Universidade Federal de Pernambuco
Centro de Tecnologia e Geociências
Departamento de Engenharia Biomédica
Programa de Pós-Graduação em Engenharia Biomédica

Pedro Vitor Soares Gomes de Lima

Sistema inteligente para apoio ao diagnóstico de câncer de pele e hanseníase usando análise digital de imagens e redes neurais artificiais profundas

Recife
2023

Pedro Vitor Soares Gomes de Lima

Sistema inteligente para apoio ao diagnóstico de câncer de pele e hanseníase usando análise digital de imagens e redes neurais artificiais profundas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do grau de Mestre em Engenharia Biomédica.

Área de concentração: Computação Biomédica.

Orientador: Prof. Wellington Pinheiro dos Santos

Recife

2023

Catalogação na fonte
Bibliotecário Gabriel Luz, CRB-4 / 2222

L732s Lima, Pedro Vitor Soares Gomes de.
Sistema inteligente para apoio ao diagnóstico de câncer de pele e hanseníase usando análise digital de imagens e redes neurais artificiais profundas / Pedro Vitor Soares Gomes de Lima. 2023.
99 f.: il.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Biomédica, Recife, 2023.
Inclui referências.

1. Engenharia biomédica. 2. Sistema de apoio ao diagnóstico. 3. Hanseníase. 4. Câncer de pele. 5. Redes neurais artificiais profundas. I. Santos, Wellington Pinheiro dos (Orientador). II. Título.

UFPE

610.26 CDD (22. ed.) BCTG / 2023 - 110

Pedro Vitor Soares Gomes de Lima

Sistema inteligente para apoio ao diagnóstico de câncer de pele e hanseníase usando análise digital de imagens e redes neurais artificiais profundas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do grau de Mestre em Engenharia Biomédica.

Aprovado em: 30 de março de 2023.

BANCA EXAMINADORA

Prof. Dr. Wellington Pinheiro dos Santos (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Ricardo Emmanuel de Souza (Examinador Interno)
Universidade Federal de Pernambuco

Prof[a]. Dr[a]. Ellany Gurgel Cosme do Nascimento (Examinadora Externa)
Universidade do Estado do Rio Grande do Norte

AGRADECIMENTOS

À minha família, por sempre acreditar em mim e me apoiar. Em especial, à minha mãe e ao meu pai, que estiveram ao meu lado em todas as etapas da minha vida.

Aos amigos do curso de mestrado, Anielle, Eduarda, Flor, Leandro e Luysa, pela amizade criada apesar da distância das aulas remotas. Aos amigos do grupo de pesquisa, pela colaboração e atenção, em especial, Anderson e Ingrid.

Ao amigo de infância, Alex Augusto, que nos deixou no mesmo período em que este trabalho estava sendo finalizado. Você sempre será minha fonte de inspiração.

Aos professores e servidores do Departamento de Engenharia Biomédica, por todo o suporte ao longo da minha trajetória. Em especial, ao professor Wellington pela sua orientação, paciência e comprometimento.

À banca examinadora, Ellany e Ricardo, por aceitar o convite e pela atenção despendida.

À universidade, pela sua responsabilidade social e também por, apesar das dificuldades, continuar com seu processo de formação do ser humano, do profissional e do cidadão.

A todos que, de alguma maneira, contribuíram para a realização deste trabalho.

RESUMO

O uso de inteligência artificial cresce continuamente em diversos setores, e no setor da saúde não é diferente. Aplicações de inteligência computacional ajudam aos profissionais da saúde no processo de diagnóstico, permitindo que este seja mais rápido e mais confiável. Segundo a Organização Mundial da Saúde (OMS), as doenças negligenciadas são aquelas doenças endêmicas que ocorrem em regiões subdesenvolvidas e que não possuem apelo suficiente para receberem investimentos do Estado ou do mercado. O Brasil é um dos países com mais casos de hanseníase, esta sendo uma das doenças classificadas como negligenciadas, chegando a uma taxa de detecção geral de novos casos de 13,23 por 100 mil habitantes em 2019. Com uma realidade diferente, mas também preocupante, o câncer é tido como uma das principais causas de morte e uma importante barreira para o aumento da expectativa de vida em todos os países do mundo. No Brasil, para cada ano do triênio 2020-2022 estima-se 625 mil novos casos de câncer, sendo o câncer de pele não melanoma o de maior incidência (177 mil). Este trabalho busca contribuir na detecção precoce da hanseníase e do câncer de pele com a criação de uma estrutura de apoio ao diagnóstico utilizando redes neurais artificiais profundas para extração de atributos, combinado a um algoritmo de classificação, de modo a reconhecer as doenças em imagens com lesões de pele. A fim de uma maior assertividade, foram propostos 7 diferentes cenários de experimentos envolvendo as doenças citadas, levando em conta tanto aspectos computacionais quanto relatos de profissionais da saúde. Foram analisadas as redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3 combinadas a diversas configurações do classificador Random Forest. Com base nas métricas de avaliação definidas, observou-se que a arquitetura Inception v3 é aquela que normalmente gera melhores resultados e que, considerando um trade off entre performance e custo computacional, a configuração apropriada para o classificador é aquela com 100 árvores. Como consequência disso, o objetivo da pesquisa foi atingido, chegando em um eficiente sistema de apoio ao diagnóstico por imagem e não invasivo da hanseníase e do câncer de pele adaptado à realidade do Brasil.

Palavras-chave: sistema de apoio ao diagnóstico; hanseníase; câncer de pele; redes neurais artificiais profundas.

ABSTRACT

The use of artificial intelligence is continuously growing in various industries, including healthcare. Computational intelligence applications help healthcare professionals in the diagnostic process, and are faster and more reliable. According to the World Health Organization (WHO), neglected diseases are infectious diseases common in low-income populations in developing regions that do not have sufficient appeal to receive government or market investments. One of the diseases classified as neglected is Hansen's disease, with Brazil being one of the countries with the most cases, reaching an overall new cases detection rate of 13.23 per 100,000 inhabitants in 2019. With a different reality, but also worrying, cancer is considered one of the leading causes of death and a critical barrier to increasing world life expectancy. In Brazil, for each year of the triennium 2020-2022, there will be an estimated 625 thousand new cancer cases, with non-melanoma skin cancer being the one with the highest incidence (177 thousand). This study aims to contribute to the early detection of Hansen's disease and skin cancer by creating a clinical decision support system using deep neural networks for feature extraction, combined with a classification algorithm, to recognize diseases by analyzing skin lesion images. To be more assertive, we proposed seven different experimental scenarios, considering both computational aspects and reports from healthcare professionals. We investigated the deep neural networks LeNet, SqueezeNet, NASNetMobile, ResNet50, and Inception v3 combined with different hyperparameters of the Random Forest classifier. Based on the defined evaluation metrics, we observed that the deep neural network Inception v3 usually generates better results and that, considering a trade-off between performance and computational cost, the appropriate configuration for the classifier is one with 100 trees. As a result, we achieved the research objective, arriving at an efficient clinical decision support system for the imaging and non-invasive diagnosis of Hansen's disease and skin cancer adapted to the reality of Brazil.

Keywords: clinical decision support system; hansen's disease; skin cancer; convolutional neural network.

LISTA DE ILUSTRAÇÕES

Figura 1 Representação das três camadas da pele: epiderme (*epidermis*), derme (*dermis*) e hipoderme (*subcutis*). E os três principais tipos de células da epiderme: células escamosas (*squamous cells*), melanócitos (*melanocyte*) e células basais (*basal cells*). 35

Figura 2 Arquitetura típica de uma rede neural convolucional. A primeira etapa consiste na extração de atributos de uma imagem de entrada, seguida pela etapa de classificação a partir do entendimento desses atributos. 36

Figura 3 Arquitetura da rede neural convolucional LeNet. São duas camadas de convolução (*convolutions*), duas camadas de subamostragem (*subsampling*) e três camadas totalmente conectadas para fazer as classificações, onde cada plano corresponde a um mapa de características. 37

Figura 4 Princípio básico e estrutura da rede neural convolucional SqueezeNet. 38

Figura 5 Princípio de funcionamento e estrutura da rede neural convolucional ResNet. 39

Figura 6 Ilustração do funcionamento da sobreamostragem de dados pelo SMOTE. Até que as classes possuam a mesma proporção, pontos artificiais são adicionados em segmentos de linhas que juntam os pontos da classe minoritária e os k vizinhos mais próximos selecionados aleatoriamente. 40

Figura 7 Demonstração do funcionamento de classificação pelo método *Random Forest*. O resultado do modelo é aquele mais frequente dentre os resultados de n árvores de decisão que são treinadas com diferentes conjuntos de dados e com os atributos presentes definidos de forma aleatória. 42

Figura 8 Amostras do banco de imagens referentes aos tipos de câncer de pele. 43

Figura 9 Amostras do banco de imagens referentes aos tipos de hanseníase. 43

Figura 10 Amostras do banco de imagens referentes às doenças que fazem diagnóstico diferencial com a hanseníase. 44

Figura 11 Cenários propostos para a classificação de uma imagem entre hanseníase e câncer de pele. 46

Figura 12 Cenário proposto para a classificação de uma imagem como câncer de pele, onde cada tipo da doença é considerado independente. 46
Figura 13 Cenários propostos para a classificação de uma imagem como hanseníase. 47
Figura 14 Processo de treinamento, validação e teste utilizando a técnica de validação cruzada com k-dobra ($k = 6$) para escolher o modelo de melhor desempenho, seguido pelo teste final com amostras até então desconhecidas pelo modelo..... 49
Figura 15 Fluxograma dos experimentos que visam escolher a rede neural profunda que proporciona o melhor desempenho do sistema de apoio ao diagnóstico. 51
Figura 16 Fluxograma dos experimentos que visam definir os hiperparâmetros do classificador do sistema de apoio ao diagnóstico. 52
Figura 17 O sistema desenvolvido considera que o suporte ao diagnóstico médico se dá após processar a imagem da lesão do paciente e associá-la a uma das possíveis doenças..... 53
Figura 18 Acurácia da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 57
Figura 19 Especificidade da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 57
Figura 20 Sensibilidade da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 58
Figura 21 Área sob a curva ROC da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 58

Figura 22 Índice kappa da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 59
Figura 23 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 60
Figura 24 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 60
Figura 25 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 61
Figura 26 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 61
Figura 27 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 62
Figura 28 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 63
Figura 29 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 63

Figura 30 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 64
Figura 31 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 64
Figura 32 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 65
Figura 33 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 66
Figura 34 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 66
Figura 35 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 67
Figura 36 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 67
Figura 37 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 68

Figura 38 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 69
Figura 39 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 69
Figura 40 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 70
Figura 41 Área sob a cura ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 70
Figura 42 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 71
Figura 43 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 72
Figura 44 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 72
Figura 45 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 73

Figura 46 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 73
Figura 47 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 74
Figura 48 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 75
Figura 49 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 75
Figura 50 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 76
Figura 51 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 76
Figura 52 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 77
Figura 53 Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 78

Figura 54 Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 78
Figura 55 Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 79
Figura 56 Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 79
Figura 57 Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos. 80
Figura 58 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11a. 81
Figura 59 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11b. 82
Figura 60 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 12. 83
Figura 61 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13a. 83

Figura 62 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13b. 84
Figura 63 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13c. 85
Figura 64 Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13d. 86
Figura 65 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 11a. 87
Figura 66 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 11b. 87
Figura 67 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 12. 88
Figura 68 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13a. 88
Figura 69 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13b. 89
Figura 70 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13c. 90
Figura 71 Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13d. 91

LISTA DE TABELAS

Tabela 1 Panorama de trabalhos relacionados ao apoio ao diagnóstico de doenças de pele, especificamente os que de alguma forma englobam o câncer de pele e/ou a hanseníase. São detalhados as doenças incluídas, os métodos utilizados, as bases de imagens, os resultados obtidos, além de observações julgadas pertinentes pelo autor. 30

Tabela 2 Quantidade de imagens por doença que compõem a base de imagens utilizada. 45

Tabela 3 Quantidade de atributos extraídos por cada rede neural profunda utilizada. 48

Tabela 4 Expressões matemáticas das métricas escolhidas para avaliar o desempenho dos classificadores. As predições podem ser classificadas como TP (Verdadeiro Positivo), TN (Verdadeiro Negativo), FP (Falso Positivo) ou FN (Falso Negativo). A análise de confiabilidade dos resultados é realizada a partir da concordância observada (acurácia) ρ_o e da concordância esperada ρ_e..... 54

Tabela 5 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda LeNet e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos..... 55

Tabela 6 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda SqueezeNet e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 55

Tabela 7 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda NASNetMobile e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 56

Tabela 8 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda ResNet50 e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 56

Tabela 9 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos. 56

Tabela 10 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 60

Tabela 11 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 62

Tabela 12 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 65

Tabela 13 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 68

Tabela 14 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 71
Tabela 15 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 74
Tabela 16 Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos..... 77
Tabela 17 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11a..... 81
Tabela 18 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11b. 81
Tabela 19 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 12..... 82
Tabela 20 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13a..... 83

Tabela 21 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13b. 84
Tabela 22 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13c. 85
Tabela 23 Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13d. 85

SUMÁRIO

1 **INTRODUÇÃO** 20
1.1 MOTIVAÇÃO E JUSTIFICATIVA 20
1.2 OBJETIVOS 22
1.3 ORGANIZAÇÃO DO TRABALHO 22
2 **TRABALHOS RELACIONADOS** 23
3 **FUNDAMENTAÇÃO TEÓRICA** 31
3.1 HANSENÍASE 31
3.2 CÂNCER DE PELE 33
3.3 APRENDIZADO PROFUNDO 35
3.4 BALANCEAMENTO DE DADOS 39
3.5 CLASSIFICAÇÃO 41
4 **MATERIAIS E MÉTODOS** 43
4.1 BASE DE DADOS 43
4.2 CENÁRIOS PROPOSTOS 44
4.3 EXTRAÇÃO DE ATRIBUTOS 46
4.4 TREINAMENTO, VALIDAÇÃO E TESTE 48
4.5 ARRANJO EXPERIMENTAL 49
4.6 MÉTRICAS DE AVALIAÇÃO 53
5 **RESULTADOS E DISCUSSÃO** 55
5.1 AVALIAÇÃO DOS MODELOS DE REDE PROFUNDA 55
5.2 INVESTIGAÇÃO DAS CONFIGURAÇÕES DO CLASSIFICADOR 59
5.3 TEMPO DE TREINAMENTO E TESTE DURANTE A VALIDAÇÃO CRUZADA 80
5.4 TESTE DO MELHOR MODELO 86
5.5 CONSIDERAÇÕES FINAIS 90
6 **CONCLUSÃO** 94
REFERÊNCIAS 95

1 INTRODUÇÃO

1.1 MOTIVAÇÃO E JUSTIFICATIVA

O uso de Inteligência Artificial (IA) cresce continuamente em diversos setores, e no setor da saúde não é diferente. O intuito é que a IA aumente a capacidade dos humanos em fornecer cuidados de saúde, seja facilitando a detecção precoce, permitindo um diagnóstico mais preciso ou contribuindo na descoberta de novos tratamentos (FOGEL; KVEDAR, 2018; HE, J. et al., 2019). Aplicações de inteligência computacional ajudam aos profissionais da saúde no processo de diagnóstico, permitindo que este seja mais rápido e mais confiável. A utilização de métodos de aprendizado de máquina (*machine learning*) e, ainda mais recente, aplicações com aprendizado profundo (*deep learning*), vêm trazendo bons resultados e permite que poderosas ferramentas de apoio ao diagnóstico por imagem e sinais possam ser construídas, por exemplo, extraindo características de imagens e criando um sistema de classificação (YU; BEAM; KOHANE, 2018). São diversas as doenças conhecidas e classificadas pelo ser humano, podendo ser causadas de diferentes formas, algumas são comumente encontradas pelo mundo, já outras possuem baixa prevalência. Contribuir na detecção precoce das doenças definidas como negligenciadas foi a motivação inicial deste trabalho.

Doenças negligenciadas são aquelas doenças endêmicas, em geral transmissíveis, que ocorrem em regiões subdesenvolvidas que não possuem apelo suficiente para que recebam investimentos do Estado ou do mercado, seja por baixa prevalência, seja simplesmente por serem doenças das camadas menos favorecidas da população (WORLD HEALTH ORGANIZATION, 2021). De acordo com o relatório do G-Finder 2019 (CHAPMAN et al., 2019), plataforma que monitora o investimento global em pesquisa e desenvolvimento de produtos para doenças negligenciadas, o financiamento para combater essas doenças foi quase 10% menor em 2018 do que em 2009, caindo US$34 milhões (-9,1%). O Brasil, por sua vez, apresenta queda em investimentos nas cinco doenças negligenciadas definidas pelo Ministério da Saúde como prioritárias (dengue, doença de Chagas, leishmaniose, hanseníase e esquistossomose).

No Brasil, a taxa de detecção geral de novos casos de hanseníase foi de 13,23 por 100 mil habitantes em 2019, sendo um dos países com mais casos da doença. A região Nordeste aparece como a terceira região do país com maior taxa de detecção geral de

novos casos (19,97 por 100 mil habitantes) e o estado de Pernambuco como o sexto estado do país, e terceiro da região Nordeste, com uma taxa de 26,24 por 100 mil habitantes (BRASIL, 2021). O ano de 2020 apresentou a maior redução da taxa de detecção geral (desde 2011), o que profissionais médicos relacionam aos efeitos da sobrecarga dos serviços de saúde e restrições vigentes durante a pandemia de Covid-19 (BRASIL, 2022). Além disso, outro fator relevante para o alto índice de subnotificação da doença é o estigma que seus pacientes carregam (RIBEIRO; SILVA; OLIVEIRA, 2018). Embora seja uma doença com potencial incapacitante, a hanseníase tem bom prognóstico quando diagnosticada precocemente, podendo não deixar sequelas. No entanto, como doença negligenciada, ainda não se dispõe facilmente de ferramentas de apoio, o que faz com que o diagnóstico de caso de hanseníase seja essencialmente clínico, sem exames complementares.

Com uma realidade diferente, outro tipo de doença que também merece atenção é o câncer, sendo uma das principais causas de morte e uma importante barreira para o aumento da expectativa de vida em todos os países do mundo (BRAY et al., 2021). Estima-se que em 2020 cerca de 19,3 milhões de novos casos e quase 10,0 milhões de mortes por câncer ocorreram em todo o mundo. Em geral, crescem a incidência e a mortalidade por câncer, o que pode ser entendido como um reflexo tanto do envelhecimento e do crescimento da população quanto das mudanças na prevalência e distribuição dos principais fatores de risco para o câncer, muitos associados ao desenvolvimento socioeconômico (SUNG et al., 2021). No Brasil, a estimativa para cada ano do triênio 2020-2022 indica 625 mil casos novos de câncer. O de maior incidência deve ser o câncer de pele não melanoma (177 mil), seguido por câncer de mama e câncer de próstata (66 mil cada) e cólon e reto (41 mil). Na região Nordeste, o câncer de pele não melanoma é mais incidente em mulheres (risco estimado de 63,02 casos novos a cada 100 mil mulheres) e o câncer de próstata ocupa a primeira posição entre os homens (72,35/100 mil) (INSTITUTO NACIONAL DO CÂNCER, 2019).

Sabendo dos impactos significativos dessas doenças, o intuito deste trabalho é contribuir para o diagnóstico precoce delas, já que esse pode ser considerado um fator importante em relação à resposta ao tratamento e à diminuição de sequelas. Isso será feito com base no aprendizado profundo (estado da arte em inteligência artificial) combinado à técnicas de classificação, de modo a criar uma estrutura de apoio ao diagnóstico não invasivo e capaz de atender a realidade encontrada no Brasil.

1.2 OBJETIVOS

Construir um modelo de aprendizado de máquina para apoio ao diagnóstico não invasivo de câncer de pele e hanseníase, a partir da análise automatizada de imagens fotográficas, baseado em arquitetura de rede neural profunda e classificador Random Forest. Como objetivos específicos, têm-se:

1. Investigar as doenças hanseníase e câncer de pele e construir um banco de imagens fotográficas;
2. Avaliar modelos de redes profundas para extração de atributos de imagens;
3. Investigar classificadores para compor uma arquitetura de apoio à detecção e classificação das doenças;
4. Construir uma solução para triagem e apoio ao diagnóstico da hanseníase e do câncer de pele.

1.3 ORGANIZAÇÃO DO TRABALHO

O trabalho está estruturado da forma que segue:

- No Capítulo 2 são apresentados trabalhos relacionados à pesquisa desenvolvida. São traçados paralelos, levantados pontos de atenção e também é destacado o diferencial do presente trabalho;
- No Capítulo 3 são abordados os conceitos fundamentais para o desenvolvimento e entendimento da pesquisa;
- No Capítulo 4 são descritas a base de dados utilizada, as maneiras como os experimentos foram conduzidos, as configurações definidas para cada etapa e as métricas de avaliação escolhidas;
- No Capítulo 5 são realizadas discussões em cima dos resultados obtidos nas diferentes configurações testadas, sendo finalizado com a escolha de uma delas para compor o sistema de apoio ao diagnóstico;
- No Capítulo 6 é realizada a conclusão do trabalho, deixando claro a relevância da pesquisa, pontuando observações relevantes e sugerindo trabalhos futuros.

2 TRABALHOS RELACIONADOS

Neste capítulo, foram revisadas as publicações mais recentes relacionadas ao apoio ao diagnóstico de doenças de pele. O foco será em aplicações de diagnóstico por imagem e com preferência para aqueles trabalhos que tratam especificamente de câncer de pele e/ou hanseníase, mas também serão considerados aqueles que contam com outras doenças de pele, desde que câncer de pele e/ou hanseníase também estejam incluídas.

O trabalho de (DE GOMA; DEVARAJ, 2020) é um exemplo de aplicação que leva em conta o reconhecimento de lesões de hanseníase, além de considerar acne, dermatite atópica, queratose pilar, psoríase e verrugas. Isso se dá pois o trabalho foca em doenças de pele comuns nas Filipinas. Antes de testar a classificação por redes neurais artificiais e máquina de vetores de suporte, a imagem de entrada passa por um pré-processamento (transformação para escala de cinza, filtro de mediana, entre outros), depois é segmentada e, enfim, são extraídos os atributos cor, borda e textura.

Para a hanseníase, ambos os classificadores atingiram uma precisão e sensibilidade superiores a 90%, com destaque para as redes neurais artificiais que chegou em 96,55% e 93,33%, respectivamente. Ou seja, para o caso específico que é abordado no trabalho, os resultados são bons, principalmente por propor um algoritmo que é capaz de normalizar os valores do pré-processamento de acordo com a cor da pele do paciente. No estudo, algumas das doenças consideradas têm características comuns que tornam difícil distinguir significativamente uma da outra. Pensando na operacionalização da solução, isso pode ser um limitante, e, soluções como adicionar outros atributos para o classificador (e.g. idade - algumas das doenças são mais comuns em determinadas faixas etárias) ou adicionar outros filtros, podem impactar consideravelmente na realização.

Já o estudo de (BAWEJA; PARHAR, 2016) se refere à utilização de uma arquitetura baseada em rede neural convolucional para o reconhecimento de lesões de hanseníase. A base de imagens utilizada é composta por imagens igualmente distribuídas entre lesões de hanseníase e não-hanseníase. Estas são imagens extraídas da base pública DermnetNz [1] e de resultados de pesquisa do Google, após tratamento manual. De um total de 150, é utilizada a proporção de 60% das imagens para treinamento, 20% para validação cruzada e 20% para teste.

[1] `https://dermnetnz.org/`

As imagens passam por duas camadas convolucionais, compostas de conjuntos de filtros, de modo que a rede proposta aprenda a detectar características específicas das imagens. As saídas da segunda camada convolucional são conectadas a dois neurônios, cada um dando uma probabilidade da imagem ser um exemplo positivo ou um exemplo negativo de hanseníase. A arquitetura final usa um otimizador *Adaptive Moment Estimation* (ADAM) e resulta em um valor ótimo de 3 e 6 para a quantidade de filtros na primeira e na segunda camada convolucional, respectivamente, obtendo uma acurácia média de 91,6%.

Na proposta realizada, é considerado positivo atingir uma alta acurácia média em um método de diagnóstico não invasivo e que é capaz de perceber se uma lesão está de fato relacionada à hanseníase. Porém, os resultados obtidos deixam abertos alguns questionamentos quanto à metodologia utilizada. Três pontos, que são ponderados no presente projeto de mestrado, podem ser levantados: referente à base de imagens, (i) não fica claro quais são exatamente as doenças consideradas como não-hanseníase, a quantidade de imagens para cada uma delas e se realmente fazem diagnóstico diferencial com hanseníase; e (ii) a quantidade total de imagens é baixa, o que não garante uma grande abrangência para os diversos tipos de lesões possíveis; outro ponto importante é em relação à análise dos resultados obtidos, (iii) a única métrica de qualidade é a acurácia e, como será observado na Seção 4.6, é importante que a análise seja feita por um conjunto de métricas que podem dar um panorama real da performance do método.

Em (JOSHI et al., 2019) é proposto um sistema para detectar se uma imagem com lesão na pele pode ser considerada saudável ou não; se não, o sistema classifica entre melanoma, eczema ou hanseníase. A imagem passa por pré-processamento (aumento de contraste e conversão de cor para escala cinza), segmentação (*global thresholding*) e extração de atributos (avaliação da qualidade da imagem e matriz de coocorrência de nível de cinza - GLCM). Enfim, a classificação é realizada a partir de uma árvore de decisão.

O sistema proposto é interessante pois desenvolve uma Interface Gráfica do Usuário (GUI), onde é possível observar as imagens e os valores resultantes de cada operação realizada na imagem original. Considerando 45 imagens de testes, o sistema apresentou uma acurácia de 87%. Porém, apesar de já ser considerada uma baixa quantidade de imagens para teste, não foi explicitada a quantidade total de imagens utilizadas no trabalho, nem a origem delas. Outro ponto a ser comentado é que apenas foi apresentado o valor

de acurácia já citado, nenhuma outra métrica de qualidade foi utilizada.

O trabalho de (ESTEVA et al., 2017) é uma referência quanto ao uso de redes neurais convolucionais na classificação de câncer de pele. A proposta conta com uma base de dados com 129.450 imagens clínicas, incluindo 3.374 imagens de dermatoscopia, originadas de 18 repositórios online de acesso aberto com curadorias clínicas diferentes. Foi desenvolvida uma rede profunda de arquitetura GoogleNet Inception v3 com desempenho correspondente ao de dermatologistas em três tarefas principais de diagnóstico: classificação de melanoma, classificação de melanoma usando dermatoscopia e classificação de carcinoma, restringindo as comparações à classificação baseada em imagem.

Foi possível observar a eficácia das redes profundas em dermatologia, tanto aplicando a condições gerais da pele (precisão geral de $72,1 \pm 0,9\%$), quanto a cânceres específicos (precisão geral de $55,4 \pm 1,\%$). Importante salientar que o diagnóstico de um dermatologista se baseia em fatores além da inspeção visual e dermatoscópica de uma lesão isolada, então, a solução ideal é aquela que integra a capacidade de classificar imagens de lesões de pele de uma rede profunda com a experiência de um dermatologista. Porém, ainda é preciso avaliar o desempenho de acordo com o que é encontrado na prática típica de um ambiente clínico.

Em (HASAN et al., 2019) são utilizadas redes neurais convolucionais para classificação de câncer de pele em duas classes: benigno e maligno. Foram utilizadas 23.907 imagens extraídas da base ISIC [2]. Três tipos de camadas estão presentes na rede neural convolucional: camada de convolução, camada de agrupamento e camada totalmente conectada. A avaliação do método proposto foi realizada a partir da precisão, sensibilidade e medida F, obtendo 83,25%, 84% e 83,25%, respectivamente.

Além de contar com uma base de imagens pública de qualidade, propor um sistema que utiliza redes neurais convolucionais para classificação é uma boa estratégia. Porém, obter imagens de dermatoscopia com alta qualidade não combina com o objetivo e a proposta da presente pesquisa de mestrado. Então, pensando na operacionalização dessa solução, as imagens usadas rotineiramente terão características diferentes daquelas com que a rede foi treinada.

Como já mostrado, existem bons trabalhos que classificam imagens de câncer de pele a partir de redes neurais convolucionais. Porém, conforme o questionamento proposto

[2] `https://www.isic-archive.com/`

em (TSCHANDL et al., 2019), não são muitos os trabalhos que consideram os tipos de câncer de pele não pigmentado e não melanoma, sendo esses mais comuns e mais difíceis de diagnosticar. No trabalho, eles comparam a precisão de um classificador baseado em redes neurais convolucionais com a de médicos com diferentes níveis de experiência.

A base de imagens utilizada para treinamento é composta por 7.895 imagens dermatoscópicas e 5.829 fotos em close de lesões, extraídas em uma clínica primária de câncer de pele na Austrália entre janeiro de 2008 e julho de 2017. Eles combinaram duas redes neurais convolucionais, uma treinada apenas com as imagens dermatoscópicas (melhor resultado: InceptionV3) e outra apenas com as fotos em close das lesões (melhor resultado: ResNet50). Essa rede neural convolucional combinada foi testada em um conjunto de 2.072 casos desconhecidos e comparada com os resultados de 95 médicos.

A abordagem é interessante por contar com uma grande base de imagens e analisar o diagnóstico com uma série de profissionais da saúde, obtendo uma rede neural convolucional capaz de classificar lesões não pigmentadas com a mesma precisão que avaliadores especialistas e com maior precisão do que avaliadores menos experientes. Porém, o fato da rede neural convolucional ser treinada com imagens de lesões de pacientes de uma clínica da Austrália acaba enviesando a capacidade de detecção do sistema proposto e diminuindo a abrangência de aplicação. Além disso, por mais que seja comum ter dados desbalanceados por conta da prevalência de cada doença, nenhum pré-processamento foi realizado, e o resultado disso foi que a rede neural convolucional superou os avaliadores humanos em classes malignas comuns, mas não atingiu a precisão de avaliadores humanos em raras lesões malignas não pigmentadas.

Entrando nos trabalhos que tratam mais especificamente da detecção de câncer de pele melanoma, o sistema proposto em (SHALU; KAMBOJ, 2018) é criado a partir da base de imagens MED-NODE (100 imagens do tipo nevo e 70 imagens do tipo melanoma), estas passam por pré-processamento, segmentação e extração de atributos. A informação resultante serve como entrada para um classificador, os avaliados foram: naïve bayes, árvore de decisão e k-ésimo vizinho mais próximo (KNN - *k-nearest neighbors*). O melhor resultado é obtido pela árvore de decisão, atingindo uma precisão de 82,35%.

