

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Aprimoramento da Detecção de Câncer em Bases de Imagens Limitadas: Uma Abordagem Baseada em Ensemble e Aumento de Dados

Fernando Humberto de Almeida Moraes Neto

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Fernando Humberto de Almeida Moraes Neto

Aprimoramento da Detecção de Câncer em Bases de Imagens Limitadas: Uma Abordagem Baseada em Ensemble e Aumento de Dados

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
Março de 2025

Fernando Humberto de Almeida Moraes Neto

Improving Cancer Detection in Limited Image Bases: An
Ensemble-Based and Data Augmentation Approach

Doctoral dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP
and to the Department of Statistics – DEs-UFSCar, in
partial fulfillment of the requirements for the degree of
the Doctorate Interagency Program Graduate in
Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
March 2025



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Fernando Humberto de Almeida Moraes Neto, realizada em 18/02/2025.

Comissão Julgadora:

Prof. Dr. Adriano Kamimura Suzuki (ICMC/USP)

Prof. Dr. Francisco Louzada Neto (USP)

Prof. Dr. Ricardo Ferreira da Rocha (UFBA)

Prof. Dr. Erlandson Ferreira Saraiva (UFMS)

Prof. Dr. Almir Galvão Vieira Bittencourt (A.C.Camargo)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M827a Moraes Neto, Fernando Humberto de Almeida
Aprimoramento da Detecção de Câncer em Bases de
Imagens Limitadas: Uma Abordagem Baseada em
Ensemble e Aumento de Dados / Fernando Humberto de
Almeida Moraes Neto; orientador Adriano Kamimura
Suzuki. -- São Carlos, 2025.
109 p.

Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2025.

1. Rede Neural Convolucional. 2. Transferência
de aprendizado. 3. Ensemble. 4. Data augmentation.
5. Undersampling and Oversampling. I. Kamimura
Suzuki, Adriano, orient. II. Título.

*Aos meus pais, Mauricio (in memoriam) e Maria,
por todo esforço, apoio e incentivo.*

AGRADECIMENTOS

Aos meus pais, Mauricio (*in memoriam*) e Maria, pela incessante motivação e apoio durante toda a minha jornada acadêmica.

À minha irmã, Crislaine, pela amizade genuína e pela parceria inabalável ao longo da vida.

À Hellen, minha companheira, por sua dedicação, carinho, paciência, conselhos valiosos e por todo o suporte. Este trabalho não seria o mesmo sem a sua presença e contribuição.

A minha família, pelos momentos de descontração e pelo suporte incondicional.

Aos amigos e colegas de doutorado, pela convivência enriquecedora e pela amizade. Agradeço especialmente a Camila e Michele pelo apoio constante.

Ao meu amigo e orientador de mestrado, Ricardo Rocha, por todo o suporte durante a minha formação.

A todos os meus professores, pela dedicação e pela generosa partilha de conhecimento.

Ao meu orientador, Adriano Kamimura, pela orientação precisa e por todo o apoio ao longo desta pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro essencial que possibilitou a realização deste trabalho e a concretização dos meus objetivos acadêmicos.

Por fim, agradeço a todos que, de alguma maneira, contribuíram para a conclusão desta etapa.

“Uma mente necessita de livros da mesma forma que uma espada necessita de uma pedra de amolar, se quisermos que se mantenha afiada.”
(George R. R. Martin)

RESUMO

MORAES NETO, F. H. A. . **Aprimoramento da Detecção de Câncer em Bases de Imagens Limitadas: Uma Abordagem Baseada em Ensemble e Aumento de Dados.** 2025. 109 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

Recentemente os modelos de *Deep Learning* têm ganhado muito destaque na área de análise de imagem, porém necessitam de muitos dados para terem bons desempenhos. Isto dificulta a aplicação em bases de dados médicas visto que existe uma dificuldade em se obter muitas observações equiparadamente, entre os diferentes casos como, por exemplo, patológicos e normais. Para bases de dados pequenas pode-se utilizar modelos de transferência de aprendizado, ensemble e aumento de dados para se ter melhores desempenhos na tarefa de classificar imagens, já para base de dados desbalanceadas pode-se utilizar técnicas de reamostragem como *undersampling* e *oversampling*. Neste trabalho propomos uma nova abordagem baseada no ensemble de modelos de transferência de aprendizado com diferentes pesos para melhorar a predição de observações da classe minoritária em pequenas bases de dados desbalanceadas considerando reamostragem e aumento de dados. Dois experimentos foram realizados, o primeiro com a base de dados do Hospital A.C.Camargo Cancer Center, que é um hospital oncológico especializado no diagnóstico, tratamento e pesquisa de câncer. Esta base de dados nunca foi utilizada para nenhum tipo de análise e tem-se como objetivo classificar imagens de mamografia como malignas ou benignas. O segundo experimento é uma base de dados de imagens de câncer de pele benignas ou malignas do ISIC 2016. A partir dos resultados obtidos nota-se que as combinações com melhores métricas são obtidas com a utilização de reamostragem e as piores sem a utilização desta técnica. O que comprova que a utilização da metodologia de reamostragem (*undersampling* e *oversampling*) melhora o desempenho da combinação em prever a classe minoritária. Ainda, com a utilização de aumento de dados, as combinações propostas tem um melhor desempenho do que a não utilização desta técnica.

Palavras-chave: Rede Neural Convolutacional, Transferência de aprendizado, Ensemble, Aumento de Dados, Undersampling e Oversampling.

ABSTRACT

MORAES NETO, F. H. A. . **Improving Cancer Detection in Limited Image Bases: An Ensemble-Based and Data Augmentation Approach.** 2025. 109 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2025.

Recently, Deep Learning models have gained a lot of prominence in the area of image analysis, but they need a lot of data to perform well. This makes the application in medical databases difficult, since there is a difficulty in obtaining many observations equally, between different cases, such as pathological and normal. For small databases, models of learning transfer, ensemble and data augmentation can be used to have better performance in the task of classifying images, while for unbalanced databases, resampling techniques such as undersampling and oversampling can be used. In this work we propose a new approach based on the ensemble of transfer learning models with different weights to improve the prediction of minority class observations in small unbalanced databases considering resampling and data augmentation. At the end, two experiments are carried out, the first with the database from the Hospital A.C.Camargo Cancer Center, which is an oncology hospital specialized in the diagnosis, treatment and research of cancer. This database has never been used for any type of analysis and its objective is to classify mammography images as malignant or benign. The second experiment is a database of benign or malignant skin cancer images from ISIC 2016. From the results obtained, it is noted that the combinations with the best metrics are obtained with the use of resampling and the worst without the use of this technique. This proves that the use of the resampling methodology (undersampling and oversampling) improves the performance of the combination in predicting the minority class. Yet, with the use of data augmentation, the proposed combinations have a better performance than not using this technique.

Keywords: Convolutional Neural Network, Transfer Learning, Ensemble, Data augmentation, Undersampling and Oversampling.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 – Mapa de rede de coautorias com agrupamentos por co-citação dos autores referente a ETCNN. | 28 |
| Figura 2 – Ilustração de um neurônio artificial. | 32 |
| Figura 3 – Redes <i>Feed forward</i> | 33 |
| Figura 4 – Rede MLP. | 33 |
| Figura 5 – Função de ativação sigmoideal. | 35 |
| Figura 6 – Função de ativação ReLU. | 35 |
| Figura 7 – Exemplo de uma ANN. | 37 |
| Figura 8 – Arquitetura geral de uma CNN. | 39 |
| Figura 9 – Fluxograma da classificação de imagens. | 40 |
| Figura 10 – Exemplo do processo de convolução. | 41 |
| Figura 11 – Processo de convolução, dado uma imagem I (3×3) com zero <i>padding</i> e uma máscara de convolução θ (3×3) que produz uma saída s (3×3). | 42 |
| Figura 12 – Processo de convolução 3D, dado uma imagem I com zero <i>padding</i> e uma máscara de convolução θ que produz uma saída s | 43 |
| Figura 13 – Processo do <i>polling</i> , dado uma imagem I (6×6) e um determinado <i>polling</i> (3×3) que produz uma saída (2×2). | 44 |
| Figura 14 – Exemplo de <i>pooling</i> máximo e médio. | 44 |
| Figura 15 – Exemplo de um <i>ensemble</i> com três modelos individuais (A, B, C). | 46 |
| Figura 16 – Exemplo de <i>Ensemble</i> - moda. | 46 |
| Figura 17 – Diferença entre <i>oversampling</i> e <i>undersampling</i> | 47 |
| Figura 18 – Exemplo de uma ANN com e sem <i>dropout</i> , onde n observa-se uma taxa de dropout de 40%, resultando na remoção de 2 neurônios de um total de 5. | 50 |
| Figura 19 – Imagem original. | 50 |
| Figura 20 – Transformações da imagem na Figura 19. | 51 |
| Figura 21 – Fluxograma do funcionamento da GAN. | 52 |
| Figura 22 – Exemplo da utilização do método de <i>early stopping</i> | 52 |
| Figura 23 – Taxa de erro (%) dos modelos do ILSVRC. | 56 |
| Figura 24 – Representação da modelagem proposta e passo a passo, veja que os pontos importantes aumento de dados, <i>undersampling</i> e <i>oversampling</i> estão aplicados após a divisão do treinamento e teste, os 13 modelos são treinados com esses conjuntos e em seguida as 13 combinações são realizadas que geram a predição final de cada um dos <i>k-fold</i> 's. | 58 |

| | |
|--|----|
| Figura 25 – Exemplos de imagens da base de dados de mamografia. | 62 |
| Figura 26 – Pré-processamento da imagem. | 63 |
| Figura 27 – Aumento de dados com a utilização da imagem final na Figura 26. | 63 |
| Figura 28 – Representação da modelagem proposta e passo a passo, adaptado para concatenação da ANN com o modelo TCNN. | 64 |
| Figura 29 – Resultado dos 13 modelos utilizados sem aumento de dados. | 66 |
| Figura 30 – Resultado das 13 combinações utilizadas sem aumento de dados. | 67 |
| Figura 31 – Resultado dos 13 modelos utilizados com aumento de dados. | 67 |
| Figura 32 – Resultado das 13 combinações utilizadas com aumento de dados. | 68 |
| Figura 33 – Resultado do F2 score dos 13 modelos utilizados. | 68 |
| Figura 34 – Resultado do F2 score das 13 combinações utilizadas. | 69 |
| Figura 35 – Resultado do F2 score dos 13 modelos utilizados com a utilização do método <i>undersampling</i> | 69 |
| Figura 36 – Resultado do F2 score das 13 combinações utilizadas com a utilização do método <i>undersampling</i> | 70 |
| Figura 37 – Resultado do F2 score dos 13 modelos utilizados com a utilização do método <i>oversampling</i> | 70 |
| Figura 38 – Resultado do F2 score das 13 combinações utilizadas com a utilização do método <i>oversampling</i> | 71 |
| Figura 39 – Imagens da base de dados do ISIC 2016. | 73 |
| Figura 40 – Pré-processamento da imagem. | 73 |
| Figura 41 – Aumento de dados com a imagem final na Figura 40. | 74 |
| Figura 42 – Resultados dos 13 modelos usando aumento de dados. | 75 |
| Figura 43 – Resultados das 13 combinações usando aumento de dados. | 75 |
| Figura 44 – Resultados do F2 score dos 13 modelos utilizados. | 76 |
| Figura 45 – Resultados do F2 score das 13 combinações utilizadas. | 76 |
| Figura 46 – Resultados do F2 score dos 13 modelos utilizados com o método <i>undersampling</i> | 77 |
| Figura 47 – Resultados do F2 score das 13 combinações utilizadas com o método <i>undersampling</i> | 77 |
| Figura 48 – Resultados do F2 score dos 13 modelos utilizados com o método <i>oversampling</i> | 78 |
| Figura 49 – Resultados do F2 score das 13 combinações utilizadas com o método <i>oversampling</i> | 78 |
| Figura 50 – Resultados do F2 score dos 13 modelos utilizados aumento de dados com a GAN com o <i>Oversampling</i> e <i>Undersampling</i> | 80 |
| Figura 51 – Resultados do F2 score das 13 combinações utilizados aumento de dados com a GAN com o <i>Oversampling</i> e <i>Undersampling</i> | 80 |
| Figura 52 – Resultados do F2 score dos 13 modelos utilizados (10%). | 84 |
| Figura 53 – Resultados do F2 score das 13 combinações utilizados (10%). | 85 |

| | |
|---|----|
| Figura 54 – Resultados do F2 score dos 13 modelos utilizados com o método <i>undersampling</i> (10%). | 85 |
| Figura 55 – Resultados do F2 score das 13 combinações utilizadas com o método <i>undersampling</i> (10%). | 86 |
| Figura 56 – Resultados do F2 score dos 13 modelos utilizados com o método <i>oversampling</i> (10%). | 86 |
| Figura 57 – Resultados do F2 score das 13 combinações utilizadas com o método <i>oversampling</i> (10%). | 87 |
| Figura 58 – Resultados do F2 score dos 13 modelos utilizados (30%). | 88 |
| Figura 59 – Resultados do F2 score das 13 combinações utilizadas (30%). | 88 |
| Figura 60 – Resultados do F2 score dos 13 modelos utilizados com o método <i>undersampling</i> (30%). | 89 |
| Figura 61 – Resultados do F2 score das 13 combinações utilizadas com o método <i>undersampling</i> (30%). | 89 |
| Figura 62 – Resultados do F2 score dos 13 modelos utilizados com o método <i>oversampling</i> (30%). | 90 |
| Figura 63 – Resultados do F2 score das 13 combinações utilizadas com o método <i>oversampling</i> (30%). | 90 |
| Figura 64 – Resultados do F2 score dos 13 modelos utilizados (40%). | 91 |
| Figura 65 – Resultados do F2 score das 13 combinações utilizadas (40%). | 92 |
| Figura 66 – Resultados do F2 score dos 13 modelos utilizados (50%). | 93 |
| Figura 67 – Resultados do F2 score das 13 combinações utilizadas (50%). | 93 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 – Matriz de confusão 2×2 | 48 |
| Tabela 2 – Medidas de avaliação. | 49 |
| Tabela 3 – Modelos e suas predições | 57 |
| Tabela 4 – Número de elementos pertencentes a cada classe na base de dados de mamografia. | 62 |
| Tabela 5 – Número de elementos pertencentes a cada classe na base de dados de mamografia com a utilização do <i>undersampling</i> e do <i>oversampling</i> no primeiro <i>fold</i> | 63 |
| Tabela 6 – Exemplo de uma matriz de confusão considerando um modelo que classifica dois tipos de tumor maligno e benigno. | 65 |
| Tabela 7 – Número de elementos pertencentes a cada uma das classes da base de dados do ISIC 2016. | 73 |
| Tabela 8 – Número de elementos pertencentes a cada classe no conjunto de dados ISIC usando <i>undersampling</i> e <i>oversampling</i> em <i>fold-1</i> | 74 |
| Tabela 9 – Comparação de outros trabalhos do ISIC 2016 no conjunto de teste. | 79 |

SUMÁRIO

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 23 |
| 1.1 | Revisão de literatura | 26 |
| 1.2 | Objetivos | 29 |
| 1.3 | Organização do Trabalho | 30 |
| 2 | CONTEÚDOS PRELIMINARES | 31 |
| 2.1 | Redes Neurais Artificiais | 31 |
| 2.1.1 | <i>Função de ligação</i> | 34 |
| 2.1.2 | <i>Treinamento das ANN's</i> | 36 |
| 2.1.2.1 | <i>Backpropagation</i> | 36 |
| 2.1.3 | <i>Método do Gradiente Adaptativo</i> | 37 |
| 2.2 | Redes Neurais Convolucionais | 38 |
| 2.2.1 | <i>Arquitetura geral</i> | 39 |
| 2.2.1.1 | <i>Processamento de imagens</i> | 39 |
| 2.2.1.2 | <i>Processo de convolução</i> | 40 |
| 2.2.1.3 | <i>Pooling</i> | 43 |
| 2.2.1.4 | <i>Achatamento</i> | 45 |
| 2.2.2 | <i>Técnicas utilizadas com CNN's</i> | 45 |
| 2.2.2.1 | <i>Transferência de Aprendizado</i> | 45 |
| 2.2.2.2 | <i>Ensemble</i> | 45 |
| 2.2.2.3 | <i>Undersampling e Oversampling</i> | 47 |
| 2.2.3 | <i>Métricas de avaliação</i> | 48 |
| 2.2.4 | <i>Métodos de evitar overfitting</i> | 49 |
| 2.2.4.1 | <i>Dropout</i> | 49 |
| 2.2.4.2 | <i>Aumento de Dados</i> | 50 |
| 2.2.4.3 | <i>Aumento de dados por rede adversária generativa</i> | 51 |
| 2.2.4.4 | <i>Early stopping</i> | 52 |
| 2.3 | Conclusão | 53 |
| 3 | MODELO PROPOSTO | 55 |
| 3.1 | Modelos de transferência de aprendizado | 55 |
| 3.2 | <i>Ensembles Propostos</i> | 57 |
| 3.3 | Modelo final proposto | 58 |

| | | |
|-------|---|----|
| 3.4 | Conclusão | 59 |
| 4 | APLICAÇÕES | 61 |
| 4.1 | Câncer de Mama | 61 |
| 4.1.1 | <i>Base de dados</i> | 61 |
| 4.1.2 | <i>Configurações do Experimento</i> | 64 |
| 4.1.3 | <i>Resultados</i> | 66 |
| 4.2 | Câncer de pele | 72 |
| 4.2.1 | <i>Base de dados</i> | 72 |
| 4.2.2 | <i>Configurações do Experimento</i> | 74 |
| 4.2.3 | <i>Resultados</i> | 74 |
| 4.2.4 | <i>Aumento de dados com GAN</i> | 79 |
| 4.3 | Conclusão | 81 |
| 5 | ESTUDO DE SIMULAÇÃO | 83 |
| 5.1 | Estrutura da simulação | 83 |
| 5.2 | Simulação | 84 |
| 5.2.1 | <i>Simulação 1</i> | 84 |
| 5.2.2 | <i>Simulação 3</i> | 87 |
| 5.2.3 | <i>Simulação 4</i> | 91 |
| 5.2.4 | <i>Simulação 5</i> | 92 |
| 5.3 | Conclusão | 94 |
| 6 | CONSIDERAÇÕES FINAIS E PESQUISAS FUTURAS | 97 |
| 6.1 | Considerações Finais | 97 |
| 6.2 | Sugestões para pesquisas futuras | 98 |
| | REFERÊNCIAS | 99 |

INTRODUÇÃO

[Backes e Junior \(2019\)](#) definem a visão computacional como sendo a área de pesquisa que tenta repassar para as máquinas a capacidade de captar imagens, retirar ruídos, extrair informações, relacionar com outras imagens vistas anteriormente, dentre outros. A visão computacional inspira-se na capacidade humana de interpretar e descrever a estrutura tridimensional do ambiente ao redor. Esse campo utiliza técnicas de processamento digital de imagens para preparar as imagens de entrada, facilitando sua interpretação por algoritmos de visão computacional ([FERNANDES, 2013](#)). Segundo [Ribeiro \(2021\)](#), nos últimos anos modelos de Visão Computacional estão sendo muito utilizados para auxílio no diagnóstico por imagens na comunidade científica. Ainda, pelo mesmo autor, na área médica, uma das principais aplicações da visão computacional é a extração de informações a partir de exames de imagem, oferecendo uma análise mais detalhada e precisa de patologias específicas.

Pesquisas que propõem a utilização de *softwares* para o diagnóstico médico são também as principais responsáveis por permitir os avanços observados na visão computacional e medicina. Segundo [Calas, Gutflen e Pereira \(2012\)](#), foi realizado um estudo com um *software Computer Aided Detection* (CAD) de apoio para comparação de duas abordagens: a primeira envolve o uso do software pelo radiologista, enquanto a segunda realiza a classificação sem esse recurso. O estudo revelou que a sensibilidade dos radiologistas foi de 79% sem o auxílio do *software* e aumentou para 84% com a ajuda do *software*. [Leibig et al. \(2022\)](#) em seu trabalho verificam que radiologistas auxiliados por um modelo que classifica imagens de mamografia, realizam o diagnóstico do câncer de mama com mais sucesso do que quando trabalham sozinhos.

Segundo [DALVI \(2018\)](#) o aumento expressivo no número de casos de câncer, somado à escassez de métodos eficazes para sua cura, especialmente em estágios avançados, representa uma preocupação global para a saúde pública. Dessa forma, investir em métodos diagnósticos torna-se essencial para possibilitar a detecção precoce da doença, aumentando as chances de tratamento bem-sucedido. A radiografia se destaca como um dos principais métodos de detecção

da doença. As imagens geradas por esse exame são analisadas por radiologistas, cujas decisões diagnósticas podem ser influenciadas por diversos fatores. Uma discussão sobre essas influências é apresentada em (FREITAS, 1988).

Estima-se que, em 2023, aproximadamente 30 milhões de novos casos de câncer foram diagnosticados globalmente (SUNG *et al.*, 2023). Devido à natureza do câncer, a detecção precoce possibilita tratamentos menos invasivos e aumenta significativamente a taxa de sucesso, resultando em maior chance de cura. Diversas iniciativas têm sido adotadas para facilitar o diagnóstico precoce do câncer (INCA, 2019).

Uma abordagem para o diagnóstico do câncer de mama, conforme descrito por Buciu e Gacsadi (2011), envolve a captura de mamografias em pacientes com alto risco. A mamografia permite a detecção de alterações suspeitas antes do surgimento de sintomas visíveis, e sua análise é realizada por radiologistas, que identificam manualmente possíveis sinais de câncer na mama. Técnicas computacionais, como as Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks* ou CNNs), surgem como ferramentas auxiliares para melhorar a precisão na classificação de mamografias, apoiando os especialistas no diagnóstico.

A CNN (Rede Neural Convolucional) é uma evolução das Redes Neurais Artificiais, ou *Artificial Neural Networks* (ANNs), amplamente aplicada na análise de imagens, pois leva em conta a estrutura bidimensional dos dados. Algumas CNN's se tornaram populares na classificação de imagens nos últimos anos como, por exemplo, Alexnet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) Densenet (HUANG *et al.*, 2017), VGG (SIMONYAN; ZISSERMAN, 2014) e a Resnet (HE *et al.*, 2016a). Contudo esses modelos necessitam de muitos dados para apresentarem bons desempenhos, o que dificulta a utilização da técnica em pequenas bases de dados e/ou bases desbalanceadas. Na literatura existem algumas abordagens para se trabalhar com pequenas bases de dados, como a utilização de aumento de dados, transferência de aprendizado e o *ensemble*. Já para bases de dados desbalanceadas pode-se utilizar técnicas de *oversampling* e *undersampling*, que são técnicas de reamostragem.

Segundo Yang (2008) o processo de ajustar um modelo através de uma rede já treinada é chamado de transferência de aprendizado. Segundo Chaves (2019), esse método é muito utilizado em modelos de *Deep Learning*, visto que é necessário uma grande quantidade de dados e de recursos computacionais para treinar esse tipo de modelo. Neto (2017) exemplifica esse processo de transferência de aprendizagem como uma pessoa que está vendo um objeto e o descreve para outra pessoa. Por sua vez, a segunda pessoa que não está vendo o objeto descrito, mediante as características faladas irá traçar um perfil com as características que ele conhece e das recebidas pela primeira pessoa e associar a um objeto que ele próprio já conhece e que possua características semelhantes.

Diversos trabalhos mostram a eficácia da aplicação dos modelos de transferência de aprendizado. Como exemplo tem-se Aneja e Aneja (2019) que utilizam sete modelos de transferência de aprendizado para reconhecer dígitos manuscritos do alfabeto Devanagari, com o

melhor modelo tendo 98% de precisão. [Morid, Borjali e Fiol \(2021\)](#) fazem uma revisão bibliográfica de 102 artigos com o objetivo de auxiliar futuros estudos de modelos de transferência de aprendizado.

Mesmo com a utilização de transferência de aprendizado, para pequenas bases de dados esses modelos ainda não apresentam bons desempenhos. De acordo com [Goodfellow, Bengio e Courville \(2016\)](#), uma das melhores estratégias para garantir uma boa generalização do modelo é aumentar a quantidade de dados utilizados no treinamento ou seja, generalizar bem para exemplos que ele nunca viu antes. Existem várias abordagens para realizar esse aumento de dados. Dentre elas destaca-se aplicar combinações de operações sobre as imagens originais. São exemplos deste recurso a translação, achatamento, alongamento, distorção, espelhamento, rotação, adição de ruído, dentre outros.

Uma abordagem adicional para aprimorar o desempenho de um modelo isolado é empregar a combinação de múltiplos modelos, conhecida como *ensemble*. De acordo com [Ferreira \(2012\)](#), um *ensemble* consiste em um modelo que agrega vários outros modelos, conhecidos como componentes, cujas previsões são combinadas para realizar a classificação de uma amostra que o modelo ainda não viu durante o treinamento. Essa combinação visa aumentar a precisão e a robustez do modelo final, aproveitando as diferentes características dos modelos componentes para reduzir o erro de predição e melhorar a generalização. [Antipov, Berrani e Dugelay \(2016\)](#) e [Minetto, Segundo e Sarkar \(2019\)](#) são exemplos de autores que se utilizam de *ensemble* em seus trabalhos para melhorar o desempenho dos modelos individuais e realizar aplicações. Conforme observado por [Chen e Shakhnarovich \(2014\)](#), ainda não há um consenso sobre a quantidade ideal de modelos a serem incluídos em um *ensemble*, sobre o grau de redução de erro necessário ou sobre qual método específico de combinação entre modelos oferece os melhores resultados.

Mesmo no contexto de pequenas bases de dados podemos ter bases de dados desbalanceadas, que é um problema muito comum no contexto de classificação. Segundo [Gama et al. \(2012\)](#) uma base de dados desbalanceada refere-se a um conjunto de dados onde as classes não estão distribuídas de maneira uniforme. Uma forma simples de solucionar esse desbalanceamento é retirar aleatoriamente observações da classe majoritária até se obter o mesmo número de observações que a classe minoritária (*undersampling*), ou aumentar o número de dados da classe minoritária até obter o mesmo número de observações que a classe majoritária (*oversampling*). [Mohammed, Rawashdeh e Abdullah \(2020\)](#) realizam um estudo sobre a utilização dessas técnicas em diversos cenários, obtendo que o *oversampling* tem um desempenho melhor do que o *undersampling* para diferentes classificadores.

Desta forma, uma revisão de literatura é feita com base em trabalhos que utilizam combinações de modelos de transferência de aprendizado com enfoque em diagnóstico médico e utilização de técnicas de aumento de dados e reamostragem.

1.1 Revisão de literatura

As CNN's já são muito utilizadas para detecção de câncer em imagens médicas, como, por exemplo, para detecção de câncer de colon (HASAN *et al.*, 2024; KARTHIKEYAN; JOTHILAKSHMI; SUTHIR, 2024), câncer de pulmão (MAJUMDER *et al.*, 2024; LEE *et al.*, 2024), câncer gastrointestinal (WU *et al.*, 2024; AUZINE *et al.*, 2024), câncer de mama (MAHORO; AKHLOUFI, 2024; CHAKRAVARTHY *et al.*, 2024) e câncer de pele (KUMAR *et al.*, 2024; ALI *et al.*, 2024), dentre outros. As CNN's se destacam como um recurso inovador que proporcionam um avanço científico em diversas áreas do conhecimento. As camadas convolucionais garantem o armazenamento de uma quantidade paramétrica menor, o que fornece a redução da memória utilizada em modelos de estudos.

Na classificação de imagens médicas existe uma dificuldade em se obter muitas observações equiparadamente. Por isso, devido a proporção desigual (desbalanceada) de casos patológicos e normais, uma CNN pode ter dificuldade em acertar os dados da classe com a menor quantidade de observações. É muito comum encontrar trabalhos de dados médicos que são desbalanceados e algumas estratégias são utilizadas para suavizar esse desbalanceamento. Segundo Celik (2023) duas técnicas de reamostragem para classes desbalanceadas são: *undersampling* e *oversampling*. No *undersampling* a redução da amostra de classes (subamostragem) é realizada balanceando o número de outras classes no conjunto de dados com a contagem mínima de amostras de classes, já no *oversampling* o balanceamento é feito tornando o número de outras classes no conjunto de dados igual ao número mais alto da classe.

Qu *et al.* (2020) utilizaram *undersampling* e *oversampling* para mitigar a classe desbalanceada em um problema de classificação utilizando CNN em um conjunto de dados de radiografia de tórax. O autor obteve que tais métodos foram eficazes para mitigar o desbalanceamento. Pereira e Saraiva (2021) utilizaram dois conjuntos desbalanceados de radiografia da mama e fizeram uma comparação da utilização de *undersampling* e *oversampling* com a não utilização dessas técnicas, obtendo que a utilização dessas técnicas de reamostragem são uma maneira eficaz de combater o impacto do desbalanceamento dos dados de treinamento.

Um outro problema bastante comum em *Deep Learning* é a generalização do modelo, visto que esses modelos são excessivamente complexos, com muitos parâmetros em comparação com o número de observações de treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016). Isto significa que em alguns casos uma CNN pode ter bom desempenho nos dados de treinamento, mas ter dificuldade em prever novas observações. Para resolver essa questão, é comum aplicar técnicas de aumento de dados no conjunto de treinamento. Em um estudo de classificação de lesão de pele, Li *et al.* (2018) utilizaram uma CNN e aumento de dados no conjunto de treinamento, comparando o resultado com e sem o aumento, ainda realizando uma análise da influência de diferentes números de amostras do aumento de dados no desempenho de diferentes classificadores. O aumento de dados também pode ser utilizado no *ensemble* de modelos de transferência de aprendizado. Em seu trabalho Chatterjee *et al.* (2022) realizam uma

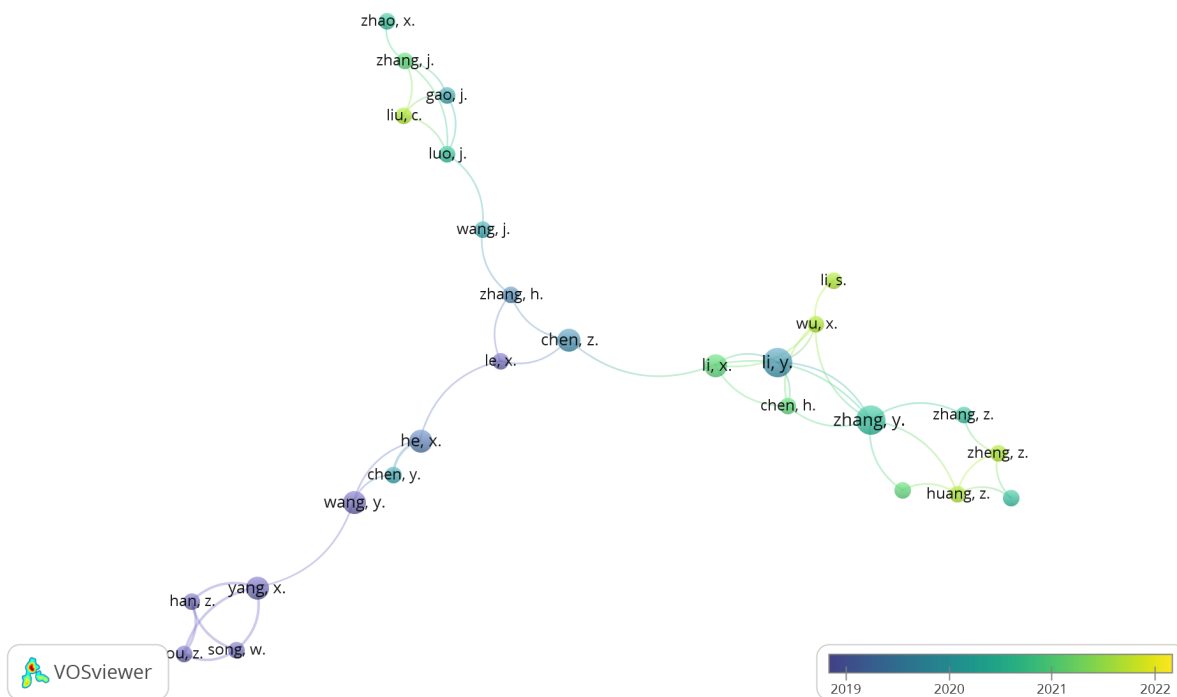
combinação de modelos de transferência de aprendizado para detecção de objetos em imagens de veículos aéreos não tripulados, com aumento de dados para solucionar o problema do número de observações.

Uma alternativa a utilização da CNN quando se tem poucas observações é utilizar transferência de aprendizado. Segundo [Thomas e Jerome \(2024\)](#) esta é uma técnica de aprendizado de máquina que nos permite utilizar modelos pré-treinados para novas tarefas. Isto pode ser particularmente útil para tarefas de imagens médicas, onde os dados rotulados são frequentemente escassos.

No contexto médico, alguns modelos de transferência de aprendizado também tem sido aplicados. [Beghriche et al. \(2024\)](#) e [RS et al. \(2024\)](#), utilizaram modelos de transferência de aprendizado no contexto de detecção de câncer de mama. O primeiro estudo utilizou as redes ResNet50, MobileNetV2, DenseNet121 e Xception para transferência de aprendizado, já o segundo artigo utilizou apenas a ResNet152v2. Em ambos os trabalhos os modelos apresentaram bons desempenhos com a utilização da abordagem.

Além disso, a combinação de diferentes modelos de transferência de aprendizado pode ser uma estratégia eficaz. Na literatura muitos são os trabalhos que descrevem a combinação de modelos de transferência de aprendizado (ETCNN). Para enfatizar este fato, bases de dados do repositório online Scopus foram utilizadas. Segundo [Burnham \(2006\)](#) o Scopus é uma base de dados que reúne resumos e citações de artigos de periódicos acadêmicos, servindo como uma importante fonte para pesquisas científicas e acadêmicas. No qual, segundo [Ribeiro e Tavares \(2017\)](#) é possível obter a mensuração da produção acadêmica de temas, áreas do conhecimento, com foco na contagem de autoria e coautoria, mediante as análises de publicações, citações, co-citações, dentre outros fatores. Deste modo, a [Figura 1](#) destaca o mapa de rede que ressalta os principais autores de forma qualitativa, com o número de citações, e quantitativas, com o número de publicações, que desenvolveram estudos sobre a combinação de modelos de transferência de aprendizado, mostrando o avanço literário do tema nos últimos anos.

Figura 1 – Mapa de rede de coautorias com agrupamentos por co-citação dos autores referente a ETCNN.



Fonte: Elaborada pelo autor.

