



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Diagnóstico de câncer de mama baseado em métodos de aprendizado de máquina

Caio de Oliveira Mendes

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientadora
Prof.a Dr.a Roberta Barbosa Oliveira

Brasília
2023



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Diagnóstico de câncer de mama baseado em métodos de aprendizado de máquina

Caio de Oliveira Mendes

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof.a Dr.a Roberta Barbosa Oliveira (Orientadora)
CIC/UnB

Prof. Dr. Dibio Leandro Borges Prof. Dr. Marcus Vinicius Lamar
Universidade de Brasília Universidade de Brasília

Prof. Dr. João Luiz Azevedo de Carvalho
Coordenador do Curso de Engenharia da Computação

Brasília, 13 de dezembro de 2023

Dedicatória

Dedico este trabalho à minha parceira Laura Muglia, por todo o seu suporte e companheirismo que foram essenciais para a execução deste projeto. Você me inspira todos os dias e sou grato por sua parceria.

Agradecimentos

Agradeço à minha família, em especial à minha mãe, por todo o carinho, paciência, suporte e amor durante toda essa jornada.

Agradeço à minha orientadora Roberta Barbosa Oliveira, por toda a paciência, tempo e esforço dedicado em me orientar neste trabalho.

Agradeço a todos os meus colegas de curso e amigos, que me apoiaram e proporcionaram um ambiente enriquecedor e colaborativo.

Agradeço à Universidade de Brasília, por seu ensino de qualidade e por me fornecer a base necessária que me guiará durante toda minha trajetória profissional.

Por fim, agradeço à todos que, de alguma forma, contribuíram para a elaboração deste trabalho, direta ou indiretamente.

Resumo

O câncer de mama é o câncer mais diagnosticado no mundo todo além de ser o mais letal entre mulheres. Por essa razão, o diagnóstico precoce é fundamental. O diagnóstico se baseia em métodos de biópsia para obtenção de imagens histopatológicas que são analisadas por médicos histopatologistas. Essa análise é extremamente demorada devido a quantidade de detalhes presentes nessas imagens, sendo comum haver divergência entre os médicos que estão analisando. Desse modo, o presente trabalho objetiva propor um modelo *ensemble* utilizando a estratégia *stacking*, em que classificadores base constituídos pela rede *ResNet50*, uma rede CNN simples e um preditor *Random Forest* são unidos e combinados utilizando um meta-classificador de regressão linear, para contribuir para servir de apoio para o diagnóstico médico. De modo a avaliar o desempenho do modelo proposto, as métricas de acurácia, precisão, revocação e *F1 Score* foram utilizadas dentro de uma análise comparativa com outras redes CNN consideradas *state-of-the-art*. A partir da utilização das redes *ResNet50*, *VGG-16* e *InceptionV3 state-of-the-art*, foi possível realizar uma análise comparativa envolvendo os fatores de ampliação de 40X, 100X, 200X e 400X das imagens histopatológicas e o comportamento das redes ao utilizar classes balanceadas e desbalanceadas. O modelo proposto apresentou resultados promissores, atingindo 89.84% de acurácia, 86.78% de precisão, 94.19% de revocação e 90.33% de *F1 Score* para fator de ampliação de 40X. Os resultados obtidos, no geral, foram superiores aos resultados das outras redes *CNN state-of-the-art* utilizadas para cada fator de ampliação.

Palavras-chave: Aprendizado de Máquina, Ensemble de Classificadores, Redes Neurais Convolucionais, Classificação de Imagens, Câncer de Mama

Abstract

Breast cancer is the most diagnosed cancer worldwide and remains the deadliest among women. For this reason, early diagnosis is crucial. Diagnosis is based on biopsy methods to obtain histopathological images that are analyzed by histopathologists. This analysis is extremely time-consuming due to the amount of detail in these images, often leading to divergent opinions among analyzing physicians. Therefore, this work aims to propose an ensemble model using the stacking strategy, where base classifiers consisting of the ResNet50 network, a simple CNN, and a Random Forest predictor are combined using a meta-classifier of linear regression to contribute as support for medical diagnosis. Accuracy, precision, recall, and F1 Score metrics were used to evaluate the performance of the proposed model, in a comparative analysis with other state-of-the-art CNN networks. Utilizing ResNet50, VGG-16, and InceptionV3 state-of-the-art networks, a comparative analysis was conducted involving magnification factors of 40X, 100X, 200X and 400X of histopathological images, considering the behavior of the networks when using balanced and unbalanced classes. The proposed model showed promising results, achieving 89.84% accuracy, 86.78% precision, 94.19% recall, and 90.33% F1 Score for a 40X magnification factor. The overall results exceeded those of other state-of-the-art CNNs used for each magnification factor.

Keywords: Machine Learning, Ensemble of Classifiers, Convolutional Neural Network, Image Classification, Breast Cancer

Sumário

1	Introdução	1
1.1	Objetivos do trabalho	2
1.2	Estrutura do documento	3
2	Fundamentação Teórica	4
2.1	Aprendizado de máquina e suas aplicações na área médica	4
2.2	Técnicas de aprendizado de máquina para o diagnóstico de câncer de mama	5
2.2.1	Redes Neurais Artificiais	5
2.2.2	Redes Neurais Convolucionais	7
2.2.3	Modelos de Redes Neurais Convolucionais	9
2.2.4	Árvores de Decisão	11
2.2.5	Aprendizado por <i>Ensemble</i>	12
2.2.6	<i>Random Forest</i>	15
2.2.7	<i>Regressão Linear</i>	16
2.2.8	Aprendizado por Transferência	16
2.3	Técnicas de Balanceamento de Dados	16
2.4	Métricas de Avaliação de Desempenho	18
3	Revisão de Literatura	21
4	Metodologia	26
4.1	Base de dados	26
4.2	Pré-processamento de imagens	28
4.3	Experimentos Propostos	30
4.3.1	Modelo Ensemble proposto	30
4.3.2	Modelos state-of-the-art	33
5	Resultados	35
5.1	Ferramentas utilizadas	35
5.2	Experimentos do Modelo Ensemble Proposto	35

5.2.1	Sem aumento de dados e sem <i>Oversampling</i>	36
5.2.2	Com aumento de dados e sem <i>Oversampling</i>	37
5.2.3	Com <i>Oversampling</i>	40
5.3	Experimentos dos Modelos state-of-the-art	41
5.3.1	Sem aumento de dados e sem <i>Oversampling</i>	41
5.3.2	Com aumento de dados e sem <i>Oversampling</i>	44
5.3.3	Com <i>Oversampling</i>	45
5.4	Síntese dos experimentos	47
6	Conclusão e Trabalhos Futuros	57
	Referências	59

Lista de Figuras

2.1	Estrutura geral de um <i>Perceptron</i> (Fonte: [1]).	6
2.2	Exemplo de um MLP (Fonte: [1]).	7
2.3	Exemplo da operação de convolução (Fonte: [2]).	8
2.4	Exemplo de operação de <i>Pooling</i> máximo (Fonte: [2]).	9
2.5	Camadas do modelo VGG-16 (Fonte: [3]).	10
2.6	Camadas do modelo <i>ResNet50</i> (Fonte: [4]).	10
2.7	Camadas do modelo InceptionV3 (Fonte: [5]).	11
2.8	Exemplo de uma árvore de decisão simples (Fonte: [6]).	12
2.9	Exemplo de modelo de <i>ensemble</i> por votos majoritários (Fonte: [7]).	13
2.10	Exemplo de estratégia de <i>ensemble bagging</i> (Fonte: [7]).	13
2.11	Exemplo de modelo de <i>ensemble Adaboost</i> (Fonte: [7]).	14
2.12	Exemplo de estratégia de <i>ensemble stacking</i> (Fonte: [7]).	15
2.13	Exemplo de matriz de confusão para 2 classes (Fonte: Elaborada pelo autor).	18
4.1	Imagens de tumores malignos da base de dados: (a) 40X, (b) 100X, (c) 200X e (d) 400X (Fonte: [8]).	27
4.2	Imagens de tumores benignos da base de dados: (a) 40X, (b) 100X, (c) 200X e (d) 400X (Fonte: [8]).	28
4.3	Exemplo de imagem gerada pela técnica de <i>Data Augmentation</i>	30
4.4	Treinamento dos modelos base.	31
4.5	Treinamento do meta-classificador.	32
4.6	Modelo <i>ensemble</i> proposto.	32
5.1	Matriz de confusão do modelo ensemble sem aumento de dados e sem <i>Oversampling</i>	38
5.2	Matriz de confusão do modelo ensemble com aumento de dados e sem <i>Oversampling</i>	39
5.3	Matriz de confusão do modelo ensemble com aumento de dados e com <i>Oversampling</i>	41

5.4	Matriz de confusão do modelo <i>VGG-16</i> sem aumento de dados e sem <i>Oversampling</i>	43
5.5	Matriz de confusão do modelo <i>ResNet50</i> sem aumento de dados e sem <i>Oversampling</i>	44
5.6	Matriz de confusão do modelo <i>InceptionV3</i> sem aumento de dados e sem <i>Oversampling</i>	45
5.7	Curvas de acurácia e perda do modelo <i>VGG-16</i> sem aumento de dados e sem <i>Oversampling</i>	46
5.8	Curvas de acurácia e perda do modelo <i>ResNet50</i> sem aumento de dados e sem <i>Oversampling</i>	47
5.9	Curvas de acurácia e perda do modelo <i>InceptionV3</i> sem aumento de dados e sem <i>Oversampling</i>	48
5.10	Matriz de confusão do modelo <i>VGG-16</i> com aumento de dados e sem <i>Oversampling</i>	50
5.11	Matriz de confusão do modelo <i>ResNet50</i> com aumento de dados e sem <i>Oversampling</i>	50
5.12	Matriz de confusão do modelo <i>InceptionV3</i> com aumento de dados e sem <i>Oversampling</i>	51
5.13	Curvas de acurácia e perda do modelo <i>VGG-16</i> com aumento de dados e sem <i>Oversampling</i>	51
5.14	Curvas de acurácia e perda do modelo <i>ResNet50</i> com aumento de dados e sem <i>Oversampling</i>	52
5.15	Curvas de acurácia e perda do modelo <i>InceptionV3</i> com aumento de dados e sem <i>Oversampling</i>	52
5.16	Matriz de confusão do modelo <i>VGG-16</i> com aumento de dados e com <i>Oversampling</i>	53
5.17	Matriz de confusão do modelo <i>ResNet50</i> com aumento de dados e com <i>Oversampling</i>	53
5.18	Matriz de confusão do modelo <i>InceptionV3</i> sem aumento de dados e com <i>Oversampling</i>	54
5.19	Curvas de acurácia e perda do modelo <i>VGG-16</i> com aumento de dados e com <i>Oversampling</i>	54
5.20	Curvas de acurácia e perda do modelo <i>ResNet50</i> com aumento de dados e com <i>Oversampling</i>	55
5.21	Curvas de acurácia e perda do modelo <i>InceptionV3</i> sem aumento de dados e com <i>Oversampling</i>	55

Lista de Tabelas

3.1	Trabalhos abordados neste capítulo.	25
4.1	Estrutura da base de dados.	27
4.2	Estrutura da base de dados.	29
4.3	Tabela de hiperparâmetros utilizados no modelo <i>ensemble</i>	33
4.4	Tabela de hiperparâmetros utilizados nos modelos <i>state-of-the-art</i>	34
4.5	Tabela de Experimentos Propostos.	34
5.1	Tabela de hiperparâmetros para os modelos base.	36
5.2	Métricas obtidas dos classificadores base individuais e do modelo <i>ensemble</i> proposto sem aumento de dados e sem <i>Oversampling</i>	37
5.3	Métricas obtidas dos classificadores base e do modelo <i>ensemble</i> proposto com aumento de dados e sem <i>Oversampling</i>	39
5.4	Métricas obtidas dos classificadores base e do modelo <i>ensemble</i> proposto com aumento de dados e com <i>Oversampling</i>	40
5.5	Tabela de melhores hiperparâmetros para os modelos <i>state-of-the-art</i>	42
5.6	Métricas obtidas dos modelos <i>state-of-the-art</i> sem aumento de dados e sem <i>Oversampling</i>	42
5.7	Métricas obtidas dos modelos <i>state-of-the-art</i> com aumento de dados e sem <i>Oversampling</i>	49
5.8	Métricas obtidas dos modelos <i>state-of-the-art</i> com <i>Oversampling</i>	49
5.9	Resumo dos melhores resultados obtidos nos experimentos.	56

Lista de Abreviaturas e Siglas

Adaboost Adaptative Boosting.

ANN Artificial Neural Network.

BGD Batch Gradient Descent.

BGR Blue, Green and Red.

BreakHis Breast Cancer Histopathological Database.

CNN Convolutional Neural Network.

DCNN Deep Convolutional Neural Network.

DNN Dense Neural Network.

DRDA-Net Dense Residual Dual-shuffle Attention Network.

FN Falso Negativo.

FP Falso Positivo.

INCA Instituto Nacional do Câncer.

IOT Internet of Things.

KNNs K-Nearest Neighbors.

MBGD Mini-Batch Gradient Descent.

MLP Multilayer Perceptron.

MTRRE-Net Multi-scale Dual Residual Recurrent Network.

ReLU Rectified Linear Unit.

RGB Red, Green and Blue.

SGD Stochastic Gradient Descent.

SMOTE Synthetic Minority Over-sampling Technique.

SOB Surgical Open Biopsy.

SSAE Stacked Sparse Autoencoder.

SSDHO-based DNN Shuffled Shepherd Deer Hunting Optimization-based Deep Neural Network.

VGG Visual Geometry Group.

VN Verdadeiro Negativo.

VP Verdadeiro Positivo.

Capítulo 1

Introdução

De acordo com a Agência Internacional de Pesquisa sobre o Câncer [9], o câncer de mama se tornou o câncer mais diagnosticado em todo o mundo. Estima-se que, em 2020, surgiram mais de 2 milhões de novos casos e quase 685 mil mortes no mundo inteiro. O câncer de mama é a principal mortalidade de câncer para mulheres e a quinta maior causa de câncer na população em geral. Segundo o Instituto Nacional do Câncer (INCA), estima-se a ocorrência de 704 mil casos de câncer de mama anualmente até 2025 no Brasil [10].

Existem diferentes tipos de câncer de mama e isso é determinado de acordo com os tipos de células mamárias afetadas. Os tipos mais comuns são os **carcinomas ductais** e os **adenocarcinomas**, encontrados nos ductos mamários ou nos lóbulos, respectivamente. Não existe uma causa específica para câncer de mama, mas existem fatores de riscos que podem contribuir para o surgimento da doença, como por exemplo alcoolismo, sobrepeso, sedentarismo, primeira gravidez após os 30 anos e uso de contraceptivos orais. Além disso, fatores genéticos e hereditários também estão associados.

Se sintomas de câncer de mama surgirem, como um caroço ou massa, ou os resultados de exames de imagem, como por exemplo a mamografia, sugerirem que exista a possibilidade de câncer de mama, uma biópsia pode ser necessária. A biópsia do tecido mamário é a única forma de ter certeza se de fato é câncer ou não. Existem diferentes tipos de biópsia do tecido mamário e a escolha entre elas depende de inúmeros fatores como a localização do tumor, o seu tamanho, a saúde geral do paciente e o quão suspeito o tumor é.

Os principais tipos de biópsia são: aspiração por agulha fina, biópsia por agulha grossa ou biópsia cirúrgica. No primeiro tipo, uma pequena parte de tecido mamário é removida da área suspeita por uma agulha fina. A biópsia por agulha grossa consiste em utilizar uma agulha grossa acoplada em uma pistola para retirar pedaços maiores de tecido mamário. Esse tipo é o preferido por retirar mais amostras de tecido sem precisar de

cirurgia. Quando os resultados dos dois primeiros tipos não são conclusivos, é necessário realizar uma biópsia cirúrgica, onde a área suspeita é retirada parcial ou integralmente para análise.

A taxa de sobrevivência do câncer de mama aumenta consideravelmente quando o diagnóstico é realizado de forma precoce. No entanto, o diagnóstico definitivo depende da análise das imagens das lâminas de tecido mamário obtidas por meio da biópsia e a análise dessas imagens por médicos histopatologistas demanda tempo. A acurácia destas análises dependem da experiência do médico que está analisando [11].

Nesse contexto, a Inteligência Artificial com métodos de aprendizagem de máquina podem ser úteis fornecendo ferramentas e técnicas que auxiliem na classificações de tumores a partir de imagens de tecido mamário ainda em estágios iniciais, exemplificado nos trabalhos de Chattopadhyay *et al.* [12] e Sheeba *et al.* [13]. Desse modo, as análises dessas imagens com o auxílio de técnicas de aprendizado de máquina podem ser realizadas de maneira mais rápida e assertiva [14].

Dentro das técnicas disponíveis de aprendizagem de máquina, as Redes Neurais Convolucionais, ou do inglês Convolutional Neural Network (CNN), apresentam grande destaque em tarefas de classificação [15]. Além disso, os modelos *ensemble*, que combinam as predições de vários modelos de classificação individuais, são robustos e apresentam excelentes capacidades de generalização se comparados com a utilização de modelos individualmente, como indicam os trabalhos de Majumdar, Pramanik e Sarkar [16] e Karthik, Menaka e Siddharth [17].

Assim, o presente trabalho propõe a elaboração de um modelo *ensemble* cuja finalidade principal é a de realizar a classificação de imagens de tecido mamário em classes benignas e malignas. Além disso, este trabalho busca realizar uma análise comparativa com outras redes convolucionais consideradas *state-of-the-art* para esse tipo de problema.

1.1 Objetivos do trabalho

Este trabalho possui o objetivo principal de propor um modelo para auxiliar na classificação de imagens de lâminas de tecido mamário, sendo capaz de diagnosticar amostras malignas e benignas. Para atingir o objetivo principal, foram considerados os seguintes objetivos específicos:

- Analisar a eficácia de diferentes modelos de redes CNN *state-of-the-art* para classificação de amostras de imagens de lâminas de tecido mamário;

- Analisar o impacto em diferentes modelos de redes CNN com a utilização de estratégias para lidar com o desbalanceamento de classes em problema de classificação entre diferentes tipos de câncer de mama;
- Analisar o impacto de diferentes fatores de ampliação de imagens de lâminas de tecido mamário em diferentes modelos de redes CNN;
- Desenvolver e avaliar o comportamento de um modelo *ensemble* com classes balanceadas e desbalanceadas para problema de classificação binária para diferentes fatores de ampliação.

1.2 Estrutura do documento

Esta seção apresenta a estrutura da elaboração deste trabalho. O Capítulo 2 visa fornecer uma base teórica sólida necessária para o pleno entendimento do trabalho proposto. O Capítulo 3 apresenta uma revisão aprofundada da literatura relacionada a métodos de aprendizagem de máquina voltados para o câncer de mama. No Capítulo 4 são descritos de forma detalhada os métodos e procedimentos necessários para a confecção dos experimentos realizados. Em sequência, o Capítulo 5 apresenta de forma objetiva os dados e resultados obtidos durante os experimentos realizados, com uma discussão e análise do que foi encontrado. Por fim, o Capítulo 6 destaca as principais contribuições deste trabalho e apresenta sugestões para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

O objetivo deste capítulo é explanar sobre os principais conceitos teóricos que são indispensáveis para se obter uma compreensão abrangente desta monografia. Inicialmente são apresentadas uma breve introdução sobre os conceitos de aprendizagem de máquina e aplicações na área médica. Em seguida, conceitos e técnicas de aprendizado de máquina são apresentados, como diferentes tipos de classificadores, redes neurais convolucionais, árvores de decisão e técnicas de *ensemble*.

2.1 Aprendizado de máquina e suas aplicações na área médica

Na literatura, existem diversas definições diferentes para aprendizado de máquina. Mitchell [18], define aprendizado de máquina como “a capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência”. Por outro lado, Franco [19] define aprendizado de máquina como “uma subárea da inteligência artificial, cujo objetivo é ensinar os computadores a aprenderem de forma automática”. As duas definições apresentadas têm o enfoque no processo de obtenção de experiência por computadores, de maneira a tornar possível a utilização da experiência passada a fim de executar uma determinada tarefa de maneira mais precisa.

Com o passar dos anos, as técnicas de aprendizado de máquina estão ficando mais elaboradas e existem inúmeras aplicações bem-sucedidas na solução de problemas reais do nosso cotidiano. Um exemplo disso, embora nem sempre notamos, é o algoritmo presente por trás de grandes plataformas de comércio eletrônico [1]. Com isso, técnicas de aprendizado de máquina são empregadas de modo a utilizar o histórico passado de compras de um determinado usuário a fim de promover recomendações personalizadas de outros produtos. Outro exemplo mais palpável presente no nosso cotidiano é o uso de

assistentes de voz em *Smartphones* [1], em que técnicas de aprendizado de máquina são utilizadas para entender comandos de voz dos usuários e executar tarefas.

O aprendizado de máquina também possui aplicações na área médica: é possível utilizar técnicas para diagnóstico de câncer por meio da análise de dados de expressão gênica, permitindo uma avaliação mais abrangente do genoma humano de modo a detectar anomalias precocemente [1]. Além disso, o câncer e outras doenças sérias podem ter seu diagnóstico facilitado por meio de técnicas de aprendizado de máquina que auxiliem na interpretação de imagens médicas, como radiografias, imagens de biópsias e ressonâncias magnéticas [9]. Os modelos de aprendizado de máquina podem, também, ser empregados para ajudar a identificar tendências na saúde de um indivíduo e prever a ocorrência de doenças crônicas [19].

2.2 Técnicas de aprendizado de máquina para o diagnóstico de câncer de mama

Para o diagnóstico de câncer de mama, técnicas de aprendizado de máquina têm desempenhado um papel fundamental no campo de medicina diagnóstica. Nesse contexto, diversas abordagens de aprendizado de máquina são apresentadas nesta seção.

2.2.1 Redes Neurais Artificiais

As redes neurais artificiais, também conhecidas como Artificial Neural Network (ANN), surgiram inspiradas pelas redes de neurônios encontrados em nossos cérebros [7] e representam o núcleo de vários algoritmos de aprendizagem profunda, como classificação de imagens, reconhecimento de voz e recomendação de vídeos.