Nos resultados obtidos, a especificidade do sistema desenvolvido em geral é maior do que a sensibilidade, o que significa que o sistema é melhor em identificar os casos benignos. Uma abordagem para diminuir essa diferença de performance é adicionar mais

recursos de pré-processamento das imagens. O problema dessa estratégia é que são necessárias diversas etapas antes de chegar a uma classificação, o que implica em uma grande quantidade de parâmetros a serem configurados e pode implicar em um alto custo computacional. Outras técnicas, mostradas mais a frente, são boas alternativas para isso e chegam a resultados semelhantes ou até melhores.

A proposta em (ASTORINO et al., 2020) é uma aplicação de Aprendizado de Múltiplas Instâncias (MIL - *Multiple Instance Learning*) para detecção de melanoma. Em vez de receber um conjunto de instâncias que são rotuladas individualmente, o objetivo do MIL é discriminar entre conjuntos de itens positivos e negativos. Esses conjuntos são chamados de bolsas e os itens dentro das bolsas são chamados de instâncias, uma bolsa é positiva se pelo menos uma de suas instâncias for positiva e é negativa se todas as suas instâncias forem negativas. A ideia é que, a partir de diversas bolsas rotuladas, seja possível (i) induzir um conceito que rotulará instâncias individuais corretamente ou (ii) aprender a rotular bolsas sem induzir o conceito.

Da base de imagens PH^2, foram utilizadas 40 imagens de melanomas e as 80 de nevos comuns. Os pacientes de quem as fotos foram tiradas correspondem ao fototipo II ou III (classificação do tipo de pele de Fitzpatrick). O algoritmo MIL buscar discriminar as imagens entre melanomas (imagens positivas) e nevos comuns (imagens negativas). Considerando diferentes testes, o melhor resultado obtido chegou a uma acurácia de 92,50%, sensibilidade de 97,50% e especificidade de 87,50%.

Como geralmente um diagnóstico baseado em imagem não depende da imagem inteira, mas apenas de algumas partes dela, considerar um classificador MIL se encaixa bem na problemática, por conta da sua capacidade de extrair informações globais (melanomas ou nevos comuns) da detecção local (sub-regiões da imagem). Os resultados parecem promissores e a utilização desta técnica pode ser interessante. Como pontos de melhoria, considerar etapas de pré-processamento das imagens pode melhorar o desempenho do algoritmo; além disso, a base de imagens utilizada é bastante limitada, tanto pela quantidade de imagens quanto pela abrangência de diferentes fototipos de pele.

Já o trabalho de (PEREZ; AVILA; VALLE, 2019) avalia os fatores que afetam a escolha de uma arquitetura de rede profunda para análise de lesões de pele, mais especificamente, melanoma. Foram consideradas 9 arquiteturas de rede (DenseNet, Dual Path Nets, Inception-v4, Inception-ResNet-v2, MobileNetV2, PNASNet, ResNet, SENet,

Xception) em 5 divisões criadas da base de dados do *ISIC Challenge 2017*[3]. De 2.750 imagens, produziu-se cinco combinações aleatórias de divisões de treino (1.750), validação (500) e teste (500), uma partição aproximada de 60–20–20%. O objetivo é a analisar a classificação entre melanoma ou não melanoma (queratose seborreica e nevo melanocítico).

A proposta é criar conjuntos com vários modelos, de forma a melhorar e estabilizar as precisões, já que a precisão de modelos de aprendizado de máquina tende a flutuar devido a fatores não controlados, como a escolha do conjunto de treinamento, e até mesmo condições aleatórias, como a inicialização das redes. Neste trabalho, foram reproduzidas as descobertas do desafio citado, e verificou-se que as correlações até então relatadas desapareceram por completo quando são consideradas apenas as redes de melhor desempenho: fatores não controlados parecem mais influentes do que a escolha de arquiteturas. Onde, não apenas nenhuma característica da rede (e.g. quantidade de parâmetros) se correlaciona bem com as métricas escolhidas, mas também parece impossível distinguir a maioria das redes umas das outras em um teste estatístico. Os modelos avaliados chegaram em um valor de área sob a curva ROC entre 84 e 91%, próximo ao primeiro lugar do desafio (87,4%).

Em (MENEGOLA et al., 2017) é avaliada a técnica de transferência de aprendizado (reaproveitamento do conhecimento de modelos treinados em diferentes tarefas) como alternativa para a necessidade de grandes quantidades de dados de treinamento nas soluções que utilizam redes neurais profundas. Foram verificados os resultados de dois casos: ajuste fino do modelo de aprendizado profundo após a transferência de conhecimento e escolha de modelos mais profundos (com maior custo computacional).

A base de imagens para treinamento e testes tem duas origens: *Interactive Atlas of Dermoscopy* (comercial) e *ISBI Challenge 2016* (pública). Apenas imagens dermatoscópicas são incluídas e o intuito é a classificação melanoma *vs* outras. Referente à transferência de aprendizado, a origem dos dados vem de um conjunto de dados do *Kaggle Challenge for Diabetic Retinopathy Detection* (mais de 35.000 imagens de retina de alta resolução obtidas em condições variadas); e do conjunto de dados *ImageNet Large Scale Visual Recognition Challenge 2012* (1 milhão de imagens de treinamento rotuladas em 1.000 categorias).

Sempre passando por uma camada de decisão pelo SVM (máquina de vetores de

[3] `https://challenge.isic-archive.com/landing/2017/`

suporte), foram avaliados os seguintes cenários: sem transferência de aprendizado (construção do modelo do zero), tranferência da base *Retinopathy*, transferência da base *ImageNet*, tranferência dupla (da *Retinopathy* e da *ImageNet*) e mais o ajuste fino ou não nos exemplos que utilizam transferência. Verificou-se que uma transferência de outra tarefa médica específica (*Retinopathy*) não é eficaz, mesmo que precedida de transferência de uma tarefa geral (*ImageNet*), e que o ajuste fino tem grande impacto e deve ser considerado uma necessidade. Os resultados obtidos favorecem modelos mais profundos, pré-treinados sobre *ImageNet*, com ajuste fino, atingindo uma área sob a curva ROC de 80,7% e 84,5% para os dois conjuntos de dados de lesões de pele avaliados. Estratificando pela dificuldade de diagnóstico, bons resultados são obtidos ao lidar com o espectro inferior e médio de dificuldade, assim como já acontece na arte atual, e lesões difíceis continuam sendo de difícil diagnóstico.

A pesquisa realizada para separar os trabalhos apresentados neste capítulo pode ser resumida na Tabela 1. Foi possível observar que a maioria destes trabalhos segue uma das duas abordagens: (i) pré-processamento das imagens (segmentação, extração de atributos etc) seguido de algum algoritmo de classificação ou (ii) utilização de redes neurais artificiais profundas para processamento e classificação das imagens. Já a pesquisa que será apresentada nos próximos capítulos une as duas técnicas: as redes neurais artificiais para extrair os atributos das imagens, que posteriormente passam por um algoritmo de classificação.

A presente pesquisa também busca lidar com outro ponto de bastante importância: a base de imagem utilizada. Em geral, elas possuem alguma das seguintes limitações: pouca abrangência quanto à tonalidade da pele, baixa quantidade de imagens e necessidade de recursos para reproduzir (e.g. imagens dermatoscópicas de alta qualidade). Esses fatores de fato contribuem para que uma técnica escolhida tenha uma boa performance, mas dificilmente é um retrato da realidade que será encontrada (WEN et al., 2022).

Tabela 1: Panorama de trabalhos relacionados ao apoio ao diagnóstico de doenças de pele, especificamente os que de alguma forma englobam o câncer de pele e/ou a hanseníase. São detalhados as doenças incluídas, os métodos utilizados, as bases de imagens, os resultados obtidos, além de observações julgadas pertinentes pelo autor.

Trabalho	Doenças incluídas	Técnicas utilizadas	Base de imagens	Nº de imagens	Resultados	Observações
(DE GOMA; DEVARAJ, 2020)	Hanseníase, acne, dermatite atópica, queratose pilar, psoríase e verrugas	Algoritmos de pré-processamento e segmentação. Classificação: redes neurais artificiais e máquina de vetores de suporte	Imagens registradas por dermatologistas	686	Precisão de 96,55% e sensibilidade de 93,33% para as redes neurais artificiais	O algoritmo é capaz de normalizar os valores do pré-processamento de acordo com a cor da pele do paciente
(BAWEJA; PARHAR, 2016)	Hanseníase, câncer de pele e queimaduras leves	Rede neural convolucional InceptionV3	DermnetNz e resultados de pesquisa do Google	150	Acurácia média de 91,6%	Baixa quantidade de imagens
(JOSHI et al., 2019)	Melanoma, eczema e hanseníase	Algoritmos de pré-processamento, segmentação e extração de atributos. Classificação: árvore de decisão	Não informado	45 para teste (total não informado)	Acurácia de 87%	Baixa quantidade de imagens e origem desconhecida. Nenhuma outra métrica de qualidade além da acurácia foi utilizada
(ESTEVA et al., 2017)	2.032 doenças (avaliado: classificação de melanoma e classificação de carcinoma)	Rede neural convolucional InceptionV3	18 repositórios online de acesso aberto com curadorias clínicas diferentes	129.450	Precisão: 72,1±0,9% para condições gerais da pele 55,4±1,0% para cânceres específicos	Ainda é preciso avaliar o desempenho de acordo com o que é encontrado na prática típica de um ambiente clínico
(HASAN et al., 2019)	Predição de câncer (lista de doenças não foi informada)	Redes neurais convolucionais	ISIC	23.907	Precisão: 83,25%; Sensibilidade: 84% Medida F: 83,25%	Faz-se necessário a obtenção de imagens de dermatoscopia com alta qualidade
(TSCHANDL et al., 2019)	Ceratose actínica, Carcinoma basocelular, Verruga seborreica, Dermatofibromas, Melanoma, Carcinoma epidermoide, Queratoacantoma	Combinação de duas redes neurais convolucionais (InceptionV3 e ResNet50)	Extraídas em uma clínica primária de câncer de pele na Austrália	13.724	Área sob a curva ROC: 74,2±1,3%; Sensibilidade: 80,5±1,5%; Acurácia: 37,6±0,9%	Pouca abrangência devido à origem das imagens
(SHALU; KAMBOJ, 2018)	Câncer de pele (nevo e melanoma)	Algoritmos de pré-processamento, segmentação e extração de atributos. Classificação: naïve bayes, árvore de decisão e KNN	MED-NODE	170	Precisão: 82,35% (árvore de decisão)	São necessárias diversas etapas antes de chegar a uma classificação, o que implica em uma grande quantidade de parâmetros a serem configurados e pode implicar em um alto custo computacional
(ASTORINO et al., 2020)	Melanoma e nevo comum	Aprendizado de Múltiplas Instâncias	PH2	120	Acurácia: 92,50%; Sensibilidade: 97,50%; Especificidade: 87,50%	A base de imagens utilizada é bastante limitada, tanto pela quantidade de imagens quanto pela abrangência de diferentes fototipos de pele
(PEREZ; AVILA; VALLE, 2019)	Melanoma, queratose seborreica e nevo melanocítico	Redes neurais convolucionais	ISIC Challenge 2017	2.750	Área sob a curva ROC entre 84 e 91%	Fatores não controlados (escolha do conjunto de treinamento, e até mesmo condições aleatórias, como a inicialização das redes) parecem mais influentes do que a escolha de arquiteturas
(MENEGOLA et al., 2017)	Melanoma, carcinoma basocelular, nevos, dermatofibroma, lentigo, melanose, queratose seborreica, lesões vasculares	Redes neurais convolucionais com transferência de aprendizado e ajuste fino	Interactive Atlas of Dermoscopy (comercial) e ISBI Challenge 2016 (pública)	2.000+	Área sob a curva ROC: 84,5%	Uma transferência de outra tarefa médica específica não foi eficaz. O ajuste fino tem grande impacto e deve ser considerado uma necessidade

Fonte: Próprio Autor.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 HANSENÍASE

A hanseníase é uma doença crônica, causada pela bactéria *Mycobacterium leprae*, que pode afetar qualquer pessoa. Esta quase sempre evolui e torna-se transmissível se não tratada na forma inicial. Geralmente isso ocorre de forma lenta e progressiva, podendo levar a incapacidades físicas (BRASIL, 2017). Foi identificada no ano de 1873 pelo cientista Armauer Hansen e é considerada uma das doenças mais antigas, com registro de casos há mais de 4000 anos, na China e na Índia. Porém, ainda é uma doença complexa de erradicar por estar fortemente relacionada à pobreza, falta de serviços de saúde e educação, além do estigma que seus pacientes carregam (SANTACROCE et al., 2021).

A bactéria *Mycobacterium leprae* é transmitida por secreções das vias aéreas superiores e por gotículas de saliva, em contato próximo com doentes que ainda não iniciaram tratamento e estão em fases adiantadas da doença. Ou seja, a transmissão acontece pelo ar, e não pelos objetos utilizados pelo paciente. De maneira geral, a hanseníase se caracteriza pela alteração da sensibilidade térmica, dolorosa, tátil e força muscular, além do aparecimento de manchas na pele e até de caroços, podendo gerar incapacidades permanentes. O tratamento cura a doença, interrompe sua transmissão e previne incapacidades físicas (BRASIL, 2017).

O diagnóstico médico envolve sobretudo a avaliação clínica do paciente e depende do reconhecimento dos sinais e sintomas, portanto, só se torna possível após a manifestação da doença. As lesões cutâneas são geralmente a primeira manifestação clínica observada. Se necessário um exame laboratorial, o mais utilizado para auxiliar no diagnóstico é a baciloscopia (coleta da serosidade cutânea), mas que pode ser negativa nas fases iniciais da doença. O diagnóstico precoce é de suma importância para evitar transmissão, complicações e deficiências (CHEN et al., 2022).

Para fins de tratamento, a Organização Mundial da Saúde divide a hanseníase em: paucibacilar (PB), com poucos ou nenhum bacilo nos exames, e multibacilar (MB), com muitos bacilos. Porém, para melhor compreensão das formas clínicas da doença, neste trabalho será considerada a classificação de Madrid (SIXTH INTERNATIONAL CONGRESS OF LEPROSY, 1953), que divide a hanseníase em: indeterminada (PB), tuberculoide (PB), dimorfa (MB) e virchowiana (MB). Com as seguintes características:

- **Hanseníase Indeterminada** (*Indeterminate Leprosy*): mesmo que não seja percebida, todos os pacientes passam por essa fase no início da doença. A lesão de pele geralmente é única, mais clara do que a pele ao redor (mancha), não é elevada (sem alteração de relevo), apresenta bordas mal delimitadas, e é seca (BRASIL, 2017);

- **Hanseníase Tuberculoide** (*Tuberculoid Leprosy*): comumente manifesta-se por uma placa (mancha elevada em relação à pele ao redor) totalmente anestésica ou por placa com bordas elevadas, bem delimitadas e centro claro (forma de anel ou círculo), porém, quando compromete mais de um nervo, geralmente a apresentação é assimétrica (BRASIL, 2017);

- **Hanseníase Dimorfa** (*Borderline Leprosy*): compreende a maior parte dos casos de hanseníase e normalmente se caracteriza por várias manchas de pele avermelhadas ou esbranquiçadas, com bordas elevadas, mal delimitadas na periferia, ou por múltiplas lesões bem delimitadas semelhantes à lesão tuberculoide, porém com uma borda externa pouco definida (BRASIL, 2017);

- **Hanseníase Virchowiana** (*Lepromatous Leprosy*): é a forma mais contagiosa da doença. Os sinais e sintomas podem ser discretos, pois o paciente normalmente apresenta lesões difusas e, em geral, simétricas, podendo manifestar-se por meio de lesões pardacentas, de limites imprecisos. Neste tipo, é comum observar a pele avermelhada, seca, infiltrada e com os poros dilatados (BRASIL, 2017).

Entende-se por diagnóstico diferencial uma lista de doenças que podem explicar os sinais e sintomas de um paciente durante o exame clínico. Como a hanseníase exibe diferentes manifestações dermatológicas e neurológicas dentro de um amplo espectro clínico, realizar o diagnóstico diferencial é de suma importância para prevenir erros e ajudar no correto diagnóstico da doença (KUNDAKCI; ERDEM, 2019). A seguir, o que será considerado para o diagnóstico diferencial de cada tipo de hanseníase:

- **Diagnóstico diferencial da hanseníase indeterminada:** Pitiríase Alba (*Pityriasis Alba*), Dermatite Seborreica (*Seborrhoeic Dermatitis*), Nevo Acrômico (*Naevus Achromic*), Nevo Anêmico (*Naevus Anaemicus*), Vitiligo;

- **Diagnóstico diferencial da hanseníase tuberculoide:** Dermatofitose do Corpo (*Tinea Corporis*), Granuloma Anular (*Granuloma Annulare*), Sarcoidose (*Sarcoi-*

dosis), Tuberculose Cutânea Verrucosa (*Tuberculosis Verrucosa Cutis*), Psoríase (*Psoriasis*);

- **Diagnóstico diferencial da hanseníase dimorfa:** Esclerodermia Linear (*Linear Scleroderma*), Dermatofitose do Corpo (*Tinea Corporis*), Parapsoríase em Placas (*Parapsoriasis En Plaques*), Psoríase (*Psoriasis*);
- **Diagnóstico diferencial da hanseníase virchowiana:** Neurofibromatose (*Neurofibromatosis*), Xantoma Tuberoso (*Xanthoma Tuberosum*), Condrodermatite Nodular Crônica da Hélice (*Chondrodermatitis Nodularis Chronica Helicis*).

3.2 CÂNCER DE PELE

Os tecidos do corpo humano são formados por células normais capazes de se multiplicar por meio de um processo natural e contínuo. Crescer, multiplicar-se e morrer de maneira ordenada é o ciclo da maioria das células normais. O crescimento das células cancerosas é diferente do crescimento das células normais, elas multiplicam-se de maneira descontrolada em vez de morrerem, formando outras novas células anormais. O câncer se caracteriza pela perda do controle da divisão celular e pela capacidade de migrar para outros tecidos (processo conhecido por metástase) (INSTITUTO NACIONAL DO CÂNCER, 2011). Os diferentes tipos de câncer correspondem aos vários tipos de células do corpo. Logo, o câncer de pele corresponde ao crescimento anormal e descontrolado das células da pele.

As células da pele são altamente suscetíveis a danos no DNA, uma prova disso é que uma das principais causas do câncer de pele é a constante exposição à radiação ultravioleta (natural e artificial), sendo responsável pelas mutações que transformam células saudáveis em células cancerígenas. Isso significa que os cânceres de pele podem ser evitados com mudança de comportamento, evitando a exposição aos raios UV. A maneira mais simples e eficaz de se livrar do câncer de pele é com a remoção do tumor, evitando assim a recorrência local. Porém, isso não necessariamente significa estar curado, a depender do momento em que o câncer foi descoberto e diagnosticado, ele pode ter se espalhado de forma a complicar o plano de tratamento (LAIKOVA et al., 2019).

O tempo é crítico, por isso a detecção precoce recebe grande atenção em razão da premissa de que quanto mais cedo o câncer for identificado, maiores são as chances

de cura. Duas estratégias são possíveis quanto a isso: (i) realização de exames de rotina com o objetivo de encontrar o câncer pré-clínico ou as lesões pré-cancerígenas e (ii) o diagnóstico precoce, identificando o câncer em pessoas que apresentam sinais e sintomas suspeitos da doença (WORLD HEALTH ORGANIZATION, 2020).

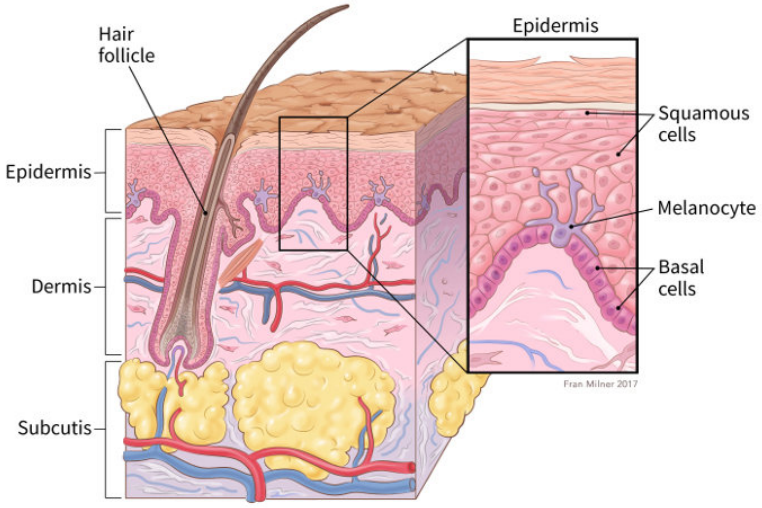
A Figura 1 apresenta as três camadas que compõem a pele: epiderme, derme e hipoderme. Na epiderme (camada superior) é onde começa a maioria dos cânceres de pele. A classificação varia de acordo com as células que lhe deram origem, sendo os mais comuns o melanoma e o não melanoma (mais frequentes: carcinoma basocelular e carcinoma epidermoide) (AMERICAN CANCER SOCIETY, 2019). Abaixo segue uma explicação sobre cada um desses tipos e as indicações na Figura 1 das células afetadas.

- **Melanoma**: tem origem nos melanócitos (apresentados como *melanocyte* na Figura 1), as células que produzem o pigmento que confere cor à pele e a protege dos efeitos perigosos da radiação UV, e pode aparecer em qualquer parte do corpo (pele ou mucosas), na forma de manchas, pintas ou sinais. Apesar de ser o tipo mais grave, este possui um bom prognóstico se detectado em sua fase inicial (AMERICAN CANCER SOCIETY, 2019);

- **Carcinoma basocelular** (*Basal Cell Carcinoma*): é tipo mais comum de câncer de pele e começa na camada mais profunda da epiderme, nas chamadas células basais (apresentadas como *basal cells* na Figura 1). O carcinoma basocelular se desenvolve lentamente e geralmente se manifesta em áreas expostas ao sol, especialmente na região da cabeça e pescoço (AMERICAN CANCER SOCIETY, 2019);

- **Carcinoma epidermoide** (*Squamous Cell Carcinoma*): este tipo de câncer se manifesta na camada mais superficial da pele, nas células escamosas (apresentadas como *squamous cells* na Figura 1). Geralmente aparece em áreas do corpo mais expostas à radiação solar, como o rosto, cabeça, pescoço, braços, mãos e pés. Diferentemente do carcinoma basocelular, o carcinoma epidermoide é considerado mais agressivo e invasivo, com maior tendência de crescimento nas camadas mais profundas da pele e de disseminação para outros órgãos (AMERICAN CANCER SOCIETY, 2019).

Figura 1: Representação das três camadas da pele: epiderme (*epidermis*), derme (*dermis*) e hipoderme (*subcutis*). E os três principais tipos de células da epiderme: células escamosas (*squamous cells*), melanócitos (*melanocyte*) e células basais (*basal cells*).



Fonte: (AMERICAN CANCER SOCIETY, 2019).

3.3 APRENDIZADO PROFUNDO

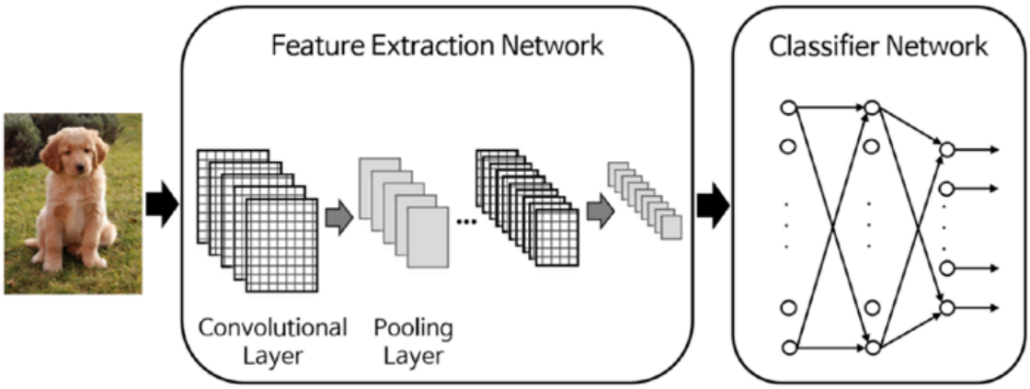
O sistema desenvolvido neste trabalho lida com imagens de lesões de pele, logo, é condição básica tratá-las adequadamente. Para que algoritmos de aprendizado de máquina possam ser utilizados se faz necessário que as imagens sejam representadas numericamente, e isso pode ser feito transformando-as em um mapa de características. Assim, as informações originais permanecem preservadas e ainda é possível que esse mapa de características evidencie partes mais relevantes das imagens.

Quando se trata de extração de atributos, o estado da arte é o aprendizado profundo (*deep learning*), mais especificamente redes neurais convolucionais (CNN - *convolutional neural networks*), onde os modelos conseguem identificar diferentes níveis de representação das imagens. A rede neural convolucional é um tipo de rede neural *feedforward* (dados fluem em uma única direção) que, em resumo, parte de uma imagem como entrada e atribui importância (pesos e vieses) a diferentes áreas, de modo a aprender suas características e propiciar uma classificação mais precisa (JOGIN et al., 2018).

A representação do funcionamento de uma rede neural convolucional pode ser vista na Figura 2. A etapa de extração de atributos (*feature extraction network*) acontece com base em duas operações: convolução (*convolution*) e agrupamento (*pooling*). Entende-se a convolução como um filtro que, em pequenos passos, percorre toda uma imagem e capta

traços marcantes, convertendo a imagem em um mapa de características (*feature map*). Já o *pooling* é utilizado para redução de escala, criando um mapa de características condensadas das informações que saem da camada convolucional. Por fim, as características extraídas na primeira etapa servem de entrada para um classificador baseado em redes neurais artificiais (*classifier network*) (KIM, 2017).

Figura 2: Arquitetura típica de uma rede neural convolucional. A primeira etapa consiste na extração de atributos de uma imagem de entrada, seguida pela etapa de classificação a partir do entendimento desses atributos.

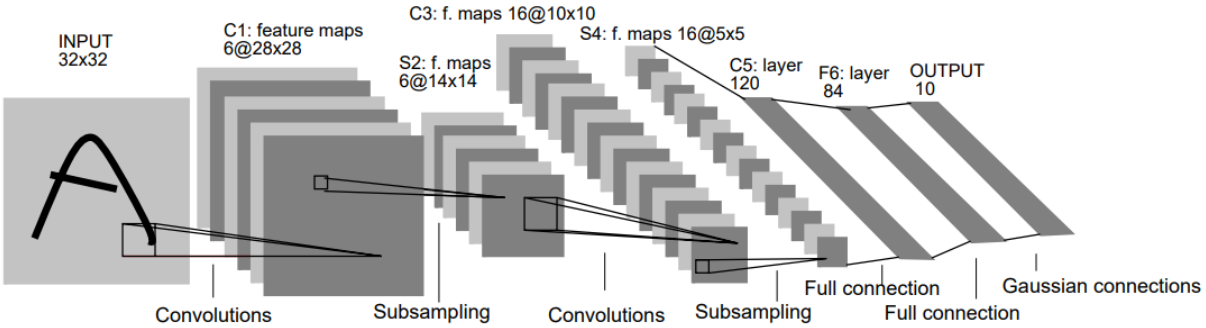


Fonte: (KIM, 2017).

Uma das primeiras arquiteturas de CNN a ganhar destaque pelo desempenho em tarefas de visão computacional foi a LeNet (LECUN et al., 1998). Em sua publicação, ela foi usada para o reconhecimento de caracteres manuscritos. Como mostrado na Figura 3, esta arquitetura conta com um total de sete camadas, onde pesos e vieses são definidos ao longo do treinamento. A primeira camada $C1$ (convolucional) possui 6 mapas de características (contém 156 parâmetros treináveis e 122.304 conexões), a segunda camada $S2$ (subamostragem) também possui 6 mapas de características, mas reduz as dimensões de 28x28 para 14x14 (contém 12 parâmetros treináveis e 5.880 conexões). A terceira camada $C3$ (convolucional) possui 16 mapas de características (contém 1.516 parâmetros treináveis e 151.600 conexões) e a quarta camada $S4$ (subamostragem) também possui 16 mapas de características, mas reduz as dimensões de 10x10 para 5x5 (contém 32 parâmetros treináveis e 2.000 conexões). As quinta $C5$, sexta $F6$ e sétima camadas são totalmente conectadas e são responsáveis pela classificação da imagem (com um total de 58.284 parâmetros treináveis).

Em geral, as propostas de arquiteturas de redes neurais convolucionais buscam principalmente melhorar a acurácia do modelo. A rede convolucional SqueezeNet se con-

Figura 3: Arquitetura da rede neural convolucional LeNet. São duas camadas de convolução (*convolutions*), duas camadas de subamostragem (*subsampling*) e três camadas totalmente conectadas para fazer as classificações, onde cada plano corresponde a um mapa de características.



Fonte: (LECUN et al., 1998).