Note que a intensidade das conexões entre os autores indica o número de publicações em conjunto a combinação de modelos de transferência de aprendizado, enquanto o tamanho dos “nós” referem-se ao volume de publicações nos últimos anos sobre este tema e os *clusters* indicam os grupos coautorais. Como presenciado na Figura 1, autores como [Xu et al. \(2017\)](#) e [Zhang et al. \(2019\)](#) se destacam entre os autores que abordam a técnica em seus estudos com um maior número de citações em outras pesquisas e uma maior quantidade de publicações referentes ao tema. Desta forma, alguns trabalhos envolvendo ETCNN são apresentados a seguir com o intuito de enfatizar a importância do tema e de sua aplicação no âmbito acadêmico.

Um desses trabalhos foi o de [Zhou et al. \(2021\)](#), que utilizam uma combinação de três modelos de transferência de aprendizado (AlexNet, GoogleNet e ResNet) para prever o diagnóstico de COVID-19 através de imagens de radiografia do tórax. Seus resultados foram combinados utilizando a votação, obtendo uma classificação melhor com a combinação do que no modelo individual. [Zhu et al. \(2020\)](#) fazem uma comparação da utilização da estrutura clássica de aprendizado de transferência e a utilização do *ensemble* no contexto de previsão da resposta a drogas anti-câncer. Os autores concluem que a utilização do *ensemble* tem um melhor desempenho. [Zhang et al. \(2019\)](#) utilizam um ETCNN com três modelos para classificação de sentimento e concluem que decisões de modelos combinados podem melhorar ainda mais a precisão da classificação de sentimentos.

A combinação de modelos de transferência de aprendizado é feita utilizando uma função

de generalização, em que a mais comum no contexto de classificação é a moda. Um exemplo dessa combinação, empregando a moda, é demonstrado em [Minetto, Segundo e Sarkar \(2019\)](#). Recentemente outros ensembles vem sendo propostos, como o de [Taherkhani, Cosma e McGinnity \(2020\)](#) que integraram o *Adaptive Boosting* (AdaBoost) a uma CNN, projetando um novo método nomeado de AdaBoost-CNN.

Alguns trabalhos mais recentes fazem a comparação da utilização dos modelos individuais com a utilização do *ensemble* nos modelos de transferência de aprendizado. [Mondal et al. \(2021\)](#) utilizam um ETCNN para diagnosticar leucemia linfoblástica a partir de imagens microscópicas. O autor compara os resultados considerando vários aumentos de dados, pesos e pré processamentos, obtendo que a combinação tem melhor desempenho do que os modelos considerados individualmente. [Ahmad, Khan e Javed \(2021\)](#) utilizam um ETCNN comparando o *ensemble* proposto com os modelos da literatura com o objetivo de prever Covid-19 através de imagens de radiografia de tórax, também obtendo que a combinação tem melhor desempenho. [Gültekin e Uğur \(2021\)](#) utilizam transferência de aprendizado e propõe novas funções generalizadoras no *ensemble* para aumentar a performance de CNN's em tarefas de classificação em pequenos conjuntos de dados. O método proposto pelos autores apresenta melhores precisões de classificação do que a utilização de apenas um modelo de transferência de aprendizado. [Xu et al. \(2017\)](#) apresentam uma nova arquitetura utilizando três CNN's para detecção de objetos RGB-D, fazendo uma comparação com as redes individuais VGG16 e Alexnet.

A maioria das abordagens citadas anteriormente para utilização de uma combinação de modelos de transferência de aprendizado em pequenas bases de dados tem boas predições em bases de dados balanceadas, mas o mesmo pode não ocorrer em bases de dados desbalanceadas. Dessa maneira é importante utilizar técnicas para balanceamento da base de dados.

O principal objetivo deste estudo é mostrar que a utilização de ETCNN aliada a técnicas de reamostragem e aumento de dados, é uma abordagem altamente eficaz para aprimorar a predição de observações da classe minoritária em bases de dados desbalanceadas. Embora diversos estudos recentes explorem ETCNN, a metodologia proposta neste trabalho se diferencia ao integrar a reamostragem, o ensemble e o aumento de dados no treinamento, buscando maximizar a capacidade preditiva do modelo em contextos com restrição de dados.

1.2 Objetivos

- Realizar uma revisão teórica sobre Redes Neurais Convolucionais, focando na utilização da combinação de modelos de transferência de aprendizado, além do estudo da aplicação de aumento de dados, *undersampling* e *oversampling* nesta combinação para bases de dados desbalanceadas. Estes tópicos subsidiaram todo o desenvolvimento dos resultados aqui apresentados.

- Propor uma nova abordagem para melhorar o desempenho de CNN's em bases de dados pequenas e desbalanceadas com a utilização de métodos de transferência de aprendizado, *ensemble*, aumento de dados e técnicas de reamostragem (*undersampling* e *oversampling*).
- Desenvolver uma combinação de modelos de transferência de aprendizado com a utilização de reamostragem e aumento de dados para detecção precoce de câncer em imagens de radiografia da mama e pele.
- Comparar o método proposto sem a utilização da combinação, aumento de dados, *undersampling* e *oversampling*.

1.3 Organização do Trabalho

Os capítulos deste trabalho são organizados e divididos da seguinte forma:

- No Capítulo 2 é realizada uma revisão teórica sobre as Redes Neurais Artificiais e Convolucionais. Ainda, são comentados o aumento de dados, *ensemble*, *undersampling* e *oversampling* nas CNN's. Esta revisão foi necessária para apoiar o desenvolvimento do restante deste trabalho.
- No Capítulo 3 é apresentado o modelo proposto que ressalta a composição entre modelos de transferência de aprendizado, sendo destacada a utilização de aumento de dados, *undersampling* e *oversampling*.
- No Capítulo 4 são exibidas duas aplicações. A primeira aplicação é em um conjunto de dados de imagens de mama para detecção de câncer maligno ou benigno cedido pelo A.C.Camargo Cancer Center e em um conjunto de dados de imagens de lesões de pele para detecção de câncer maligno ou benigno. A segunda aplicação é em um conjunto de imagens de pele para detecção de câncer maligno ou benigno em que a base de dados é do ISIC 2016.
- No Capítulo 5 é realizada uma simulação com o objetivo de avaliar o comportamento das combinações e modelos individuais considerando as reamostragens em cenários desbalanceados e balanceados.
- No Capítulo 6 é apresentado um resumo dos principais resultados presentes neste trabalho. Além disso, são listadas algumas propostas futuras.

CONTEÚDOS PRELIMINARES

Neste capítulo será apresentada uma revisão teórica sobre os principais temas que subsidiam este trabalho. Dentre os tópicos destacam as Redes Neurais Artificiais (RNAs) sendo fundamentado segundo os estudos de [Haykin e Network \(2004\)](#) e [Fausett *et al.* \(1994\)](#). Além das Redes Neurais Convolucionais, embasado nos trabalhos de [Goodfellow, Bengio e Courville \(2016\)](#) e [Ferreira \(2019\)](#), com a intuito de promover uma base teórica para o modelo proposto, exibido no Capítulo 3.

2.1 Redes Neurais Artificiais

O desenvolvimento das RNAs foi motivado pelo interesse em compreender e modelar o funcionamento do cérebro humano, conforme descrito em [Fausett *et al.* \(1994\)](#). O cérebro é composto por bilhões de neurônios interconectados, cada um estabelecendo, em média, cerca de 100 conexões com outros neurônios. Essas conexões ocorrem por meio de sinapses, que transmitem impulsos elétricos e viabilizam a comunicação neural.

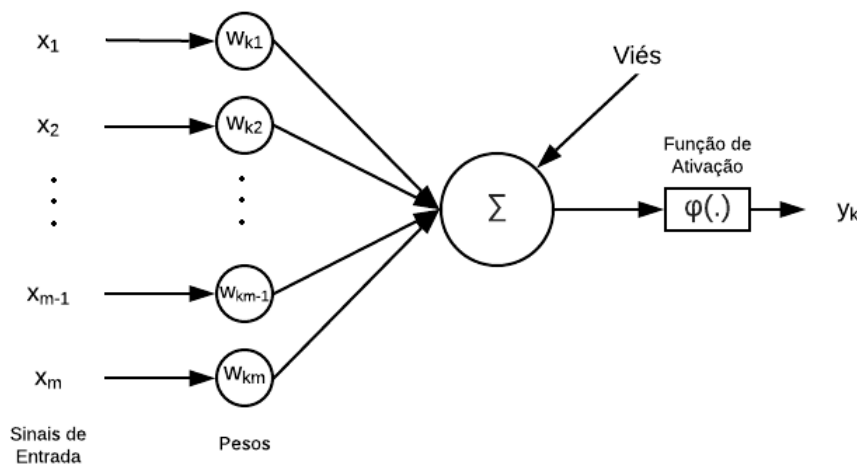
Em termos estruturais, cada neurônio biológico possui três componentes principais: dendritos, corpo celular e axônio. Conforme descrito em [Ferneda \(2006\)](#), os dendritos são responsáveis por receber estímulos ao longo do tempo e transmitir essas informações para o corpo celular, onde o processamento dos sinais ocorre. Quando os estímulos atingem um nível mínimo, ou limiar, o corpo celular gera um novo impulso, transmitido pelo axônio para outros neurônios através das sinapses.

Analogamente, uma Rede Neural Artificial é composta por múltiplas unidades de processamento interligadas por um grande número de conexões ponderadas ([SEEMAN, 2006](#)). Cada neurônio artificial, representado na Figura 2, recebe m sinais de entrada (x_1, \dots, x_m) , cada um associado a um peso (w_{k1}, \dots, w_{km}) . O neurônio realiza uma soma ponderada dos sinais recebidos, $\sum_{i=1}^m x_i w_{ki}$, e decide, com base em uma função de ativação $\varphi(u)$, se gera uma saída,

y_k . O índice k indica o neurônio de destino na camada atual.

O autor [Babini e Marranghello \(2008\)](#) cita que a entrada (x_1, \dots, x_m) , que têm a função dos dendritos do neurônio biológico, um único terminal de saída y , parecido com o axônio (do neurônio biológico) e um somador central mais um função de ativação, que podem ser comparados ao corpo celular (do neurônio biológico).

Figura 2 – Ilustração de um neurônio artificial.



Fonte: Adaptado de [McCulloch e Pitts \(1943\)](#).

A saída deste neurônio, conforme descrito por ([BABINI; MARRANGHELLO, 2008](#)), pode ser representada da seguinte forma:

$$y_k = \varphi \left(\sum_{i=1}^m x_i w_{ki} + b \right)$$

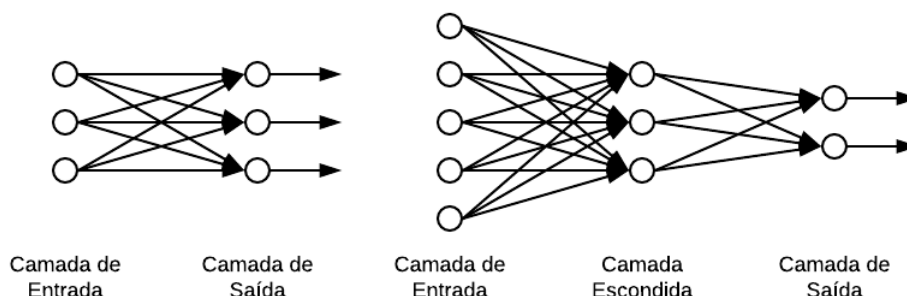
no qual b é o viés e φ é a função de ativação. De acordo com [McCulloch e Pitts \(1943\)](#) e [Babini e Marranghello \(2008\)](#), o neurônio terá sua saída ativada somente se $\sum_{i=1}^m x_i w_{ki} \geq \theta$, onde θ é o limiar. Essa característica do neurônio de chamada de propriedade tudo-ou-nada [Haykin \(1999\)](#).

De acordo com [Haykin \(1999\)](#), a organização dos neurônios em uma Rede Neural está diretamente vinculada ao algoritmo de aprendizado empregado em seu treinamento. As arquiteturas de RNAs mais frequentemente adotadas são as de tipo *Feed forward* e Recorrentes.

Em redes do tipo *Feed forward*, os sinais propagam-se de maneira unidirecional: fluem apenas da camada de entrada para a camada de saída, sem retorno ([ZANINI, 2004](#)). Essas redes não possuem feedback das saídas para as entradas e são estruturadas em camadas, podendo ter uma ou mais camadas intermediárias. A Figura 3 ilustra essa arquitetura, na qual todos os sinais

se deslocam da esquerda para a direita. A segunda rede possui uma camada intermediária ou oculta.

Figura 3 – Redes *Feed forward*.

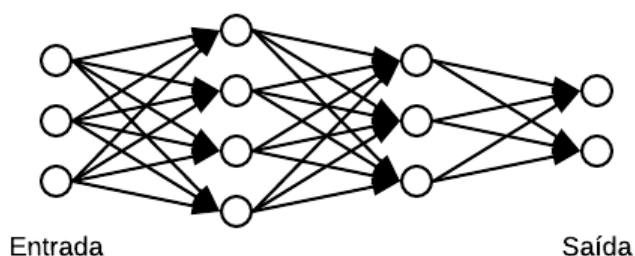


Fonte: Adaptado de [Haykin \(1999\)](#)

Uma limitação das ANN's de camada única é a incapacidade de aprender funções não lineares ([FERREIRA, 2017](#)). Entretanto, modelos capazes de lidar com não linearidades podem ser construídos utilizando Redes Neurais de múltiplas camadas. A vantagem de se ter mais camadas em uma rede neural está relacionada à sua capacidade de representar e aprender padrões complexos ([HAYKIN, 1999](#)).

Segundo [Haykin \(1999\)](#), redes neurais com arquitetura feedforward que possuem ao menos uma camada oculta são conhecidas como Perceptrons Multicamadas (Multilayer Perceptron - MLP) (Figura 4). De acordo com [Silva et al. \(1998\)](#), essas redes têm se mostrado eficazes na resolução de problemas que envolvem não linearidades. O treinamento das MLPs é supervisionado e, geralmente, realizado por meio do algoritmo backpropagation.

Figura 4 – Rede MLP.



Fonte: Adaptado de [Haykin \(1999\)](#)

Segundo [Haykin e Network \(2004\)](#) para obter a saída final dessas redes, é necessário calcular a saída de todas as unidades presentes em cada camada. Ao considerar uma rede com t

camadas intermediárias, a saída da primeira camada é dada por:

$$y_j^1 = \varphi \left(\sum_{k=1}^m x_k w_{jk}^1 + b_j^1 \right), \quad j = 1, \dots, m$$

onde m representa a quantidade de neurônios em cada camada, w_{jk}^1 são os pesos que conectam o k -ésimo neurônio da camada de entrada ao j -ésimo neurônio da primeira camada oculta, b_j^1 é o termo de bias da primeira camada.

Para calcular a saída final da rede, é necessário propagar as ativações ao longo das camadas intermediárias. A saída da segunda camada é dada por:

$$y_j^2 = \varphi \left(\sum_{k=1}^n \left(\sum_{i=1}^m x_i w_{ki}^1 + b_k^1 \right) w_{jk}^2 + b_j^2 \right),$$

onde m representa o número de neurônios na segunda camada oculta, w_{jk}^2 são os pesos conectando a primeira camada oculta à segunda camada e b_j^2 é o bias da segunda camada.

Esse processo continua até a camada de saída, garantindo que cada camada transforme os sinais recebidos e aplique a função de ativação correspondente. Mais detalhes sobre a saída dessas camadas podem ser encontrados em (HAYKIN, 1999).

2.1.1 Função de ligação

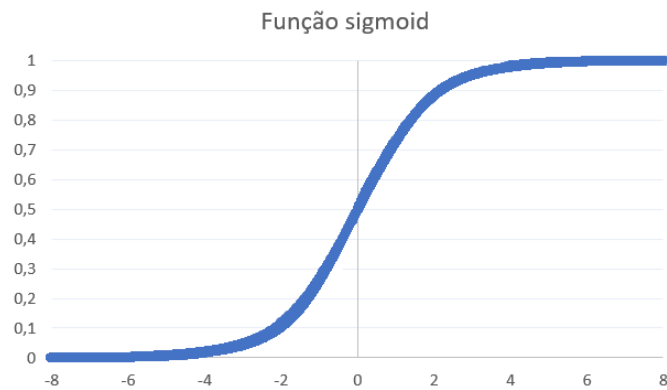
A seleção de uma função de ativação ou ligação específica resulta em diferentes tipos de neurônios. Existem diversas funções de ativação, que podem ser qualquer função que seja derivável (SHARMA; SHARMA; ATHAIYA, 2017). A seguir, são apresentadas algumas das funções de ativação mais comuns.

A função sigmoide (Figura 5), também conhecida como logística, é uma função de ativação não-linear frequentemente utilizada em Perceptrons Multicamadas. Esta é dada por:

$$\varphi(u) = \frac{1}{1 + e^{-\alpha u}}$$

em que $\alpha > 0$ é o parâmetro de inclinação da função sigmoide e $u \in \mathbb{R}$.

Figura 5 – Função de ativação sigmoidal.



Fonte: Elaborada pelo autor.

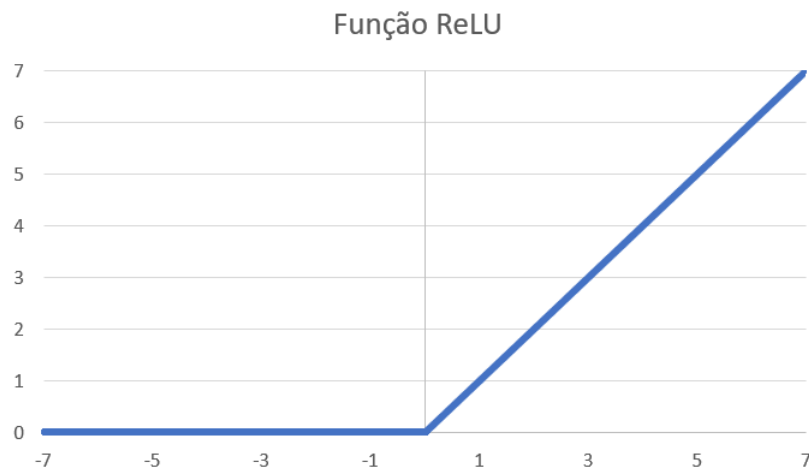
A função de ativação *Rectified Linear Unit* (ReLU) (Figura 6), apresentada por [Nair e Hinton \(2010\)](#), é amplamente utilizada em arquiteturas modernas de Redes Neurais. Ainda segundo o mesmo autor uma de suas principais vantagens é a ativação esparsa dos neurônios, pois aqueles que recebem entradas negativas não são ativados (*output zero*), tornando o processamento mais eficiente e acelerando o treinamento da rede.

A função ReLU é dada por:

$$\varphi(u) = \max(0, u).$$

onde, $u \in \mathbb{R}$.

Figura 6 – Função de ativação ReLU.



Fonte: Elaborada pelo autor.

Segundo [Nair e Hinton \(2010\)](#), é muito mais fácil computar como maior ou igual a 0 o

valor de entrada.

A função de ligação Softmax (Equação 2.1) é empregada em problemas de classificação. Esta aceita um vetor de entradas $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ e produz um vetor de probabilidades $p = (p_1, \dots, p_n) \in [0, 1]^n$.

$$p_j = \frac{e^{x_j}}{\sum_{j=1}^n e^{x_j}}, \quad (2.1)$$

em que, n é o número de classes e p_j é a probabilidade da imagem pertencer a classe j .

Outros exemplos de funções de ativação, tanto lineares quanto não lineares, podem ser encontrados em (HAYKIN, 1999).

2.1.2 Treinamento das ANN's

O treinamento de uma Rede Neural Artificial (ANN) tem como objetivo ajustar seus parâmetros para que um conjunto específico de entradas resulte em saídas desejadas ou, pelo menos, em previsões consistentes (BUENO *et al.*, 2006).

Esse processo envolve a determinação dos valores apropriados para os parâmetros da rede, com o objetivo de reduzir os erros cometidos durante as previsões, mais detalhes sobre esse processo podem ser vistos em Izbicki e Santos (2020). Para isso, utiliza-se a minimização de uma função de custo ou perda L .

No caso de problemas de classificação, a função de perda L é definida conforme a Equação 2.2, onde $y_j = 1$ se j corresponde à classe correta e $y_j = 0$ caso contrário:

$$L = - \sum_{j=1}^C y_j \log p_j. \quad (2.2)$$

Os valores ótimos para os pesos e o viés podem ser determinados por meio da minimização da função de custo. Para isso, um dos métodos mais utilizados é o gradiente descendente, que ajusta iterativamente os parâmetros da rede visando reduzir o erro. Para mais detalhes sobre esse processo, consulte (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.1.2.1 Backpropagation

O algoritmo de Retropropagação dos Erros (*Backpropagation*), inicialmente proposto por (NESTEROV, 1983), é o método muito utilizado para o treinamento de Redes Neurais Artificiais (ANNs). Segundo Haykin e Network (2004) esse algoritmo se fundamenta em uma regra de aprendizado que ajusta os pesos para reduzir o erro durante o treinamento, seguindo duas etapas principais: *forward* e *backward*.

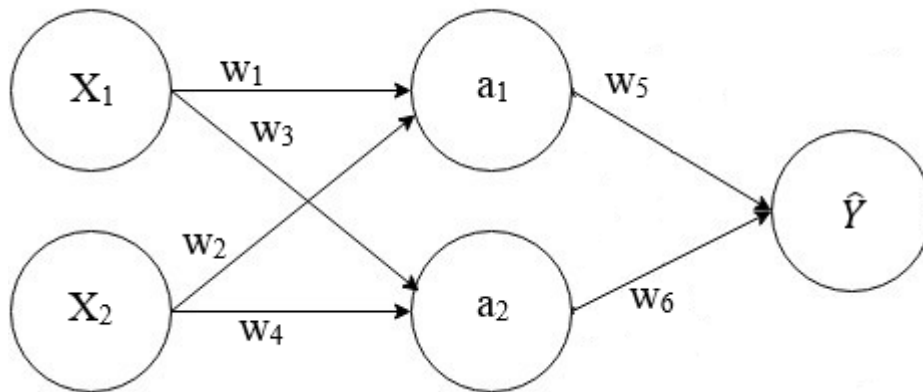
Durante a fase *forward*, a rede realiza o processamento de um padrão de entrada e determina sua saída associada. Em contraste, na fase *backward*, a rede emprega a discrepância entre a saída esperada e a saída obtida para ajustar os pesos das conexões (BUENO *et al.*, 2006).

Para que o *backpropagation* seja aplicado com eficácia, a função de ativação deve ser diferenciável, pois o ajuste dos pesos requer o cálculo da derivada parcial do erro em relação a cada peso w_i .

Para exemplificar o método, considere a Figura 7, na qual temos:

$$a_1 = g(x_1 w_1 + x_2 w_2 + b_1), \quad \hat{y} = g(a_1 w_5 + a_2 w_6 + b_2).$$

Figura 7 – Exemplo de uma ANN.



Fonte: Adaptado de Haykin (1999).

Na fase *forward*, calcula-se o valor estimado \hat{Y} , resultando em um erro que será retropropagado. Em seguida, utiliza-se a regra da cadeia para derivar a função de perda em relação ao peso w_5 :

$$\nabla_{w_5} = \left(\frac{1}{2} (a_1^2 - y)^2 \right)' = (a_1^2 - y) a^2 (1 - a_1^2) a_1^2$$

Esse mesmo procedimento pode ser seguido para calcular as derivadas em relação aos demais pesos, de forma análoga.

2.1.3 Método do Gradiente Adaptativo

Segundo Duchi, Hazan e Singer (2011), o método Gradiente Adaptativo (*Adagrad*) é um algoritmo baseado no gradiente descendente que ajusta a taxa de aprendizado de forma individual para cada parâmetro, com a fórmula:

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{S_t + \epsilon}} \frac{\partial L}{\partial w_t}$$

em que,

$$S_t = S_{t-1} + \left(\frac{\partial L}{\partial w_t} \right)^2,$$

onde w_t representa o valor atual do peso da rede no instante t , enquanto w_{t+1} é o novo valor atualizado; α é a taxa de aprendizado que controla o tamanho do passo na direção do gradiente, S_t é usado para adaptar a taxa de aprendizado de cada parâmetro individualmente; o denominador $\sqrt{S_t + \epsilon}$ ajusta a taxa de aprendizado com base na acumulação do gradiente, e ϵ é um pequeno valor positivo, geralmente 10^{-6} , inserido para evitar divisões por zero.

2.2 Redes Neurais Convolucionais

Conforme afirmado por [Goodfellow, Bengio e Courville \(2016\)](#), as Redes Neurais Convolucionais são uma classe de arquiteturas de Aprendizado Profundo que incluem a operação de convolução em pelo menos uma de suas camadas. Essas redes são propostas com objetivo de processar dados em formato de múltiplos vetores, sendo que um exemplo comum de sua aplicação é em imagens coloridas, uma vez que, estas são compostas por três vetores que representam a intensidade dos pixels em três canais de cores distintos.

As camadas convolucionais presentes na CNN's são responsáveis pela aplicação de filtros em toda a imagem para que haja a redução de sua complexidade. Desta forma, quando uma imagem é filtrada, a saída será outra imagem que pode ser entendida como um mapa de características que indicará se certas características foram detectadas na imagem de entrada, como afirma [Witten, Frank e Hall \(2017\)](#). Esta capacidade de extrair características relevantes através dos *kernels* é uma das vantagens de se utilizar as CNN's, além do fato de dependerem de um número menor de parâmetros de ajustes ao se comparar com as MLP com o mesmo número de camadas intermediárias. Estas vantagens são destacadas por [Araújo et al. \(2017\)](#).

Dentre as aplicações onde normalmente se empregam CNN's destaca-se o processamento de imagem. Segundo [Ponti e Costa \(2018\)](#), a operação de convolução possibilita a filtragem de imagens levando em conta sua estrutura bidimensional, sendo amplamente utilizada no processamento de informações visuais, especialmente em imagens. [Liu et al. \(2019\)](#) e [Shen, Wu e Suk \(2017\)](#) também reconhecem as CNN's como uma ferramenta poderosa para tarefas preditivas usando-as em imagens médicas. Além disso, estas redes têm se mostrado extremamente precisas para análise de séries temporais e classificação de imagens, fato afirmado por [Goodfellow, Bengio e Courville \(2016\)](#) e [Cao et al. \(2020\)](#).

Muitos são os trabalhos que utilizam CNN's para evidenciar aplicações, sendo que estas redes são bem-sucedidas em aplicações práticas, como destaca [Goodfellow, Bengio e Courville](#)

(2016). Um exemplo é o trabalho de Freitas (2018), que, diante da dificuldade na detecção de cáries em radiografias, desenvolveu um modelo de CNN para classificar cáries dentárias, alcançando uma taxa de acerto de 78,1%. Da mesma forma, Mendes *et al.* (2018) elaboraram três modelos de CNN para segmentar imagens de ultrassom de lesões mamárias, com o melhor modelo apresentando uma acurácia de 95,93%, visando auxiliar os profissionais da área.

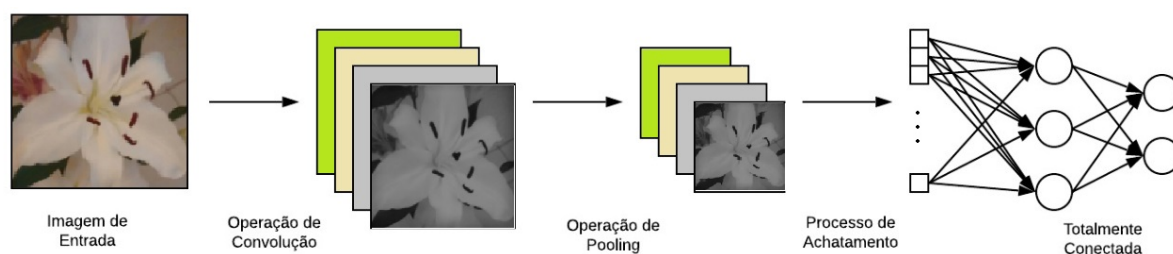
Outros exemplos recentes na área médica que utilizam CNNs para a detecção de câncer em imagens médicas incluem: câncer de cólon (HASAN *et al.*, 2024; KARTHIKEYAN; JOTHILAKSHMI; SUTHIR, 2024), câncer de pulmão (MAJUMDER *et al.*, 2024; LEE *et al.*, 2024), câncer gastrointestinal (WU *et al.*, 2024; AUZINE *et al.*, 2024), câncer de mama (MAHORO; AKHLOUFI, 2024; CHAKRAVARTHY *et al.*, 2024) e câncer de pele (KUMAR *et al.*, 2024; ALI *et al.*, 2024), entre outros.

2.2.1 Arquitetura geral

Uma Rede Convolutiva (Figura 8) é composta por camadas com diferentes funções. Conforme afirmado por Ebermam e Krohling (2016), a camada de convolução desempenha um papel crucial na extração de características da imagem, enquanto a camada de *pooling*, reduz a dimensão da imagem. Em seguida, a camada de achatamento transforma essas características em um formato adequado para uma rede totalmente conectada, que é responsável por interpretar as informações extraídas.

Deste modo, o processo de uma rede convolutiva acontece conforme ilustrado na Figura 8. Observe que primeiro uma imagem de entrada é compilada com filtros, reduzindo assim seu tamanho, fazendo com que uma nova matriz mostre os resultados das operações de convolução. Esses resultados são reduzidos por meio do *pooling*, que extrai os valores máximos de pequenas regiões de interesse (*max pooling*). Por fim o resultado ocorrerá a transmissão das características no processo de achatamento para a rede totalmente conectada.

Figura 8 – Arquitetura geral de uma CNN.



Fonte: Elaborada pelo autor.

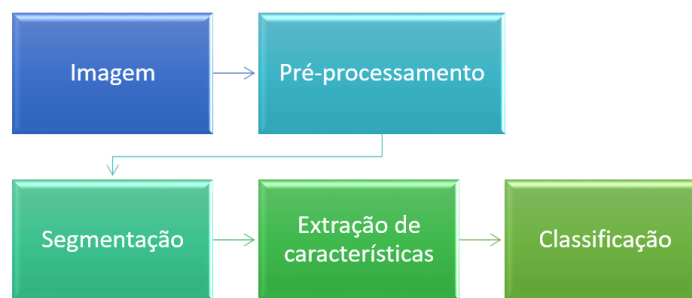
2.2.1.1 Processamento de imagens

A principal aplicação das Redes Neurais Convolucionais (CNN's) é o processamento de imagens, onde uma imagem é representada como uma matriz de *pixels*. Cada *pixel* possui

valores que refletem sua intensidade, variando de 0 a 255, o que corresponde à intensidade das cores nos *pixels*. Imagens coloridas, por sua vez, são representadas em formato de um array tridimensional.

A qualidade dos dados tem um impacto significativo nas técnicas de Aprendizado Profundo, podendo ser impactada por diferentes ruídos como, por exemplo, nomes nas imagens, bordas desnecessárias, dentre outros. Segundo [Rehman e Saba \(2014\)](#), o pré-processamento melhora a imagem real para uma análise posterior adequada. Essa fase tem como objetivo reduzir ou remover inconsistências na base de dados, tornando as informações mais apropriadas e simples para a análise posterior. Entre as atividades envolvidas no pré-processamento de dados estão a limpeza das imagens, o preenchimento, a alteração de resolução, cortes e outras operações.

Figura 9 – Fluxograma da classificação de imagens.



Fonte: Elaborada pelo autor.

As técnicas utilizadas para as análises de imagens tem como principal objetivo a extração de características da imagem para identificar padrões e obter uma melhor predição. A Figura 9 ilustra um fluxograma que descreve o processo de classificação de imagens. Note que a análise da imagem começa com a sua digitalização, seguida pela etapa de pré-processamento, cujo objetivo é aprimorar a qualidade da imagem digitalizada. Em seguida, realiza-se a segmentação, que envolve a identificação das regiões de interesse, cujas características serão extraídas. Por fim, ocorre a classificação das imagens.

2.2.1.2 Processo de convolução

Conforme mencionado por [Goodfellow, Bengio e Courville \(2016\)](#), a convolução é uma operação que envolve duas funções com argumentos de valores reais. Essa operação pode ser definida de duas maneiras: contínua e discreta.

Considere x e w como funções que mapeiam $\mathbb{R} \rightarrow \mathbb{R}$. A definição da convolução contínua é dada por:

$$(x * w)(t) = \int_{-\infty}^{\infty} x(a) w(t - a) da.$$

E a convolução discreta é dada por:

$$(x * w)(t) = \sum_{a=-\infty}^{\infty} x(a) w(t-a). \quad (2.3)$$

Onde o primeiro argumento x na operação de convolução representa a entrada e w refere-se ao o *kernel*. O resultado dessa operação é chamado de mapa de características, ou *feature map* (GOODFELLOW; BENGIO; COURVILLE, 2016).

O procedimento de convolução na equação (2.3), utilizando um *kernel* de bidimensiol K e uma imagem X pode ser expresso da seguinte forma:

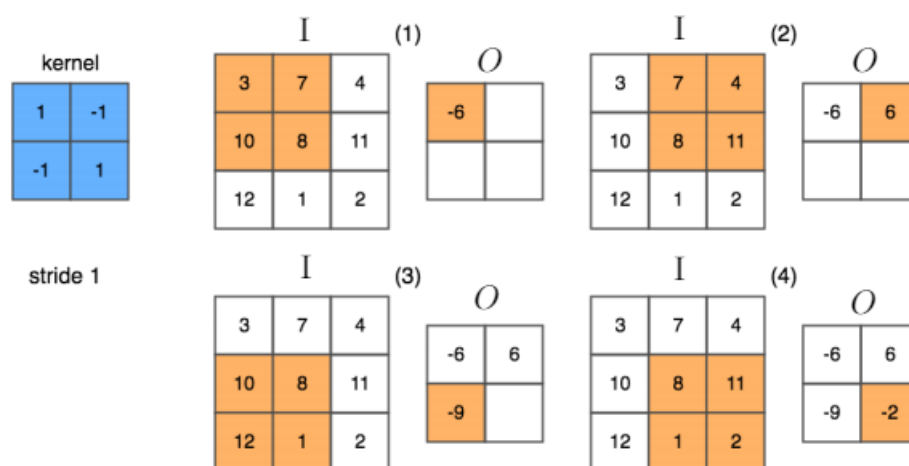
$$S(i, j) = (K * X)(i, j) = \sum_m \sum_n X(m, n) K(i-m, j-n), \quad (2.4)$$

onde, $S(i, j)$ representa a imagem de saída.

A operação de convolução em imagens pode ser entendida como o movimento do *kernel*/filtro sobre todas as posições da imagem, onde se calcula o produto entre o filtro e a imagem (FERREIRA, 2017). Um exemplo desse processo de convolução é ilustrado na Figura 10, que considera uma imagem de 3×3 e um filtro de 2×2 . Para o primeiro passo, a multiplicação do filtro com a imagem (matriz), utilizando a equação (2.4), é dada por:

$$O = (K * I)(i, j) = 3 - 7 - 10 + 8 = -6$$

Figura 10 – Exemplo do processo de convolução.