O primeiro modelo de um neurônio artificial é o chamado de *Perceptron*, proposto por Rosenblatt [20] em 1958. O *perceptron* possui uma estrutura simples: as entradas de dados passam por células detentora de pesos, nos quais as entradas irão ser multiplicadas. Em seguida, somadas a um valor chamado de *bias*, que representa um parâmetro adicional importante que ajudará no processo de decisão. Por fim, a saída gerada por essa operação passa por uma função de ativação, que nada mais é uma função que recebe a saída de um neurônio e determina se ele deve ser ativado ou não. Existem diferentes funções de ativação, como a função Rectified Linear Unit (ReLU) e a *Softmax*. A função de ativação ReLU retorna zero para valores de entrada negativos e retorna o valor de entrada quando este for positivo. A função de ativação *Softmax*, por outro lado, consiste em utilizar uma função de normalização exponencial para gerar probabilidades. A Figura 2.1 ilustra a arquitetura simplificada de um *Perceptron*. Nesta figura, X representa os valores das

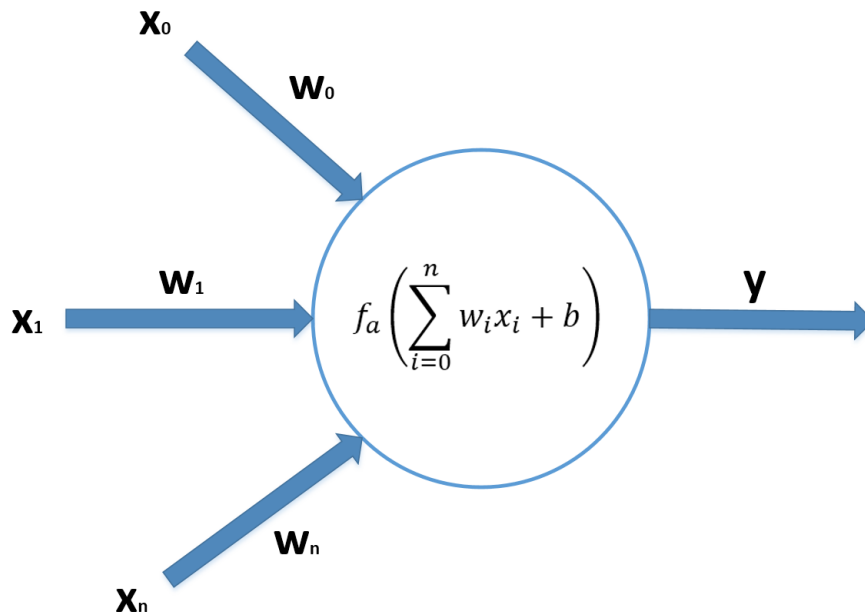


Figura 2.1: Estrutura geral de um *Perceptron* (Fonte: [1]).

entradas, W representa os pesos do neurônio que irão ser multiplicados pelas entradas fornecidas, b representa o *bias* a ser somado no valor resultante e Y representa a saída final do *Perceptron*.

Uma modelo mais complexo é o *Multilayer Perceptron (MLP)*, que é constituído por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Todas as camadas exceto a última estão conectadas na próxima camada da sequência. As camadas ocultas realizam operações similares ao que foi visto no *Perceptron*, processando as informações recebidas das camadas anteriores. A camada de saída representa a saída final do modelo e o número de neurônios nessa camada depende do problema aplicado. Em um problema de classificação binária, por exemplo, há um único neurônio na camada de saída indicando a classe gerada.

Os pesos dos neurônios de todas as camadas da MLP são modificados durante o treinamento de acordo com o algoritmo de *backpropagation*. Esse algoritmo realiza o ajuste dos pesos de forma a minimizar a função de perda e esse ajuste ocorre das camadas superiores até as inferiores. Existem variações no algoritmo de *backpropagation*, como o Stochastic Gradient Descent (SGD), o Batch Gradient Descent (BGD) e o Mini-Batch Gradient Descent (MBGD), que diferem na forma como os pesos são atualizados (em lotes completos, em lotes pequenos ou em uma amostra única por vez). A Figura 2.2 ilustra a arquitetura de uma MLP.

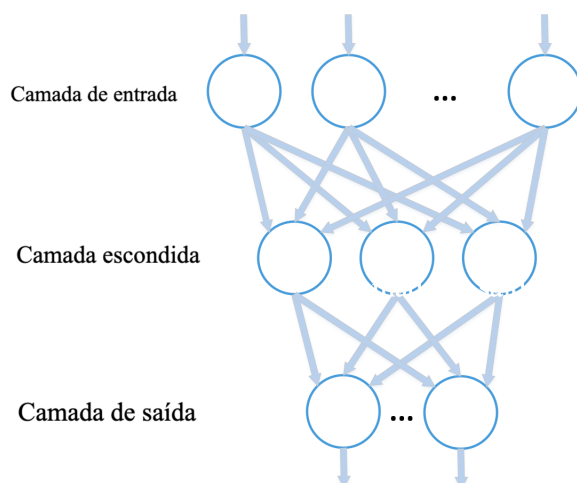


Figura 2.2: Exemplo de um MLP (Fonte: [1]).

2.2.2 Redes Neurais Convolucionais

A CNN foi proposta inicialmente por Lecun *et al.* [21] e representa um tipo especializado de uma rede neural. As CNNs obtêm grande sucesso em aplicações práticas, sobretudo com dados de séries temporais ou com imagens. Goodfellow, Bengio e Courville [22] definem uma CNN como “simplesmente uma rede neural que utilizam a operação matemática de convolução ao invés de multiplicação matricial em pelo menos uma de suas camadas”. No Capítulo 4 uma rede CNN simples será descrita para a construção do modelo proposto neste trabalho. Uma rede CNN é constituída por camadas convolucionais, camadas de *Pooling* e por uma camada totalmente conectada que são detalhadas a seguir.

Camadas convolucionais

A camada convolucional de uma CNN é uma das camadas mais importantes, uma vez que é responsável por extrair *features* das entradas fornecidas. Para isso, as camadas convolucionais dispõem-se de uma espécie de filtro, que são um *array* multidimensional de parâmetros adaptados pelo algoritmo de aprendizagem [22]. Desse modo, é possível extrair características importantes das imagens.

Os filtros possuem tamanhos reduzidos e percorrem os dados de entrada em largura, altura e profundidade. Para cada uma dessas posições, a operação de convolução é realizada e uma saída resultante dessa operação é gerada, apresentando dimensões reduzidas. Os filtros são ajustados a cada iteração no processamento da entrada de maneira a extrair *features* mais complexas. A Figura 2.3 apresenta a operação de convolução utilizando um filtro de dimensões 2×2 em uma entrada de dimensões 3×3 . Para cada posição do filtro

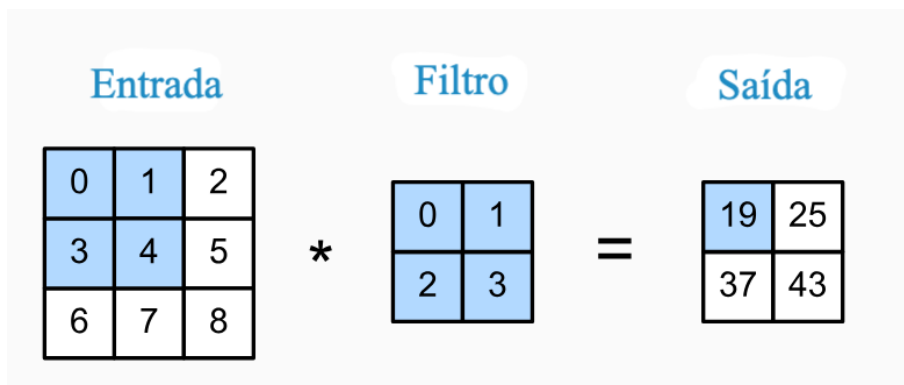


Figura 2.3: Exemplo da operação de convolução (Fonte: [2]).

sobre o conjunto de dados da imagem, a operação de convolução é realizada gerando a saída correspondente.

Camadas de *Pooling*

As camadas de *Pooling* são camadas responsáveis por reduzir o tamanho dos dados resultantes das camadas de convolução baseado em algum fator pré-definido. Desse modo, os dados obtêm uma outra representação permitindo que a rede aprenda com uma maior diversidade de dados apresentados, reduzindo, assim, o *overfitting* [23].

Há duas técnicas principais muito utilizadas para aplicar o *Pooling*: o *Pooling* máximo e o *Pooling* por média. No *Pooling* máximo, os dados de entrada dessa camada são subdivididos em subpartes e, em cada subparte, o valor máximo é extraído. O *Pooling* por média funciona de maneira análoga: os dados de entrada são subdivididos em subpartes e, em cada subparte, o valor da média dos dados é extraído. Desse modo, as dimensões são reduzidas ainda mais pelo fator do filtro utilizado nesta camada.

A Figura 2.4 ilustra o funcionamento da operação de *Pooling*. O conjunto original de dados de dimensões 3×3 é subdividido em 4 regiões de tamanho 2×2 e, em seguida, o valor máximo de cada uma dessas subregiões é extraído. Rodrigues [24] afirma que este procedimento é benéfico para a rede neural, pois melhora a regularização da rede e reduz custo de memória e processamento.

Camada Totalmente Conectada

A última camada de uma CNN é uma camada totalmente conectada. Esta camada é responsável por utilizar as *features* extraídas das camadas anteriores para enfim classificar os dados de imagem em uma determinada categoria. Para a construção dessa camada, é utilizado um MLP.

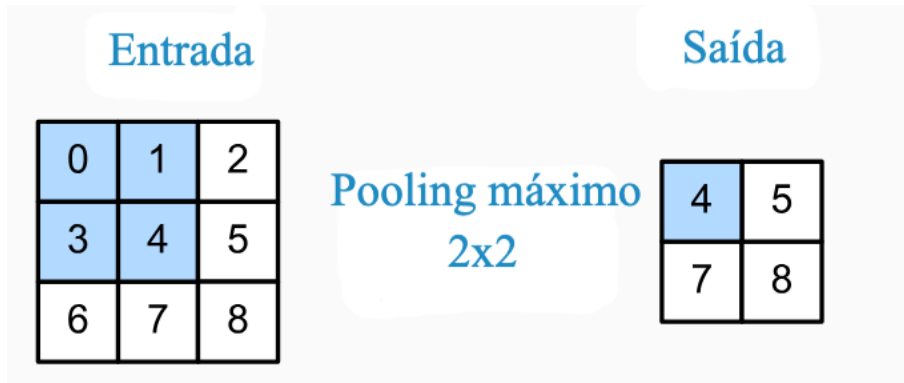


Figura 2.4: Exemplo de operação de *Pooling* máximo (Fonte: [2]).

2.2.3 Modelos de Redes Neurais Convolucionais

Os modelos de CNN, como foi visto na Subseção 2.2.2, desempenham um papel importantíssimo na área de aprendizado de máquina, impulsionando avanços significativos em tarefas como classificação de imagens, detecção de objetos e segmentação. Nesta seção, alguns modelos de CNN, como a *VGG-16*, a *ResNet50* e a *InceptionV3*, serão abordados. Essas redes CNN são consideradas *state-of-the-art* para o problema de classificação de câncer de mama e serão utilizadas para uma análise comparativa com o modelo *ensemble* proposto neste trabalho.

VGG-16

Simonyan e Zisserman [15] propõem um modelo de CNN fortemente utilizada para tarefas de visão computacional e aprendizado profundo. Esse modelo foi desenvolvido pelo Visual Geometry Group (VGG) da Universidade de Oxford e possui um total de 16 camadas, sendo 13 delas camadas convolucionais, 5 camadas de Pooling e 3 camadas totalmente conectadas. A Figura 2.5 ilustra a arquitetura da rede *VGG-16*.

ResNet50

He *et al.* [25] propõem uma arquitetura de rede neural baseada no funcionamento das redes residuais. A principal característica desse tipo de rede é o uso de blocos residuais. A arquitetura da *ResNet50* é composta por 50 camadas convolucionais, que incluem camadas de convolução, camadas de normalização, funções de ativação e camadas de *pooling*. A Figura 2.6 ilustra a arquitetura da rede *ResNet50*.

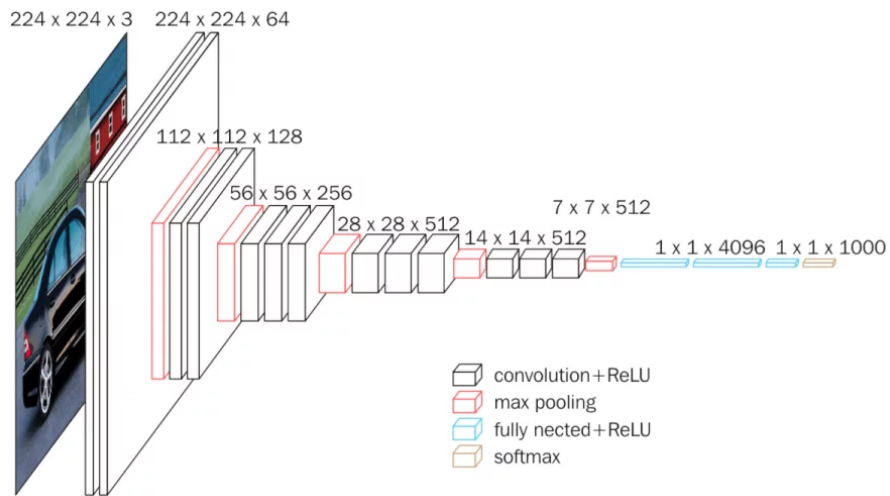


Figura 2.5: Camadas do modelo VGG-16 (Fonte: [3]).



Figura 2.6: Camadas do modelo *ResNet50* (Fonte: [4]).

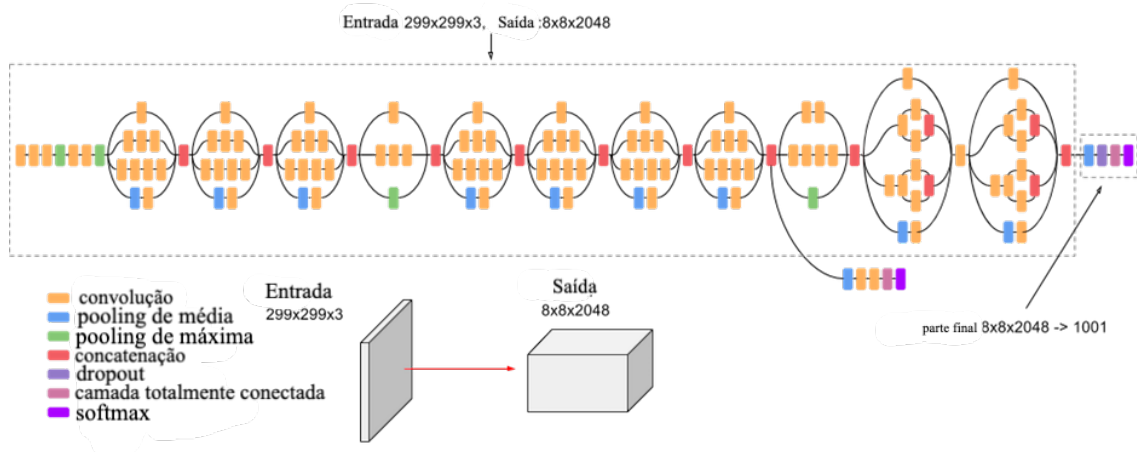


Figura 2.7: Camadas do modelo InceptionV3 (Fonte: [5]).

InceptionV3

Szegedy *et al.* [26] propõem a rede *InceptionV3* para a realização de tarefas de visão computacional. O funcionamento desta rede é baseado no uso de módulos *Inception*. Esses módulos representam a aplicação de operações de convolução com vários tamanhos de filtros diferentes. Com isso, a rede tem a capacidade de aprender e extrair características em diferentes escalas. A Figura 2.7 ilustra a arquitetura desta rede.

2.2.4 Árvores de Decisão

Árvores de decisão representam um dos algoritmos mais famosos de aprendizado de máquina e baseia-se na estratégia de dividir para conquistar para resolver um problema de decisão, em que um problema complexo é dividido em muitos problemas mais simples [27]. Cada um desses subproblemas são combinados em um formato de árvore até atingir a melhor solução para o problema original, que pode ser de classificação ou regressão.

Existem muitos algoritmos diferentes na implementação de uma árvore de decisão, como o algoritmo *CART* [28] e o algoritmo *C4.5* [29]. Estes algoritmos operam com base em um mesmo princípio de funcionamento, em que o conjunto de dados original é inserido e um atributo é escolhido desse conjunto de dados de forma a maximizar o critério de divisão. Essa regra de divisão é baseada em uma medida que indica o quão bem esse atributo discrimina as classes [27].

Em seguida, para cada partição de dados geradas na divisão descrita anteriormente, esse mesmo processo é realizado novamente de forma recursiva, gerando, assim, a árvore de decisão. A Figura 2.8 exemplifica o funcionamento de uma árvore de decisão simples que realiza a classificação binária de um indivíduo ir ou não à praia.

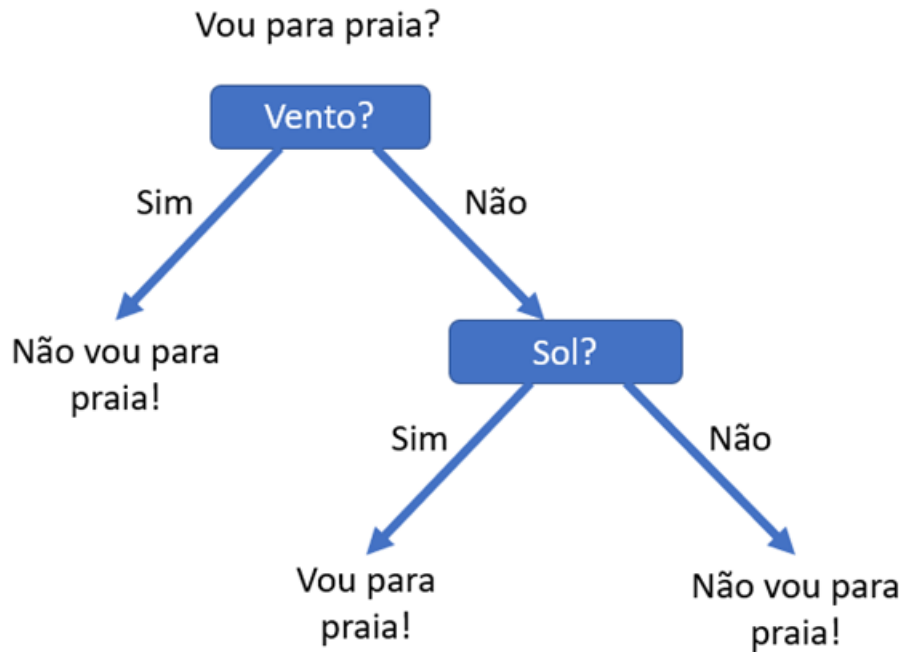


Figura 2.8: Exemplo de uma árvore de decisão simples (Fonte: [6]).

2.2.5 Aprendizado por *Ensemble*

O aprendizado por *ensemble*, também chamado de *Ensemble Learning*, é uma técnica de aprendizado de máquina que consistem em combinar um agrupamento de modelos preditores com o objetivo de atingir um melhor resultado se comparado a utilizar os preditores individualmente [7]. Um exemplo simples da utilização do aprendizado por *ensemble* é a construção de um modelo que combina vários classificadores de tal forma que a classe resultante desse modelo é a classe que obteve mais votos considerando todos estes classificadores. A Figura 2.9 ilustra esse processo, em que a saída resultante é a que teve mais votos considerando todos os classificadores envolvidos. Existem outras três principais estratégias de *ensemble*: *bagging*, *boosting* e *stacking*.

Bagging

Bonaccorso [1] define a estratégia *bagging*, também chamado de *bootstrap aggregating*, como cada classificador sendo treinado de forma independente considerando apenas um subconjunto da base de dados original. O objetivo do *bagging* é o de evitar ter mais de um classificador especialista no mesmo subconjunto de dados. Um exemplo famoso de *bagging* é o algoritmo de *Random Forest*, que irá ser discutido na Subseção 2.2.6. A Figura 2.10 ilustra o funcionamento desse método.

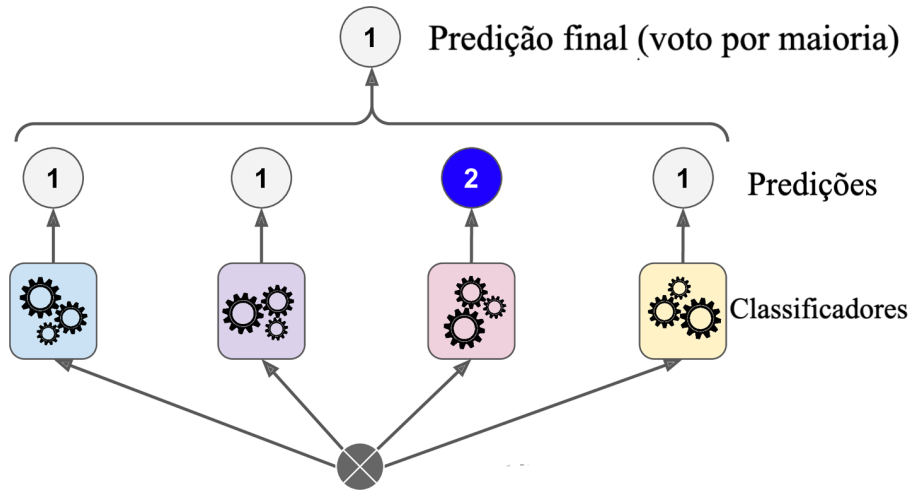


Figura 2.9: Exemplo de modelo de *ensemble* por votos majoritários (Fonte: [7]).

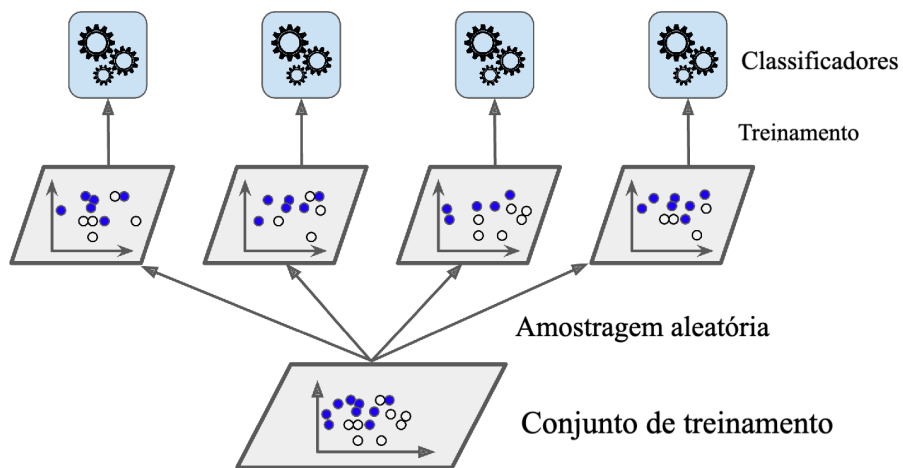


Figura 2.10: Exemplo de estratégia de *ensemble bagging* (Fonte: [7]).