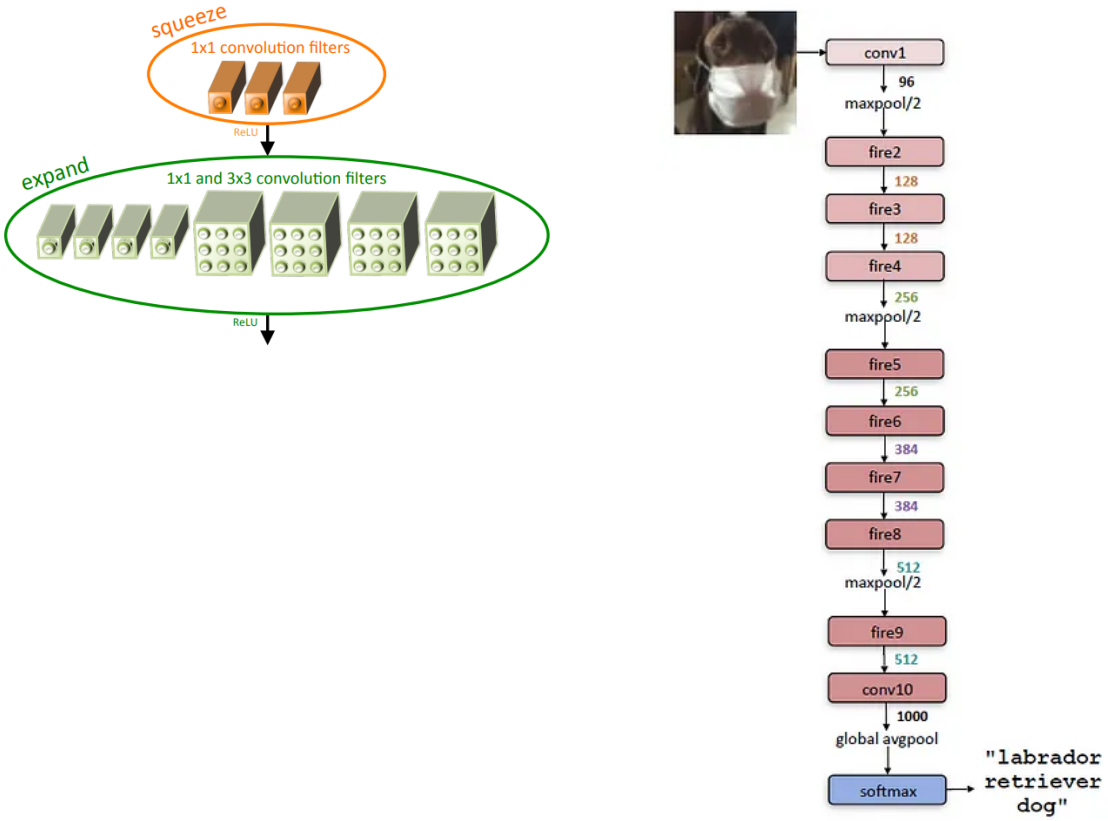
centra em propor uma arquitetura menor, diminuindo a quantidade de parâmetros a serem treinados, e manter o bom desempenho encontrado na literatura (IANDOLA et al., 2016). O princípio básico dessa arquitetura pode ser visto na Figura 4a, onde uma camada de compressão (*squeeze*), apenas com filtros 1x1, e uma camada de expansão (*expand*), com filtros 1x1 e 3x3, se combinam em uma estrutura denominada *fire module*. O objetivo desta é maximizar a performance dada uma quantidade limitada de parâmetros, preservando o desempenho e diminuindo a quantidade de parâmetros quando comparado a outras CNNs encontradas na literatura. A estrutura completa da SqueezeNet pode ser vista na Figura 4b, sendo composta por uma camada de convolução ($conv1$), seguida por oito *fire modules* ($fire2-9$) e finalizando com mais uma camada de convolução ($conv10$).

Desenvolvida pela Google Brain (equipe de pesquisa de aprendizagem profunda no Google), a arquitetura NASNet usa o conceito de Aprendizagem por Reforço para encontrar a melhor arquitetura CNN. NAS significa *Neural Architecture Search* (Busca por Arquitetura Neural) e tem como objetivo escolher a melhor combinação de parâmetros do espaço de busca de tamanhos de filtros, número de camadas, etc. A recompensa após cada ação de busca é a acurácia da arquitetura para aquele determinado conjunto de dados (ZOPH et al., 2018). Neste trabalho também é proposta uma nova técnica de regularização chamada *ScheduledDropPath*, onde caminhos são descartados com uma probabilidade que aumenta linearmente ao longo do treinamento, o que melhora significativamente a generalização nos modelos NASNet. O número total de parâmetros treináveis pode chegar em algumas dezenas de milhões. A versão reduzida é chamada NASNetMobile e possui 5,3 milhões de parâmetros.

Figura 4: Princípio básico e estrutura da rede neural convolucional SqueezeNet.

(a) O princípio básico da arquitetura SqueezeNet se baseia na estrutura denominada *fire module*, composta por uma camada de compressão (*squeeze*) e uma camada de expansão (*expand*).

(b) Estrutura da arquitetura SqueezeNet, composta por uma camada de convolução (*conv*1), seguida por oito *fire modules* ($fire2-9$) e finalizando com mais uma camada de convolução (*conv*10).



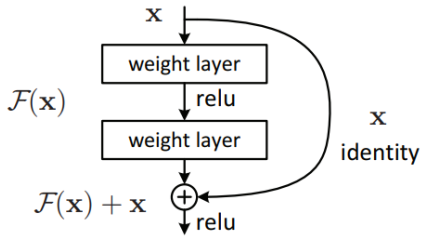
Fonte: (IANDOLA et al., 2016).

A arquitetura da rede neural convolucional ResNet, proposta em (HE, K. et al., 2016), introduz o conceito denominado de blocos residuais (*residual blocks*). A ideia aqui é utilizar a técnica de conexão de atalho (*skip connections*), ignorando algumas camadas intermediárias e, assim, formar um bloco residual. Essa abordagem é ilustrada na Figura 5a, onde a entrada é denota por x e o mapeamento original $\mathcal{F}(x)$ é reformulado em $\mathcal{F}(x) + x$. Dessa forma, se alguma camada prejudicar o desempenho da arquitetura, ela acabará sendo ignorada pela regularização. A hipótese é que é mais fácil otimizar o mapeamento residual do que otimizar o mapeamento original. Uma rede ResNet é feita empilhando vários desses blocos residuais. Por exemplo, a rede ResNet34 tem 34 camadas usando esses blocos, como mostrado na Figura 5b.

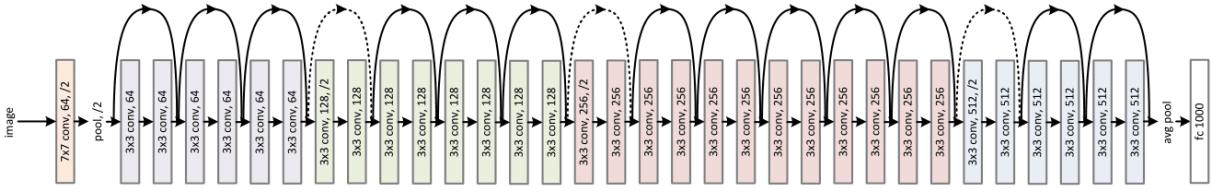
A arquitetura convolucional do Inception foi introduzida em 2015 como GoogLeNet e depois também passou a ser chamada de Inception v1. Posteriormente, outras variações

Figura 5: Princípio de funcionamento e estrutura da rede neural convolucional ResNet.

(a) Representação de um bloco residual. A entrada é denota por x e o mapeamento original $\mathcal{F}(x)$ é reformulado em $\mathcal{F}(x) + x$.



(b) Estrutura da arquitetura ResNet. Exemplo considerando 34 camadas de blocos residuais.



Fonte: (HE, K. et al., 2016), Adaptado.

foram sendo criadas buscando o refinamento de alguma propriedade (SZEGEDY; IOFFE et al., 2017). A rede Inception v3 foi apresentada em (SZEGEDY; VANHOUCKE et al., 2016) e acrescenta a ideia de fatoração das convoluções, com o objetivo de reduzir a quantidade de parâmetros utilizados sem diminuir a eficiência da rede. Um exemplo disso é substituir uma convolução 5x5 (25 parâmetros) por duas convoluções 3x3 ($2x9 = 18$ parâmetros), reduzindo em 28% a quantidade de parâmetros utilizados. Mais duas particularidades são importantes: (i) apenas um classificador auxiliar é utilizado, em vez de dois como na primeira versão, e este agora tem a função de regularizador; (ii) é proposta uma maneira eficiente para o processo de redução do tamanho do mapa de características, que convencionalmente é feito por *max pooling* mas que pode ser muito custoso ou muito ambicioso. Tudo isso é consolidado de modo a compor a arquitetura final da rede Inception v3, resultando em um total de 48 camadas.

3.4 BALANCEAMENTO DE DADOS

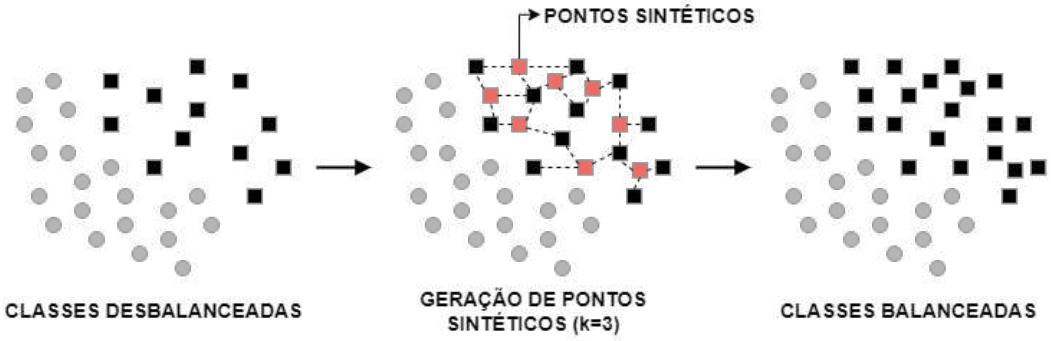
Considerando que a quantidade de informação para cada classe (i.e. doenças) seja diferente, usar a base de dados dessa forma pode prejudicar na construção dos modelos e, assim, nas suas predições. Portanto, antes de passar os dados para o classificador, se faz necessário o balanceamento dos dados.

Para tal, duas abordagens são possíveis: (i) subamostragem, isto é, usar como

referência a classe minoritária e excluir dados das demais; ou (ii) sobreamostragem, isto é, usar como referência a classe majoritária e acrescentar dados nas demais. A primeira opção é mais simples, porém, está atrelada à perda de informação, o que influencia negativamente na performance dos classificadores. Já na sobreamostragem, apesar de normalmente ter um custo computacional maior, nenhuma informação é descartada. Neste trabalho, a estratégia escolhida é a de sobreamostragem.

É comum em técnicas de sobreamostragem que o número de dados da classe minoritária aumente, mas que não seja fornecida nenhuma informação nova ou variação para o modelo de aprendizado de máquina. A abordagem escolhida para este trabalho é a proposta em (CHAWLA et al., 2002), onde as classes de menor incidência são sobreamostradas com exemplos sintéticos. *Synthetic Minority Oversampling Technique* (SMOTE) junta os pontos das classes minoritárias com segmentos de linha e, em seguida, adiciona pontos artificiais nessas linhas. Logo, dados sintéticos são criados entre pontos das classes minoritárias e os k vizinhos mais próximos selecionados aleatoriamente. O procedimento é repetido várias vezes até que as classes minoritárias tenham a mesma proporção da classe majoritária. Ou seja, o gerador de dados sintéticos cria dados que se assemelham à forma ou aos valores dos dados existentes, em vez de simplesmente copiá-los. A Figura 6 ilustra o funcionamento do SMOTE.

Figura 6: Ilustração do funcionamento da sobreamostragem de dados pelo SMOTE. Até que as classes possuam a mesma proporção, pontos artificiais são adicionados em segmentos de linhas que juntam os pontos da classe minoritária e os k vizinhos mais próximos selecionados aleatoriamente.



Fonte: Próprio Autor.

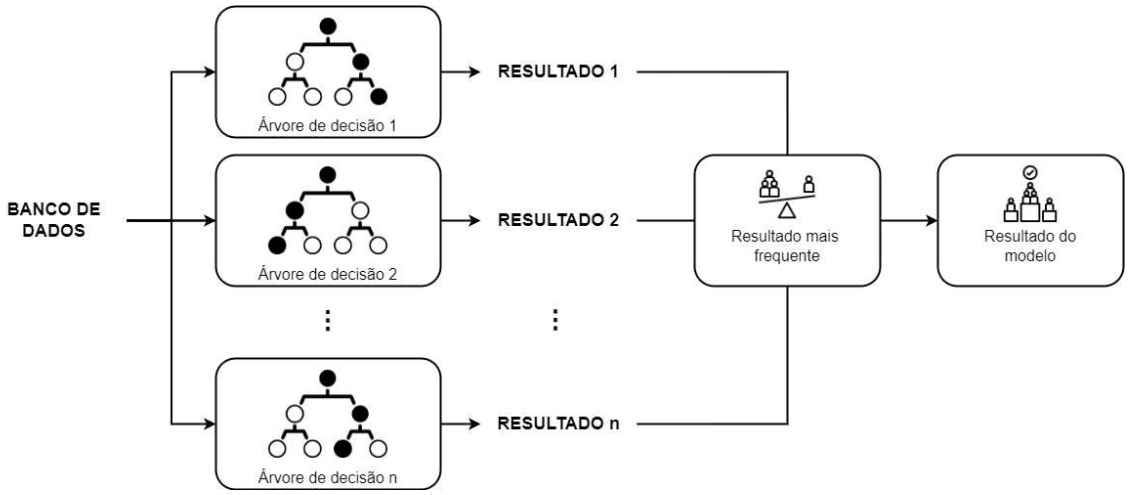
3.5 CLASSIFICAÇÃO

Entende-se por classificador um determinado algoritmo capaz de identificar, com uma certa limitação, a qual categoria pertence uma determinada amostra. Para tal, se faz necessária a utilização de um conjunto de dados para treinar o modelo e torná-lo capaz de realizar as predições. Neste trabalho, o apoio ao diagnóstico médico se dará a partir da avaliação dos dados pelo método *Random Forest*, tendo como base de dados o resultado do pré-processamento da base de imagens.

O método *Random Forest* tem como base o conceito de árvore de decisão. A árvore de decisão funciona de modo a encontrar, dentre todos os atributos, quais são mais relevantes para a classificação em uma determinada categoria. Este algoritmo desenvolve gradativamente uma espécie de árvore onde um nó é expandido assim que houver evidência estatística suficiente de que existe um recurso de divisão ideal e o processo continua até que seja suficiente a classificação em uma classe. Porém, apesar de fácil interpretação e visualização, as árvores de decisão podem se tornar excessivamente complexas, tendem a não ser muito reproduzíveis com dados futuros e pequenas mudanças nos dados podem resultar em árvores de decisão um pouco ou mesmo muito diferentes.

Sabendo disso, o *Random Forest*, conforme ilustrado na Figura 7, consiste em uma quantidade n de árvores de decisão individuais operando como um conjunto. Dada uma amostra da base de dados, cada árvore realiza sua classificação e a categoria mais frequente se torna a previsão do modelo. Essas árvores de decisão são treinadas com diferentes conjuntos de dados e os atributos presentes em cada uma delas são definidos de forma aleatória. Dessa forma, a diversificação das árvores de decisão no *Random Forest* é capaz de fornecer um modelo mais robusto em termos de performance, visto que é diminuída a propagação de erro de árvores individuais.

Figura 7: Demonstração do funcionamento de classificação pelo método *Random Forest*. O resultado do modelo é aquele mais frequente dentre os resultados de n árvores de decisão que são treinadas com diferentes conjuntos de dados e com os atributos presentes definidos de forma aleatória.



Fonte: Próprio Autor.

4 MATERIAIS E MÉTODOS

4.1 BASE DE DADOS

O Atlas Dermatológico (SILVA, 1995) é uma base pública que conta com 11.606 imagens de doenças dermatológicas em geral (última atualização: 03/10/2022). A base de dados utilizada neste trabalho é composta pela extração das imagens referentes às doenças apresentadas na Seção 3. Algumas amostras são apresentadas na Figura 8 (câncer de pele), na Figura 9 (hanseníase) e na Figura 10 (diagnóstico diferencial com a hanseníase). Na Tabela 2 é descrita em detalhes a quantidade de imagens consideradas para cada doença, onde os valores se referem ao total disponível no momento do acesso ao site.

Figura 8: Amostras do banco de imagens referentes aos tipos de câncer de pele.

(a) Melanoma. (b) Carcinoma basocelular. (c) Carcinoma epidermoide.

Fonte: (SILVA, 1995).

Figura 9: Amostras do banco de imagens referentes aos tipos de hanseníase.

(a) Hanseníase Indeterminada. (b) Hanseníase Tuberculoide. (c) Hanseníase Dimorfa. (d) Hanseníase Virchowiana.

Fonte: (SILVA, 1995).

Figura 10: Amostras do banco de imagens referentes às doenças que fazem diagnóstico diferencial com a hanseníase.

(a) Pitiríase Alba. (b) Xantoma Tuberoso. (c) Vitiligo. (d) Psoríase.

(e) Granuloma Anular. (f) Dermatite Seborreica. (g) Sarcoidose. (h) Condrodermatite Nodular Crônica da Hélice.

(i) Nevo Anêmico. (j) Neurofibromatose. (k) Esclerodermia Linear. (l) Nevo Acrômico.

(m) Parapsoríase em Placas. (n) Dermatofitose do Corpo. (o) Tuberculose Cutânea Verrucosa.

Fonte: (SILVA, 1995).

4.2 CENÁRIOS PROPOSTOS

A partir da base de imagens construída para este trabalho, são diversas as possibilidades de experimentos a serem realizados. Dessa forma, os cenários aqui propostos levam em conta aspectos computacionais e, sobretudo, são baseados em relatos de profissionais da área da saúde. Um total de 7 cenários são considerados, divididos da seguinte forma:

- **Hanseníase vs Câncer de pele:** A Figura 11 ilustra dois possíveis casos: na Figura 11a os tipos das duas doenças são agrupados, chegando assim em 2 classes (hanseníase e câncer de pele) e na Figura 11b cada tipo de doença é considerada independente, obtendo-se 7 classes (4 tipos de hanseníase e 3 tipos de câncer de

Tabela 2: Quantidade de imagens por doença que compõem a base de imagens utilizada.

Doença	**Tipo**	**Nº de imagens**
Hanseníase	Virchowiana	276
	Tuberculoide	232
	Dimorfa	145
	Indeterminada	31
Câncer de pele	Carcinoma basocelular	391
	Carcinoma epidermoide	134
	Melanoma	120
Psoríase		125
Dermatofitose do Corpo		110
Neurofibromatose		81
Vitiligo		60
Granuloma Anular		55
Dermatite Seborreica		48
Sarcoidose		34
Condrodermatite Nodular Crônica da Hélice		23
Parapsoríase em Placas		15
Xantoma Tuberoso		10
Esclerodermia Linear		9
Nevo Acrômico		6
Tuberculose Cutânea Verrucosa		5
Nevo Anêmico		4
Pitiríase Alba		2
Total		1916

Fonte: (SILVA, 1995).

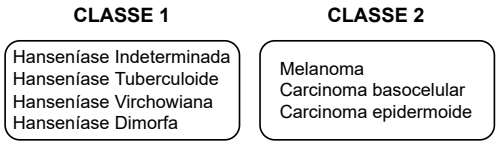
pele);

- **Câncer de pele:** A Figura 12 apresenta o único caso considerando ao se tratar da classificação de uma imagem com suspeita de câncer de pele, esta deve ser classificada entre 3 possíveis classes (cada tipo de câncer de pele);

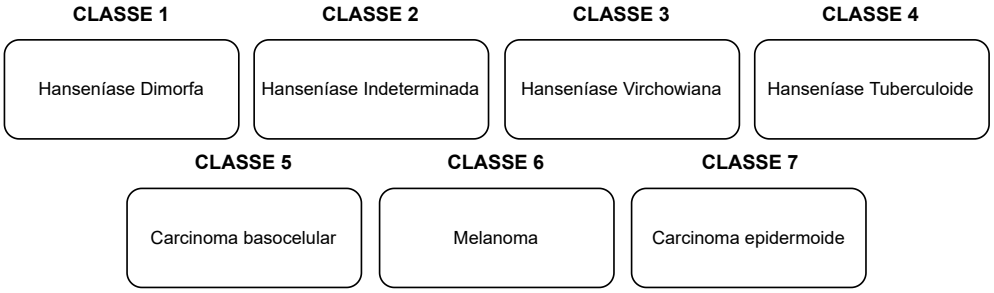
- **Hanseníase e diagnóstico diferencial:** A Figura 13 explica as quatro situações consideradas no diagnóstico da hanseníase. Na Figura 13a uma imagem deve ser classificada entre 4 classes (cada tipo de hanseníase). Além dessas, para a Figura 13b são considerados cada conjunto de doenças que fazem diagnóstico diferencial com cada tipo de hanseníase, totalizando 8 possíveis classes. Já na Figura 13c todas essas são consideradas de forma independente, chegando em um total de 14 classes (doenças com pelo menos 10 imagens disponíveis). Por fim, a Figura 13d considera cada tipo de hanseníase e agrupa em uma única classe todas as doenças que fazem diagnóstico diferencial, obtendo 5 possíveis classes.

Figura 11: Cenários propostos para a classificação de uma imagem entre hanseníase e câncer de pele.

(a) Os tipos de cada doença são agrupados.

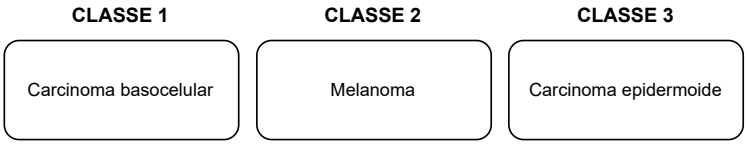


(b) Os tipos de cada doença são considerados independentes.



Fonte: Próprio Autor.

Figura 12: Cenário proposto para a classificação de uma imagem como câncer de pele, onde cada tipo da doença é considerado independente.



Fonte: Próprio Autor.

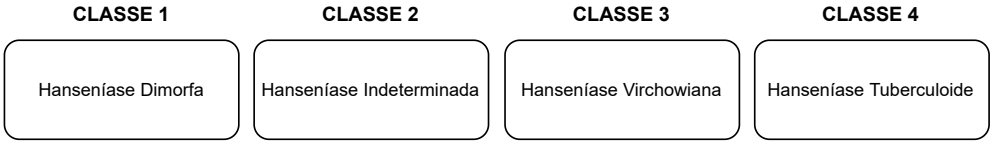
4.3 EXTRAÇÃO DE ATRIBUTOS

A extração de atributos das imagens é executada no *software* Weka 3.8 (FRANK; HALL; WITTEN, 2016) a partir da biblioteca *WekaDeeplearning4j* (LANG et al., 2019). O funcionamento se resume em:

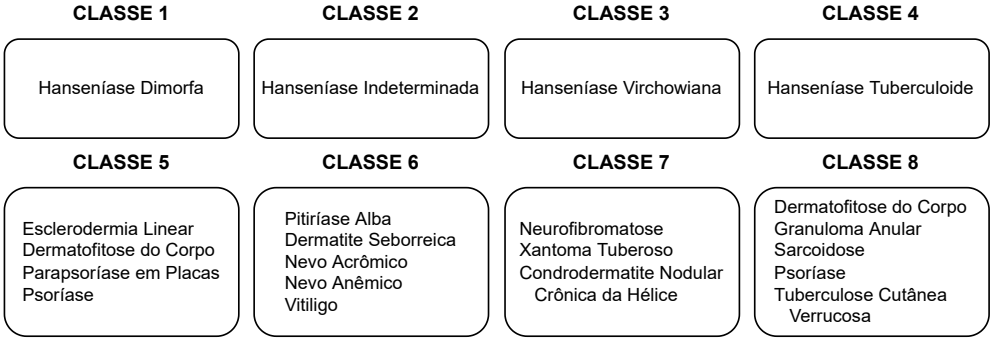
1. Criar um arquivo de extensão *.arff* listando as imagens e suas respectivas classes;
2. Abrir o arquivo criado no *software* e escolher o filtro *Dl4jMlpFilter*;
3. Definir a rede neural profunda desejada (parâmetro *Pretrained zoo model*);
4. Informar que o conteúdo do arquivo *.arff* é do tipo imagem e selecionar o diretório onde estas estão armazenadas (parâmetro *Instance iterator*);
5. Aplicar o filtro e obter os atributos das imagens de acordo com a rede escolhida.

Figura 13: Cenários propostos para a classificação de uma imagem como hanseníase.

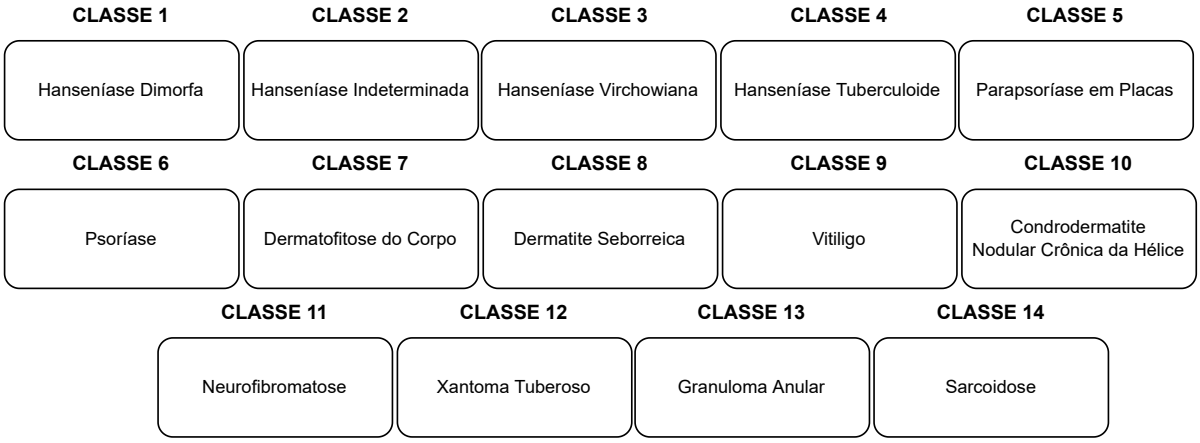
(a) Cada tipo de hanseníase é considerada independente.



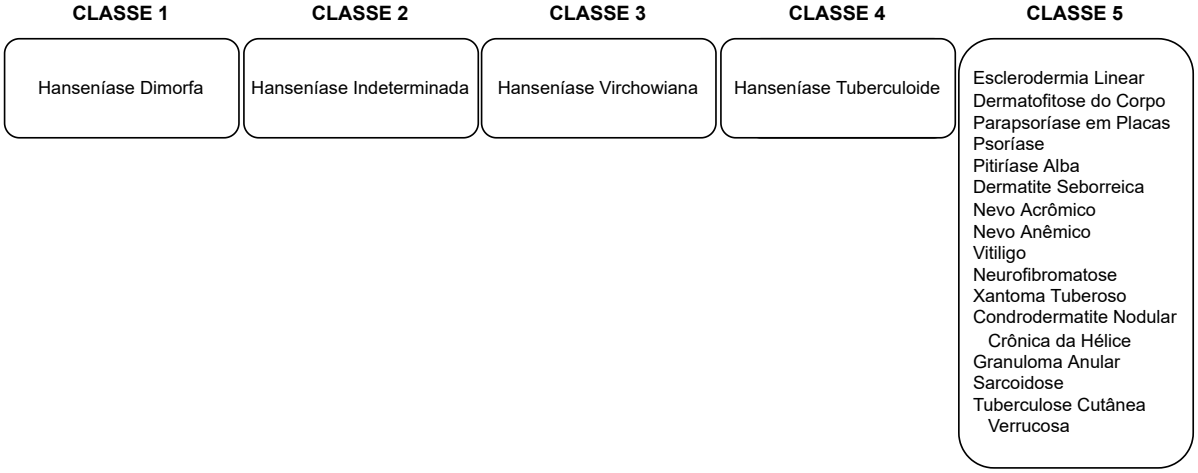
(b) Cada tipo de hanseníase e seus respectivos conjuntos de doenças que fazem diagnóstico diferencial.



(c) Cada tipo de hanseníase e todas as doenças que fazem diagnóstico diferencial são independentes.



(d) Cada tipo de hanseníase e um grupo com as doenças que fazem diagnóstico diferencial.



Fonte: Próprio Autor.

Além disso, padronizou-se as dimensões das imagens para 256x256 pixels e o restante das configurações foram as padrões. As redes neurais profundas utilizadas neste trabalho foram descritas na Seção 3.3 e a quantidade de atributos extraídos por cada uma delas é apresentada na Tabela 3.

Tabela 3: Quantidade de atributos extraídos por cada rede neural profunda utilizada.

Rede neural profunda	**Nº de atributos extraídos**
LeNet	500
SqueezeNet	1000
NASNetMobile	1056
ResNet50	2048
Inception v3	2048

Fonte: Próprio Autor.

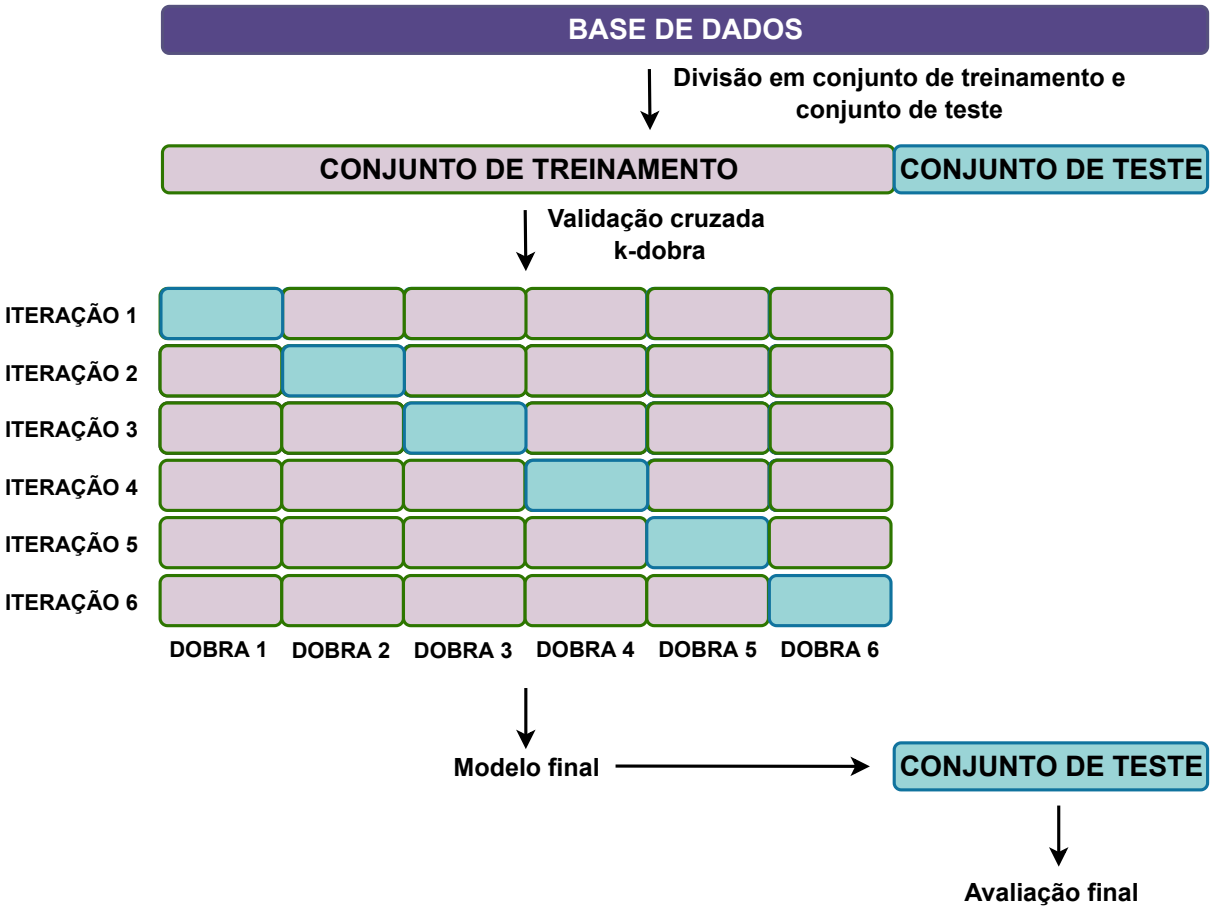
4.4 TREINAMENTO, VALIDAÇÃO E TESTE

O objetivo de treinar e validar um modelo de aprendizado de máquina é avaliar a sua capacidade de generalização. Para tal, a técnica utilizada neste trabalho será a validação cruzada com *k-fold* (ou k-dobra). Esta funciona da seguinte maneira:

1. Separar o conjunto de treinamento em k pedaços (dobras);
2. Repetir para cada pedaço:
   (a) Treina o modelo nos $k - 1$ pedaços restantes;
   (b) Testa o modelo no pedaço escolhido.
3. Gerar métricas com base nos resultados de todas as combinações.

