vezes nas páginas 7, 9 e 28.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, Springer, v. 115, n. 3, p. 211–252, 2015. Citado 2 vezes nas páginas 13 e 15.

- SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 4510–4520. Citado na página 15.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015. Citado na página 13.
- SOLÍS-SÁNCHEZ, L. O. et al. Identification and evaluation on diabetic foot injury by computer vision. *IEEE International Conference on Industrial Technology (ICIT)*, p. 758–762, 2016. Citado na página 1.
- SRIVASTAVA, N. et al. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, v. 15, n. 56, p. 1929–1958, 2014. Disponível em: <<http://jmlr.org/papers/v15/srivastava14a.html>>. Citado 2 vezes nas páginas 9 e 17.
- SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016. Disponível em: <<http://arxiv.org/abs/1602.07261>>. Citado na página 14.
- SZEGEDY, C. et al. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. Disponível em: <<http://arxiv.org/abs/1409.4842>>. Citado na página 14.
- SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 2818–2826. Citado na página 14.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 6105–6114. Citado na página 15.
- VEREDAS, F.; MESA, H.; MORENTE, L. Binary tissue classification on wound images with neural networks and bayesian classifiers. *IEEE Transactions on Medical Imaging*, v. 2, n. 29, p. 410–427, 2010. Citado 3 vezes nas páginas 5, 9 e 28.
- VOGADO, L. et al. Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 21, n. 9, p. 2989–2989, 2021. Citado na página 13.
- WANG, J.; PEREZ, L. et al. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, v. 11, p. 1–8, 2017. Citado na página 13.
- WANG, L. et al. Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification. *IEEE Transactions on Biomedical Engineering*, IEEE, v. 64, n. 9, p. 2098–2109, 2016. Citado 3 vezes nas páginas 5, 6 e 9.
- WU, S. et al. L1-norm batch normalization for efficient training of deep neural networks. *CoRR*, abs/1802.09769, 2018. Disponível em: <<http://arxiv.org/abs/1802.09769>>. Citado na página 17.
- YOSINSKI, J. et al. How transferable are features in deep neural networks? In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: [s.n.], 2014. v. 2, p. 3320–3328. Citado na página 15.



**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO ELETRÔNICA NO
REPOSITÓRIO INSTITUCIONAL - RI/UFPI**

1. Identificação do material bibliográfico:

- Tese Dissertação Monografia TCC Artigo Livro
 Capítulo de Livro Material Cartográfico ou Visual Música
 Obra de Arte Partitura Peça de Teatro Relatório de pesquisa
 Comunicação e Conferência Artigo de periódico Publicação seriada
 Publicação de Anais de Evento

2. Identificação do Trabalho Científico:

Curso de Graduação: _____

Programa de pós-graduação: Pós-Graduação em Ciência da Computação

Outro: _____

Autor(a): Francisco das Chagas Torres dos Santos

E-mail: fsantos@ufpi.edu.br

Orientador (a) Prof. Rodrigo de Melo Souza Veras

Instituição: Universidade Federal do Piauí

Membro da banca: Profa. Andrea Gomes Campos Bianchi

Instituição: Universidade Federal de Ouro Preto, Minas

Gerais, Brasil

Membro da banca: Prof. André Macedo Santana

Instituição: Universidade Federal do Piauí

Membro da banca: Prof. Vinicius Ponte Machado

Instituição: Universidade Federal do Piauí

Membro da banca: _____

Instituição: _____

Membro da banca: _____

Instituição: _____

Título obtida: Mestre

Data da defesa: 20 / 04 / 2022

Título do trabalho: DFU-VGG, uma Nova e Aprimorada Rede VGG-19 para Classificação de Úlceras do Pé Diabético

Agência de fomento (em caso de aluno bolsista): _____

3. Informações de acesso ao documento no formato eletrônico:

Liberação para publicação:

Total: [x]

Parcial: []. Em caso de publicação parcial especifique a(s) parte(s) ou o(s) capítulos(s) a serem publicados: _____

.....

TERMO DE AUTORIZAÇÃO

Em atendimento ao Artigo 6º da Resolução CEPEX nº 264/2016 de 05 de dezembro de 2016, autorizo a Universidade Federal do Piauí - UFPI, a disponibilizar gratuitamente sem ressarcimento dos direitos autorais, o texto integral ou parcial da publicação supracitada, de minha autoria, em meio eletrônico, no Repositório Institucional (RI/UFPI), no formato especificado* para fins de leitura, impressão e/ou *download* pela *internet*, a título de divulgação da produção científica gerada pela UFPI a partir desta data.

Local: Teresina-PI Data: 10 / 06 / 2022

Assinatura do(a) autor(a): _____

* **Texto** (PDF); **imagem** (JPG ou GIF); **som** (WAV, MPEG, MP3); **Vídeo** (AVI, QT).