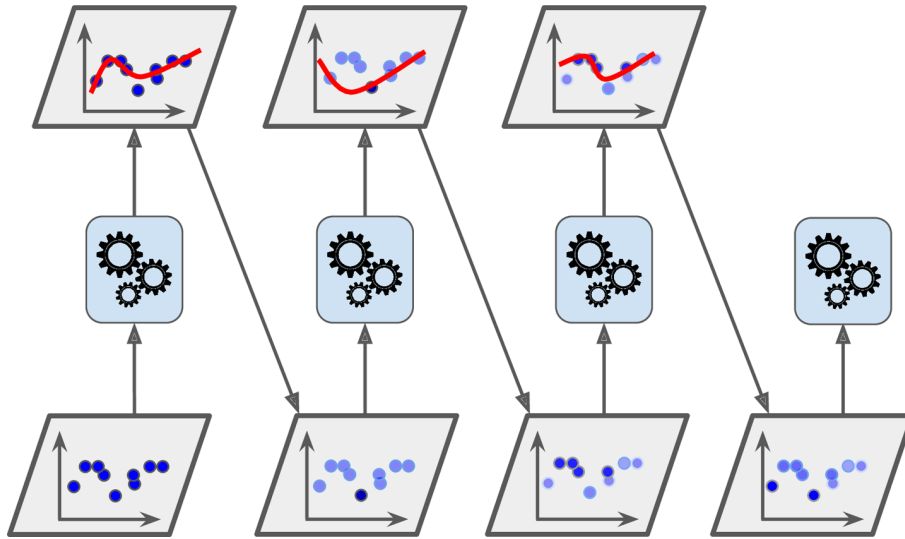


Figura 2.11: Exemplo de modelo de *ensemble Adaboost* (Fonte: [7]).

Boosting

Outra estratégia de *ensemble* bastante utilizado é o *boosting*, no qual cada preditor utilizado dentro do modelo é treinado de maneira sequencial, cada um visando corrigir os erros do seu predecessor. Existem muitos métodos de *boosting* disponíveis, sendo os mais famosos o *Adaptive Boosting (Adaboost)* [30] e o *Gradient Boosting* [31].

O *Adaboost* funciona da seguinte maneira: inicialmente um preditor inicial é treinado usando todo o conjunto de dados e, em seguida, as instâncias de treinamento que foram classificadas de forma incorreta tem seu peso aumentado. O próximo classificador irá utilizar os mesmos pesos obtidos do classificador anterior atualizados e irá repetir o processo de treinamento utilizando o conjunto de dados e aumentar os pesos nas instâncias que forem classificadas incorretamente. Esse processo se repete até que todos os preditores forem treinados. A Figura 2.11 exemplifica o processo de treinamento utilizando o algoritmo *Adaboost*.

O *Gradient Boosting* funciona de maneira muito similar ao *Adaboost*. A diferença entre os dois métodos reside no fato de que, ao invés de modificar os pesos a cada iteração como o *Adaboost* faz, o método do *Gradient Boosting* tenta fazer com que o preditor atual ajuste seus pesos baseado nos erros residuais do preditor anterior.

Stacking

A estratégia *stacking*, também chamado de *stacked generalization* [32], é baseado em um conjunto de classificadores heterogêneos. Nesse contexto, cada classificador distinto é treinado de forma independente e suas previsões geradas servem como entrada de um preditor final, chamado de meta-classificador. O meta-classificador utiliza as previsões

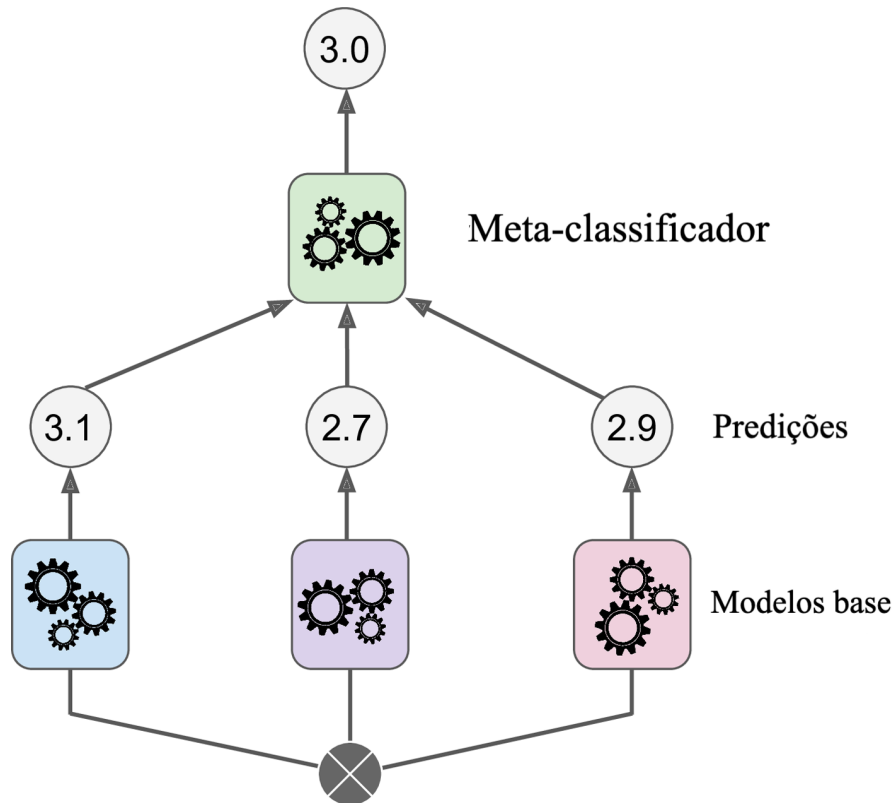


Figura 2.12: Exemplo de estratégia de *ensemble stacking* (Fonte: [7]).

anteriores para gerar uma predição final. A Figura 2.12 exemplifica um modelo por *ensemble* utilizando a estratégia de *stacking*. A utilização de um meta-classificador difere da estratégia de votação do modelo básico de *ensemble* apresentado nesta seção. A estratégia de votação combina diretamente as predições dos modelos base, enquanto no *stacking* essas predições são utilizadas como entrada no meta-classificador, que é treinado para combinar essas predições da melhor maneira possível.

2.2.6 *Random Forest*

Floresta aleatória, também chamada de *Random Forest* [33], é um dos algoritmos de aprendizado de máquina mais poderosos disponíveis atualmente [7]. Como visto na Subseção 2.2.4 e na Subseção 2.2.5, o algoritmo de floresta aleatória representa um modelo de aprendizagem por *ensemble* utilizando a estratégia *bagging*, em que cada preditor envolvido é uma árvore de decisão.

Neste algoritmo, a base de dados original é dividida em N subconjuntos distintos com reposição, gerando N classificadores de árvores de decisão diferentes. Desse modo, há uma outra diferença no algoritmo de floresta aleatória se comparado ao algoritmo de árvore de decisão: enquanto o algoritmo de árvore de decisão tenta encontrar a melhor *feature*

possível para dividir o conjunto de dados, o algoritmo de floresta aleatória busca encontrar a melhor *feature* que divide cada subconjunto aleatório da base de dados original.

2.2.7 *Regressão Linear*

Em aprendizado de máquina, regressão linear é uma técnica utilizada encontrar a melhor função linear que represente os dados de treinamento fornecidos e que possa ser utilizada para realizar previsões futuras. Seu funcionamento consiste em encontrar os melhores coeficientes da equação de uma reta de modo a minimizar a diferença entre os valores de treinamento e os valores reais.

2.2.8 **Aprendizado por Transferência**

Aprendizado por transferência, também conhecido como *Transfer Learning*, é uma técnica de aprendizado de máquina que visa utilizar o conhecimento adquirido durante a execução de uma determinada tarefa a fim de resolver uma outra tarefa relacionada [34]. A vantagem de se utilizar esta técnica está na economia de tempo e recursos, visto que, em muitos casos, o aprendizado obtido anteriormente é bem útil para vários problemas distintos. Esses efeitos são ainda mais acentuados quando o problema alvo possui dados escassos.

Ajuste Fino

O Ajuste fino ou *Fine Tuning* é uma outra etapa após a aplicação da técnica de aprendizado por transferência [22]. A ideia do *Fine Tuning* é de melhorar ainda mais o desempenho final do modelo. Para isso, esse processo é realizado da seguinte forma: o modelo inicialmente é inicializado com os pesos obtidos por meio do aprendizado de transferência, sendo treinado deixando todos esses pesos fixos, com exceção dos pesos presentes nas camadas finais que irão variar neste treinamento. Por fim, as outras camadas deixarão de ter estes pesos fixos e, por tanto, estarão suscetíveis a variações de um novo treinamento a ser realizado. Esse treinamento final representa o ajuste final do modelo ao problema alvo.

2.3 **Técnicas de Balanceamento de Dados**

Ao lidar com diferentes bases de dados em problemas de aprendizado de máquina, é muito comum encontrarmos bases de dados que estão desbalanceadas, ou seja, que contêm muito mais rótulos de uma determinada classe do que de outra. Bases de dados desbalanceadas é algo que pode afetar consideravelmente a capacidade de generalização de um modelo,

de modo a classificar muito bem uma determinada classe em detrimento de outras [35]. Desse modo, pode ocorrer o fenômeno de *overfitting*, isto é, quando o modelo se ajusta perfeitamente para o conjunto de dados de treinamento mas apresentando péssimo desempenho no conjunto de dados de teste [36]. Existem muitas técnicas para lidar com o desbalanceamento de dados, como por exemplo o *Oversampling* e o *Undersampling* [35].

Oversampling

O *Oversampling*, também chamado de sobre-amostragem, é uma técnica utilizada para balancear o número de amostras. Para isso, a classe com menos amostras presentes na base de dados, chamada de classe minoritária, tem seu número de amostras aumentado de modo a ficar equilibrado com as outras classes presentes.

Existem muitas formas de realizar o *Oversampling*, como por exemplo replicar aleatoriamente amostras da classe minoritária inserindo algumas variações de modo a aumentar seu número. Outra maneira de fazer isso é utilizando a técnica *Synthetic Minority Oversampling Technique (SMOTE)*. O *SMOTE* é uma técnica proposta por Chawla *et al.* [37] e seu funcionamento é baseado na geração de amostras sintéticas da classe minoritária de modo a aumentar o seu número. De forma simples, o algoritmo de *SMOTE* funciona da seguinte maneira: o algoritmo realiza a detecção de quais amostras pertencem à classe minoritária da base de dados, seleciona uma dessas amostras de forma aleatória e seleciona os K-vizinhos mais próximos dessa amostra e os combina realizando uma média ponderada. O processo é repetido até que a base de dados fique equilibrada.

Note que a técnica de *Oversampling* não deve ser confundida com a técnica de *Data Augmentation* ou aumento de dados. Essa técnica, por outro lado, consiste em utilizar as amostras já existentes para aplicar transformações que gerem uma maior variabilidade. O *Data Augmentation* é útil para gerar um modelo de dados mais diversificado e melhorar a capacidade de generalização do modelo.

Undersampling

O *Undersampling* é uma abordagem mais simples do que o *Oversampling*. O seu propósito consiste na remoção de amostras da classe majoritária de modo a equilibrar a base de dados. Essa remoção é realizada de maneira aleatória e torna o modelo menos enviesado pela classe majoritária. No entanto, o *Undersampling* pode levar a perda de informações importantes que estavam presentes nas amostras excluídas.

		Classe Verdadeira	
		Positivo	Negativo
Classe Predita	Positivo	VP	FP
	Negativo	FN	VN

Figura 2.13: Exemplo de matriz de confusão para 2 classes (Fonte: Elaborada pelo autor).

2.4 Métricas de Avaliação de Desempenho

Esta seção tem o objetivo de mostrar as principais métricas de avaliação de desempenho para avaliar os modelos abordados neste trabalho. Para o problema proposto de realizar a classificação binária de câncer de mama, as seguintes métricas podem ser utilizadas: Acurácia, Precisão, Revocação e *F1 Score*. Essas métricas são calculadas a partir da matriz de confusão.

Matriz de Confusão

Uma matriz de confusão é uma estrutura de dados no formato de tabela com o objetivo de sumarizar o desempenho de classificação de um determinado modelo [38]. Essa estrutura normalmente apresenta o tamanho de $n \times n$, em que n se refere ao número de classes que fazem parte do problema de classificação. Um dos eixos da matriz de confusão representa a classe real do problema, enquanto o outro eixo representa a classe predita pelo modelo utilizado. A Figura 2.13 representa uma matriz de confusão e as células das classes verdadeiras e das classes preditas.

Cada posição na matriz de confusão representa um determinado significado. Para o problema de classificação de câncer de mama, a classe positiva é representada pelas amostras benignas e a classe negativa é representada pelas amostras malignas. Observando a Figura 2.13, temos os seguintes rótulos:

- Verdadeiro Positivo (VP): As amostras desta célula representam os dados que pertencem à classe positiva e que foram classificados corretamente pelo modelo;

- Falso Positivo (FP): As amostras desta célula representam os dados que pertencem à classe negativa e que foram classificados como pertencente à classe positiva pelo modelo;
- Verdadeiro Negativo (VN): As amostras desta célula representam os dados que pertencem à classe negativa e que foram classificados corretamente pelo modelo;
- Falso Negativo (FN): As amostras desta célula representam os dados que pertencem à classe positiva e que foram classificados como pertencente à classe negativa pelo modelo.

Acurácia

A acurácia é uma das métricas mais importantes para problemas de classificação. Ela representa a taxa percentual de todas as amostras que foram classificadas corretamente pelo modelo. A acurácia não deve ser levada em consideração por si só, dependendo da análise conjunta com outras métricas [7]. O principal problema da acurácia ocorre quando uma classe é melhor generalizada por um modelo do que a outra, fazendo com que essa métrica apresente valores altos da mesma forma. A Equação 2.1 representa o cálculo da acurácia.

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN}. \quad (2.1)$$

Precisão

A Precisão, por outro lado, representa a taxa percentual de todas as amostras que foram classificadas como positivas pelo modelo que são realmente positivas. Desse modo, a precisão é bem útil quando é desejado minimizar a quantidade de Falsos Positivos. De maneira análoga a métrica de acurácia, a precisão, por si só, não é suficiente para categorizar o desempenho de um modelo. Um modelo pode ter alta precisão e ter um desempenho ruim por apresentar muitos Falsos Negativos. A Equação 2.2 representa o cálculo da precisão.

$$Precisão = \frac{VP}{VP + FP}. \quad (2.2)$$

Revocação

A métrica de Revocação, também é conhecida como Sensibilidade ou Taxa de Verdadeiros Positivos, é uma métrica muito importante de avaliação de modelos de classificação. Esta métrica representa a taxa percentual de todas as amostras positivas que foram calculadas pelo modelo em relação ao total de instâncias reais da classe positiva. A Equação 2.3 representa o cálculo desta métrica.

$$Revocação = \frac{VP}{VP + FN}. \quad (2.3)$$

F1 Score

O *F1 Score* é uma métrica importante que combina a precisão e a revocação em uma medida única de desempenho de um modelo de classificação. Dessa forma, o *F1 Score* é uma média harmônica entre essas duas medidas, calculado de acordo com a Equação 2.4. Esta métrica é muito importante quando queremos ter previsões positivas de alta qualidade e, ao mesmo tempo, uma boa capacidade de capturar de maneira eficaz todas as instâncias de classe positiva.

$$F1\ Score = 2 \times \frac{precisão \times revocação}{precisão + revocação}. \quad (2.4)$$

Capítulo 3

Revisão de Literatura

O objetivo deste capítulo é o de apresentar uma breve revisão de literatura acerca das principais técnicas de aprendizado de máquina para tarefas de classificação de câncer de mama. O primeiro artigo de grande importância para a elaboração deste trabalho é publicado por Chattopadhyay *et al.* [14]. Nesse trabalho é proposto um modelo de *Deep Learning*, chamado de *Multi-scale Dual Residual Recurrent Network (MTRRE-Net)*, com o objetivo de atenuar o problema do *vanishing gradient* para redes neurais profundas. A base de dados BreakHis utilizada neste trabalho possui um total de 7909 imagens, divididas entre benignas e malignas e com isso os autores conseguem obter uma acurácia de 97.12%, precisão de 97%, revocação de 96% e 96% de *F1 Score*. Comparando com outras redes de *Deep Learning* famosas, os autores verificam que o modelo proposto supera as outras redes de *Deep Learning* com uma boa margem.

Bhausahab e Kashyap [39] propõem um novo método de classificação de câncer de mama chamado de *Shuffled Shepherd Deer Hunting Optimization-based Deep Neural Network (SSDHO-based DNN)*. O método proposto é desenvolvido para realizar classificação multi-classe, classificando o tumor em uma dentre seis classes distintas. Para isso, duas bases de dados diferentes são utilizadas: uma contendo as imagens originais em conjunto com os arquivos de *ground truth* e outra contendo um total de 162 slides com os quais foram geradas 277524 outras imagens menores. Os autores conseguem atingir uma acurácia de 95.71%, sensibilidade de 79.03%, precisão de 82.32% e especificidade de 94.26%. Os resultados obtidos são comparados com os resultados obtidos de outras técnicas, como *Deep Convolutional Neural Network (DCNN)*, *Stacked Sparse Autoencoder (SSAE)* e *Fully Convolutional autoencoder* e foi verificado que o método proposto supera os demais.

Majumdar, Pramanik e Sarkar [16] publicam um artigo propondo um modelo *ensemble* baseado em ranqueamento para detecção de câncer de mama que combina três modelos de CNN diferentes aplicados com *transfer learning*. O modelo proposto utiliza a função gama para realizar a classificação binária das imagens. A base de dados utilizada contém

um total de 7909 amostras dentro das categorias benigna e maligna. As amostras são de 82 pacientes e possuem diferentes tipos de ampliação: 40X, 100X, 200X e 400X. Com isso, os autores conseguem atingir 99.16%, 98.24%, 98.67% e 96.16% de acurácia, respectivamente para as ampliações citadas anteriormente. Os resultados obtidos superam com grande margem outras abordagens *state-of-the-art*.

Outro trabalho relevante para a classificação de câncer de mama é proposto por Karthik, Menaka e Siddharth [17]. Neste trabalho, é proposto um modelo *ensemble* de duas arquiteturas CNN integradas. As características das imagens são extraídas paralelamente por duas arquiteturas CNN e combinadas para atingir melhor desempenho. A base de dados utilizada possui um total de 7909 amostras de imagens histopatológicas de câncer de mama obtidas por biópsia. O modelo proposto é comparado com outros modelos *state-of-the-art* e atinge uma acurácia de 99.55%, superando os outros modelos comparados.

Chattopadhyay *et al.* [12] apresentam um novo modelo de *Deep Learning* chamado *Dense Residual Dual-shuffle Attention Network (DRDA-Net)*. O modelo proposto é construído de modo a evitar o problema do *vanishing gradient* e *overfitting* e, além disso, é construído com atenção aos canais das imagens, o que permite ao modelo aprender padrões mais complexos. Diferentes variações do modelo proposto e outros modelos considerados *state-of-the-art* são comparados na tarefa de classificação de câncer de mama, e os resultados mostram que o modelo proposto é superior aos demais comparados. O modelo proposto também permite que ele seja utilizado para obter resultados em tempo real. Para isso, é utilizada a base de dados *BreaKHis*, contendo imagens de 82 pacientes com resolução de 752×582 *pixels* e diferentes fatores de ampliação. Para as imagens com aplicação de 200X, o modelo atinge uma acurácia de 97.43%, precisão de 96%, revocação de 99% e *F1 Score* de 97.44%.

Sheeba *et al.* [13] apresentam um trabalho de um modelo *ensemble* combinando métodos de *Deep Learning* para classificação de câncer de mama e extração de características. As imagens utilizadas nesse trabalho são coletadas por um microscópio conectado a um módulo baseado em Internet of Things (IOT). Posteriormente, as imagens coletadas passam por um filtro Gaussiano para redução de ruído e são segmentadas utilizando uma CNN de contorno ativo antes de serem classificadas. Os resultados atingidos são comparados com outras técnicas já existentes e o modelo proposto obteve uma acurácia de classificação de 96%, acurácia de detecção de 92%, precisão média de 82%, sensibilidade de 92%, especificidade de 91% e *root mean squared error* de 70%.

Oyelade *et al.* [40], em 2022, propõem um modelo de *Deep Learning* para classificação multi-classe e localização de anormalidades aplicado em imagens histopatológicas utilizando duas combinações de arquiteturas CNN. Para isso, são utilizadas e combinadas

duas bases de dados diferentes: o *BreaKHis*, contendo 7909 imagens histopatológicas de 82 pacientes distintos e o *BACH*, contendo por volta de 400 imagens. O modelo é comparado com outras técnicas de *Deep Learning*, e obteve os seguintes resultados: 97.08% de acurácia, sendo resultados superiores se comparado com outras técnicas mencionadas no trabalho.

Yu *et al.* [41] criam um modelo de classificação de câncer de mama baseado em árvores de decisão convolucionais, capaz de realizar a tarefa de classificação binária em amostras malignas e benignas e, também, de realizar a tarefa de classificação multi-classe, permitindo classificar dentre oito categorias diferentes de câncer de mama. A base de dados utilizada neste trabalho é a base *BreaKHis*, contendo 7909 amostras de 82 pacientes. A base de dados é dividida em oito categorias com quatro ampliações diferentes. Os resultados obtidos são comparados com outros modelos *state-of-the-art* e o modelo apresentado consegue atingir 99.75% de acurácia para a tarefa de classificação binária e 95.69% de acurácia para a tarefa de classificação multi-classe para oito classes distintas. Os resultados apresentados superam com grande margem os outros modelos *state-of-the-art*.

Joseph *et al.* [42] aplicam técnicas para melhorar a tarefa de classificação multi-classe para câncer de mama, de modo a melhorar a acurácia e reduzir o *overfitting* amplamente presente nesse problema. Para isso, utilizam técnicas criadas manualmente para extração das características das imagens, *data augmentation* para diminuir o *overfitting* e uma Dense Neural Network (DNN) para realizar a multi-classificação. As técnicas apresentadas são aplicadas na base de dados *BreaKHis* e comparadas com outros trabalhos relacionados. Desse modo, o trabalho apresentado atinge uma acurácia de 97.89% para ampliação de 40X, 97.60% de acurácia para imagens com ampliação de 100X, 96.10% de acurácia para imagens com ampliação de 200X e 96.84% de acurácia para imagens com ampliação de 400X. Os resultados obtidos são superiores aos outros trabalhos comparados.