Dessa forma, é possível ter uma ideia do quão suscetível o modelo pode ser a variações nos dados. Por fim, aquele modelo de melhor desempenho deve ser testado com um conjunto de dados que em nenhum momento foi utilizado na sua construção, de forma a obter uma estimativa imparcial da sua eficácia. As etapas de validação cruzada e teste são ilustradas na Figura 14, usando como exemplo $k = 6$.

Figura 14: Processo de treinamento, validação e teste utilizando a técnica de validação cruzada com k-dobra ($k = 6$) para escolher o modelo de melhor desempenho, seguido pelo teste final com amostras até então desconhecidas pelo modelo.



Fonte: Próprio Autor.

4.5 ARRANJO EXPERIMENTAL

Todos os experimentos foram realizados no *software* Weka 3.8 (FRANK; HALL; WITTEN, 2016), e os resultados obtidos foram exportados para arquivos de extensão *.csv* para serem analisados a partir do Python 3.10.8, utilizando as bibliotecas *matplotlib*, *pandas* e *seaborn*. Baseado nos objetivos apresentados na Seção 1.2, os experimentos foram divididos em duas partes:

1. Avaliação dos modelos de rede profunda

   - **Objetivo:** verificar a rede neural profunda que proporciona o melhor desempenho do sistema;
   - **Cenário do experimento:** Figura 11b;

- **Divisão das amostras:** a validação cruzada será executada considerando todas as imagens referentes às doenças abrangidas, ou seja, sem a etapa final de teste;
- **Extração de atributos:** redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3;
- **Balanceamento dos dados:** SMOTE com $k = 5$ vizinhos;
- **Validação cruzada:** 30 repetições de 10 dobras;
- **Classificador:** Random Forest com 50, 100, 150, 200, 250 e 300 árvores;
- **Fluxograma dos experimentos:** Figura 15.

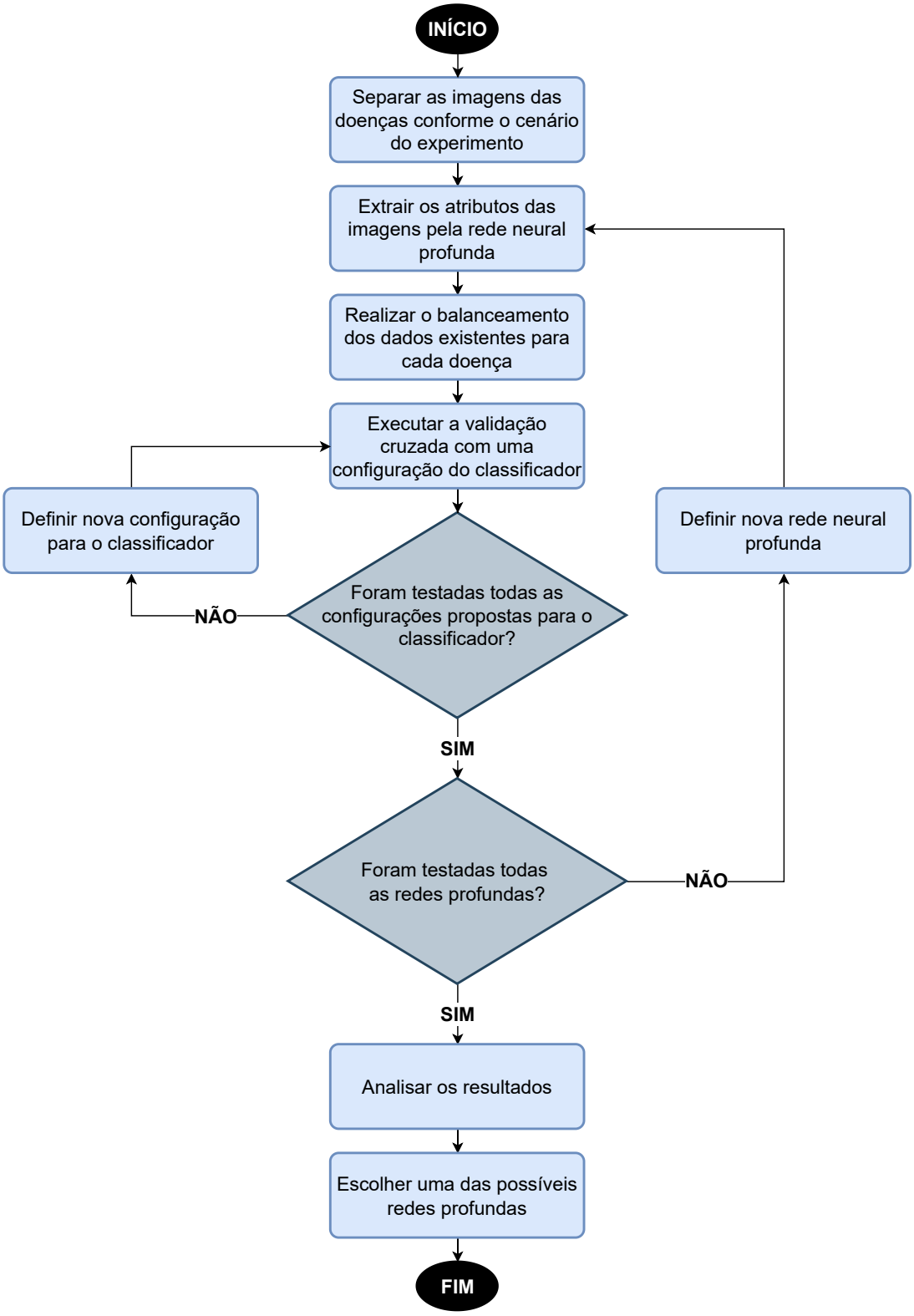
2. Investigação das configurações do classificador

- **Objetivo:** definir os hiperparâmetros do classificador;
- **Cenário do experimento:** todos os 7 cenários apresentados na Seção 4.2;
- **Divisão das amostras:** 80% das imagens para a validação cruzada e 20% para o teste;
- **Extração de atributos:** rede neural profunda escolhida na etapa anterior;
- **Balanceamento dos dados:** SMOTE com $k = 3$ vizinhos;
- **Validação cruzada:** 30 repetições de 10 dobras;
- **Classificador:** Random Forest com 100, 200, 300, 400 e 500 árvores;
- **Fluxograma dos experimentos:** Figura 16.

A configuração escolhida para a validação cruzada implica em $30 * 10 = 300$ resultados obtidos para cada uma das combinações existentes. Esses valores foram definidos de modo a propiciar resultados estatisticamente significativos para a análise.

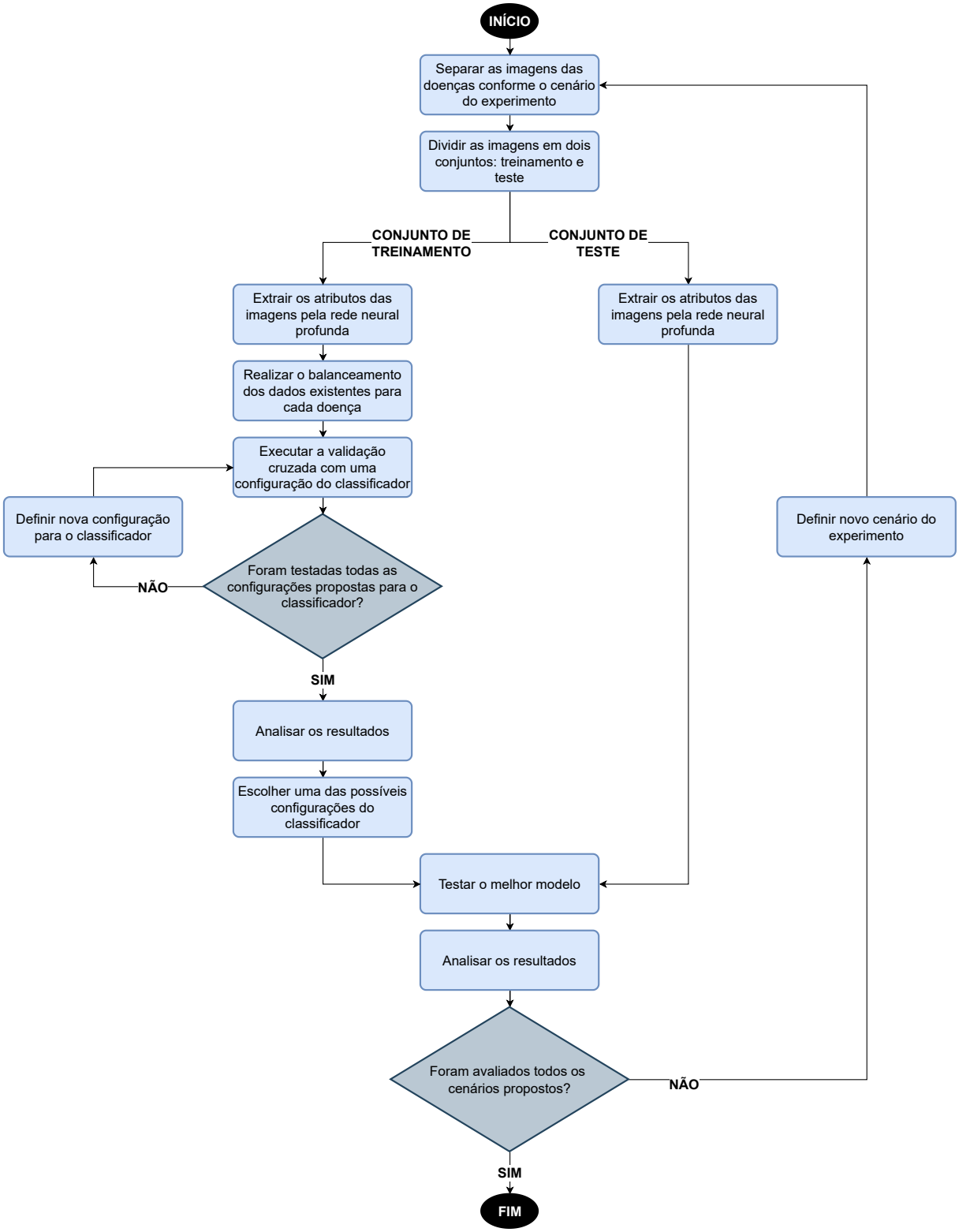
O intuito é que, ao final de todos os experimentos, seja definida uma configuração para cada etapa (extração de atributos e classificação). Isso permitirá a construção de um sistema de apoio a detecção e classificação de doenças, funcionando tal como mostrado na Figura 17, onde a imagem da lesão do paciente é processada pelo sistema e classificada em uma das possíveis doenças.

Figura 15: Fluxograma dos experimentos que visam escolher a rede neural profunda que proporciona o melhor desempenho do sistema de apoio ao diagnóstico.



Fonte: Próprio Autor.

Figura 16: Fluxograma dos experimentos que visam definir os hiperparâmetros do classificador do sistema de apoio ao diagnóstico.



Fonte: Próprio Autor.

Figura 17: O sistema desenvolvido considera que o suporte ao diagnóstico médico se dá após processar a imagem da lesão do paciente e associá-la a uma das possíveis doenças.

Fonte: Próprio Autor.

4.6 MÉTRICAS DE AVALIAÇÃO

As métricas descritas abaixo foram escolhidas para avaliar as performances dos modelos após passar pela etapa de classificação. As expressões matemáticas de cada uma delas podem ser vistas na Tabela 4. Além disso, outra métrica será o tempo decorrido tanto no treinamento quanto no teste de um determinado modelo durante a validação cruzada, a fim de analisar o custo computacional correspondente. Por fim, após o treinamento, os modelos serão avaliados a partir de matrizes de confusão, sendo esta uma maneira eficiente e visual de comparar as previsões com os valores reais.

Referente às previsões de um determinado modelo, TN é o número de exemplos negativos corretamente classificados (Verdadeiro Negativo - *True Negative*), FP é o número de exemplos negativos classificados incorretamente como positivos (Falso Positivo - *False Positive*), FN é o número de exemplos positivos classificados incorretamente como negativos (Falso Negativo - *False Negative*) e TP é o número de exemplos positivos corretamente classificados (Verdadeiro Positivo - *True Positive*).

- **Acurácia:** representa a taxa de acerto geral. Ou seja, dentre todas as previsões de um determinado método, quantas destas foram classificadas como Verdadeiros Positivos (TP) e Verdadeiros Negativos (TN);

- **Especificidade:** corresponde à taxa de Verdadeiros Negativos (TN) dentre todos aqueles que deveriam ser classificados como negativo. Ou seja, mostra a capacidade do método em detectar corretamente o que é negativo;

- **Sensibilidade:** corresponde à taxa de Verdadeiros Positivos (TP) dentre todos

aqueles que deveriam ser classificados como verdadeiro. Ou seja, mostra a capacidade do método em detectar corretamente o que é positivo;

- **Área sob a curva ROC:** a curva ROC é uma representação gráfica da relação entre especificidade e sensibilidade. Calcular a área sob essa curva significa mensurar a capacidade discriminativa de um método. Ou seja, medir a capacidade de um teste classificar corretamente entre as possíveis classes. Quando este valor é próximo a 1, indica que o classificador é capaz de discriminar os casos. Se o classificador não for capaz de distinguir entre as classes, a área será igual a 0,5;

- **Índice Kappa:** é calculado com base na matriz de confusão e compara a previsão de um método com o resultado esperado. Em contraste com o cálculo da acurácia, o índice kappa leva em consideração o desequilíbrio na distribuição de classes (nesse caso, a quantidade de previsões positivas e negativas dos métodos). Em sua expressão matemática, entende-se ρ_o como a concordância observada (acurácia) e ρ_e como a concordância esperada.

Tabela 4: Expressões matemáticas das métricas escolhidas para avaliar o desempenho dos classificadores. As predições podem ser classificadas como TP (Verdadeiro Positivo), TN (Verdadeiro Negativo), FP (Falso Positivo) ou FN (Falso Negativo). A análise de confiabilidade dos resultados é realizada a partir da concordância observada (acurácia) ρ_o e da concordância esperada ρ_e.

Métricas	**Expressões Matemáticas**
Acurácia	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensibilidade	$\frac{TP}{TP+FN}$
Especificidade	$\frac{TN}{TN+FP}$
Área sob a curva ROC	$AUC = \int TPR\ d(FPR)$
Kappa	$\frac{\rho_o - \rho_e}{1-\rho_e}$, onde $\rho_e = \frac{(TP+FP)(TP+FN)+(FN+TN)(FP+TN)}{(TP+FP+FN+TN)^2}$

Fonte: Próprio Autor.

5 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados e discutidos todos os resultados obtidos ao longo dos experimentos. Serão utilizadas tabelas para exibir os valores de média e desvio padrão, além de gráficos boxplots para cada métrica definida na Seção 4.6. Primeiro serão analisadas as redes profundas que realizam a extração de atributos das imagens, seguido da análise das diferentes configurações do classificador e no final será escolhida a melhor combinação de rede profunda e classificador tendo em vista a problemática tratada.

5.1 AVALIAÇÃO DOS MODELOS DE REDE PROFUNDA

Primeiramente são apresentados a média e o desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos das imagens pelas redes neurais profundas LeNet (Tabela 5), SqueezeNet (Tabela 6), NASNetMobile (Tabela 7), ResNet50 (Tabela 8) e Inception v3 (Tabela 9), todos com balanceamento de dados pelo SMOTE com $k = 5$ vizinhos.

Tabela 5: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda LeNet e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 50 árvores	0,7735	0,0246	0,9438	0,0159	0,5229	0,0782	0,8454	0,0315	0,7358	0,0287
RF 100 árvores	0,8035	0,0239	0,9537	0,0139	0,5408	0,0785	0,8663	0,0311	0,7707	0,0279
RF 150 árvores	0,8171	0,0224	0,9583	0,0137	0,5516	0,0743	0,8735	0,0308	0,7866	0,0262
RF 200 árvores	0,8228	0,0214	0,9592	0,0130	0,5606	0,0738	0,8776	0,0305	0,7933	0,0250
RF 250 árvores	0,8268	0,0212	0,9607	0,0126	0,5638	0,0735	0,8806	0,0299	0,7979	0,0248
RF 300 árvores	0,8292	0,0215	0,9612	0,0119	0,5651	0,0742	0,8818	0,0298	0,8008	0,0251

Fonte: Próprio Autor.

Tabela 6: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda SqueezeNet e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 50 árvores	0,7686	0,0256	0,9456	0,0151	0,6132	0,0802	0,9081	0,0232	0,7300	0,0299
RF 100 árvores	0,7835	0,0245	0,9506	0,0139	0,6281	0,0746	0,9162	0,0222	0,7474	0,0286
RF 150 árvores	0,7895	0,0250	0,9522	0,0139	0,6372	0,0780	0,9190	0,0219	0,7544	0,0292
RF 200 árvores	0,7922	0,0242	0,9531	0,0136	0,6393	0,0754	0,9208	0,0215	0,7576	0,0283
RF 250 árvores	0,7946	0,0235	0,9538	0,0133	0,6428	0,0791	0,9218	0,0214	0,7604	0,0274
RF 300 árvores	0,7961	0,0234	0,9540	0,0137	0,6468	0,0776	0,9223	0,0213	0,7621	0,0273

Fonte: Próprio Autor.

Tabela 7: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda NASNetMobile e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 50 árvores	0,8256	0,0217	0,9530	0,0143	0,6350	0,0809	0,9078	0,0245	0,7966	0,0254
RF 100 árvores	0,8566	0,0199	0,9609	0,0133	0,6744	0,0773	0,9237	0,0210	0,8327	0,0232
RF 150 árvores	0,8676	0,0194	0,9626	0,0130	0,6920	0,0747	0,9289	0,0205	0,8455	0,0227
RF 200 árvores	0,8738	0,0184	0,9642	0,0122	0,7001	0,0728	0,9317	0,0199	0,8528	0,0215
RF 250 árvores	0,8780	0,0179	0,9649	0,0120	0,7082	0,0720	0,9336	0,0195	0,8577	0,0209
RF 300 árvores	0,8805	0,0181	0,9657	0,0120	0,7113	0,0682	0,9347	0,0193	0,8606	0,0211

Fonte: Próprio Autor.

Tabela 8: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda ResNet50 e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 50 árvores	0,7862	0,0224	0,9526	0,0149	0,5680	0,0774	0,8902	0,0255	0,7505	0,0261
RF 100 árvores	0,8139	0,0220	0,9604	0,0132	0,5918	0,0764	0,9054	0,0242	0,7829	0,0256
RF 150 árvores	0,8256	0,0219	0,9646	0,0122	0,6048	0,0761	0,9110	0,0237	0,7966	0,0255
RF 200 árvores	0,8302	0,0217	0,9658	0,0120	0,6090	0,0778	0,9137	0,0235	0,8019	0,0253
RF 250 árvores	0,8338	0,0217	0,9668	0,0118	0,6123	0,0769	0,9153	0,0234	0,8061	0,0253
RF 300 árvores	0,8356	0,0214	0,9673	0,0118	0,6125	0,0782	0,9165	0,0234	0,8081	0,0250

Fonte: Próprio Autor.

Tabela 9: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

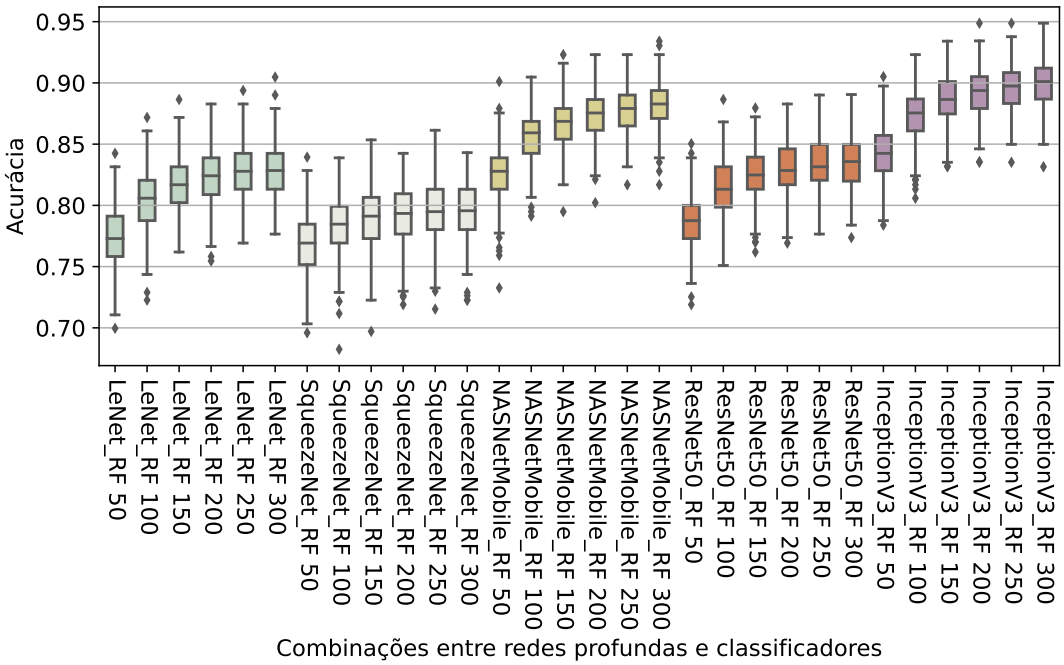
CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 50 árvores	0,8425	0,0217	0,9645	0,0123	0,6487	0,0783	0,9261	0,0197	0,8163	0,0253
RF 100 árvores	0,8741	0,0203	0,9707	0,0112	0,6864	0,0760	0,9418	0,0184	0,8532	0,0237
RF 150 árvores	0,8859	0,0193	0,9730	0,0112	0,7100	0,0739	0,9480	0,0178	0,8669	0,0225
RF 200 árvores	0,8926	0,0192	0,9741	0,0112	0,7216	0,0728	0,9502	0,0189	0,8747	0,0225
RF 250 árvores	0,8962	0,0188	0,9747	0,0108	0,7283	0,0730	0,9520	0,0168	0,8789	0,0220
RF 300 árvores	0,8990	0,0184	0,9752	0,0105	0,7341	0,0733	0,9531	0,0168	0,8822	0,0215

Fonte: Próprio Autor.

Também são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 18), especificidade (Figura 19), sensibilidade (Figura 20), área sob curva ROC (Figura 21) e índice kappa (Figura 22), considerando a extração de atributos das imagens pelas redes neurais profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3, além do balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.

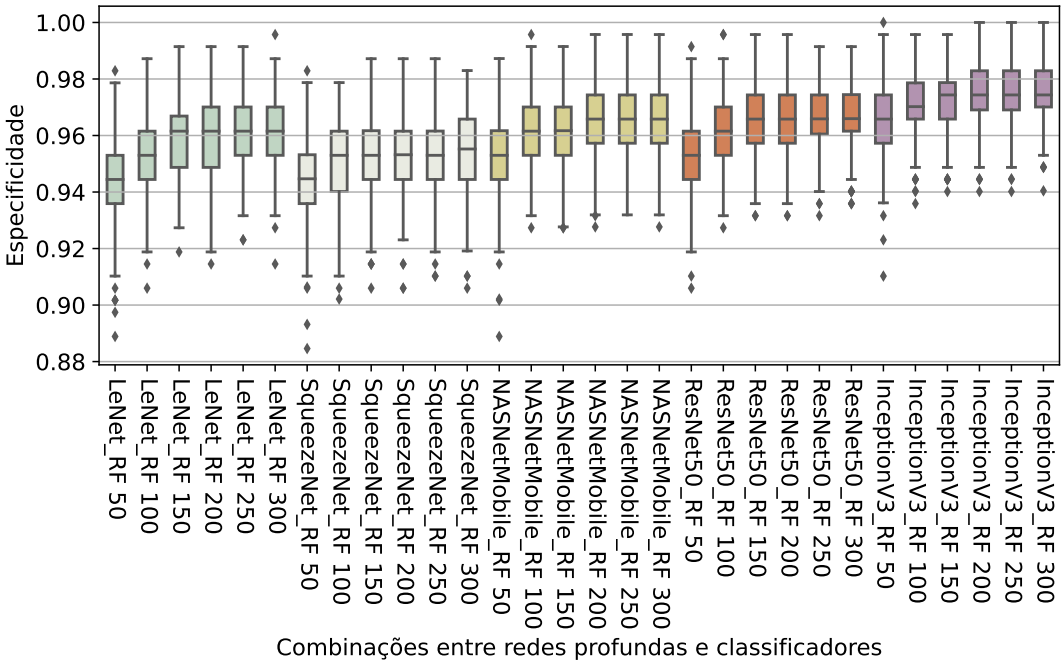
Comparando uma mesma configuração do classificador com as diferentes formas de extração de atributos, é possível observar que na maioria das vezes os resultados obtidos a partir da rede Inception v3 são os melhores ou estão entre os melhores. Nas tabelas, observa-se uma média sempre superior e um desvio padrão em geral inferior à maioria. Nas figuras, a distribuição do resultados normalmente se encontra mais próxima do 100%.

Figura 18: Acurácia da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.



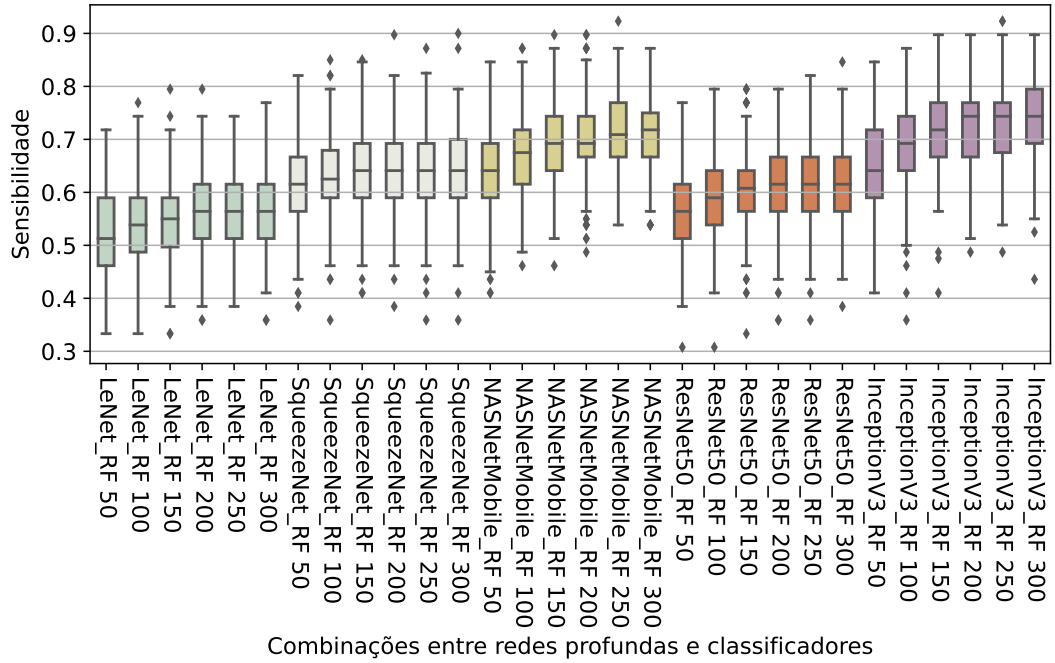
Fonte: Próprio Autor.

Figura 19: Especificidade da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.



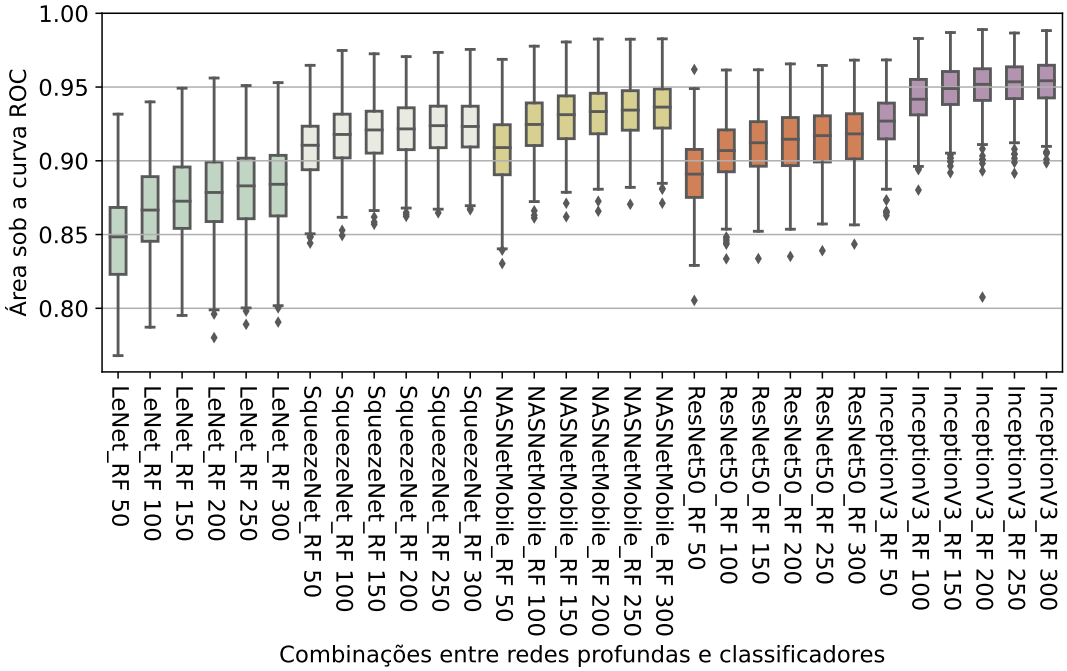
Fonte: Próprio Autor.

Figura 20: Sensibilidade da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.