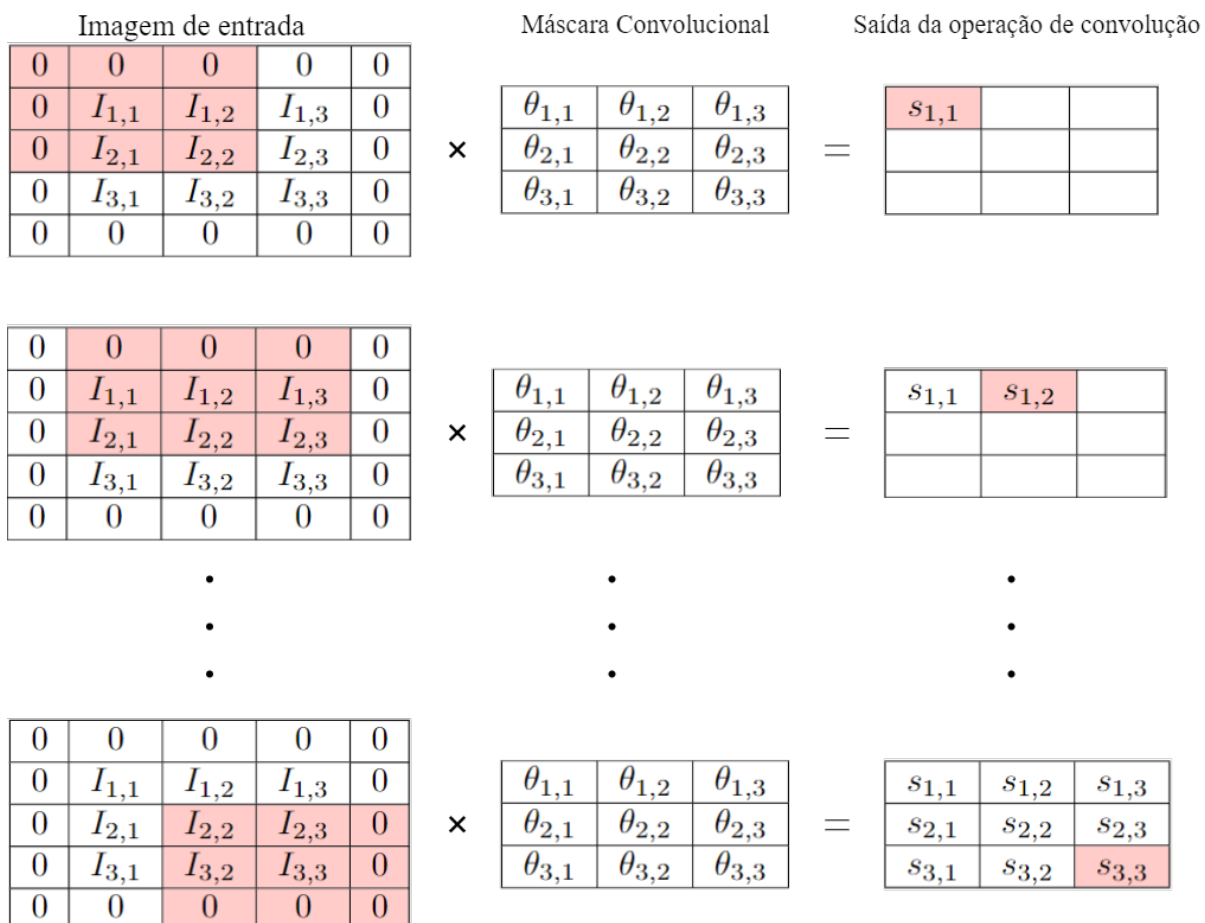
Fonte: Elaborada pelo autor.

Segundo Goodfellow, Bengio e Courville (2016) a saída do processo de convolução depende das variáveis: *padding* e o *stride*. O *stride* indica a quantidade de *pixels* que o filtro avança após cada operação; por exemplo, no caso da Figura 10, o filtro se deslocou 1 *pixel* a cada

operação, uma vez que o *stride* é igual a 1. Por outro lado, no *padding* são adicionados *pixels* simetricamente ao redor da imagem. Comumente, os *pixels* adicionados ao redor da imagem têm valor 0 (*0-padding*). Esse procedimento garante que as bordas da imagem original sejam incluídas na operação de convolução.

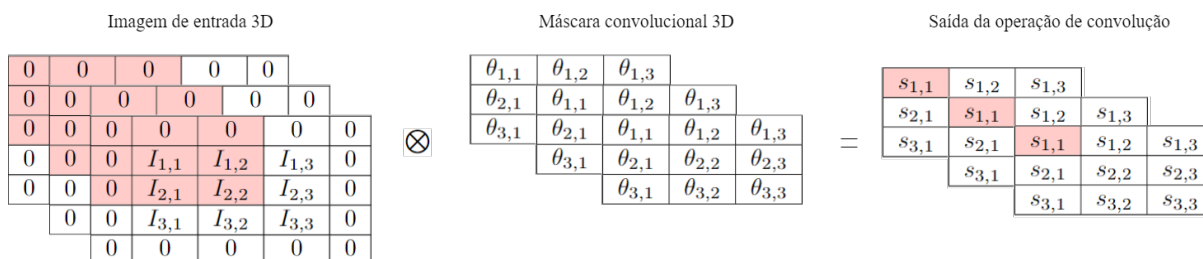
A Figura 11 demonstra o processo de convolução, utilizando uma imagem I de dimensão 3×3 com zero *padding*, juntamente com um filtro de convolução θ também de 3×3 , resultando em uma saída s de 3×3 . Por outro lado, a Figura 12 demonstra a aplicação da convolução em uma imagem que possui três canais.

Figura 11 – Processo de convolução, dado uma imagem I (3×3) com zero *padding* e uma máscara de convolução θ (3×3) que produz uma saída s (3×3).



Fonte: Adaptado de Nielsen (2015).

Figura 12 – Processo de convolução 3D, dado uma imagem I com zero *padding* e uma máscara de convolução θ que produz uma saída s .



Fonte: Adaptado de Nielsen (2015).

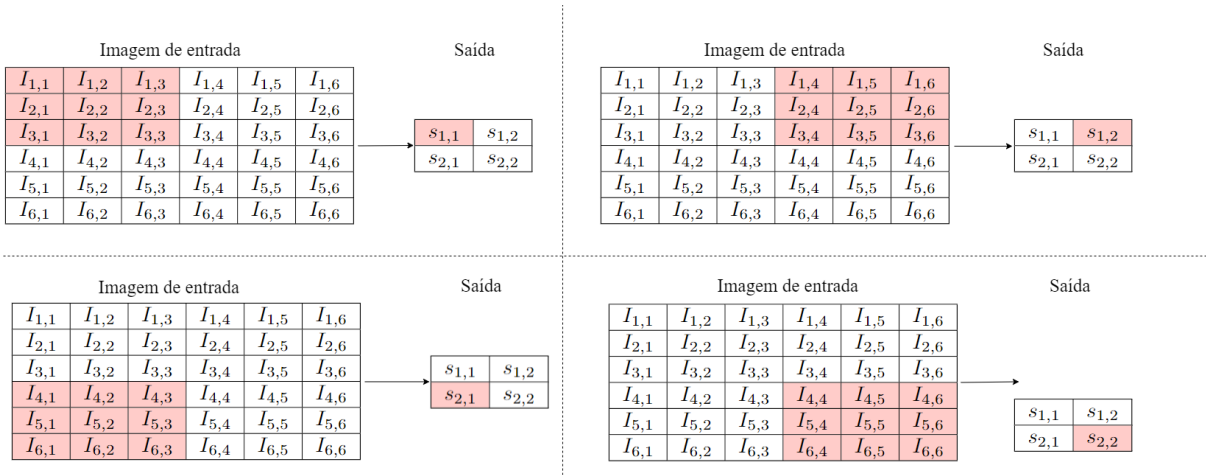
2.2.1.3 Pooling

De acordo com Goodfellow, Bengio e Courville (2016), uma camada comum em uma rede convolucional é composta por três etapas. Na primeira etapa, a camada realiza múltiplas convoluções simultaneamente, gerando um conjunto de ativações lineares. Na segunda etapa, cada uma dessas ativações é processada por uma função de ativação não linear, conhecida como estágio do detector. Por fim, na terceira etapa, uma função de *pooling* é aplicada para transformar ainda mais a saída da camada.

Normalmente, uma camada de convolução é seguida por uma camada de *pooling*. Essa camada tem como função diminuir a dimensionalidade dos mapas de características, substituindo a saída da rede em um local específico por uma estatística resumida das saídas adjacentes. Conforme mencionado por Ebermam e Krohling (2016), essa camada atua de maneira semelhante a uma camada convolutiva, embora o *stride* geralmente corresponda ao tamanho do filtro, resultando em uma redução de 75% nos mapas de características

A Figura 13 mostra o processo do *pooling* de tamanho 3×3 com um *stride* de mesmo tamanho, na qual a imagem inicial que era de tamanho 6×6 é reduzida a apenas 2×2 .

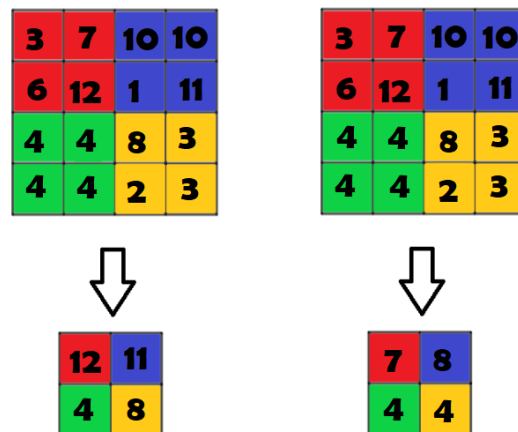
Figura 13 – Processo do *polling*, dado uma imagem I (6 × 6) e um determinado *polling* (3 × 3) que produz uma saída (2 × 2).



Fonte: Adaptado de Nielsen (2015).

Segundo Goodfellow, Bengio e Courville (2016), os filtros mais comuns utilizados na operação de *pooling* são o *pooling* máximo (ou *max pooling*) e o *pooling* médio (ou *average pooling*). No caso do *pooling* máximo, seleciona-se o maior valor entre os *pixels* que compõem o filtro, enquanto no *pooling* médio é calculada a média dos valores. A Figura 14 ilustra esses dois tipos de *pooling*, utilizando um filtro de tamanho 2 × 2. Na imagem à esquerda, podemos observar o *pooling* máximo, que é definido pelo valor máximo encontrado nos quadrados destacados em vermelho, ou seja, $\text{Máximo}(3, 7, 6, 12) = 12$. Por sua vez, o *pooling* médio é destacado na imagem da direita, no qual será determinado através da média dos valores dos quadrados vermelhos, sendo dado por: $(3 + 7 + 12 + 6)/4 = 7$.

Figura 14 – Exemplo de *pooling* máximo e médio.



Fonte: Elaborada pelo autor.

2.2.1.4 Achatamento

Depois de realizar as operações de convolução e *pooling* nas imagens de entrada, os resultados obtidos refletem as características extraídas dessas imagens. Em seguida, essas características são encaminhadas para uma camada de achatamento (ou *flatten*), que tem como objetivo preparar essas informações para a classificação das imagens.

Segundo [Ferreira \(2017\)](#), a camada de achatamento converte as características extraídas da imagem em uma estrutura adequada para uma Rede Neural Artificial, na qual todas as suas unidades estão conectadas à camada anterior.

2.2.2 Técnicas utilizadas com CNN's

2.2.2.1 Transferência de Aprendizado

A transferência de aprendizado é especialmente útil quando a conjunto de imagens disponível é menor do que o utilizado no treino de um modelo pré-treinado. Segundo [Araújo et al. \(2017\)](#), para aproveitar uma CNN já treinada como extrator de características, é necessário substituir a última camada do modelo. Esta camada é responsável por calcular e classificar a imagem de entrada em uma das classes que foram definidas.

Adicionalmente, para a aplicação dessa técnica, é essencial utilizar redes que tiveram bom desempenho nas tarefas para as quais foram projetados. Muitos modelos de Redes Neurais Convolucionais já desenvolvidas estão acessíveis em pacotes ou bibliotecas, permitindo sua utilização nesse procedimento. [Menegola et al. \(2017\)](#) ressaltam sobre transferência de aprendizado, sendo que esta técnica tem apresentado bons resultados e é uma alternativa muito vantajosa. Segundo [Thomas e Jerome \(2024\)](#), essa técnica de aprendizado de máquina nos permite aproveitar modelos pré-treinados para novas tarefas, sendo particularmente útil em tarefas de imagens médicas, onde dados rotulados frequentemente são escassos.

No contexto médico, a transferência de aprendizado tem sido aplicada em diversos modelos. Por exemplo, [Beghriche et al. \(2024\)](#) e [RS et al. \(2024\)](#) utilizaram essa abordagem para a detecção de câncer de mama. O primeiro estudo empregou as redes ResNet50, MobileNetV2, DenseNet121 e Xception, enquanto o segundo utilizou apenas a ResNet152v2. Em ambos os trabalhos, os modelos demonstraram bom desempenho ao aplicar essa estratégia.

2.2.2.2 Ensemble

Uma abordagem efetiva em Aprendizado de Máquina para melhorar o desempenho de um modelo é o uso do *Ensemble* ([FERREIRA, 2012](#)). Este pode ser definido como:

$$\Psi = \{h_1, h_2, h_3, \dots, h_L\},$$

onde h representa cada um dos L modelos que compõem o *ensemble* Ψ .

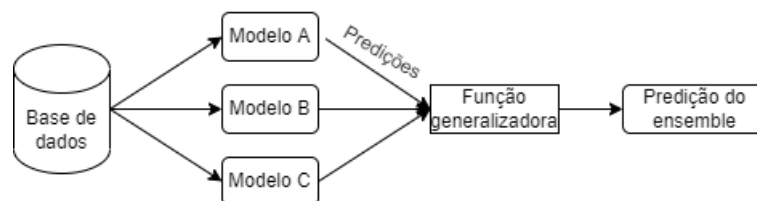
Ainda no conjunto Ψ , há um vetor de pesos $w = [w_1, \dots, w_L]^T$ que realiza a ponderação das saídas dos modelos, resultando na saída y_i^Ψ do *ensemble*, onde $w_j \geq 0$ e $\sum w = 1$. Assim, a saída do *ensemble* é uma combinação das saídas de h_j e w_j expressa pela seguinte fórmula:

$$y_i^\Psi = g(\Psi, x_i, w | \phi),$$

onde, ϕ representa o conjunto de parâmetros da função de combinação. Mais detalhes sobre esse método podem ser vistos em (LAPPALAINEN; MISKIN, 2000).

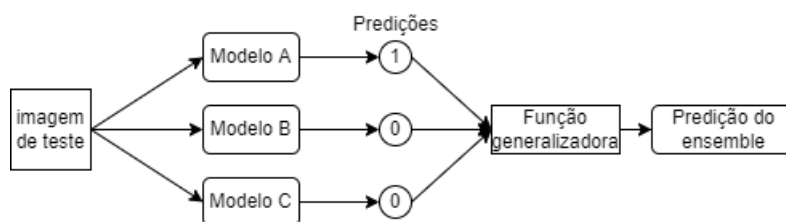
Entre as técnicas de ensemble, destacam-se a moda, a média, Bagging e Boosting. Para o processamento de imagens, a moda é frequentemente utilizada como uma função de generalização em problemas de classificação, na qual a previsão final do combinação corresponde à moda das saídas dos modelos individuais. Um exemplo dessa abordagem, com a combinação de modelos, é apresentado na Figura 16, onde há três modelos (A, B e C) com previsões de 1, 0 e 0 respectivamente. Assim, o resultado da combinação é dado por $\text{moda}(1, 0, 0) = 0$.

Figura 15 – Exemplo de um *ensemble* com três modelos individuais (A, B, C).



Fonte: Elaborada pelo autor.

Figura 16 – Exemplo de *Ensemble* - moda.



Fonte: Elaborada pelo autor.

Nos últimos anos, diversas pesquisas têm explorado técnicas de ensemble com o objetivo e reduzir a taxa de erro das Redes Neurais Convolucionais (CNNs). Por exemplo, o estudo de Krizhevsky, Sutskever e Hinton (2012) implementou um ensemble que combina cinco CNNs utilizando a média como função de generalização, alcançando o primeiro lugar no ILSVRC 2012. De maneira similar, Antipov, Berrani e Dugelay (2016) aplicaram um ensemble que utiliza a média como função generalizadora para o reconhecimento de gênero, obtendo o melhor desempenho ao combinar apenas três CNNs. Além disso, Minetto, Segundo e Sarkar (2019)

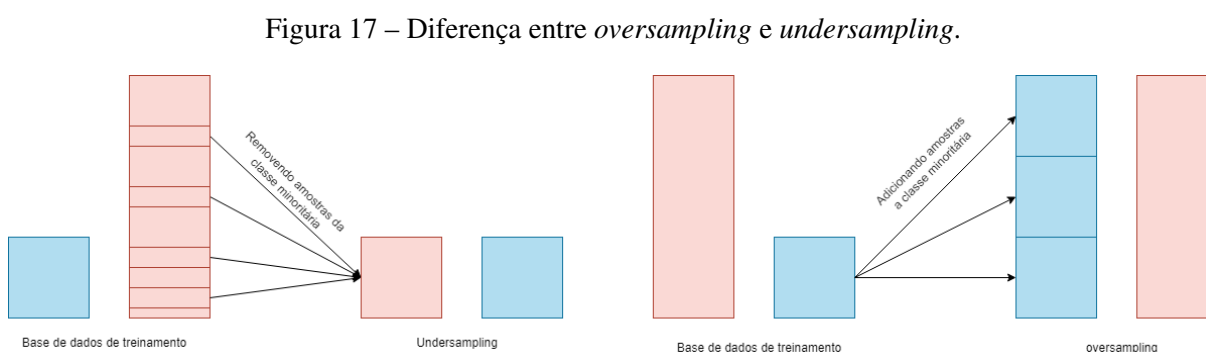
descreveram um ensemble de CNNs que emprega a moda para a classificação geoespacial de terras.

No contexto da aprendizagem por transferência, [Aneja e Aneja \(2019\)](#) empregou sete modelos para reconhecer dígitos manuscritos do alfabeto Devanagari, obtendo uma precisão de 98% com o melhor modelo. Por outro lado, [Rao et al. \(2024\)](#) combinaram os modelos Inception V3, VGG-19 e VGG-16 com Máquinas de Vetores de Suporte e analisaram seu desempenho em termos de acurácia, sensibilidade, especificidade e AUC. O melhor modelo alcançou uma acurácia de 95%.

2.2.2.3 Undersampling e Oversampling

Segundo [Shelke, Deshmukh e Shandilya \(2017\)](#), *oversampling* e *undersampling* são técnicas usadas para balancear a distribuição de classes de um conjunto de dados, estes são dois métodos principais de amostragem nos quais as amostras são reduzidos da classe majoritária ou amostras são adicionadas à classe minoritária. Conforme destacado por [He e Garcia \(2009\)](#), a aplicação de métodos de amostragem em contextos de aprendizado geralmente envolve a modificação de conjuntos de dados desbalanceados por meio de diversos mecanismos, com o objetivo de alcançar uma distribuição mais equilibrada.

A Figura 17 ilustra a diferença entre as duas técnicas *oversampling* and *undersampling*. Na base de dados gerada pelo *undersampling* a escolha das observações da classe majoritária que constituirá ela é feita de forma aleatória, no *oversampling* a classe minoritária é replicada até igualar a minoritária.



Fonte: Elaborada pelo autor.

Este processo pode ser replicado para base de dados de imagem. No *oversampling* basta replicar as imagens da classe minoritária até obter o mesmo número de imagens da classe majoritária, já para o *undersampling*, basta retirar aleatoriamente imagens da classe majoritária.

Diversos trabalhos têm explorado técnicas de *undersampling* e *oversampling* para balanceamento da base de dados nas CNN's: [Zhang, Wang e Liu \(2019\)](#), [Darwish, Ezzat e Hassanien](#)

(2020) e Buda, Maki e Mazurowski (2018), que implementam estratégias de reamostragem no conjunto de treinamento com o intuito de melhorar as previsões na classe minoritária.

Um trabalho mais recente, conduzido por Pereira e Saraiva (2021), investigou dois conjuntos desbalanceados de radiografias de mama e comparou os efeitos do *undersampling* e *oversampling* com a abordagem sem reamostragem. Os resultados evidenciaram que a aplicação dessas técnicas é uma estratégia eficaz para mitigar o impacto do desbalanceamento nos dados de treinamento.

2.2.3 Métricas de avaliação

Para avaliar o desempenho do modelo em uma tarefa de classificação supervisionada, são utilizadas diversas métricas derivadas da matriz de confusão, apresentada na Tabela 1. Essa matriz permite analisar a qualidade das previsões do modelo, considerando acertos e erros em cada classe.

Tabela 1 – Matriz de confusão 2×2 .

| | Valor Verdadeiro | |
|---------------|------------------|-----|
| | Sim | Não |
| Valor Predito | VP | FP |
| | FN | VN |

Em que VP, FP, VN e FN representam os Verdadeiros Positivos, Falsos Positivos, Verdadeiros Negativos e Falsos Negativos, respectivamente.

A matriz de confusão serve para avaliar o desempenho em problemas de classificação e pode ser ampliada para lidar com mais classes.

Para avaliar a adequação do modelo em problemas de classificação, utilizam-se diversas métricas de desempenho fundamentadas na matriz de confusão. Algumas medidas de avaliação são apresentadas a seguir: recall (R), precisão (P), acurácia (ACC), valor preditivo negativo (NPV), eficiência (E), especificidade (S) e F beta score que são apresentadas na Tabela 2. Mais detalhes sobre essas métricas podem ser vistos em Heyburn *et al.* (2018) e Idrees *et al.* (2017).

Conforme destacado por Menardi e Torelli (2014) e Jeni, Cohn e Torre (2013), a acurácia pode ser comprometida em situações onde as categorias estão desbalanceadas, tornando outras métricas, como recall e precisão, mais eficazes. Segundo Jeni, Cohn e Torre (2013), a precisão é a proporção de exemplos classificados como positivos que estão corretos, enquanto o recall avalia quantos exemplos da classe positiva foram corretamente identificados. Além dessas métricas que se concentram na classe positiva, também existem aquelas que avaliam a classe negativa. De acordo com Matta (2021), o valor preditivo negativo representa a proporção de verdadeiros negativos em relação a todas as previsões negativas, e a especificidade é a razão entre os verdadeiros negativos e todas as observações realmente negativas. O F-Beta score, que

é uma média harmônica entre recall e precisão, é especialmente útil em conjuntos de dados desbalanceados (LEE; YANG; YOO, 2021). Conforme afirmado por Alaiz-Moreton *et al.* (2019), o parâmetro β ajusta o peso da precisão na métrica combinada: valores de $\beta < 1$ favorecem a precisão, enquanto valores de $\beta > 1$ priorizam o recall. Por fim, segundo Zaki, Jr e Meira (2014), a eficiência é calculada como a média aritmética entre recall e especificidade.

Tabela 2 – Medidas de avaliação.

| Medidas | Fórmula |
|----------------|---|
| Acurácia | $\frac{VP + VN}{VP + VN + FP + FN}$ |
| Recall | $\frac{VP}{VP + FN}$ |
| Precisão | $\frac{VP}{VP + FP}$ |
| Especificidade | $\frac{VN}{FP + VN}$ |
| NPV | $\frac{VN}{VN + FN}$ |
| Eficiência | $\frac{S + R}{2}$ |
| F Beta Score | $\frac{(1 + \beta^2) * P * R}{\beta^2 * P + R}$ |

2.2.4 Métodos de evitar overfitting

Ao ajustar um modelo, podem ocorrer três cenários distintos: o modelo pode apresentar um ótimo ajuste aos dados de treinamento, mas ter um desempenho insatisfatório nos dados de teste (superajuste ou *overfitting*), não se adaptar adequadamente aos dados de treinamento (subajuste ou *underfitting*), ou ainda alcançar um ajuste razoável aos dados de treinamento e demonstrar boa previsão nos dados de teste.

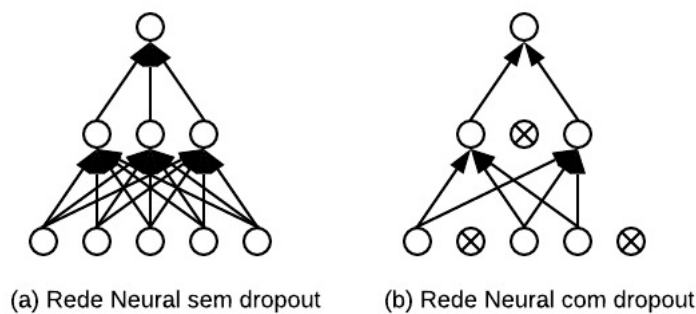
Segundo Goodfellow, Bengio e Courville (2016), há dois desafios centrais no Aprendizado de Máquina: *underfitting* e *overfitting*. Há diversos modos de evitar tais desafios, como por exemplo: Aumentar a quantidade de dados do conjunto treinamento para reduzir o *overfitting*, *Dropout*, *Early stopping*, dentre outros. Algumas dessas técnicas se baseiam na função perda.

2.2.4.1 Dropout

Segundo Ebermam e Krohling (2016) a técnica de *Dropout* é amplamente utilizada para evitar o *overfitting* em redes neurais, especialmente em cenários onde o modelo pode se ajustar excessivamente aos dados de treinamento, perdendo a capacidade de generalizar para novos dados.

Como apresentado por [Srivastava et al. \(2014\)](#), o *Dropout* (Figura 18) consiste em desativar aleatoriamente um conjunto de neurônios durante o treinamento, com uma probabilidade p . Isso significa que, para cada iteração, alguns neurônios são "removidos" da rede, enquanto os restantes continuam ativos com probabilidade $(1 - p)$. Após essa remoção temporária de unidades, o modelo é ajustado normalmente com o conjunto reduzido de neurônios, melhorando sua capacidade de generalização.

Figura 18 – Exemplo de uma ANN com e sem *dropout*, onde n observa-se uma taxa de dropout de 40%, resultando na remoção de 2 neurônios de um total de 5.



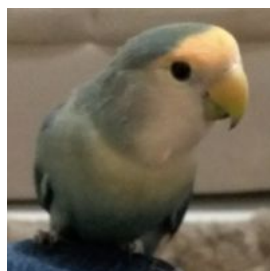
Fonte: Elaborada pelo autor.

2.2.4.2 Aumento de Dados

Segundo [Tanaka e Aranha \(2019\)](#), ao analisar informações médicas algoritmos de classificação pode ter dificuldades em identificar as classes minoritárias, podendo classificar todas as observações minoritárias como erradas. Ainda segundo o mesmo autor, para evitar este problema é possível aumentar os dados através da criação de novas dados.

Considere a Figura 19, que ilustra a criação de novas imagens, as quais são apresentadas na Figura 20. Essas novas imagens foram geradas a partir de rotações verticais e horizontais, baseadas em princípios de geometria, além de variações nos tons de cores e inclusão de ruídos. Variações sobre aumento de dados podem ser vistas em [Wang, Perez et al. \(2017\)](#).

Figura 19 – Imagem original.



Fonte: Elaborada pelo autor.

Figura 20 – Transformações da imagem na Figura 19.



Fonte: Elaborada pelo autor.

Com base nisso, podemos concluir que, ao aplicar essas mesmas transformações a um conjunto de 20 imagens, obteríamos uma base de dados aumentada contendo 240 imagens.

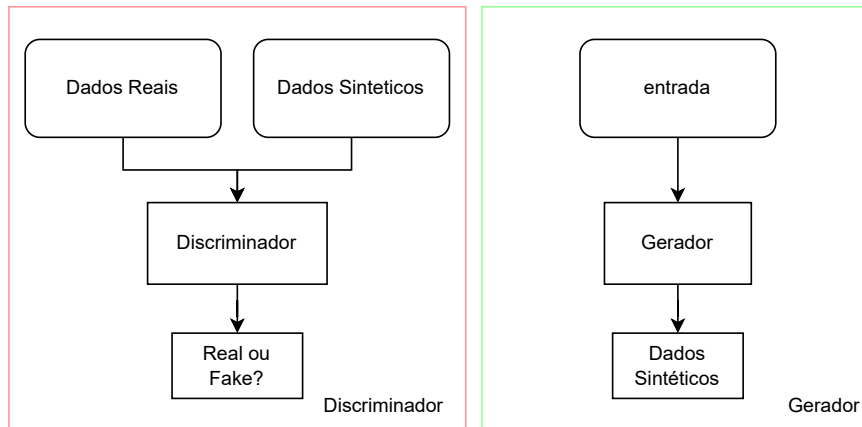
2.2.4.3 Aumento de dados por rede adversária generativa

Outro método utilizado para aumento de dados é a Rede Adversária Generativa (do inglês *Generative Adversarial Network (GAN)*) que têm sido amplamente utilizados em visão computacional para aumento de dados (MIRZA; OSINDERO, 2014). A primeira rede Adversarial Generativa foi proposta por Goodfellow *et al.* (2014) como estrutura para estimar modelos generativos. De acordo com Huang *et al.* (2020) essas redes não exigem que os dados sigam quaisquer suposições específicas ao modelar dados complexos, mesmo com algumas distribuições implícitas, sendo naturalmente adequadas para a tarefa de geração de dados.

Vários autores pesquisam o uso de GAN's para aumento de dados, Dai *et al.* (2022) e Lee e Park (2021) usam aumento de dados baseado em GAN's para gerar amostras de dados artificiais para classes menores. Ali-Gombe e Elyan (2019) propõe uma nova abordagem de aumento de dados usando um GAN para lidar com o problema de desequilíbrio de classes.

Segundo Reis (2020) GANs são compostos por uma combinação de duas redes distintas. As redes que compõem uma GAN são o gerador e o discriminador. O gerador é a rede responsável para capturar uma determinada distribuição de dados suficientemente bem para ser capaz de gerar novos dados após tal distribuição. O discriminador é uma rede responsável por detectar e dados falsos, sendo os dados reais uma amostra do conjunto de dados, e os falsos sintetizados por o gerador. Na Figura 21 é apresentada esta relação o retângulo vermelho representa a rede discriminadora e a verde a geradora.

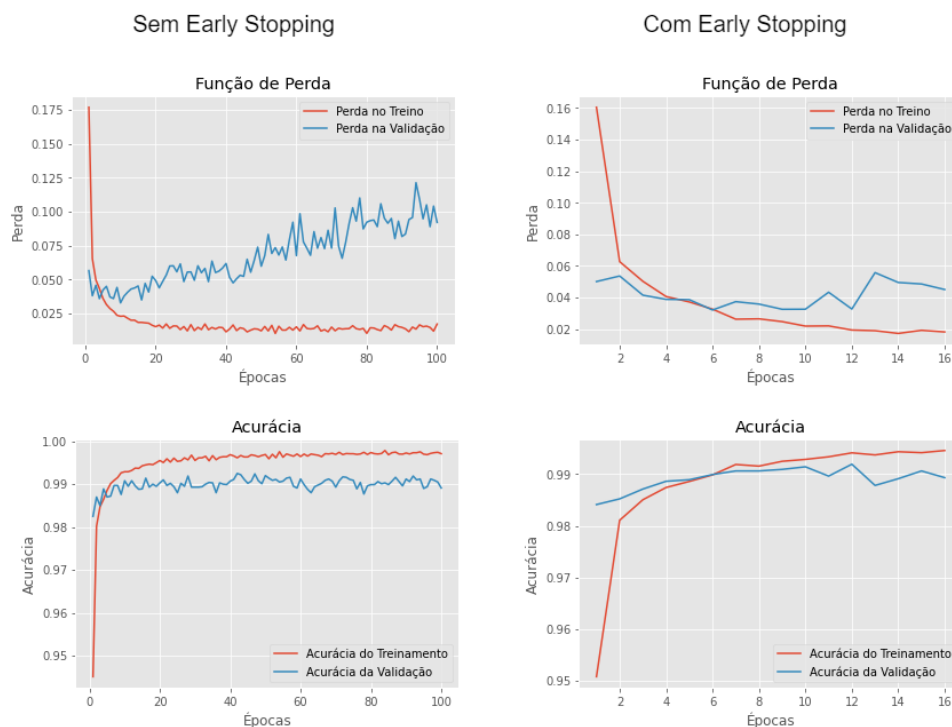
Figura 21 – Fluxograma do funcionamento da GAN.



Fonte: Elaborada pelo autor.

2.2.4.4 Early stopping

O Early stopping é uma forma de regularização baseada na escolha de quando parar de executar um algoritmo iterativo (RASKUTTI; WAINWRIGHT; YU, 2014). Nas redes neurais, no final de cada época, as métricas são calculadas nos dados de validação, quando o modelo para de melhorar em um número determinado de épocas, o treinamento é finalizado.

Figura 22 – Exemplo da utilização do método de *early stopping*

Fonte: Elaborada pelo autor.

A Figura 22 ilustra um exemplo da aplicação do *early stopping*, os dois modelos tem a mesma estrutura e foram treinados na mesma base de dados, diferindo apenas na utilização do *early stopping*. A esquerda está um modelo que não teve a aplicação do *early stopping* e treinou por 100 épocas, a direita o modelo faz a utilização da técnica e treina por apenas 16 épocas, o *early stopping* foi configurado para parar de treinar caso o modelo não melhorasse em até 10 épocas. No primeiro modelo é possível notar que conforme as épocas passam o modelo melhora nos dados de treino na função de perda e na acurácia, mas na validação esse padrão não é visto, o que não acontece no *early stopping* pois o modelo é encerrado.

2.3 Conclusão

Neste capítulo foi apresentada uma revisão bibliográfica necessária para o desenvolvimento desta tese, contendo as principais definições e referências utilizadas.

- Na Seção 2.1 foi mostrada a ideia da RNA. O conceito das funções de ligação, método de treinamento e algumas funções de perda.
- Na Seção 2.2 foi exibido um breve estudo sobre CNN, focando a sua arquitetura geral e em seguida algumas técnicas que podem ser utilizadas com essas redes.

MODELO PROPOSTO

Neste capítulo é proposto o modelo do estudo que é composto por uma combinação de modelos de transferência de aprendizado.

3.1 Modelos de transferência de aprendizado

Os modelos de transferência de aprendizado utilizados neste trabalho foram: VGG16 (SIMONYAN; ZISSERMAN, 2014), Densenet 121, Densenet 169, Densenet 201 (HUANG *et al.*, 2017), Inception Resnetv2 (SZEGEDY *et al.*, 2017), Resnet 50, Resnet 101, Resnet 152 (HE *et al.*, 2016b), Resnet 50 v2, Resnet 101 v2, Resnet 152 v2 (HE *et al.*, 2016a), Xception (CHOLLET, 2017) e MobileNet (HOWARD *et al.*, 2017). Esses modelos estão no pacote Keras Chollet *et al.* (2015) no python e maiores informações podem ser encontradas em Sarkar, Bali e Ghosh (2018).