Maleki, Raahemi e Nasiri [43] propõem um novo modelo de *transfer learning* para classificar tumores de câncer de mama em imagens histopatológicas. Para chegar no modelo ótimo, 18 combinações de modelos pré-treinados e algoritmos de *gradient boosting* são analisados. A base de dados utilizada neste trabalho é a base *BreaKHis*, que contém aproximadamente 7909 amostras divididas em 4 ampliações diferentes: 40X, 100X, 200X e 400X. O modelo proposto é comparado com outros modelos *state-of-the-art* e atinge maior acurácia se comparado com eles. Os resultados obtidos são de 93.6% de acurácia para ampliação de 40X, 91.3% de acurácia para ampliação de 100X, 93.8% de acurácia para ampliação de 200X e 89.1% de acurácia para ampliação de 400X.

Os trabalhos apresentados neste capítulo evidenciam o avanço das técnicas de aprendizado de máquina para a classificação de câncer de mama. Os modelos apresentados alcançam altas taxas de acurácia, precisão e revocação, superando consideravelmente

abordagens anteriores. A Tabela 3.1 apresenta uma síntese dos trabalhos discutidos no presente capítulo, detalhando o ano de publicação, autores, problemas de classificação abordados, técnicas utilizadas e resultados alcançados. O presente trabalho tem como objetivo contribuir para o avanço das técnicas de aprendizado de máquina na classificação de câncer de mama, propondo a construção de um modelo *ensemble* com uma arquitetura distinta dos demais trabalhos apresentados neste capítulo.

Tabela 3.1: Trabalhos abordados neste capítulo.

Ano de publicação	Autores	Problemas de classificação	Técnicas	Resultados
2022	Chattopadhyay <i>et al.</i> [14]	Classificação binária	<i>MTRRE-Net</i>	Acurácia: 97.12% Precisão: 97% Revocação: 96% <i>F1 Score</i> : 96%
2023	Bhauasaheb e Kashyap [39]	Classificação multi-classe	<i>SSDHO-based DNN</i>	Acurácia: 95.71% Sensibilidade: 79.03% Precisão: 82.32% Especificidade: 94.26%
2022	Majumdar, Pramanik e Sarkar [16]	Classificação binária	<i>Ensemble CNN</i>	com Acurácia: 40X: 99.16% 100X: 98.24% 200X: 98.67% 400X: 96.16%
2022	Karthik, Menaka e Siddharth [17]	Classificação binária	<i>Ensemble CNN</i>	com Acurácia: 99.55%
2022	Chattopadhyay <i>et al.</i> [12]	Classificação binária	<i>DRDA-Net</i>	Acurácia: 97.43% Precisão: 96% Revocação: 99% <i>F1 Score</i> : 97.44%
2022	Sheeba <i>et al.</i> [13]	Classificação binária	CNN	Acurácia de classificação: 96% Acurácia de detecção: 92% Precisão média: 82% Sensibilidade: 92% Especificidade: 91% <i>Root Mean Squared Error</i> : 70%
2022	Oyelade <i>et al.</i> [40]	Classificação multi-classe	DNN	Acurácia: 97.08%
2023	Yu <i>et al.</i> [41]	Classificação binária e multi-classe	Árvores de decisão convolucionais	Acurácia classificação binária: 99.75% Acurácia classificação multi-classe: 95.69%
2022	Joseph <i>et al.</i> [42]	Classificação multi-classe	DNN	Acurácia: 40X: 97.89% 100X: 97.60% 200X: 96.10% 400X: 96.84%
2023	Maleki, Raahemi e Nasiri [43]	Classificação binária	<i>DenseNet201-XGBoost</i>	Acurácia: 40X: 93.6% 100X: 91.3% 200X: 93.8% 400X: 89.1%

Capítulo 4

Metodologia

O presente capítulo descreve em detalhes a abordagem utilizada para abordar o problema de classificação de imagens histopatológicas de tumores mamários. Inicialmente, a base de dados empregada é discutida, incluindo sua origem, natureza e características essenciais. Em seguida, as técnicas de pré-processamento de dados adotadas são apresentadas, e, por fim, os experimentos propostos para o modelo sugerido neste trabalho e para os modelos *state-of-the-art* utilizados.

4.1 Base de dados

Neste trabalho, a base de dados *Breast Cancer Histopathological Database (BreakHis)* [8] é utilizada com o objetivo de classificar imagens de microscópicas de tumores presentes em tecidos mamários. Essa base de dados foi construída em colaboração com o Laboratório de Anatomia e Citopatologia do Paraná, Brasil e suas imagens foram coletadas durante um estudo clínico conduzido de janeiro a dezembro de 2014. Todas as amostras foram coletadas pelo método *Surgical Open Biopsy (SOB)*, também chamado de mastectomia parcial ou biópsia excisional cirúrgica aberta. Nesse procedimento, uma amostra de tecido grande é retirada em comparação com outros métodos, sendo realizado em um hospital com anestesia geral [8]. Após a coleta, as amostras foram preparadas para estudo histológico e rotuladas por patologistas.

A base de dados contém um total de 7909 imagens representando lâminas de biópsia de tecido mamário de um total de 82 pacientes utilizando quatro tipos de fatores de ampliação: 40X, 100X, 200X e 400X, como ilustram as Figuras 4.1 e 4.2. O quadrado amarelo presente na Figura 4.1 representa o trecho onde o zoom será aplicado de modo a gerar os fatores de ampliação. Dessas imagens, 2480 são referentes a amostras benignas e 5429 são referentes a amostras malignas. Todas as imagens estão no formato PNG, de 700×460 *pixels* e 3 canais RGB. A Tabela 4.1 representa como as amostras estão

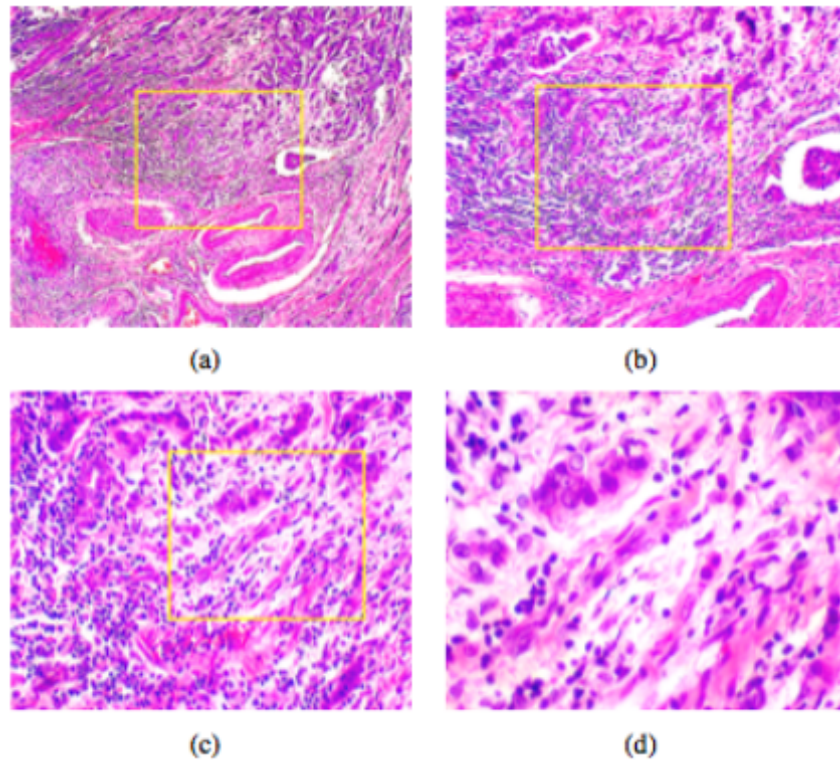


Figura 4.1: Imagens de tumores malignos da base de dados: (a) 40X, (b) 100X, (c) 200X e (d) 400X (Fonte: [8]).

distribuídas dentro da base de dados. Note que, pela Tabela 4.1, há uma quantidade bem maior de imagens de classes malignas do que benignas, apresentando quase o dobro de amostras malignas do que benignas. Isso evidencia a necessidade de utilizarmos métodos de balanceamento de dados. A base de dados é separada de maneira aleatória e estratificada em três conjuntos distintos: 70% das amostras para o conjunto de treinamento, 20% das amostras para o conjunto de testes e 10% das amostras para o conjunto de validação. Para que não ocorra contaminação entre os três conjuntos, cada paciente tem todas as suas amostras inseridas exclusivamente em um único conjunto.

Tabela 4.1: Estrutura da base de dados.

Fator de ampliação	Amostras benignas	Amostras malignas	Total
40X	652	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total de imagens	2480	5429	7909

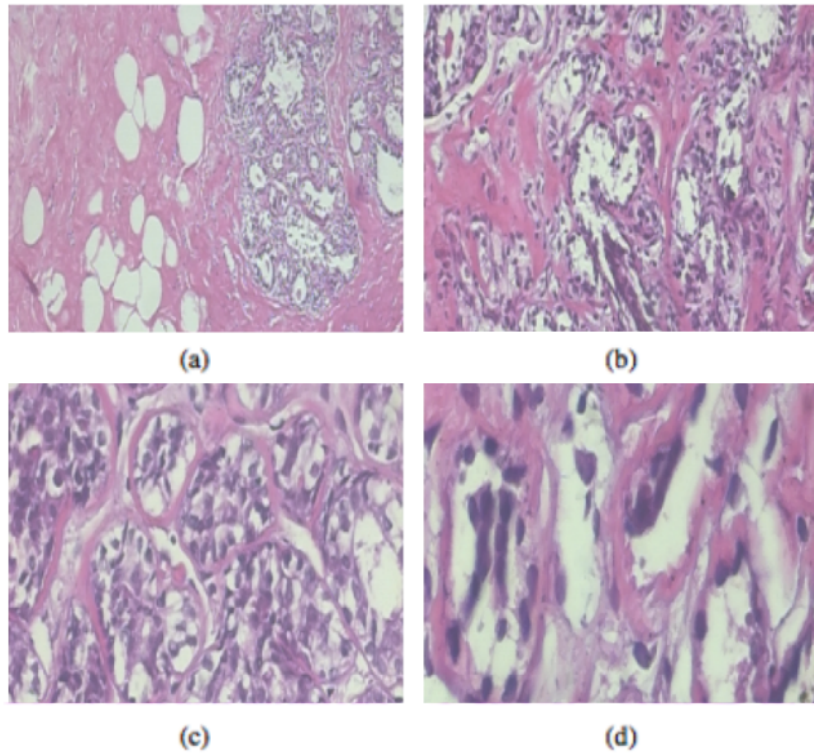


Figura 4.2: Imagens de tumores benignos da base de dados: (a) 40X, (b) 100X, (c) 200X e (d) 400X (Fonte: [8]).

4.2 Pré-processamento de imagens

Para a realização dos experimentos propostos neste trabalho, foi necessário aplicar algumas transformações de pré-processamento nas imagens da base de dados apresentadas na Seção 4.1, de maneira a obter uma melhor adequação dos dados de entrada aos modelos utilizados.

Neste trabalho, CNNs previamente treinadas, como a *ResNet50* [25], a *VGG-16* [15] e a *InceptionV3* [26] requerem que as imagens de entrada apresentem uma resolução específica. Desse modo, as imagens da base de dados foram redimensionadas pelo método de interpolação bilinear para apresentarem a resolução de $224 \times 224 \times 3$ pixels, compatíveis com estes modelos. Além do redimensionamento, para cada uma destes modelos de CNN previamente treinados com pesos da base de dados *ImageNet* [44], é necessário aplicar um pré-processamento personalizado e específico para cada modelo.

Para as redes *VGG-16* e *ResNet50*, é necessário que as imagens de entrada sejam convertidas do formato Red, Green and Blue (RGB) para o formato Blue, Green and Red (BGR) e que cada canal de cor tenha o seu centro ajustado com base no conjunto de dados da *ImageNet* [44], sem realizar escalonamento. A rede *InceptionV3*, por outro lado, tem a necessidade de que as imagens de entrada tenham os valores dos seus pixels

convertidos entre -1 e 1.

Balanceamento e aumento dos dados

A base de dados BreakHis, como visto na Seção 4.1, é desbalanceada, possuindo muito mais amostras da classe maligna do que da classe benigna. Desse modo, é necessário utilizar técnicas de balanceamento dos dados, como o *Oversampling*. Para isso, neste trabalho a técnica SMOTE [37] foi utilizada, gerando amostras sintéticas da classe minoritária de modo a igualar o seu número com a quantidade de amostras da classe majoritária, conforme foi discutido na Seção 2.3. A Tabela 4.2 representa como as amostras são distribuídas dentro da base de dados após o balanceamento.

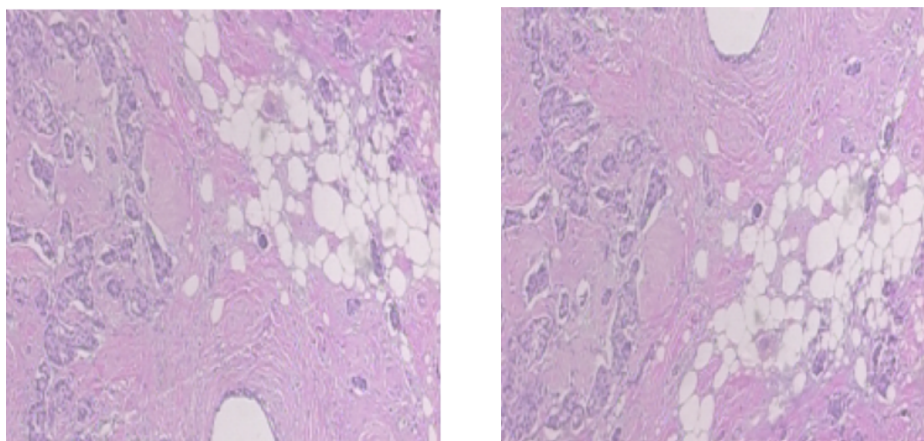
Tabela 4.2: Estrutura da base de dados.

Fator de ampliação	Amostras benignas	Amostras malignas	Total
40X	1370	1370	2740
100X	1437	1437	2874
200X	1390	1390	2780
400X	1232	1232	2464
Total de imagens	5429	5429	10858

Para que as amostras utilizadas apresentem maior diversidade de modo a diminuir a incidência de *overfitting*, técnicas de aumento de dados, ou do inglês *Data Augmentation*, foram empregadas. Para isso, ainda na etapa de pré-processamento, diversas transformações foram aplicadas às imagens do conjunto de treinamento. Essas modificações visam criar variações controladas, permitindo que o modelo aprendesse padrões mais robustos ao longo do processo de treinamento. As seguintes transformações foram utilizadas:

- Rotações aleatórias de até 15°;
- Deslocamento lateral de até 10%;
- Deslocamento vertical de até 10%;
- Espelhar a imagem horizontalmente, com probabilidade de ocorrência aleatória;
- Espelhar a imagem verticalmente, com probabilidade de ocorrência aleatória.

Estas transformações foram aplicadas diretamente a partir das imagens originais do conjunto de treinamento da base de dados. A Figura 4.3 ilustra as imagens antes e depois da aplicação do processo de *Data Augmentation*.



(a) Imagem original

(b) Imagem espelhada verticalmente

Figura 4.3: Exemplo de imagem gerada pela técnica de *Data Augmentation*.

4.3 Experimentos Propostos

Para o problema deste trabalho de realizar a classificação binária de amostras de imagens de lâminas de tumores mamários nas classes benigno e maligno, dois grandes grupos de experimentos foram propostos.

O primeiro grupo de experimentos tem o objetivo de construir um modelo *ensemble* utilizando a estratégia *stacking* e refiná-lo de modo a obter o melhor desempenho para a tarefa de classificação. O segundo grupo de experimentos tem o objetivo de construir e utilizar modelos considerados *state-of-the-art* e refiná-los para que eles tenham o melhor desempenho possível dentro da base de dados utilizada.

Para avaliar o desempenho dos modelos utilizados nos dois grandes grupos de experimentos, as métricas de avaliação de desempenho discutidas na Seção 2.4 foram utilizadas. Desse modo, é possível comparar o desempenho do modelo proposto com os outros modelos considerados *state-of-the-art*.

4.3.1 Modelo Ensemble proposto

Neste trabalho, foi proposto um modelo *ensemble* utilizando a estratégia *stacking*. Para isso, utilizamos três classificadores base: Uma CNN simples completa [21], um preditor *Random Forest* [33] e uma rede *ResNet50* [25], um dos modelos *state-of-the-art* considerados. Para o meta-classificador, o preditor de regressão linear [7] foi utilizado. Como a rede *ResNet50* foi utilizada como parte do modelo *ensemble*, adicionar outra rede CNN complexa, como a *InceptionV3*, pode aumentar a redundância do modelo sem fornecer benefícios significativos. Desse modo, foi preferível o uso de uma rede CNN simples.

As imagens de entrada da base de dados foram separadas em três grupos: teste, treinamento e validação. Para a construção do modelo *ensemble*, três etapas foram necessárias. A primeira delas consiste em utilizar os conjuntos de treinamento e validação para realizar o treinamento de todos os modelos base, como ilustra a Figura 4.4. A técnica de *Early Stopping* foi utilizada com paciência de 5 épocas para os modelos base *ResNet50* e para a rede CNN simples utilizada. Desse modo, caso o modelo não apresente melhora após 5 épocas, o treinamento é encerrado antecipadamente e o melhor *score* é preservado.

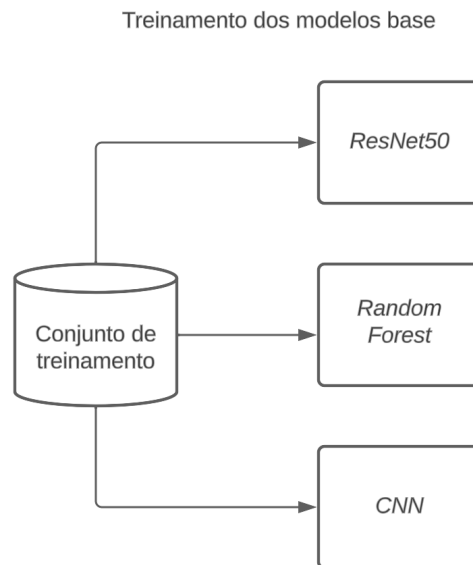


Figura 4.4: Treinamento dos modelos base.

A segunda etapa consiste em utilizar os modelos base já treinados para obter as previsões do conjunto de validação. Em seguida, essas previsões serão utilizadas para a fase de treinamento do meta-classificador utilizado, que é constituído de um regressor linear simples. A Figura 4.5 ilustra este processo.

Por fim, com o modelo *ensemble* treinado, utilizamos os classificadores base novamente para obter as previsões no conjunto de dados de teste e, para obter as métricas de acurácia, precisão, revocação e *F1 Score* finais do modelo proposto, as previsões como entrada do meta-classificador foram utilizadas, obtendo, assim, as métricas finais desejadas. A Figura 4.6 ilustra o funcionamento do modelo *ensemble* proposto.

Com o modelo *ensemble* construído, três experimentos distintos foram conduzidos para cada fator de ampliação diferente: $40X$, $100X$, $200X$, e $400X$. O primeiro experimento utiliza as imagens pré-processadas de entrada da base de dados apresentada na Seção 4.1 sem passar pelo processo de *Data Augmentation* e *Oversampling*. O segundo experimento utiliza essas imagens com o processo de *Data Augmentation* na etapa de

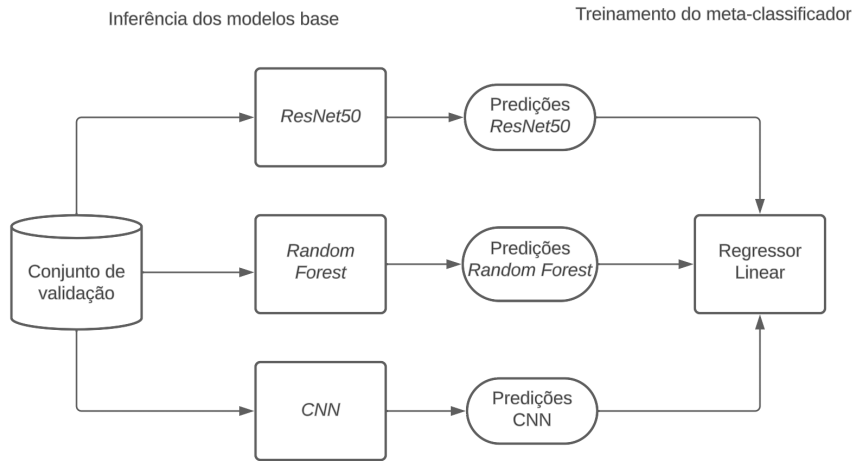


Figura 4.5: Treinamento do meta-classificador.

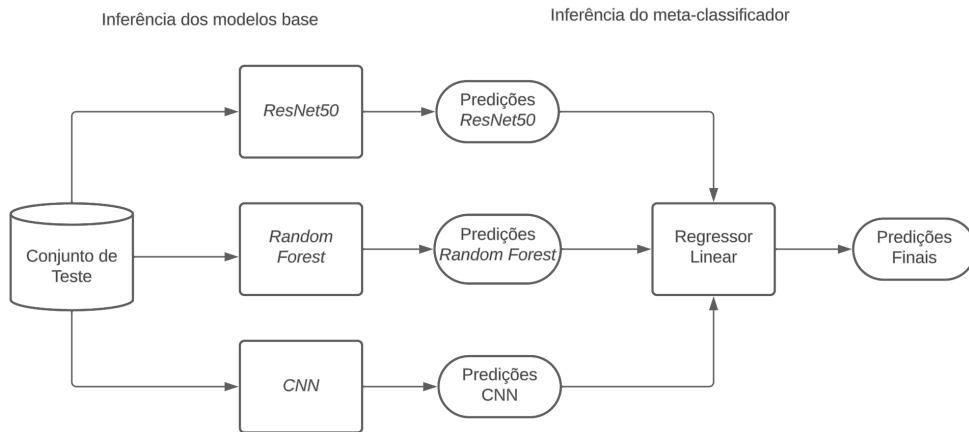


Figura 4.6: Modelo *ensemble* proposto.

pré-processamento, porém sem *Oversampling*. Por fim, o terceiro experimento utiliza essas imagens com *Data Augmentation* ou não, a depender do que resultou em resultados melhores dos experimentos anteriores, e utilizando SMOTE.