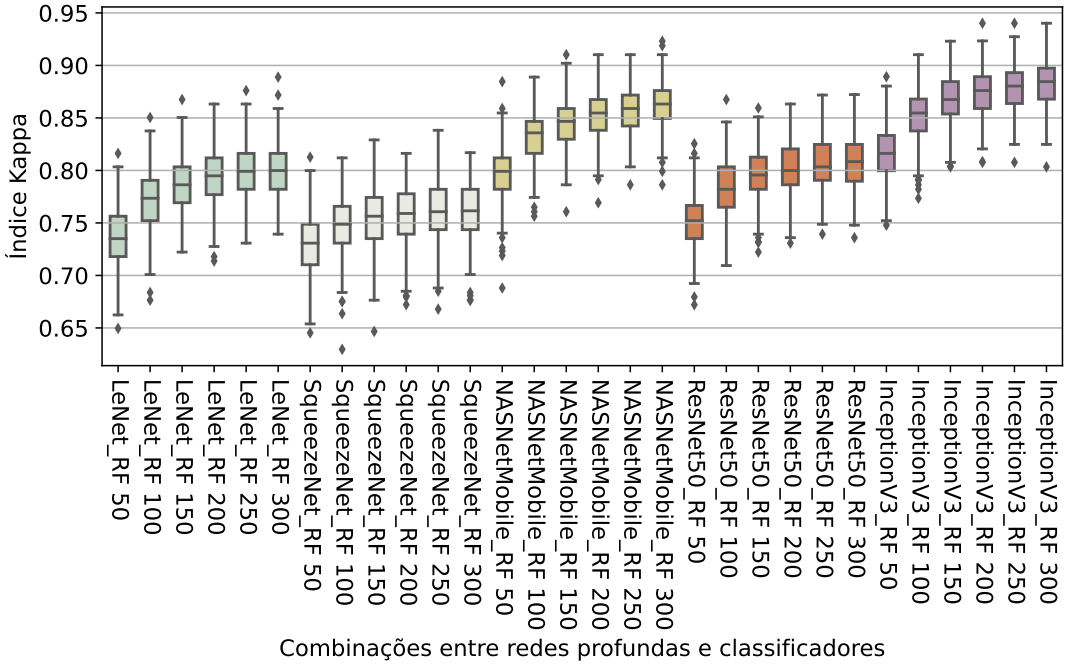
Fonte: Próprio Autor.

Figura 21: Área sob a curva ROC da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.



Fonte: Próprio Autor.

Figura 22: Índice kappa da classificação das imagens com Random Forest de 50, 100, 150, 200, 250 e 300 árvores após extração de atributos pelas redes profundas LeNet, SqueezeNet, NASNetMobile, ResNet50 e Inception v3; e balanceamento dos dados pelo SMOTE com $k = 5$ vizinhos.



Fonte: Próprio Autor.

Referente às diferentes configurações do classificador, o Random Forest com 50 árvores sempre foi aquele de pior desempenho. Porém, entre uma configuração de 100 árvores e de 300 árvores, a diferença de desempenho ultrapassou 3% apenas duas vezes: sensibilidade referente à rede NASNetMobile (Tabela 7) e à rede Inception v3 (Tabela 9). Por isso, outras configurações do classificador serão avaliadas antes de qualquer definição.

5.2 INVESTIGAÇÃO DAS CONFIGURAÇÕES DO CLASSIFICADOR

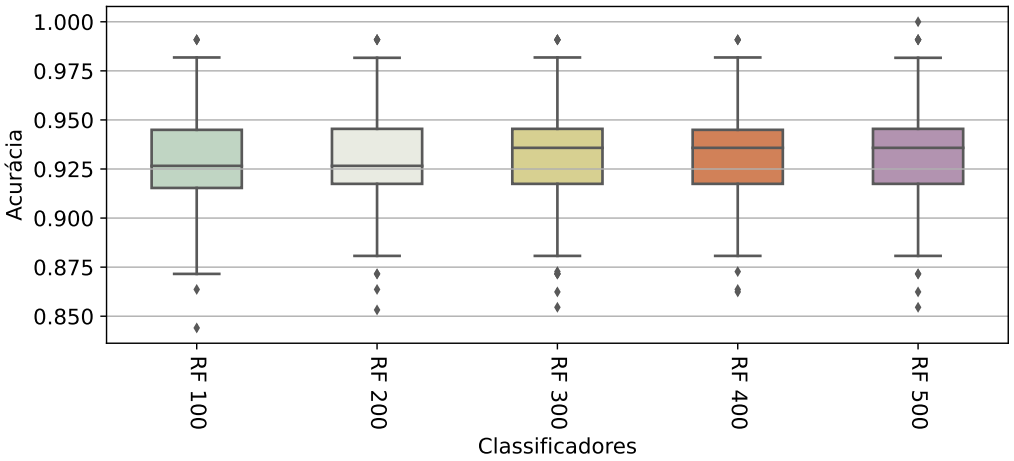
Para os experimentos considerando o cenário representado pela Figura 11a (tipos de hanseníase e câncer de pele agrupados em 2 classes distintas), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 23), especificidade (Figura 24), sensibilidade (Figura 25), área sob curva ROC (Figura 26) e índice kappa (Figura 27), além dos respectivos valores de média e desvio padrão (Tabela 10), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 10: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9290	0.0241	0.9332	0.0336	0.9249	0.0358	0.9833	0.0090	0.8581	0.0482
RF 200 árvores	0.9302	0.0238	0.9344	0.0341	0.9260	0.0350	0.9846	0.0082	0.8604	0.0475
RF 300 árvores	0.9315	0.0245	0.9366	0.0341	0.9265	0.0359	0.9850	0.0081	0.8631	0.0490
RF 400 árvores	0.9312	0.0236	0.9367	0.0330	0.9256	0.0345	0.9852	0.0080	0.8623	0.0473
RF 500 árvores	0.9313	0.0240	0.9380	0.0330	0.9246	0.0358	0.9853	0.0079	0.8626	0.0481

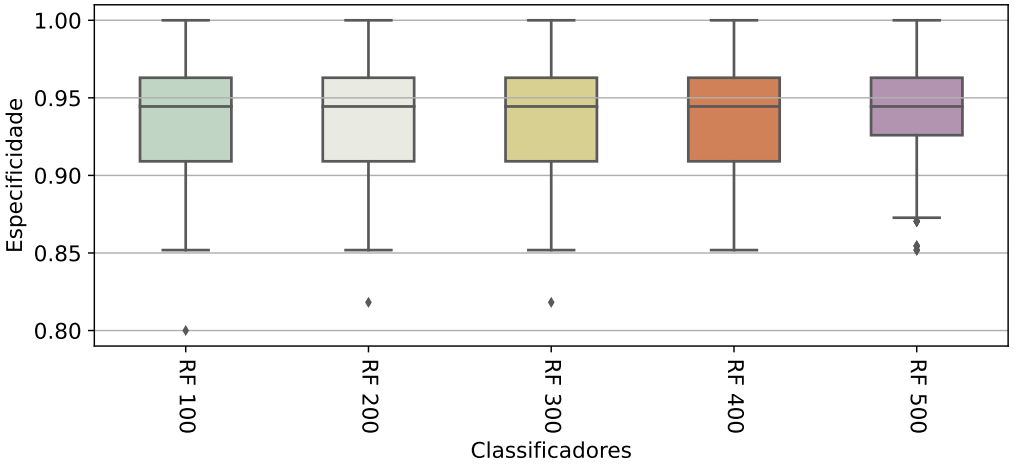
Fonte: Próprio Autor.

Figura 23: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



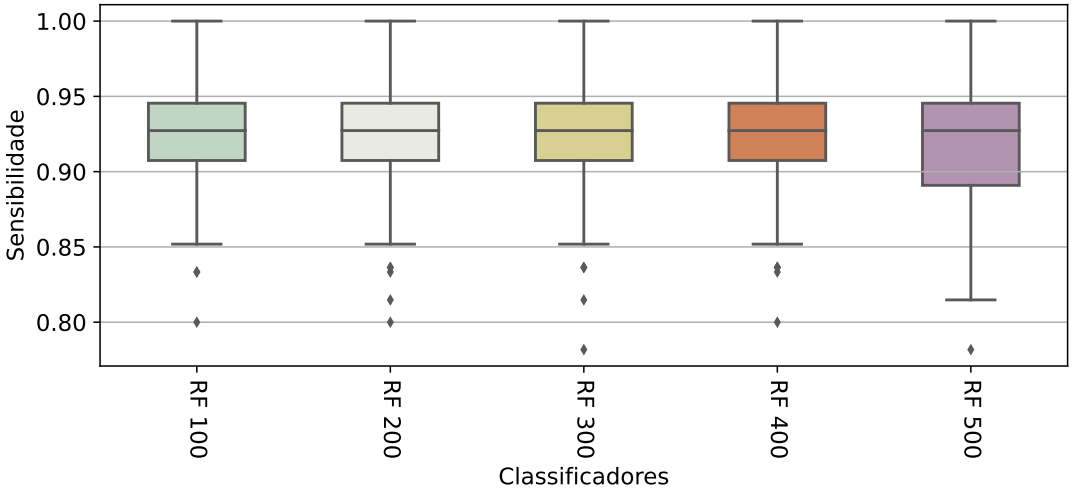
Fonte: Próprio Autor.

Figura 24: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



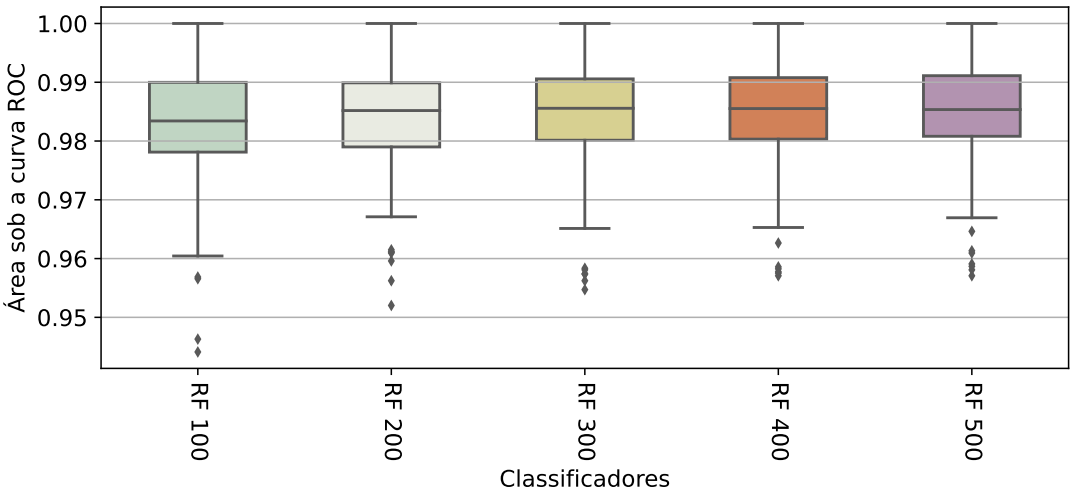
Fonte: Próprio Autor.

Figura 25: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



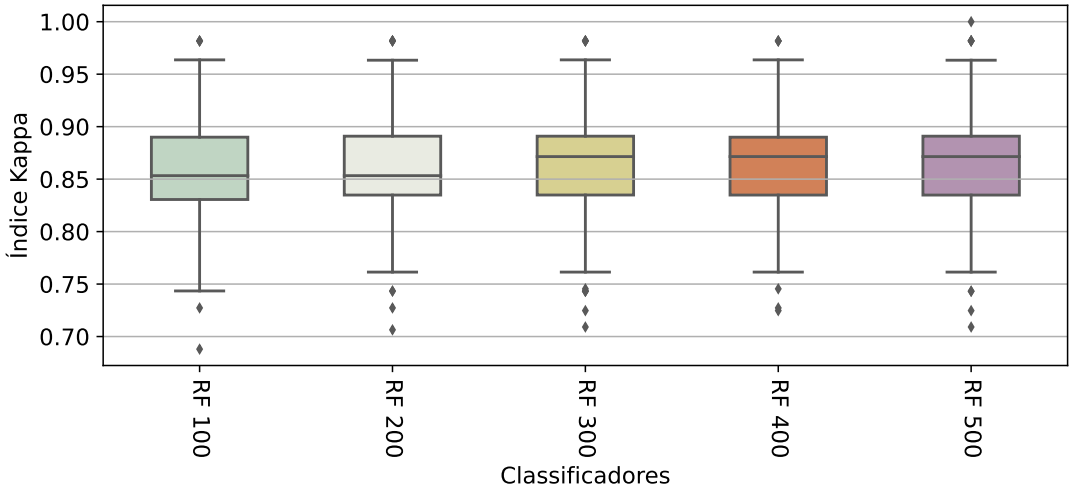
Fonte: Próprio Autor.

Figura 26: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 27: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

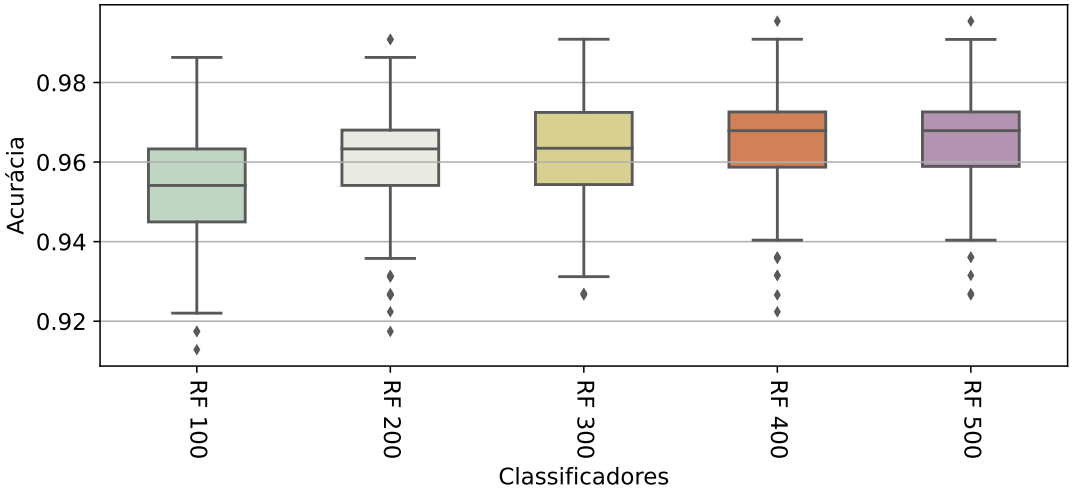
Para os experimentos considerando o cenário representado pela Figura 11b (tipos de hanseníase e câncer de pele considerados independentes), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 28), especificidade (Figura 29), sensibilidade (Figura 30), área sob curva ROC (Figura 31) e índice kappa (Figura 32), além dos respectivos valores de média e desvio padrão (Tabela 11), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 11: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9533	0.0136	0.9936	0.0064	0.9818	0.0227	0.9991	0.0013	0.9455	0.0159
RF 200 árvores	0.9609	0.0134	0.9955	0.0056	0.9860	0.0194	0.9994	0.0010	0.9544	0.0156
RF 300 árvores	0.9637	0.0123	0.9963	0.0049	0.9881	0.0187	0.9995	0.0008	0.9577	0.0143
RF 400 árvores	0.9657	0.0118	0.9968	0.0047	0.9887	0.0171	0.9996	0.0007	0.9600	0.0138
RF 500 árvores	0.9668	0.0115	0.9967	0.0045	0.9887	0.0180	0.9996	0.0008	0.9613	0.0134

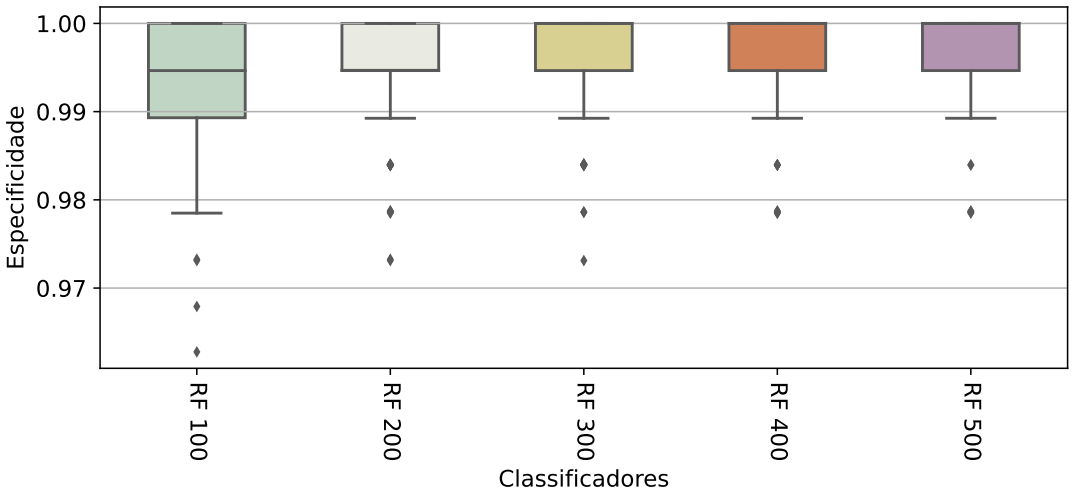
Fonte: Próprio Autor.

Figura 28: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



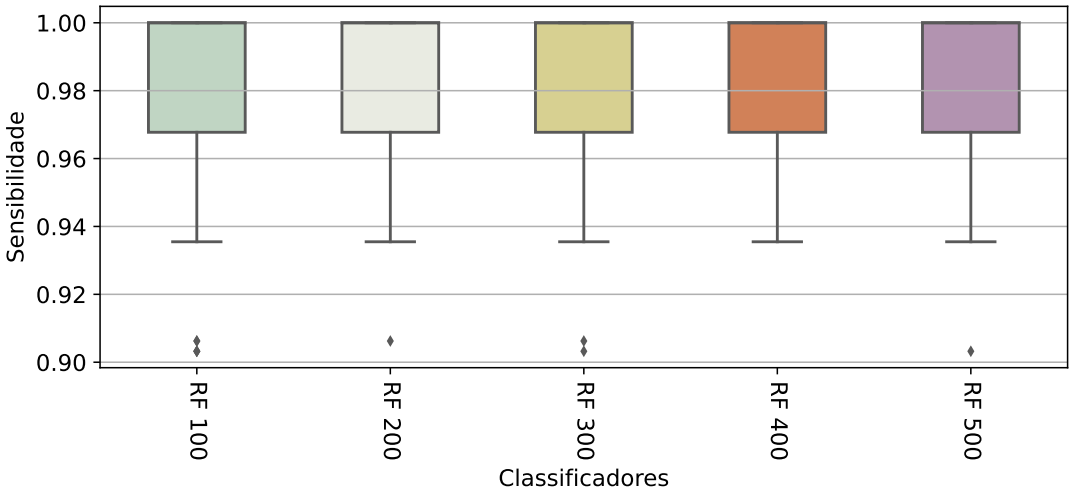
Fonte: Próprio Autor.

Figura 29: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



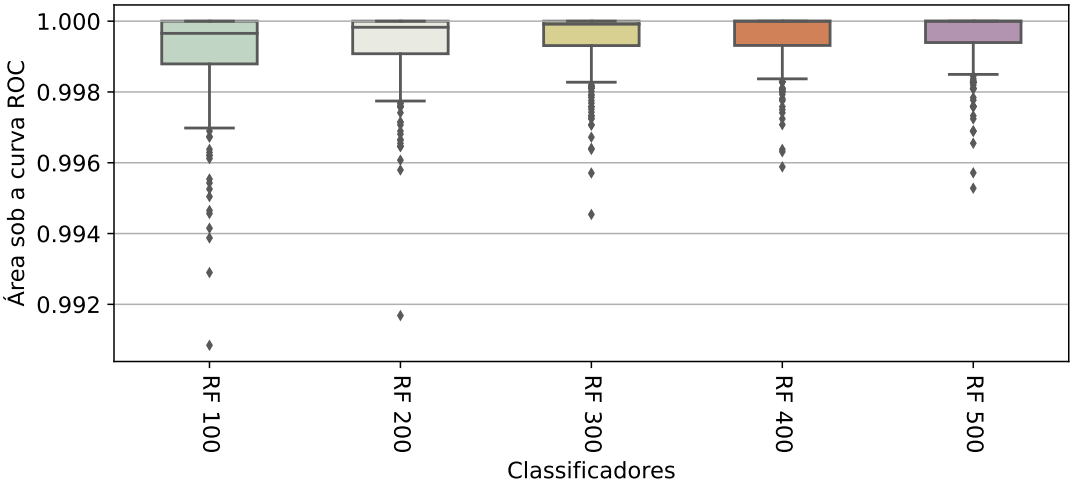
Fonte: Próprio Autor.

Figura 30: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



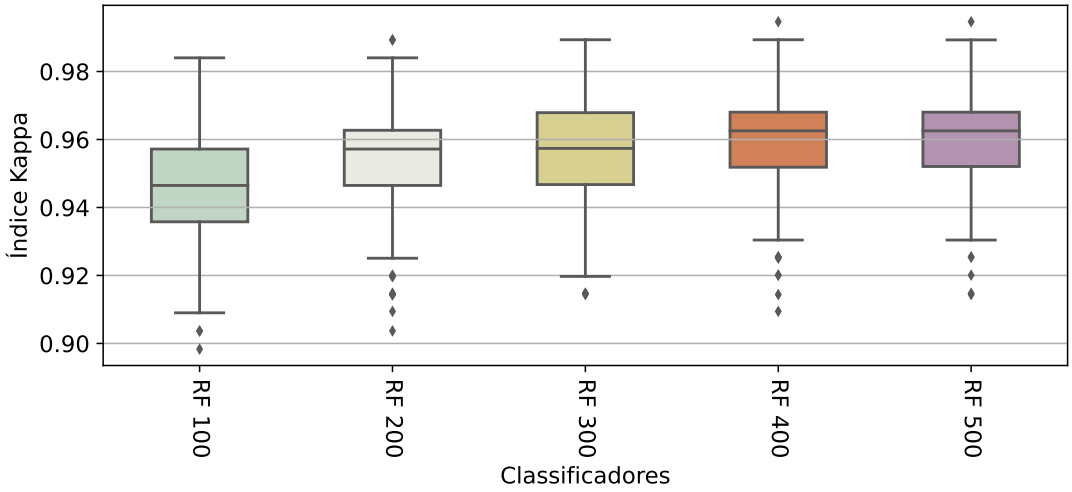
Fonte: Próprio Autor.

Figura 31: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 32: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 11b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

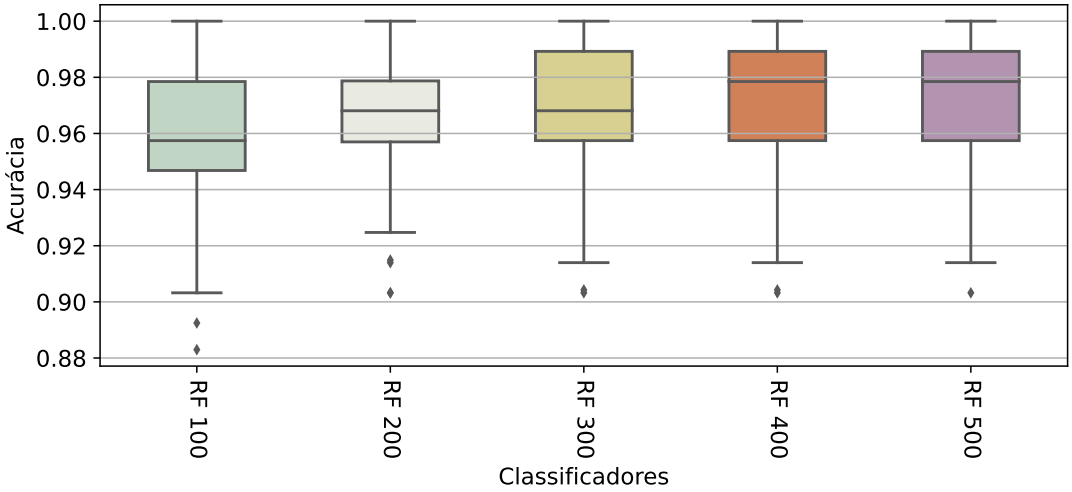
Para os experimentos considerando o cenário representado pela Figura 12 (tipos de câncer de pele considerados independentes), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 33), especificidade (Figura 34), sensibilidade (Figura 35), área sob curva ROC (Figura 36) e índice kappa (Figura 37), além dos respectivos valores de média e desvio padrão (Tabela 12), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 12: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9604	0.0204	0.9763	0.0197	0.9341	0.0449	0.9942	0.0052	0.9405	0.0307
RF 200 árvores	0.9682	0.0186	0.9805	0.0172	0.9481	0.0428	0.9963	0.0039	0.9522	0.0279
RF 300 árvores	0.9705	0.0188	0.9813	0.0176	0.9532	0.0414	0.9968	0.0035	0.9557	0.0283
RF 400 árvores	0.9718	0.0185	0.9821	0.0173	0.9556	0.0409	0.9972	0.0033	0.9577	0.0278
RF 500 árvores	0.9721	0.0189	0.9824	0.0170	0.9555	0.0416	0.9973	0.0030	0.9582	0.0283

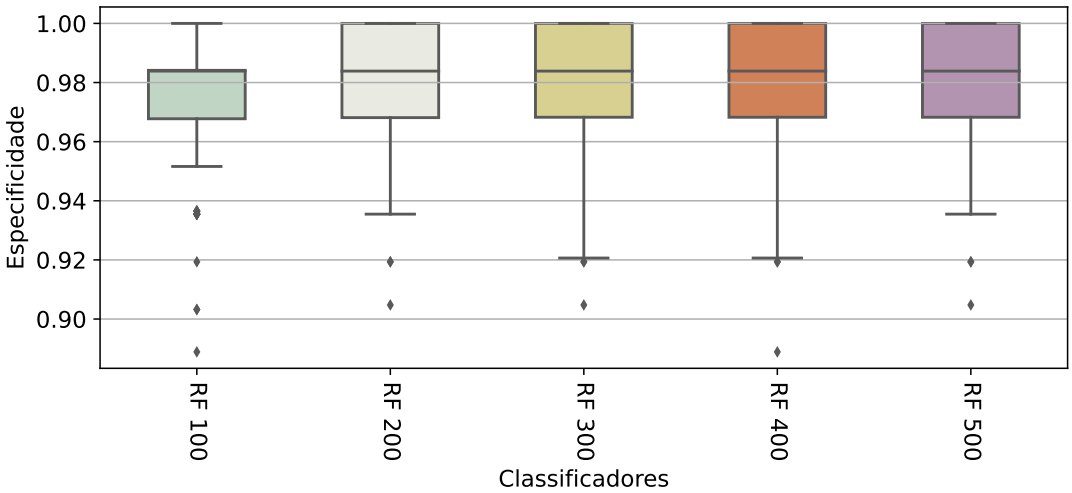
Fonte: Próprio Autor.

Figura 33: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



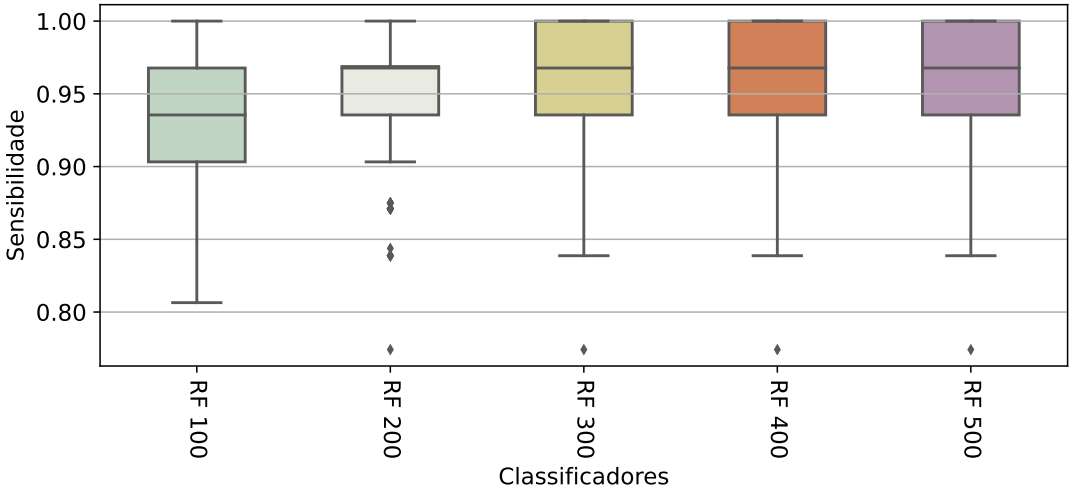
Fonte: Próprio Autor.

Figura 34: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



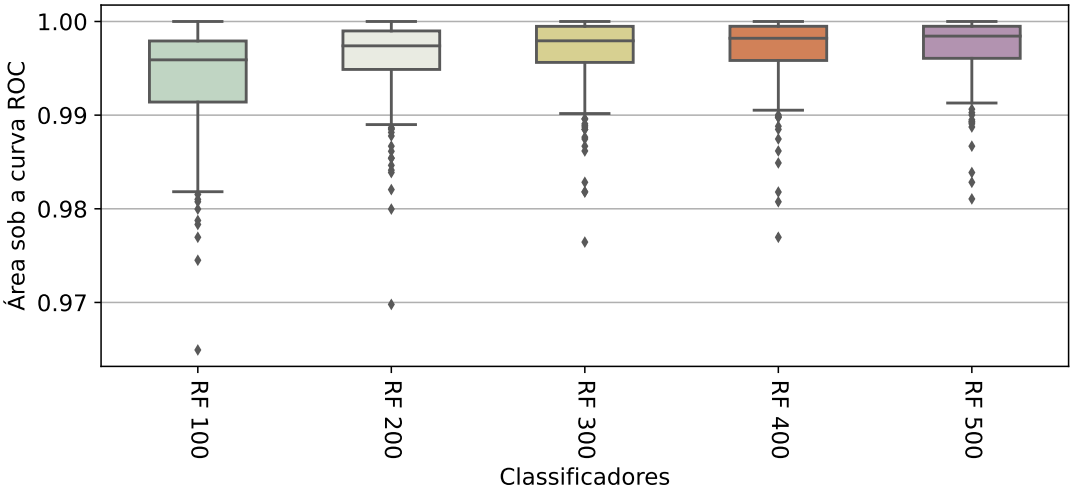
Fonte: Próprio Autor.