Os modelos utilizados já obtiveram boas performances em classificar imagens de grandes bases de dados. A VGG-16 (*Visual Geometry Group architecture with 16 layer*), por exemplo, é uma CNN que foi publicada em 2015 e foi treinada no banco de dados *ImageNet* (base de dados com aproximadamente 14 milhões de imagens) e, segundo Deng *et al.* (2009) a partir do *ImageNet* surgiu o *ImageNet Large Scale Visual Recognition Competition (ILSVRC)* que tem como objetivo medir o progresso da área de análise de imagem e visão computacional. A VGG-16 tem uma arquitetura com filtros de convolução muito pequenos de tamanho 3×3 , que mostram melhora significativa em relação aos modelos propostos sem essa arquitetura (THECKEDATH; SEDAMKAR, 2020). Ainda segundo Theckedath e Sedamkar (2020), a VGG16 dá excelentes precisões mesmo quando os conjuntos de dados de imagem são pequenos.

A Densenet também foi utilizada para predição no *ImageNet* e segundo Wang e Zhang (2020) oferece concatenações de todos os mapas de recursos de camadas, o que significa que todos os mapas de recursos se propagam para as camadas posteriores e se conectam às novas

mapas de recursos gerados. Cada versão DenseNet (121, 169 e 201) terá diferentes números da unidade repetida dos seus sub blocos (ABDANI; ZULKIFLEY, 2019).

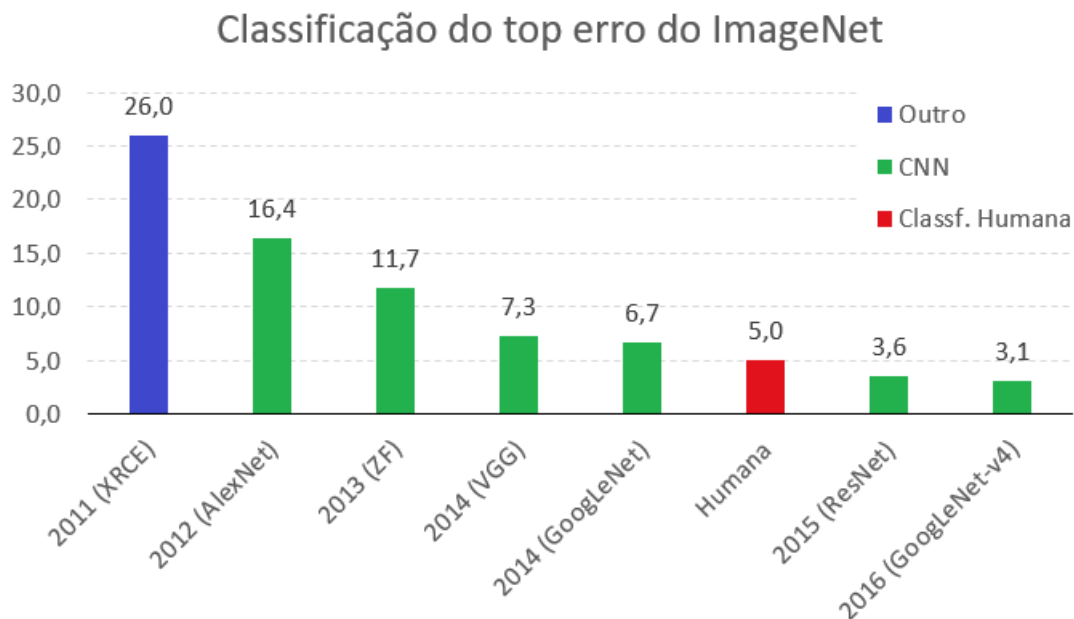
Já para as redes residuais (Resnet) é composta por blocos residuais (HE *et al.*, 2016b). A Resnet pode ser nomeada pelo seu nível de profundidade, Resnet 50 possui 50 camadas, Resnet 101 possui 101 camadas e Resnet 152 possui 152 camadas. Segundo Pan *et al.* (2021) as ResNet V2, são versões modificadas das ResNet com desempenho melhores no conjunto de dados *ImageNet*.

A inception Resnet V2 também treinada com a base de dados do Imagenet e tem uma arquitetura que utiliza os truques e decisões de uma rede inception com variantes de conexão residual (FERREIRA *et al.*, 2018), que é baseada na conexão residual da resnet.

Kassani *et al.* (2019) afirmam que a Xception é uma extensão da arquitetura Inception, também utilizando conexões residuais e a MobileNet consiste em camadas de convolução separáveis. Uma alta taxa de precisão pode ser alcançada na arquitetura MobileNet com um pequeno número de hiperparâmetros.

Todos esses modelos demonstraram um bom desempenho na classificação da base de dados *ImageNet*. As redes neurais convolucionais (CNNs) ganharam grande popularidade após o Desafio de Reconhecimento Visual em Grande Escala do *ImageNet*, que tinha como objetivo avaliar o avanço na área de análise de imagens.

Figura 23 – Taxa de erro (%) dos modelos do ILSVRC.



Fonte: Elaborada pelo autor.

A Figura 23 ilustra a taxa de erro na classificação dos modelos ganhadores do ILSVRC desde 2011, incluindo a classificação humana, sendo que apenas o primeiro modelo não é uma

rede neural convolucional. O objetivo do ILSVRC era identificar o melhor modelo para classificar a base de dados *ImageNet* ou um subconjunto dessa base. Após 2014, os modelos que venceram o ILSVRC apresentaram taxas de acerto superiores às dos humanos na base de dados *ImageNet*. Essa figura evidencia a eficácia das CNNs como modelos para a classificação de imagens.

3.2 Ensembles Propostos

Os *ensembles* utilizados neste trabalho consideram a combinação de 13 modelos de transferência de aprendizado. As combinações propostas aplicam peso à classe minoritária. Na primeira combinação foi considerada: se pelo menos um dos modelos classificar a observação como a classe minoritária, então a predição da observação será igual à classe minoritária. Consequentemente as outras combinações são: Combinação i : Se pelo menos i modelos classificarem como a classe com menos observações, então a predição será considerada 1, $i = 1, \dots, 13$, considerando que há apenas duas classes (0 e 1) e a classe 1 possui menos observações.

Matematicamente sejam M_1, M_2, \dots, M_{13} os modelos de classificação utilizados, cada um produzindo uma previsão binária $M_j(\mathbf{x}) \in \{0, 1\}$, onde \mathbf{x} é a observação que está sendo classificada. A classe 1 é a classe minoritária. Para uma combinação i , a previsão do conjunto $P_i(\mathbf{x})$ é definida como:

$$P_i(\mathbf{x}) = \begin{cases} 1, & \text{se } \sum_{j=1}^{13} M_j(x) \geq i, \\ 0, & \text{caso contrário,} \end{cases}$$

para $i = 1, 2, \dots, 13$. Assim, a decisão da combinação segue um esquema de votação majoritária onde um limite i é definido. Se a soma cumulativa das previsões do modelo atingir ou exceder i , então $P_i(\mathbf{x}) = 1$, caso contrário, $P_i(\mathbf{x}) = 0$.

Tabela 3 – Modelos e suas predições

| Modelos | Predições | Limiar (i) | Predição Final $P_i(\mathbf{x})$ |
|--------------------|-----------|------------|----------------------------------|
| VGG-16 | 0 | 1 | 1 |
| DenseNet 121 | 1 | 2 | 1 |
| DenseNet 169 | 0 | 3 | 1 |
| DenseNet 201 | 1 | 4 | 1 |
| Inception ResNetv2 | 0 | 5 | 1 |
| ResNet 50 | 1 | 6 | 1 |
| ResNet 101 | 0 | 7 | 0 |
| ResNet 152 | 1 | 8 | 0 |
| ResNet 50 v2 | 0 | 9 | 0 |
| ResNet 101 v2 | 1 | 10 | 0 |
| ResNet 152 v2 | 0 | 11 | 0 |
| Xception | 1 | 12 | 0 |
| MobileNet | 0 | 13 | 0 |

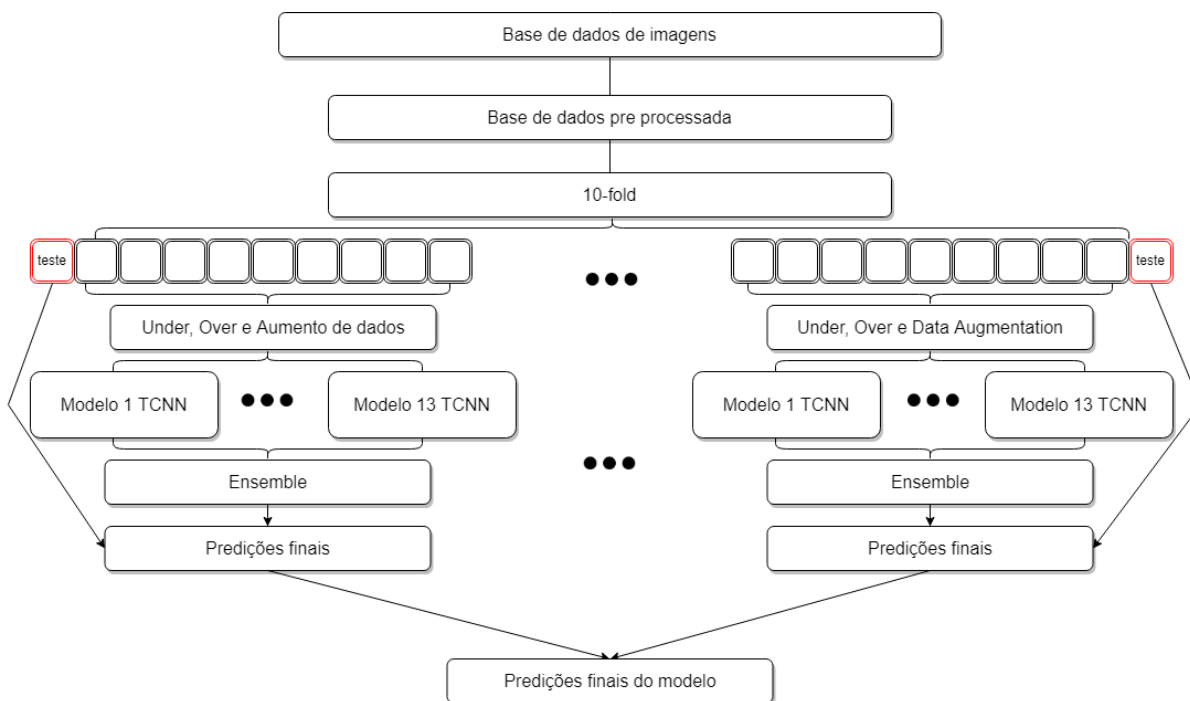
Por exemplo, na Tabela 3, seis modelos preveem 1 (DenseNet121 e 201, ResNet50, 152 e 101v2 e Xception). Se o limite for $i = 6$ (Combinação 6), então $P_6(\mathbf{x}) = 1$; no entanto, se o limite for $i = 7$ (Combinação 7), então $P_7(\mathbf{x}) = 0$.

3.3 Modelo final proposto

Neste estudo, utilizamos ETCNN com diferentes pesos (considerando o limiar de 1 a 13), reamostragem e aumento de dados para melhorar o desempenho das CNN's em pequenas bases de dados desbalanceados. Para isso, combinamos 13 modelos de transferência de aprendizado com 13 combinações.

Este trabalho se diferencia de pesquisas anteriores por sua abordagem voltada especificamente para bases de imagens limitadas, um cenário desafiador na detecção de câncer por meio de aprendizado profundo. Diferente de estudos que utilizam grandes bases de dados ou modelos individuais, nossa proposta emprega um *ensemble* de modelos de transferência de aprendizado, combinando diferentes arquiteturas para aprimorar a detecção de casos na classe minoritária. Além disso, exploramos de maneira sistemática o impacto do aumento de dados e da reamostragem no desempenho do modelo, buscando melhorar a generalização sem comprometer a robustez das predições. Assim, nossa contribuição reside na integração dessas estratégias para otimizar a detecção de câncer em contextos de dados restritos, oferecendo uma solução mais eficaz para bases médicas com disponibilidade limitada e desbalanceada de imagens.

Figura 24 – Representação da modelagem proposta e passo a passo, veja que os pontos importantes aumento de dados, *undersampling* e *oversampling* estão aplicados após a divisão do treinamento e teste, os 13 modelos são treinados com esses conjuntos e em seguida as 13 combinações são realizadas que geram a predição final de cada um dos *k-fold*'s.



Fonte: Elaborada pelo autor.

O fluxograma da metodologia proposta deste estudo é mostrado na Figura 24. Como pode ser visto, toda a metodologia é dividida principalmente em cinco etapas: Pré-processamento

de dados que é o mesmo discutido na seção 2.2.1.1, divisão dos *fold's*, aumento de dados, *oversampling* e *undersampling*, extração dos resultados das 13 CNN's selecionadas, classificação por meio do *ensemble* e, por finalmente, a avaliação do modelo.

Observe que a aplicação do aumento de dados, *undersampling* e *oversampling* é feita apenas no conjunto de treinamento de cada um dos *fold's*.

O método de validação utilizado é o K-fold, onde a base de dados é dividida em k partes uma dessas é usada para teste e as demais para treino (ANGUITA *et al.*, 2012). Neste trabalho utilizamos um 10-fold, então um fold é usado para teste e os outros 9 para treino esse processo é repetido 10 vezes.

3.4 Conclusão

Neste capítulo foi introduzido o método em estudo, no qual, o método proposto são 13 combinações de 13 modelos de transferência de aprendizado, a Seção 3.1 apresenta os modelos de transferência utilizados, a Seção 3.2 mostra as combinações utilizadas e, por fim, a última Seção apresenta a estrutura que será considerada na modelagem.

APLICAÇÕES

Neste capítulo é apresentada uma aplicação da modelagem proposta a um conjunto de dados com imagens de mamografias com câncer de mama (4.1) e imagens de câncer de pele (4.2). Todos os códigos para as análises estão disponíveis em <https://github.com/FernandoMoraes>.

4.1 Câncer de Mama

Com o objetivo de verificar a performance da modelagem proposta foi empregada uma pequena base de dados desbalanceada disponibilizada pelo hospital oncológico A.C. Camargo Cancer Center. Esta base de dados nunca foi utilizada para nenhum tipo de análise e tem como finalidade classificar imagens de mamografia como tumor maligno ou benigno. Um dos objetivos da aplicação é criar um modelo que auxilie os radiologistas a classificarem novas imagens.

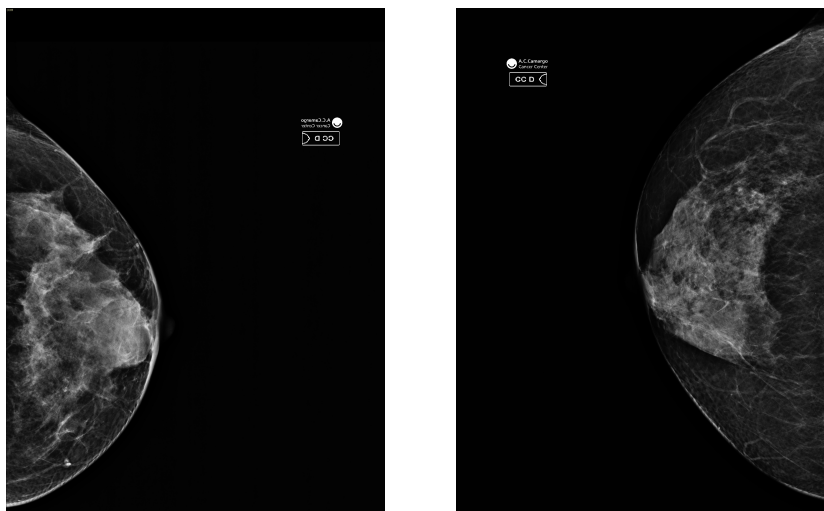
Nesta aplicação, utilizamos 13 combinações de modelos focados na classificação binária (Maligno ou Benigno) para exames de mamografia com a base de dados desbalanceada. Para verificar se as combinações utilizadas eram eficazes, comparamos o resultado das combinações com o resultado individual de cada um dos 13 modelos utilizados. Os cenários considerados são: (I) sem técnica de reamostragem (com e sem aumento de dados); (II) com a utilização de *undersampling* para balanceamento (com e sem aumento de dados); e (III) com *oversampling* para balanceamento (com e sem aumento de dados). Todas as análises dessa aplicação foram realizadas na GPU da plataforma Google Colab, utilizando a linguagem Python.

4.1.1 Base de dados

A base de dados é composta por 414 imagens de mamografias digitalizadas, com resoluções variando entre 3400 a 4800 *pixels* em altura e largura. Essas imagens estão divididas em duas classes: 91 imagens de tumor maligno e 323 tumor benigno. A distribuição de frequências

por tipo de classificação das imagens é apresentada na Tabela 4 e um exemplo delas é apresentado na Figura 25. A base de dados apresenta ainda características referentes as imagens como idade, densidade mamaria, tipo de calcificações, tamanho, profundidade e método utilizado.

Figura 25 – Exemplos de imagens da base de dados de mamografia.



(a) Benigno

(b) Maligno

Fonte: Base de dados do A.C. Camargo Cancer Center.

Tabela 4 – Número de elementos pertencentes a cada classe na base de dados de mamografia.

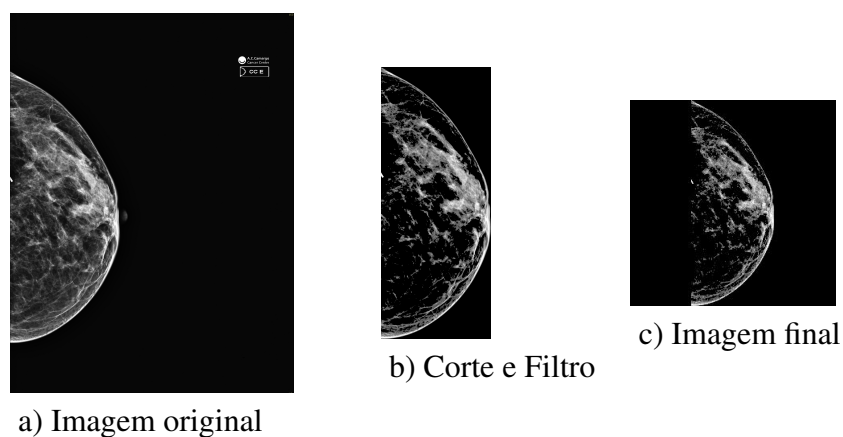
| Classe | Número de elementos |
|---------|---------------------|
| Maligno | 91 |
| Benigno | 323 |
| Total | 414 |

As imagens originais da base de dados (Figura 25) apresentavam algum ruído e um tamanho excessivo. Portanto, foi necessário realizar um corte para isolar apenas a região da mama, garantindo que o foco da análise estivesse na área de interesse para a classificação. Posteriormente, foi aplicado um filtro para eliminar as partes menos relevantes para a análise. Por fim, a imagem final foi redimensionada para ter dimensões 256×256 . A Figura 26 apresenta um exemplo desse pré-processamento a imagem original tinha aproximadamente 3500×4700 , com o corte o tamanho diminuiu para aproximadamente 1200×3500 e a imagem final tem tamanho 256×256 .

Com o objetivo de obter uma bom ajuste do modelo utilizou-se neste estudo o aumento de dados no conjunto de treinamento. Os dados foram aumentados com os métodos tradicionais de aumento de dados. Algumas imagens aumentadas utilizando a imagem final da Figura 26 estão apresentadas na Figura 27, o aumento de dados utilizados incluem rotação, espelhamento, corte. O aumento de dados foi utilizado com o objetivo de balanceamento então foram geradas

imagens até obtermos o mesmo número de imagens em ambas as classes. Ainda, neste estudo foram considerados em cada um dos cenários a não utilização de aumento de dados, ao final uma comparação entre a utilização e a não utilização do aumento é realizada.

Figura 26 – Pré-processamento da imagem.



Fonte: Adaptado da base de dados do A.C.Camargo Cancer Center.

Figura 27 – Aumento de dados com a utilização da imagem final na Figura 26.



Fonte: Adaptado da base de dados do A.C.Camargo Cancer Center.

Nesta aplicação foram utilizados o *undersampling* e *oversampling* no conjunto de treinamento com o objetivo de balancear a base de dados sob estudo e comparar as metodologias no caso com e sem reamostragem, esses processos são feitos de forma aleatória. Um exemplo do processo de reamostragem é apresentado na Tabela 5, apenas para o primeiro *fold* do modelo. Foi usado o *k-fold* com 10 *fold's* para a validação do modelo.

Tabela 5 – Número de elementos pertencentes a cada classe na base de dados de mamografia com a utilização do *undersampling* e do *oversampling* no primeiro *fold*.

| Classe | Original | <i>undersampling</i> | <i>oversampling</i> |
|---------|----------|----------------------|---------------------|
| Maligna | 81 | 81 | 291 |
| Benigna | 291 | 81 | 291 |
| Total | 372 | 182 | 582 |

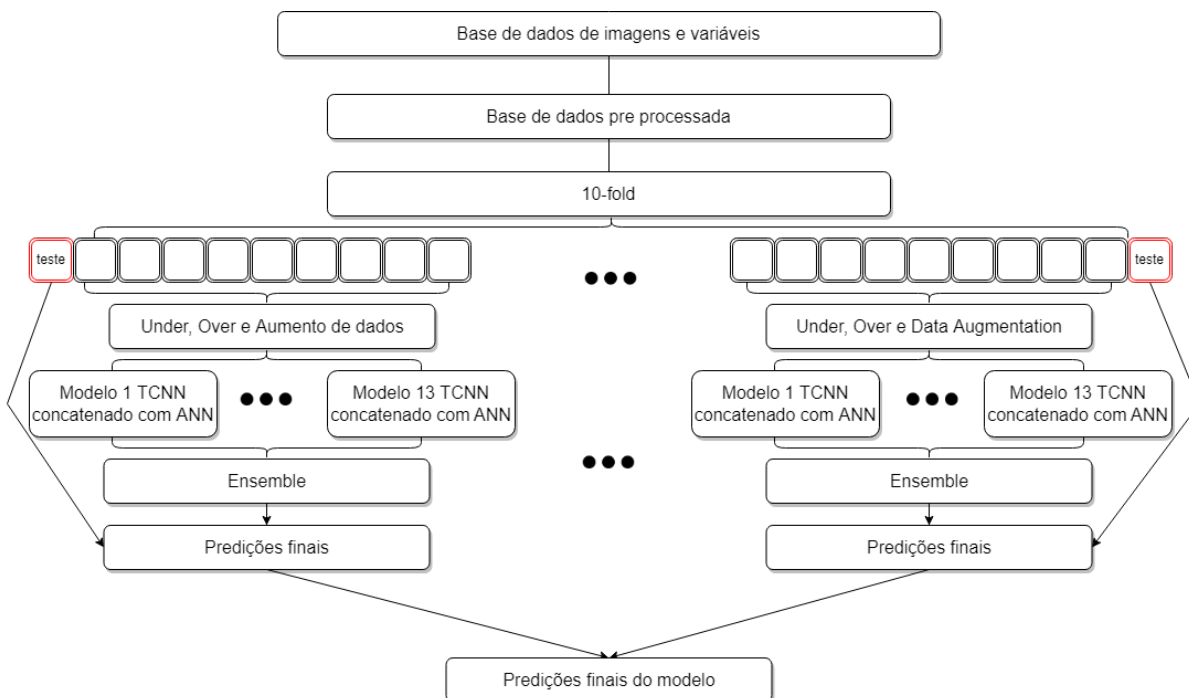
Por fim, foi realizada uma comparação da utilização do aumento de dados sem utilização da amostragem e com a utilização do *undersampling* e *oversampling*.

4.1.2 Configurações do Experimento

Como a base de dados utilizada possui imagens e suas características, para treinamento utilizamos a concatenização do modelo de transferência de aprendizado e de uma Rede Neural Artificial que foi treinada através das características da imagem.

Como citado anteriormente, a metodologia é dividida principalmente em cinco etapas, na primeira etapa as imagens da base de dados são melhoradas através de diferentes estratégias de pré-processamento apresentadas acima. Em seguida, é realizada a divisão dos *fold's*. Posteriormente os experimentos são divididos considerando ou não a implementação das técnicas de reamostragem e de aumento de dados para comparações posteriores. Após isso, os modelos de transferência de aprendizado são treinados com as imagens de entrada e as suas características do conjunto de treinamento, seguidamente a ANN e CNN são combinadas através da concatenização. Depois do treinamento dos modelos o desempenho é validado com o conjunto de teste e, por fim, as diferentes combinações são utilizadas.

Figura 28 – Representação da modelagem proposta e passo a passo, adaptado para concatenização da ANN com o modelo TCNN.



Fonte: Elaborada pelo autor.

Observe que a aplicação do aumento de dados, *undersampling* e *oversampling* é feita apenas no conjunto de treinamento de cada um dos *fold's*. Ainda, o mesmo processo foi realizado

apenas sem o aumento de dados para comparação dos resultados. O *oversampling*, *undersampling* e a não utilização dessas técnicas são aplicadas separadamente e comparadas posteriormente.

As funções de ativação utilizadas nos modelos foram ReLU e Sigmoid, sendo a função Sigmoid aplicada apenas na última camada. O Adagrad foi utilizado para treinar a CNN, processo que foi conduzido ao longo de 100 épocas. Além disso, utilizou-se o dropout para prevenir overfitting e o critério de parada Early Stopping. A função de perda adotada foi a Entropia Cruzada, apropriada para problemas de classificação binária.

Na base de dados utilizada, o melhor modelo é aquele que possui melhor taxa de acerto da classe minoritária (maligna), dado que o pior erro é prever que um indivíduo possui tumor benigno quando este possui um tumor maligno. Isto indica que é preciso ter maior sensibilidade (com menos observações malignas sendo classificadas como benignas), no entanto é necessário ter um bom recall para classificar a classe benigna corretamente e assim obter uma melhor precisão. A precisão e o valor preditivo negativo dependem das observações classificadas malignas quando são benignas (falso positivo) e benignas como malignas (falso negativo), quanto maior esses erros menores são essas métricas. Por fim, pode-se analisar a eficiência que leva em conta a ponderação entre a especificidade e o recall. Uma maneira mais eficiente de analisar os dados desbalanceados é utilizar o F beta score, pode-se analisar considerando pesos entre duas métricas. Na Tabela 6, há dois tipos de erro que o modelo pode cometer classificar um tumor maligno quando ele é benigno (FP) e classificar um tumor benigno quando este é maligno (FN) que é o pior erro que o modelo pode cometer, então é necessário minimizar esse erro que o modelo comete.

Tabela 6 – Exemplo de uma matriz de confusão considerando um modelo que classifica dois tipos de tumor maligno e benigno.

| Valor Predito | Valor Verdadeiro | |
|---------------|------------------|---------|
| | Benigno | Maligno |
| | Benigno | VP |
| Maligno | FN | VN |

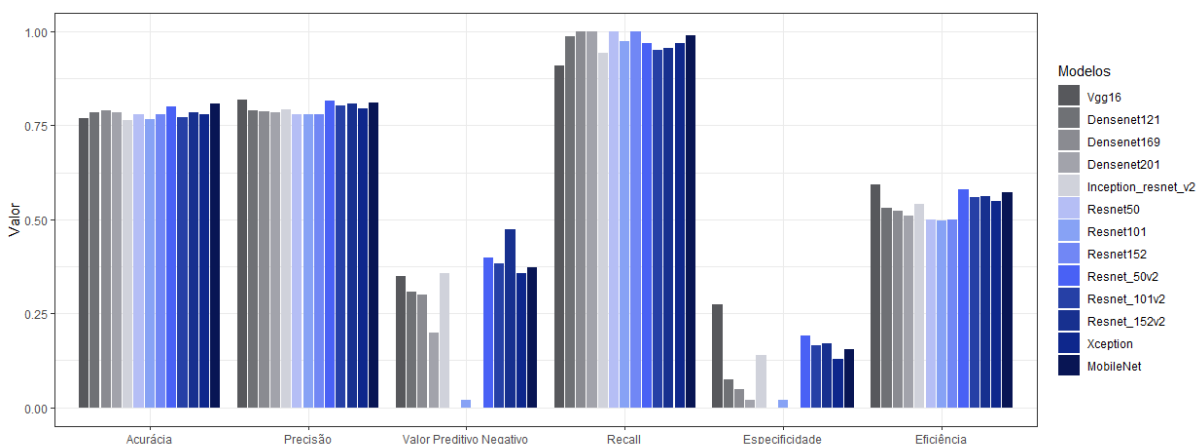
Diversos trabalhos utilizam o F beta score quando as classes da variável resposta são desbalanceadas. Por exemplo, [Lin e Lin \(2020\)](#) utilizam F beta score com um beta igual a 2, ou seja, o F2 score para avaliar o modelo em seu trabalho porque esta métrica considera tanto o recall quanto a precisão, dando mais ênfase ao recall. Segundo [Devarriya et al. \(2020\)](#), para a classificação do câncer de mama o F2 score dá mais importância para a classe minoritária (classe maligna) na classificação. Ainda, pelo mesmo autor, na medicina a identificação de instâncias minoritárias é mais importante em comparação com os dados da maioria, pois os dados da minoria contêm informações mais úteis. Alguns trabalhos que também utilizaram F2 score para dar peso no recall são [Prasetyo et al. \(2021\)](#) e [Lin e Wibowo \(2021\)](#). Por isso neste trabalho será utilizado o F2 score.

4.1.3 Resultados

Como a base de dados utilizada é desbalanceada, a classificação não tem boa predição nas observações malignas, como pode ser visto na Figura 29, onde o recall nos modelos individuais é próximo de 1 em todos os modelos (prevê corretamente quase todas as observações da classe benigna) e a especificidade é próxima de 0, ou seja não prevê quase nenhuma das observações que são malignas como malignas. Analisando a Figura 30 temos uma especificidade maior e um recall mais baixo nas primeiras combinações, além disso é possível notar que o valor preditivo negativo é maior em algumas combinações do que nos modelos individuais. Verifica-se no cenário I, sem aumento de dados (Figuras 29 e 30), que as combinações são 1 e 2, apresentam melhor eficiência, especificidade e valor preditivo negativo com relação as demais combinações. Observe que caso seja necessário tomar um maior valor preditivo negativo a melhor combinação seria a 2 e caso seja melhor tomar uma especificidade mais alta seria a combinação 1. É importante verificar que essa escolha se deve ao tipo de erro que é necessário minimizar. O modelo VGG16 apresenta melhor especificidade e eficiência conjuntamente. Portanto, é possível notar que a primeira combinação tem um melhor desempenho em prever observações da classe maligna do que os modelos individuais.

Note que alguns modelos individuais e combinações no cenário I sem aumento de dados são triviais, ou seja, as combinações acertam apenas a classe positiva errando a classe negativa. Com o aumento de dados não há nenhuma combinação ou modelo individual triviais. Este resultado mostra que o desbalanceamento da base de dados influencia o resultado quando o modelo proposto é um modelo de CNN. A utilização da combinação proposta se mostrou eficiente nesse contexto com melhor opção para prever observações malignas.

Figura 29 – Resultado dos 13 modelos utilizados sem aumento de dados.

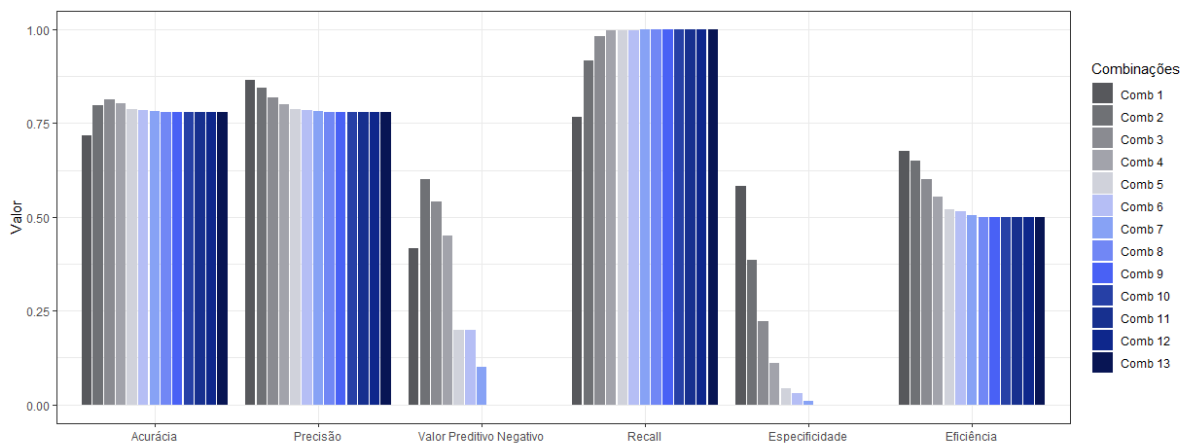


Fonte: Elaborada pelo autor.

No cenário I, com aumento de dados (Figura 31 e 32), tem-se que a combinação 12, apresenta melhor eficiência, especificidade e recall conjuntamente. Nos modelos individuais

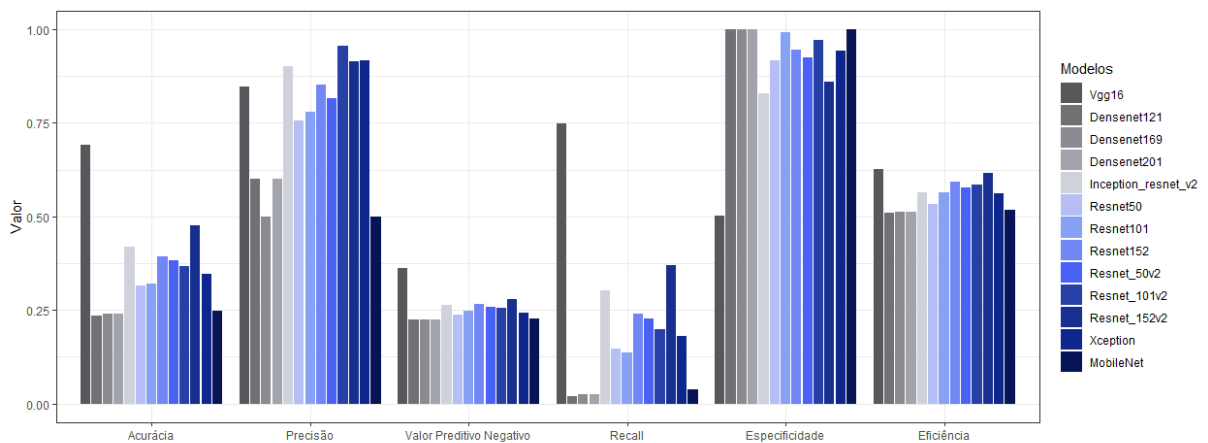
observa-se que o modelo Resnet 152V2 é o melhor comparado aos demais modelos individuais, com melhor eficiência, especificidade e recall conjuntamente. A combinação 12 tem melhor eficiência, valor preditivo negativo e maior recall do que o melhor modelo individual. É interessante ver o *trade-off* entre o recall e a especificidade nas combinações permite escolher a melhor abordagem considerando o custo-benefício. Conforme o número das combinações aumenta o valor da recall aumenta e a especificidade diminui (Figuras 30 e 32).