A técnica de *Random Search* [45] foi utilizada em todos os experimentos para obter os melhores hiperparâmetros dos modelos base utilizados. Para a rede CNN simples, foram considerados como hiperparâmetros o número de camadas convolucionais, o tamanho do *kernel* e o número de neurônios da camada totalmente conectada. Para o preditor *Random Forest*, foi considerado o número máximo de profundidade da árvore como hiperparâmetro do *Random Search*. Por fim, para a rede *ResNet50*, foram considerados como hiperparâmetros a taxa de aprendizado, o número de neurônios da camada totalmente conectada e o otimizador aplicado.

O *Random Search* foi executado antes da etapa de treinamento considerando três

combinações distintas de hiperparâmetros. Para cada combinação de hiperparâmetros, 10 épocas foram executadas a fim de se avaliar o desempenho daquele conjunto de hiperparâmetros. A Tabela 4.3 ilustra todos os hiperparâmetros utilizados na construção do modelo *ensemble*.

Tabela 4.3: Tabela de hiperparâmetros utilizados no modelo *ensemble*.

Modelo	Hiperparâmetro	Faixa de Valores
CNN	Numero de camadas convolucionais	1, 2, 3
	Tamanho do <i>kernel</i>	3, 5, 7
	Número de neurônios	64, 128, 192, 256
<i>ResNet50</i>	Taxa de aprendizado	1e-4, 1e-3, 1e-2
	Número de neurônios	64, 128, 192, 256
	Algoritmo de otimização	<i>Adam</i> , <i>SGD</i> , <i>Adamax</i> , <i>RMSProp</i>
<i>Random Forest</i>	Profundidade máxima	5, 10, 15

4.3.2 Modelos *state-of-the-art*

Neste trabalho, os seguintes modelos considerados *state-of-the-art* [12] foram utilizados: as redes *ResNet50*, *VGG-16* e *InceptionV3*. Para cada uma dessas redes, experimentos análogos aos discutidos na Seção 4.3.1 foram conduzidos.

As imagens de entrada da base de dados apresentada na Seção 4.1 foram separadas novamente nos grupos de teste, treinamento e validação. Três experimentos distintos foram realizados para as redes consideradas, incluindo os fatores de ampliação de 40X, 100X, 200X e 400X.

Assim como no modelo *ensemble* proposto, o primeiro grupo de experimentos utilizará as imagens de entrada sem passar pelo processo de *Data Augmentation* e *Oversampling*. O segundo grupo utilizará as imagens de entrada com o processo de *Data Augmentation* sem utilizar *Oversampling*. O terceiro grupo utilizará as imagens de entrada com *Data Augmentation* ou não, a depender dos resultados dos experimentos anteriores e utilizando *SMOTE*.

Em todos os experimentos realizados, a técnica de *Random Search* [45] foi utilizada, de modo a buscar os melhores hiperparâmetros para cada um dos modelos. Os hiperparâmetros considerados foram a taxa de aprendizado, o número de neurônios da camada totalmente conectada e o otimizador aplicado. O *Random Search* foi executado antes da etapa de treinamento considerando três combinações distintas de hiperparâmetros. Para cada combinação de hiperparâmetros, 10 épocas foram executadas a fim de se avaliar o desempenho daquele conjunto de hiperparâmetros. A Tabela 4.4 ilustra os hiperparâmetros utilizados na construção dos modelos *state-of-the-art*.

Tabela 4.4: Tabela de hiperparâmetros utilizados nos modelos *state-of-the-art*.

Hiperparâmetro	Faixa de Valores
Taxa de aprendizado	1e-4, 1e-3, 1e-2
Número de neurônios	64, 128, 192, 256
Algoritmo de otimização	<i>Adam, SGD, Adamax, RMSProp</i>

Além disso, é importante salientar que a técnica de *Early Stopping* foi utilizada para todos os modelos *state-of-the-art* com paciência de 5 épocas. Conforme descrito na Subseção 4.3.2, caso o modelo não apresente melhora após 5 épocas, o treinamento é encerrado antecipadamente e o melhor *score* é preservado. Os modelos *state-of-the-art* utilizados foram treinados utilizando a técnica de aprendizado de transferência e ajuste fino, como discutido na Subseção 2.2.8.

A Tabela 4.5 ilustra um resumo de todos os experimentos realizados na condução deste trabalho. Após a execução de todos os experimentos, os resultados do modelo *ensemble* construído será comparado com os melhores cenários obtidos dos modelos *state-of-the-art*.

Tabela 4.5: Tabela de Experimentos Propostos.

Modelo	Experimento I - sem <i>Data Augmentation</i> e sem <i>Oversampling</i>	Experimento II - com <i>Data Augmentation</i> e sem <i>Oversampling</i>	Experimento III - com <i>Oversampling</i>
ResNet50	40X	40X	40X
	100X	100X	100X
	200X	200X	200X
	400X	400X	400X
VGG-16	40X	40X	40X
	100X	100X	100X
	200X	200X	200X
	400X	400X	400X
InceptionV3	40X	40X	40X
	100X	100X	100X
	200X	200X	200X
	400X	400X	400X
Ensemble Proposto	40X	40X	40X
	100X	100X	100X
	200X	200X	200X
	400X	400X	400X

Capítulo 5

Resultados

Este capítulo tem o intuito de apresentar os resultados dos experimentos discutidos no Capítulo 4 e as análises correspondentes no modelo *ensemble* proposto e nas redes *CNN* utilizadas. Desse modo, os dados obtidos serão comparados utilizando ou não aumento de dados e *Oversampling*. Além disso, as ferramentas utilizadas para a execução de todos os experimentos propostos serão apresentadas.

5.1 Ferramentas utilizadas

Os experimentos realizados neste trabalho e descritos no Capítulo 4 foram realizados utilizando a linguagem de programação *Python* e bibliotecas que forneceram a abstração necessária para realizar a construção e o treinamento dos modelos descritos¹.

Para que os modelos sejam executados de forma mais eficiente e com menor tempo de treinamento necessário, a plataforma *Google Colab* foi utilizada. Os experimentos foram realizados em uma máquina virtual com as seguintes especificações: CPU Intel(R) Xeon(R) CPU @ 2.00GHz, 12 GiB de memória RAM e GPU Nvidia A100 32 GiB. Em alguns experimentos, a GPU Nvidia V100 16 GiB foi utilizada mantendo os outros componentes inalterados.

5.2 Experimentos do Modelo Ensemble Proposto

Neste trabalho, conforme foi discutido no Capítulo 4, o primeiro grupo de experimentos são referentes ao modelo Ensemble construído. Os experimentos conduzidos nesta seção visam extrair o melhor desempenho possível do modelo ensemble construído, para, então, comparar com o desempenho de outras redes previamente treinadas.

¹As bibliotecas utilizadas foram: *Tensorflow*, *Keras*, *Scikit-learn*, *Numpy*, *Matplotlib*, *Glob*, *CV2*, *TQDM*, *Keras Tuner* e *Imblearn*.

5.2.1 Sem aumento de dados e sem *Oversampling*

Como discutido na Subseção 4.3.1, os experimentos iniciais conduzidos no modelo *ensemble* proposto objetivam alcançar o melhor desempenho possível por meio da abordagem utilizando múltiplos classificadores. Além disso, objetiva-se verificar a influência do fator de ampliação no conjunto de imagens da base de dados e o desempenho de cada classificador presente no modelo proposto de forma individual.

Nesse grupo de experimentos, a rede CNN simples utilizada e a rede *ResNet50* foram treinadas considerando um total de 100 épocas, utilizando a técnica de *early stopping* com paciência de 5 épocas. O classificador *Random Forest* foi treinado considerando 100 estimadores distintos. A técnica de *early stopping* foi realizada a fim de se evitar o sobreajuste dos modelos. A Tabela 5.1 mostra os melhores hiperparâmetros obtidos para cada modelo.

Tabela 5.1: Tabela de hiperparâmetros para os modelos base.

Modelo	Hiperparâmetro	Melhores valores
CNN	Numero de camadas convolucionais	3
	Tamanho do <i>kernel</i>	3
	Número de neurônios	64
<i>ResNet50</i>	Taxa de aprendizado	1e-3
	Número de neurônios	64
	Algoritmo de otimização	<i>Adam</i>
<i>Random Forest</i>	Profundidade máxima	15

A Tabela 5.2 mostra os resultados obtidos para o conjunto de teste nesse grupo de experimentos para o modelo *ensemble* final proposto e as métricas de cada modelo base individualmente, para cada fator de ampliação. Dentro de cada modelo base, as melhores métricas foram destacadas em negrito.

Ao observar os dados da Tabela 5.2, verifica-se que, para o fator de ampliação de 40X, a rede *ResNet50* obteve melhor resultado, apresentando 76.52% de acurácia e 85.43% de *F1 Score*. O desempenho foi muito próximo ao modelo *ensemble* proposto, que atingiu 75.73% de acurácia e 85.35% de *F1 Score*. Para o fator de ampliação de 100X, o melhor desempenho foi do modelo *Random Forest*, que atingiu 75.96% de acurácia e 83.57% de *F1 Score*. O modelo *Random Forest* também atingiu melhor desempenho para o fator de ampliação de 200X, com acurácia de 87.50% e 91.30% de *F1 Score*. Por fim, para o fator de ampliação de 400X, a rede CNN foi a que atingiu melhor resultado, com acurácia de 85.52% e *F1 Score* de 90.50%.

Além disso, é possível observar que, dentre todos os modelos apresentados na Tabela 5.2, o modelo *Random Forest* foi o que mais se destacou em relação aos outros nas métricas de acurácia, precisão e *F1 Score*. No entanto, o modelo *ensemble* apresentou a métrica

Tabela 5.2: Métricas obtidas dos classificadores base individuais e do modelo *ensemble* proposto sem aumento de dados e sem *Oversampling*.

Modelo	Ampliação	Acurácia	Precisão	Revocação	F1 Score
<i>CNN</i>	40X	0.7124	0.7124	1.0000	0.8320
	100X	0.7059	0.7059	1.0000	0.8276
	200X	0.7396	0.7396	1.0000	0.8503
	400X	0.8552	0.8901	0.9205	0.9050
<i>ResNet50</i>	40X	0.7652	0.7654	0.9667	0.8543
	100X	0.7519	0.7973	0.8696	0.8319
	200X	0.8646	0.8791	0.9472	0.9119
	400X	0.8130	0.8414	0.9243	0.8809
<i>Random Forest</i>	40X	0.7256	0.8347	0.7667	0.7992
	100X	0.7596	0.8074	0.8659	0.8357
	200X	0.8750	0.9403	0.8873	0.9130
	400X	0.7960	0.8504	0.8826	0.8661
<i>Ensemble Proposto</i>	40X	0.7573	0.7486	0.9926	0.8535
	100X	0.7519	0.7973	0.8696	0.8319
	200X	0.8646	0.8791	0.9472	0.9119
	400X	0.8244	0.8217	0.9773	0.8927

de revocação com desempenho bastante superior. Isso indica que o modelo *Random Forest* obteve grande capacidade de classificar corretamente as instâncias benignas, entre aquelas que o modelo previu como benignas. Por outro lado, o modelo *ensemble* conseguiu identificar corretamente todas as instâncias positivas reais.

A Figura 5.1 representa a matriz de confusão do modelo *ensemble* proposto para as ampliações de 40X, 100X, 200X e 400X. Nessa figura, é possível perceber a grande capacidade do modelo de classificar corretamente as amostras malignas e de forma errônea as amostras benignas. Isso possivelmente é causado pelo desbalanceamento da base de dados utilizada, que não passou por um processo de balanceamento no conjunto de dados de treinamento. Mesmo assim, para as amostras de 200X, o modelo teve bom desempenho para ambas as classes.

5.2.2 Com aumento de dados e sem *Oversampling*

Com o objetivo de aumentar a capacidade de generalização do modelo, este grupo de experimentos é responsável por aplicar a técnica de aumento de dados para verificar como o modelo *ensemble* proposto se comporta com variações adicionadas ao conjunto de imagens. Além disso, busca-se observar novamente a influência do fator de ampliação nas imagens da base de dados utilizada.

A Tabela 5.3 mostra que, para o fator de ampliação de 40X, a rede *ResNet50* obteve melhor resultado, apresentando 77.27% de acurácia e 84.91% de *F1 Score*. Para o fator de

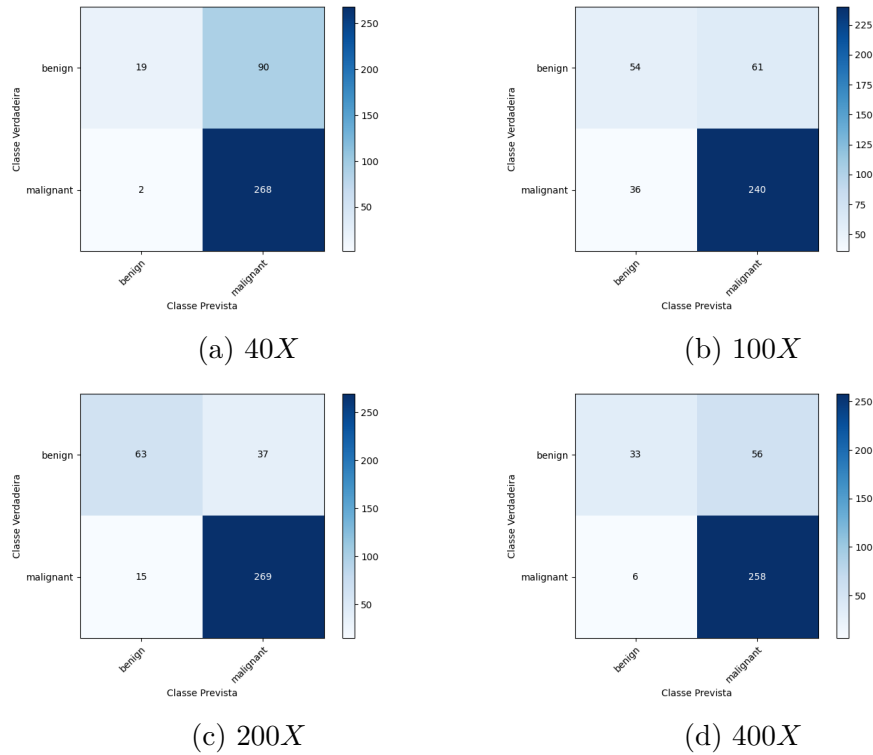


Figura 5.1: Matriz de confusão do modelo ensemble sem aumento de dados e sem *Over-sampling*.

ampliação de 100X, o modelo *Random Forest* atingiu melhor desempenho, com 78.91% de acurácia e 85.76% de *F1 Score*. O modelo *Random Forest* também atingiu melhor desempenho para o fator de ampliação de 200X, com acurácia de 90.89% e *F1 Score* de 93.65%. Por fim, para o fator de ampliação de 400X, o melhor desempenho foi da rede CNN simples, com acurácia de 85.79% e *F1 Score* de 90.74%.

Além disso, em comparação com os dados exibidos na Tabela 5.2, é possível visualizar que a utilização de aumento de dados foi benéfica para os classificadores base de uma maneira geral. No entanto, isso não se refletiu em melhores resultados para o modelo *ensemble* proposto em que os resultados obtidos foram bem próximos.

A Figura 5.2 representa a matriz de confusão do modelo *ensemble* proposto para as ampliações de 40X, 100X, 200X e 400X. Nessa figura, ainda é possível perceber a grande capacidade do modelo de classificar corretamente as amostras malignas e de forma errônea as amostras benignas, resultante do desbalanceamento da base de dados. No entanto, é possível notar uma ligeira melhora na classificação de amostras benignas, sobretudo para as ampliações de 200X e 400X.

Tabela 5.3: Métricas obtidas dos classificadores base e do modelo *ensemble* proposto com aumento de dados e sem *Oversampling*.

Modelo	Ampliação	Acurácia	Precisão	Revocação	F1 Score
<i>CNN</i>	40X	0.6761	0.7623	0.7984	0.7799
	100X	0.7813	0.8852	0.7941	0.8372
	200X	0.8880	0.9057	0.9472	0.9260
	400X	0.8579	0.8877	0.9280	0.9074
<i>ResNet50</i>	40X	0.7727	0.8123	0.8893	0.8491
	100X	0.7500	0.8014	0.8603	0.8298
	200X	0.7969	0.8679	0.8556	0.8617
	400X	0.7983	0.8784	0.8485	0.8632
<i>Random Forest</i>	40X	0.7216	0.8444	0.7510	0.7950
	100X	0.7891	0.8215	0.8971	0.8576
	200X	0.9089	0.9663	0.9085	0.9365
	400X	0.8381	0.8848	0.9015	0.8931
<i>Ensemble Proposto</i>	40X	0.7415	0.8045	0.8458	0.8247
	100X	0.7813	0.8013	0.9191	0.8562
	200X	0.8698	0.8774	0.9577	0.9158
	400X	0.8381	0.8581	0.9394	0.8969

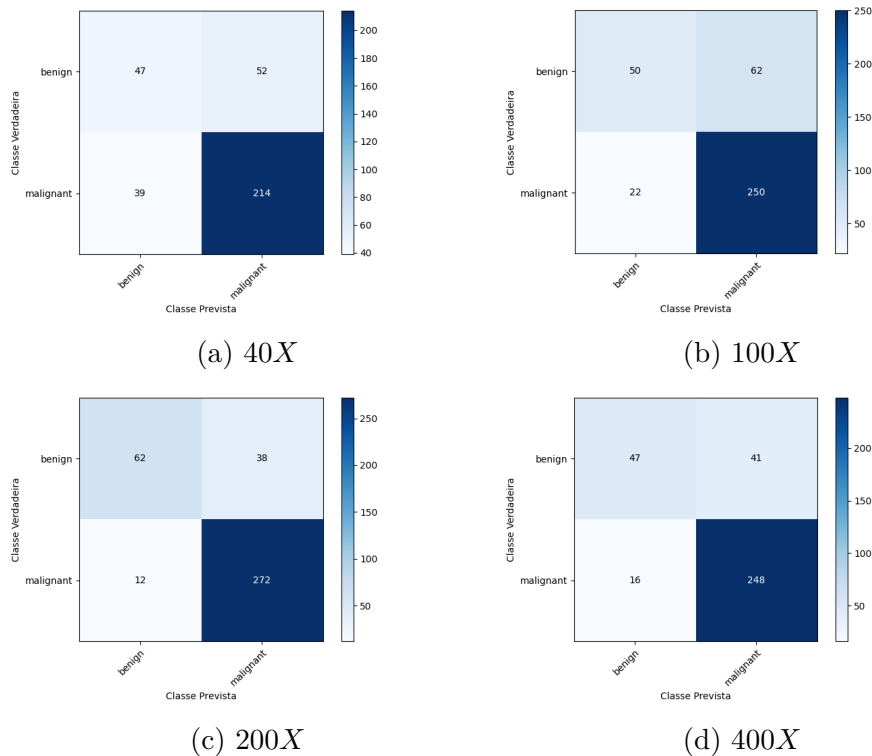


Figura 5.2: Matriz de confusão do modelo ensemble com aumento de dados e sem *Oversampling*.

5.2.3 Com *Oversampling*

Este grupo de experimentos visa aplicar a técnica SMOTE para corrigir o desbalanceamento do conjunto de dados e melhorar o desempenho geral do modelo. Assim, objetiva-se verificar como o modelo *ensemble* proposto se comporta com as classes balanceadas e a influência do fator de ampliação. Como observado nas Subseções 5.2.1 e 5.2.2, utilizar aumento de dados trouxe ganhos para os resultados do modelo *ensemble*. Desse modo, os experimentos conduzidos nessa seção utilizam a estratégia de *Data Augmentation*.

Como mostra a Tabela 5.4, é possível verificar que, para o fator de ampliação de 40X, o modelo *ensemble* atingiu o melhor resultado, superando os modelos base utilizados em sua construção, com acurácia de 89.84% e *F1 Score* de 90.33%. Para o fator de ampliação de 100X, o melhor desempenho foi do modelo *Random Forest*, que atingiu 87.13% de acurácia e 85.71% de *F1 Score*. A rede CNN simples atingiu 93.20% de acurácia e 93.26% de *F1 Score* para o fator de ampliação de 200X, sendo o melhor desempenho para esse fator de ampliação. Por fim, para o fator de ampliação de 400X, o modelo *Random Forest* atingiu melhor desempenho, com acurácia de 90.63% e *F1 Score* de 89.79%.

Tabela 5.4: Métricas obtidas dos classificadores base e do modelo *ensemble* proposto com aumento de dados e com *Oversampling*.

Modelo	Ampliação	Acurácia	Precisão	Revocação	<i>F1 Score</i>
<i>CNN</i>	40X	0.8945	0.8696	0.9302	0.8989
	100X	0.8364	0.8996	0.7574	0.8224
	200X	0.9320	0.9143	0.9517	0.9326
	400X	0.8809	0.8265	0.9605	0.8885
<i>ResNet50</i>	40X	0.7324	0.6613	0.9612	0.7836
	100X	0.7500	0.7073	0.8529	0.7733
	200X	0.8511	0.8032	0.9257	0.8601
	400X	0.7969	0.7633	0.8538	0.8060
<i>Random Forest</i>	40X	0.8125	0.9655	0.6512	0.7778
	100X	0.8713	0.9630	0.7721	0.8571
	200X	0.8585	0.9948	0.7175	0.8337
	400X	0.9063	0.9724	0.8340	0.8979
<i>Ensemble Proposto</i>	40X	0.8984	0.8678	0.9419	0.9033
	100X	0.7739	0.7110	0.9228	0.8032
	200X	0.8548	0.8045	0.9331	0.8640
	400X	0.8184	0.7326	0.9960	0.8442

Em relação aos experimentos das Subseções 5.2.1 e 5.2.2, é possível notar uma considerável melhora no desempenho de todos os modelos após a aplicação da técnica SMOTE, o que é devido a um melhor desempenho em classificar as amostras que eram originalmente minoritárias. Desse modo, ao utilizar *oversampling*, verifica-se um modelo com melhor capacidade de generalização.

A Figura 5.3 representa a matriz de confusão do modelo *ensemble* proposto para as ampliações de 40X, 100X, 200X e 400X. Nessa figura, fica ainda mais claro a grande capacidade do modelo de classificar corretamente as amostras malignas e benignas, em todos os fatores de ampliação possíveis, e apresentando excelentes resultados.