Figura 35: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



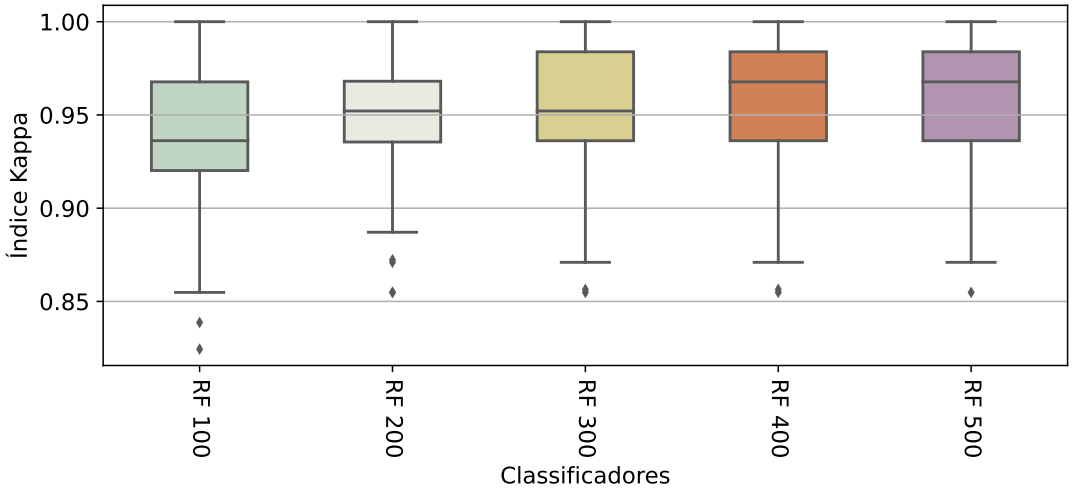
Fonte: Próprio Autor.

Figura 36: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 37: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 12 após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

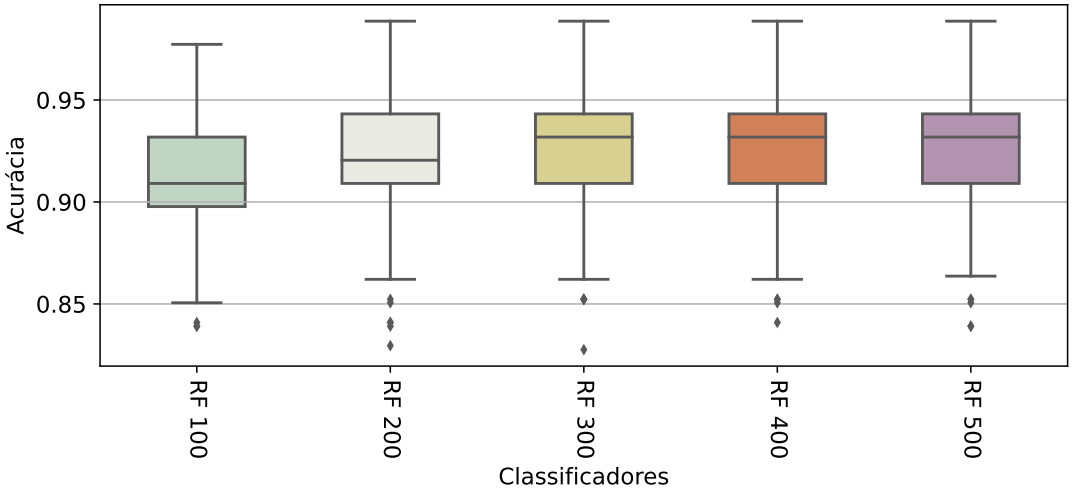
Para os experimentos considerando o cenário representado pela Figura 13a (tipos de hanseníase considerados independentes), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 38), especificidade (Figura 39), sensibilidade (Figura 40), área sob curva ROC (Figura 41) e índice kappa (Figura 42), além dos respectivos valores de média e desvio padrão (Tabela 13), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 13: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9135	0.0282	0.9692	0.0212	0.9357	0.0533	0.9932	0.0071	0.8846	0.0375
RF 200 árvores	0.9230	0.0279	0.9755	0.0203	0.9431	0.0499	0.9947	0.0058	0.8973	0.0373
RF 300 árvores	0.9251	0.0272	0.9774	0.0189	0.9416	0.0526	0.9951	0.0056	0.9001	0.0363
RF 400 árvores	0.9272	0.0269	0.9787	0.0190	0.9439	0.0506	0.9954	0.0054	0.9030	0.0358
RF 500 árvores	0.9283	0.0266	0.9783	0.0185	0.9454	0.0493	0.9955	0.0053	0.9044	0.0355

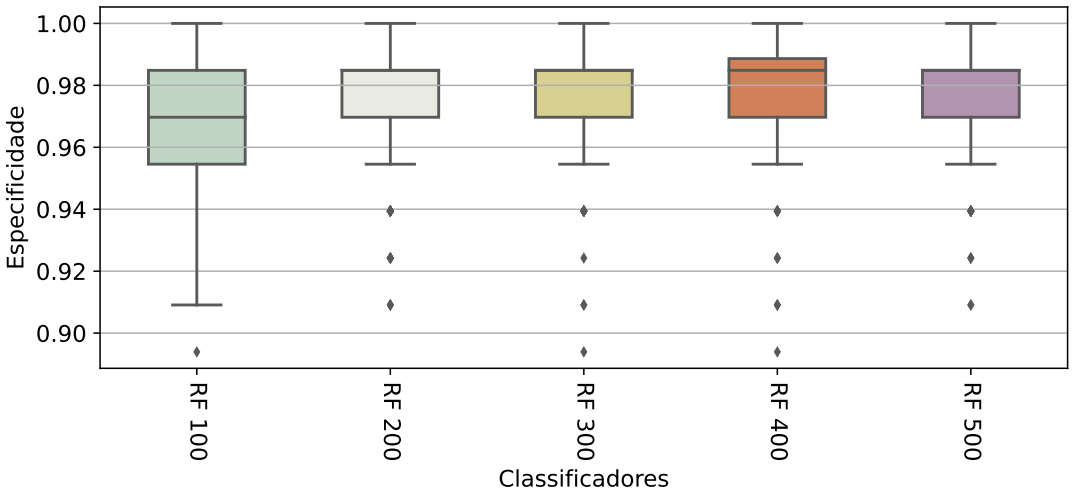
Fonte: Próprio Autor.

Figura 38: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



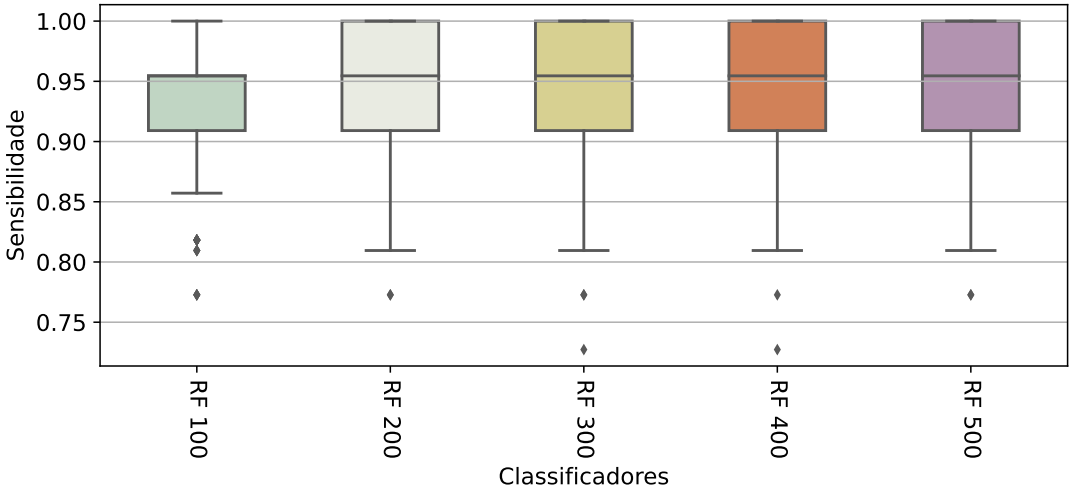
Fonte: Próprio Autor.

Figura 39: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



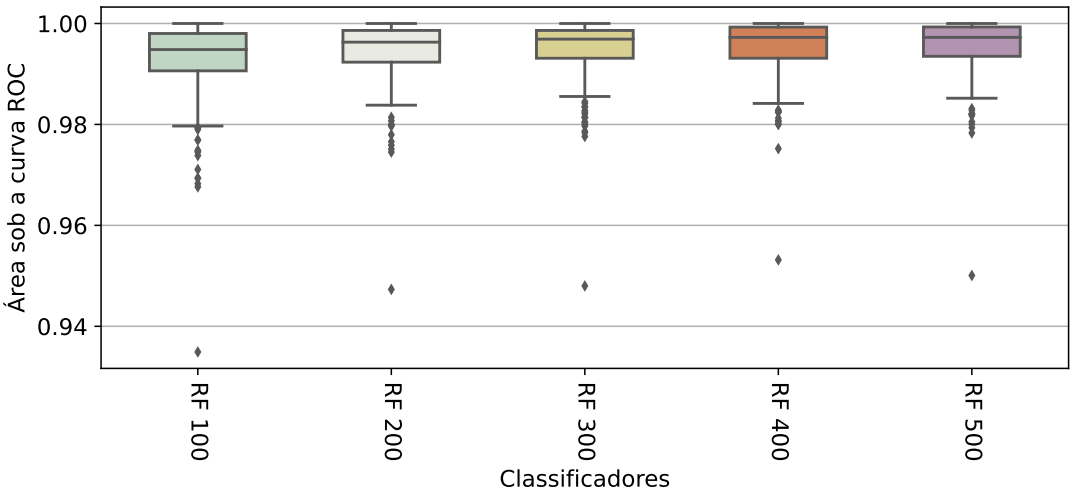
Fonte: Próprio Autor.

Figura 40: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



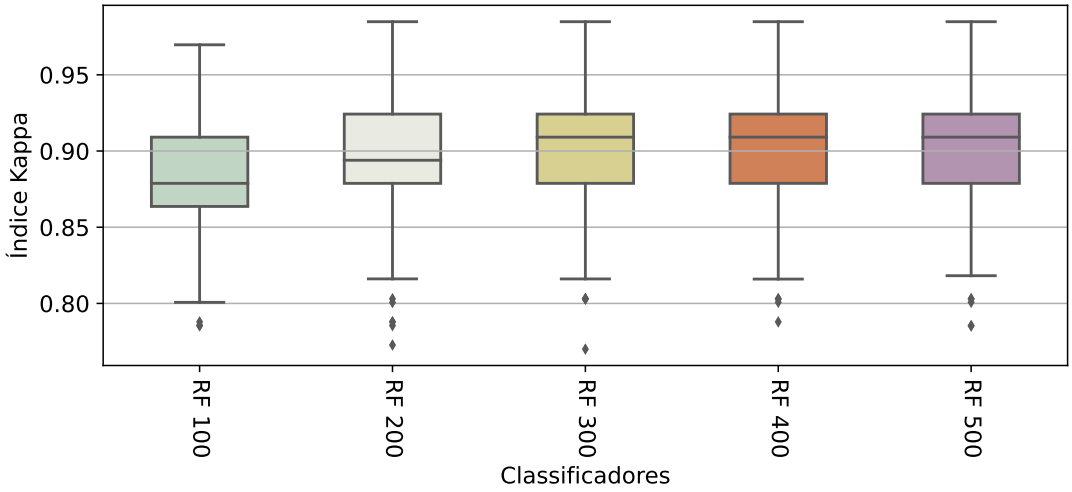
Fonte: Próprio Autor.

Figura 41: Área sob a cura ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 42: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13a após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

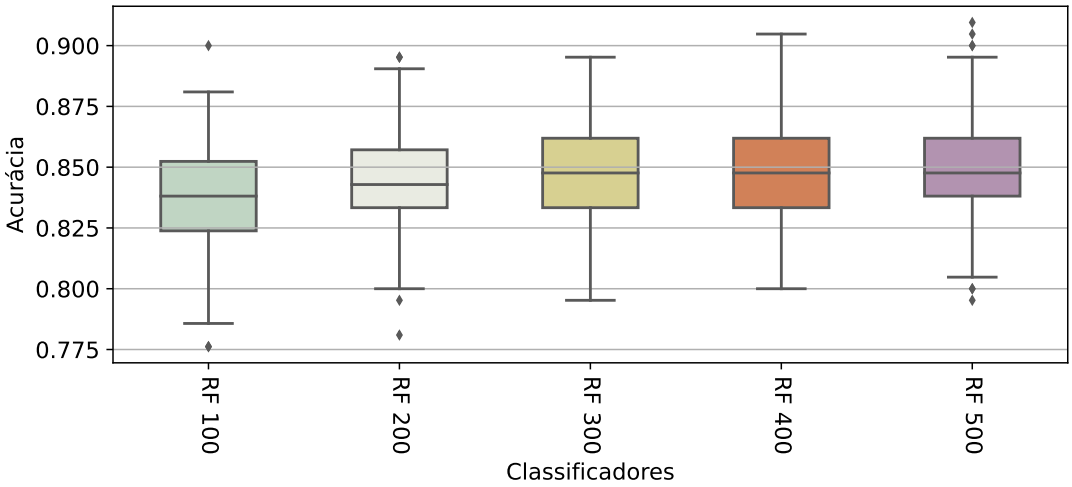
Para os experimentos considerando o cenário representado pela Figura 13b (tipos de hanseníase e seus respectivos conjuntos de doenças que fazem diagnóstico diferencial), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 43), especificidade (Figura 44), sensibilidade (Figura 45), área sob curva ROC (Figura 46) e índice kappa (Figura 47), além dos respectivos valores de média e desvio padrão (Tabela 14), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 14: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.8377	0.0205	0.9783	0.0114	0.9600	0.0385	0.9957	0.0047	0.8145	0.0235
RF 200 árvores	0.8449	0.0202	0.9802	0.0108	0.9711	0.0348	0.9970	0.0035	0.8228	0.0231
RF 300 árvores	0.8477	0.0198	0.9801	0.0110	0.9755	0.0328	0.9974	0.0032	0.8259	0.0226
RF 400 árvores	0.8487	0.0202	0.9811	0.0108	0.9770	0.0322	0.9976	0.0031	0.8271	0.0231
RF 500 árvores	0.8501	0.0203	0.9809	0.0105	0.9788	0.0307	0.9977	0.0030	0.8286	0.0232

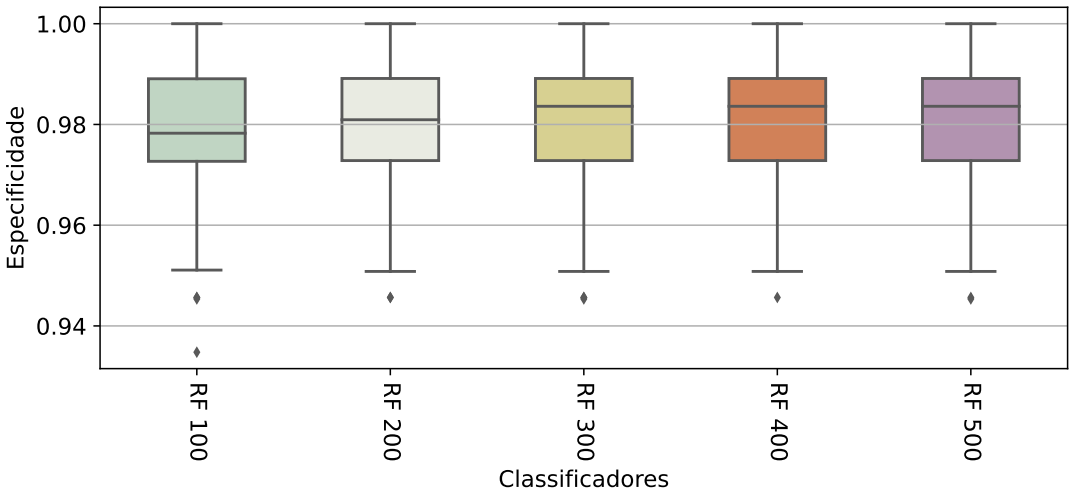
Fonte: Próprio Autor.

Figura 43: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



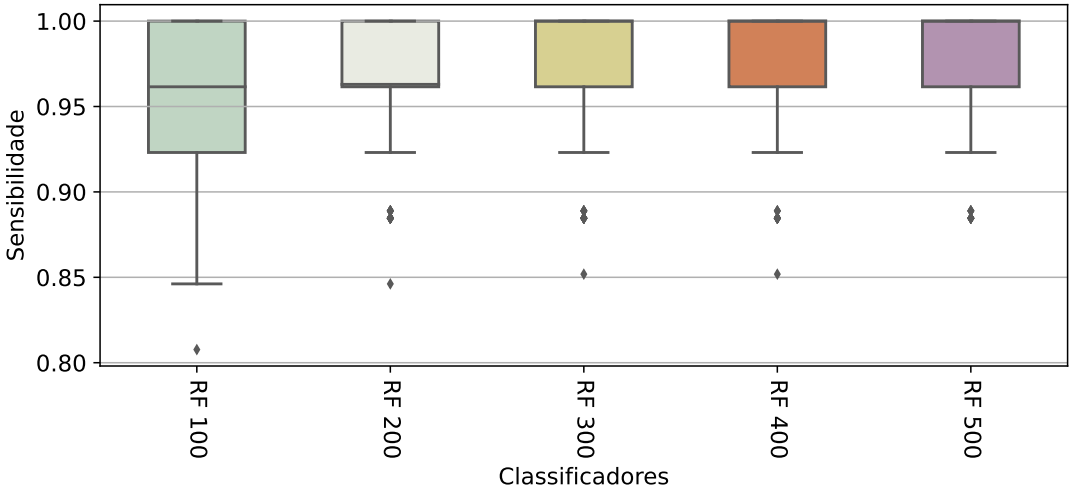
Fonte: Próprio Autor.

Figura 44: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



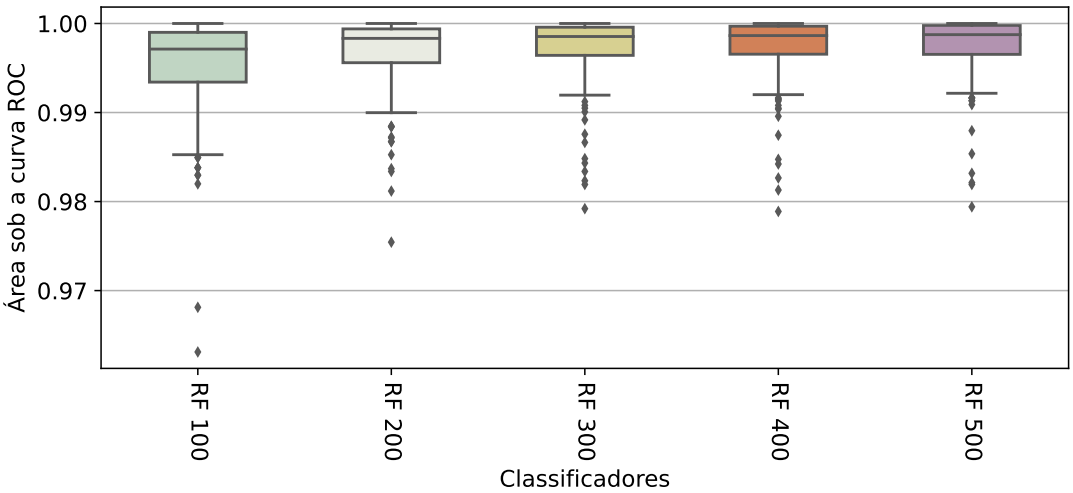
Fonte: Próprio Autor.

Figura 45: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



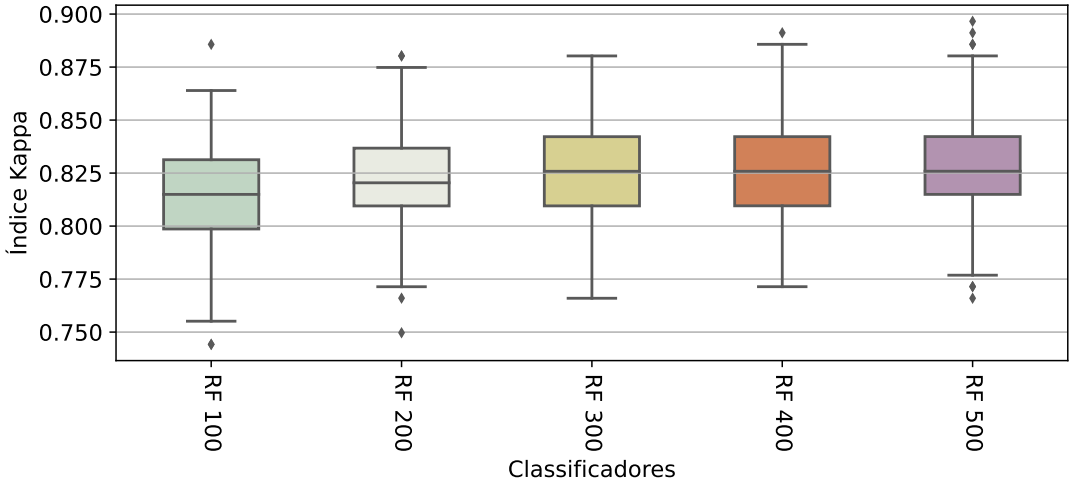
Fonte: Próprio Autor.

Figura 46: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 47: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13b após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

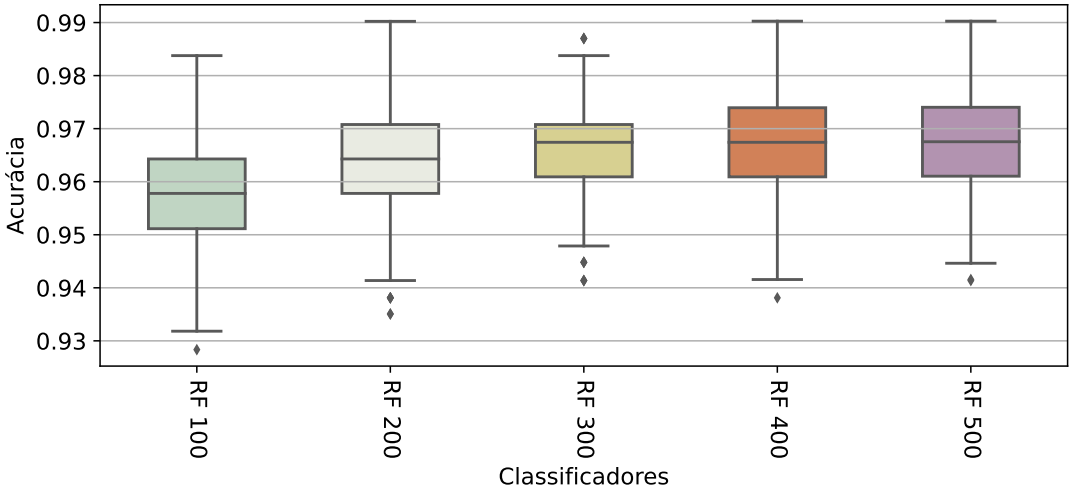
Para os experimentos considerando o cenário representado pela Figura 13c (tipos de hanseníase e todas as doenças que fazem diagnóstico diferencial são independentes), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 48), especificidade (Figura 49), sensibilidade (Figura 50), área sob curva ROC (Figura 51) e índice kappa (Figura 52), além dos respectivos valores de média e desvio padrão (Tabela 15), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 15: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9578	0.0107	0.9912	0.0059	0.9430	0.0497	0.9970	0.0040	0.9545	0.0116
RF 200 árvores	0.9642	0.0100	0.9930	0.0052	0.9541	0.0467	0.9982	0.0024	0.9614	0.0107
RF 300 árvores	0.9659	0.0093	0.9935	0.0051	0.9603	0.0435	0.9985	0.0022	0.9633	0.0100
RF 400 árvores	0.9668	0.0091	0.9937	0.0048	0.9639	0.0419	0.9987	0.0018	0.9643	0.0098
RF 500 árvores	0.9679	0.0092	0.9938	0.0047	0.9654	0.0407	0.9987	0.0017	0.9654	0.0099

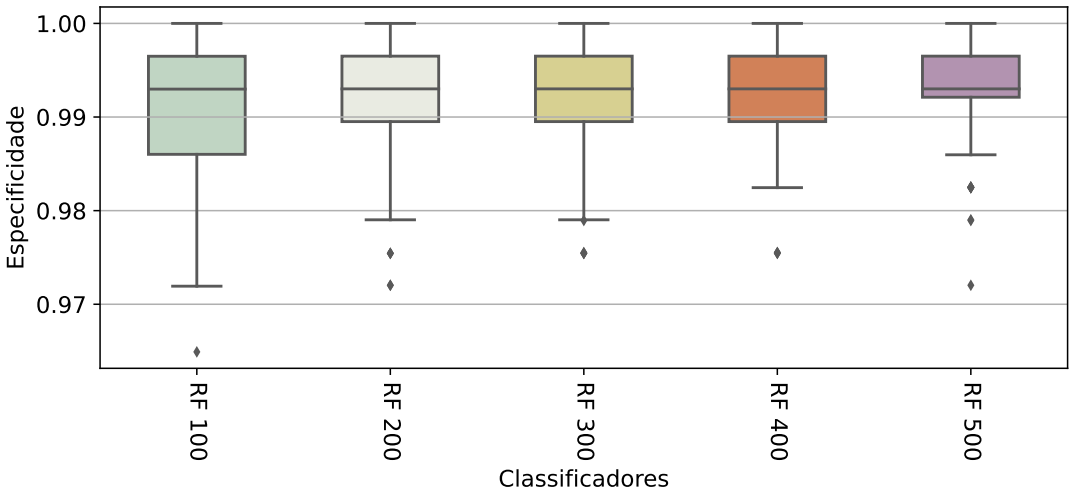
Fonte: Próprio Autor.

Figura 48: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



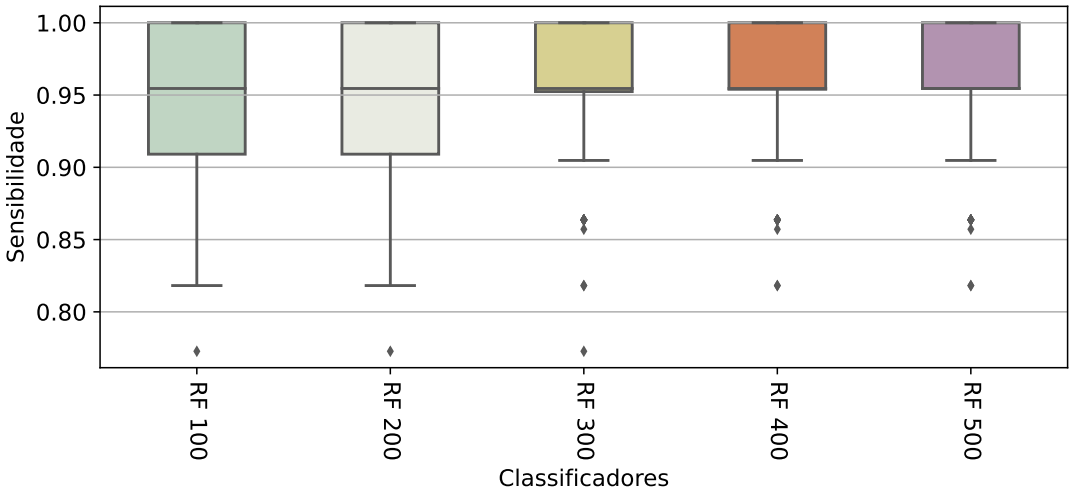
Fonte: Próprio Autor.

Figura 49: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



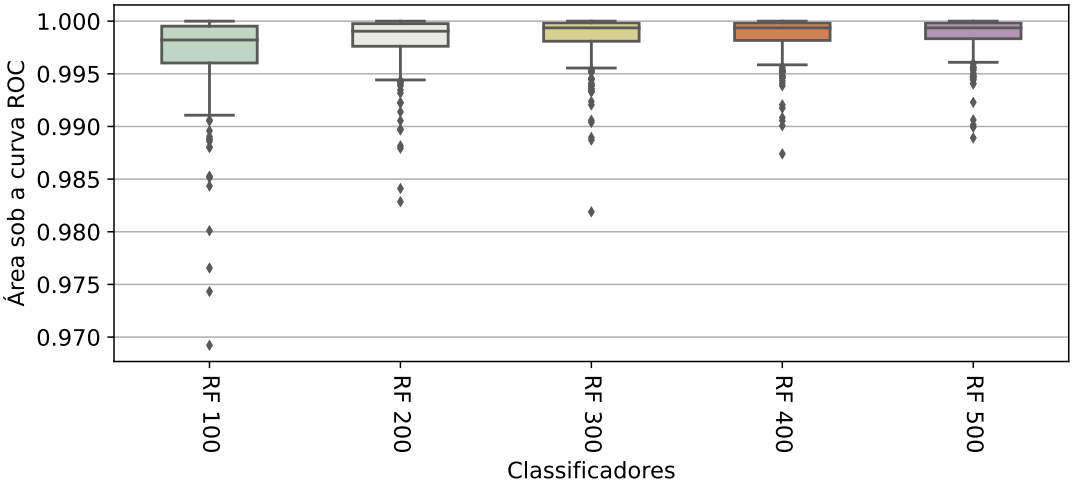
Fonte: Próprio Autor.

Figura 50: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



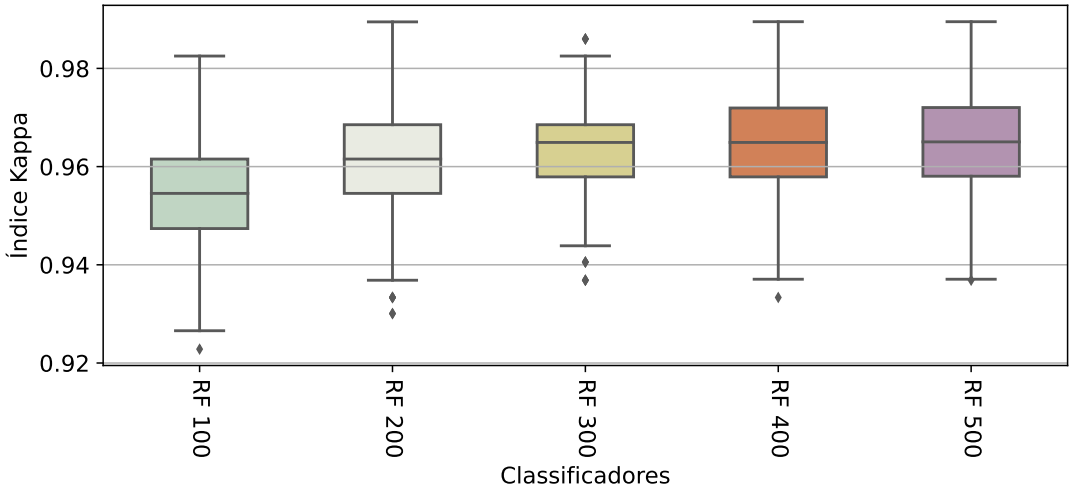
Fonte: Próprio Autor.