Figura 30 – Resultado das 13 combinações utilizadas sem aumento de dados.



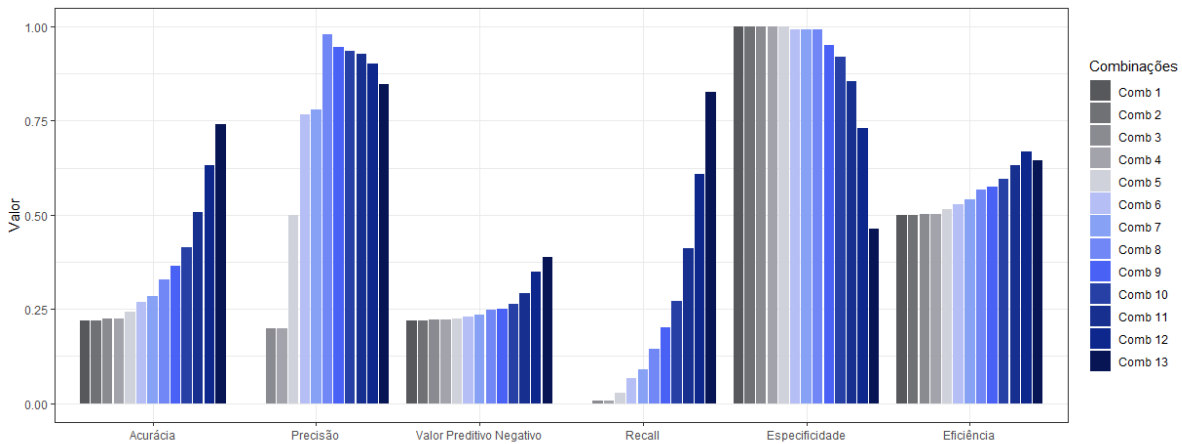
Fonte: Elaborada pelo autor.

Figura 31 – Resultado dos 13 modelos utilizados com aumento de dados.



Fonte: Elaborada pelo autor.

Figura 32 – Resultado das 13 combinações utilizadas com aumento de dados.

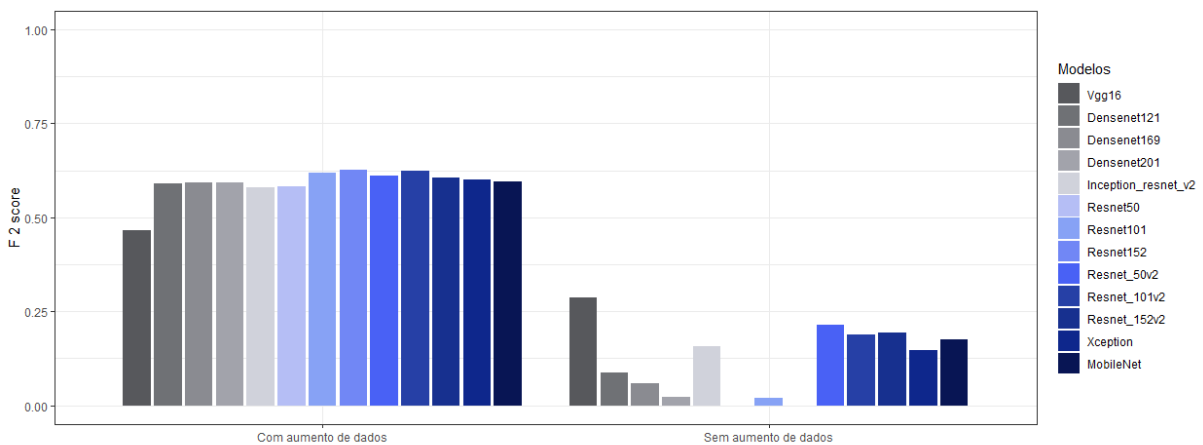


Fonte: Elaborada pelo autor.

Já para o F2 score no cenário I (Figura 33 e 34) sendo considerada a classe maligna como referência e beta como sendo 2, temos que modelo com melhor F2 score sem aumento de dados é o VGG16 e com aumento de dados é o Resnet 152, para as combinações sem aumento de dados a com melhor F2 score é a combinação 1 e a com aumento de dados é a combinação 8 ou 11 diferindo apenas por 0,10. Observe que nesse caso o F2 score é bem próximo do modelo individual com aumento de dados e da combinação.

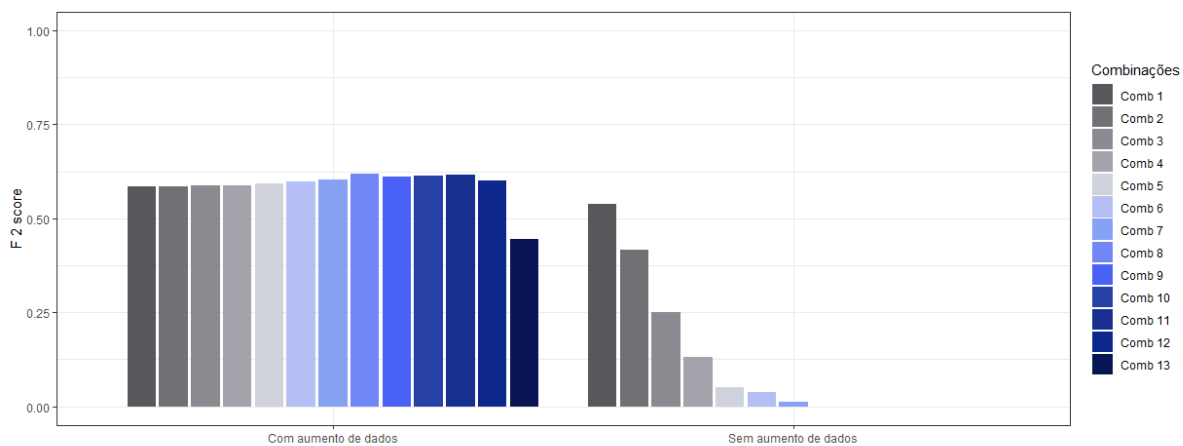
Para o F2 score no cenário II (Figura 35 e 36) e a classe com menos observações sendo considerada como referência temos que o modelo com melhor F2 score sem aumento de dados é o VGG16 e com aumento de dados é o Resnet 50 V2, para as combinações sem aumento de dados a com melhor F2 score é a combinação 3 e a com aumento de dados é a combinação 7.

Figura 33 – Resultado do F2 score dos 13 modelos utilizados.

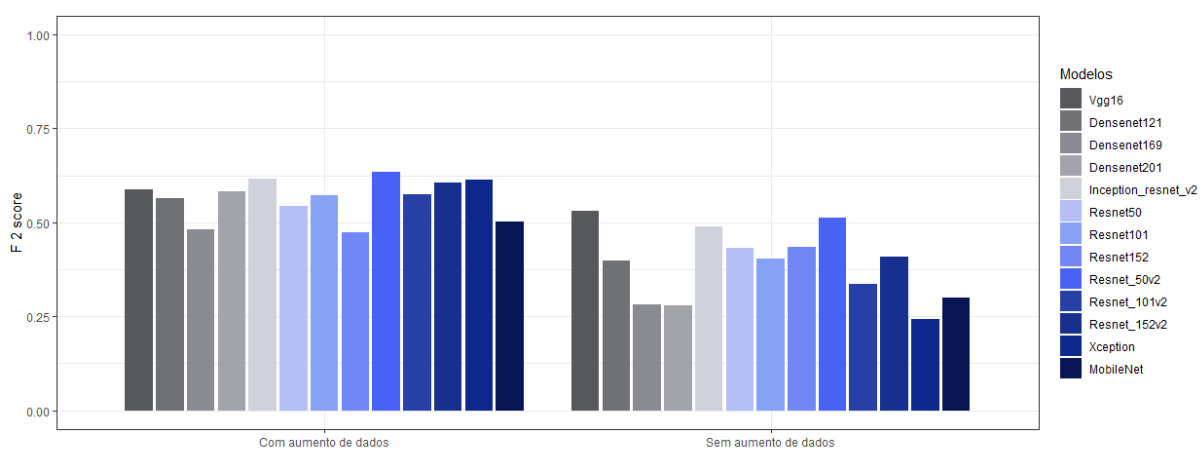


Fonte: Elaborada pelo autor.

Figura 34 – Resultado do F2 score das 13 combinações utilizadas.

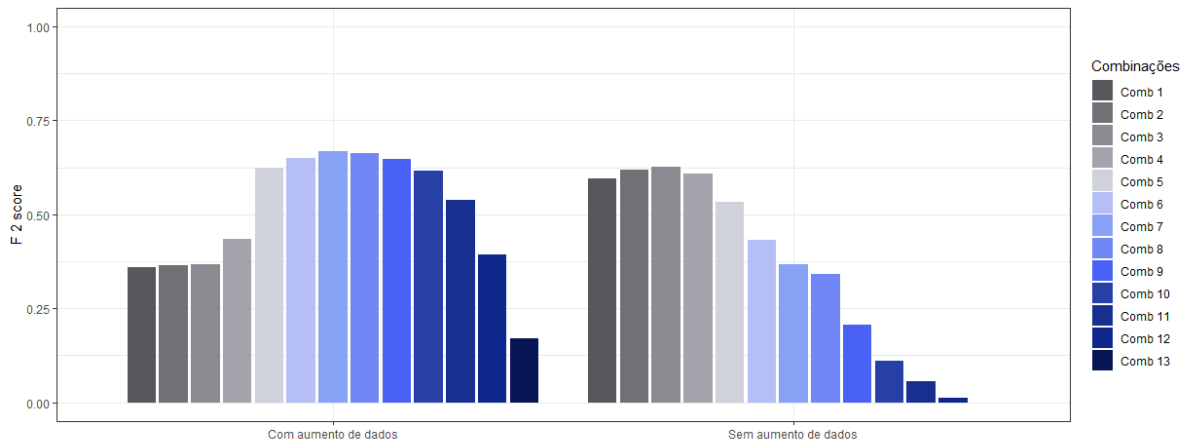


Fonte: Elaborada pelo autor.

Figura 35 – Resultado do F2 score dos 13 modelos utilizados com a utilização do método *undersampling*.

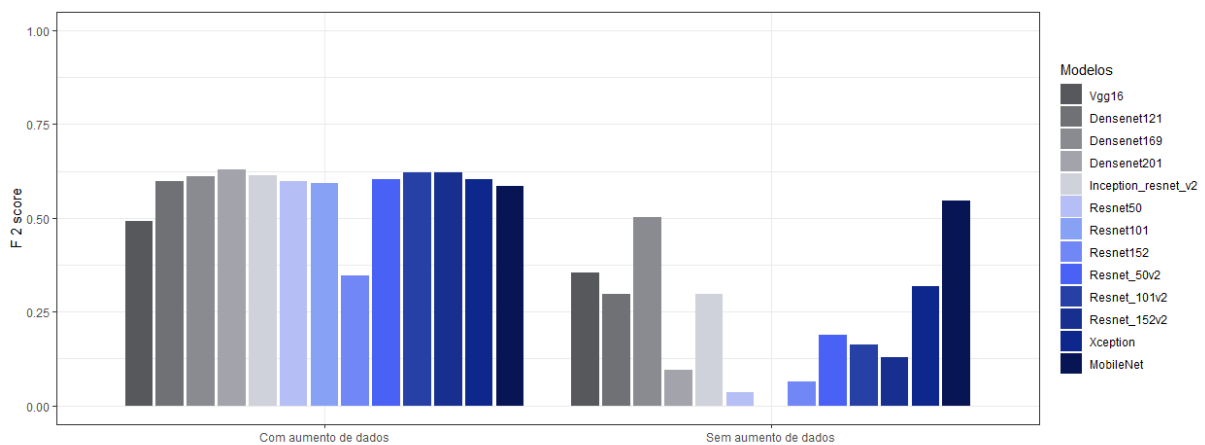
Fonte: Elaborada pelo autor.

Figura 36 – Resultado do F2 score das 13 combinações utilizadas com a utilização do método *undersampling*.



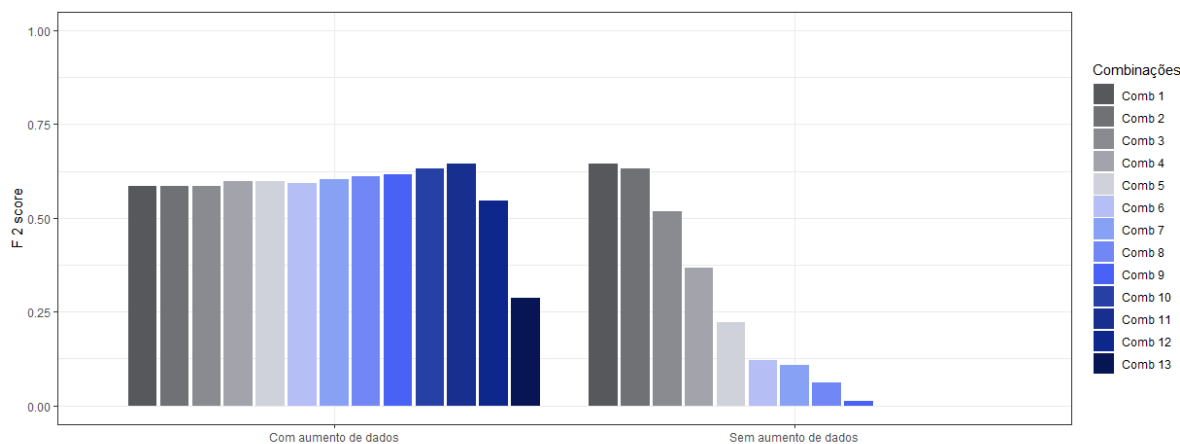
Fonte: Elaborada pelo autor.

Figura 37 – Resultado do F2 score dos 13 modelos utilizados com a utilização do método *oversampling*.



Fonte: Elaborada pelo autor.

Figura 38 – Resultado do F2 score das 13 combinações utilizadas com a utilização do método *oversampling*.



Fonte: Elaborada pelo autor.

Para o F2 score no cenário III (Figura 37 e 38) apenas considerando a classe com menos observações como referência temos que o modelo com melhor F2 score sem aumento de dados é o Mobilenet e com aumento de dados é o Densenet 201, para as combinações sem aumento de dados a com melhor F2 score é a combinação 1 e a sem aumento de dados é a combinação 11.

Analisando os três cenários conjuntamente, é possível notar que sem a utilização de reamostragem e aumento de dados há uma dificuldade em identificar elementos da classe maligna. No aumento de dados há uma melhora no acerto dessa classe. Nota-se uma melhora com relação a taxa de acerto das observações da classe maligna na utilização do método de *oversampling* e *undersampling*, o mesmo acontece na utilização das combinações propostas. E, por fim, as combinações comparadas aos modelos individuais obtiveram melhores métricas em todos os cenários. A combinação com maior F2 score está no cenário II com aumento de dados com aproximadamente 67%, a terceira combinação com maior F2 score se encontra no método *oversampling* com aumento de dados com aproximadamente 64,45%.

Portanto, conclui-se que as melhores F2 score se encontram com a utilização do ETCNN com reamostragem e aumento de dados. Em todos os cenários o ETCNN teve melhores F2 scores que os modelos individuais com e sem a utilização dos métodos de reamostragem *undersampling* e *oversampling*. Observa-se que em cada um dos cenários com e sem utilização de aumento de dados as melhores combinações diferiram, com a utilização de aumento de dados combinações maiores são melhores e sem aumento de dados combinações menores são melhores.

Ainda, é importante observar que apesar do melhor modelo no *undersampling* ter aproximadamente 67% de F2 score, essa métrica é baixa. Comparando o resultado com o [Devarriya et al. \(2020\)](#) que também utilizou uma base de dados desbalanceada de câncer de mama de 444 imagens benignas e 239 malignas, obtendo um resultado de F2 score maior que 84% em todos os 6 modelos utilizados. Outro trabalho que utiliza F2 score no mesmo contexto é o

Calderon-Ramirez *et al.* (2022) que tem uma base de dados de aproximadamente 4.000 imagens, com 5% sendo imagens de tumor maligno. Nesse trabalho, o modelo com maior F2 score obteve um valor de aproximadamente 40%, em que o autor não utilizou nenhuma metodologia de reamostragem.

É importante destacar que, embora a diferença entre os resultados com e sem a combinação possa parecer pequena, como no caso de um aumento de apenas 1%, no contexto médico essa variação pode representar vidas salvas.

Os dados desbalanceados afetam a predição da categoria minoritária da base de dados, isto fica evidente no cenário I, onde os modelos de transferência de aprendizado tiveram poucos acertos nessa categoria. O método proposto têm um grande espaço para melhorias comparado ao trabalho de Devarriya *et al.* (2020). O número de observações da classe maligna da base de dados utilizada do A.C.Camargo Câncer Center possui menos que a metade do trabalho citado anteriormente. Por esse motivo, acredita-se que com mais observações dessa classe a métrica F2 score pode apresentar melhorias. Mesmo com essas limitações a utilização da combinação com aumento de dados e reamostragem alcançou melhor F2 score, sendo uma boa alternativa à não utilização da combinação, aumento de dados e reamostragem.

Por fim, é importante observar que a escolha da combinação vai depender da base de dados, da metodologia de balanceamento e do número de modelos utilizados. A melhor combinação neste trabalho não necessariamente é a melhor para outras bases de dados.

4.2 Câncer de pele

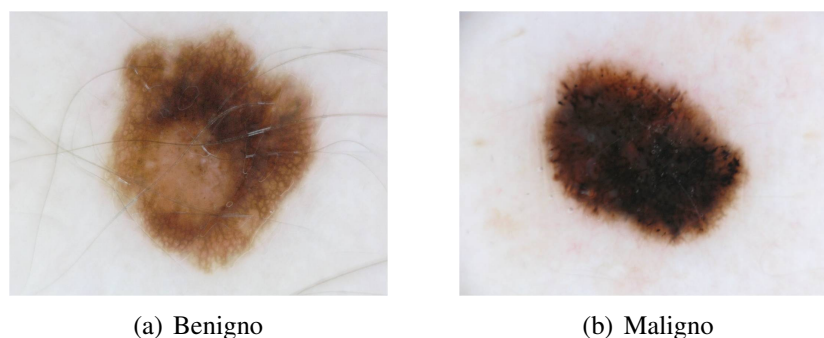
Para verificar a metodologia proposta, utilizamos um pequeno conjunto de dados desbalanceado fornecido pelos ISIC Challenge Datasets 2016 (GUTMAN *et al.*, 2016). E tem como objetivo classificar as imagens das lesões dermatoscópicas em malignas ou benignas.

Nesta aplicação, utilizamos ainda as 13 combinações de modelos focados na classificação binária (Maligno ou Benigno) para as imagens de lesões dermatoscópicas. Para verificar se as combinações utilizadas eram eficazes, comparamos o resultado das combinações com o resultado individual de cada um dos 13 modelos utilizados. Os cenários considerados são os mesmos da aplicação anterior.

4.2.1 Base de dados

O conjunto de dados empregado consiste em 900 imagens destinadas ao treinamento e 379 imagens para teste, categorizadas em duas classes: malignas e benignas. A distribuição das imagens é apresentada na Tabela 7 e um exemplo delas é apresentado na Figura 39. Para treinar o modelo foram utilizadas apenas as imagens de treinamento, enquanto as imagens de teste foram utilizadas para comparação com outros trabalhos que utilizaram a mesma base de dados.

Figura 39 – Imagens da base de dados do ISIC 2016.



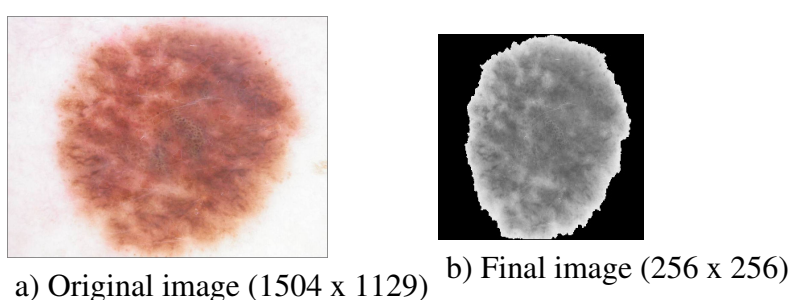
Fonte: Gutman *et al.* (2016).

Tabela 7 – Número de elementos pertencentes a cada uma das classes da base de dados do ISIC 2016.

| Classe | Número de elementos (treino) | Número de elementos (teste) |
|---------|------------------------------|-----------------------------|
| Maligno | 173 | 75 |
| Benigno | 727 | 304 |
| Total | 900 | 379 |

Nesta aplicação, as imagens do conjunto de dados continham algum ruído e possuíam diferentes tamanhos de largura e altura. Por essa razão, foi necessário realizar um recorte para focar apenas na área da lesão na imagem. Em seguida, foi aplicado um filtro para reduzir as partes menos relevantes para a análise. Por fim, a imagem final foi redimensionada para ter dimensões 256×256 . A figura 40 apresenta um exemplo deste pré-processamento da imagem original.

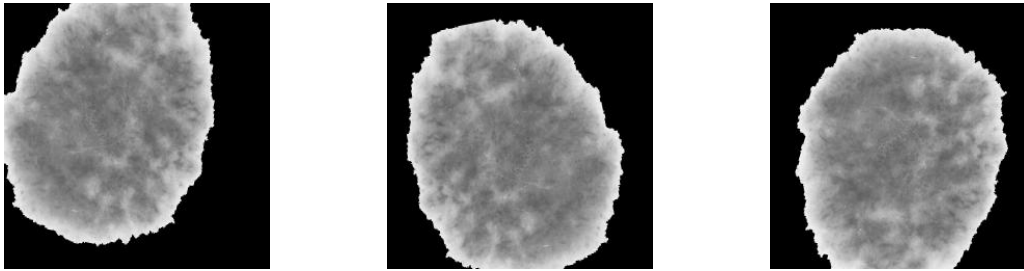
Figura 40 – Pré-processamento da imagem.



Fonte: Adaptado de Gutman *et al.* (2016).

Para obter uma boa generalização do modelo, neste trabalho foi utilizado o aumento de dados no conjunto de treinamento. Os dados foram aumentados com métodos tradicionais de aumento de dados. Algumas imagens ampliadas usando a imagem final da Figura 40 são mostradas na Figura 41. Ainda neste estudo foi considerado o não uso de aumento de dados em cada um dos cenários, ao final é realizada uma comparação entre o uso e o não uso de aumento.

Figura 41 – Aumento de dados com a imagem final na Figura 40.



Fonte: Adaptado de Gutman *et al.* (2016).

Nesta aplicação foram utilizados *undersampling* e *oversampling* no conjunto de treinamento com o objetivo de balancear o conjunto de dados em estudo e comparar as metodologias no caso com e sem reamostragem. Um exemplo de como foi realizado o processo de reamostragem é mostrado na Tabela 8, apenas para o primeiro fold do modelo. O k-fold com 10 fold's foi utilizado para validação do modelo.

Tabela 8 – Número de elementos pertencentes a cada classe no conjunto de dados ISIC usando undersampling e oversampling em fold-1.

| Classe | Original | Undersampling | Oversampling |
|---------|----------|---------------|--------------|
| Maligno | 173 | 151 | 659 |
| Benigno | 727 | 151 | 659 |
| Total | 900 | 352 | 1318 |

4.2.2 Configurações do Experimento

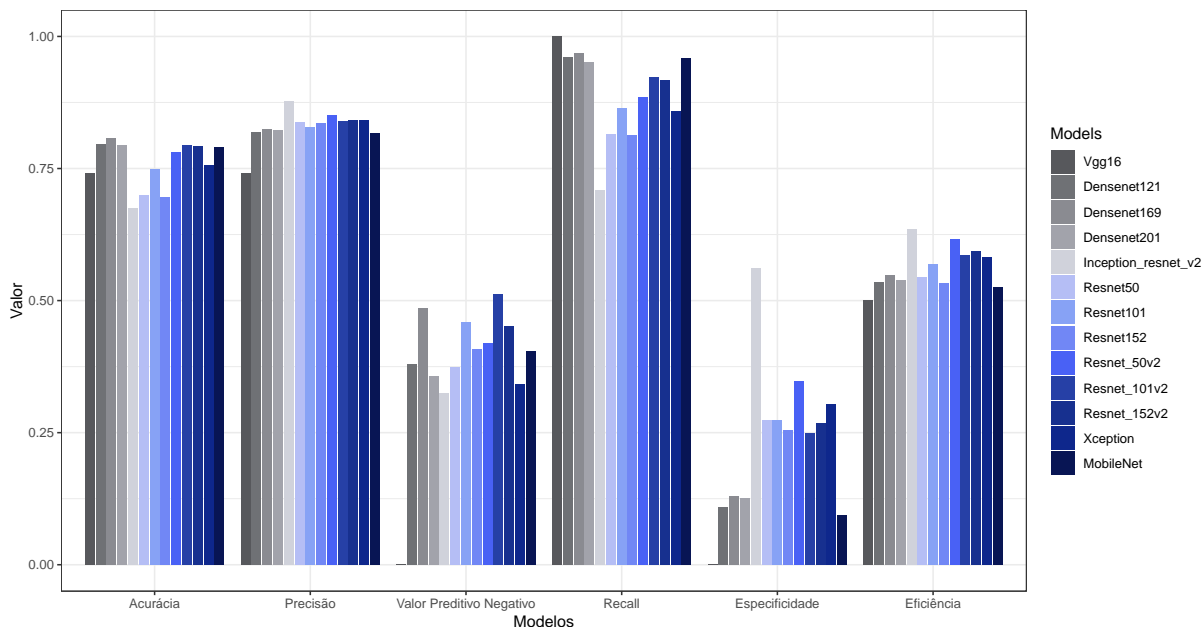
A base de dados utilizada neste experimento consiste exclusivamente em imagens. Diferentemente da aplicação anterior, onde foi realizada a concatenação de um modelo de transferência de aprendizado com uma Rede Neural Artificial, como descrito na aplicação 4.1, nesta abordagem não se utilizou essa estratégia. As demais configurações aplicadas neste experimento foram mantidas idênticas às do experimento relacionado ao câncer de pele. O fluxograma dessa configuração pode ser visto na Figura 24.

4.2.3 Resultados

Esta seção apresenta os resultados e suas respectivas discussões referentes aos experimentos propostos com o ISIC Challenge Dataset 2016. Neste trabalho foram utilizadas 13 combinações de modelos focados na classificação binária (maligno ou benigno) para imagens de lesões dermatoscópicas com conjunto de dados desbalanceado. Para verificar se as combinações utilizadas foram eficazes, comparamos o resultado das combinações com o resultado individual de cada um dos 13 modelos utilizados. Os cenários considerados são: (I) sem técnica de reamostragem (com e sem aumento de dados); (II) com uso de *undersampling* para balanceamento (com

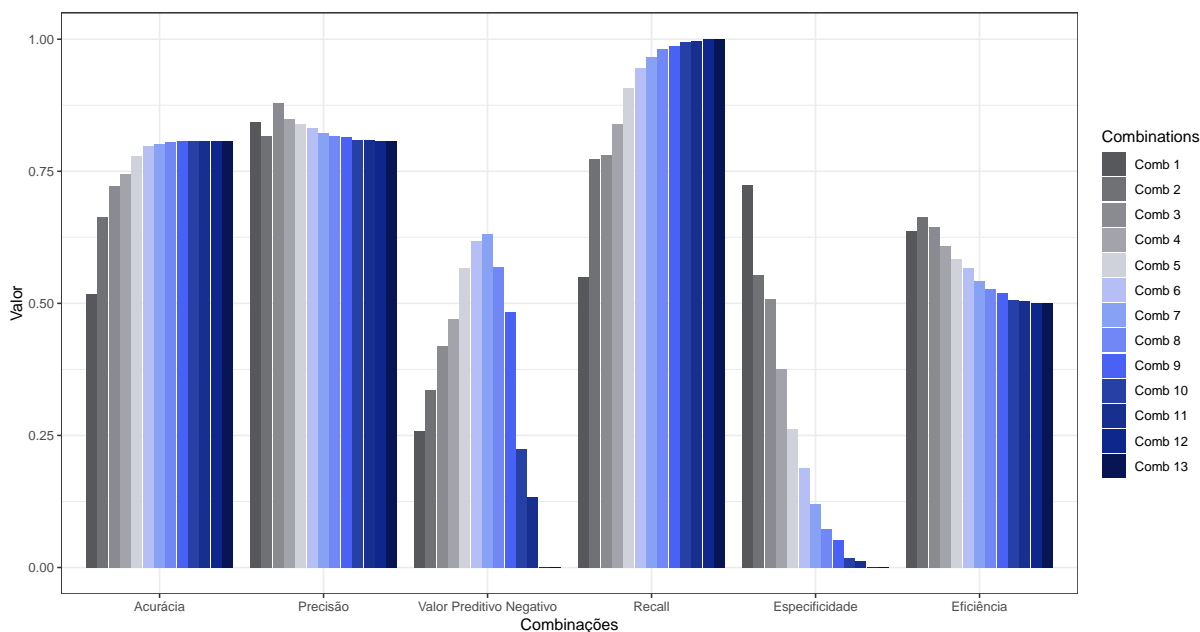
e sem aumento de dados); e (III) com *oversampling* para balanceamento (com e sem aumento de dados).

Figura 42 – Resultados dos 13 modelos usando aumento de dados.



Fonte: Elaborada pelo autor.

Figura 43 – Resultados das 13 combinações usando aumento de dados.

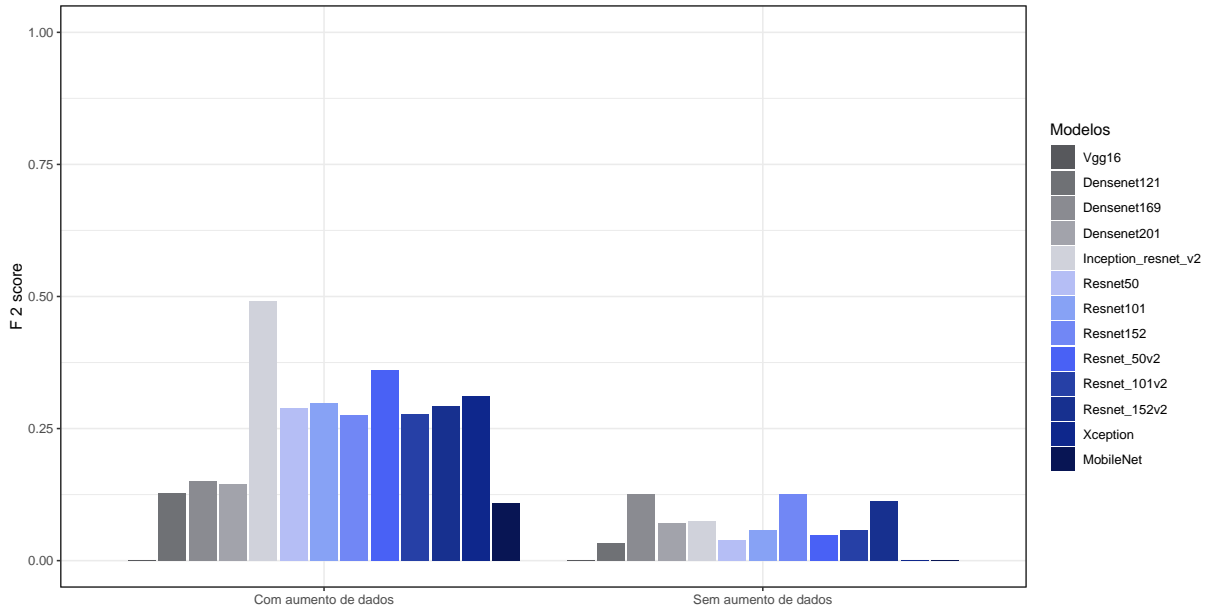


Fonte: Elaborada pelo autor.

No cenário I, com aumento de dados (Figura 42 e 43), as combinações de 5 a 8 apresentam melhores valores preditivos negativos, já as combinações de 1 a 3 apresentam melhores eficiência

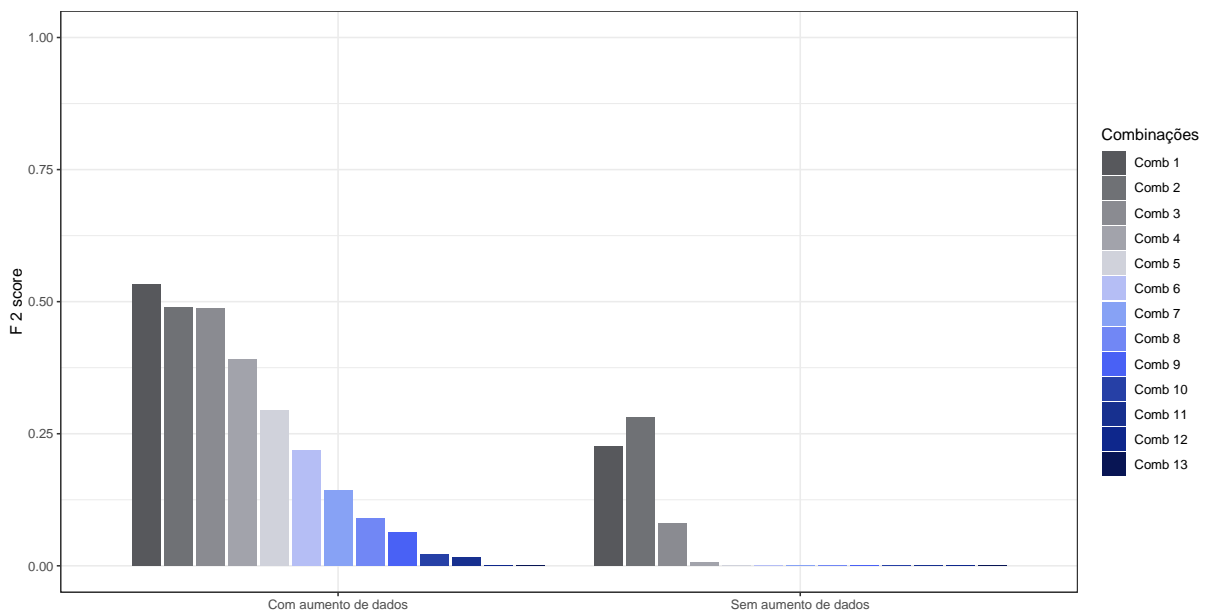
e precisão juntas. Nos modelos individuais, observa-se que o modelo Densenet201 é o modelo com melhor especificidade e o Densenet 269 com melhor acurácia.

Figura 44 – Resultados do F2 score dos 13 modelos utilizados.