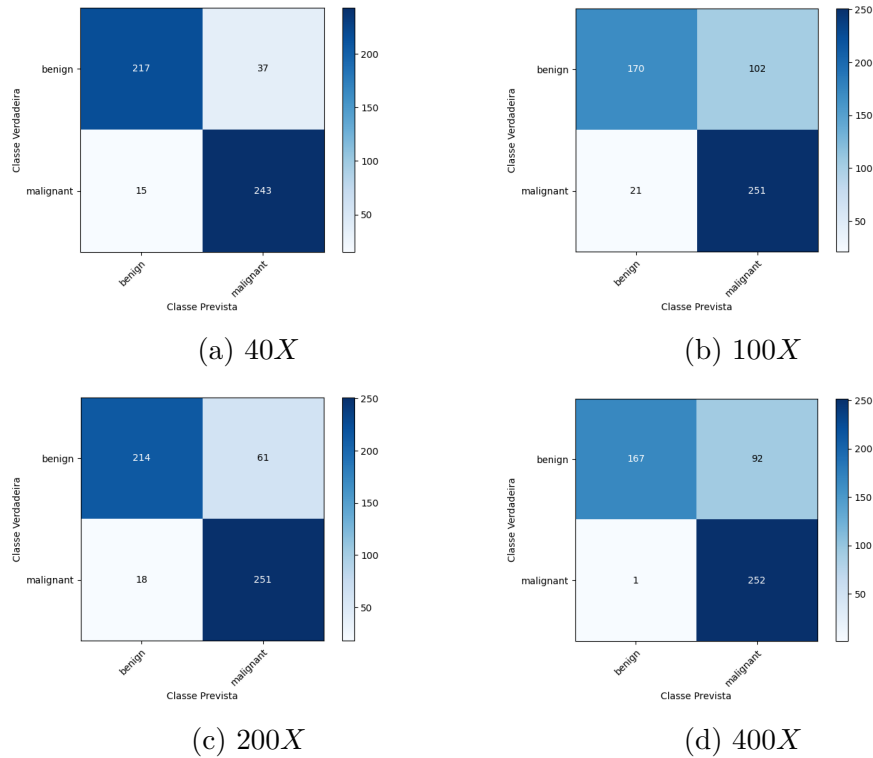


Figura 5.3: Matriz de confusão do modelo ensemble com aumento de dados e com *Over-sampling*.

5.3 Experimentos dos Modelos state-of-the-art

Os experimentos conduzidos nesta seção visam obter os resultados de modelos considerados *state-of-the-art* para o problema de classificação de imagens. Desse modo, experimentos com as redes *VGG-16*, *ResNet50* e *InceptionV3* para diferentes configurações do conjunto de dados foram conduzidos.

5.3.1 Sem aumento de dados e sem *Oversampling*

Como discutido na Subseção 4.3.2, os experimentos conduzidos para as redes *VGG-16*, *ResNet50* e *InceptionV3* objetivam verificar o desempenho das redes consideradas *state-of-the-art* para uma comparação com os resultados obtidos do modelo *ensemble* proposto.

Além disso, pretende-se verificar a influência do fator de ampliação no conjunto de imagens da base de dados para essas redes.

Nesse grupo de experimentos, as redes foram treinadas considerando um total de 100 épocas, utilizando a técnica de parada antecipada com paciência de 5 épocas. Isso foi realizado a fim de se evitar o sobreajuste dos modelos. Os melhores hiperparâmetros obtidos para as redes *state-of-the-art* estão descritos na Tabela 5.5. A Tabela 5.6 mostra os resultados obtidos no conjunto de teste para os modelos *state-of-the-art*, para cada fator de ampliação. Dentro de cada modelo base, as melhores métricas foram destacadas em negrito.

Tabela 5.5: Tabela de melhores hiperparâmetros para os modelos *state-of-the-art*.

Modelo	Hiperparâmetro	Melhores valores
<i>ResNet50</i>	Taxa de aprendizado	1e-3
	Número de neurônios	64
	Algoritmo de otimização	<i>Adam</i>
<i>VGG-16</i>	Taxa de aprendizado	1e-4
	Número de neurônios	64
	Algoritmo de otimização	<i>Adam</i>
<i>InceptionV3</i>	Taxa de aprendizado	1e-3
	Número de neurônios	128
	Algoritmo de otimização	<i>Adamax</i>

Tabela 5.6: Métricas obtidas dos modelos *state-of-the-art* sem aumento de dados e sem *Oversampling*.

Modelo	Ampliação	Acurácia	Precisão	Revocação	<i>F1 Score</i>
<i>VGG-16</i>	40X	0.8153	0.8311	0.9296	0.8320
	100X	0.7698	0.8039	0.8913	0.8276
	200X	0.8438	0.8784	0.9155	0.8503
	400X	0.7592	0.8482	0.8258	0.8558
<i>ResNet50</i>	40X	0.7810	0.7791	0.9667	0.8320
	100X	0.7596	0.7898	0.8986	0.8276
	200X	0.8594	0.8966	0.9155	0.8503
	400X	0.7904	0.9025	0.8068	0.8558
<i>InceptionV3</i>	40X	0.7282	0.7951	0.8333	0.8320
	100X	0.7033	0.7367	0.9022	0.8276
	200X	0.4818	0.6923	0.5387	0.8503
	400X	0.7025	0.7934	0.8144	0.8558

Ao observar os dados da Tabela 5.6, verifica-se que as redes *VGG-16* e *ResNet50* atingiram resultados promissores. A rede *InceptionV3*, por outro lado, não apresentou um desempenho muito bom para a base de dados utilizada. Dos modelos apresentados,

a rede *ResNet50* obteve maior destaque, apresentando uma acurácia de 85.94% para o fator de ampliação de 200X.

As Figuras 5.4, 5.5 e 5.6 representam a matriz de confusão dos modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Nas Figuras 5.4, da rede *VGG-16*, e 5.5, da rede *ResNet50*, é possível verificar uma boa capacidade de classificação para a classe maligna do conjunto de dados e uma classificação ruim para a classe benigna. Isso se deve ao fato da base de dados estar desbalanceada. Para a ampliação de 200X, essas redes apresentam um bom desempenho. A Figura 5.6, da rede *InceptionV3*, por outro lado, apresenta uma classificação ruim para as classes benignas e boa para as classes malignas com exceção da ampliação de 400X, em que o modelo apresentou baixo desempenho para ambas as classes.

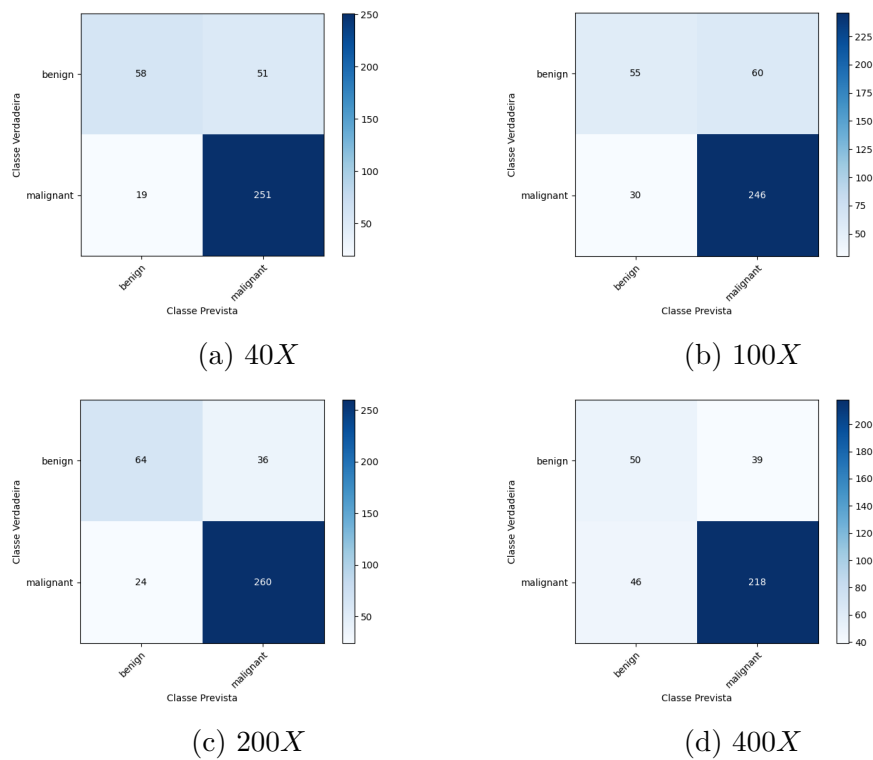


Figura 5.4: Matriz de confusão do modelo *VGG-16* sem aumento de dados e sem *Over-sampling*.

As Figuras 5.7, 5.8 e 5.9 representam as curvas de acurácia e perda para os modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Na Seção 5.2 as curvas de acurácia e perda para os modelos base constituintes do modelo *ensemble* não foram apresentadas para destacar o desempenho do modelo *ensemble* em si. Nos três modelos apresentados, as curvas de validação acompanharam proporcionalmente as curvas do conjunto de treinamento, no entanto, com uma boa distância entre elas. Isso sugere uma boa capacidade de generalização dos modelos sem apresentar *overfitting*. Um efeito observado nas curvas

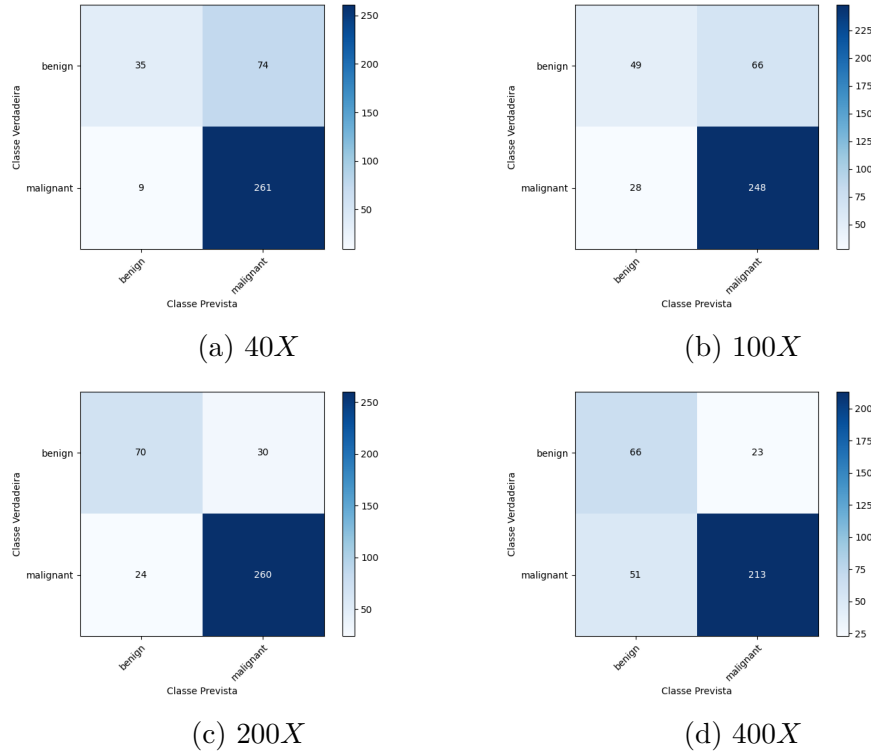


Figura 5.5: Matriz de confusão do modelo *ResNet50* sem aumento de dados e sem *Oversampling*.

do modelo *InceptionV3* é a grande instabilidade presentes nas curvas de acurácia e *loss*, apresentando diversos picos no gráfico. Isso indica que a rede utilizada está com dificuldade de convergir para o conjunto de dados fornecido ou uma necessidade de ajuste nos hiperparâmetros.

5.3.2 Com aumento de dados e sem *Oversampling*.

Com o objetivo de aumentar a capacidade de generalização do modelo, este grupo de experimentos é responsável por aplicar a técnica de aumento de dados para verificar como as redes *VGG-16*, *ResNet50* e *InceptionV3* se comportam com variações adicionadas ao conjunto de imagens. Além disso, busca-se observar novamente a influência do fator de ampliação da base de dados utilizada.

A Tabela 5.7 mostra os resultados obtidos nesse grupo de experimentos para os modelos *state-of-the-art*, para cada fator de ampliação. Ao observar os dados da tabela, verifica-se que novamente as redes *VGG-16* e *ResNet50* apresentaram promissores resultados, o que não ocorreu com a rede *InceptionV3*. A rede que se destacou para esse grupo de experimentos foi a rede *VGG-16*, que apresentou melhor acurácia, precisão, revocação e *F1 Score* para o conjunto de dados com ampliação de 400X.

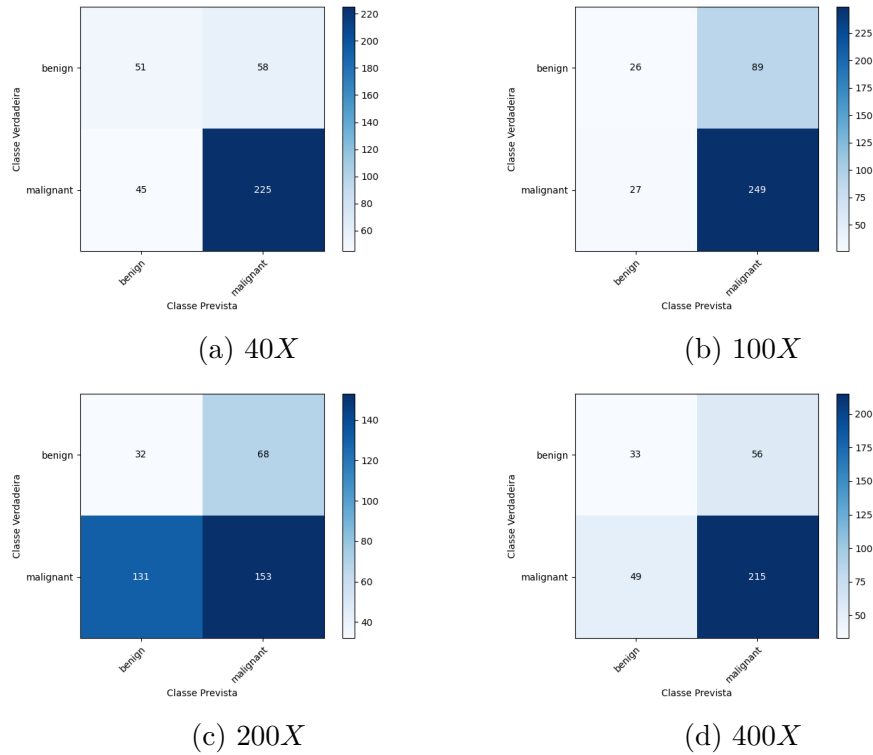


Figura 5.6: Matriz de confusão do modelo *InceptionV3* sem aumento de dados e sem *Oversampling*.

Em comparação com os experimentos da Subseção 5.3.1, verifica-se que o uso de aumento de dados melhora o desempenho das redes analisadas de forma geral, com exceção da rede *InceptionV3* que teve uma piora considerável em seu desempenho.

As Figuras 5.10, 5.11 e 5.12 representam a matriz de confusão dos modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Nas Figuras 5.10 e 5.11 é possível verificar uma boa capacidade de classificação para a classe maligna do conjunto de dados e uma melhora considerável para classificação das amostras benignas. A Figura 5.12, por outro lado, apresenta uma classificação ruim para as classes benignas e uma piora para as classes malignas, se comparado ao experimento da Subseção 5.3.1.

As Figuras 5.13, 5.14 e 5.15 representam as curvas de acurácia e perda para os modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Pode-se concluir que as curvas observadas apresentam comportamento próximo das curvas apresentadas nos experimentos da Subseção 5.3.1.

5.3.3 Com *Oversampling*

Este grupo de experimentos visa aplicar a técnica SMOTE para corrigir o desbalanceamento do conjunto de dados e melhorar o desempenho geral dos modelos. Essa técnica visa reduzir o número de classificações incorretas, sobretudo na classe minoritária do con-

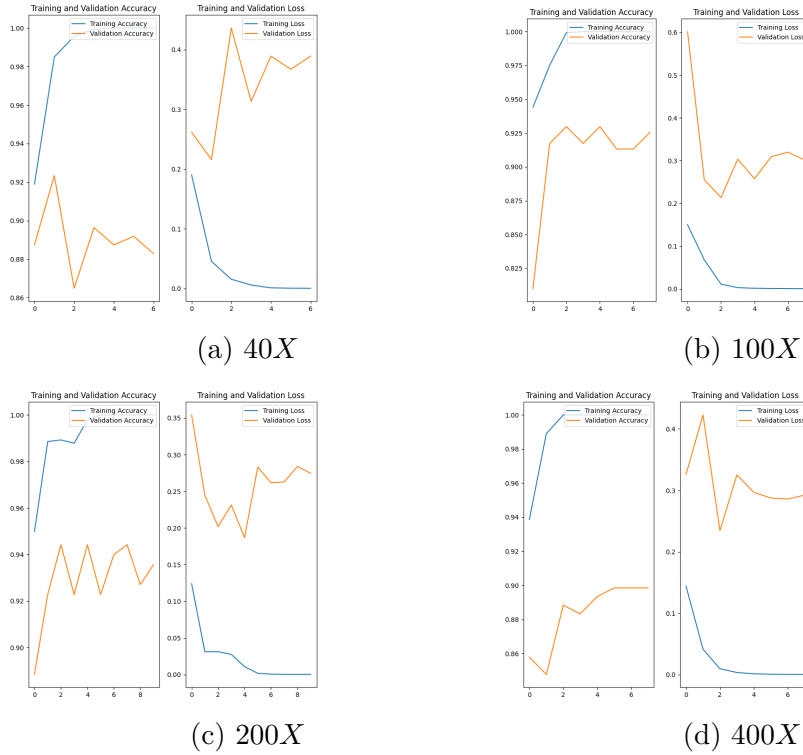


Figura 5.7: Curvas de acurácia e perda do modelo *VGG-16* sem aumento de dados e sem *Oversampling*.

junto de imagens. Desse modo, objetiva-se verificar como os modelos se comportam com classes balanceadas e a influência do fator de ampliação.

Os experimentos presentes nesta seção utilizam *oversampling* e aumento de dados para todos os modelos, com exceção da rede *InceptionV3*. Isso foi feito porquê, para essa rede, o desempenho foi pior ao utilizar aumento de dados, conforme visto nos experimentos das Subseções 5.3.1 e 5.3.2.

A Tabela 5.8 mostra os resultados obtidos nesse grupo de experimentos para os modelos *state-of-the-art*, para cada fator de ampliação. Ao observar os dados da tabela, verifica-se que novamente as redes *VGG-16* e *ResNet50* apresentaram excelentes resultados, o que não ocorreu com a rede *InceptionV3*. A rede que se destacou para esse grupo de experimentos foi a rede *ResNet50*, que apresentou melhor acurácia para o conjunto de dados com ampliação de 200X.

Em comparação com os experimentos das Subseções 5.3.1 e 5.3.2, verifica-se que o uso de *oversampling* piora o desempenho das redes analisadas de forma geral. Para a rede *InceptionV3*, no entanto, há uma considerável melhora nas métricas de acurácia e precisão.

As Figuras 5.16, 5.17 e 5.18 representam a matriz de confusão dos modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Na Figura 5.16, que representa o modelo

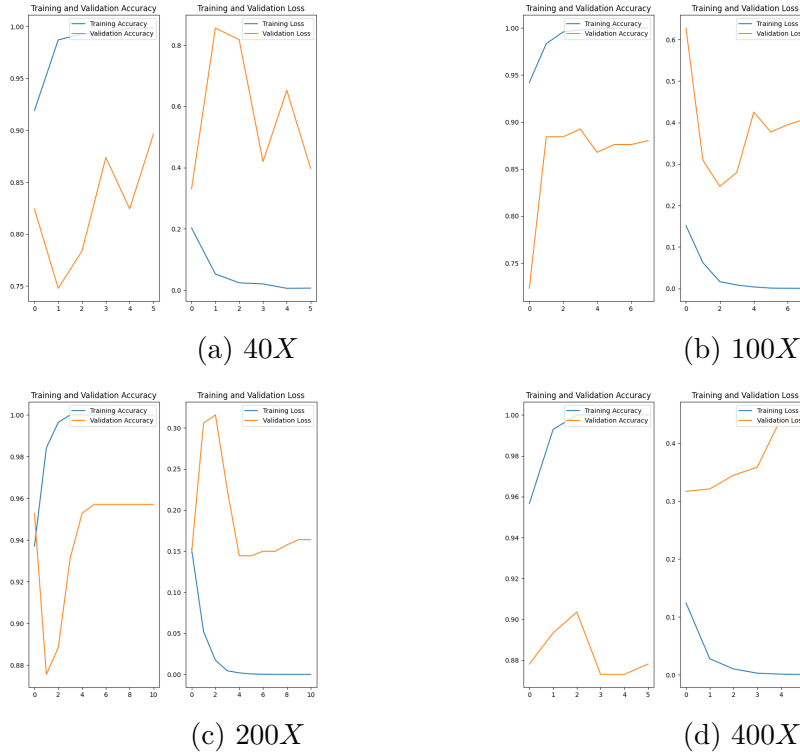


Figura 5.8: Curvas de acurácia e perda do modelo *ResNet50* sem aumento de dados e sem *Oversampling*.

VGG-16, verifica-se uma melhora na classificação de classes malignas e benignas para os fatores de ampliação de 100X e 200X, e uma piora nas outras ampliações. Por outro lado, nas Figuras 5.17 e 5.18, que representam os modelos *ResNet50* e *InceptionV3*, respectivamente, há uma melhora na classificação de classes malignas e benignas nas ampliações de 100X, 200X e 400X, enquanto apresenta uma piora considerável para 40X.

As Figuras 5.19, 5.20 e 5.21 representam as curvas de acurácia e perda para os modelos *VGG-16*, *ResNet50* e *InceptionV3*, respectivamente. Pode-se concluir que as curvas observadas apresentam comportamento próximo das curvas apresentadas nos experimentos das Subseções 5.3.1 e 5.3.2.

5.4 Síntese dos experimentos

Após a execução de todos os experimentos propostos, os melhores resultados coletados de cada modelo para todos os fatores de ampliação foram sintetizados na Tabela 5.9. Os valores em **negrito** destacam os melhores valores para cada fator de ampliação.