Figura 51: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 52: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13c após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

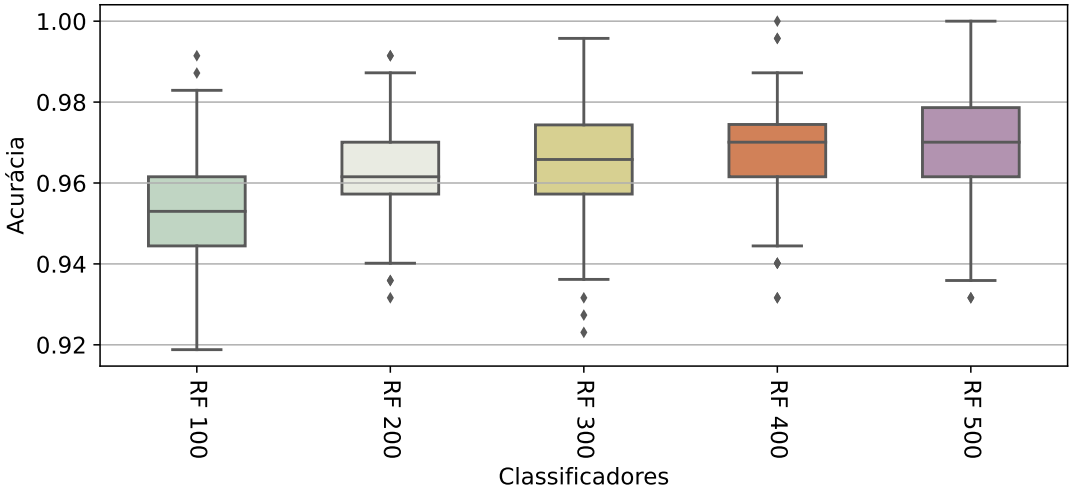
Para os experimentos considerando o cenário representado pela Figura 13d (cada tipo de hanseníase e um grupo com as doenças que fazem diagnóstico diferencial), são representados graficamente em boxplot os resultados obtidos para acurácia (Figura 53), especificidade (Figura 54), sensibilidade (Figura 55), área sob curva ROC (Figura 56) e índice kappa (Figura 57), além dos respectivos valores de média e desvio padrão (Tabela 16), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores após extração de atributos das imagens pela rede neural profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

Tabela 16: Média e desvio padrão da acurácia, especificidade, sensibilidade, área sob curva ROC e índice kappa do classificador Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.

CLASSIFICADOR	ACURÁCIA		ESPECIFICIDADE		SENSIBILIDADE		ÁREA SOB A CURVA ROC		ÍNDICE KAPPA	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	0.9530	0.0132	0.9918	0.0073	0.9897	0.0150	0.9997	0.0007	0.9412	0.0165
RF 200 árvores	0.9626	0.0116	0.9947	0.0055	0.9923	0.0129	0.9998	0.0004	0.9533	0.0145
RF 300 árvores	0.9655	0.0118	0.9956	0.0052	0.9938	0.0117	0.9999	0.0004	0.9569	0.0148
RF 400 árvores	0.9675	0.0117	0.9963	0.0047	0.9938	0.0114	0.9999	0.0004	0.9593	0.0146
RF 500 árvores	0.9686	0.0113	0.9963	0.0045	0.9942	0.0113	0.9999	0.0003	0.9608	0.0141

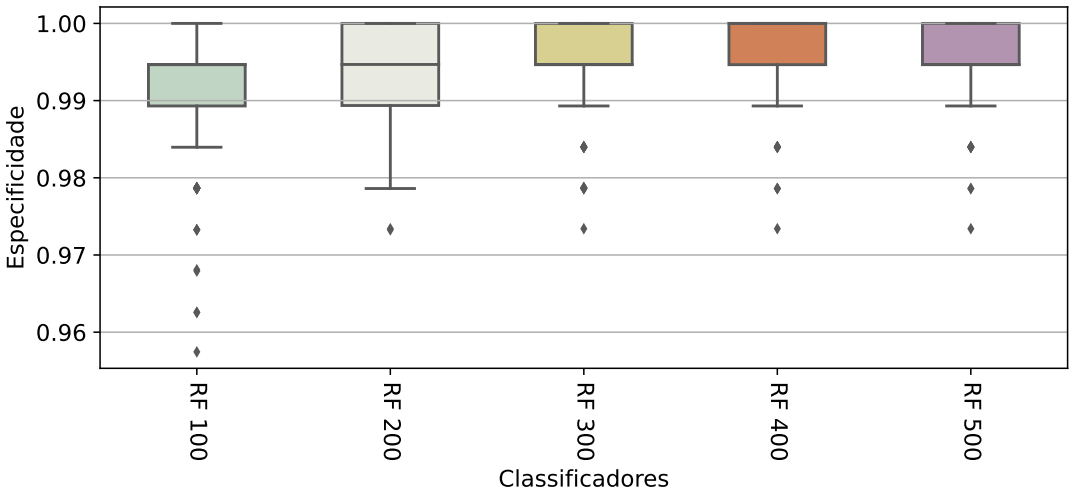
Fonte: Próprio Autor.

Figura 53: Acurácia da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



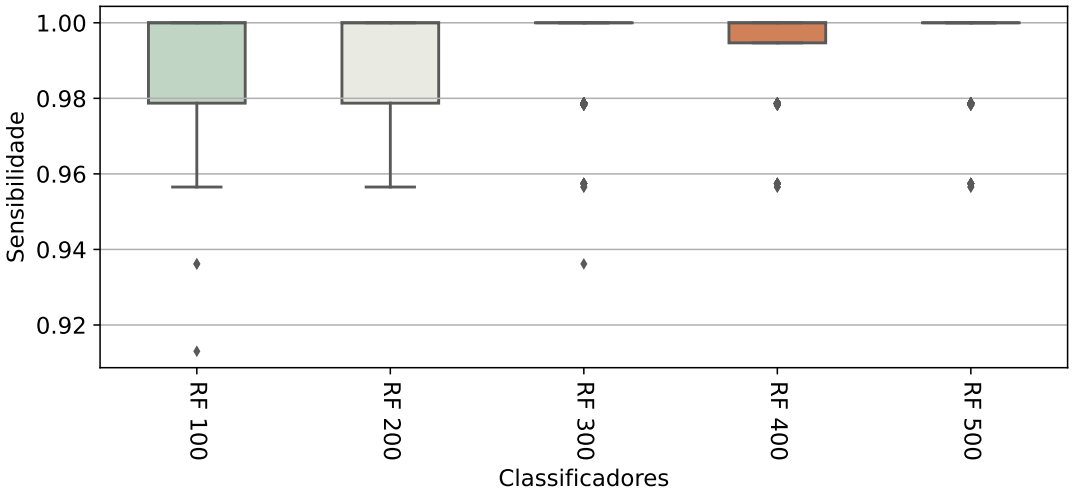
Fonte: Próprio Autor.

Figura 54: Especificidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



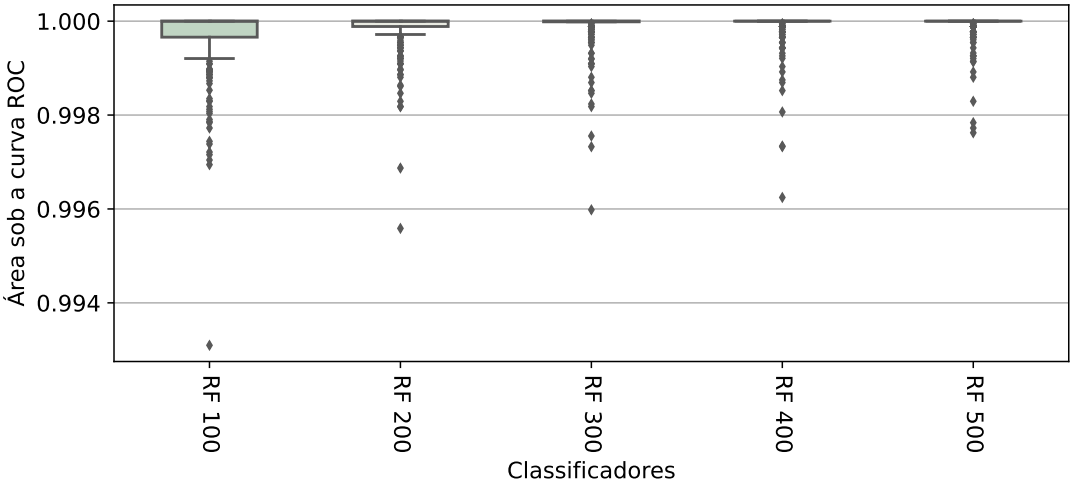
Fonte: Próprio Autor.

Figura 55: Sensibilidade da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



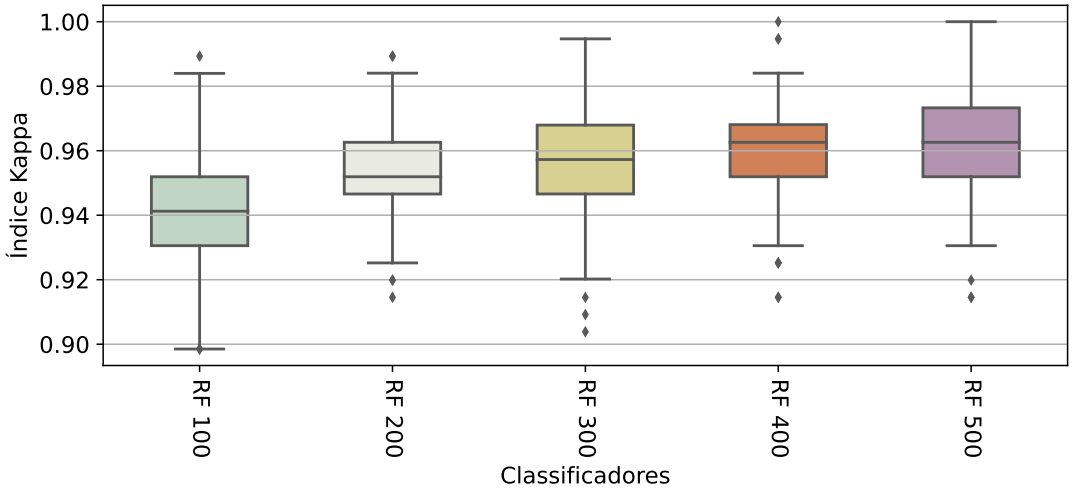
Fonte: Próprio Autor.

Figura 56: Área sob a curva ROC da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Figura 57: Índice kappa da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores para o cenário representado pela Figura 13d após extração de atributos pela rede profunda Inception v3 e balanceamento dos dados pelo SMOTE com $k = 3$ vizinhos.



Fonte: Próprio Autor.

Após execução dos experimentos para diferentes cenários, foi possível observar que em todos eles a diferença de desempenho para uma configuração de 100 árvores e de 500 árvores nunca ultrapassou 3% e só ultrapassou 2% duas vezes: sensibilidade referente ao cenário da Figura 12 (Tabela 12) e ao cenário da Figura 13c (Tabela 15). Esta análise já indica que escolher um classificador Random Forest com 100 árvores seria suficiente para o objetivo do trabalho. Porém, será realizada uma análise complementar: verificar o tempo de treinamento e teste durante a validação cruzada para cada configuração do classificador.

5.3 TEMPO DE TREINAMENTO E TESTE DURANTE A VALIDAÇÃO CRUZADA

Para os experimentos considerando o cenário representado pela Figura 11a (tipos de hanseníase e câncer de pele agrupados em 2 classes distintas), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 58), além dos respectivos valores de média e desvio padrão (Tabela 17), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

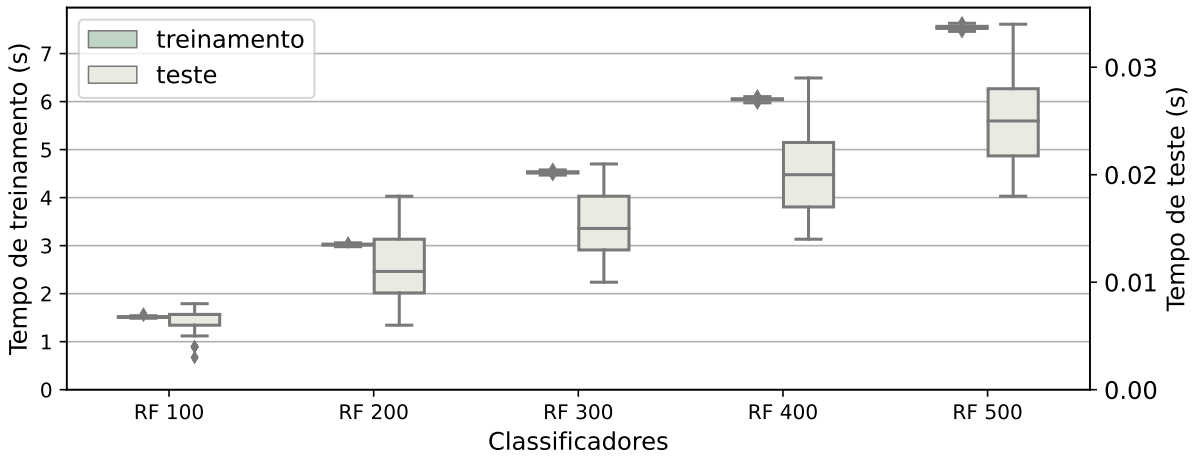
Para os experimentos considerando o cenário representado pela Figura 11b (tipos de

Tabela 17: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11a.

CLASSIFICADOR	TEMPO DE TREINAMENTO (s)		TEMPO DE TESTE (s)	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	1.5135	0.0124	0.0063	0.0012
RF 200 árvores	3.0204	0.0158	0.0111	0.0030
RF 300 árvores	4.5243	0.0212	0.0154	0.0030
RF 400 árvores	6.0408	0.0274	0.0197	0.0033
RF 500 árvores	7.5465	0.0337	0.0247	0.0037

Fonte: Próprio Autor.

Figura 58: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11a.



Fonte: Próprio Autor.

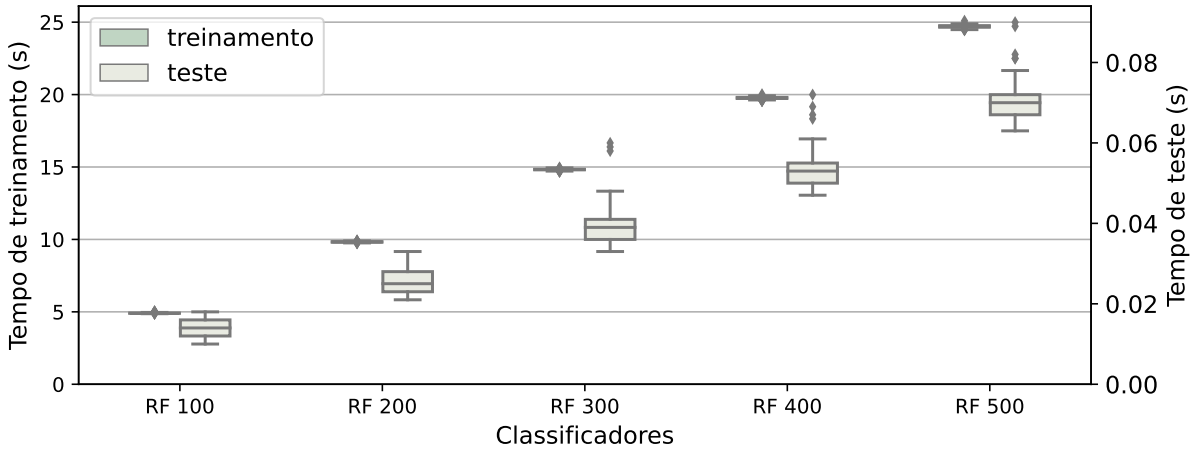
hanseníase e câncer de pele considerados independentes), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 59), além dos respectivos valores de média e desvio padrão (Tabela 18), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

Tabela 18: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11b.

CLASSIFICADOR	TEMPO DE TREINAMENTO (s)		TEMPO DE TESTE (s)	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	4.9176	0.0216	0.0138	0.0021
RF 200 árvores	9.8302	0.0298	0.0255	0.0026
RF 300 árvores	14.8310	0.0475	0.0387	0.0037
RF 400 árvores	19.7679	0.0624	0.0528	0.0036
RF 500 árvores	24.7031	0.0880	0.0699	0.0038

Fonte: Próprio Autor.

Figura 59: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 11b.



Fonte: Próprio Autor.

Para os experimentos considerando o cenário representado pela Figura 12 (tipos de câncer de pele considerados independentes), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 60), além dos respectivos valores de média e desvio padrão (Tabela 19), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

Tabela 19: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 12.

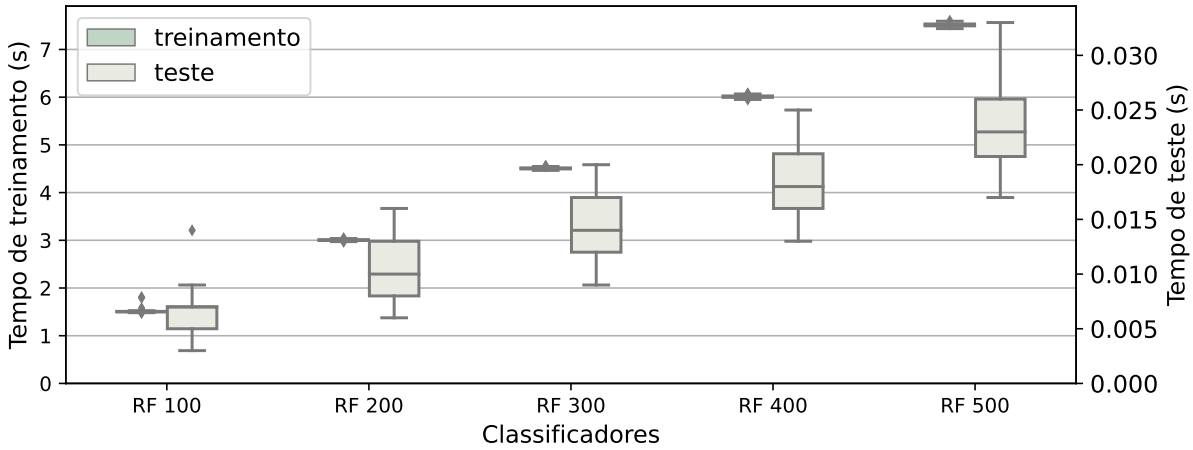
CLASSIFICADOR	**TEMPO DE TREINAMENTO (s)**		**TEMPO DE TESTE (s)**	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	1.5072	0.0258	0.0060	0.0014
RF 200 árvores	3.0046	0.0131	0.0105	0.0025
RF 300 árvores	4.5061	0.0178	0.0144	0.0027
RF 400 árvores	6.0108	0.0227	0.0185	0.0029
RF 500 árvores	7.5141	0.0278	0.0233	0.0033

Fonte: Próprio Autor.

Para os experimentos considerando o cenário representado pela Figura 13a (tipos de hanseníase considerados independentes), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 61), além dos respectivos valores de média e desvio padrão (Tabela 20), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

Para os experimentos considerando o cenário representado pela Figura 13b (tipos

Figura 60: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 12.



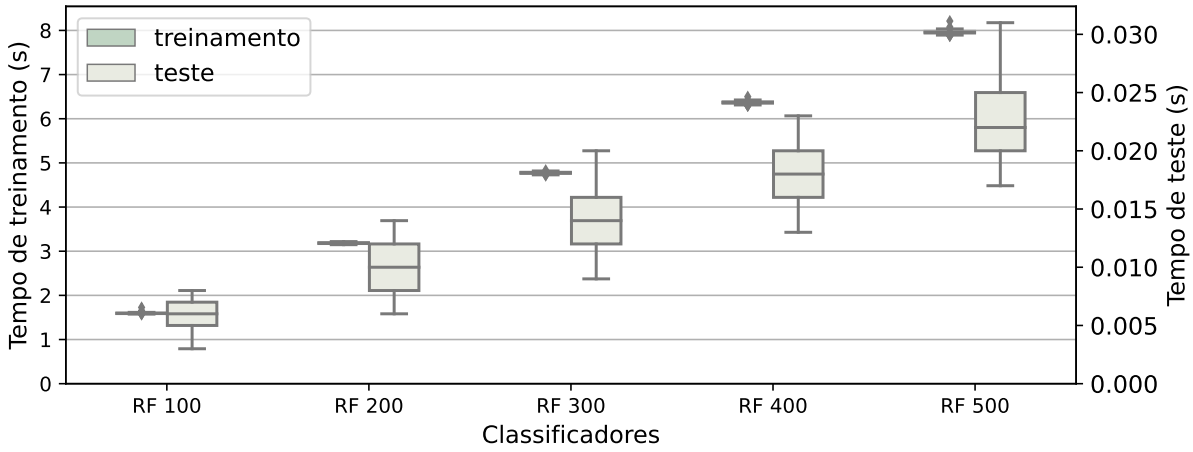
Fonte: Próprio Autor.

Tabela 20: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13a.

CLASSIFICADOR	**TEMPO DE TREINAMENTO (s)**		**TEMPO DE TESTE (s)**	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	1.5953	0.0132	0.0058	0.0014
RF 200 árvores	3.1831	0.0132	0.0100	0.0022
RF 300 árvores	4.7750	0.0183	0.0137	0.0024
RF 400 árvores	6.3669	0.0245	0.0180	0.0026
RF 500 árvores	7.9614	0.0325	0.0226	0.0029

Fonte: Próprio Autor.

Figura 61: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13a.



Fonte: Próprio Autor.

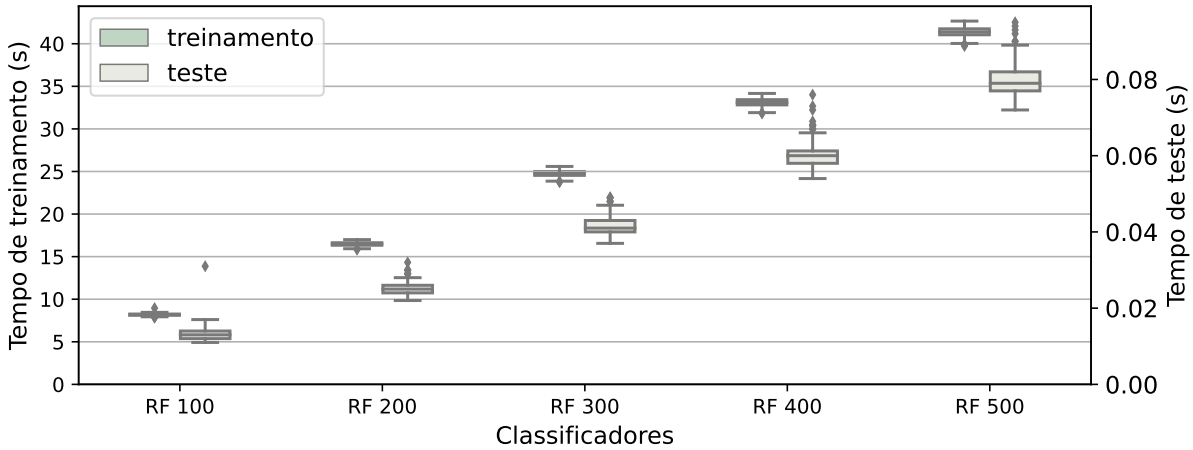
de hanseníase e seus respectivos conjuntos de doenças que fazem diagnóstico diferencial), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 62), além dos respectivos valores de média e desvio padrão (Tabela 21), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

Tabela 21: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13b.

CLASSIFICADOR	**TEMPO DE TREINAMENTO (s)**		**TEMPO DE TESTE (s)**	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	8.1879	0.1166	0.0129	0.0015
RF 200 árvores	16.4775	0.2148	0.0252	0.0015
RF 300 árvores	24.7636	0.3171	0.0416	0.0023
RF 400 árvores	33.1070	0.4136	0.0599	0.0032
RF 500 árvores	41.3816	0.5059	0.0796	0.0040

Fonte: Próprio Autor.

Figura 62: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13b.



Fonte: Próprio Autor.

Para os experimentos considerando o cenário representado pela Figura 13c (tipos de hanseníase e todas as doenças que fazem diagnóstico diferencial são independentes), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 63), além dos respectivos valores de média e desvio padrão (Tabela 22), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

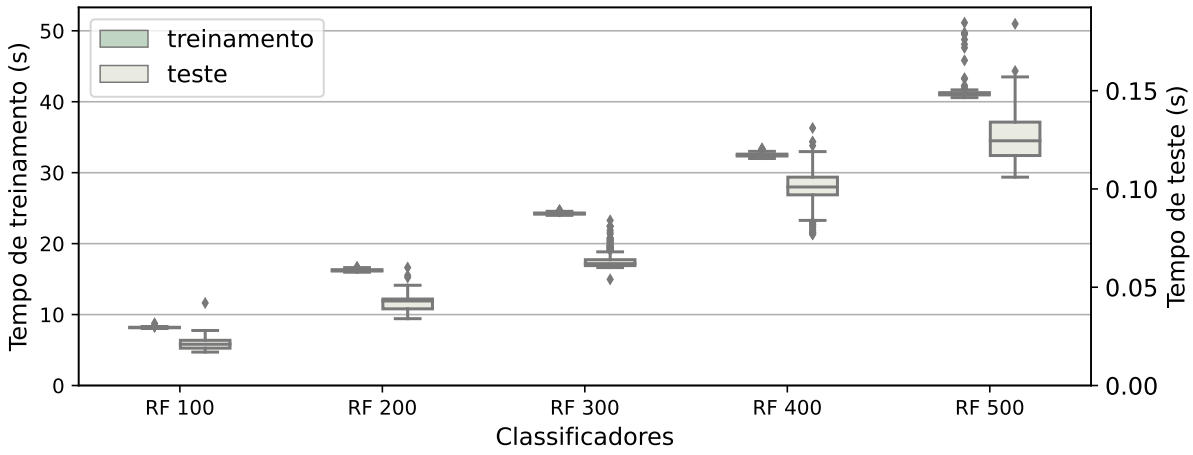
Para os experimentos considerando o cenário representado pela Figura 13d (cada

Tabela 22: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13c.

CLASSIFICADOR	TEMPO DE TREINAMENTO (s)		TEMPO DE TESTE (s)	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	8.1808	0.0680	0.0210	0.0026
RF 200 árvores	16.2407	0.1376	0.0419	0.0035
RF 300 árvores	24.2507	0.1289	0.0636	0.0041
RF 400 árvores	32.5053	0.2440	0.1007	0.0091
RF 500 árvores	41.3438	1.2921	0.1261	0.0123

Fonte: Próprio Autor.

Figura 63: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13c.



Fonte: Próprio Autor.

tipo de hanseníase e um grupo com as doenças que fazem diagnóstico diferencial), são representados graficamente em boxplot os tempos de treinamento e teste (Figura 64), além dos respectivos valores de média e desvio padrão (Tabela 23), da classificação das imagens com Random Forest de 100, 200, 300, 400 e 500 árvores.

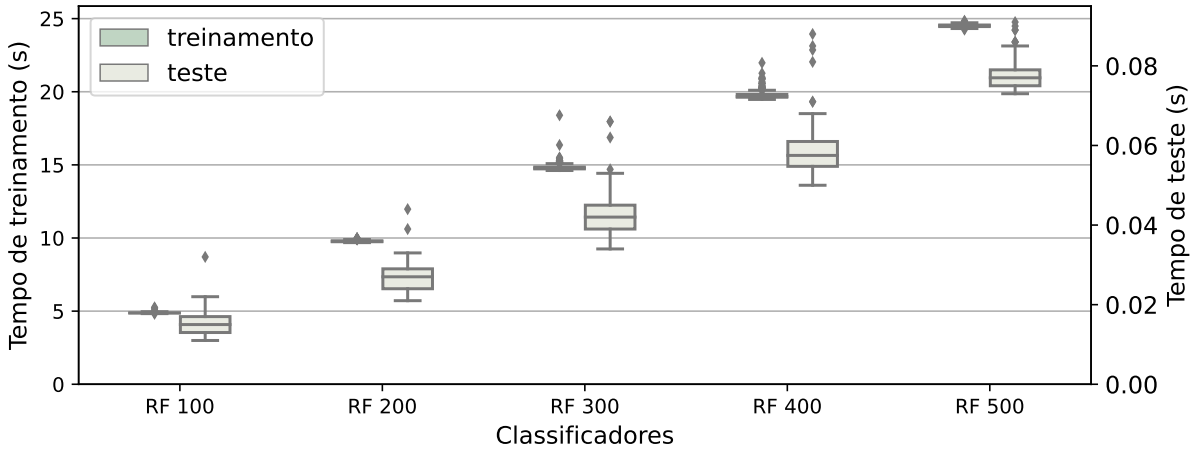
Tabela 23: Média e desvio padrão do tempo de treinamento e do tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13d.

CLASSIFICADOR	TEMPO DE TREINAMENTO (s)		TEMPO DE TESTE (s)	
	Média	Desvio Padrão	Média	Desvio Padrão
RF 100 árvores	4.8947	0.0426	0.0148	0.0026
RF 200 árvores	9.7832	0.0543	0.0271	0.0031
RF 300 árvores	14.8376	0.2702	0.0419	0.0044
RF 400 árvores	19.7879	0.3061	0.0580	0.0052
RF 500 árvores	24.5174	0.0863	0.0776	0.0032

Fonte: Próprio Autor.

É possível observar que treinar um classificador Random Forest com 500 árvores

Figura 64: Tempo de treinamento e tempo de teste, ambos em segundos, durante a validação cruzada de 10 dobras considerando o classificador Random Forest de 100, 200, 300, 400 e 500 árvores no cenário representado pela Figura 13d.



Fonte: Próprio Autor.

levou cerca de 5x mais tempo do que com a configuração de 100 árvores (considerando o processo de validação cruzada de 10 dobras). Já na etapa de teste, esse número varia de 4x a 6x. Por isso, além da análise das outras métricas na seção anterior, define-se neste trabalho que a utilização do classificador Random Forest com 100 árvores é a melhor opção considerando um *trade off* entre performance e custo computacional.