Fonte: Elaborada pelo autor.

Figura 45 – Resultados do F2 score das 13 combinações utilizadas.

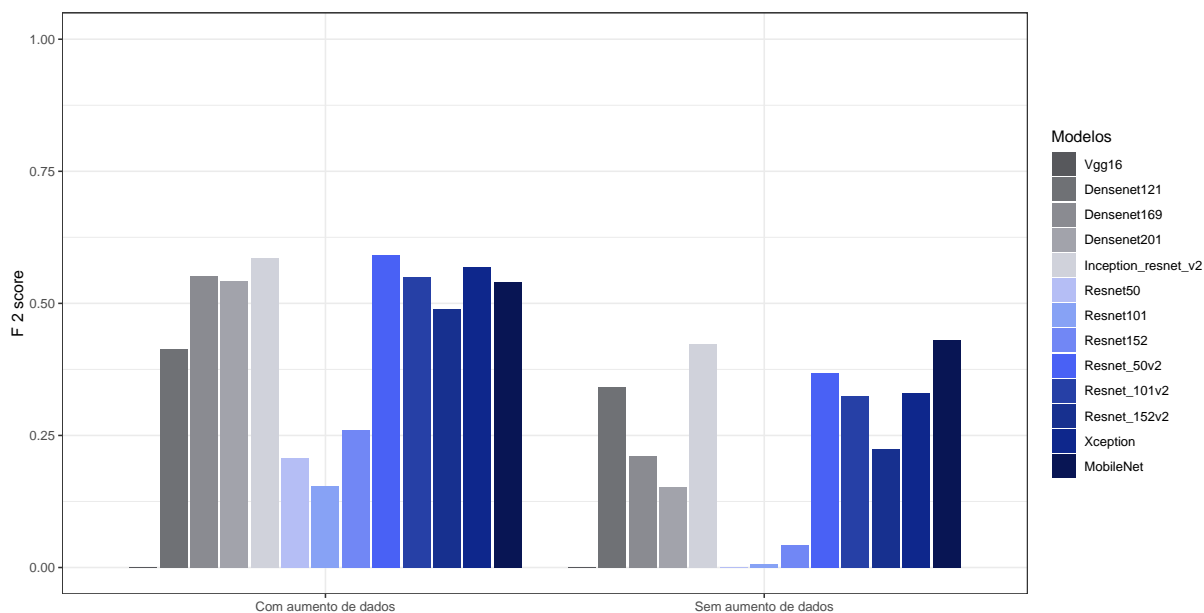


Fonte: Elaborada pelo autor.

Quanto ao F2 score no cenário I (Figura 44 e 45) considerando a classe com menor número de observações como referência, temos que o modelo com melhor F2 score sem aumento

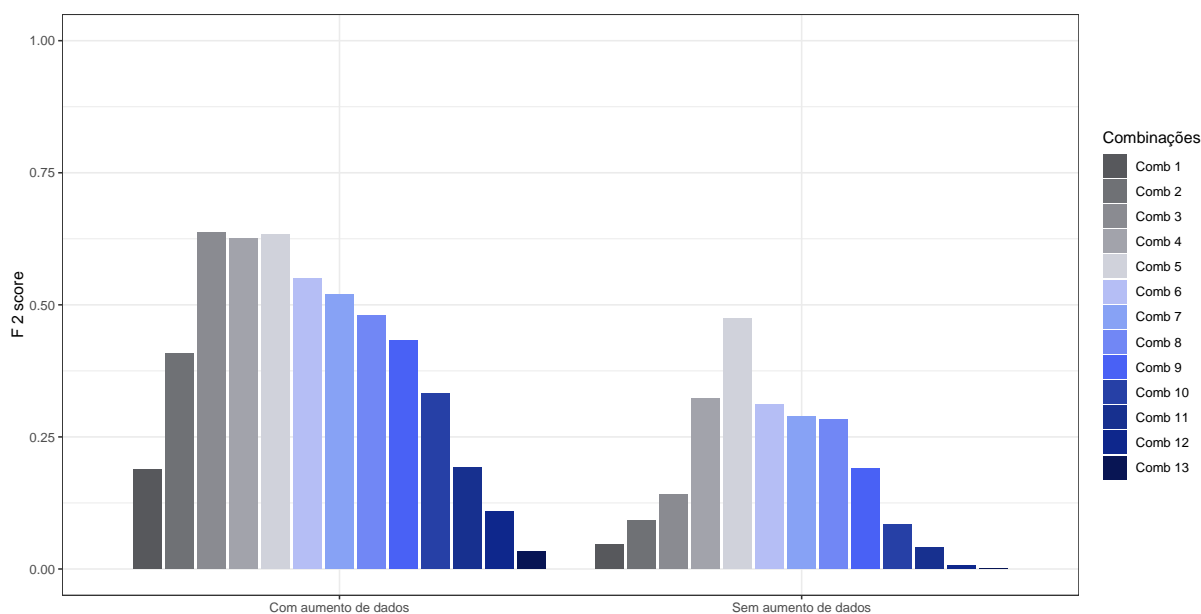
com aumento de dados é Densenet121 e com aumento de dados é Densenet201, para as combinações aquela com melhor F2 score é a combinação 1 com aumento de dados e a combinação 2 sem aumento de dados.

Figura 46 – Resultados do F2 score dos 13 modelos utilizados com o método *undersampling*



Fonte: Elaborada pelo autor.

Figura 47 – Resultados do F2 score das 13 combinações utilizadas com o método *undersampling*

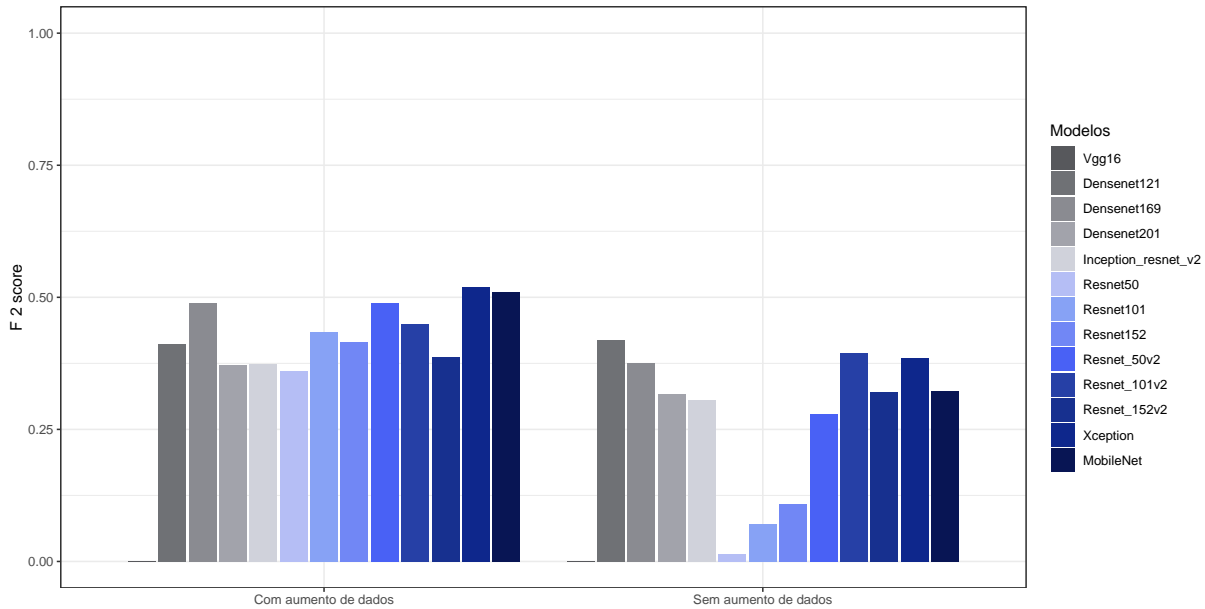


Fonte: Elaborada pelo autor.

Para o F2 score no cenário II (Figura 46 e 47) e a classe com menor número de obser-

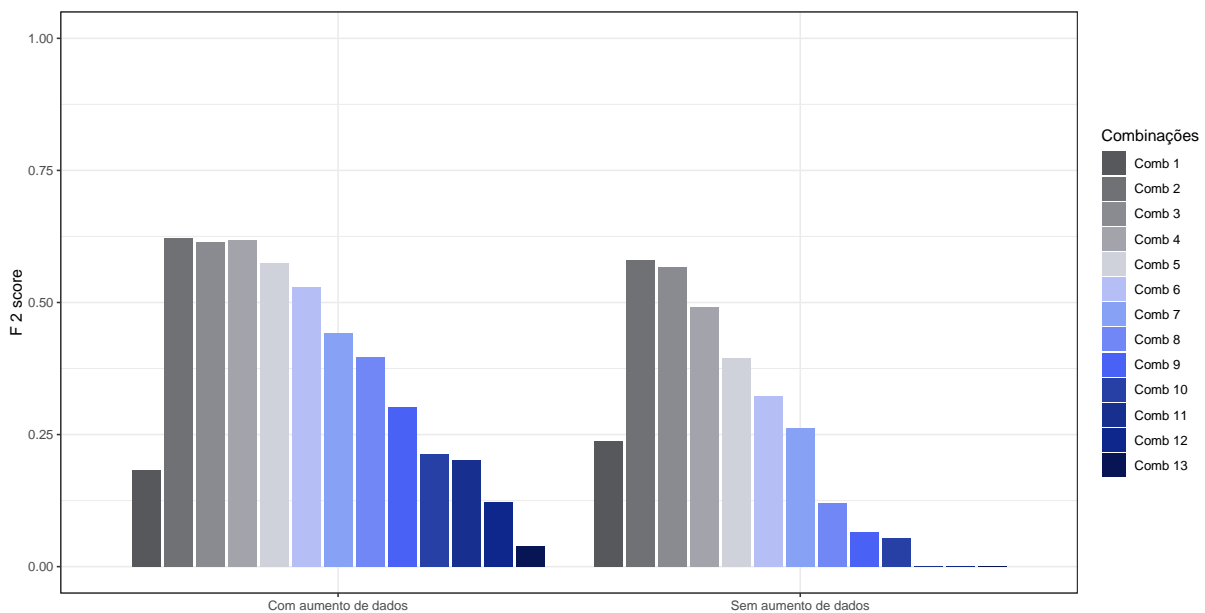
vações sendo considerada como referência, o modelo com melhor F2 score sem aumento de dados é o Mobilenet e com aumento de dados é o Resnet 50 V2, para as combinações aquela com melhor F2 score é a combinação 3 com aumento de dados e 5 sem aumento de dados.

Figura 48 – Resultados do F2 score dos 13 modelos utilizados com o método *oversampling*



Fonte: Elaborada pelo autor.

Figura 49 – Resultados do F2 score das 13 combinações utilizadas com o método *oversampling*



Fonte: Elaborada pelo autor.

Para o F2 score no cenário III (Figura 48 e 49) considerando apenas a classe com menor

número de observações como referência, temos que o modelo com melhor F2 score é o VGG16 sem aumento de dados e Xception com aumento de dados, para combinações aquela com a melhor F2 score é a combinação 2 com e sem aumento de dados.

Analisando os três cenários em conjunto observa-se que, sem reamostragem ou aumento de dados, há dificuldade em identificar elementos da classe maligna. A classificação de observações malignas melhora com a utilização de métodos de *undersampling*, *oversampling* e aumento de dados. As combinações dos modelos de transferência de aprendizado superam os modelos individuais em todos os cenários. E, por fim, as combinações comparadas aos modelos individuais obtiveram melhores métricas em todos os cenários.

A combinação com maior F2 score foi observada no cenário II, com aumento de dados de aproximadamente 64%, seguida pelo no cenário III com aumento de dados de 62%. Portanto, conclui-se que os melhores F2 score são encontrados com a utilização do ETCNN com reamostragem e aumento de dados. Em todos os cenários, o ETCNN obteve melhores F2 scores do que os modelos individuais com e sem utilização de métodos de reamostragem.

Além disso, é importante notar que embora o melhor modelo em *undersampling* tenha aproximadamente 64% de F2 score, esta métrica é baixa quando comparando o resultado com o de Zhou, Ren e Qi (2020) na Tabela 9. Esta tabela apresenta alguns trabalhos que utilizaram este mesmo banco de dados. O método proposto tem espaço para melhorias, como a utilização de técnicas de aumento de dados mais avançadas. Os resultados obtidos indicam que a metodologia proposta pode ser promissora em aplicações futuras, especialmente no contexto de dados desbalanceados.

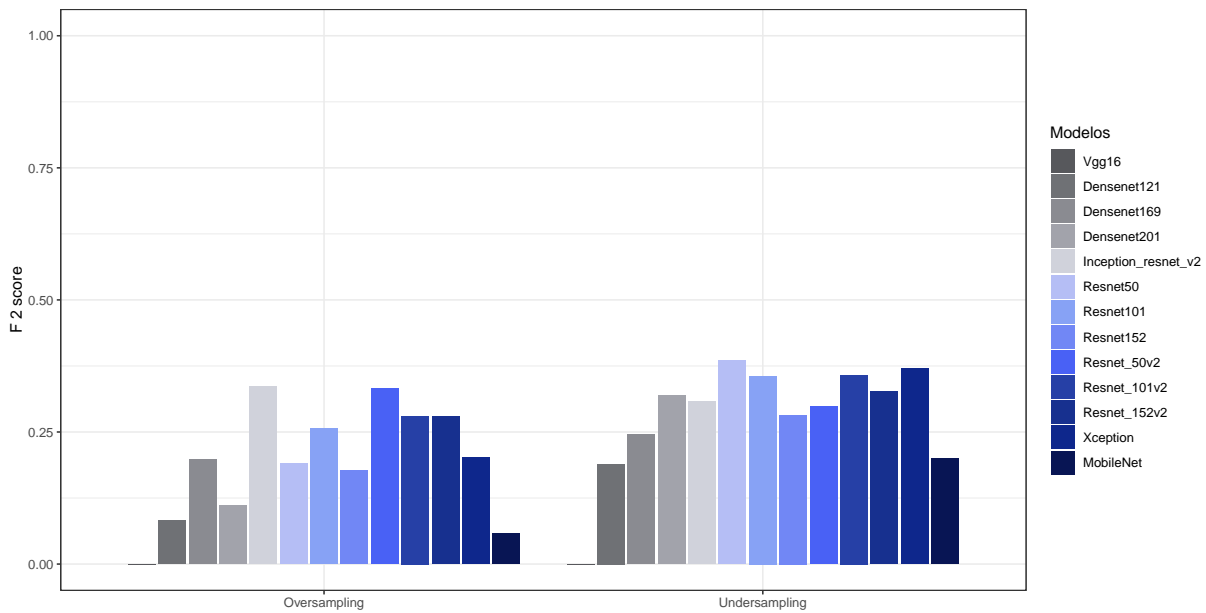
Tabela 9 – Comparação de outros trabalhos do ISIC 2016 no conjunto de teste.

| Autores | F2 score | Especificidade | Precisão |
|--------------------------------|----------|----------------|----------|
| (ZHOU; REN; QI, 2020) | 0,49 | 0,99 | - |
| | 0,69 | 0,95 | - |
| (ALIZADEH; MAHLOOJIFAR, 2021) | - | 0,93 | - |
| (CASSIDY <i>et al.</i> , 2022) | - | 0,58 | 0,22 |
| | - | 0,94 | 0,21 |
| (GUTMAN <i>et al.</i> , 2016) | - | 0,92 | - |
| | - | 0,94 | - |
| | - | 0,93 | - |
| (LOPEZ <i>et al.</i> , 2017) | - | 0,79 | 0,80 |
| (CODELLA <i>et al.</i> , 2017) | - | 0,55 | 0,93 |
| <i>undersampling</i> | 0,54 | 0,96 | 0,20 |

4.2.4 Aumento de dados com GAN

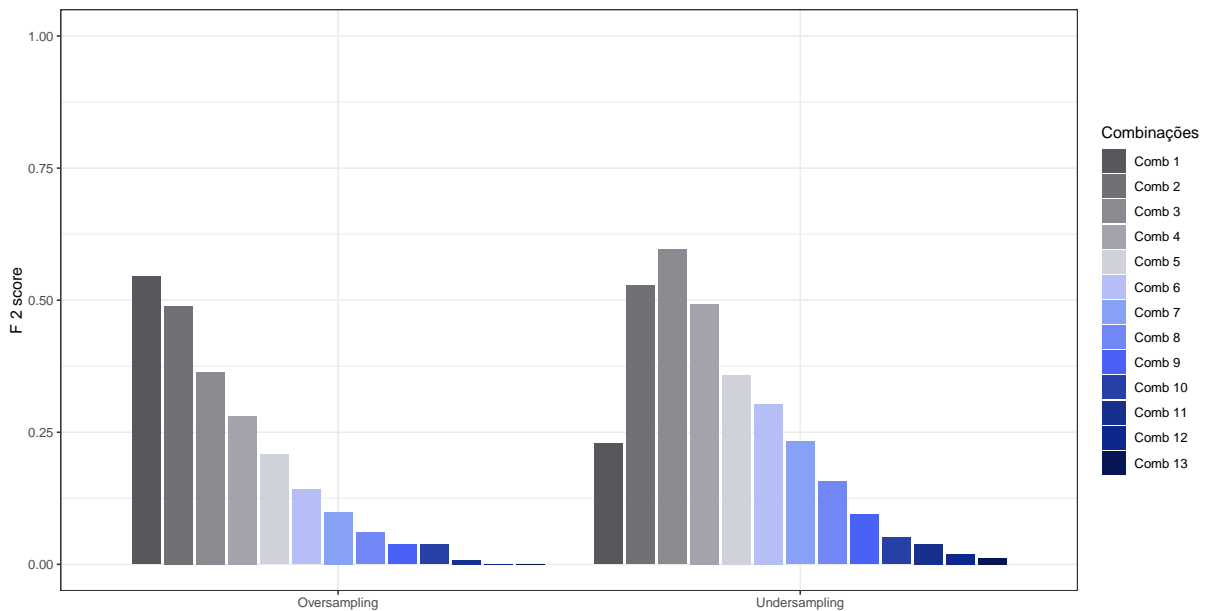
Nesta seção os resultados sobre o ISIC 2016 são apresentados considerando o aumento de dados através de uma GAN. O fluxograma da metodologia utilizada neste estudo é apresentado na Figura 24.

Figura 50 – Resultados do F2 score dos 13 modelos utilizados aumento de dados com a GAN com o *Oversampling* e *Undersampling*



Fonte: Elaborada pelo autor.

Figura 51 – Resultados do F2 score das 13 combinações utilizadas aumento de dados com a GAN com o *Oversampling* e *Undersampling*



Fonte: Elaborada pelo autor.

As Figuras 50 e 51 apresentam os resultados referentes aos F2 no *undersampling* e *oversampling*. No *undersampling*, o modelo com melhor F2 sem aumento com dados aumento é

Resnet 50 e no *oversampling* o Inception Resnet v2. Considerando as combinações o melhor F2 score é a combinação 3 no *undersampling* e combinação 1 no *oversampling*.

A partir desse resultado, conclui-se que os melhores F2 são obtidos pela utilização do ETCNN com reamostragem e aumento de dados utilizando GAN. Além disso, é importante notar que embora os melhores modelos em *undersampling* tem aproximadamente 60% de pontuação F2 score, essa métrica é um pouco mais baixa do que a de aumento de dados comum.

Já para o conjunto de teste no *undersampling* o modelo alcançou métricas de F2 score 0,53, Especificidade 0,88 e Precisão 0,20, bem similares ao aumento de dados comum realizado anteriormente. Comparando o resultado obtido com os trabalhos apresentados na Tabela 9, o método proposto tem espaço para melhorias. Tal tabela apresenta alguns trabalhos que utilizaram a mesma base de dados, onde as métricas se referem ao conjunto de teste disponível.

4.3 Conclusão

A principal contribuição destas aplicações é a proposta de uma ETCNN com reamostragem e aumento de dados para melhorar a predição de tumor maligno nas bases de dados utilizadas. Para a aplicação da metodologia foram utilizadas imagens de mamografia e imagens de lesões de pele para diagnosticar tipos de câncer, pois a detecção precoce do tumor maligno é crucial para evitar que ele se espalhe para outros lugares do corpo.

Inicialmente foi aplicado o pré-processamento às imagens brutas, para que apenas a área da mama e a área da lesão da pele seja avaliada para detecção de câncer. Em seguida a base de dados foi dividida em conjuntos de treino e teste, de modo que na base de treino foi aplicada três diferentes abordagens: sem reamostragem, *undersampling* e *oversampling*. Essas partes foram utilizadas com e sem aumento de dados. A próxima etapa consistiu na transferência de aprendizado e no total 13 *ensembles* foram realizados. Os resultados apontam que as combinações propostas tem melhores desempenhos em todas as abordagens para detectar tumor maligno.

A partir dos resultados obtidos nota-se que as combinações com melhores métricas são obtidas com a utilização de reamostragem e as piores sem a utilização desta técnica. Dessa forma, comprova-se que a utilização da metodologia de reamostragem *undersampling* e *oversampling* conjuntamente com a combinação dos modelos de transferência de aprendizado melhora a previsão da classe minoritária. Ainda, com a utilização de aumento de dados, as combinações propostas tem um melhor desempenho na predição de tumor maligno do que a não utilização desta técnica.

ESTUDO DE SIMULAÇÃO

Neste capítulo são descritos e apresentados os cenários e os resultados dos estudos de simulação realizados. Esses estudos tiveram como objetivo avaliar o comportamento das combinações propostas em diferentes cenários, considerando tanto situações desbalanceadas quanto balanceadas.

5.1 Estrutura da simulação

Os experimentos foram conduzidos alterando a proporção de observações da classe maligna na base de dados ISIC, utilizada no capítulo anterior. Além disso seguiram a mesma estrutura que foi apresentada na Figura 24.

A Tabela 5.1 ilustra a estrutura das bases de dados simuladas.

| Simulação | Dados malignos (%) | Dados benignos (%) |
|-----------|--------------------|--------------------|
| 1 | 10 | 90 |
| 2 | 20 | 80 |
| 3 | 30 | 70 |
| 4 | 40 | 60 |
| 5 | 50 | 50 |

Toda a simulação foi realizada utilizando a linguagem Python, desde a seleção das imagens que compuseram cada um das simulações até o pré-processamento descrito na Seção 4.2.1. As imagens malignas utilizadas nas simulações foram retiradas das bases de dados ISIC 2017 e ISIC 2018. A base de dados original do ISIC 2016 apresenta uma proporção inicial de 80,77% de imagens benignas, com um total de 727 imagens benignas em um conjunto de 900 imagens. Dessa forma, o cenário de simulação 2 reflete o resultado descrito no Subcapítulo 4.2.

As simulações 1 e 3 que tem dados são desbalanceados consideraram os cenários com e sem reamostragem, conforme discutido nos capítulos anteriores da seguinte forma:

- (I) Sem técnica de reamostragem (com e sem aumento de dados);
- (II) Com uso de *undersampling* para balanceamento (com e sem aumento de dados);
- (III) Com *oversampling* para balanceamento (com e sem aumento de dados).

Nas simulações 4 e 5, que representam dados aproximadamente balanceados, respectivamente, foi considerado apenas o caso sem reamostragem, a fim de permitir uma comparação direta.

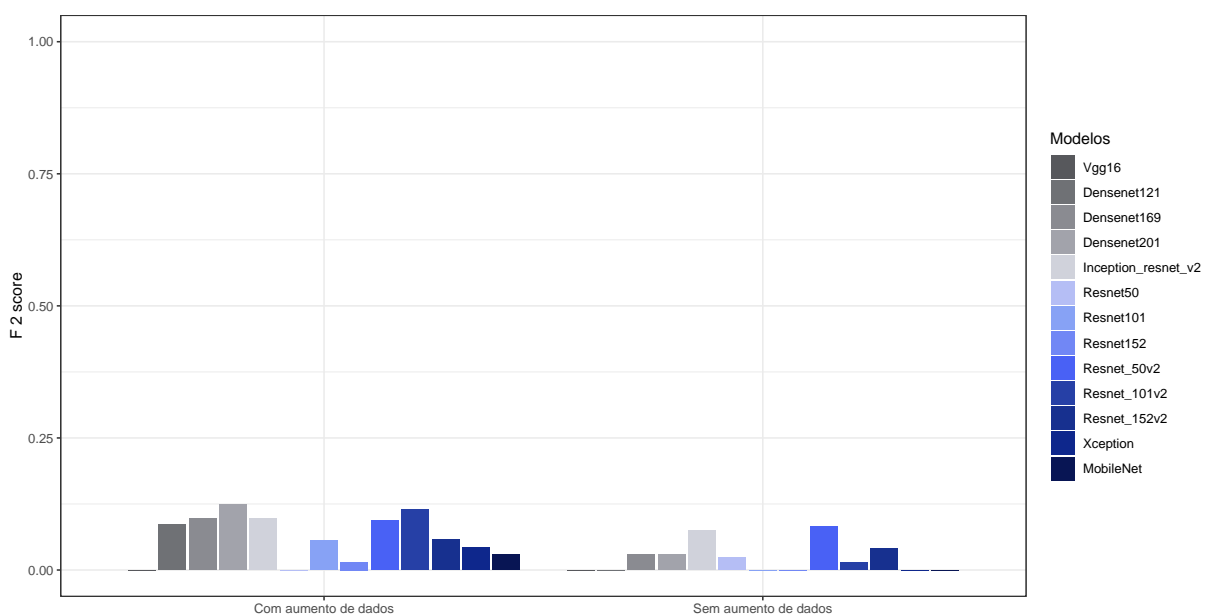
Para a simulação 1, foram removidas aleatoriamente imagens malignas até que a proporção de 90% de imagens benignas e 10% de imagens malignas fosse atingida. Para as demais simulações, foram adicionadas aleatoriamente imagens malignas até alcançar as proporções desejadas. Por exemplo, no cenário de simulação 5 (50% de imagens benignas e 50% de imagens malignas), foram utilizadas 727 imagens benignas e 727 imagens malignas.

5.2 Simulação

Esta seção apresenta os resultados das simulações, considerando a estrutura dos experimentos propostos neste trabalho, utilizando as bases de dados do ISIC 2016, 2017 e 2018.

5.2.1 Simulação 1

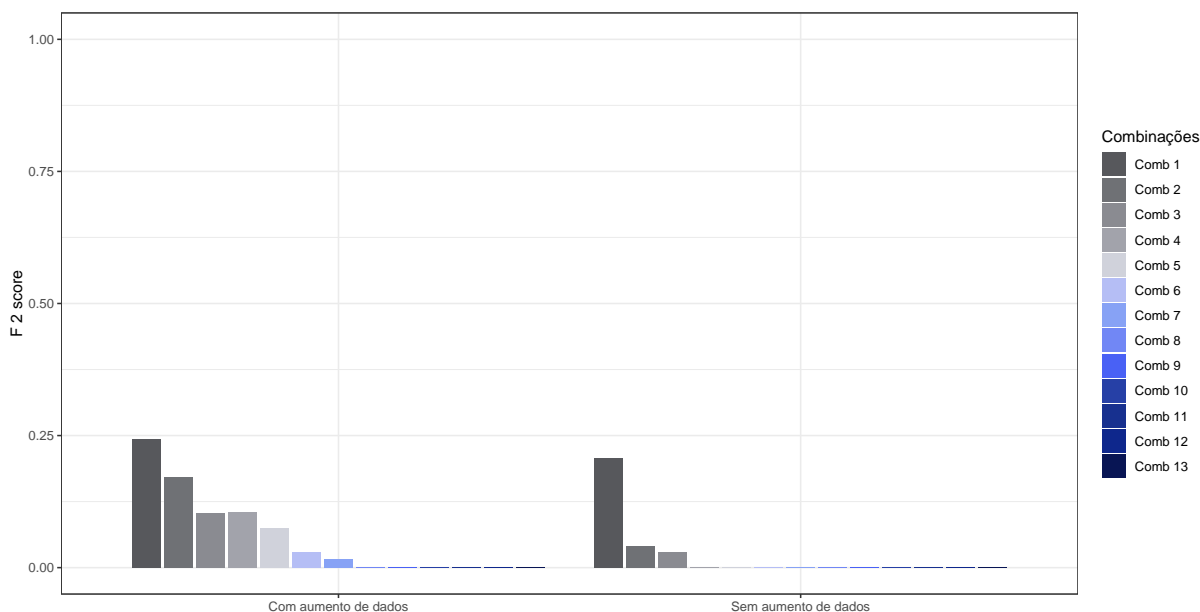
Figura 52 – Resultados do F2 score dos 13 modelos utilizados (10%).



Fonte: Elaborada pelo autor.

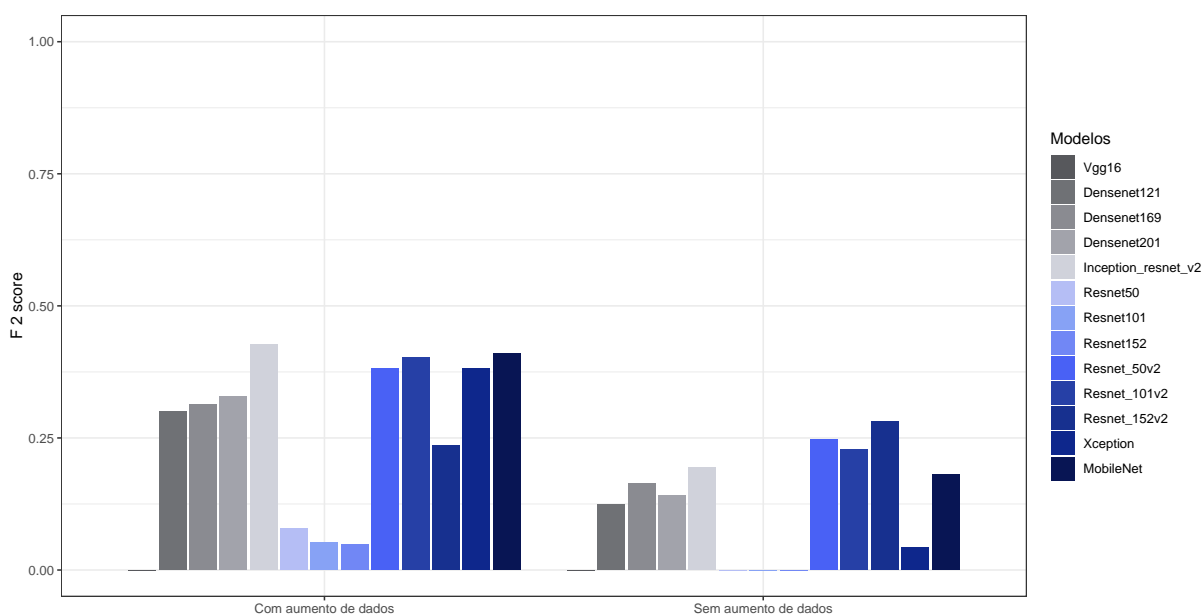
Na simulação 1, temos 10% das imagens sendo malignas e 90% benignas. Os resultados dessa simulação são apresentados a seguir. No cenário I, para os modelos individuais (Figura 52), observa-se que os modelos apresentam dificuldades em detectar a classe maligna, resultando em F2-scores baixos. Já para as combinações (Figura 53), a combinação 1 tem melhor F2 score com aumento de dados e sem aumento de dados.

Figura 53 – Resultados do F2 score das 13 combinações utilizadas (10%).



Fonte: Elaborada pelo autor.

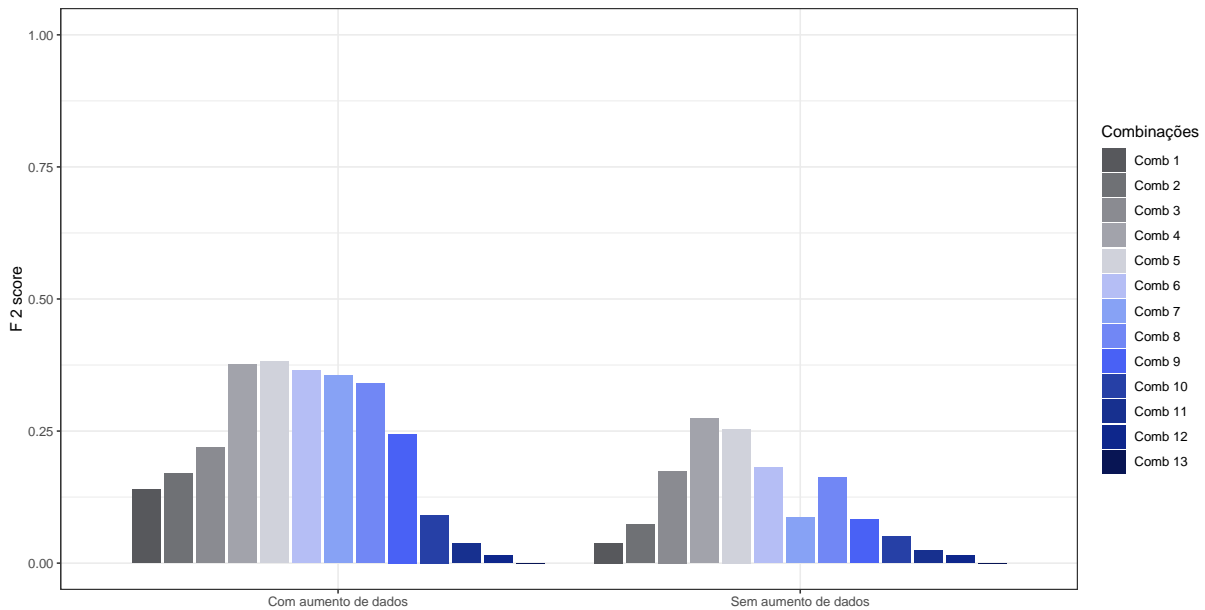
Figura 54 – Resultados do F2 score dos 13 modelos utilizados com o método *undersampling* (10%).



Fonte: Elaborada pelo autor.