Analisando a Tabela 5.9, pode-se concluir que o modelo *ensemble* apresentou desempenho superior aos outros modelos *state-of-the-art* para todos os fatores de ampliação. Embora o modelo *ensemble* proposto tenha atingido uma acurácia um pouco inferior em

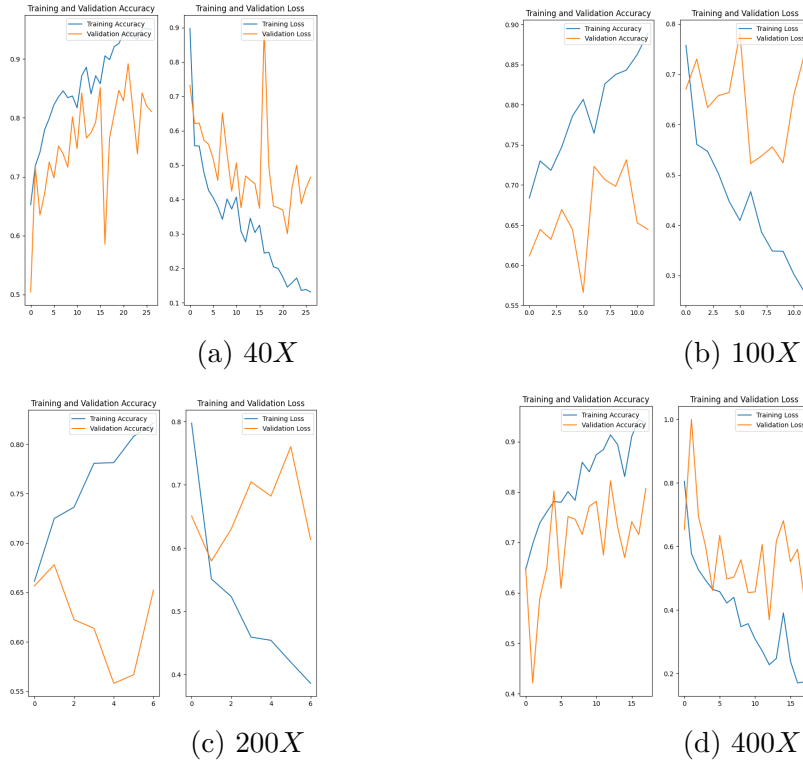


Figura 5.9: Curvas de acurácia e perda do modelo *InceptionV3* sem aumento de dados e sem *Oversampling*.

relação à rede *ResNet50* para o fator de ampliação de $100X$, a métrica de *F1 Score* foi bastante para o modelo *ensemble*. Foi para o fator de ampliação de $40X$ que o modelo *ensemble* atingiu melhor desempenho, com acurácia de 89.84%, precisão de 86.78%, revocação de 94.19% e *F1 Score* de 90.33%. Para o fator de ampliação de $100X$, o modelo *ensemble* atingiu acurácia de 78.13%, precisão de 80.13%, revocação de 91.91% e *F1 Score* de 85.62%, sendo inferior apenas ao obtido para a rede *ResNet50*, que atingiu 79.17% de acurácia, 80.38% de precisão, 93.38% de revocação e 82.93% de *F1 Score*.

Para o fator de ampliação de $200X$, o modelo *ensemble* atingiu 86.98% de acurácia, 87.74% de precisão, 95.77% de revocação e 91.58% de *F1 Score*. Por fim, o modelo *ensemble* proposto obteve 83.81% de acurácia, 85.81% de precisão, 93.94% de revocação e 89.69% de *F1 Score*.

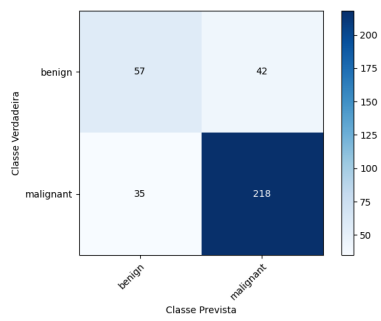
Mesmo com desempenho superior aos demais modelos *state-of-the-art* para todos os fatores de ampliação, o modelo *ensemble* proposto não atingiu desempenho superior ao modelo *Random Forest* utilizado em sua construção, como é possível observar nas Tabelas 5.2, 5.3 e 5.4. De modo geral, modelos *ensemble* possuem desempenho superior aos modelos base utilizados em sua construção e o motivo de não ter superado o desempenho do modelo base *Random Forest* pode ser um indício de que os outros modelos base utilizados podem não ter gerado benefícios significativos para o modelo *ensemble* como um todo.

Tabela 5.7: Métricas obtidas dos modelos *state-of-the-art* com aumento de dados e sem *Oversampling*.

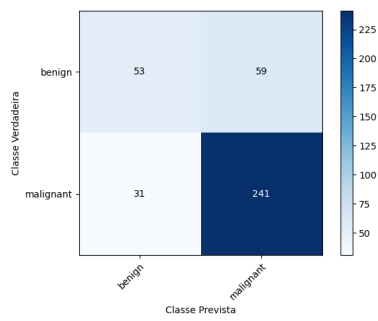
Modelo	Ampliação	Acurácia	Precisão	Revocação	<i>F1 Score</i>
<i>VGG-16</i>	40X	0.7812	0.8385	0.8617	0.8364
	100X	0.7656	0.8033	0.8860	0.8293
	200X	0.8151	0.8610	0.8944	0.8503
	400X	0.8295	0.8669	0.9129	0.8571
<i>ResNet50</i>	40X	0.7614	0.8007	0.8893	0.8364
	100X	0.7917	0.8038	0.9338	0.8293
	200X	0.8099	0.8728	0.8697	0.8503
	400X	0.8125	0.8561	0.9015	0.8571
<i>InceptionV3</i>	40X	0.6705	0.7686	0.7747	0.8364
	100X	0.6641	0.7241	0.8493	0.8293
	200X	0.7318	0.8390	0.7887	0.8503
	400X	0.7216	0.7804	0.8750	0.8571

Tabela 5.8: Métricas obtidas dos modelos *state-of-the-art* com *Oversampling*.

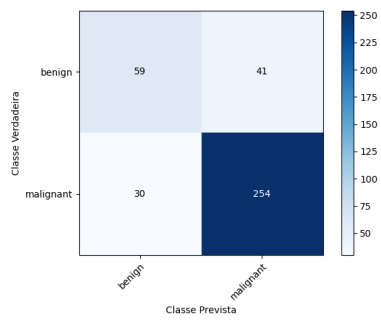
Modelo	Ampliação	Acurácia	Precisão	Revocação	<i>F1 Score</i>
<i>VGG-16</i>	40X	0.7461	0.6730	0.9651	0.6719
	100X	0.7721	0.7342	0.8529	0.6667
	200X	0.7958	0.7561	0.8732	0.6667
	400X	0.6477	0.5933	0.9394	0.6667
<i>ResNet50</i>	40X	0.7441	0.6693	0.9729	0.6701
	100X	0.7702	0.7442	0.8235	0.6667
	200X	0.8309	0.8062	0.8662	0.6617
	400X	0.8086	0.7719	0.8696	0.6614
<i>InceptionV3</i>	40X	0.7222	0.6786	0.8444	0.6667
	100X	0.6286	0.6212	0.6594	0.6667
	200X	0.7359	0.7043	0.8134	0.6667
	400X	0.6856	0.6713	0.7272	0.6667



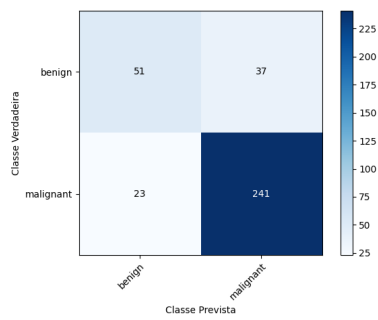
(a) 40X



(b) 100X

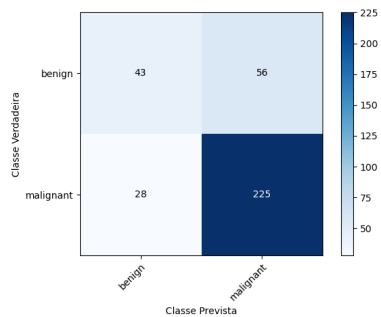


(c) 200X

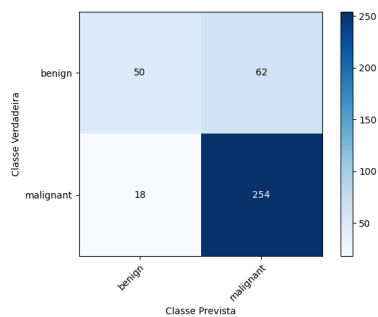


(d) 400X

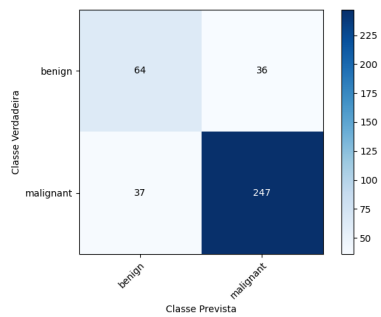
Figura 5.10: Matriz de confusão do modelo *VGG-16* com aumento de dados e sem *Over-sampling*.



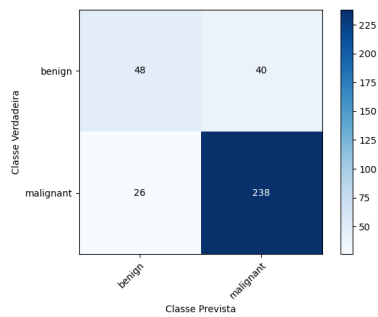
(a) 40X



(b) 100X



(c) 200X



(d) 400X

Figura 5.11: Matriz de confusão do modelo *ResNet50* com aumento de dados e sem *Oversampling*.

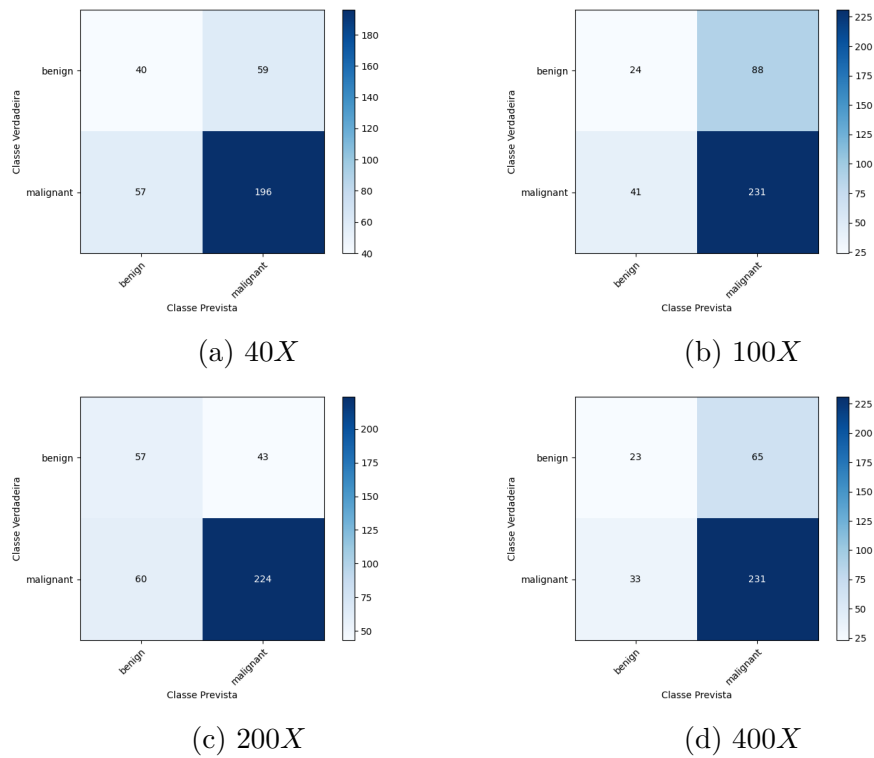


Figura 5.12: Matriz de confusão do modelo *InceptionV3* com aumento de dados e sem *Oversampling*.

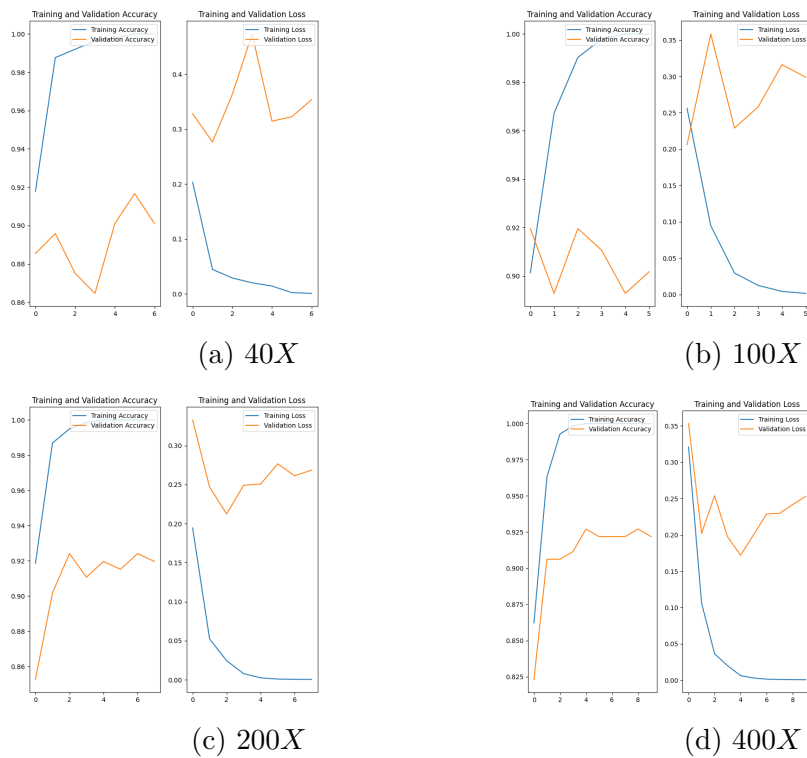


Figura 5.13: Curvas de acurácia e perda do modelo *VGG-16* com aumento de dados e sem *Oversampling*.

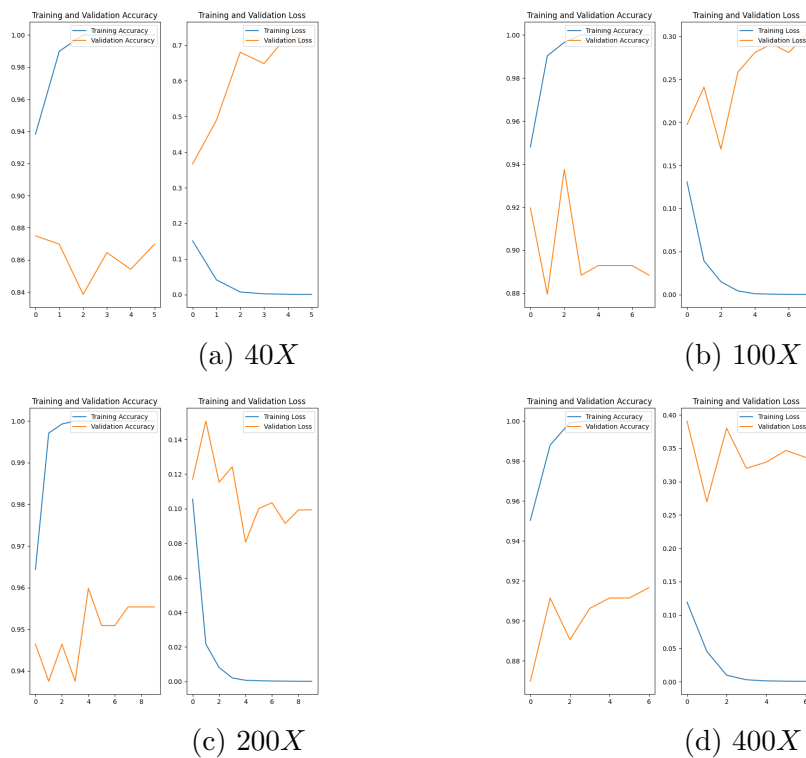


Figura 5.14: Curvas de acurácia e perda do modelo *ResNet50* com aumento de dados e sem *Oversampling*.

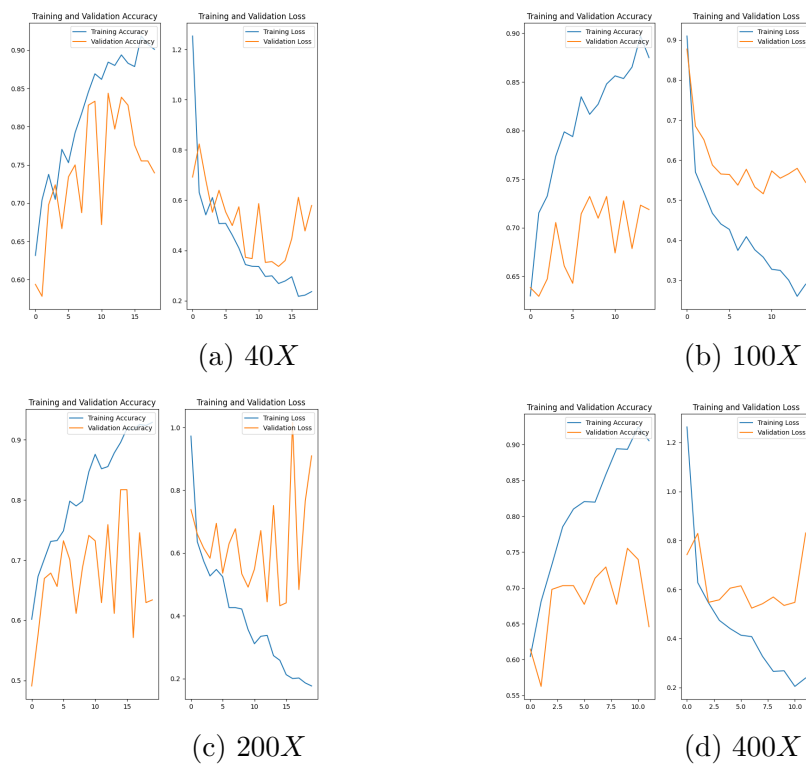


Figura 5.15: Curvas de acurácia e perda do modelo *InceptionV3* com aumento de dados e sem *Oversampling*.

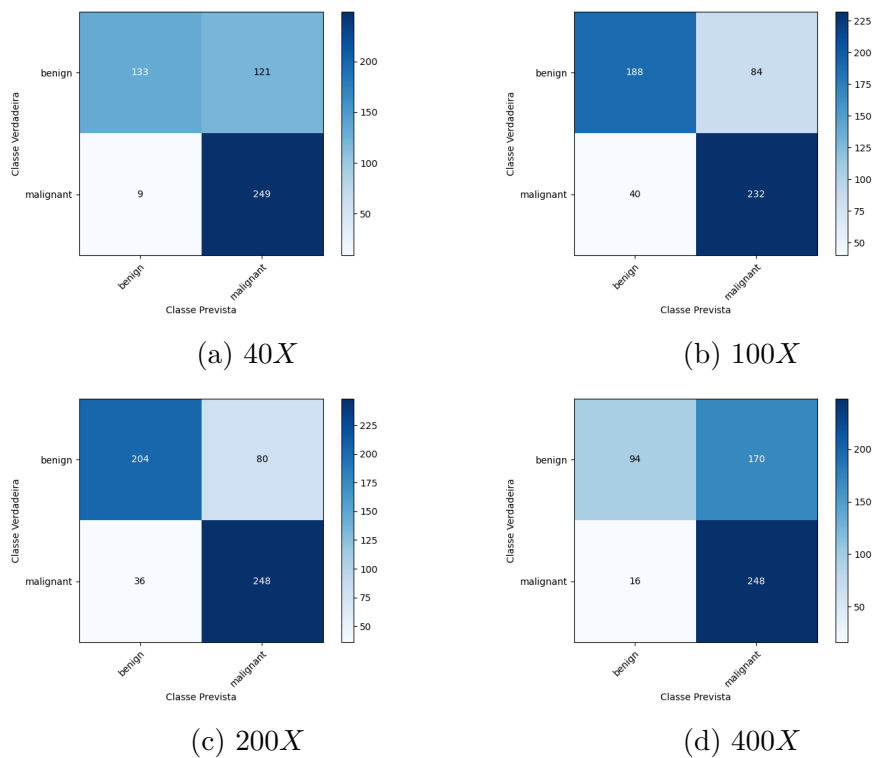


Figura 5.16: Matriz de confusão do modelo *VGG-16* com aumento de dados e com *Oversampling*.

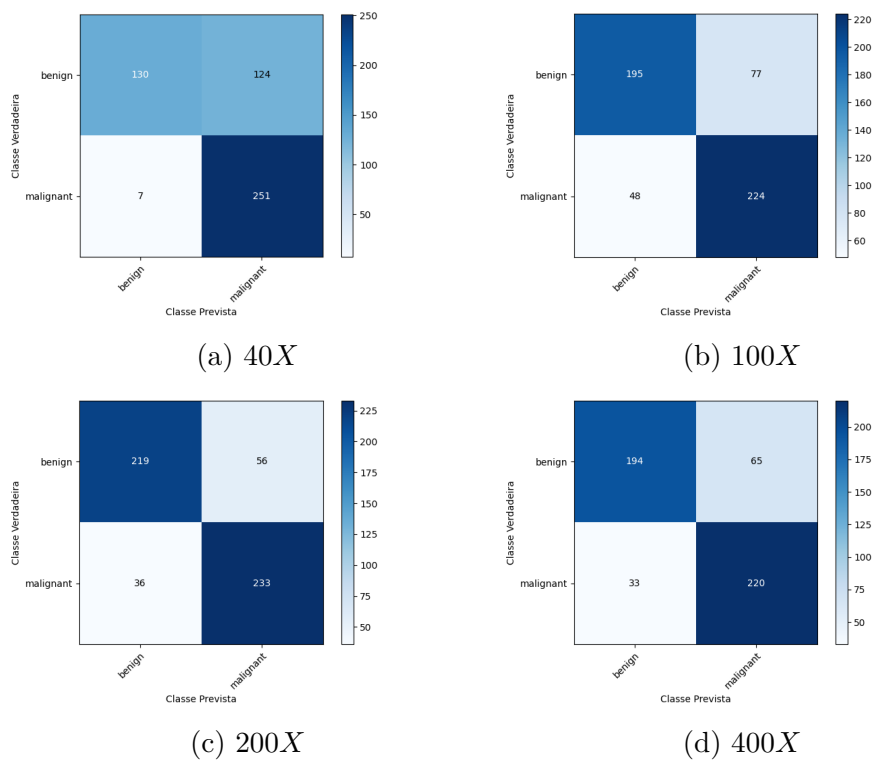


Figura 5.17: Matriz de confusão do modelo *ResNet50* com aumento de dados e com *Oversampling*.