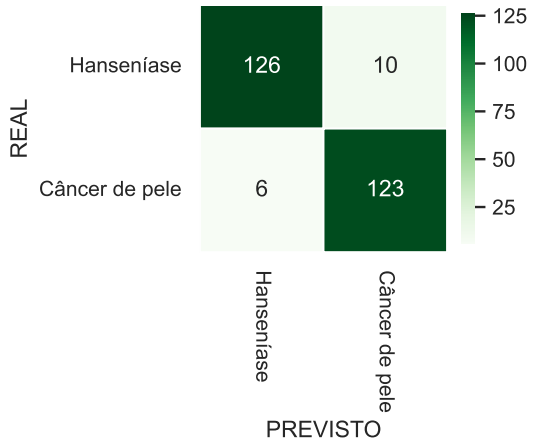
5.4 TESTE DO MELHOR MODELO

Enfim, foi possível chegar em uma combinação de rede neural profunda e de classificador adequada para um sistema de apoio a detecção e classificação das doenças consideradas neste trabalho. O melhor modelo tem o seguinte arranjo:

- **Rede neural profunda:** Inception v3;
- **Classificador:** Random Forest de 100 árvores.

A última rodada de experimentos tem o objetivo de analisar o comportamento do modelo escolhido na classificação de imagens até então não utilizadas. A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 11a (tipos de hanseníase e câncer de pele agrupados em 2 classes distintas) pode ser vista na Figura 65.

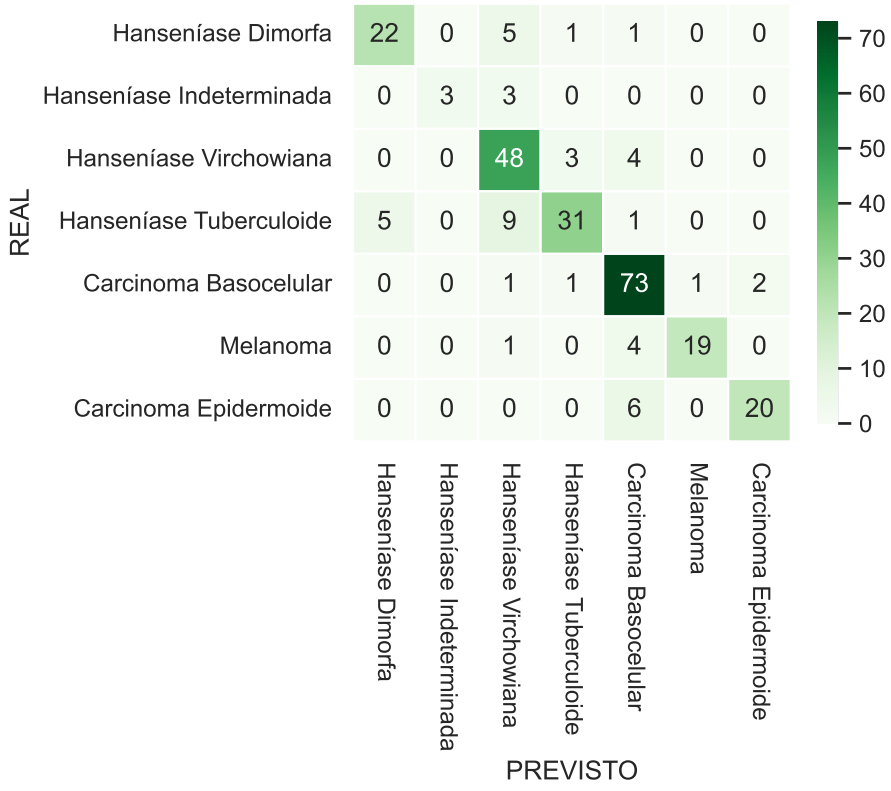
Figura 65: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 11a.



Fonte: Próprio Autor.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 11b (tipos de hanseníase e câncer de pele considerados independentes) pode ser vista na Figura 66.

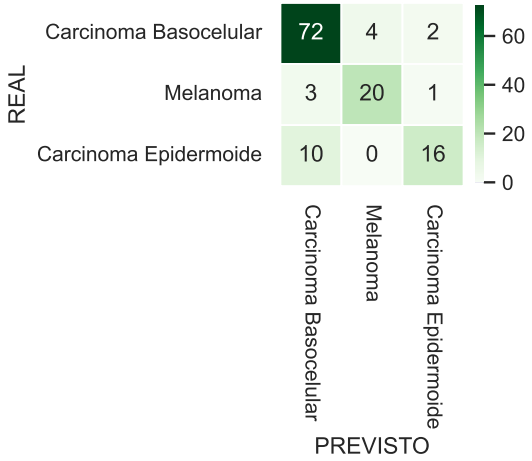
Figura 66: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 11b.



Fonte: Próprio Autor.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 12 (tipos de câncer de pele considerados independentes) pode ser vista na Figura 67.

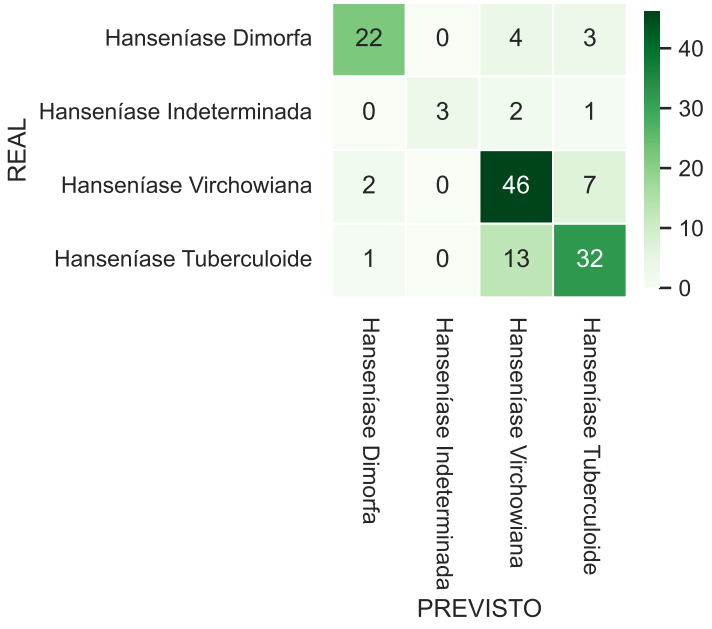
Figura 67: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 12.



Fonte: Próprio Autor.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 13a (tipos de hanseníase considerados independentes) pode ser vista na Figura 68.

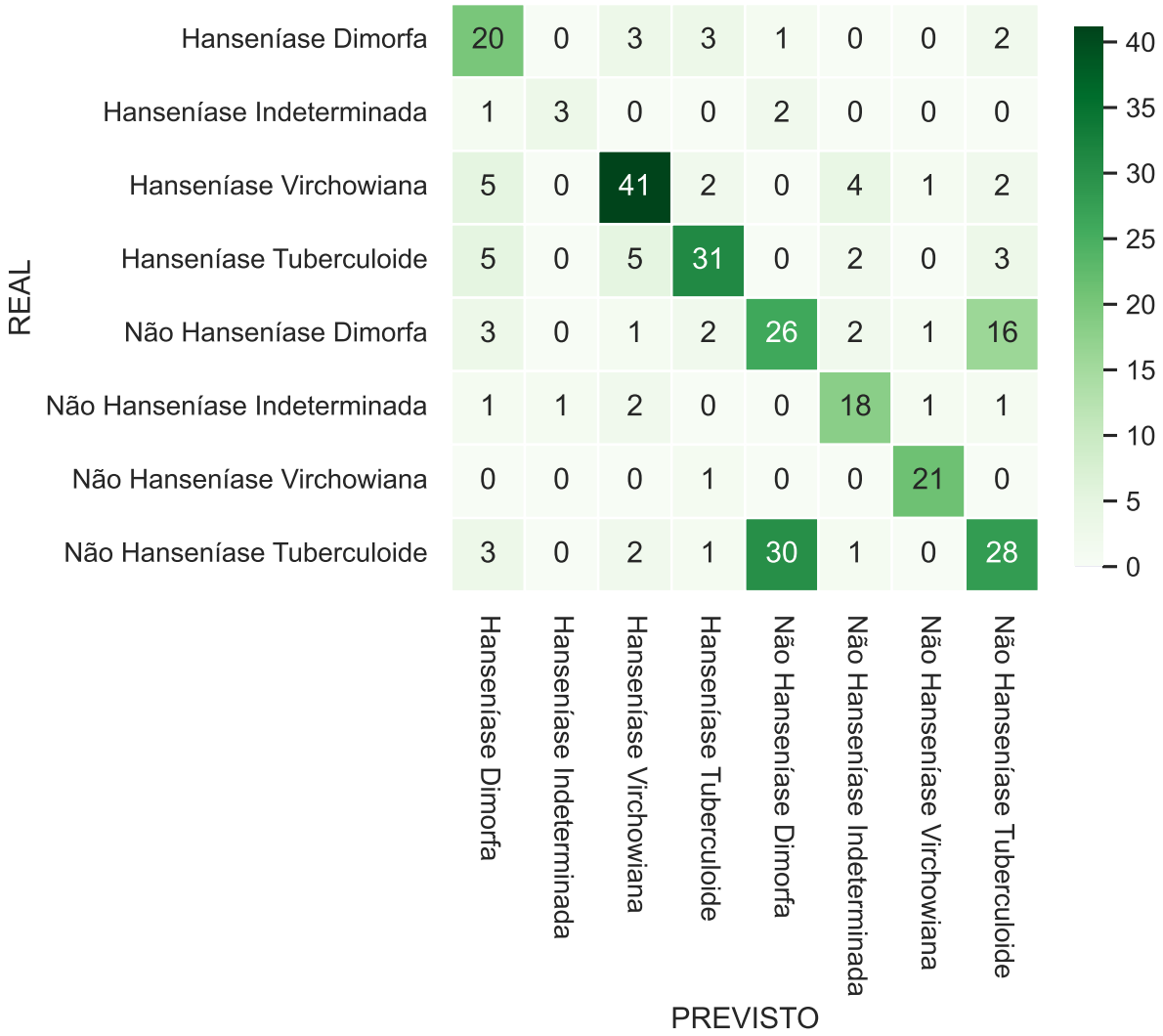
Figura 68: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13a.



Fonte: Próprio Autor.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 13b (tipos de hanseníase e seus respectivos conjuntos de doenças que fazem diagnóstico diferencial) pode ser vista na Figura 69.

Figura 69: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13b.

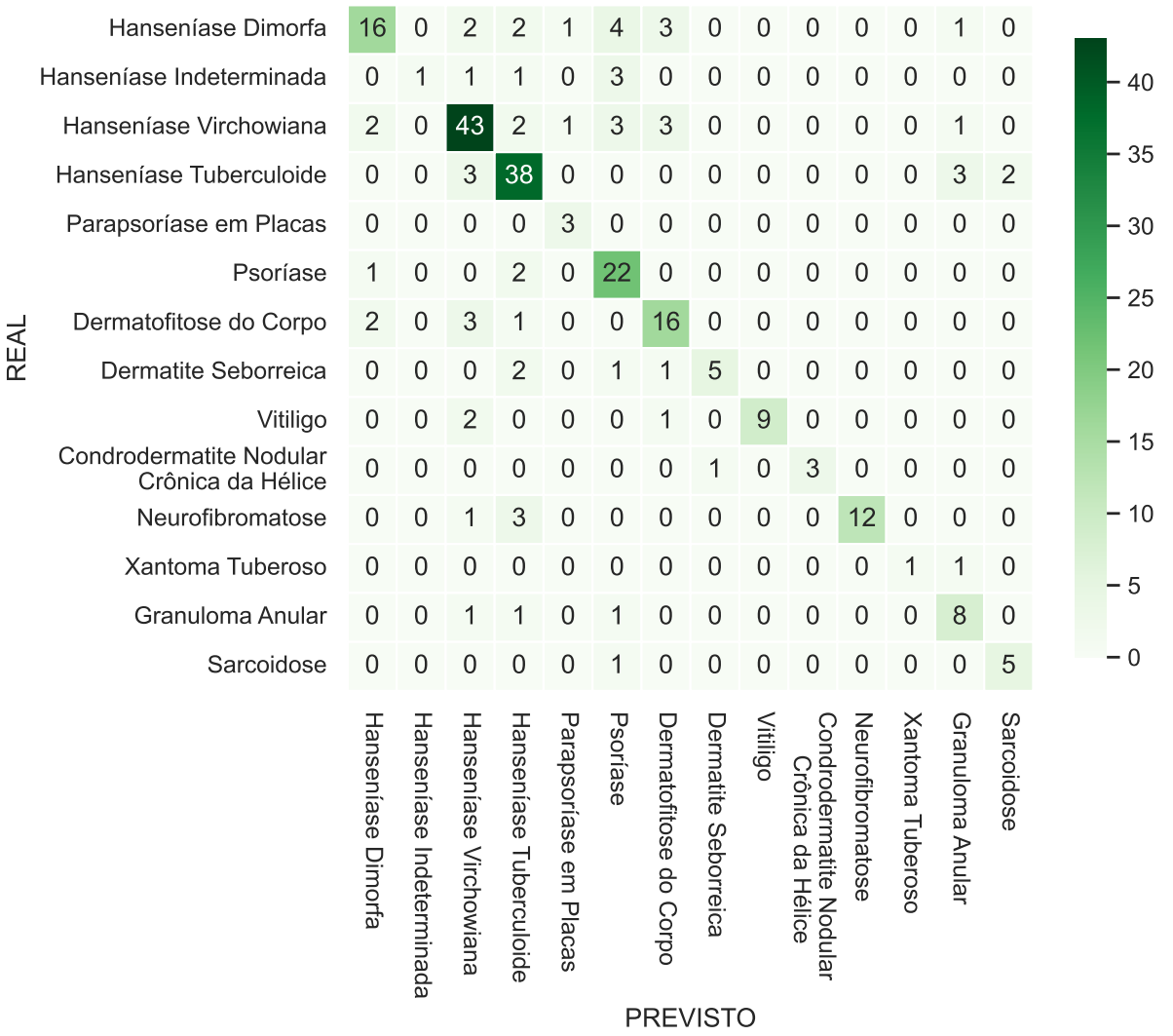


Fonte: Próprio Autor.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 13c (tipos de hanseníase e todas as doenças que fazem diagnóstico diferencial são independentes) pode ser vista na Figura 70.

A matriz de confusão referente ao teste do melhor modelo no cenário representado pela Figura 13d (cada tipo de hanseníase e um grupo com as doenças que fazem diagnóstico diferencial) pode ser vista na Figura 71.

Figura 70: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13c.

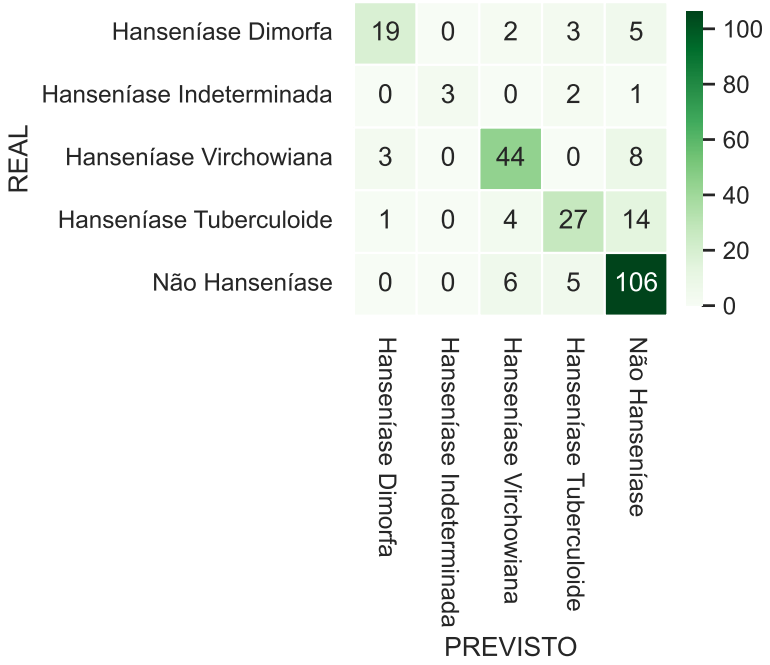


Fonte: Próprio Autor.

5.5 CONSIDERAÇÕES FINAIS

Sobre os experimentos realizados na Seção 5.1, a arquitetura de rede profunda escolhida é uma das duas que extrai um maior número de atributos da imagem (Inception v3 - 2048 atributos extraídos). Porém, ao verificar a performance das outras redes profundas nota-se que essa correlação não existe, um exemplo disso é que a arquitetura em que o desempenho mais se aproximou daquela escolhida (NASNetMobile - 1056 atributos extraídos), em geral teve um desempenho melhor que a ResNet50 (2048 atributos extraídos). Não foi possível encontrar um único fator determinante na escolha da arquitetura de rede (e.g. número de parâmetros, número de atributos extraídos), por isso o intuito foi

Figura 71: Matriz de confusão resultante do teste com o melhor modelo escolhido no cenário representado pela Figura 13d.



Fonte: Próprio Autor.

avaliar diferentes técnicas desenvolvidas em diferentes épocas para entender a que melhor se encaixa na problemática desta pesquisa.

Já a escolha da configuração do classificador se deu tanto pela análise de desempenho das classificações quanto pelo custo operacional de sua utilização. Apesar do modelo só precisar ser construído uma vez, o que permitiria considerar modelos custosos computacionalmente, também é importante entender que mais memória e mais processamento continuarão sendo necessários para realizar as predições. E isso está ligado não só ao custo de operação em si, como ao custo financeiro de dispor de uma infraestrutura adequada. Como a diferença entre as configurações de menor e de maior número de árvores não ultrapassou 3% em nenhuma das métricas e escolher a de maior número implicaria em cerca de 5x mais custo computacional, aquela configuração de menor número de árvores foi a escolhida.

Os testes após a escolha do melhor modelo de classificação trazem resultados interessantes. Naqueles cenários que envolvem tanto a hanseníase quanto o câncer de pele (Figura 65 e Figura 66), é possível observar que os erros de classificação acontecem geralmente entre os tipos da mesma doença. Como esperado, já que estas se apresentam na pele de forma distinta.

Para os cenários que incluem os tipos de câncer de pele (Figura 66 e Figura 67) aconteceram poucos erros de classificação para o tipo melanoma, como esperado, devido ao tipo da célula afetada (melanócitos para o melanoma e queratinócitos para o carcinoma basocelular e carcinoma epidermoide). Porém, erros de classificação ocorreram sobretudo para o carcinoma epidermoide, sempre confundido com o carcinoma basocelular. Uma possível explicação pode ser dada considerando que (i) estes dois tipos, enquadrados como não-melanoma, se apresentam na pele das mais diversas formas, por vezes até semelhantes a depender do estágio da doença, e (ii) por ser mais comum, a quantidade de imagens disponíveis para o carcinoma basocelular é quase 3 vezes maior do que para o carcinoma epidermoide. Ou seja, o modelo pode não ter aprendido com tanta eficiência os padrões relacionados ao carcinoma epidermoide.

A classificação entre os tipos da hanseníase (Figura 68) mostra uma boa performance para a classificação da hanseníase dimorfa, mas não tão boa para a hanseníase indeterminada. Mais uma vez, um possível problema de generalização do modelo de classificação devido à quantidade de imagem disponível para a doença (cerca de 4,5% das imagens referentes à hanseníase são para o tipo indeterminada). Além das classificações apresentadas na Seção 3.1, os tipos tuberculoide e virchowiana também são considerados como polos da hanseníase, ou seja, são dois tipos estáveis e opostos. Logo, os erros de classificação observados entre esses dois tipos não eram esperados. A base de imagens foi descartada como possível causa pois esses erros não se repetiram na mesma proporção para os outros cenários.

Quanto à Figura 69, onde são considerados os tipos de hanseníase e seus respectivos conjuntos de doenças que fazem diagnóstico diferencial, o que mais chama atenção são os erros de classificação entre as classes *Não Hanseníase Dimorfa* e *Não Hanseníase Tuberculoide*. A causa desse problema fica clara quando observadas as doenças que fazem parte dessas classes (apresentadas na Seção 3.1), visto que tanto a psoríase quanto a dermatofitose do corpo estão inclusas em ambas. Esse foi um dos cenários considerados dado o ponto de vista médico, já que não é eficaz no aspecto computacional.

Para o cenário em que todos os tipos de hanseníase e todas as doenças que fazem diagnóstico diferencial são independentes (Figura 70), novamente a questão da quantidade de imagens disponíveis é mencionada. Inclusive, algumas doenças precisaram ser desconsideradas ao longo dos experimentos por conta da dificuldade em dividir as amostras em

treinamento e teste, foram elas: esclerodermia linear, nevo acrômico, tuberculose cutânea verrucosa, nevo anêmico e pitiríase alba. Apesar desse problema conhecido, foi observado um bom desempenho na classificação de algumas doenças com poucas imagens disponíveis, destacam-se: parapsoríase em placas, condrodermatite nodular crônica da hélice e sarcoidose. Uma possível razão é que essas imagens tenham características muito específicas, além da forma de apresentação na pele, e que acabam facilitando a classificação. Como por exemplo: as poucas imagens disponíveis são de uma mesma lesão na mesma região do corpo de uma mesma pessoa.

Por fim, no cenário que compara cada tipo de hanseníase e um grupo contendo todas as doenças que fazem diagnóstico diferencial (Figura 71), o problema de generalização para a hanseníase indeterminada permanece, o desempenho referente à hanseníase tuberculoide diminui se comparado aos outros cenários e o modelo consegue distinguir bem as doenças pertencentes à classe *Não Hanseníase*.

6 CONCLUSÃO

O propósito deste trabalho se resumiu em modelar um sistema para o apoio ao diagnóstico por imagem e não invasivo de câncer de pele e hanseníase, sendo este capaz de auxiliar na detecção precoce e, assim, contribuir para a redução da mortalidade e das sequelas envolvidas. O trabalho conta não só com o estado da arte em inteligência artificial, como também foi desenhado considerando a realidade do Brasil (sobretudo do Nordeste) e leva em conta o diagnóstico diferencial da hanseníase, tudo isso de modo a chegar em um resultado efetivo para a problemática tratada.

A ideia de definir diferentes cenários de experimentos se deu com o objetivo de avaliar tanto os aspectos computacionais quanto as necessidades médicas para o diagnóstico das doenças envolvidas. Porém, sendo um próximo passo construir um sistema web ou aplicativo móvel onde de fato o modelo definido nesta pesquisa será implementado e ficará disponível como uma ferramenta de apoio ao diagnóstico, se faz necessário definir como as doenças serão agrupadas (um dos cenários apresentados ou uma combinação deles), de modo que as necessidades médicas e computacionais sejam atendidas.

A principal dificuldade encontrada esteve relacionada com a criação da base de imagens necessária para o desenvolvimento do trabalho. De início não foi simples encontrar uma base pública que satisfizesse o objetivo da pesquisa e que tivesse imagens de qualidade adequada. Além disso, ficou claro que representar uma doença com poucas imagens tem um impacto negativo no desempenho do modelo construído. Geralmente isso ocorre com doenças de baixa prevalência, então a solução foi contornar o problema utilizando técnicas adequadas, como o balanceamento de dados.

Por fim, como trabalho futuro, além da já citada construção de um sistema web ou aplicativo móvel, sugere-se que mais imagens sejam adicionadas na base atual, se possível que estas abranjam outras realidades existentes. Além disso, outras arquiteturas de rede profunda fazem parte do estado da arte, então avaliá-las pode trazer resultados ainda mais significativos. E sugere-se também a realização de experimentos com outros tipos de classificadores.

REFERÊNCIAS

AMERICAN CANCER SOCIETY. **Cancer A-Z. Skin cancer**. Atlanta: ACS, 2019. Disponível em: https://www.cancer.org/cancer/skin-cancer.html. Acesso em: 10/01/2022.

ASTORINO, Annabella et al. Melanoma detection by means of multiple instance learning. **Interdisciplinary Sciences: Computational Life Sciences**, Springer, v. 12, n. 1, p. 24–31, 2020.

BAWEJA, Harjatin Singh; PARHAR, Tanvir. Leprosy lesion recognition using convolutional neural networks. **2016 International Conference on Machine Learning and Cybernetics (ICMLC)**, v. 1, p. 141–145, 2016.

BRASIL. **Boletim Epidemiológico de Hanseníase**. Brasília-DF: Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, 2021. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2021/boletim-hanseniase-_-25-01.pdf.

_____._____. Brasília-DF: Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, 2022. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2022/boletim-epidemiologico-de-hanseniase-_-25-01-2022.pdf.

_____. **Guia prático sobre a hanseníase**. Brasília-DF: Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis, 2017. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/guia_pratico_hanseniase.pdf. ISBN 978-85-334-2542-2.

BRAY, Freddie et al. The ever-increasing importance of cancer as a leading cause of premature death worldwide. **Cancer**, Wiley Online Library, 2021.

CHAPMAN, Nick et al. 2019 G-FINDER Report. Neglected Disease Research and Development: Uneven Progress. Policy Cures Research, 2019.

CHAWLA, Nitesh et al. SMOTE: Synthetic Minority Over-sampling Technique. **J. Artif. Intell. Res. (JAIR)**, v. 16, p. 321–357, jun. 2002.

CHEN, Kou-Huang et al. Leprosy: A review of epidemiology, clinical diagnosis, and management. **Journal of Tropical Medicine**, Hindawi, v. 2022, 2022.

DE GOMA, JOEL CASUAYAN; DEVARAJ, MADHAVI. Recognizing Common Skin Diseases in the Philippines Using Image Processing and Machine Learning Classification. **2020 the 3rd International Conference on Computing and Big Data**, p. 68–72, 2020.

ESTEVA, Andre et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, Nature Publishing Group, v. 542, n. 7639, p. 115–118, 2017.

FOGEL, Alexander L; KVEDAR, Joseph C. Artificial intelligence powers digital medicine. **NPJ Digital Medicine**, Nature Publishing Group, v. 1, n. 1, p. 1–4, 2018.

FRANK, Eibe; HALL, Mark A.; WITTEN, Ian H. **The WEKA Workbench**. 2016. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.

HASAN, Mahamudul et al. Skin cancer detection using convolutional neural network. **Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence**, p. 254–258, 2019.

HE, Jianxing et al. The practical implementation of artificial intelligence technologies in medicine. **Nature Medicine**, Nature Publishing Group, v. 25, n. 1, p. 30–36, 2019.

HE, Kaiming et al. Deep residual learning for image recognition. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 770–778, 2016.

IANDOLA, Forrest N et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. **arXiv preprint arXiv:1602.07360**, 2016.

INSTITUTO NACIONAL DO CÂNCER. **ABC do Câncer: Abordagens Básicas para o Controle do Câncer**. Rio de Janeiro: Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva, 2011. ISBN 978-85-7318-188-3.

_____. **Estimativa 2020 : incidência de câncer no Brasil**. Rio de Janeiro: Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva, 2019. ISBN 978-85-7318-389-4.

JOGIN, Manjunath et al. Feature extraction using convolution neural networks (CNN) and deep learning. **2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)**, p. 2319–2323, 2018.

JOSHI, Amruta et al. Skin disease detection and classification. **International Journal of Advanced Engineering Research and Science (IJAERS)**, v. 6, n. 5, p. 396–400, 2019.

KIM, Phil. Convolutional neural network. **MATLAB deep learning: with machine learning, neural networks and artificial intelligence**, Springer, p. 121–147, 2017.

KUNDAKCI, Nihal; ERDEM, Cengizhan. Leprosy: A great imitator. **Clinics in dermatology**, Elsevier, v. 37, n. 3, p. 200–212, 2019.

LAIKOVA, Kateryna V et al. Advances in the understanding of skin cancer: ultraviolet radiation, mutations, and antisense oligonucleotides as anticancer drugs. **Molecules**, MDPI, v. 24, n. 8, p. 1516, 2019.

LANG, Steven et al. WekaDeeplearning4j: A deep learning package for Weka based on Deeplearning4j. **Knowledge-Based Systems**, Elsevier, v. 178, p. 48–50, 2019.

LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.

MENEGOLA, Afonso et al. Knowledge transfer for melanoma screening with deep learning. **2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)**, p. 297–300, 2017.

PEREZ, Fábio; AVILA, Sandra; VALLE, Eduardo. Solo or Ensemble? Choosing a CNN Architecture for Melanoma Classification. **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**, p. 2775–2783, 2019.

RIBEIRO, Mara Dayanne Alves; SILVA, Jefferson Carlos Araujo; OLIVEIRA, Sabrynna Brito. Estudo epidemiológico da hanseníase no Brasil: reflexão sobre as metas de eliminação. **Revista Panamericana de Salud Pública**, SciELO Public Health, v. 42, e42, 2018.

SANTACROCE, Luigi et al. Mycobacterium leprae: A historical study on the origins of leprosy and its social stigma. **Le Infezioni in Medicina**, Edizioni Internazionali srl, v. 29, n. 4, p. 623, 2021.

SHALU; KAMBOJ, Aman. A color-based approach for melanoma skin cancer detection. **2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)**, p. 508–513, 2018.

SILVA, Samuel Freire da. **Atlas Dermatológico.** 1995. Disponível em: `http://www.atlasdermatologico.com.br/`. Acesso em: 01/11/2021.

SIXTH INTERNATIONAL CONGRESS OF LEPROSY. The technical resolution on classification. **International Journal of Leprosy and Other Mycobacterial Diseases**, v. 21, n. 4, p. 504–516, 1953.

SUNG, Hyuna et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 71, n. 3, p. 209–249, 2021.

SZEGEDY, Christian; IOFFE, Sergey et al. Inception-v4, inception-resnet and the impact of residual connections on learning. **Proceedings of the AAAI conference on artificial intelligence**, v. 31, n. 1, 2017.

SZEGEDY, Christian; VANHOUCKE, Vincent et al. Rethinking the inception architecture for computer vision. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 2818–2826, 2016.

TSCHANDL, Philipp et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. **JAMA dermatology**, American Medical Association, v. 155, n. 1, p. 58–65, 2019.

WEN, David et al. Characteristics of publicly available skin cancer image datasets: a systematic review. **The Lancet Digital Health**, Elsevier, v. 4, n. 1, e64–e74, 2022.

WORLD HEALTH ORGANIZATION. **Neglected tropical diseases**. 2021. Disponível em:`https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases`. Acesso em: 25/04/2021.

_____. **WHO report on cancer: setting priorities, investing wisely and providing care for all**. Geneva: World Health Organization, 2020. ISBN 978-92-4-000129-9.

YU, Kun-Hsing; BEAM, Andrew L; KOHANE, Isaac S. Artificial intelligence in healthcare. **Nature biomedical engineering**, Nature Publishing Group, v. 2, n. 10, p. 719–731, 2018.

ZOPH, Barret et al. Learning transferable architectures for scalable image recognition. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 8697–8710, 2018.