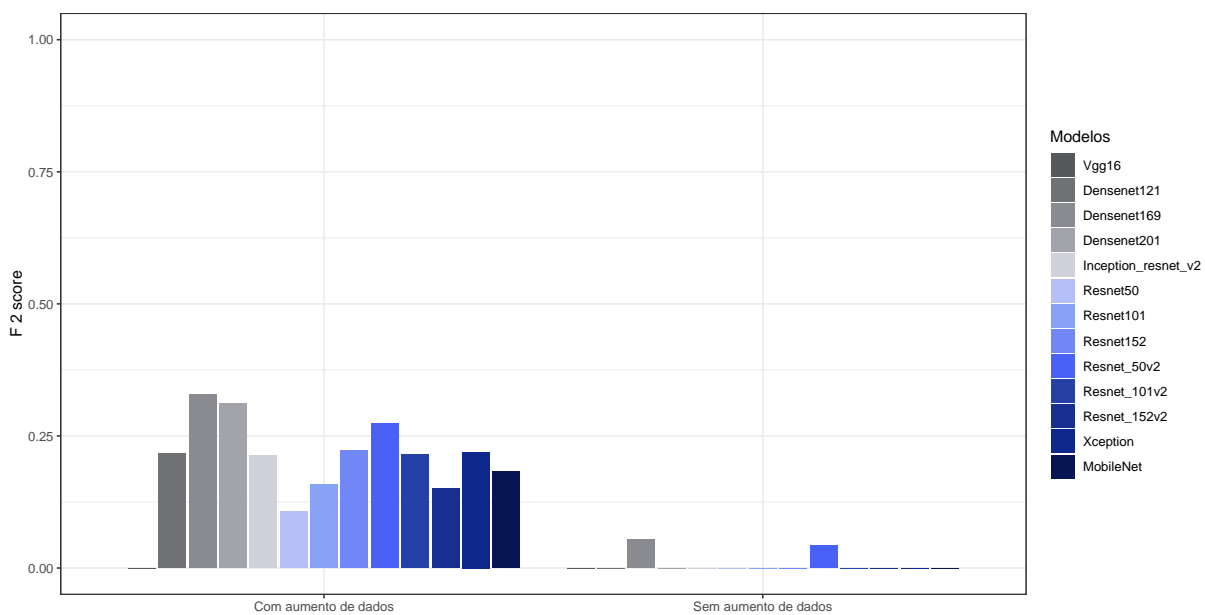
No cenário II, para os modelos individuais (Figura 54), observa-se que os modelos individuais tem um melhor F2 score do que o cenário I, com F2-scores um pouco mais altos. Já para as combinações Figura 55, a combinação 5 tem melhor F2 score com aumento de dados e a combinação 4 sem aumento de dados.

Figura 55 – Resultados do F2 score das 13 combinações utilizadas com o método undersampling (10%).



Fonte: Elaborada pelo autor.

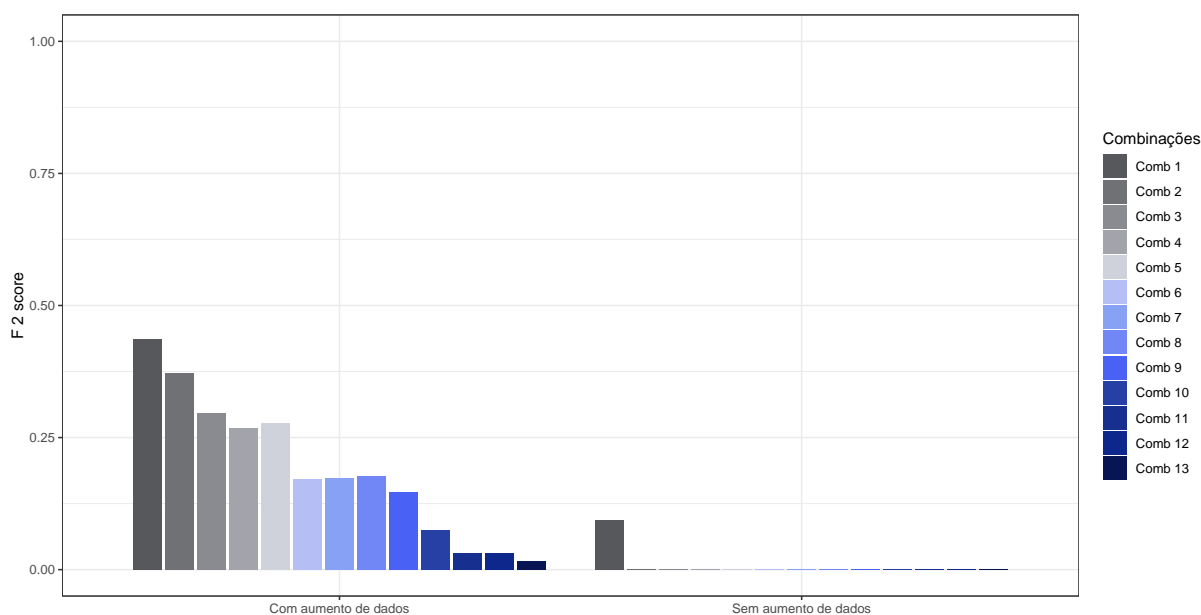
Figura 56 – Resultados do F2 score dos 13 modelos utilizados com o método *oversampling* (10%).



Fonte: Elaborada pelo autor.

No cenário III, para os modelos individuais (Figura 56), observa-se que no geral os modelos individuais tem um melhor F2 score do que o cenário I, com F2-scores um pouco mais altos, mas um desempenho um pouco pior que o cenário II. Já para as combinações (Figura 57), a combinação 1 tem melhor F2 score com aumento de dados.

Figura 57 – Resultados do F2 score das 13 combinações utilizadas com o método *oversampling* (10%).



Fonte: Elaborada pelo autor.

Analisando os três cenários em conjunto, é possível perceber que sem a utilização de reamostragem e aumento de dados há dificuldade na identificação de elementos da classe maligna. Ao aumentar os dados há uma melhoria na precisão desta classe. As combinações comparadas aos modelos individuais obtiveram melhores métricas em todos os cenários. A combinação com maior pontuação F2 está no cenário III com aumento de dados de aproximadamente 45%, a segunda combinação com maior pontuação F2 é encontrada no método de *undersampling* com aumento de dados de aproximadamente 38%.

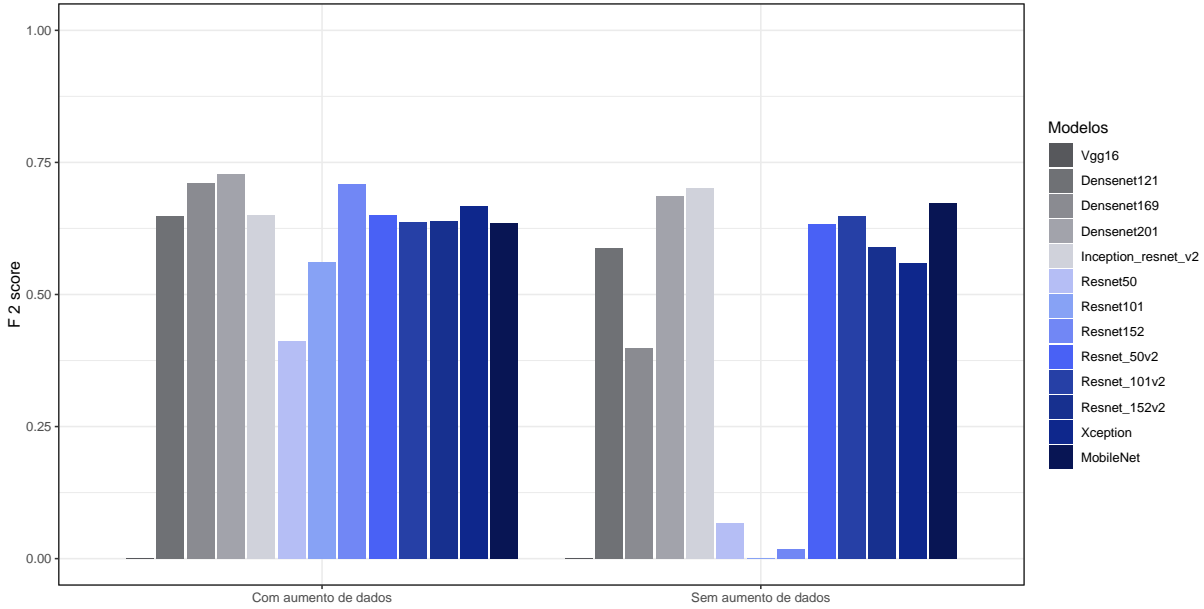
Portanto, conclui-se que os melhores F2 scores são encontrados com a utilização do ETCNN com reamostragem e aumento de dados. Em todos os cenários, o ETCNN obteve melhores pontuações F2 do que os modelos individuais com e sem utilização de métodos de reamostragem.

5.2.2 Simulação 3

Na simulação 3, temos 30% das imagens sendo malignas e 70% benignas. Os resultados dessa simulação são apresentados a seguir. No cenário I, para os modelos individuais (Figura 58), observa-se que os modelos tem uma maior facilidade em detectar a classe maligna, com

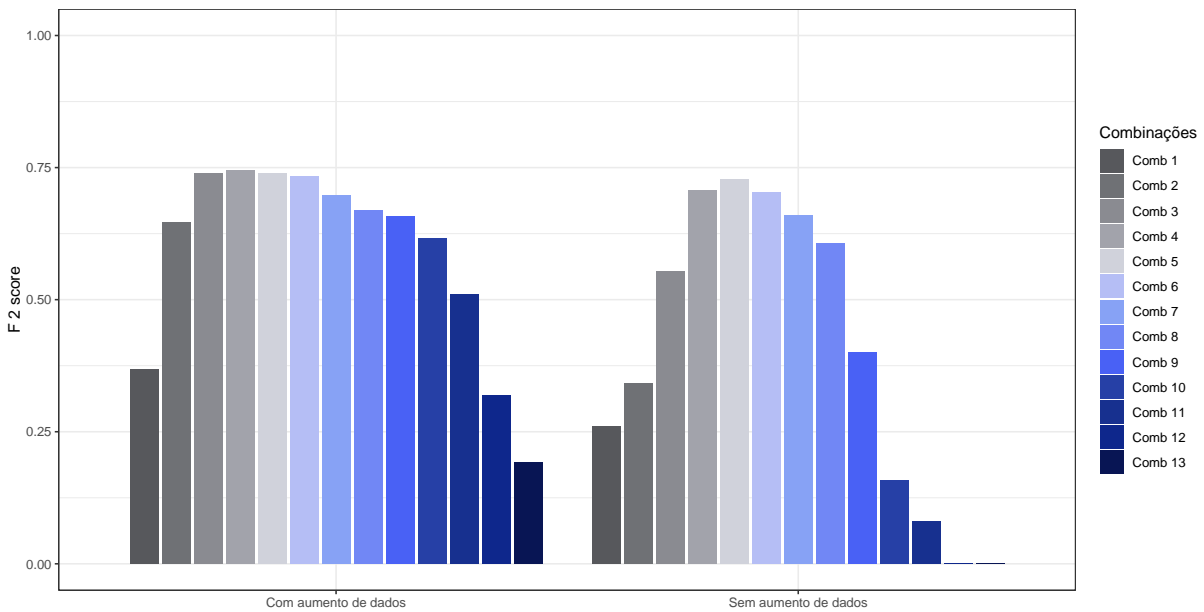
F2-scores maiores que na simulação anterior. Já para as combinações (Figura 59), a combinação 4 tem melhor F2 score com aumento de dados e a combinação 6 sem aumento de dados.

Figura 58 – Resultados do F2 score dos 13 modelos utilizados (30%).



Fonte: Elaborada pelo autor.

Figura 59 – Resultados do F2 score das 13 combinações utilizadas (30%).

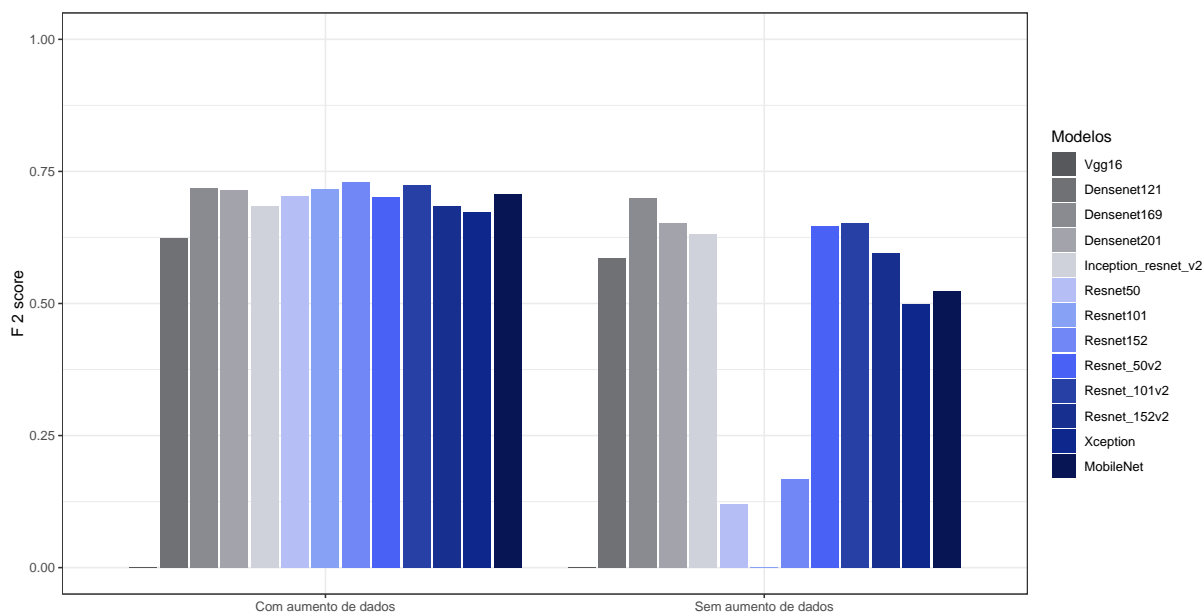


Fonte: Elaborada pelo autor.

No cenário II, para os modelos individuais (Figura 60), observa-se que os modelos no geral tem uma maior facilidade em detectar a classe maligna, com F2-scores maiores que no

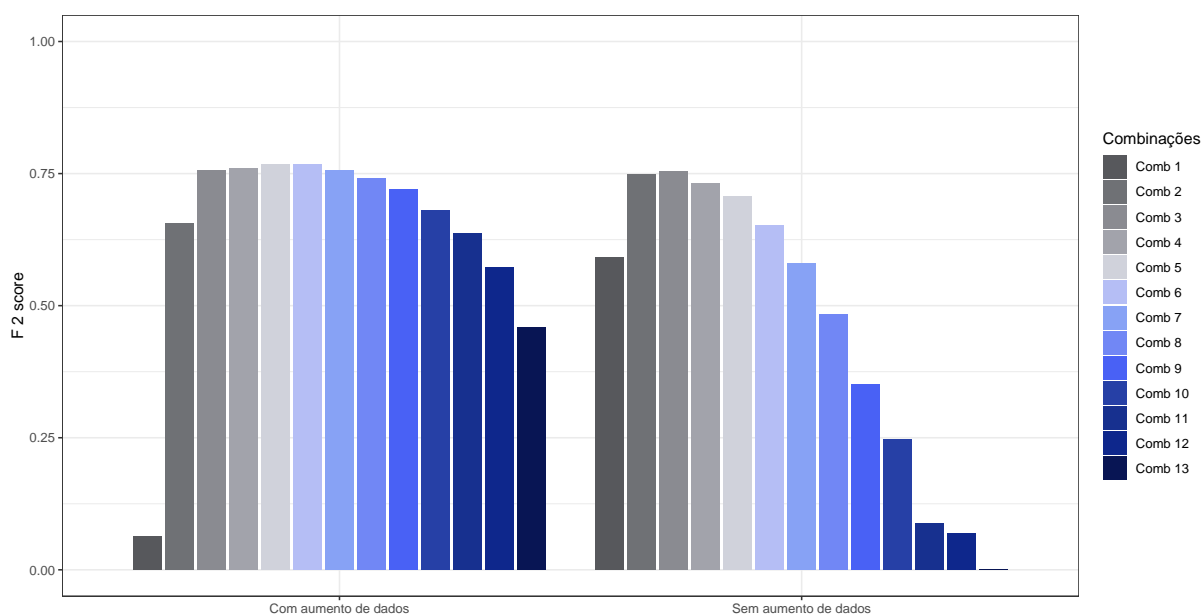
cenário sem reamostragem. Já para as combinações (Figura 61), a combinação 6 tem melhor F2 score com aumento de dados e a combinação 3 sem aumento de dados.

Figura 60 – Resultados do F2 score dos 13 modelos utilizados com o método undersampling (30%).



Fonte: Elaborada pelo autor.

Figura 61 – Resultados do F2 score das 13 combinações utilizadas com o método undersampling (30%).

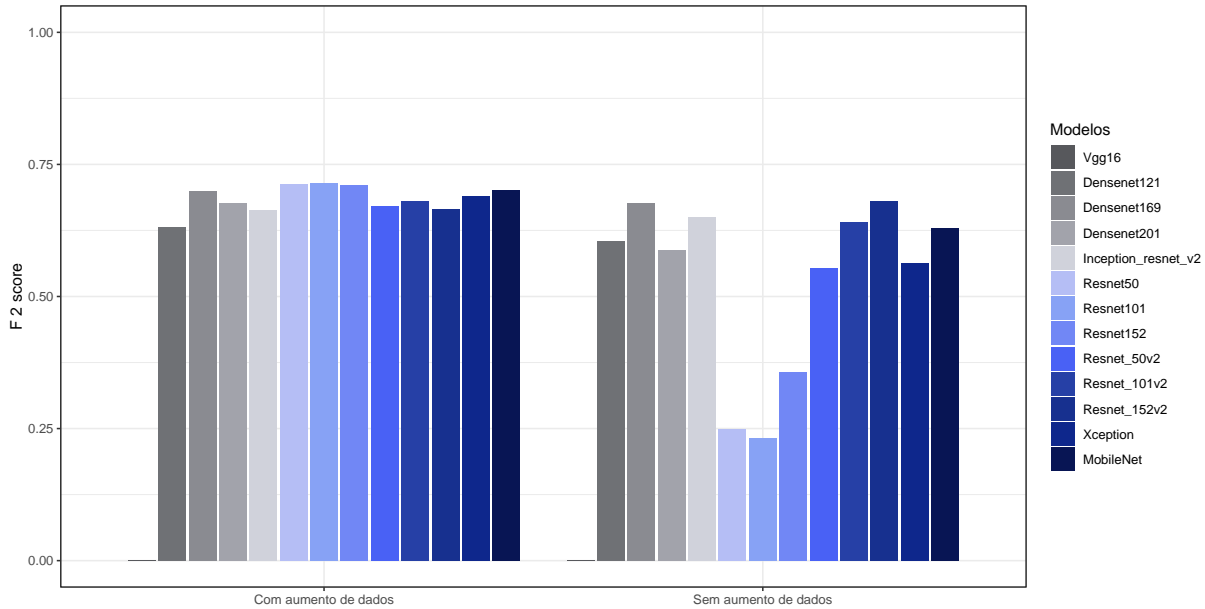


Fonte: Elaborada pelo autor.

No cenário III, para os modelos individuais (Figura 62), observa-se que os modelos tem uma maior facilidade em detectar a classe maligna, com F2-scores maiores que no cenário sem

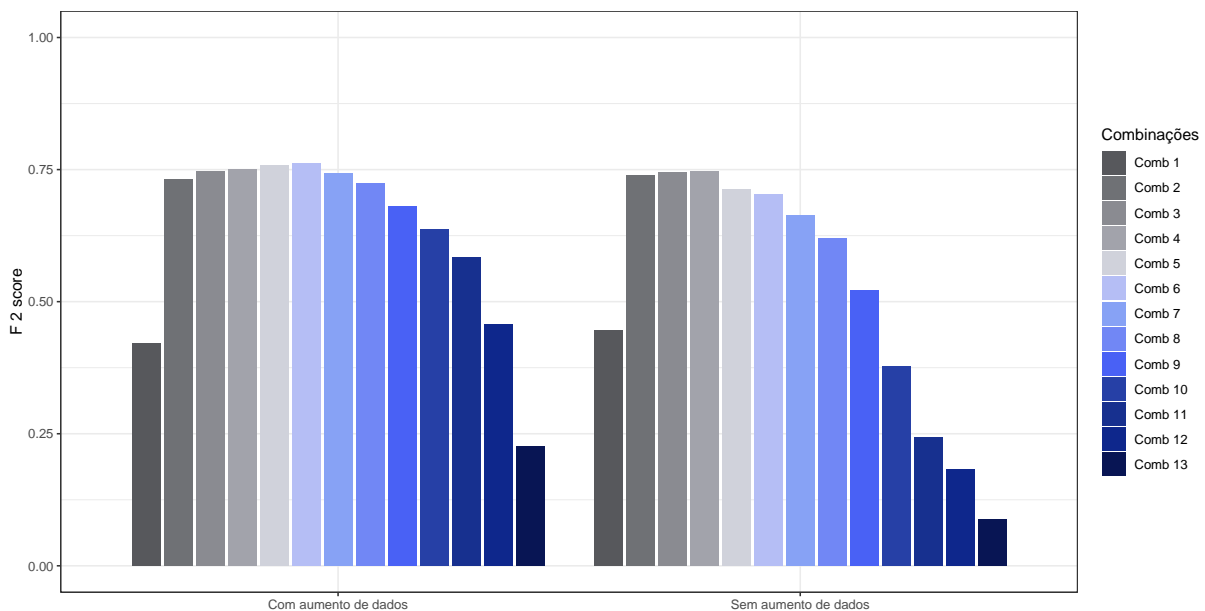
reamostragem. Já para as combinações (Figura 63), a combinação 6 tem melhor F2 score com aumento de dados e a combinação 4 sem aumento de dados.

Figura 62 – Resultados do F2 score dos 13 modelos utilizados com o método *oversampling* (30%).



Fonte: Elaborada pelo autor.

Figura 63 – Resultados do F2 score das 13 combinações utilizadas com o método *oversampling* (30%).



Fonte: Elaborada pelo autor.

Analisando os três cenários em conjunto, é possível perceber que sem a utilização de reamostragem e aumento de dados há dificuldade na identificação de elementos da classe maligna.

Ao aumentar os dados há uma melhoria na precisão desta classe. As combinações comparadas aos modelos individuais obtiveram melhores métricas em todos os cenários. A combinação com maior pontuação F2 está no cenário II com aumento de dados de aproximadamente 77%, a segunda combinação com maior pontuação F2 é encontrada no método de *oversampling* com aumento de dados de aproximadamente 76%.

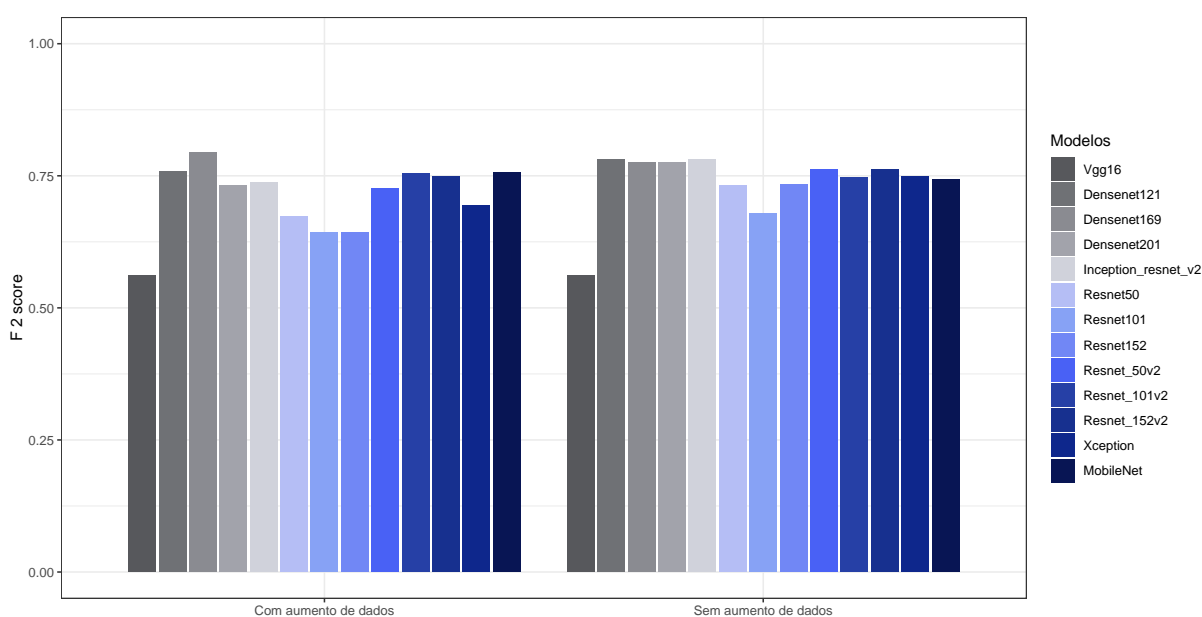
Portanto, conclui-se que os melhores F2 scores são encontrados com a utilização do ETCNN com reamostragem e aumento de dados. Em todos os cenários, o ETCNN obteve melhores pontuações F2 do que os modelos individuais com e sem utilização de métodos de reamostragem.

5.2.3 Simulação 4

Na simulação 4, temos 40% das imagens sendo malignas e 60% benignas. Os resultados dessa simulação são apresentados a seguir. A simulação 4 por ser um cenário quase balanceado não foi feito nenhum tipo de reamostragem, a Figura 64 mostra o resultados dos F2 scores para os modelos individuais, é possível notar com aumento de dados o melhor F2 score é obtido no modelo Densenet 169 e sem aumento de dados é o Densenet 121.

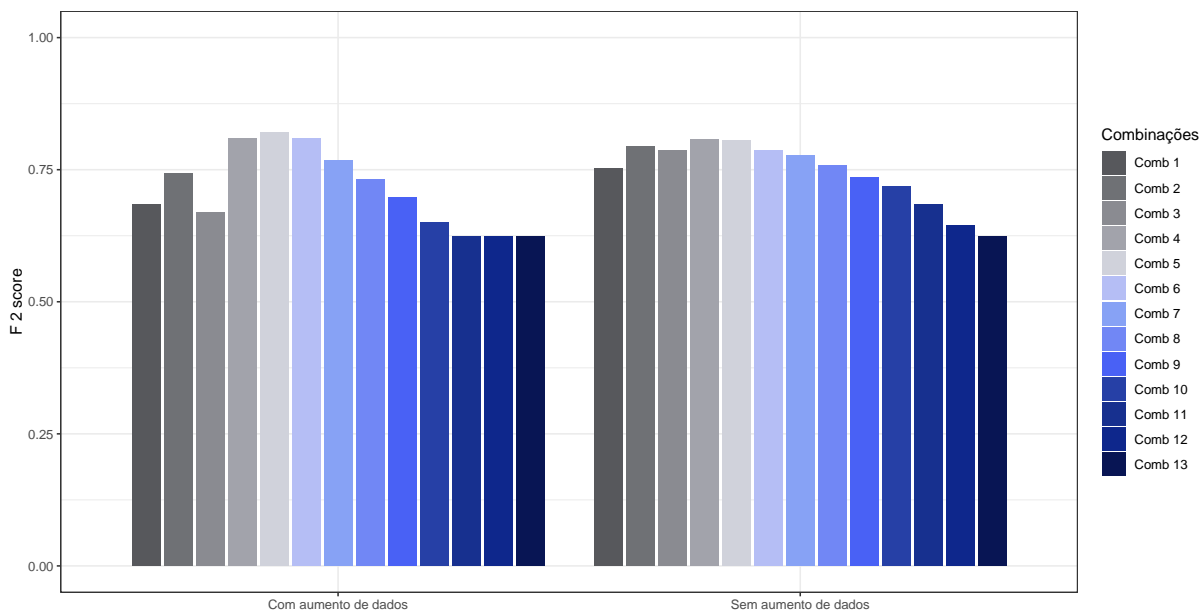
Já a Figura 65 apresenta os resultados das combinações, onde a melhor combinação com aumento de dados é a combinação 5 com aproximadamente 82% e a sem aumento de dados é a combinação 4.

Figura 64 – Resultados do F2 score dos 13 modelos utilizados (40%).



Fonte: Elaborada pelo autor.

Figura 65 – Resultados do F2 score das 13 combinações utilizadas (40%).



Fonte: Elaborada pelo autor.

Analisando os resultados, é possível perceber que sem a utilização de aumento de dados há dificuldade na identificação de elementos da classe maligna. As combinações comparadas aos modelos individuais obtiveram melhores métricas em todos os cenários. A combinação com maior pontuação F2 é com a utilização do aumento de dados de com aproximadamente 82%.

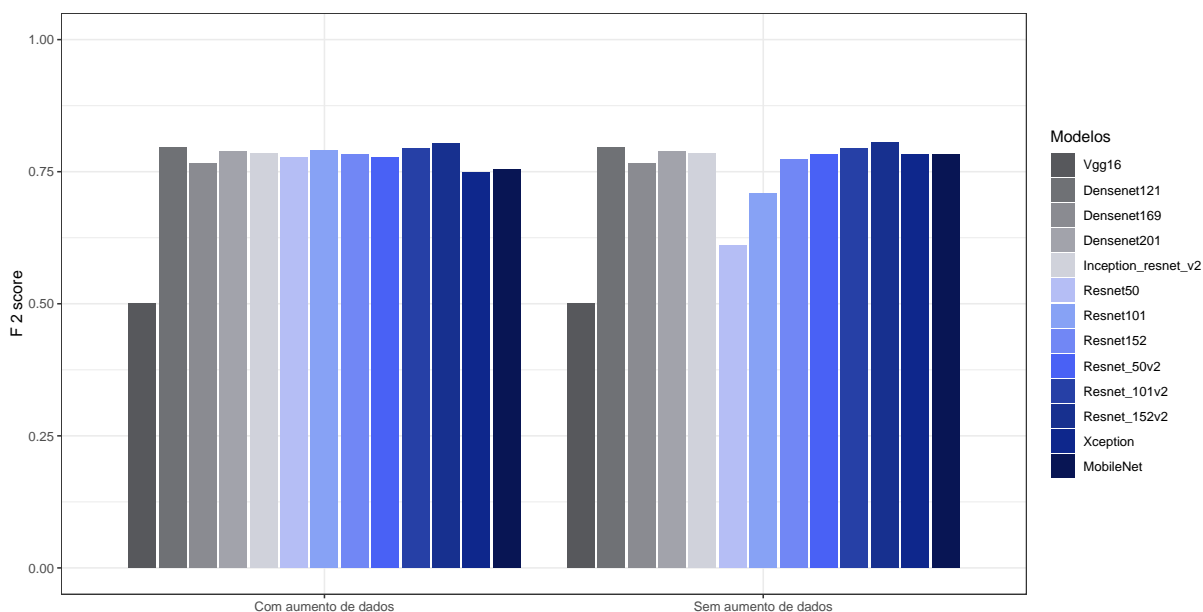
Portanto, conclui-se que com a utilização de aumento de dados as combinações tem melhor F2 score.

5.2.4 Simulação 5

Na simulação 5, temos 50% das imagens sendo malignas e 50% benignas. Os resultados dessa simulação são apresentados a seguir. A simulação 4 por ser um cenário quase balanceado não foi feito nenhum tipo de reamostragem, a Figura 66 mostra o resultados dos F2 scores para os modelos individuais, é possível notar que o Resnet 152 v2 tem o melhor F2 score com e sem aumento de dados

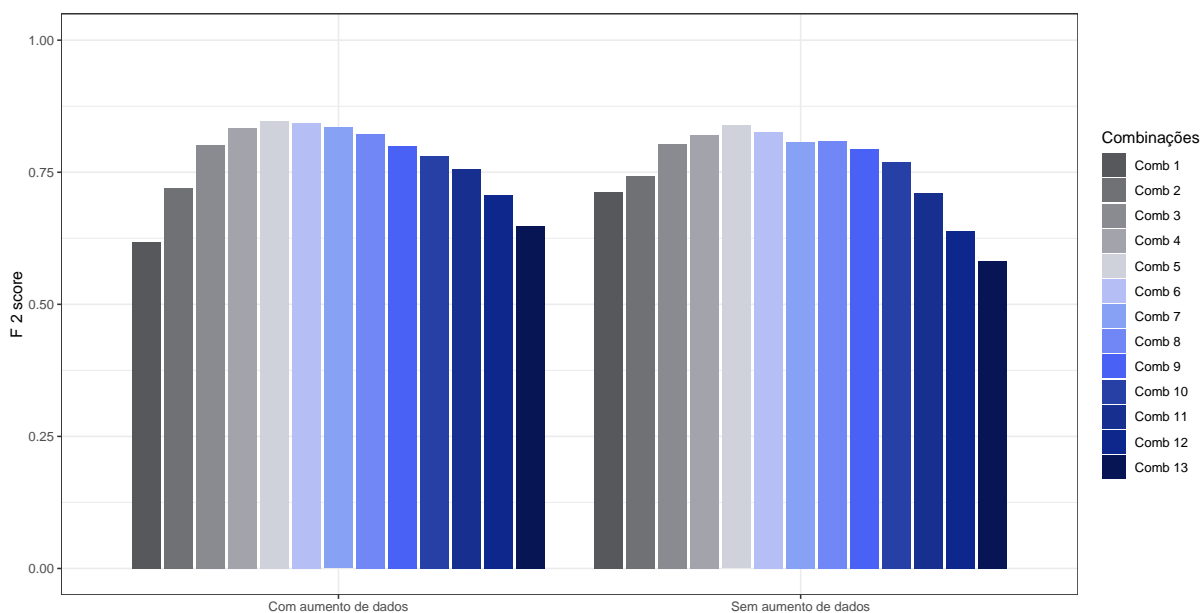
Já a Figura 67 apresenta os resultados das combinações, onde a melhor combinação com e sem aumento de dados é a combinação 5, tendo valores de F2 score bem próximos 84% e 83% aproximadamente.

Figura 66 – Resultados do F2 score dos 13 modelos utilizados (50%).



Fonte: Elaborada pelo autor.

Figura 67 – Resultados do F2 score das 13 combinações utilizadas (50%).



Fonte: Elaborada pelo autor.

Analisando os resultados, é possível perceber que sem o aumento de dados não há mais uma dificuldade na identificação de elementos da classe maligna. As combinações comparadas aos modelos individuais obtiveram melhores métricas. A combinação com maior pontuação F2 é com a utilização do aumento de dados de com aproximadamente 85%.

Portanto, conclui-se que com a utilização de aumento de dados as combinações tem melhor F2 score.

5.3 Conclusão

Esta seção apresentou os resultados das simulações que avaliam o impacto de técnicas de reamostragem e aumento de dados sobre o desempenho de modelos de transferência de aprendizado e suas combinações aplicados a cenários de diagnóstico de tumores, considerando as bases de dados do ISIC 2016, 2017 e 2018.

A principal contribuição desta simulação consiste na análise do comportamento das métricas de desempenho dos modelos individuais e das combinações, considerando diferentes estratégias de reamostragem *undersampling* e *oversampling* em cenários desbalanceados.

Observou-se que, nos cenários mais desbalanceados, os modelos individuais apresentaram dificuldade para prever observações da classe maligna (classe com menos observações), o que indica a necessidade de técnicas adicionais para equilibrar o desempenho entre as classes. O uso de combinações de modelos com técnicas de aumento de dados e reamostragem mostrou-se particularmente eficaz na melhora das métricas de F2 score (que é sensível à detecção de falsos negativos), ou seja, reduzir o número de diagnósticos considerados benignos quando na realidade são malignos.