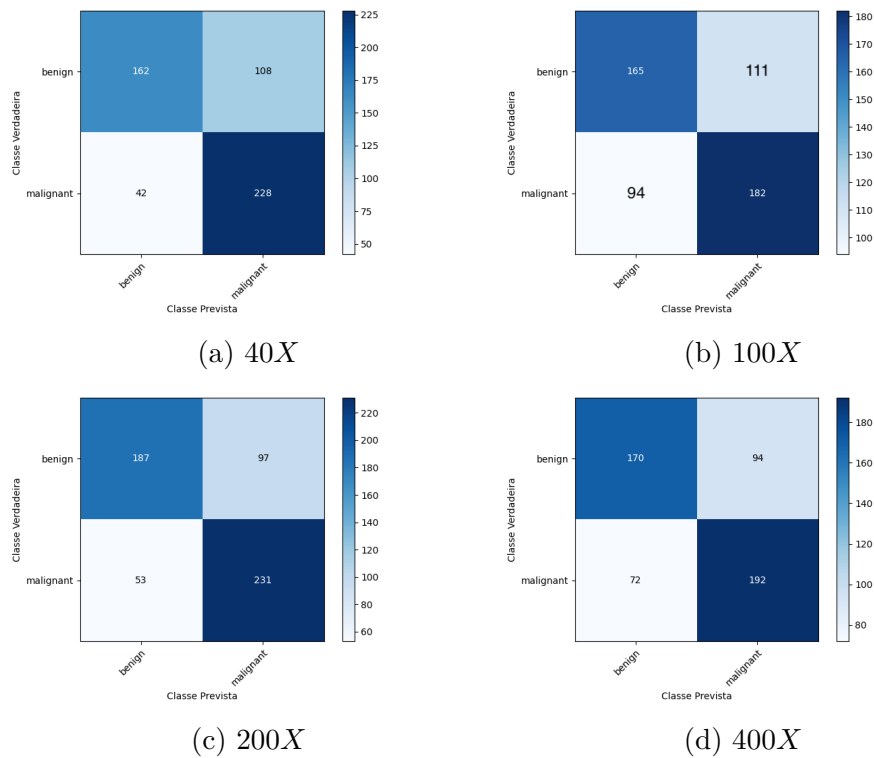


Figura 5.18: Matriz de confusão do modelo *InceptionV3* sem aumento de dados e com *Oversampling*.

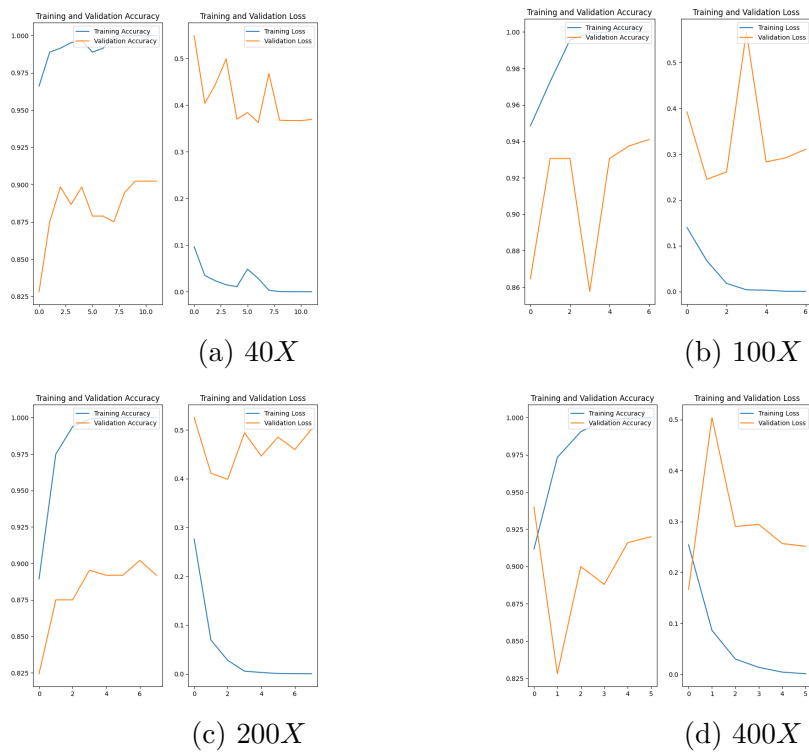


Figura 5.19: Curvas de acurácia e perda do modelo *VGG-16* com aumento de dados e com *Oversampling*.

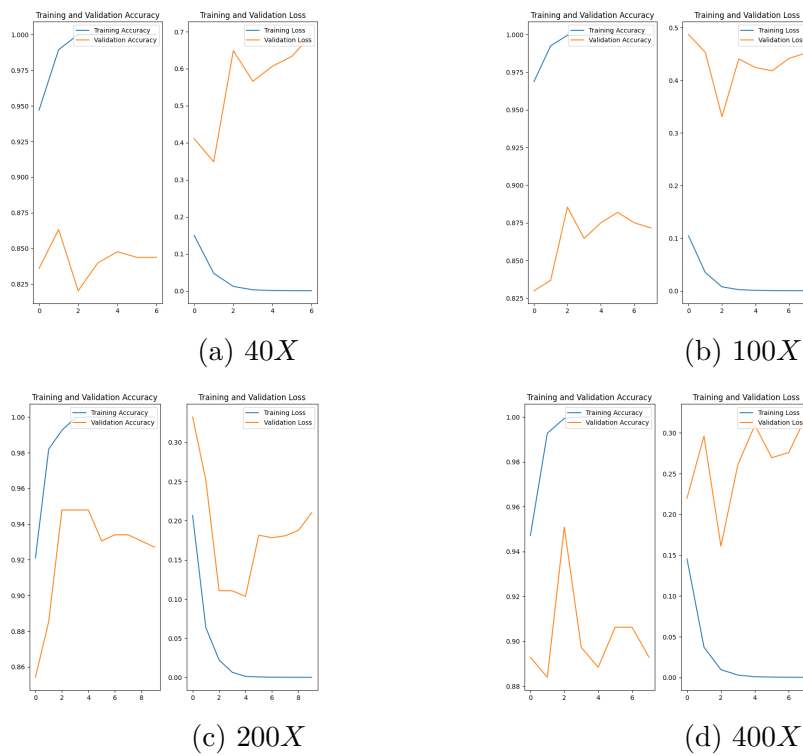


Figura 5.20: Curvas de acurácia e perda do modelo *ResNet50* com aumento de dados e com *Oversampling*.

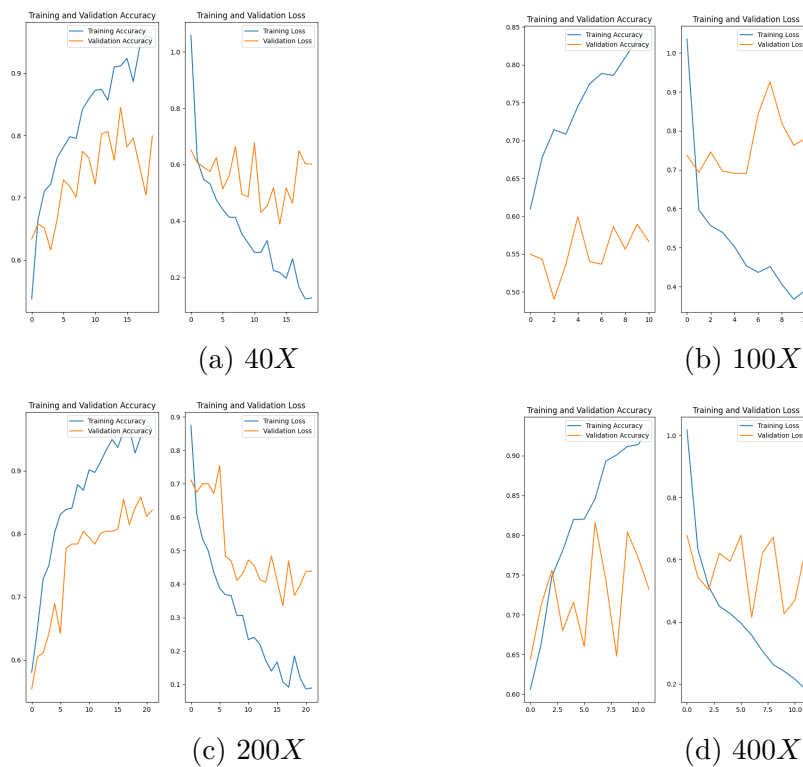


Figura 5.21: Curvas de acurácia e perda do modelo *InceptionV3* sem aumento de dados e com *Oversampling*.

Tabela 5.9: Resumo dos melhores resultados obtidos nos experimentos.

Modelo	Ampliação	Acurácia	Precisão	Revocação	<i>F1 Score</i>
<i>VGG-16</i>	40X	0.8153	0.8311	0.9296	0.8320
	100X	0.7698	0.8039	0.8913	0.8276
	200X	0.8438	0.8784	0.9155	0.8503
	400X	0.8295	0.8669	0.9129	0.8571
<i>ResNet50</i>	40X	0.7810	0.7791	0.9667	0.8320
	100X	0.7917	0.8038	0.9338	0.8293
	200X	0.8594	0.8966	0.9155	0.8503
	400X	0.8125	0.8561	0.9015	0.8571
<i>InceptionV3</i>	40X	0.7282	0.7951	0.8333	0.8320
	100X	0.7033	0.7367	0.9022	0.8276
	200X	0.7318	0.8390	0.7887	0.8503
	400X	0.7216	0.7804	0.8750	0.8571
<i>Ensemble</i>	40X	0.8984	0.8678	0.9419	0.9033
	100X	0.7813	0.8013	0.9191	0.8562
	200X	0.8698	0.8774	0.9577	0.9158
	400X	0.8381	0.8581	0.9394	0.8969

Capítulo 6

Conclusão e Trabalhos Futuros

Com o objetivo de realizar uma classificação binária para a tarefa de classificação de tumores com base em imagens de lâminas de tecido mamário, este trabalho propõe um modelo *ensemble* utilizando a estratégia *stacking*. Esse modelo é composto por algoritmos de aprendizagem de máquina tradicionais e modelos de CNN. Por fim, esse modelo proposto foi comparado com outras redes CNN existentes.

O primeiro grupo de experimentos descritos no Capítulo 4 busca avaliar o modelo *ensemble* proposto em três cenários distintos: sem aumento de dados e sem *oversampling*, com aumento de dados e sem *oversampling* e com *oversampling*, utilizando aumento de dados ou não a depender do melhor resultado obtido. Dentro de cada um destes cenários, os fatores de ampliação de 40X, 100X, 200X e 400X da base de dados também foram comparados. O melhor resultado obtido para o modelo *ensemble* foi para o fator de ampliação de 40X, atingindo acurácia de 89.84%, precisão de 86.78%, revocação de 94.19% e *F1 Score* de 90.33%, utilizando aumento de dados e *oversampling*.

No segundo grupo de experimentos realizados neste trabalho, três redes distintas de redes CNN consideradas *state-of-the-art* foram avaliadas nos mesmos cenários distintos anteriormente. Das três redes utilizadas, o melhor resultado obtido foi por meio da rede *ResNet50* para ampliação de 200X sem aumento de dados e sem *oversampling*. Esse modelo atingiu uma acurácia de 85.94%, precisão de 89.66%, revocação de 91.55% e *F1 Score* de 85.03%. Utilizar aumento de dados melhorou o desempenho dos modelos no geral com exceção da rede *InceptionV3*.

Observando o modelo *ensemble* proposto em comparação com os outros modelos *state-of-the-art*, verifica-se que ele atingiu resultados promissores para esse problema, superando todos os outros modelos *state-of-the-art* utilizados para os fatores de ampliação de 40X, 200X e 400X. Utilizar um modelo *ensemble* apresenta algumas limitações: seu custo computacional é mais levado por requerer o treinamento e manutenção de vários modelos e também depende da diversidade dos modelos base. Se os modelos base forem muito

semelhantes, o modelo *ensemble* pode não apresentar melhorias significativas. Os resultados apresentados neste trabalho para as redes utilizadas não são comparáveis diretamente aos trabalhos citados no Capítulo 3 que utilizam a mesma base de dados, visto que foi necessário realizar uma operação manual na base de dados de modo a não misturar dados de pacientes entre os conjuntos de treinamento, validação e teste. Esta operação manual não foi realizada nos trabalhos anteriores mencionados.

As imagens histopatológicas obtidas por biópsia, em comparação com outros tipos de imagens úteis para o diagnóstico de câncer de mama, como as obtidas por mamografia ou ultrassonografia, apresentam um custo de obtenção mais alto. Além disso, há um possível viés no diagnóstico com base em imagens histopatológicas, visto que os pacientes que passaram pelo procedimento de biópsia apresentam grandes chances de possuírem tumores malignos.

Para pesquisas e trabalhos futuros, uma abordagem distinta seria considerar bases de dados diferentes do que foi utilizado neste trabalho, para que se tenha uma maior quantidade de amostras. Outra abordagem seria melhorar o modelo *ensemble* apresentado, testando o desempenho com outros meta-classificadores e modelos-base a fim de se atingir o melhor desempenho possível. O algoritmo de K-Nearest Neighbors (KNNs) seria uma possível sugestão para o meta-classificador no lugar do algoritmo de regressor linear. Por fim, trabalhos futuros poderão ser realizados considerando múltiplas classes de tumores, como as classes Adenose, Fibroadenoma, Tumor Filoide, Adenoma Tubular, Carcinoma, Carcinoma Lobular, Carcinoma Mucinoso e Carcinoma Papilar presentes na base de dados descrita na Seção 4.1. Os trabalhos de Oyelade *et al.* [40] e Joseph *et al.* [42] consideram múltiplas classes.

Referências

- [1] Bonaccorso, Giuseppe: *Machine learning algorithms*. Packt Publishing, Birmingham, England, 2023. ix, 4, 5, 6, 7, 12
- [2] Zhang, Aston, Zachary C Lipton, Mu Li e Alexander J Smola: *Dive into deep learning*. Cambridge University Press, Cambridge, England, 2023. ix, 8, 9
- [3] Hassan, Muneeb ul: *VGG16 - Convolutional Network for Classification and Detection*. <https://neurohive.io/en/popular-networks/vgg16/>. Acesso em: 18-10-2023. ix, 10
- [4] Hernandez, Sara, Rubén Pozo e Luis Gómez: *Deep Neural Networks for Driver Identification Using Accelerometer Signals from Smartphones*, páginas 206–220. maio 2019, ISBN 978-3-030-20481-5. ix, 10
- [5] Narein, T: *Inception V3 Model Architecture*. <https://iq.opengenus.org/inception-v3-model-architecture/>. Acesso em: 18-10-2023. ix, 11
- [6] *Como funciona o algoritmo Árvore de Decisão*. <://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>. ix, 12
- [7] Geron, Aurelien: *Hands-on machine learning with scikit-learn and TensorFlow*. O'Reilly Media, Sebastopol, CA, USA, 2017. ix, 5, 12, 13, 14, 15, 19, 30
- [8] Spanhol, Fabio A., Luiz S. Oliveira, Caroline Petitjean e Laurent Heutte: *A dataset for breast cancer histopathological image classification*. IEEE Transactions on Biomedical Engineering, 63(7):1455–1462, 2016. ix, 26, 27, 28
- [9] International Agency for Research on Cancer: *Breast cancer awareness month 2021*. <https://www.iarc.who.int/featured-news/breast-cancer-awareness-month-2021>. Acesso em: 17-11-2023. 1, 5
- [10] Instituto Nacional do Câncer: *Inca estima 704 mil casos de câncer por ano no Brasil até 2025*. <https://www.gov.br/inca/pt-br/assuntos/noticias/2022/inca-estima-704-mil-casos-de-cancer-por-ano-no-brasil-ate-2025>. Acesso em: 17-11-2023. 1
- [11] Belsare, A D: *Histopathological image analysis using image processing techniques: An overview*. Signal Image Process., 3(4):23–36, 2012. 2

- [12] Chattopadhyay, Soham, Arijit Dey, Pawan Kumar Singh e Ram Sarkar: *Drda-net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images*. Computers in Biology and Medicine, 145:105437, 2022, ISSN 0010-4825. <https://www.sciencedirect.com/science/article/pii/S0010482522002293>. 2, 22, 25, 33
- [13] Sheeba, Adlin, P. Santhosh Kumar, M. Ramamoorthy e S. Sasikala: *Microscopic image analysis in breast cancer detection using ensemble deep learning architectures integrated with web of things*. Biomedical Signal Processing and Control, 79:104048, 2023, ISSN 1746-8094. <https://www.sciencedirect.com/science/article/pii/S1746809422005225>. 2, 22, 25
- [14] Chattopadhyay, Soham, Arijit Dey, Pawan Kumar Singh, Diego Oliva, Erik Cuevas e Ram Sarkar: *Mtrre-net: A deep learning model for detection of breast cancer from histopathological images*. Computers in Biology and Medicine, 150:106155, 2022, ISSN 0010-4825. <https://www.sciencedirect.com/science/article/pii/S0010482522008630>. 2, 21, 25
- [15] Zisserman, Andrew Karen Simonyan e: *Very deep convolutional networks for large-scale image recognition*. arXiv preprint, 2015. 2, 9, 28
- [16] Majumdar, Samriddha, Payel Pramanik e Ram Sarkar: *Gamma function based ensemble of cnn models for breast cancer detection in histopathology images*. Expert Systems with Applications, 213:119022, 2023, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417422020401>. 2, 21, 25
- [17] Karthik, R., R. Menaka e M.V. Siddharth: *Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks*. Biocybernetics and Biomedical Engineering, 42(3):963–976, 2022, ISSN 0208-5216. <https://www.sciencedirect.com/science/article/pii/S0208521622000705>. 2, 22, 25
- [18] Mitchell, T.: *Machine Learning*. McGraw-Hill, 1ª edição, 1997. 4
- [19] Franco, C. R.: *Inteligência Artificial*. Londrina: Editora e distribuidora educacional, 1ª edição, 2014. 4, 5
- [20] Rosenblatt, F.: *The perceptron: a probabilistic model for information storage and organization in the brain*. Psychol. Rev., 65(6):386–408, 1958. 5
- [21] Lecun, Y., L. Bottou, Y. Bengio e P. Haffner: *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11):2278–2324, 1998. 7, 30
- [22] Goodfellow, Ian, Yoshua Bengio e Aaron Courville: *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. Acesso em: 09-10-2023. 7, 16
- [23] *CS231n Convolutional Neural Networks for Visual Recognition*. <https://cs231n.github.io/convolutional-networks/>. Acesso em: 09-10-2023. 8
- [24] Rodrigues, D. A.: *Deep learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento automotivo*. Universidade Federal da Paraíba, 2018. 8

- [25] He, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun: *Deep residual learning for image recognition*. Em *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016. 9, 28, 30
- [26] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens e Zbigniew Wojna: *Rethinking the inception architecture for computer vision*. Em *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016. 11, 28
- [27] al., Katti Faceli et: *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC, 1ª edição, 2011. 11
- [28] Breiman, L., Friedman J. Olshen R. e Stone C.: *Classification and Regression Trees*. Wadsworth International Group, 1ª edição, 1984. 11
- [29] Quinlan, J. R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1ª edição, 1993. 11
- [30] Freund, Yoav e Robert E Schapire: *A decision-theoretic generalization of on-line learning and an application to boosting*. Em *Lecture Notes in Computer Science*, páginas 23–37. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995. 14
- [31] Friedman, Jerome H: *Greedy function approximation: a gradient boosting machine*. *Annals of statistics*, páginas 1189–1232, 2001. 14
- [32] Wolpert, David H: *Stacked generalization*. *Neural Netw.*, 5(2):241–259, 1992. 14
- [33] Friedman, Jerome H: *Random forests*. *Annals of statistics*, páginas 1189–1232, 2001. 15, 30
- [34] Sarkar, Dipanjan, Raghav Bali e Tamoghna Ghosh: *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing, Birmingham, England, 2018. 16
- [35] He, Haibo e E A Garcia: *Learning from imbalanced data*. *IEEE Trans. Knowl. Data Eng.*, 21(9):1263–1284, 2009. 17
- [36] Ying, Xue: *An overview of overfitting and its solutions*. *Journal of Physics: Conference Series*, 1168(2):022022, feb 2019. <https://dx.doi.org/10.1088/1742-6596/1168/2/022022>. 17
- [37] Chawla, N V, K W Bowyer, L O Hall e W P Kegelmeyer: *SMOTE: Synthetic minority over-sampling technique*. 2011. 17, 29
- [38] Ting, Kai Ming: *Confusion Matrix*, páginas 209–209. Springer US, Boston, MA, 2010, ISBN 978-0-387-30164-8. https://doi.org/10.1007/978-0-387-30164-8_157. 18
- [39] Bhausahab, Deshmukh Pramod e Kanchan Lata Kashyap: *Shuffled shepherd deer hunting optimization based deep neural network for breast cancer classification using breast histopathology images*. *Biomedical Signal Processing and Control*, 83:104570, 2023, ISSN 1746-8094. <https://www.sciencedirect.com/science/article/pii/S1746809423000034>. 21, 25

- [40] Oyelade, Olaide N., Absalom E. Ezugwu, Hein S. Venter, Seyedali Mirjalili e Amir H. Gandomi: *Abnormality classification and localization using dual-branch whole-region-based cnn model with histopathological images*. *Computers in Biology and Medicine*, 149:105943, 2022, ISSN 0010-4825. <https://www.sciencedirect.com/science/article/pii/S0010482522006783>. 22, 25, 58
- [41] Yu, Dianzhi, Jianwu Lin, Tengbao Cao, Yang Chen, Mingfei Li e Xin Zhang: *Secs: An effective cnn joint construction strategy for breast cancer histopathological image classification*. *Journal of King Saud University - Computer and Information Sciences*, 35(2):810–820, 2023, ISSN 1319-1578. <https://www.sciencedirect.com/science/article/pii/S1319157823000253>. 23, 25
- [42] Ameh Joseph, Agaba, Mohammed Abdullahi, Sahalu Balarabe Junaidu, Hayatu Hassan Ibrahim e Haruna Chiroma: *Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer)*. *Intelligent Systems with Applications*, 14:200066, 2022, ISSN 2667-3053. <https://www.sciencedirect.com/science/article/pii/S2667305322000072>. 23, 25, 58
- [43] Maleki, Alireza, Mohammad Raahemi e Hamid Nasiri: *Breast cancer diagnosis from histopathology images using deep neural network and xgboost*. *Biomedical Signal Processing and Control*, 86:105152, 2023, ISSN 1746-8094. <https://www.sciencedirect.com/science/article/pii/S1746809423005852>. 23, 25
- [44] Deng, Jia, Wei Dong, Richard Socher, Li Jia Li, Kai Li e Li Fei-Fei: *Imagenet: A large-scale hierarchical image database*. Em *2009 IEEE conference on computer vision and pattern recognition*, páginas 248–255. Ieee, 2009. 28
- [45] Bergstra, James e Yoshua Bengio: *Random search for hyper-parameter optimization*. *J. Mach. Learn. Res.*, 13(null):281–305, feb 2012, ISSN 1532-4435. 32, 33