Quando a base de dados é mais balanceada (Simulação 4 e 5), os modelos tendem a melhorar naturalmente o desempenho na detecção de tumores malignos visto que há novas imagens da classe maligna com novos padrões. No entanto as combinações de modelos superaram os modelos individuais, especialmente com o uso de aumento de dados. Isso sugere que, mesmo em cenários próximos do balanceamento, a aplicação de técnicas de *ensemble* e aumento de dados contribui para um melhor desempenho na predição final.

O treinamento dos modelos foi realizado em GPUs do Google Colab. No entanto, devido às limitações da versão gratuita, optamos por utilizar o Google Colab Pro+ para garantir um melhor desempenho e maior tempo de execução dos experimentos.

No total foram ajustados 1.430 modelos sem aumento de dados, considerando três simulações com reamostragem ($130 \times 3 \times 3 = 1.170$) e duas simulações sem reamostragem ($130 \times 2 = 260$). Com a aplicação de técnicas de aumento de dados, esse número dobrou, totalizando 2.860 modelos analisados ao longo do estudo.

Os resultados deste estudo indicaram que determinadas combinações de modelos apresentaram métricas superiores em comparação com abordagens individuais. Na detecção de câncer em imagens médicas, a minimização de falsos negativos é crucial, pois um erro nesse sentido pode levar a um diagnóstico tardio e consequências clínicas graves. Assim, o desempenho de um modelo deve ser avaliado além de métricas tradicionais, considerando sua aplicabilidade em

ambientes médicos reais.

Além disso, o uso de reamostragem e aumento de dados demonstrou benefícios no equilíbrio entre classes, mas a generalização para novas bases de dados ainda precisa ser validada em estudos futuros, especialmente quando aplicado a diferentes modalidades de exames médicos.

Um bom modelo para a detecção de câncer não é necessariamente aquele com as melhores métricas, mas sim aquele que equilibra desempenho, interpretabilidade e aplicabilidade clínica. Esse trabalho contribui para essa discussão ao propor uma abordagem que não apenas melhora métricas quantitativas, mas também busca estratégias que aumentem a confiabilidade dos modelos em cenários médicos reais.

CONSIDERAÇÕES FINAIS E PESQUISAS FUTURAS

6.1 Considerações Finais

A detecção precoce de câncer a partir de imagens médicas representa um grande desafio devido à escassez de dados e ao desbalanceamento dessas bases. Neste trabalho foi proposto um modelo para prever observações da classe minoritária em bases de dados desbalanceadas. A abordagem empregada, baseada em um *ensemble* de modelos de transferência de aprendizado combinado com reamostragem e aumento de dados, demonstrou uma melhoria da predição de casos malignos (classe minoritária). Ao longo dos experimentos, verificou-se que, no geral na maioria dos cenários analisados a utilização dessas estratégias resultou em métricas superiores em comparação aos modelos individuais, especialmente em termos de F2-score.

A detecção precoce de tumores malignos pode salvar vidas, a metodologia proposta busca minimizar a ocorrência de falsos negativos que é um aspecto crítico para diagnósticos mais confiáveis. Mesmo utilizando bases de dados limitadas, a estratégia adotada neste estudo conseguiu alcançar melhores F2-scores nas combinações, reforçando a importância da combinação de técnicas para superar os desafios impostos pelo desbalanceamento dos dados.

Além disso, observou-se que a reamostragem e o aumento de dados também impactam a capacidade do modelo em identificar corretamente a classe minoritária. Os resultados indicam que, em cenários mais desbalanceados, a combinação dessas técnicas proporciona uma melhora expressiva na capacidade de predição. Já em cenários mais balanceados, as combinações de modelos mantiveram sua eficácia, o que sugere que a combinação com aumento de dados e reamostragem é uma alternativa robusta para diferentes contextos de diagnóstico por imagem.

6.2 Sugestões para pesquisas futuras

Há várias pesquisas que podem dar continuidade ao trabalho desenvolvido nesta tese:

Primeiramente, a realização de estudos adicionais sobre o modelo desenvolvido, com foco em uma análise mais aprofundada das técnicas de *ensemble*, incluindo a utilização de outras funções generalizadoras, como o *stacking*. Além disso, recomenda-se explorar novos métodos de reamostragem.

Outra possibilidade é ampliar o modelo com a inclusão de mais modelos de transferência de aprendizado, bem como integrar modelos mais recentes. Ainda, propõe-se a implementação de modelos de transferência de aprendizado já pré-treinados especificamente no contexto médico.

Também seria interessante avaliar o impacto do uso de imagens com resoluções mais altas, o que pode contribuir para a extração de características mais detalhadas.

Por fim, considerar métricas de custo-benefício na avaliação dos modelos pode fornecer uma abordagem mais realista e aplicável, auxiliando na escolha da melhor estratégia para diferentes cenários clínicos e computacionais.

REFERÊNCIAS

ABDANI, S. R.; ZULKIFLEY, M. A. Densenet with spatial pyramid pooling for industrial oil palm plantation detection. In: IEEE. **2019 International conference on mechatronics, robotics and systems engineering (MoRSE)**. [S.l.], 2019. p. 134–138. Citado na página [56](#).

AHMAD, F.; KHAN, M. U. G.; JAVED, K. Deep learning model for distinguishing novel coronavirus from other chest related infections in x-ray images. **Computers in biology and medicine**, Elsevier, v. 134, p. 104401, 2021. Citado na página [29](#).

ALAIZ-MORETON, H.; AVELEIRA-MATA, J.; ONDICOL-GARCIA, J.; MUÑOZ-CASTAÑEDA, A. L.; GARCÍA, I.; BENAVIDES, C. Multiclass classification procedure for detecting attacks on mqtt-iot protocol. **Complexity**, Hindawi, v. 2019, 2019. Citado na página [49](#).

ALI-GOMBE, A.; ELYAN, E. Mfc-gan: Class-imbalanced dataset classification using multiple fake class generative adversarial network. **Neurocomputing**, Elsevier, v. 361, p. 212–221, 2019. Citado na página [51](#).

ALI, R.; MANIKANDAN, A.; LEI, R.; XU, J. A novel spsa based hyper-parameter optimized fcedn with adaptive cnn classification for skin cancer detection. **Scientific Reports**, Nature Publishing Group UK London, v. 14, n. 1, p. 9336, 2024. Citado nas páginas [26](#) e [39](#).

ALIZADEH, S. M.; MAHLOOJIFAR, A. Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features. **International Journal of Imaging Systems and Technology**, Wiley Online Library, v. 31, n. 2, p. 695–707, 2021. Citado na página [79](#).

ANEJA, N.; ANEJA, S. Transfer learning using cnn for handwritten devanagari character recognition. In: IEEE. **2019 1st International Conference on Advances in Information Technology (ICAIT)**. [S.l.], 2019. p. 293–296. Citado nas páginas [24](#) e [47](#).

ANGUITA, D.; GHELARDONI, L.; GHIO, A.; ONETO, L.; RIDELLA, S. *et al.* The 'k' in k-fold cross validation. In: **ESANN**. [S.l.: s.n.], 2012. v. 102, p. 441–446. Citado na página [59](#).

ANTIPOV, G.; BERRANI, S.-A.; DUGELAY, J.-L. Minimalistic cnn-based ensemble model for gender prediction from face images. **Pattern recognition letters**, Elsevier, v. 70, p. 59–65, 2016. Citado nas páginas [25](#) e [46](#).

ARAÚJO, F. H.; CARNEIRO, A. C.; SILVA, R. R.; MEDEIROS, F. N.; USHIZIMA, D. M. Redes neurais convolucionais com tensorflow: Teoria e prática. **SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. III Escola Regional de Informática do Piauí. Livro Anais-Artigos e Minicursos**, Sociedade Brasileira de Computação, v. 1, p. 382–406, 2017. Citado nas páginas [38](#) e [45](#).

AUZINE, M. M.; KHAN, M. H.-M.; BAICHO, S.; SAHIB, N. G.; BISSOONAUTH-DAIBOO, P.; GAO, X.; HEETUN, Z. Development of an ensemble cnn model with explainable ai for the

- classification of gastrointestinal cancer. **Plos one**, Public Library of Science San Francisco, CA USA, v. 19, n. 6, p. e0305628, 2024. Citado nas páginas 26 e 39.
- BABINI, M.; MARRANGHELLO, N. **Introdução às Redes Neurais Artificiais**. 1. ed. [S.l.]: Cultura Acadêmica Editora, 2008. Citado na página 32.
- BACKES, A. R.; JUNIOR, J. J. d. M. S. **Introdução à visão computacional usando Matlab**. [S.l.]: Alta Books Editora, 2019. Citado na página 23.
- BEGHRICHE, T.; DJERIOUI, M.; BRIK, Y.; ATTALLAH, B. Deep-modified transfer learning-based cnn networks for enhanced breast cancer prediction. **Studies in Engineering and Exact Sciences**, v. 5, n. 2, p. e5383–e5383, 2024. Citado nas páginas 27 e 45.
- BUCIU, I.; GACSADI, A. Directional features for automatic tumor classification of mammogram images. **Biomedical Signal Processing and Control**, Elsevier, v. 6, n. 4, p. 370–378, 2011. Citado na página 24.
- BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. **Neural networks**, Elsevier, v. 106, p. 249–259, 2018. Citado na página 48.
- BUENO, E. I. *et al.* **Utilização de Redes Neurais Artificiais na Monitoração e Detecção de Falhas em sensores do reator IEA-R1**. Tese (Doutorado) — Dissertação (Mestrado), Universidade de São Paulo–IPEN, 98 p., São Paulo, 2006. Citado nas páginas 36 e 37.
- BURNHAM, J. F. Scopus database: a review. **Biomedical digital libraries**, BioMed Central, v. 3, n. 1, p. 1–8, 2006. Citado na página 27.
- CALAS, M. J. G.; GUTFILEN, B.; PEREIRA, W. C. de A. Cad e mamografia: por que usar esta ferramenta? **Radiologia Brasileira**, Radiologia Brasileira, v. 45, n. 1, p. 46–52, 2012. Citado na página 23.
- CALDERON-RAMIREZ, S.; MURILLO-HERNANDEZ, D.; ROJAS-SALAZAR, K.; ELIZONDO, D.; YANG, S.; MOEMENI, A.; MOLINA-CABELLO, M. A real use case of semi-supervised learning for mammogram classification in a local clinic of costa rica. **Medical & biological engineering & computing**, Springer, v. 60, n. 4, p. 1159–1175, 2022. Citado na página 72.
- CAO, K.; CHOI, K.-n.; JUNG, H.; DUAN, L. Deep learning for facial beauty prediction. **Information**, Multidisciplinary Digital Publishing Institute, v. 11, n. 8, p. 391, 2020. Citado na página 38.
- CASSIDY, B.; KENDRICK, C.; BRODZICKI, A.; JAWOREK-KORJAKOWSKA, J.; YAP, M. H. Analysis of the isic image datasets: Usage, benchmarks and recommendations. **Medical image analysis**, Elsevier, v. 75, p. 102305, 2022. Citado na página 79.
- CELIK, A. Diagnosis of the diseases using resampling methods with machine learning algorithms. In: **Proceedings of the Bulgarian Academy of Sciences**. [S.l.: s.n.], 2023. v. 76, n. 7, p. 1065–1076. Citado na página 26.
- CHAKRAVARTHY, S.; BHARANIDHARAN, N.; KHAN, S. B.; KUMAR, V. V.; MAHESH, T.; ALMUSHARRAF, A.; ALBALAWI, E. Multi-class breast cancer classification using cnn features hybridization. **International Journal of Computational Intelligence Systems**, Springer, v. 17, n. 1, p. 191, 2024. Citado nas páginas 26 e 39.

CHATTERJEE, S.; HAZRA, D.; BYUN, Y.-C.; KIM, Y.-W. Enhancement of image classification using transfer learning and gan-based synthetic data augmentation. **Mathematics**, MDPI, v. 10, n. 9, p. 1541, 2022. Citado na página 26.

CHAVES, E. d. L. Detecção de câncer de mama por meio de imagens infravermelhas utilizando redes neurais convolucionais. **Universidade Federal de Uberlândia**, 2019. Citado na página 24.

CHEN, L.; SHAKHNAROVICH, G. Learning ensembles of convolutional neural networks. **The University of Chicago**, 2014. Citado na página 25.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 1251–1258. Citado na página 55.

CHOLLET, F. *et al.* **Keras**. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>. Citado na página 55.

CODELLA, N. C.; NGUYEN, Q.-B.; PANKANTI, S.; GUTMAN, D. A.; HELBA, B.; HALPERN, A. C.; SMITH, J. R. Deep learning ensembles for melanoma recognition in dermoscopy images. **IBM Journal of Research and Development**, IBM, v. 61, n. 4/5, p. 5–1, 2017. Citado na página 79.

DAI, W.; LI, D.; TANG, D.; WANG, H.; PENG, Y. Deep learning approach for defective spot welds classification using small and class-imbalanced datasets. **Neurocomputing**, Elsevier, v. 477, p. 46–60, 2022. Citado na página 51.

DALVI, R. **Detecção e Classificação de Câncer a partir de Mamografias Digitalizadas e Redes Neurais Convolucionais**. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2018. Citado na página 23.

DARWISH, A.; EZZAT, D.; HASSANIEN, A. E. An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. **Swarm and evolutionary computation**, Elsevier, v. 52, p. 100616, 2020. Citado na página 48.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: **IEEE. 2009 IEEE conference on computer vision and pattern recognition**. [S.l.], 2009. p. 248–255. Citado na página 55.

DEVARRIYA, D.; GULATI, C.; MANSHARAMANI, V.; SAKALLE, A.; BHARDWAJ, A. Unbalanced breast cancer data classification using novel fitness functions in genetic programming. **Expert Systems with Applications**, Elsevier, v. 140, p. 112866, 2020. Citado nas páginas 65, 71 e 72.

DUCHI, J.; HAZAN, E.; SINGER, Y. Adaptive subgradient methods for online learning and stochastic optimization. **Journal of machine learning research**, v. 12, n. 7, 2011. Citado na página 37.

EBERMAM, E.; KROHLING, R. Uma introdução compreensiva às redes neurais convolucionais: Um estudo de caso para reconhecimento de caracteres alfabéticos. **Revista de Sistemas de Informação da FSMA**, n. 22, p. 49–59, 2016. Citado nas páginas 39, 43 e 49.

- FAUSETT, L. V. *et al.* **Fundamentals of neural networks: architectures, algorithms, and applications**. [S.l.]: prentice-Hall Englewood Cliffs, 1994. v. 3. Citado na página 31.
- FERNANDES, B. J. T. Redes neurais com extração implícita de características para reconhecimento de padrões visuais. Universidade Federal de Pernambuco, 2013. Citado na página 23.
- FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, SciELO Brasil, v. 35, n. 1, 2006. Citado na página 31.
- FERREIRA, A. E. T. Estimacão do ângulo de direção por vídeo para veículos autônomos utilizando redes neurais convolucionais multicanais. 2017. Citado nas páginas 33, 41 e 45.
- FERREIRA, C. A.; MELO, T.; SOUSA, P.; MEYER, M. I.; SHAKIBAPOUR, E.; COSTA, P.; CAMPILHO, A. Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In: SPRINGER. **International conference image analysis and recognition**. [S.l.], 2018. p. 763–770. Citado na página 56.
- FERREIRA, E. J. **Método baseado em rotação e projeção otimizadas para a construção de ensembles de modelos**. Tese (Doutorado) — Universidade de São Paulo, 2012. Citado nas páginas 25 e 45.
- FERREIRA, M. D. **Designing convolutional neural network architectures based on dynamical system concepts**. Tese (Doutorado) — Universidade de São Paulo, 2019. Citado na página 31.
- FREITAS, A. d. Radiologia odontológica. In: **Radiologia odontológica**. [S.l.: s.n.], 1988. p. 610–610. Citado na página 24.
- FREITAS, R. d. S. Desenvolvimento de modelo de rede neural convolucional para classificação de cáries dentárias: abordagem usando algoritmo genético como otimizador. 2018. Citado na página 39.
- GAMA, J.; CARVALHO, A. C. P. d. L. F. d.; FACELI, K.; LORENA, A. C.; OLIVEIRA, M. Extração de conhecimento de dados: data mining. 2012. Citado na página 25.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado nas páginas 25, 26, 31, 36, 38, 39, 40, 41, 43, 44 e 49.
- GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAI, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. **Advances in neural information processing systems**, v. 27, 2014. Citado na página 51.
- GÜLTEKIN, T.; UĞUR, A. A hybrid approach based on transfer and ensemble learning for improving performances of deep learning models on small datasets. **Turkish Journal of Electrical Engineering and Computer Sciences**, v. 29, n. 7, p. 3197–3211, 2021. Citado na página 29.
- GUTMAN, D.; CODELLA, N. C.; CELEBI, E.; HELBA, B.; MARCHETTI, M.; MISHRA, N.; HALPERN, A. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). **arXiv preprint arXiv:1605.01397**, 2016. Citado nas páginas 72, 73, 74 e 79.

HASAN, M. A.; HAQUE, F.; SABUJ, S. R.; SARKER, H.; GONI, M. O. F.; RAHMAN, F.; RASHID, M. M. An end-to-end lightweight multi-scale cnn for the classification of lung and colon cancer with xai integration. **Technologies**, MDPI, v. 12, n. 4, p. 56, 2024. Citado nas páginas 26 e 39.

HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 1999. Citado nas páginas 32, 33, 34, 36 e 37.

HAYKIN, S.; NETWORK, N. A comprehensive foundation. **Neural networks**, v. 2, n. 2004, p. 41, 2004. Citado nas páginas 31, 33 e 36.

HE, H.; GARCIA, E. A. Learning from imbalanced data. **IEEE Transactions on knowledge and data engineering**, Ieee, v. 21, n. 9, p. 1263–1284, 2009. Citado na página 47.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778. Citado nas páginas 24 e 55.

_____. Identity mappings in deep residual networks. In: SPRINGER. **European conference on computer vision**. [S.l.], 2016. p. 630–645. Citado nas páginas 55 e 56.

HEYBURN, R.; BOND, R. R.; BLACK, M.; MULVENNA, M.; WALLACE, J.; RANKIN, D.; CLELAND, B. Machine learning using synthetic and real data: similarity of evaluation metrics for different healthcare datasets and for different algorithms. In: WORLD SCIENTIFIC. **Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018)**. [S.l.], 2018. p. 1281–1291. Citado na página 48.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017. Citado na página 55.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 4700–4708. Citado nas páginas 24 e 55.

HUANG, Y.; JIN, Y.; LI, Y.; LIN, Z. Towards imbalanced image classification: a generative adversarial network ensemble learning method. **IEEE Access**, IEEE, v. 8, p. 88399–88409, 2020. Citado na página 51.

IDREES, F.; RAJARAJAN, M.; CONTI, M.; CHEN, T. M.; RAHULAMATHAVAN, Y. Pindroid: A novel android malware detection system using ensemble learning methods. **Computers & Security**, Elsevier, v. 68, p. 36–46, 2017. Citado na página 48.

INCA. **Instituto Nacional do Câncer**. 2019. <<https://www.inca.gov.br/>>. Acessado: 11-01-2020. Citado na página 24.

IZBICKI, R.; SANTOS, T. M. dos. **Aprendizado de máquina: uma abordagem estatística**. 2020. Citado na página 36.

JENI, L. A.; COHN, J. F.; TORRE, F. D. L. Facing imbalanced data—recommendations for the use of performance metrics. In: IEEE. **2013 Humaine association conference on affective computing and intelligent interaction**. [S.l.], 2013. p. 245–251. Citado na página 48.

- KARTHIKEYAN, A.; JOTHILAKSHMI, S.; SUTHIR, S. Colorectal cancer detection based on convolutional neural networks (cnn) and ranking algorithm. **Measurement: Sensors**, Elsevier, v. 31, p. 100976, 2024. Citado nas páginas 26 e 39.
- KASSANI, S. H.; KASSANI, P. H.; KHAZAEINEZHAD, R.; WESOLOWSKI, M. J.; SCHNEIDER, K. A.; DETERS, R. Diabetic retinopathy classification using a modified xception architecture. In: IEEE. **2019 IEEE international symposium on signal processing and information technology (ISSPIT)**. [S.l.], 2019. p. 1–6. Citado na página 56.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. Citado nas páginas 24 e 46.
- KUMAR, A.; KUMAR, M.; BHARDWAJ, V. P.; KUMAR, S.; SELVARAJAN, S. A novel skin cancer detection model using modified finch deep cnn classifier model. **Scientific Reports**, Nature Publishing Group UK London, v. 14, n. 1, p. 11235, 2024. Citado nas páginas 26 e 39.
- LAPPALAINEN, H.; MISKIN, J. W. Ensemble learning. In: **Advances in independent component analysis**. [S.l.]: Springer, 2000. p. 75–92. Citado na página 46.
- LEE, B.; LEE, J.; LEE, J.-O.; HWANG, Y.; BAHN, H.-K.; PARK, I.; JHEON, S.; LEE, D.-S. Breath analysis system with convolutional neural network (cnn) for early detection of lung cancer. **Sensors and Actuators B: Chemical**, Elsevier, v. 409, p. 135578, 2024. Citado nas páginas 26 e 39.
- LEE, J.; PARK, K. Gan-based imbalanced data intrusion detection system. **Personal and Ubiquitous Computing**, Springer, v. 25, p. 121–128, 2021. Citado na página 51.
- LEE, N.; YANG, H.; YOO, H. A surrogate loss function for optimization of f beta score in binary classification with imbalanced data. **arXiv preprint arXiv:2104.01459**, 2021. Citado na página 49.
- LEIBIG, C.; BREHMER, M.; BUNK, S.; BYNG, D.; PINKER, K.; UMUTLU, L. Combining the strengths of radiologists and ai for breast cancer screening: a retrospective analysis. **The Lancet Digital Health**, Elsevier, v. 4, n. 7, p. e507–e519, 2022. Citado na página 23.
- LI, W.; CHEN, C.; ZHANG, M.; LI, H.; DU, Q. Data augmentation for hyperspectral image classification with deep cnn. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 16, n. 4, p. 593–597, 2018. Citado na página 26.
- LIN, H.-I.; LIN, P.-Y. An image quality assessment method for surface defect inspection. In: IEEE. **2020 IEEE International Conference On Artificial Intelligence Testing (AITest)**. [S.l.], 2020. p. 1–6. Citado na página 65.
- LIN, H.-I.; WIBOWO, F. S. Image data assessment approach for deep learning-based metal surface defect-detection systems. **IEEE Access**, IEEE, v. 9, p. 47621–47638, 2021. Citado na página 65.
- LIU, S.; WANG, Y.; YANG, X.; LEI, B.; LIU, L.; LI, S. X.; NI, D.; WANG, T. Deep learning in medical ultrasound analysis: a review. **Engineering**, Elsevier, v. 5, n. 2, p. 261–275, 2019. Citado na página 38.

- LOPEZ, A. R.; NIETO, X. Giro-i; BURDICK, J.; MARQUES, O. Skin lesion classification from dermoscopic images using deep learning techniques. In: IEEE. **2017 13th IASTED international conference on biomedical engineering (BioMed)**. [S.l.], 2017. p. 49–54. Citado na página 79.
- MAHORO, E.; AKHLOUFI, M. A. Breast cancer classification on thermograms using deep cnn and transformers. **Quantitative InfraRed Thermography Journal**, Taylor & Francis, v. 21, n. 1, p. 30–49, 2024. Citado nas páginas 26 e 39.
- MAJUMDER, S.; GAUTAM, N.; BASU, A.; SAU, A.; GEEM, Z. W.; SARKAR, R. Menet: A mitscherlich function based ensemble of cnn models to classify lung cancer using ct scans. **Plos one**, Public Library of Science San Francisco, CA USA, v. 19, n. 3, p. e0298527, 2024. Citado nas páginas 26 e 39.
- MATTA, T. F. da. **Novos modelos de regressao binária usando funções de ligação simétricas e assimétricas**. Tese (Doutorado) — Universidade Federal da Bahia, 2021. Citado na página 48.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, p. 115–133, 1943. Citado na página 32.
- MENARDI, G.; TORELLI, N. Training and assessing classification rules with imbalanced data. **Data mining and knowledge discovery**, Springer, v. 28, n. 1, p. 92–122, 2014. Citado na página 48.
- MENDES, J. P. C. *et al.* Segmentação automática de lesões de mama em imagens de ultrassom utilizando redes neurais convolutivas. Universidade Federal do Amazonas, 2018. Citado na página 39.
- MENEGOLA, A.; FORNACIALI, M.; PIRES, R.; BITTENCOURT, F. V.; AVILA, S.; VALLE, E. Knowledge transfer for melanoma screening with deep learning. In: IEEE. **2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)**. [S.l.], 2017. p. 297–300. Citado na página 45.
- MINETTO, R.; SEGUNDO, M. P.; SARKAR, S. Hydra: An ensemble of convolutional neural networks for geospatial land classification. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 57, n. 9, p. 6530–6541, 2019. Citado nas páginas 25, 29 e 46.
- MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. **arXiv preprint arXiv:1411.1784**, 2014. Citado na página 51.
- MOHAMMED, R.; RAWASHDEH, J.; ABDULLAH, M. Machine learning with oversampling and undersampling techniques: overview study and experimental results. In: IEEE. **2020 11th international conference on information and communication systems (ICICS)**. [S.l.], 2020. p. 243–248. Citado na página 25.
- MONDAL, C.; HASAN, M. K.; AHMAD, M.; AWAL, M. A.; JAWAD, M. T.; DUTTA, A.; ISLAM, M. R.; MONI, M. A. Ensemble of convolutional neural networks to diagnose acute lymphoblastic leukemia from microscopic images. **Informatics in Medicine Unlocked**, Elsevier, v. 27, p. 100794, 2021. Citado na página 29.

- MORID, M. A.; BORJALI, A.; FIOL, G. D. A scoping review of transfer learning research on medical image analysis using imagenet. **Computers in biology and medicine**, Elsevier, v. 128, p. 104115, 2021. Citado na página 25.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: **Icml**. [S.l.: s.n.], 2010. Citado na página 35.
- NESTEROV, Y. E. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. In: **Dokl. akad. nauk Sssr**. [S.l.: s.n.], 1983. v. 269, p. 543–547. Citado na página 36.
- NETO, D. G. d. A. Transferência de conhecimento utilizando aprendizado profundo para classificação de imagens histopatológicas. **32f. TCC (Graduação em Ciência da Computação) - Universidade Federal do Ceará, Campus Quixadá, Quixadá**, 2017. Citado na página 24.
- NIELSEN, M. A. **Neural networks and deep learning**. [S.l.]: Determination press San Francisco, CA, USA, 2015. v. 25. Citado nas páginas 42, 43 e 44.
- PAN, T.-S.; HUANG, H.-C.; LEE, J.-C.; CHEN, C.-H. Multi-scale resnet for real-time underwater object detection. **Signal, Image and Video Processing**, Springer, v. 15, n. 5, p. 941–949, 2021. Citado na página 56.
- PEREIRA, J.; SARAIVA, F. Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 131, p. 107085, 2021. Citado nas páginas 26 e 48.
- PONTI, M. A.; COSTA, G. B. P. da. Como funciona o deep learning. **arXiv preprint arXiv:1806.07908**, 2018. Citado na página 38.
- PRASETIYO, B.; MUSLIM, M.; BAROROH, N. *et al.* Evaluation performance recall and f2 score of credit card fraud detection unbalanced dataset using smote oversampling technique. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2021. v. 1918, n. 4, p. 042002. Citado na página 65.
- QU, W.; BALKI, I.; MENDEZ, M.; VALEN, J.; LEVMAN, J.; TYRRELL, P. N. Assessing and mitigating the effects of class imbalance in machine learning with application to x-ray imaging. **International journal of computer assisted radiology and surgery**, Springer, v. 15, n. 12, p. 2041–2048, 2020. Citado na página 26.
- RAO, K. S.; TERLAPU, P. V.; JAYARAM, D.; RAJU, K. K.; KUMAR, G. K.; PEMULA, R.; GOPALACHARI, V.; RAKESH, S. Intelligent ultrasound imaging for enhanced breast cancer diagnosis: Ensemble transfer learning strategies. **IEEE Access**, IEEE, 2024. Citado na página 47.
- RASKUTTI, G.; WAINWRIGHT, M. J.; YU, B. Early stopping and non-parametric regression: an optimal data-dependent stopping rule. **The Journal of Machine Learning Research, JMLR.org**, v. 15, n. 1, p. 335–366, 2014. Citado na página 52.
- REHMAN, A.; SABA, T. Neural networks for document image preprocessing: state of the art. **Artificial Intelligence Review**, Springer, v. 42, n. 2, p. 253–273, 2014. Citado na página 40.

- REIS, F. A. de B. **A Generative Adversarial Network Approach to Visual Expressive Speech Synthesis with Emotion Control** *Abordagem por Rede Generativa Adversária para Síntese de Discurso Visual Expressivo com.* Tese (Doutorado) — [sn], 2020. Citado na página 51.
- RIBEIRO, A. B. N. Um novo sistema automático para detecção e classificação de nódulos pulmonares em imagens de tomografia computadorizada do tórax usando uma única rede neural convolucional. 2021. Citado na página 23.
- RIBEIRO, H. C. M.; TAVARES, V. C. M. Comportamento e particularidades da produção acadêmica do tema “contabilidade gerencial” divulgada na base de dados do isi web of science core collection de 1985 a 2014. **Revista de Educação e Pesquisa em Contabilidade (REPeC)**, v. 11, n. 1, 2017. Citado na página 27.
- RS, R. I. *et al.* Transfer learning-based cnn model for the classification of breast cancer from histopathological images. **International Journal of Advanced Computer Science & Applications**, v. 15, n. 4, 2024. Citado nas páginas 27 e 45.
- SARKAR, D.; BALI, R.; GHOSH, T. **Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras.** [S.l.]: Packt Publishing Ltd, 2018. Citado na página 55.
- SEEMAN, E. Reconhecimento de palavras usando redes neurais. 2006. Citado na página 31.
- SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **Towards Data Sci**, v. 6, n. 12, p. 310–316, 2017. Citado na página 34.
- SHELKE, M. S.; DESHMUKH, P. R.; SHANDILYA, V. K. A review on imbalanced data handling using undersampling and oversampling technique. **Int. J. Recent Trends Eng. Res**, v. 3, n. 4, p. 444–449, 2017. Citado na página 47.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. **Annual review of biomedical engineering**, Annual Reviews, v. 19, p. 221–248, 2017. Citado na página 38.
- SILVA, L. N. d. C. *et al.* Análise e síntese de estratégias de aprendizado para redes neurais artificiais. [sn], 1998. Citado na página 33.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Citado nas páginas 24 e 55.
- SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The Journal of Machine Learning Research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014. Citado na página 50.
- SUNG, H.; FERLAY, J.; SIEGEL, R. L.; LAVERSANNE, M.; SOERJOMATARAM, I.; JEMAL, A.; BRAY, F. Global cancer statistics 2023: 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, 2023. Citado na página 24.
- SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: **Thirty-first AAAI conference on artificial intelligence.** [S.l.: s.n.], 2017. Citado na página 55.

- TAHERKHANI, A.; COSMA, G.; MCGINNITY, T. M. Adaboost-cnn: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. **Neurocomputing**, Elsevier, v. 404, p. 351–366, 2020. Citado na página 29.
- TANAKA, F. H. K. d. S.; ARANHA, C. Data augmentation using gans. **arXiv preprint arXiv:1904.09135**, 2019. Citado na página 50.
- THECKEDATH, D.; SEDAMKAR, R. Detecting affect states using vgg16, resnet50 and se-resnet50 networks. **SN Computer Science**, Springer, v. 1, n. 2, p. 1–7, 2020. Citado na página 55.
- THOMAS, N. M.; JEROME, S. A. Diabetic retinopathy detection using ensembled transfer learning based thrice cnn with svm classifier. **Multimedia Tools and Applications**, Springer, p. 1–27, 2024. Citado nas páginas 27 e 45.
- WANG, J.; PEREZ, L. *et al.* The effectiveness of data augmentation in image classification using deep learning. **Convolutional Neural Networks Vis. Recognit**, v. 11, p. 1–8, 2017. Citado na página 50.
- WANG, S.-H.; ZHANG, Y.-D. Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. **ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)**, ACM New York, NY, USA, v. 16, n. 2s, p. 1–19, 2020. Citado na página 55.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining Practical Machine Learning Tools and Techniques Third Edition**. [S.l.]: Morgan Kaufmann, 2017. Citado na página 38.
- WU, R.; QIN, K.; FANG, Y.; XU, Y.; ZHANG, H.; LI, W.; LUO, X.; HAN, Z.; LIU, S.; LI, Q. Application of cnn in the diagnosis for the invasion depth of gastrointestinal cancer: a systematic review and meta-analysis. **Journal of Gastrointestinal Surgery**, Elsevier, 2024. Citado nas páginas 26 e 39.
- XU, X.; LI, Y.; WU, G.; LUO, J. Multi-modal deep feature learning for rgb-d object detection. **Pattern Recognition**, Elsevier, v. 72, p. 300–313, 2017. Citado nas páginas 28 e 29.
- YANG, Q. An introduction to transfer learning. In: SPRINGER. **ADMA**. [S.l.], 2008. p. 1. Citado na página 24.
- ZAKI, M. J.; JR, W. M.; MEIRA, W. **Data mining and analysis: fundamental concepts and algorithms**. [S.l.]: Cambridge University Press, 2014. Citado na página 49.
- ZANINI, A. Regulação econômica no setor elétrico brasileiro: uma metodologia para definição de fronteiras de eficiência e cálculo do fator x para empresas distribuidoras de energia elétrica. **Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro**, 2004. Citado na página 32.
- ZHANG, G.; WANG, M.; LIU, K. Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china. **International Journal of Disaster Risk Science**, Springer, v. 10, n. 3, p. 386–403, 2019. Citado na página 47.
- ZHANG, Y.; ZHANG, Z.; MIAO, D.; WANG, J. Three-way enhanced convolutional neural networks for sentence-level sentiment classification. **Information Sciences**, Elsevier, v. 477, p. 55–64, 2019. Citado na página 28.

ZHOU, Q.; REN, C.; QI, S. An imbalanced r-stdp learning rule in spiking neural networks for medical image classification. **IEEE Access**, IEEE, v. 8, p. 224162–224177, 2020. Citado na página 79.

ZHOU, T.; LU, H.; YANG, Z.; QIU, S.; HUO, B.; DONG, Y. The ensemble deep learning model for novel covid-19 on ct images. **Applied Soft Computing**, Elsevier, v. 98, p. 106885, 2021. Citado na página 28.

ZHU, Y.; BRETTIN, T.; EVRARD, Y. A.; PARTIN, A.; XIA, F.; SHUKLA, M.; YOO, H.; DOROSHOW, J. H.; STEVENS, R. L. Ensemble transfer learning for the prediction of anti-cancer drug response. **Scientific Reports**, Nature Publishing Group, v. 10, n. 1, p. 1–11, 2020. Citado na página 28.

